Flinders
UNIVERSITY
ADELAIDE • SOUTH AUSTRALIA

An Investigation into Indonesian Pre-Service Physics Teachers' Scientific Thinking and Conceptual Understanding of Physics

by

LINA AVIYANTI

Thesis
Submitted to Flinders University
for the degree of

Doctor of Philosophy (PhD)

College of Education, Psychology and Social Work

April 2020

DECLARATION

I certify that this thesis does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university; and that, to the best of my knowledge and belief it does not contain any material previously published or written by another person except where due reference is made in the text.

Lina Aviyanti

12 January 2020

ACKNOWLEDGEMENTS

Praise be to Allah for granting me the strength and capability to complete my PhD thesis. This thesis would not have been accomplished without the support and assistance of many people and institutions towards whom I would like to express my deepest gratitude and immeasurable appreciation.

First and foremost, I am highly indebted to the Directorate General of Higher Degrees (DGHE) – Ministry of Research, Technology, and Higher Education of the Republic of Indonesia for granting me financial support and the opportunity to pursue my doctoral study in Australia. My special thanks also go to my home institution, the Indonesia University of Education (UPI), for providing me with permission to pursue my doctorate degree abroad. I would also like to acknowledge the support of, and assistance provided by, the College of Education, Psychology and Social Work, and the International Student Support Unit (ISSU) at Flinders University.

I am extremely grateful to my principal supervisor, Dr. Carol Aldous, for her critical and insightful comments, her endless patience and commitment, and for tirelessly encouraging and guiding me throughout my candidature. I have never forgotten how she encouraged me by saying "Keep on keeping on. The end is in sight!" These words encouraged me to work harder to complete my PhD thesis. I would like to extend my sincere thanks to my co-supervisor, Dr. Penny van Deur, for her encouragement, her prompt and constructive feedback, and her tireless efforts in supervising me. I would also like to gratefully acknowledge my non-formal supervisors, Dr. Carl J. Wenning (Illinois State University) for his advice and invaluable feedback, as well as his endless commitment during the period of my study at Flinders University. My deepest gratitude also goes to Dr. I Gusti Ngurah Darmawan (The University of Adelaide) who devoted his time to giving me sustained guidance, thoughtful comments, and precious and outstanding contributions for improvements, especially in the quantitative part of the data analysis. His support and commitment have made this thesis exist. I also wish to acknowledge the contributions of my previous research supervisor, Professor Stella Vosniadou, during the initial stage of this research.

I am deeply thankful to the universities that provided me with permission to conduct this research, as well as the academic staff members who kindly assisted me during the data collection phase. I would also like to express my sincere appreciation to all the pre-service physics teachers who participated in the study. This study would not have been possible without their willingness and enthusiasm to take part in this project.

I would also like to acknowledge my professional editor, Dr. Robert Muller, who edited and proofread for my thesis without altering or improving the substantive content of the thesis.

Special thanks go to all my friends, especially from the Indonesian community in Adelaide for their support during my PhD journey at Flinders University.

Last, but certainly not least, my heartfelt gratitude goes to my beloved husband, Mochammad Arif Saiful Rohman for his patience, understanding, and encouragement, and also for his endless prayers and sacrifice during my study. To my lovely daughters, Shafira Az-Zahra, Rizqia Lailika, and Sarah Fauziah, I cannot thank you enough for your endless love and support. I am also grateful to my brothers, parents-in-law, and my extended family for their continued prayers and support at all times.

Finally, I dedicate this thesis to my father and my mother who have both left this world. May Allah the Almighty grant both of you His Jannah.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

AACTE	American Association of Colleges for Teacher Education
AGFI	Adjusted Goodness-of-Fit Index
AMOS	Analysis of Moment Structures
ANOVA	Analysis of Variance
BEMA	Brief Electricity and Magnetism Assessment
BME	Bayes Modal Estimation
C.R.	Critical Ratio
CA	Challenging Argumentation
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
CLASS	Colorado Learning Attitudes about Science Survey
COV	Control of Variables
CTT	Classical Test Theory
CWV	Conservation of Weight and Volume
EAP	Expected A-Posteriori
EBAPS	Epistemological Beliefs Assessment for Physical Science
EDI	Education Development Index
EFA	Exploratory Factor Analysis
EK	Evolving Knowledge
EM	Electricity and Magnetism
FCI	Force Concept Inventory
FKIP	Fakultas Keguruan dan Ilmu Pendidikan (Faculty of Teacher Training and Educational Studies)
GALT	Group Assessment of Logical Thinking Test
GFI	Goodness-of-Fit Index
HDI	Human Development Index
HDR	Hypothetical-Deductive Reasoning
ICT	Information and Communication Technology
IKIP	Institut Keguruan dan Ilmu Pendidikan (Teacher and Education Studies Institute)
IRT	Item Response Theory
LCTSR	Lawson's Classroom Test of Scientific Reasoning
LPTK	Lembaga Pendidikan Tenaga Keguruan (Teacher Preparation Programs or Teacher Training Institutions)
Logit	Log odds unit
MA	Madrasah Aliyah (Senior Secondary Schools)
MAK	Madrasah Aliyah Kejuruan (Islamic Vocational Senior Secondary Schools)
MDT	Mechanics Diagnostic Test
MECH	Mechanics
MI	Madrasah Ibtidaiyah (Islamic Primary Schools)
MI	Modification Indices
ML	Maximum Likelihood
MLE	Maximum Likelihood Estimation
MML	Marginal Maximum Likelihood

MMLE	Marginal Maximum Likelihood Estimation
MOEC	Ministry of Education and Culture
MORA	Ministry of Religious Affairs
MORTHE	Ministry of Research, Technology and Higher Education
MPEX	Maryland Physics Expectation
MSA	Making Scientific Argumentation
MTs	Madrasah Tsanawiyah (Islamic Junior Secondary Schools)
NKL	Nature of Knowing and Learning
OECD	Organization for Economic Cooperation and Development
MNSQ	Mean Square
PCM	Partial Credit Model
PCR	Probability and Correlation Reasoning
PCU	Physics Conceptual Understanding
PISA	Programme for International Student Assessment
PR	Proportional Reasoning
RLA	Real-Life Applicability
RMSEA	Root Mean Square Error of Approximation
RSM	Rating Scale Model
SAL	Source of Ability to Learn
SBMPTN	Seleksi Bersama Masuk Perguruan Tinggi Negeri (Selection of Joint Entrance State University)
SBREC	Social and Behavioural Research Ethics Committee
SD	Sekolah Dasar (Elementary Schools)
SEM	Structural Equation Modelling
SMA	Sekolah Menengah Atas (Senior Secondary Schools)
SMK	Sekolah Menengah Kejuruan (Vocational Senior Secondary Schools)
SMP	Sekolah Menengah Pertama (Junior Secondary Schools)
SNMPTN	Seleksi Nasional Masuk Perguruan Tinggi Negeri (National Selection of State Universities)
SPSS	Statistical Program for Social Sciences
SSK	Structure of Scientific Knowledge
STEM	Science, Technology, Engineering, and Mathematics
STKIP	Sekolah Tinggi Keguruan dan Ilmu Pendidikan (Higher Education for Teacher Training and Educational Studies)
TIMSS	Trends in International Mathematics and Science Study
TLI	Tucker-Lewis Index
TOLT	Test of Logical Thinking
UNDP	United Nations Development Program
UNESCO	United Nations Educational, Scientific, and Cultural Organization
VASS	Views About Science Survey
VIF	Variance Inflation Factor
WLE	Weighted Likelihood Estimates

ABSTRACT

Pre-service physics teachers in Indonesia exhibit relatively low levels of scientific thinking and understanding of physics concepts. The aspects of scientific thinking, made up of epistemological beliefs, argumentation, and scientific reasoning, are known to have an impact on pre-service physics teachers' ability to understand physics concepts. Demographic factors and factors related to physics teaching and learning practices are also believed to play a significant role in the development of pre-service physics teachers' scientific thinking and understanding of physics concepts. However, the extent to which these factors are interconnected, and how each factor influences the development of pre-service physics teachers' scientific thinking and understanding of physics concepts remains unclear.

Previous studies have shown that these aspects of scientific thinking, have not been examined in an integrated way but only through bivariate relationships. Consequently, using structural equation modelling (SEM) procedures, this study presents a plausible model that seeks to explain the complex set of relationships arising between each of these variables as well as several demographic factors concerning gender, year level, and university type. In addition, pre-service physics teachers' perceptions about the opportunities and barriers experienced relating to teaching and learning are also explored through interviews.

A cross-sectional mixed methods approach was used to triangulate quantitative and qualitative data sources in which a sequential explanatory design was employed. In the quantitative study, five surveys were completed by 706 Indonesian pre-service physics teachers from Year 1 to Year 4, coming from two private and two public universities, while the qualitative data were collected from face-to-face semi-structured interviews with 25 pre-service physics teachers.

The main findings provided strong evidence of the effect of pre-service physics teachers' scientific thinking (comprising epistemological beliefs, argumentation, and scientific reasoning) on enhancing their understanding of physics concepts. Furthermore, epistemological beliefs and argumentation were found to positively affect pre-service physics teachers' understanding of physics concepts mediated by their skills in scientific reasoning. The findings suggest that those with sophisticated epistemological beliefs were more skilled in argumentation and reasoning scientifically than those with naive epistemological beliefs. In addition, those with more highly developed skills in argumentation tended to be more skilled in scientific reasoning. Meanwhile, scientific reasoning was found to strongly and directly influence pre-service physics teachers' understanding of physics concepts. The higher the level of pre-service physics teachers' skills in scientific reasoning, the more likely they were to better understand physics concepts. In addition, the findings obtained from the interviews with the pre-service physics teachers indicate that the teaching methods implemented in class, and the type of examination questions used by instructors, are likely to be the most important factors influencing pre-service physics teachers' adoption of particular approaches to learning. In turn, this could have an impact on the extent to which their scientific thinking and conceptual understanding of physics is developed. The study provides meaningful contributions in terms of theoretical, methodological, and practical understanding for practitioners and policymakers, particularly in assisting their efforts to improve the quality of physics teaching and learning practices in the Indonesian context.

CHAPTER 1
INTRODUCTION

1.1 Overview

This introductory chapter aims to provide an overview of the research focus of this thesis. The chapter begins with the background and context of the study, before moving on to address the Indonesian education system and pre-service teachers' programs in higher education institutions in Indonesia. The aims of the study, the research questions, the significance of the study, and definitions of the key terms pertaining to the thesis are also presented in this chapter. Finally, the structure of the thesis will be outlined before concluding the chapter.

1.2 Background of the Study

Education plays an essential role in the development of any country and this is true of Indonesia as well. More specifically, education prepares the younger generation to face dynamic global competition as well as the technological revolution that is under way. As Faure (1972, p. 156, as cited in Keeves & Watanabe, 2013, p. 401) acknowledged, "The physical, intellectual, emotional and ethical integration of the individual into a complete man [*sic*] is a broad definition of the fundamental aim of education." This research is situated within the broad understanding of the fundamental importance of education in society. Providing outstanding education is not an easy task. It requires competent teachers, good infrastructures and materials, and no less important, the need for strong support from stakeholders. It is essential that government agencies, stakeholders, and practitioners collaborate to provide the best education for learners. The government also needs to develop strategic education policies that offer various pathways for improving the professionalism of teachers and other educators in order to meet the demands of the 21st century, which focus on high-level thinking skills as the primary outcome of education. According to the Partnership for 21st Century Skills (2007, as cited in Yue, 2019), the essential skills needed in the 21st century include creativity and innovation, critical thinking and problem-solving, communication and collaboration, information and media literacy, and ICT (Information and Communication Technology) literacy. Therefore, the government should encourage educators and teachers to teach these sets of skills. This, in turn, can help students to gain the basic competencies and skills required to tackle the complex challenges they will face in their lives in a globalised world.

To improve the quality of education in a country such as Indonesia, a comprehensive reformation of the education system is needed, particularly in relation to the quality of teachers. Teacher quality is an essential aspect of the education domain if the goals of education are to be realised, because teachers have a significant impact on their students' academic outcomes (Canales & Maldonado, 2018; Taştan et al., 2018). Teacher quality is also an important consideration because teachers have the authority to manage teaching and learning practices in the classroom and develop a conducive learning environment to foster effective learning in students. Hence, teachers need to constantly update their knowledge and develop their skills. Teachers and educators are important contributors to their students' learning when all educational resources are considered (Darling-Hammond, 2006; Taştan et al., 2018), and they are also the most important component in the educational system (Mugot & Sumbalan, 2019). In other words, teachers play an important role in developing academic content knowledge and implementing appropriate pedagogic strategies and technologies into their teaching practices in the classroom. Teachers also play an important role in motivating their students to be interested in the subject matter they are teaching, actively engaging them in learning activities, and helping them to solve complex problems aimed at improving student learning outcomes.

There is no denying that improving the quality of teachers is crucial in the education domain. Therefore, helping teachers to update their knowledge and develop their skills to meet the needs of 21st century students is urgently needed. In so doing, teachers would have the required competencies and skills to prepare quality learning experiences for their students. As teachers play a crucial role in improving the quality of education and students' academic outcomes, teacher education institutions need to provide high-quality preparation programs to develop pre-service teachers' knowledge and skills, as well as their dispositions, so that they become highly competent professional teachers in the future. As acknowledged by The American Association of Colleges for Teacher Education (AACTE) and the Partnership for 21st Century Skills, prospective teachers must be equipped with adequate knowledge and skills to successfully meet the demands of this century (Greenhill, 2010; Mugot & Sumbalan, 2019). Consequently, pre-service teachers' study programs in higher education institutions are responsible for producing high quality future teachers by equipping them with knowledge about various pedagogical approaches that focus on developing 21st century knowledge and skills, so that they can implement what they have learned in their professional careers. In turn, these prospective teachers are expected to be able to prepare their future students with the essential knowledge and skills needed to succeed in the 21st century.

As asserted by Greenhill (2010), teachers who have been trained through teacher preparation programs in higher education institutions have a positive effect on their students' learning. It is argued that the quality of student learning can be influenced by teaching and learning practices implemented in the classroom and the depth of their teacher's knowledge (Gurria, 2016). The way teachers teach their students could be affected by how they are taught during the course of their higher education studies. In other words, university teachers could be role models for their undergraduate students, because sometimes, what is implemented by these university teachers in the classroom is imitated by their undergraduate students. In addition, the responsibility of governments, teacher education institutions, and educators and other practitioners is immense in developing an effective educational system. Hence, they need to work together to facilitate and prepare prospective teachers with adequate content knowledge and thinking skills, as well as pedagogical knowledge (strategies of instruction) that are relevant to the demands of today's world. In turn, it is expected that pre-service teachers will have high competencies and be able to provide quality learning experiences to shape the potential future careers of their students, the next generation. As noted by Retnawati, Djidu, Apino, and Anazifa (2018), teacher competencies and the depth of teacher knowledge about learning content affects students in developing their knowledge and thinking skills.

Previous research has demonstrated that thinking skills are important contributors to improving students' academic performance and their everyday life success, and this has recently gained substantial attention in educational research (Dunbar & Fugelsang, 2005; Koerber, Mayer, Osterhaus, Schwippert, & Sodian, 2015; NRC, 1996). More specifically, scientific thinking skills have essential implications for science education over all levels of the education programs, from pre-kindergarten through to college (Zimmerman & Klahr, 2018, p. 2). In science education, teaching and learning practices in the classroom should teach and promote students' skills in scientific thinking to help them understand scientific phenomena occurring in the real world. Students with high-level thinking skills are more likely to be able to solve complex problems faced in real life. In addition, prior studies have acknowledged that scientific thinking plays an important role in promoting students' understanding of scientific concepts (Andayani, Hadisaputra, & Hasnawati, 2018; Ding, 2014c; Nieminen, Savinainen, & Viiri, 2012). Students with more sophisticated skills in scientific thinking are likely to have deep levels of conceptual understanding and high levels of achievement, particularly in the science domain (Cavallo, Rozman, Blickenstaff, & Walker, 2003). However, prior research has shown that the underperformance of students is one of the outcomes of ineffective teaching and learning (Dewey & Dykstra, 2008; Henderson, 2002). For instance, students' scientific thinking at both secondary school and university level has been characterised as inadequate or less than optimal (Bao, Cai, et al., 2009; Osborne, Erduran, & Simon, 2004; Putra, 2019). McGee and colleagues (2010) also identified that many teacher education programs do not equip prospective teachers with sufficient breadth of knowledge, which can negatively affect their skills in developing effective instructional practices in the classroom. Educational reforms may have called for an emphasis on developing thinking skills, rather than memorising knowledge or facts; however, it seems that there is a gap between educational policy made by stakeholders, and teaching practices taking place in the classroom. This should raise the alarm and challenge teacher preparation programs to pay more attention to, and highlight and address, the needs of prospective teachers so that they become more qualified and professional in the years to come. Thus, identifying pre-service teachers' content knowledge and scientific thinking is crucial to ensure that they are well prepared for teaching in schools. This is particularly true in the case of physics learning and teaching.

There is no denying that Indonesia is struggling to provide high-quality education for all Indonesian students, even though the government has implemented various improvements through the education reform process, one of which focuses on teacher training standards to improve the quality of Indonesian teachers (Dilas, Mackie, Huang, & Trines, 2019). More specifically, the quality of science teaching and learning has been a national issue (Hendayana, Asep, & Imansyah, 2010). In fact, the Indonesian government has paid considerable attention to teacher quality and has made great efforts to improve science teachers' quality and competence by improving teacher qualifications, teacher certification, and teacher professional development. These programs provide opportunities for teachers to extend their knowledge and practice their teaching skills in order to improve their quality. However, Indonesian students still indicate low achievement, especially in the field of science both in assessments at the national and international levels (Faisal & Martin, 2019). For instance, large-scale international studies such as TIMSS (Trends in International Mathematics and Science Study), have reported that Indonesian students were ranked at a low level among other participating countries. Likewise, another international assessment measuring student performance in mathematics, reading, and science literacy, known as PISA (Programme for International Student Assessment), also places Indonesian students at a low level of performance compared to other countries (Firman, 2016).

Since 1995, TIMSS has assessed the achievements of international students at the fourth and eighth grades in mathematics and science every four years (Martin, Mullis, Foy, & Hopper, 2016). More specifically, in 2011, TIMSS reported that eighth-grade Indonesian students in science achievement were ranked 40th out of 42 participating countries. This ranking was even below neighbouring countries, with Singapore being the highest scoring country in science achievement, while Thailand ranked 27th and Malaysia ranked 32nd. Indonesia, one of the lowest scoring countries in science achievement overall, was ranked below many other countries, and its performance in 2011 was also lower than the 2007 rankings. This decline was due to a decrease in achievement in all four content domains of the test, including physics (Martin, Mullis, Foy, & Stanco, 2012). In 2015, fourth-grade Indonesian students who participated in TIMSS were ranked 44th out of 47 participating countries in science achievement assessment (Martin et al., 2016). In addition, PISA is a large-scale assessment study of 15-year-old students in various countries around the world conducted by the Organization for Economic Cooperation and Development (OECD) every three years since 2000 (Thien, Razak, Keeves, & Darmawan, 2016). In 2015, PISA indicated that Indonesia was ranked 62nd out of 72 participating countries, below Thailand which ranked 54th, while Singapore outperformed all participating countries (Gurria, 2016). More specifically, PISA 2018 reported that Indonesian students' performance in science was ranked 70th out of 78 participating countries. This ranking was below neighbouring countries, with Singapore being the second-highest scoring country in science achievement after China, while Thailand ranked 53rd, Brunei Darussalam 50th, and Malaysia 48th (Schleicher, 2019).

These low scores indicate that Indonesian students struggled to answer the test questions given in both TIMSS and PISA. Indeed, Indonesian students’ performance on the TIMSS and PISA has remained far behind many other countries. The previous research has indicated that science learning in Indonesian schools emphasises science more as a product of knowledge, rather than as a process of scientific discovery. In addition, the existing science learning approaches encourage students to memorise knowledge or facts, and to focus on complex mathematical formulae instead of developing the ability to understand science concepts (Nurlatifah, Tukiran, & Erman, 2018). Hence, Indonesian students tended to use memorisation and to practice strategies compared to students from other countries, such as Sri Lanka and the Netherlands (Husnaini & Chen, 2019). These results suggest the importance of reforming the education system in an effort to improve the quality of education in Indonesia, which would lead to improving the quality of the students who would be more able to compete in an increasingly internationalised world. This could be achieved by improving the quality of pre- and in-service teachers in Indonesia.

Further to this, the *Education for All Global Monitoring Report* released by the United Nations Educational, Scientific, and Cultural Organization (UNESCO) reported that Indonesia's Education Development Index (EDI) rank was 69th out of 127 countries in 2011. This result was below neighbouring Brunei Darussalam (34th) and Malaysia (65th). The EDI values obtained were based on four factors, namely basic education participation, literacy levels in 15-year old students, gender equality participation, and the number of students staying until the fifth grade in school. In addition, the United Nations Development Program (UNDP) reported that Indonesia's HDI (Human Development Index) value for 2015 was ranked 113th out of 188 countries, below other ASEAN countries, such as Singapore (6th), Brunei Darussalam (30th), Malaysia (59th), and Thailand (87th). This data indicates that the quality of Indonesian students' academic performance is lower and behind that of neighbouring countries, which may reflect the poor quality of the education system in Indonesia. Certainly, the low quality of Indonesian education is affected by various factors such as a curriculum that often changes, lack of subject knowledge, and insufficient pedagogical skills of its teachers (Nugroho, Permanasari, & Firman, 2019; Rosser, 2018).

Previous studies have also indicated that the quality of Indonesian teachers has been relatively low (Chang et al., 2013; Fenanlampir, Batlolona, & Imelda, 2019). Factors contributing to the low quality of Indonesian teachers are associated with the low level of teachers' academic education, leaving many underqualified, and with limited content knowledge and underdeveloped skills (Chang et al., 2013). The low quality of teachers' mastery of content knowledge and pedagogical strategies, as well as their thinking skills, can also have an impact on the way they develop learning plans and implement teaching and learning practices in the classroom. In turn, this can affect the level of knowledge and the thinking skills of their students. As highlighted by Retnawati, Djidu, et al. (2018), teachers' knowledge and competence can contribute to improvements in student competencies and thinking skills.

In the Indonesian context, the prior research has demonstrated that students' conceptual understanding and thinking skills were relatively low and underdeveloped, especially in science education (Faisal & Martin, 2019; Husamah & Pantiwati, 2014; Irwanto, Eli, & Prodjosantoso, 2019; Putra, 2019; Rosdiana, Siahaan, & Rahman, 2019; Saputro, Sarwanto, Sukarmin, & Ratnasari, 2019). It has been argued that good thinking skills are needed when one engages in the process of constructing scientific knowledge (Suprapto, 2014). However, Rusmansyah and colleagues (2019) reported that pre-service science teachers' thinking skills were low and needed to be improved. As pre-service teachers are the teachers of tomorrow who have the responsibility of constructing and developing effective teaching and learning practices for their students, it is important that studies related to pre-service science teachers' knowledge and thinking skills are undertaken to improve the quality of science education in Indonesia.

Science education is more inquiry-based and closer to the work of real-life scientists (Breiner, Harkness, Johnson, & Koehler, 2012). More specifically, physics is one branch of the natural sciences that is closely related to various aspects of human life that are diverse and complex. Physics is also very closely related to the investigation of phenomena that occur in the universe, and it requires scientific thinking skills to discover, predict, and understand knowledge about the natural world so that valid conclusions can be drawn. Understanding the nature of scientific thinking has become an area of research development with strong underpinnings in cognitive science (Berland & McNeill, 2010; Dunbar & Fugelsang, 2005; NRC, 1996). Scientific thinking is important as a goal for science instruction which promotes science as inquiry and includes a range of activities through which students learn the scientific way of knowing and making sense of the world.

Further to this, physics is a subject which many students find challenging. Particularly in Indonesia, physics is considered as an abstract subject that is conceptually difficult to understand and boring to learn (Firdaus, Erwin, & Rosmiati, 2019). The difficulty students have with learning physics seems to be a common issue that is often found and examined by physics education researchers. Broadly speaking, students face several difficulties in understanding the basic principles of physics in topics such as mechanics, and electricity and magnetism, that can lead to misconceptions (Cahyaningrum & Hidayat, 2018; Retnawati, Arlinwibowo, Wulandari, & Pradani, 2018). It has been argued that students' conceptual understanding of physics is considered to be one of the primary goals of physics education that needs to be developed (Dervic, Glamocic, Gazibegovic-Busuladzic, & Mesic, 2018). The difficulty that students have in understanding physics concepts may be due to the poor quality of the teachers, inadequate facilities and resources, a lack of learning media and laboratory equipment, physics concepts that are considered abstract, and students' previous misconceptions about certain physics concepts (Sobremisana, 2017; Suciatmoko, Suparmi, & Sukarmin, 2018). Consequently, the reasons mentioned above would have a negative impact on students' ability to construct their knowledge and would increase the possibility of misconceptions about physics concepts.

Considering the need to enhance students' understanding of physics knowledge and thinking skills, further studies exploring various aspects that can influence the understanding of physics concepts and scientific thinking skills, as well as the relationships arising between these factors, are essential. The present study is aimed at investigating pre-service physics teachers' scientific thinking and their conceptual understanding of physics. More specifically, this study focuses on epistemological beliefs, argumentation, and scientific reasoning that are considered as aspects of scientific thinking. These aspects are believed to play an important role in promoting students' understanding of scientific concepts as well as for improving student achievement (Ding, 2014c; Franco et al., 2012; Osborne et al., 2016). Identifying students' epistemological beliefs, argumentation, and scientific reasoning is important to enable them to handle real-world tasks in their future careers, particularly those in the STEM (science, technology, engineering, and mathematics) areas as scientists or science educators (Bao, Cai, et al., 2009; Lawson, 2004).

Previous studies have examined these aspects of scientific thinking and their contributions to students' understanding of physics concepts only in a fragmented manner, through bivariate relationships that could be misleading when inferences are drawn. As noted by Nunkoo and Ramkissoon (2011, p. 1), "…bivariate statistical techniques are limited in examining relationships among different constructs simultaneously, leaving some interactions unexplained." This might cause a misinterpretation when trying to deeply understand and draw conclusions about the relationships between variables. Furthermore, no findings have been reported regarding the relationship between these aspects of scientific thinking and students' understanding of physics concepts in an integrated way. Hence, this study aims to fill this gap. Considering that there has been little research that focuses on aspects of scientific thinking (covering epistemological beliefs, argumentation, and scientific reasoning) of Indonesian pre-service teachers and its contribution to their conceptual understanding of physics, this research provides empirical evidence that contributes to the literature. Several factors related to teaching and learning practices covering the opportunities and barriers experienced by pre-service physics teachers in enhancing their scientific thinking and conceptual understanding of physics are explored in order to obtain a broader picture. Thus, the findings of this study provide broad and comprehensive insights for educators, researchers who are interested in conducting similar research, and stakeholders for the formulation of future policy in order to improve physics education in Indonesia. The next section describes the context of the study, providing an in-depth insight into Indonesian education.

1.3 Context of the Study

1.3.1 Brief overview of the Indonesian Education System

As noted previously, education plays a crucial role in improving the quality of human life and preparing people to face the challenges of their future careers in the modern world. The Law on National Education No. 20, the Year 2003 (*UU No. 20 Sistem Pendidikan Nasional, 2003*) emphasised that all Indonesian citizens have equal rights to acquire good quality education and improve their educational capacity in the process of life-long education (OECD/Asian Development Bank, 2015; MoNE, 2003). The operation of schools and higher education institutions are under the management and control of several ministries. At the central level, there are three ministries responsible for managing the education system in Indonesia, namely the Ministry of Education and Culture (MOEC), the Ministry of Religious Affairs (MORA), and the Ministry of Research, Technology and Higher Education (MORTHE). Referring to Presidential Regulation (Perpres*: Peraturan Presiden*) No. 72 the Year 2019, the MORTHE is no longer responsible for managing higher education in Indonesia as this responsibility has been given to the MOEC: (https://www.kemdikbud.go.id/main/blog/2019/10/perpres-nomor-72-tahun-2019-pendidikan-tinggi-kembali-di-bawah-naungan-kemendikbud). In addition, the national education system in Indonesia is classified into formal, non-formal, and informal education (OECD/Asian Development Bank, 2015).

Firstly, formal education comprises public and private schools, which are structured and tiered covering basic or primary education, secondary education, and higher education. In general, public schools are administered and organised by the MOEC, while public schools with a religious base, such as Islamic schools or madrasah, are managed by the MORA. However, the curriculum for all schools, whether managed by the MOEC or the MORA, is arranged by the MOEC. Private schools have the option of implementing the national curriculum or other authorised curricula. Generally, the category of the education system can be general education, vocational, academic, professional, vocational-technical, religious, or special education (MoNE, 2003). Before entering primary school, early childhood education and care are provided for children in pre-primary schools. This type of schooling comprises kindergartens, play groups, and childcare centres particularly for children from birth to six years of age in order to prepare their physical and intellectual growth before entering primary school (http://www.ibe.unesco.org). In Indonesia, it is not compulsory for pre-primary aged children to attend early childhood education.

According to the OECD/Asian Development Bank (2015), a primary school has a duration of six years (Years 1-6). Primary education or basic education is provided by general elementary schools (SD: *Sekolah Dasar*) and Islamic primary schools (MI: *Madrasah Ibtidaiyah*) for ages 7-12 years. Junior secondary schools have a three-year duration (Years 7-9), which is provided by general junior secondary schools (SMP: *Sekolah Menengah Pertama*) and Islamic junior secondary schools (MTs: *Madrasah Tsanawiyah*) for ages 13-15 years. Meanwhile, senior secondary school lasts for three years (Years 10-12) for ages 16-18 years. In addition, the vocational school equivalent has a duration of four years. Secondary education is provided by general senior secondary schools (SMA*: Sekolah Menengah Atas*) and Islamic general senior secondary schools (MA*: Madrasah Aliyah*), as well as vocational senior secondary schools (SMK*: Sekolah Menengah Kejuruan*) and Islamic vocational senior secondary schools (MAK*: Madrasah Aliyah Kejuruan*). SMA/MA and SMK/MAK are different in terms of their objectives and study content. The students at SMA/MA are being prepared to continue their studies at the university level, while the SMK/MAK focus on several forms of vocational education designed to prepare students to be part of the workforce.

Basic education, as the foundation of secondary education, is mandatory for all Indonesian students from grades 1 to 9. However, Indonesia is expanding citizen participation in education from grades 10 to 12 as well. In 2013, the MOEC announced a 12-year compulsory education program, from grade 1 in primary school to grade 12 at senior secondary school (OECD/Asian Development Bank, 2015). This means that the first nine years of education, comprising six years in primary school and three years in junior secondary school, is the basic education program that must be attended by all Indonesian citizens from the age of 6 to 15 years. After graduating from junior secondary school, students must now also continue studying at senior secondary school. This education program must be attended by all Indonesian citizens from 16 to 18 years of age. As stated by Kemendikbud (2012), competency standards are the minimum standards which students must achieve to graduate from primary, junior secondary, senior secondary, or vocational school.

Furthermore, non-formal and informal education reside outside of the formal education system, and these are commonly unstructured and do not have a certain academic qualification. As documented by the MoNE (2003, p. 16), “a non-formal education unit consists of training centres and colleges, study groups, community learning centres, majelis taklim (Islamic study group), and other education units of the similar type.” This type of education is provided for citizens as a complement to formal education. Meanwhile, informal education, which is private, can be primary education, homeschooling, and pesantren (Islamic boarding school for secondary school level students) which is provided by families or Islamic religious leaders (MoNE, 2003). Furthermore, after completing the secondary education program, students can continue studying at tertiary education or higher education institutions that include public and private universities or training institutions.

Higher education institutions in Indonesia comprise both academic and professional education. Academic education is directed at the mastery of specific subjects in the academic sense, while professional education focuses on preparing students with particular skills to face the demands of future occupations. Generally, after completing secondary education, students can undertake study for one to four years in a Diploma program in the vocational education system. Meanwhile, in academic education, students can take a three to four years Bachelor program (undergraduate), and two or more years for the Master’s program. Finally, the Doctorate program generally requires three to four years of study.

Broadly speaking, undergraduate students enrolled in Indonesia’s public universities are those who have passed a very strict selection process with certain requirements. Based on The Regulation of the Minister of Research, Technology, and Higher Education of the Republic of Indonesia Number 45 Year 2015, admission to undergraduate programs can be gained through various tests, including the National Selection of State Universities (SNMPTN*: Seleksi Nasional Masuk Perguruan Tinggi Negeri*), the Selection of Joint Entrance State University (SBMPTN*: Seleksi Bersama Masuk Perguruan Tinggi Negeri*), as well as an independent selection process or local test (managed by each university). Students who pass one of these tests have the opportunity to study at a public university. Specifically, students who are selected to study at public universities or institutions that offer teacher preparation or teacher education programs, are those who generally have good academic knowledge and skills. Meanwhile, those who fail to pass one of these pathways (tests) may choose to study at a private university. Dilas et al. (2019) pointed out that most higher education institutions in Indonesia are private universities of lower quality compared to the public universities; however, enrolment in a public university is very competitive. As this study involves undergraduate students studying at teacher education institutions, the following section specifically describes the teacher preparation programs in Indonesian higher education institutions.

1.3.2 Pre-service Teachers' Programs in Higher Education Institutions in Indonesia

In Indonesian higher education institutions, there are several types of Teacher Preparation Programs or Teacher Training Institutions (LPTK: *Lembaga Pendidikan Tenaga Keguruan*), namely the Faculty of Teacher Training and Educational Studies (FKIP: *Fakultas Keguruan dan Ilmu Pendidikan*), Higher Education for Teacher Training and Educational Studies (STKIP: *Sekolah Tinggi Keguruan dan Ilmu Pendidikan*) and the Teacher and Education Studies Institute (IKIP: *Institut Keguruan dan Ilmu Pendidikan*). More specifically, IKIP was transformed in 1999 from an institute of education to a university of education as a consequence of the extension of the mandate of the government. This was done so that the institutions could prepare better quality teachers by providing adequate academic courses and practical experience as well as enabling them to enhance teachers' academic qualifications (Jalal et al., 2009). This transformation refers to the Decree of The President of The Republic of Indonesia (*Keputusan Presiden Republik Indonesia*) No. 93/1999.

Teacher preparation programs at higher education institutions are responsible for equipping pre-service teachers with specialised knowledge and skills in order to teach future generations. Typically, prospective teachers need to hold a four-year bachelor's degree and teaching certificate from the universities of education to meet the minimum standards of competency. In addition, the universities of education also provide a one-year postgraduate teacher professional development program for new teachers to equip them with the content knowledge and pedagogical strategies needed to teach in schools. Through this program, new teachers are expected to be able to plan, implement, and evaluate their teaching (OECD/Asian Development Bank, 2015). After completing these programs, all graduates are entitled to work as teachers in the schools or they may continue their studies to obtain a master's degree.

Further to this, the university provides teacher preparation programs or education programs as well as non-education programs in mathematics, physics, chemistry, biology, accounting, etc. Undergraduate students enrolled in an education program will graduate with a Bachelor of Education degree. Meanwhile, undergraduate students enrolled in the non-education programs will obtain the degree of Bachelor of Science. As noted earlier, it takes a minimum of four years of full-time study to complete the bachelor's degree. In terms of the field of physics, pre-service physics teachers are prepared to become science teachers in the junior secondary schools or physics teachers in the senior secondary schools. Generally speaking, universities of education have study programs for both physics and physics education. However, not all universities of education offer physics education programs. According to Hendayana and colleagues (2010), the number of qualified science teachers at both the primary and secondary school levels in Indonesia is still inadequate, especially in remote areas, due to the limited number of education universities that offer science education programs. This leads to teachers with non-science/physics backgrounds being forced to teach science/physics subjects, even though they do not have expertise in these areas of study. This mismatch inhibits improvement in the quality of teaching and learning in science education, and physics education in particular, which in all likelihood, contributes to low student outcomes in physics. Physics teaching requires specific content knowledge as well as specific pedagogical content knowledge to support student learning (Jauhiainen, 2013).

Considering the significant role that pre-service physics teachers play in the teaching and learning of science (particularly in physics), this study is aimed at investigating Indonesian pre-service physics teachers' understanding of physics concepts as well as their scientific thinking skills. This is important because the current pre-service physics teachers are the future physics teachers of tomorrow who will be responsible for preparing school students with adequate physics knowledge as well as with the skills to think scientifically. Consequently, evaluations and studies of the teacher education programs offered in higher education institutions within Indonesia are needed in order to better understand the current issues and to find better solutions to these complex problems.

1.4 Aims of the Study and the Research Questions

The primary aims of this study are to investigate pre-service physics teachers' scientific thinking (comprising epistemological beliefs, argumentation, and scientific reasoning) and the conceptual understanding of physics in higher education institutions in Indonesia. Specifically, this study aims to achieve the following objectives. These are to:

1. Understand the extent to which pre-service physics teachers' demographic factors such as gender, year level, and university type, influence their scientific thinking and conceptual understanding of physics.
2. Develop a model to examine the relationships between the identified demographic factors, scientific thinking, and conceptual understanding of physics among Indonesian pre-service physics teachers.
3. Empirically validate the proposed model developed in this study using a structural equation modelling (SEM) approach.
4. Identify other factors that may contribute to pre-service physics teachers' scientific thinking and conceptual understanding of physics by asking for their perceptions about the existing teaching and learning practices they have experienced during their studies in higher education.

To achieve the aims of the study, a set of research questions have been formulated, as follows:

1. Are there any differences in demographic factors with regard to pre-service physics teachers' scientific thinking and conceptual understanding of physics?
    a. Are there any gender differences in pre-service physics teachers' epistemological beliefs, argumentation, scientific reasoning, and conceptual understanding of physics?
    b. Are there differences between year level with regard to pre-service physics teachers' epistemological beliefs, argumentation, scientific reasoning, and conceptual understanding of physics?
    c. Are there differences between university type with regard to pre-service physics teachers' epistemological beliefs, argumentation, scientific reasoning, and conceptual understanding of physics?
2. What are the relationships between pre-service physics teachers' scientific thinking (i.e., epistemological beliefs, argumentation, and scientific reasoning), conceptual understanding of physics, and their demographic factors?

3. What are pre-service physics teachers' perceptions of the relationships between scientific thinking (i.e., epistemological beliefs, argumentation, and scientific reasoning) and their conceptual understanding of physics?
4. What are pre-service physics teachers' perceptions of the opportunities and barriers in enhancing their scientific thinking and conceptual understanding of physics?

1.5 The Significance of the Study

The significance of the study can be grouped into three categories, namely theoretical, methodological, and practical significance. In terms of theoretical significance, the findings of this study provide contributions to knowledge especially in the field of physics education. Since there have been few research studies concerning some aspects of scientific thinking i.e., epistemological beliefs, argumentation, and scientific reasoning in the context of pre-service physics teachers' education programs in Indonesia, the findings of this study are expected to make a contribution to the understanding and development of knowledge in the area of research into physics education. Apart from this, there have been no research studies identifying the way in which epistemological beliefs, argumentation, and scientific reasoning act as predictor variables in ways that are considered to underpin the understanding of physics concepts. Previous studies have investigated these learning variables separately, and have commonly relied on the analysis of bivariate relationships in which the conclusions drawn by the researchers could be misleading in understanding the complex network of pre-service physics teachers' cognitive skills.

In this study, all determinant factors are comprehensively examined in a single plausible model, and pre-service physics teachers' perceptions regarding current physics teaching and learning practices are explored deeply. The findings will be useful for pre-service teachers' education programs as they can assist physics teacher educators in higher education institutions to gain a better understanding of how their students are reasoning and arguing, and how their students' epistemological beliefs influence their learning processes as they try to understand physics concepts. The findings also provide physics teacher educators with insight into pre-service physics teachers' perceptions and preferences about the actual physics teaching and learning environment. The information gained from this study could be used to assist physics teacher educators in planning physics teaching strategies to ensure effective learning among pre-service physics teachers so that they can acquire a strong conceptual understanding of physics. It is expected that pre-service teachers who have sophisticated scientific thinking (i.e., epistemological beliefs, argumentation, and scientific reasoning) and a strong conceptual understanding of physics will be better prepared to plan and implement physics learning and teaching practices for their secondary students in the future. Hence, this study provides empirical evidence and broader insights especially for physics teacher education programs to improve understanding about how these factors affect Indonesian pre-service physics teachers' scientific thinking and their conceptual understanding of physics.

In addition, this study is significant as it provides a methodological contribution as well as comprehensive data relating to pre-service physics teachers' scientific thinking and their understanding of basic concepts in physics employing a contemporary modelling approach. To achieve the research goals, several statistical procedures and techniques have been used to analyse the datasets, involving a mixed methods design, Rasch model analysis, Confirmatory Factor Analysis (CFA), and Structural Equation Modelling (SEM). Prior to conducting any analysis, all raw scores obtained in the quantitative study were transformed into interval scores to generate the Weighted Likelihood Estimates (WLE) scores. The reliability and validity of the research instruments were established by employing the Rasch model analysis and CFA approach. In addition, the SEM approach was used to examine the relationships between several variables involved in this study. The findings obtained in the qualitative study completed and enriched the quantitative study, helping to obtain a broader picture to assist with understanding the complex issues that were investigated. These statistical analysis techniques have not been widely used in physics education research, especially in the Indonesian context. Consequently, this study has made important methodological contributions in providing broad insights and valid conclusions with respect to the quantitative phase of the data analysis.

At a broader level, the findings of the current study offer practical significance that may benefit physics teachers, educators, researchers, curriculum developers, and policymakers in an attempt to promote teaching practices and the development of productive curriculum that focuses more on developing pre-service physics teachers' skills in scientific thinking and their understanding of physics concepts. This research may assist the formulation of future education policies in the Indonesian context that produce graduates who have the intended holistic competencies for better physics teaching in Indonesia. The present study could become a reference for researchers and practitioners who want to carry out further research in the same research topic. This is important considering the small number of research studies that have been conducted in Indonesia which focus on the knowledge of epistemological beliefs, argumentation, and scientific reasoning and its effect on the improvement of pre-service teachers' conceptual understanding of physics in an attempt to enhance the quality of physics education.

1.6 Definition of Terms

In consideration of the purpose of this study, some key terms to be used are described briefly as follows.

Epistemological Beliefs*.* According to Hofer (2006), epistemological beliefs is defined as individual's beliefs about the nature of knowledge and knowing. For the purposes of this study, epistemological beliefs include five sub-scales; namely, the structure of scientific knowledge, the nature of knowing and learning, real-life applicability, evolving knowledge, and the source of the ability to learn (Elby, 1999).

Argumentation*.* According to Erduran and Jiménez-Aleixandre (2008, p. 13), argumentation is defined as "the connection between claims and data through justifications or the evaluation of knowledge claims in light of the evidence, either empirical or theoretical." For this study, argumentation is indicated by the ability to make a scientific argument and to challenge arguments (Sampson & Clark, 2006).

Scientific Reasoning*.* Scientific reasoning is an essential skill required to conduct scientific inquiry (Han, 2013). For the purposes of this study, scientific reasoning is indicated by assessment scores from sub-scales that measure conservation of weight and volume, proportional reasoning, control of variables, probability and correlational reasoning, and hypothetical-deductive reasoning (Lawson, 1978, 2000a).

Physics Conceptual Understanding (PCU)*.* In this study, PCU refers to the ability to understand physics concepts, specifically in Mechanics and Electricity & Magnetism topics.

1.7 The Structure of the Thesis

This thesis is organised into nine chapters as presented in Figure 1. 1. Chapter 1 provides a general overview of the research, including the background and context of the study, the aim and research questions, the significance of the study and the definition of the key terms used in the study.

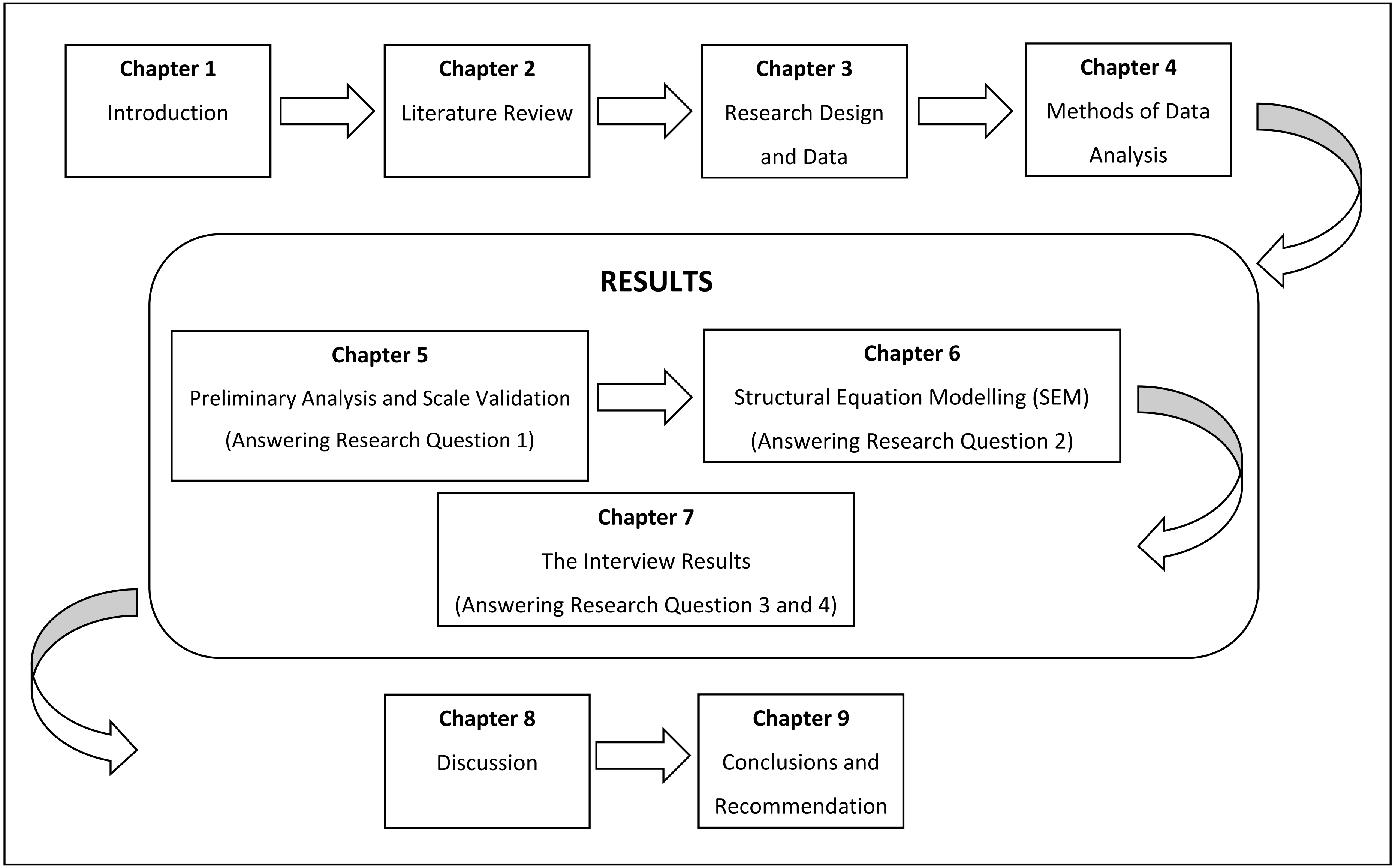


Figure 1. 1 Overview of the thesis

Chapter 2 reviews the literature in the field of study. This chapter presents a description of the nature of student learning and the conceptual understanding of physics. This is followed by a review of previous research on aspects of scientific thinking, covering scientific reasoning, argumentation, and epistemological beliefs, as well as the relationships between these variables to show how the current research relates to the work of other scholars and researchers in the field. Several aspects related to physics and teaching practices are also highlighted. Finally, an explanation of a proposed model for this study is presented.

Chapter 3 presents the research design and the data collection methods. The chapter commences with a presentation of the mixed methods research design used in the thesis, the research sites, and the participants. This is followed by a description of the methods of data collection and research instruments for both the quantitative and qualitative phases of the study. The final section presents an overview of the ethics approval process that was undertaken.

Chapter 4 outlines the methods of data analysis in both the quantitative and qualitative phases of the study. In this chapter, general methodological considerations, including missing values and tests for normality and multicollinearity are presented. This is followed by an exploration of the quantitative data analysis methods covering the reliability and validity tests of the research instruments, descriptive and inferential statistical analysis, as well as the Rasch model for item analysis. There is also a description of Confirmatory Factor Analysis (CFA) to test construct validity and the Structural Equation Modelling (SEM) approach to examine the relationships between multiple variables. The final section presents an outline of the qualitative techniques used for analysing the data obtained from the individual semi-structured interviews.

Chapter 5 presents the results of the preliminary analysis and the scale validation. This chapter begins with an overview of the demographic information of the participants, consisting of descriptive data about gender, year level, and types of university attended. This is followed by a description of the multidimensional Rasch model analysis, the person fit and item fit analyses, item discrimination, as well as an item–person map. The chapter also describes the analysis of the scale validation of the survey instruments used in the study using the Confirmatory Factor Analysis (CFA) approach. The final section presents the descriptive and inferential statistical analysis testing the normality and multicollinearity of the datasets, t-tests, and a one-way Analysis of Variance (ANOVA).

Chapter 6 reports on the results of the proposed model analysis using the Structural Equation Modelling (SEM) approach to examine the relationships between the multiple research variables. In this chapter, a number of variables used in the research model are presented first. This is followed by the results of the measurement models and the structural model analysis obtained from the hypothesised model, along with the trimming of paths from the model. Finally, the final model resulting from the SEM approach, as well as a summary of the fit indices for the model, are presented.

Chapter 7 presents the participants' demographic information for the qualitative phase of the study. This is followed by the findings generated from the interviews with pre-service physics teachers that explore their perceptions of the relationship between aspects of scientific thinking (i.e., epistemological beliefs, argumentation, and scientific reasoning) and their understanding of physics concepts. Finally, the participants' perceptions of the extent to which physics teaching and learning practices helps them to enhance their scientific thinking and conceptual understanding of physics are presented.

Chapter 8 provides a more detailed and in-depth description of both the quantitative and qualitative findings. The chapter begins with a description of the effect of demographic variables on pre-service physics teachers’ scientific thinking and conceptual understanding of physics. This is followed by an illustration of the relationships arising between the different aspects of scientific thinking (comprising epistemological beliefs, argumentation, and scientific reasoning) and the conceptual understanding of physics. Finally, pre-service physics teachers’ perceptions of physics teaching and learning factors related to the identification of any opportunities and barriers to improving their skills in scientific thinking and developing their understanding of physics concepts are presented.

Chapter 9, the final chapter, presents the conclusions of the study and recommendations for further research. The chapter begins with the research aims and the study design, and then goes on to summarise the most important findings. This is followed by the conclusions drawn from the study and an explanation of the implications covering theoretical, methodological, and practical concerns. Finally, the limitations of the study and recommendations for future research are presented.

CHAPTER 2
LITERATURE REVIEW

2.1 Overview

This chapter reviews the relevant literature that outlines several learning theories including behaviourism, cognitivism, constructivism, and humanism with a view to understanding how students learn to think scientifically and understand physics concepts. This is followed by a description of the conceptual understanding of physics and several aspects of scientific thinking comprising scientific reasoning, argumentation, and epistemological beliefs. The proposed research model that demonstrates the relationship between aspects of scientific thinking and the conceptual understanding of physics is also presented. Finally, this chapter reviews the literature in relation to several aspects of teaching and learning that are considered to contribute to the improvement of scientific thinking and the conceptual understanding of physics.

2.2 Educational Learning Viewpoints

This section presents a brief description of a number of learning theories in order to better understand the assumptions and principles that are relevant to the current study. Learning has been defined in several ways by previous theorists and researchers. According to Vosniadou, Ioannides, Dimitrakopoulou, and Papademetriou (2001, p. 382), learning is “an effortful and mindful process” in which the learner is involved in an active process, such as contributing to research, solving problems, or conducting an experiment in order to improve existing knowledge structures. In the same vein, Ertmer and Newby (2013, p. 44) pointed out that "learning is a complex process that has generated numerous interpretations and theories of how it is effectively accomplished.” Likewise, learning has been defined as the mental and practical activity of students, which is far more complex than simply absorbing and storing information or acquiring knowledge (Engeström, 1994). Through meaningful learning activities, students are able to interpret the information they obtain by linking it with previous understandings to construct new knowledge. More specifically, learning in the science can also be seen as a process of acquiring knowledge and understanding of natural phenomena which leads to changes in student behaviour (Pritchard, 2017).

Certainly, the processes of learning can take place both in and outside of the classroom. In educational research, it is important for educators to investigate teaching and learning practices that take place in the classroom, as well as students' academic outcomes. Specifically, the extent to which a teacher's teaching and learning practices have facilitated students in improving their scientific thinking skills and understanding of physics phenomena in the real world, which is one of the main goals of physics education (Phanphech, Tanitteerapan, & Murphy, 2019; Schmaltz, Jansen, & Wenckowski, 2017). As this study aims to investigate students' scientific thinking and conceptual understanding of physics, there is a core question that needs to be answered: 'What types of teaching and learning strategies, as well as learning environments, assist students to practice skills in scientific thinking and improve their ability to understand physics concepts, and what factors influence the improvement of these cognitive variables?' To be able to answer this question, it is very important to understand a number of relevant educational learning viewpoints, such as the broad theories of behaviourism, cognitivism, constructivism, and humanism. Learning theories provide insights for instructors to understand how students learn, and how to design and develop effective learning environments for students in order to overcome teaching problems in the classroom and to improve students' academic outcomes. These theories may have some overlap; however, each can be seen as a separate approach to understanding and explaining student learning (Ertmer & Newby, 2013; Guey, Cheng, & Shibata, 2010). A brief description of these four learning theories is presented below.

Behaviourism

Behaviourists emphasise that behavioural change can be explained as the association between stimulus and response. Changes in learner behaviour occur in response to stimuli from the surrounding environment (Ertmer & Newby, 2013; Kay & Kibble, 2016; Nagowah & Nagowah, 2009). In other words, behaviourists view learning as a process of response to the environmental stimulus, where learning outcomes as manifestations of learning, can be observed as changes in learner behaviour.

Behavioural learning theories point out that knowledge in the educational context is transmitted by instructors, while learners are seen as passive recipients of knowledge and information (teacher-centred) (Guney & Al, 2012; Kay & Kibble, 2016; Keesee, 2012), rather than as thinking learners. In other words, the student's role is basically passive and relying on their teachers to absorb information. Meanwhile, teachers play a highly active role in being responsible for designing a conducive learning environment and transmitting knowledge, so that students can respond appropriately and constructively to the stimulus presented (Ertmer & Newby, 2013). In addition, according to behaviourist theory, learning involves recalling or reproducing facts, and defining and applying concepts in only minimal ways (Ertmer & Newby, 2013). Therefore, the process of learning takes place passively. Specifically, when conducting experiments, students tend to use recipe laboratories with clear protocols, where they follow step-by-step procedures that are unlikely to promote deep thinking (Kay & Kibble, 2016).

Furthermore, behaviourists view learning as the acquisition of new behaviour, while ignore students' feelings and the possible engagement of deep thinking processes (Kay & Kibble, 2016; Nagowah & Nagowah, 2009). Consequently, behavioural learning theories are often criticised as not contributing to the attainment of higher-level thinking skills such as problem-solving and critical thinking, which generally require a deeper thinking process (Schunk, 1991, as cited in Ertmer & Newby, 2013).

Behaviourist learning theory seems to fall short of overcoming problems that address improvements in higher-level thinking skills and conceptual understanding of students, which are investigated in this study. Nevertheless, while teaching and learning practices that are largely dominated by instructors may not promote students' scientific thinking, they may help students to easily acquire knowledge and to understand physics concepts to some extent. This depends on how the instructors create a stimulating environment designed to assist students to not only acquire knowledge, but also understand it. As noted by Pritchard (2017), the teacher-centred approach that promotes rote learning is considered useful for learners to overcome difficulties in knowledge acquisition, but tend to be less effective for developing students' conceptual understanding.

Cognitivism

Pritchard (2017, p. 17) described cognitivism as a set of "mental processes such as learning, perceiving, remembering, using language, reasoning and solving problems." This learning theory is different from behaviourism which focuses on changes in the behaviour of learners. According to Ertmer and Newby (2013), cognitive learning theories emphasise the enhancement of students' mental processes by encouraging them to adopt an appropriate learning strategy in order to organise and connect new information with existing knowledge in their minds. From this viewpoint, each learner has different prior knowledge that can contribute to different interpretations and understandings of new information received. Hence, different learners might generate different learning outcomes, even though they are in the same learning environment. The cognitive view is also concerned with how learners know, acquire, organise, and store knowledge, as well as the way they retrieve information to promote learning and understanding of the real world (Ertmer & Newby, 2013). In this case, learners are actively engaged in processing information and knowledge as they learn, rather than simply listening to the instruction provided by the instructors. In turn, this helps them to make sense of, and understand, what they have experienced.

Furthermore, it is essential to explore the cognitive development of learners to understand their academic outcomes. Woolfolk (2005, p. 20) defined cognitive development as "gradual orderly changes by which mental processes become more complex and sophisticated." To fully understand how students develop their cognitive abilities over time, Piaget proposed four stages of cognitive development, namely the sensorimotor (0 – 2 years), preoperational (2 – 7 years), concrete operational (7 – 11 years), and formal operational (11 – adult) stages. Eggen and Kauchak (2015, p. 70) outlined these four stages of cognitive development referring to a specific age range. As this present study is aimed at investigating students at the tertiary level, it is necessary to understand the development of cognitive abilities in young adults (i.e., the final stage). In the formal operational stage, an individual's ability becomes more mature and sophisticated allowing them to think abstractly and hypothetically about complex phenomena in the world. Eggen and Kauchak (2015, p. 73) also stated that "when students cannot think abstractly, systematically, or hypothetically, they revert to memorising what they can, or, in frustration, give up completely."

Furthermore, to understand how learners 'think, learn, and develop', Eggen and Kauchak (2015) developed a framework of cognitive learning theories as shown in Figure 2. 1. The arrows in Figure 2. 1 indicate that these principles are interdependent and show that students' experiences both inside and outside of the classroom contribute to their learning and development. Students who lack experience might have difficulty in learning and developing their knowledge and skills. Eggen and Kauchak (2015, p. 286) argued that "the need to make sense of our experience may be the most basic cognitive principle." To help students make sense of an experience, they need to construct knowledge that might depend on their prior knowledge. However, it is highly likely that students' prior knowledge could lead to misconceptions about certain concepts. In addition, to be a successful learner, social interaction can contribute to learning by "providing information, building an idea and putting thoughts into words" (Eggen & Kauchak, 2015, p. 288). It is worth noting that the process of knowledge construction by learners is influenced by interactions with others who might be more knowledgeable.

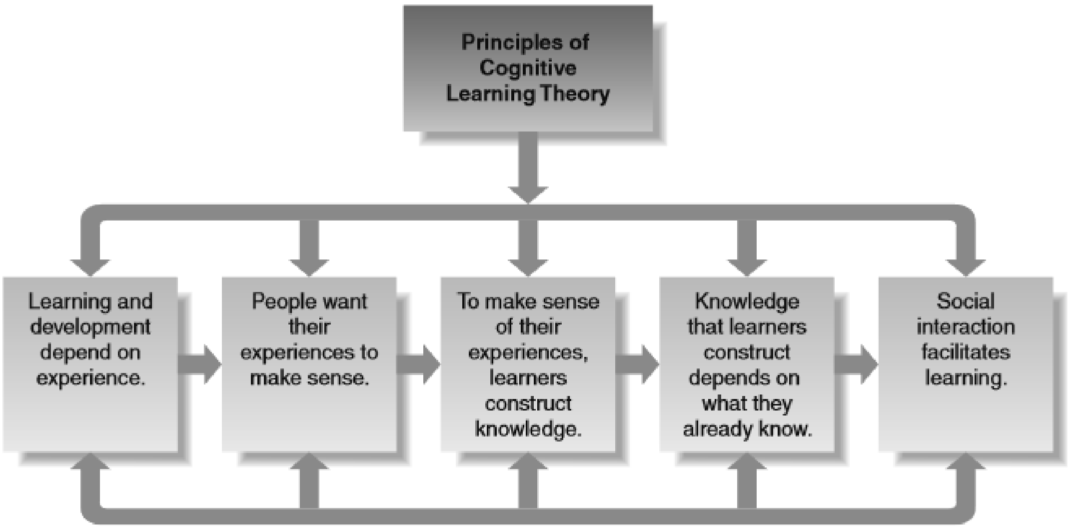


Figure 2. 1 Principles of cognitive learning theory (Eggen & Kauchak, 2015, p. 286)

As opposed to behaviourism which views students as passive receivers of environmental stimuli, in the cognitive perspective, students play an active role in the learning process (Kay & Kibble, 2016). Cognitive learning theory indicates that students do not merely absorb knowledge or information from the environment, they also actively restructure their knowledge in relation to their existing knowledge in order to understand what they are experiencing. When students make a connection between new and prior knowledge, thinking processes may occur that lead to the acquisition of new understandings or new perspectives about the world. Therefore, it is important for instructors to provide a learning environment that allows students to practice their scientific thinking skills in order to help them process information or knowledge, make connections between knowledge or facts, and to organise, store, and reproduce scientific knowledge. As noted by Ertmer and Newby (2013), cognitivism is considered appropriate for explaining complex learning forms, such as reasoning, problem-solving, and information processing. Helping students to succeed in learning is the main goal in the educational world; thus, cognitive learning theory assists educators to achieve this educational goal.

Constructivism

Constructivist learning theory has a long historical association with the works of Dewey (1929), Bruner (1961), Vygotsky (1962), and Piaget (1980), as documented by Bada and Olusegun (2015, p. 66). Educational psychologists view constructivism as a perspective on learning in which learners acquire advanced knowledge and construct their understanding based upon their prior experiences, rather than the transmission of simple knowledge from teacher to student (Bada & Olusegun, 2015). As noted by Pritchard (2017, p. 17), "constructivists view learning as a result of mental construction." In addition, constructivists emphasise that the active engagement of learners during the processes of constructing their knowledge (Veletsianos, 2016), exploring complex topics or knowledge that require skills of higher-level thinking, and collaborating with peers are essential aspects of learning (Ertmer & Newby, 2013). Therefore, a constructivist learning environment facilitates and provides opportunities for students to play an active role in learning. According to Ertmer and Newby (2013, p. 58), "both cognitivists and constructivists view the learner as being actively involved in the learning process, yet the constructivists look at the learner as more than just an active processor of information; the learner elaborates upon and interprets the given information." In the construction of knowledge, students should be encouraged to interpret new information or knowledge acquired from their new learning experiences and build meaningful connections with their prior structure of knowledge. In classroom practices, constructivist learning activities enable students to solve real-world problems and make sense of their experiences through inquiry or investigation (Veletsianos, 2016). Thus, students are required to go beyond formulaic solutions to construct more knowledge, while teachers play a vital role in facilitating students to become active participants in their learning process and help them to develop their understanding and skills. This learning theory tends to shift the approach from teacher to student (or student-centred learning), which contrasts with a traditional learning approach based on the passive transmission of knowledge from teacher to student (or teacher-centred learning). Bada and Olusegun (2015) further summarised the differences between traditional learning and constructivist learning as presented in Table 2. 1.

Referring to Table 2. 1, constructivist learning focuses more on student-centred learning rather than teacher-centred learning in the acquisition of knowledge by constructing an understanding of phenomena. Teaching and learning practices or learning environments should be designed to encourage students to actively construct their own knowledge through thinking processes. A teacher should act as a facilitator who can help students connect their prior knowledge with new ideas. Thus, in a constructivist approach, students are expected to be actively engaged in the learning process to acquire an understanding of their new learning experience based on interactions with their environment. This student-centered strategy has an impact on higher-level thinking skills.

Table 2. 1 The differences between the traditional and the constructivist classroom
(Adopted from Bada & Olusegun, 2015, pp. 68-69)

Traditional Classroom	Constructivist Classroom
Curriculum begins with the parts of the whole. Emphasises basic skills.	Curriculum emphasises big concepts, beginning with the whole and expanding to include the parts.
Strict adherence to fixed curriculum is highly valued.	Pursuit of student questions and interests is valued.
Materials are primarily textbooks and workbooks.	Materials include primary sources of material and manipulative materials.
Learning is based on repetition.	Learning is interactive, building on what the student already knows.
Teachers disseminate information to students; students are recipients of knowledge.	Teachers have a dialogue with students, helping students construct their own knowledge.
Teacher's role is directive, rooted in authority.	Teacher's role is interactive, rooted in negotiation.
Assessment is through testing, correct answers.	Assessment includes student works, observations, and points of view, as well as tests. Process is as important as product.
Knowledge is seen as inert.	Knowledge is seen as dynamic, ever changing with our experiences.
Students work primarily alone.	Students work primarily in groups.

Furthermore, in a constructive learning environment, teachers are focused on teaching that allows learners to practice complex thinking skills such as problem-solving and scientific thinking skills, in which students are actively involved in constructing and understanding knowledge, rather than being passive listeners as in more traditional classroom practices (Bada & Olusegun, 2015). Constructivist learning theory is relevant to this study because it explains how teaching and learning practices can take place in the classroom in order to enhance students' higher-level thinking skills and conceptual understanding.

Humanism

As documented by Biddulph and Carr (1999), humanistic learning theory has been outlined by scholars such as William Glasser, Carl Rogers, Abraham Maslow, and Guy Claxton. Humanism is characterised as being concerned with "… the dignity, autonomy, freedom, integrity, well-being, equity, and potential of learners" (Chen & Schmidtke, 2017, p. 119). According to Arghode, Brieger, and McLean (2017, p. 598), humanism emphasises learner "motivation and proactivity more than imposing concepts and learning." Hence, this learning theory is considered as satisfying learners' need for self-actualisation (Jingna, 2012). This theory also focuses on "... the psychological needs and values of individual learners, rather than on the process of learning" (Guey et al., 2010, p. 107). In other words, humanist learning theory basically relates more to feelings or emotions, interests, attitudes, thoughts, values, and motivations of learners and recognises that these have a powerful influence on learning processes that promote learning (Arghode et al., 2017; Guey et al., 2010).

Clearly, this learning theory demonstrates that learning is not merely a cognitive process or a behavioural change, but instead focuses on students' affective parameters that include motivation, interests, feelings, and personal development that support their learning and promote lifelong learning (Arghode et al., 2017). This learning theory emphasises that instructors should not ignore students' interests, attitudes, and motivation to learn, and considers other affective parameters that influence students' learning. As the present study is related to students' academic outcomes in physics, and recognises that physics is a subject that many students are not interested in because it is considered difficult and abstract (Firdaus et al., 2019; Sobremisana, 2017), it is important to understand the extent of students' interest, attitudes, and motivation to learn physics and how this may affect their skills to think scientifically and their ability to understand physics concepts.

In educational practice, the learning environment of schools or universities should provide for the needs of students by having a comfortable and safe learning environment that supports their learning and increases their motivation and interest in learning physics and practicing their thinking skills. Additionally, students should feel like part of the school and have the same rights as other students. The learning environment should also allow students to express their potential and help them to achieve their learning goals (Guney & Al, 2012). According to the humanist learning theory, instructors or teachers should stimulate students to explore their interests in order to develop their knowledge and skills (Chen & Schmidtke, 2017). Humanism recognises that learners play an essential role in the learning process (being student-oriented), while the instructor acts as their facilitator (Arghode et al., 2017).

Understanding these learning theories is useful as a vehicle for helping instructors to both understand and identify a variety of learning variables that may contribute to their students' academic outcomes. More specifically, the four learning theories mentioned above are useful as a framework for conducting the present study and analysing the research findings that seek to investigate university students' academic outcomes in relation to conceptual understanding of physics and several aspects of scientific thinking (i.e., scientific reasoning, argumentation, and epistemological beliefs). The following section presents a description of the multiple learning variables identified for investigation in this study.

2.3 Conceptual Understanding of Physics

Physics is closely related to the complexity of human life. Understanding physics concepts is crucial to making sense of the world and is considered to be one of the main goals of physics education (Dervic et al., 2018; Phanphech et al., 2019). To succeed in physics, one needs a good conceptual understanding of its subject matter (Kola, 2017). Educators and teachers alike need to help their students enhance their ability to conceptually understand physics, so they can better solve problems experienced during the learning process (Dervic et al., 2018). Ideally, teachers should have a deep conceptual understanding of the content being taught in order to be able to teach their students well (Andayani et al., 2018; Gürel & Süzük, 2017). If teachers have a poor understanding of physics concepts, this can lead to poor quality teaching of physics in the classroom which, in turn, can have a negative impact on students' conceptual understanding of physics (Gaigher, 2014), and could lead to students developing misconceptions about physics because of how they have been taught.

Previous researchers have used various definitions related to conceptual understanding. For instance, Scott, Asoko, Leach, Abell, and Lederman (2007, p. 35) argued that “concepts are to be understood as basic units of knowledge that can be accumulated, gradually refined, and combined to form ever richer cognitive structures.” Meanwhile, understanding can be defined as the ability of individuals to interpret, translate, or express the results of their thoughts related to the knowledge being studied (Putra, 2019). Understanding also relates to students' ability to process and reflect on ideas, connect prior knowledge with natural phenomena, make sense of previous experiences, establish new insights, draw conclusions, explain them using their own words, and transfer them to different situations (Wiggins & McTighe, 2011). Furthermore, aspects of conceptual understanding can be specified in terms of breadth and depth of knowledge. Breadth of knowledge reflects "the extent of knowledge that is distributed and represents the major sectors of a specific domain", while depth of knowledge reflects "the knowledge of scientific principles that describes the relationship among concepts" (Alao & Guthrie, 1999, p. 244). Conceptual understanding is also known as intuitive understanding, and refers to the ability to recognise fundamental concepts in a variety of different representations and applications (Richardson & McCallum, 2003). Scott et al. (2007, p. 34) asserted that "conceptual understanding is not a purely cognitive process in the individual, but also individuals as they function in social contexts." Thus, the conceptual understanding of students is more than simply the memorisation of facts, but can also be considered as students' ability to construct their own knowledge by restructuring their existing knowledge through both an individual process and social activity to acquire information about concepts from the environment. They use this knowledge to interpret their new insights and experiences. Exploration of students' conceptual understanding has revealed much about what students know and how they learn and apply their understanding in different situations. Students who have good conceptual understanding are more likely to have various abilities such as "memorizing, explaining, finding facts, stating examples, generalizing, implementing, analogizing, and expressing new concepts in other ways" (Saputro et al., 2019, p. 1).

Generally speaking, physics concepts comprise principles, laws, and theories of physics, as well as their application in everyday life. Students' conceptual understanding of physics can be indicated by their ability to apply concepts in a particular situation, as they analyse and solve complex physics problems as well as transfer knowledge to other contexts (Corpuz, 2006, p. 2). Corpuz (2006, p. 2) also acknowledged that "complex physics problems are problems that cannot be solved by just employing a plug-and-chug method of problem-solving." Furthermore, Richardson et al. (2003, p. 100) argued that in the case of physics, "students need conceptual understanding first, and some comfort in using basic skills; then a deeper approach and more sophisticated skills become meaningful." Simply put, the definition of conceptual understanding of physics in the present study is the ability to understand physics concepts, construct knowledge, interpret and apply the concepts in everyday contexts, and incorporate prior knowledge and experience with new knowledge gained from new learning environments that lead to new insights or reasonable conclusions. Certainly, the ability to understand physics concepts is more than simply memorising facts or knowledge, or even having the ability to use formulae and mathematical calculations to solve traditional physics problems, rather it is the ability to understand natural phenomena logically through scientific thinking processes.

One of the main issues which negatively influences the acquisition of knowledge and understanding, particularly in physics education, is that students may have little awareness of their level of conceptual understanding of physics, and teachers may be unaware of the poor levels of conceptual understanding of their students. Previous research has stated that students often encounter misconceptions about physics which are persistent and hinder their understanding of physics concepts (Phanphech et al., 2019; Sobremisana, 2017). Students who have a poor understanding of physics concepts are more likely to face challenges succeeding in the field of physics (Kola, 2017). Broadly speaking, students who hold fundamental misconceptions tend to resist accepting new knowledge and need much time to refine or change their ideas (Docktor & Mestre, 2014; Jiang, Wang, Wang, & Ma, 2018). In addition, students who are able to explain a particular physics theory may not necessarily understand the physics concepts that underlie the theory and be able to apply them in the real world.

The previous research has also revealed that students struggle to understand physics concepts because physics is considered to be a difficult and abstract subject (Adeyemo & Babajide, 2014; Cahyaningrum & Hidayat, 2018; Firdaus et al., 2019; Hairan, Abdullah, & Husin, 2019; Kola, 2017; Tonjo, Wirjawan, & Untung, 2017). In addition, even though students have high levels of average achievement in physics, there remains a population of students who fail to develop their understanding of physics concepts (Pollock, 2005). For example, Kim and Pak (2002) demonstrated that students still possessed difficulties in understanding basic physics concepts, even though they had worked through more than 1,000 traditional physics problems from the textbooks. As stated by McDermott (1991, as cited in Gaigher, Rogan, & Braun, 2007), the success of students in correctly answering, physics questions that use formulae and numerical calculations does not necessarily imply that an appropriate level of understanding of physics concepts has been achieved.

Further to this, the concepts of mechanics, and electricity and magnetism are basic concepts that are important to be mastered and understood by both teachers and students. These concepts can explain scientific phenomena that occur in everyday life. However, previous research has revealed that students and teachers found these concepts quite challenging and difficult to understand, which has lead to the problem of students developing misconceptions about science or physics

(Cahyaningrum & Hidayat, 2018; Docktor & Mestre, 2014; Gunstone, Mulhall, & McKittrick, 2009; Kim & Pak, 2002; Saputro et al., 2019; Tuder & Urban-Woldron, 2015). For instance, some students cannot differentiate between distance and displacement or speed and velocity. They think each pair has the same meaning and can be used interchangeably (Dilber, Karaman, & Duzgun, 2009). In addition, it is common to find that students are unaware that velocity, acceleration, and force are vectors that have a direction (Tuder & Urban-Woldron, 2015). For projectile motion and vectors, students find it difficult to visualise the independence of vertical and horizontal motion (Hestenes, Wells, & Swackhamer, 1992). Furthermore, the concepts of electricity and magnetism are considered by many students to be more abstract than mechanics, and difficult to understand. As noted by Phanphech et al. (2019), many students misunderstood and failed to understand concepts related to electricity, so they sometimes relied on their intuitive conceptions to understand these concepts. In addition, Raduta (2005) demonstrated that many students have difficulty with vector or scalar products, which are fundamental to understanding the concepts of electricity and magnetism. They also faced difficulties in understanding the concept of electromagnetic induction, how it is produced, and how to determine direction even in a simple circuit. Many students thought of the magnetic pole as being electrically charged through which a "magnet of opposite charge will pull electrons" (Raduta, 2005, p. 7). This has happened because students had difficulty understanding the interaction between the magnetic field and electric charges. Likewise, a recent study also demonstrated that pre-service science teachers had misconceptions about simple electricity circuits both in series and parallel circuits (Saputro et al., 2019). Hence, it is important to investigate pre-service teachers' conceptual understanding of physics, as they will be the physics teachers of tomorrow who will have the responsibility of teaching physics to their students in the future.

Furthermore, several instruments for evaluating students' understanding of physics concepts have been designed and developed by previous researchers. More specifically for measuring the mechanics concepts of students, educators and researchers have used the *Mechanics Baseline Test* (MBT) (Hestenes & Wells, 1992), the *Force Concept Inventory* (FCI) (Hestenes et al., 1992), and the *Force and Motion Conceptual Evaluation* (FMCE) (Thornton & Sokoloff, 1998). Meanwhile, there are various instruments to assess students' understanding of the concepts of electricity and magnetism, including the *Determining and Interpreting Resistive Electric Circuit Concepts Test* (DIRECT) (Engelhardt & Beichner, 2004), the *Conceptual Survey of Electricity and Magnetism* (CSEM)

(Maloney, O'Kuma, Hieggelke, & Van Heuvelen, 2001), and the *Brief Electricity and Magnetism Assessment* (BEMA) (Chabay & Sherwood, 1997).

The present study uses the FCI and BEMA surveys to assess pre-service physics teachers' conceptual understanding of mechanics, and electricity and magnetism topics respectively. The FCI survey is the most well-known and widely used among educators and researchers in the field of physics education to assess students' understanding of the concepts of kinematics and Newton's laws (Hairan et al., 2019; Von Korff et al., 2016). Meanwhile, BEMA is a reliable test and is also widely used among physics education researchers. It covers a broad range of basic concepts of electricity and magnetism (Ding, Chabay, Sherwood, & Beichner, 2006). As stated by Von Korff et al. (2016), instructors and researchers need to use well-designed assessment instruments in their research to ensure that their efforts to assess students' conceptual understanding are effective and lead to valid conclusions. A detailed explanation of the FCI and BEMA surveys is presented in Chapter 3.

As mentioned earlier, previous research has demonstrated that students' ability to understand physics concepts is unsatisfactory, and misconceptions about physics among students remain well documented. This suggests that there should be an effort by educators and researchers, especially in the field of physics education, to identify cognitive variables that may affect students' understanding of physics concepts. Certainly, many previous studies have sought to investigate the various factors that affect students' conceptual understanding of physics. They have even agreed that there is no single factor affecting students' learning achievement, conceptual understanding of physics, and their beliefs about physics (Marx & Cummings, 2007; Perkins, Gratny, Adams, Finkelstein, & Wieman, 2006; Pollock, 2005).

More specifically, previous studies have highlighted a number of factors that have an impact on students' understanding of physics concepts, such as teaching methods (Bigozzi, Tarchi, Fiorentini, Falsini, & Stefanelli, 2018; Kola, 2017), the learning media used in the classroom (Husnaini & Chen, 2019; Phanphech et al., 2019), the skills of scientific reasoning (Coletta & Phillips, 2015) and argumentation (Nurlatifah et al., 2018), gender differences (Henderson, Stewart, Stewart, Michaluk, & Traxler, 2017; Karim, Maries, & Singh, 2018), and students' beliefs about the nature of knowledge and knowing (Franco et al., 2012; Stathopoulou & Vosniadou, 2007). However, there has been no comprehensive research conducted that integrates and correlates these variables in a single study. More specifically, research on prospective physics teachers in Indonesia appears to be limited and under-explored, especially in relation to cognitive variables such as scientific reasoning, argumentation, and epistemological beliefs that are predicted to affect students' conceptual understanding of physics. The present study seeks to address this gap in the previous research by extending the investigation of various learning variables that might contribute to pre-service physics teachers' understanding of physics concepts by examining the relationships that exist between multiple variables through a plausible model. A range of learning variables considered to affect students' understanding of physics concepts are described in the following sections.

2.4 Scientific Thinking

Scientific thinking is one of the interesting research topics being investigated by psychologists and educators that closely relates to cognitive development as described in the work of Inhelder and Piaget (1958, as cited in Koerber et al., 2015). It is common to find in the literature that the terms scientific thinking and critical thinking are used interchangeably. According to Murtonen (2019), critical thinking is considered as a sub-component of, and a foundation for scientific thinking. In addition, scientific thinking is commonly used "… to describe evidence-based thinking in science, social science, humanities, education, and business" (Murtonen, 2019, p. 65).

Furthermore, scientific thinking is defined in terms of reasoning strategies, thinking characteristics, knowledge-seeking, coordination of theory and evidence, and conceptual change (Kuhn, 2010). The previous literature has conceptualised scientific thinking as a dimension of cognitive skills and disposition (Facione, 1990). Cognitive skills involve the activities of interpreting, analysing, evaluating, drawing conclusions, explaining, and self-regulation. Meanwhile, Siegel (1988) summarised the dispositional component as 'critical spirit' which is "… taken as one that respects authorities of truth within reason, but does not trust authority blindly; this critical spirit is challenging of authority but does not regard one's own ideas as the sole authority" (Bezuidenhout, 2011, p. 18). Facione (1990, p. 96) pointed out several affective dispositional components of scientific thinking, such as "seeking a clear statement of questions or problems, curiosity in exploring problems and seeking information, trust in the processes of reasoned inquiry, willingness to use credible sources, etc." He argued that a good scientific thinker would have these cognitive skills and some or all of the affective dispositions. In so doing, such students would be able to solve problems in their learning, perform better and more confidently in the classroom, and be able to use their reasoning ability to approach everyday problems.

According to Zimmerman (2007, p. 213), scientific thinking consists of “a complex set of cognitive and metacognitive skills.” It is crucial that students acquire and develop these skills through their learning process, which requires long-term practices. More specifically, scientific thinking skills are the mental processes used when reasoning about scientific content or engaging in scientific investigations such as conducting scientific experiments (Holyoak & Morrison, 2005), as the process of knowledge acquisition. Hence, the skills of scientific thinking are essential as a vehicle for conducting scientific inquiry and discovery. Research on scientific thinking involves an investigation of human thinking related to scientific content that requires a high level of cognitive processes in constructing and acquiring knowledge in an effort to make sense of the world.

Educational institutions should be responsible for teaching and promoting the skills of scientific thinking among students in order to achieve educational goals. As noted by Schmaltz et al. (2017), one of the goals of education is to help students develop the skills needed to support their learning process, including scientific thinking skills. More specifically, scientific thinking has important implications for science education across all levels of education programs, starting from pre-kindergarten through to university level (Zimmerman & Klahr, 2018). By promoting and enhancing students' scientific thinking, educators can ensure that their students will develop the ability to think like a scientist (Schmaltz et al., 2017), and conduct scientific activities, "such as exploring, asking questions, testing hypotheses, engaging in inquiry, and evaluating evidence" (Zimmerman & Klahr, 2018, p. 3). Practicing the skills of scientific thinking through scientific inquiry enables students to enhance their scientific understanding (Kuhn, 2010). In so doing, understanding scientific concepts is related to the thinking process carried out by students aimed at truly understanding the phenomena that occur around them (Arends, 2012). More specifically, students' scientific thinking contributes to their ability to understand physics concepts (Coletta & Phillips, 2015).

According to Zimmerman (2007, p. 173), scientific thinking is related to scientific inquiry, scientific reasoning, and problem-solving. Scientific thinking is also linked with “generating, testing, and revising theories” to acquire knowledge or understand scientific phenomena in the real-world. Scientific reasoning skills are closely related to the ability to coordinate theory and evidence (Ding, 2018; Ibrahim, Ding, Mollohan, & Stammen, 2015), where the skills of scientific reasoning are considered as a core component of scientific thinking (Ding, 2018; Kuhn, 2010). Kuhn (2010) further pointed out that the coordination of theory and evidence can also be studied with respect to epistemological beliefs. In a similar vein, Zimmerman (2007) highlighted that scientific thinking is described as involving both cognitive processes and epistemological understanding. Other researchers have stated that students' epistemological beliefs are considered as premises of scientific thinking (Hyytinen, Holma, Toom, Shavelson, & Lindblom-Ylänne, 2014). According to Kuhn (2010, p. 6) and Ding (2018, p. 1482), there are four stages of scientific thinking, namely inquiry, analysis, inference, and argument. Other literature shows that scientific thinking is conceptualised by philosophers of science as a process of argumentation (Yang, 2004). As also noted by Rahayu and Widodo (2019), argumentation skills play an important role in scientific thinking, and are the skills used to communicate empirical and causal explanations. The ability to present evidence is the foundation of argumentation skills. Previous researchers have also explored various aspects of scientific thinking such as problem-solving, hypothesis testing, reasoning strategies, argumentation skills, and epistemological beliefs related to the nature of science (Dunbar & Fugelsang, 2005; Kuhn, Iordanou, Pease, & Wirkala, 2008; Osborne et al., 2016). Overall, scientific thinking covers a combination of specific cognitive variables.

Simply put, the research literature has demonstrated that the skills of scientific thinking cover a collection of intellectual skills that are broad and varied. In the present study, the term scientific thinking is used in relation to several cognitive variables that focus on scientific reasoning, argumentation, and epistemological beliefs, in which scientific reasoning is considered different from argumentation (i.e., in the scientific context) (Berland & McNeill, 2010; Opitz, Heene, & Fischer, 2017). These three aspects of scientific thinking are believed to play a central role in science education, specifically, in enhancing students' understanding of scientific concepts (Berland & McNeill, 2010; Bybee & Fuchs, 2006; Coletta & Phillips, 2015; Ding, 2014c; Osborne et al., 2016; Zimmerman, 2007), as well as helping them to succeed, particularly in learning physics (Chen, Wang, Lu, & Hong, 2019). A description of each aspect of scientific thinking investigated in this research is presented in the following section.

2.4.1 Scientific Reasoning

In the literature, the study of students' scientific reasoning is commonly related to science, technology, engineering, and mathematics (STEM) education. Scientific reasoning has been studied as one of the main goals of science education (Coletta, Phillips, & Steinert, 2012; Ibrahim et al., 2015; Krell, Redman, Mathesius, Krüger, & van Driel, 2018). More specifically, physics education reform emphasises that it is crucial to promote and enhance students' scientific reasoning to overcome the complex challenges of the 21st century (Anderman, Sinatra, & Gray, 2012; Collins, 2014).

Scientific reasoning has been variously defined by different theorists and researchers. According to Han (2013), research on scientific reasoning was first initiated within cognitive development studies of "formal reasoning" (Piaget, 1965). In most studies, scientific reasoning has been defined as formal reasoning. This study also refers to formal reasoning that is commonly related to the academic domain instead of informal reasoning which is associated with everyday situations. In the previous literature, scientific reasoning represents the whole process of scientific inquiry such as asking questions, exploring a problem systematically, formulating and testing scientific hypotheses, making predictions, controlling and manipulating variables, conducting an experimental design, analysing and evaluating experimental outcomes, drawing valid conclusions, and developing an empirical law (Bao, Fang, et al., 2009; Lazonder & Wiskerke-Drost, 2015; Zimmerman, 2007). According to Lawson (2004, p. 308), the skills of scientific reasoning involve mental strategies, plans, or rules used to process information and draw inferences related to a scientific phenomenon beyond direct experience. While there is a range of understandings of what constitutes scientific reasoning, the literature seems to agree that scientific reasoning represents an important component of scientific inquiry used when conducting scientific investigations and supporting experimentation in order to evaluate and draw valid conclusions that lead to forming and modifying concepts and theories.

Scientific reasoning skills are also closely related to the ability to coordinate between theory and evidence. As noted by Ibrahim et al. (2015, p. 94), "a model example of theory evidence coordination includes reflecting on prior theories, searching for evidence that conflicts with one's existing theories and eliminating alternative explanations or misconceptions." Students tend to use evidence obtained through investigations to support existing theory instead of refuting the theory (Ibrahim et al., 2015). They also tend to modify the evidence to fit the theory when a conflict arises between theory and evidence obtained from experimental results. In such cases, they might repeat the experiment in an effort to reconcile the existing theory with the results of the experiment. Hence, students' scientific reasoning skills play a crucial role in coordinating theory and evidence. It is worth noting that students who have sophisticated levels of scientific reasoning skills tend to be able to differentiate and coordinate theory and evidence.

In short, the skills of scientific reasoning are those that are needed to carry out scientific investigations, to coordinate theory and evidence, and are closely related to thinking processes that lead to an enhanced understanding of the phenomenon being observed (Han, 2013; Ibrahim et al., 2015; Kuhn & Dean Jr, 2004; Lawson, 2000a). In the present study, the skills of scientific reasoning being investigated represent a set of skills referred to in Lawson's work (2000a), including the measurement of conservation of weight and volume, proportional reasoning, probability reasoning, correlational reasoning, and hypothetical-deductive reasoning, as well as controlling multiple variables to draw valid conclusions.

As stated by Han (2013), a better understanding of the nature of scientific reasoning requires extended knowledge of the process of scientific inquiry. Scientific inquiry is part of the work of a scientist and is also used in the student-centred classroom, where students are actively engaged in solving problems. Scientific reasoning skills are key intellectual skills that are crucial for students to successfully carry out scientific investigations in order to make sense of the phenomena being studied. Fostering students' scientific reasoning is a major goal in science education. This skill is essential for enabling students to acquire new knowledge, utilise scientific information to support their learning, and to solve complex scientific problems in the real-world (Engelmann, Neuhaus, & Fischer, 2016; Wenning & Vierya, 2015), in addition to handling real-world tasks in future careers, particularly those in the science, technology, engineering, and mathematics (STEM) area (Bao, Cai, et al., 2009). Hence, scientific reasoning skills are key determinants of the success of students at school and in their social life (Ding, Wei, & Liu, 2016; Lawson, 2005), and teachers should help and encourage them to have strong scientific reasoning skills.

Several instruments have been developed and employed to measure students' scientific reasoning including the *Group Assessment of Logical Thinking Test* (GALT) (Roadrangka, Yeany, & Padilla, 1982), the *Test of Logical Thinking* (TOLT) (Tobin & Capie, 1981), and *Lawson's Classroom Test of Scientific Reasoning* (LCTSR) (Lawson, 2000b). Among the various existing assessment tools, the *Lawson's Test* is a well-known and standard instrument for assessing students' scientific reasoning, especially in the science education community, and is considered a reliable and valid instrument (Bao, Fang, et al., 2009). The initial *Lawson's Test* has been revised multiple times with the latest version being released in 2000 (Ding, 2018). The *Lawson's Test* can be used to measure students' scientific reasoning at various academic levels, including at secondary school and university (Ding, Wei, & Liu, 2016). The present study assessed pre-service physics teachers' scientific reasoning using the most recent version of *Lawson's Classroom Test of Scientific Reasoning* (LCTSR). A detailed explanation regarding the *Lawson's Test* is presented in Chapter 3.

A number of previous studies have assessed students' scientific reasoning skills; however, the findings were still not encouraging. For instance, both secondary school and university students have not shown adequate levels of scientific reasoning indicating that it needs to be improved (Jufri, Setiadi, & Sripatmi, 2016; Khoirina, Cari, & Sukarmin, 2018; Rosdiana et al., 2019). Furthermore, research on the role of gender differences in students' scientific reasoning has shown mixed results. For instance, some previous research has demonstrated that gender differences had no impact on students' scientific reasoning (Novia, Syamsu, & Riandi, 2018; Piraksa, Srisawasdi, & Koul, 2014; Talib et al., 2018). In contrast, other research has indicated that male students outperform female students on the skills of scientific reasoning (Coletta et al., 2012; Nieminen, Savinainen, & Viiri, 2013). Meanwhile, research at the tertiary level in relation to the development of students' scientific reasoning assessed through the *Lawson's Test* has not demonstrated satisfactory results. For example, a study on Chinese university students reported that students' scientific reasoning did not show significant improvement over the four year levels (Ding, 2018; Ding, Wei, & Liu, 2016).

Nevertheless, a number of researchers have found that scientific reasoning is a significant predictor of students' learning performance. For instance, Han (2013) argued that teaching and training students in scientific reasoning has a long-term impact on their science achievement. In addition, students who have good scientific reasoning skills are likely to have high academic achievement (Lawson, Banks, & Logvin, 2007). Scientific reasoning skills have also been found to contribute to students' conceptual understanding of physics (Coletta & Phillips, 2015; Ding, 2014c; Nieminen et al., 2012; Pyper, 2012). Specifically, students with sophisticated levels of scientific reasoning tend to have a better conceptual understanding of physics. This could be explained as they can conduct scientific inquiry that requires a deep-thinking process, which includes designing experiments, generating and testing scientific hypotheses, controlling variables, analysing experimental outcomes, and drawing conclusions in order to make sense of the phenomena being studied. They are also able to coordinate theory and research results (or evidence) which, in turn, leads them to a deep understanding of concepts. Conversely, students who lacked scientific reasoning were found to be more likely to have limited success in their physics course (Coletta & Phillips, 2015; Sadler & Zeidler, 2005b).

Other studies have also examined the possible relationship arising between scientific reasoning and epistemological beliefs. For instance, Zeineddin and Abd-El-Khalick (2010) found that students' epistemological beliefs influenced their scientific reasoning skills. Students with sophisticated epistemological beliefs tended to have a high level of scientific reasoning. Furthermore, Acar, Patton, and White(2015) noted that “fostering argumentation would enhance student scientific reasoning” (p. 132). This suggests that there is a relationship between students’ scientific reasoning and argumentation. The existing literature offers few explanations about how scientific reasoning skills correlate with argumentation or epistemological beliefs, how each variable plays a role in affecting students' conceptual understanding of physics, and how these variables relate to one another remains unclear. Therefore, this study explores the relationships arising between these multiple variables in an integrated way to fill the gaps in the literature. The context of argumentation and epistemological beliefs are explained in the next section.

2.4.2 Argumentation

Over the past few decades, numerous studies have paid increased attention to the investigation of students’ argumentation skills in scientific contexts (Aydeniz & Dogan, 2016; Eskin & Ogan-Bekiroglu, 2013; Osborne et al., 2016). Enhancing student argumentation is one of the major learning objectives in science education at all levels of academic education (Wang & Buck, 2016). Argumentation is considered by science education researchers as one of the essential skills that students need to be able to participate in logical debate, construct and evaluate scientific knowledge and coherent arguments, and present scientific ideas and draw scientific conclusions (Eskin & Ogan-Bekiroglu, 2013; Lancaster & Cooper, 2015). As stated by Toulmin and colleagues (1984, as cited in Bathgate, Crowell, Schunn, Cannady, & Dorph, 2015, p. 1593), argumentation is a core element of scientific thinking. In addition, having the skills of argumentation enables students to generate and defend claims to explain the complex phenomena around them, and to think like a scientist (Buber & Coban, 2017). Claims are conclusions about scientific phenomena being investigated to make sense of the world, and are commonly supported by evidence or scientific data. Justification is also required using appropriate scientific principles for the connection between knowledge claims and empirical data (McNeill & Knight, 2013).

In the literature, the term argumentation is defined in various ways. For instance, argumentation is described as “a social process of constructing, supporting, and critiquing claims for the purpose of developing shared knowledge” (Manz, 2015, p. 2). Argumentation also refers to the process of dialogue between two or more people who engage in debating opposing claims, where they might hold different positions on controversial topics (Kuhn & Udell, 2003). According to Erduran and Jiménez-Aleixandre (2008, p. 13), argumentation is “the connection between claims and data through justifications or the evaluation of knowledge claims in light of the evidence, either empirical or theoretical." Meanwhile, Toulmin et al. (1984) defined argumentation as “the whole activity of making claims, challenging them, backing them up by producing reasons, criticizing those reasons, rebutting those criticisms, and so on” (p. 14). In addition, Toulmin proposed an argumentation framework which is recognised as Toulmin’s Argument Pattern (TAP) as illustrated in Figure 2. 2.

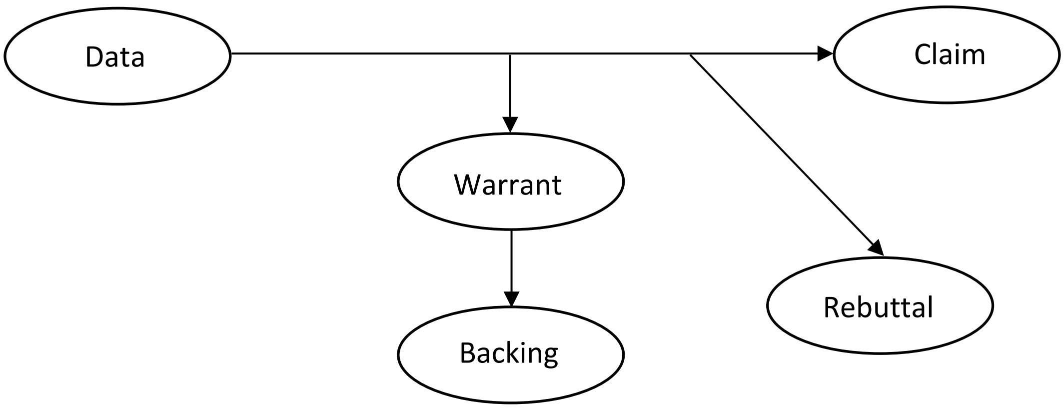


Figure 2. 2 Toulmin’s Argument Pattern

Adapted from Erduran, Simon, and Osborne (2004, p. 918)

Basically, argumentation is an activity in which one needs to put forward the data, claim, warrant, rebuttal, and backing. Based on the Toulmin model, arguments include claims which are conclusions whose merits are to be established; data are the facts used as evidence to support the claims; warrants are the reasons justifying the connection between the data and the claims; backings are basic assumptions that provide the justification for particular warrants; and rebuttals indicate specific circumstances under which claims are incorrect and invalid. Finally, the claims may include qualifiers that indicate specific circumstances under which the claim is true (Driver, Newton, & Osborne, 2000, p. 293).

The interconnection between these components of argumentation has been used to facilitate conceptualisation of the quality of argumentation and complexities of the nature of students' arguments (Venville & Dawson, 2010). Sampson and Clark (2006) explained that a stronger argument would contain more components of the argument in contrast to a weaker argument (e.g., might only contain a claim). The quality of students' argumentation is also shown in relation to the presence or absence of rebuttals in their arguments (Erduran et al., 2004). Examining students' argumentation skills is not an easy task. Researchers and educators have experienced difficulties in analysing arguments (Erduran, 2007). Meanwhile, assessing the quality of students' argumentation using the Toulmin model might require deep understanding and practice. Therefore, argumentation skills in the context of the present study refer to the works of Sampson and Clark (2006). Students' argumentation skills in scientific contexts are assessed using the *Argumentation Test* that examines their ability to make a scientific argument and to challenge arguments, as described in Chapter 3.

Further to this, the skill of argumentation is a critical element in teaching and learning practices in science education (Bathgate et al., 2015; Chen et al., 2019; Osborne et al., 2016), due to the nature of science being basically argumentative. In addition, these researchers argued that students who engage in argumentative activities in the science class tend to be able to learn more scientific knowledge than their peers who do not do this. The National Science Standards emphasise that engaging in argumentation can promote students' understanding of scientific concepts (Heng, Surif, & Seng, 2014). In the same vein, Newton and colleagues (1999, p. 554) argued that students' argumentation contributes to their deep conceptual understanding, with the authors pointing out that:

> Talking offers an opportunity for conjecture, argument, and challenge. In talking, learners will articulate reasons for supporting particular conceptual understandings and attempt to justify their views. Others will challenge, express doubts and present alternatives, so that a clearer conceptual understanding will emerge.

Clearly, involving students in argumentation enables them to engage in the construction of knowledge and develop their scientific understanding. Kuhn and Udell (2003) further argued that engaging in argumentative discussion could enhance conceptual understanding of subject matter in students of school-age children through to college.

There is no denying that argumentation is a crucial skill that must be developed in students, so they can be successful in academic and social life, and be able to compete in the 21st century world

(Frey, Ellis, Bulgren, Hare, & Ault, 2015). These skills could replace the focus of science learning from memorising knowledge to engaging students in complex scientific practices in which they construct and justify knowledge claims (Berland & McNeill, 2010). By offering students the opportunity to engage in argumentation, knowledge can be articulated, reflected upon, and modified. When students engage in argumentation which is part of the thinking process, they take an active role in constructing knowledge to support their claims or criticising other different ideas which, in turn, contribute to their conceptual understanding related to the scientific phenomena being studied (Buber & Coban, 2017). Clearly, argumentation is the skill needed to connect evidence and claims where students are involved in justifying their claims using scientific principles to support their arguments. Through this process, students will naturally gain a deep conceptual understanding of scientific concepts and strengthen their content knowledge.

A number of previous studies have assessed students' argumentation skills in relation to demographic factors and other learning variables. For instance, research on the role of gender differences in students' argumentation skills have shown mixed results. Some previous research has demonstrated that gender differences had no impact on students' argumentation skills (Chen et al., 2019; Widodo, Waldrip, & Herawati, 2016). In contrast, other research indicated that male students outperformed female students in generating a higher quality of argumentation (Hong, Lin, Wang, Chen, & Yang, 2013; Salminen & Marttunen, 2018). In addition, male students tended to actively generate more rebuttal than female students (Jeong & Davidson-Shivers, 2006, as cited in Chen et al., 2019). As summarised by Chen et al. (2019), female students tend to rarely criticise and evaluate others' views, but they prefer to show empathy, acceptance, support, and cooperation with others, while male students tend to be more critical in assessing others' points of view. Therefore, female students are likely to experience more difficulties, discomfort, and lack of confidence in generating scientific arguments that require a high-level of thinking processes and dialogue while learning science.

Furthermore, several previous studies have also assessed high school students' argumentation skills in relation to grade levels. Some studies found that there was no significant difference in the skills of argumentation of students across grade levels (Kaya, Erduran, & Cetin, 2012; Widodo et al., 2016). In contrast, Osborne et al. (2016) found that the higher the grade level of secondary school students, the higher their level of argumentation skills. In addition, previous empirical evidence has indicated that the quality of students' argumentation skills and ability to engage in argumentation activities were relatively low (Hsu, Van Dyke, Chen, & Smith, 2016; Putri & Rusdiana, 2017), meaning that they were poor at coordinating and constructing the essential relationships between evidence (or scientific data) and claim (or conclusion) (Lancaster & Cooper, 2015, p. 4).

The lack of argumentation practices due to the limited competence of teachers to develop argumentation activities in the classroom, is considered as one of the causes of this phenomenon (Buber & Coban, 2017). In turn, students lack the opportunity to participate in generating logical arguments, and presenting their ideas in argumentation activities, such as scientific discussion. In a similar vein, Kaya et al. (2012) noted that the limitations of the teachers' pedagogical skills to organise or design teaching and learning in the classroom that support the practice of argumentation, could affect students' ability to construct complex arguments and actively engage in argumentative discourse. In addition, time constraints and over-loaded curricula are also seen as barriers to creating effective classroom argumentation practices. Limited instructional resources also become an obstacle for teachers to engage students in argumentation practices in the classroom; therefore, teachers tend to focus on transferring content knowledge or explaining concepts instead of promoting students' argumentation skills (Sampson & Blanchard, 2012).

Nevertheless, a number of previous studies have demonstrated that students' argumentation skills contribute to their understanding of scientific concepts (Aydeniz & Dogan, 2016; Çınar & Bayraktar, 2014; Nurlatifah et al., 2018; Osborne et al., 2016). Venville and Dawson (2010) demonstrated a significant relationship between argumentation and conceptual understanding of students by showing that when students engaged in the process of argumentation, they showed a better understanding of the subject matter than those who did not participate in classroom-based argumentation. Students involved in argumentative practices are more likely to be engaged more deeply with the subject matter being studied and it could be expected that they would develop a better understanding of the subject matter than if they memorised facts and information. Cross and colleagues (2008) argued that "… engaging in argumentation leads to a more secure understanding of pre-existing concepts, but also allows students to hear new ideas that extends their existing knowledge and possibly eliminates misconceptions." Cross et al. (2008) further revealed that students who were involved in the process of argumentation through a computer-based program were able to address "misconceptions and develop better understandings" (p. 839). Clearly, having better argumentation skills allows students to consolidate their prior knowledge and construct new knowledge based on the ideas of others while enhancing their conceptual understanding of subject matter (Driver et al., 2000).

In addition, Venville and Dawson (2010) noted that when engaging students in argumentation activities in the classroom, the important factors to be considered were their prior experience and content knowledge. In a similar vein, some other studies agree that students' prior scientific content knowledge should be taken into account when engaging them in the argumentation process (Çelik & Kılıç, 2014; Eskin & Ogan-Bekiroglu, 2013; Osborne et al., 2016). Students' prior scientific knowledge and experience might affect the quality and complexity of the arguments generated and the way they justify scientific claims about specific phenomena being studied. In addition, without having an understanding of the scientific content, students might find it difficult to make links between claims and scientific data, to understand the ideas being presented, to elaborate ideas, and to develop new ideas or views on a given topic which, in turn, might affect their willingness to take part in argumentative activities or to share ideas (Bathgate et al., 2015).

Further to this, some previous researchers have established the impact of argumentation on students' scientific reasoning (Acar et al., 2015; Heng et al., 2014). The skills of argumentation are essential for success in conducting scientific inquiry (Chen, Wang, Lu, Lin, & Hong, 2016; Lancaster & Cooper, 2015), which is closely related to the development of scientific ideas and theories (Osborne et al., 2016). As mentioned previously, when conducting scientific inquiry, one needs scientific reasoning skills. This indicates that the skills of argumentation contribute to students' scientific reasoning.

Hence, the science classroom should not only teach students about scientific concepts, but also encourage students to engage in an argumentation discourse environment, where they need to support their claims by using evidence and justifying their ideas with rational scientific explanations (Sampson & Blanchard, 2012; Taasoobshirazi & Hickey, 2005). The role of instructors is therefore essential to help students improve their argumentation skills by developing a supportive learning environment and facilitating argumentative activities in the classroom. Yet, previous researchers have argued that argumentation is rarely integrated into teaching and learning practices across all academic educational levels (Chen et al., 2016; Sampson & Blanchard, 2012), because teachers may be reluctant to adopt argumentation practices in their class teaching (Wang & Buck, 2016). In addition, both pre- and in-service science teachers have been found to demonstrate a lack of knowledge about argumentation (Zohar, 2007, as cited in Wang & Buck, 2016). Wang and Buck (2016) further argued that the practice of argumentation would not take place in the classroom unless teachers have advanced knowledge and strong understanding of the argumentation process and know how to engage students in argumentation practices.

The empirical evidence about the extent to which argumentation skills contribute to student learning of science is still lacking (Bathgate et al., 2015). In addition, the study of argumentation for university students is still rare and there is a need for more research to be conducted. Specifically, the review of the literature indicates that limited attention has been paid to investigating prospective physics teachers' argumentation skills and the role that might be played by demographic factors such as gender, year level, and university type, particularly in the Indonesian context, and how argumentation skills contribute to pre-service physics teachers' conceptual understanding of physics. In addition, the literature has not adequately examined the extent to which argumentation skills correlate with other cognitive variables in an integrated manner. The present study seeks to bridge this gap in the literature, by examining the relationship between argumentation skills and epistemological beliefs. The following section describes epistemological beliefs.

2.4.3 Epistemological Beliefs

Over the last few decades, students' epistemological beliefs have become an active area of research within the science education research community. This cognitive variable is believed to influence students' conceptual learning in science, and their academic achievement, motivation, and intellectual development. Developing students' epistemological beliefs is one of the major goals of science education (Aslan, 2017; Bigozzi et al., 2018; Chen, Xu, Xiao, & Zhou, 2019; Kirmizigul & Bektas, 2019; Pamuk, Sungur, & Oztekin, 2017; Yang, Huang, & Tsai, 2016).

In the literature, epistemological beliefs have been variously defined. As noted by Aslan (2017, p. 38), "epistemological beliefs are those that the individual has on what knowledge is, how it is acquired, and what criteria determine knowledge." Epistemological beliefs are related to "beliefs about the definition of knowledge, how knowledge is constructed, how knowledge is evaluated, where knowledge resides, and how knowing occurs" (Hofer, 2001, p. 355). According to Hofer and Pintrich (1997), epistemological beliefs refer to the individual's beliefs about the nature of knowledge and the nature of knowing.

Historically, in 1970, William G. Perry was credited as being the pioneer who studied individual's epistemological beliefs and developed a model describing the development of epistemological beliefs (Brownlee, Purdie, & Boulton-Lewis, 2001; Hofer, 2000; Hofer & Pintrich, 1997). In Perry's model, "individuals move through some specified sequence in their ideas about knowledge and knowing, as their ability to make meaning evolves" (Hofer, 2001, p. 356). Perry developed a unidimensional model of epistemological beliefs, proposing four stages of intellectual development: 1) a *dualistic* view that is characterised by dichotomies such as right and wrong; this view characterises knowledge as being certain and that experts provide the right answers; 2) a *multiplicity* view, where there is an acknowledgement of multiple viewpoints and the possibility of uncertainty, whereby individuals may see conflicting views. This view holds the belief that knowledge is subjective; 3) a *relativistic* worldview acknowledges that some viewpoints are better than others; 4) individuals at the *commitment within relativism* acknowledge that there is no absolute or certain knowledge, and they have a more complex level of beliefs (Hofer, 2001, p. 357). Furthermore, Perry's research showed that undergraduate students' epistemological beliefs in the first year of university study tended to show that they believe in simple and unchangeable facts, and they assume that knowledge comes from omniscient authorities. By their final year, they have become relativistic thinkers who believe that knowledge is complex, tentative, and derived from reason.

Meanwhile, Schommer (1990) proposed a multidimensional model in which the dimensions of epistemological beliefs consist of more than one independent dimension. In this description, each dimension of epistemological beliefs can develop independently of the other dimensions. Schommer proposed five dimensions of epistemological beliefs comprising omniscient authority which is the belief that authorities have access to otherwise inaccessible knowledge; certain knowledge which is the belief that absolute knowledge exists and will eventually be known; simple knowledge which is the belief that knowledge consists of discrete facts; quick learning which is the belief that learning occurs in a quick or not-at-all fashion; and innate ability which is the belief that the ability to acquire knowledge is endowed at birth (Schraw, 2013, p. 2).

Other researchers have also developed epistemological belief models inspired by Perry's work. For instance, Hofer and Pintrich (1997) proposed four dimensions to identify students' epistemological beliefs that are divided into two groups. The first group is the nature of knowledge comprising the certainty of knowledge and the simplicity of knowledge dimensions. The second group is the nature of knowing consisting of the source of knowledge and the justification for knowledge dimensions. Meanwhile, Elby (2001) proposed five dimensions to investigate students' epistemological beliefs about the nature of knowledge and learning in the physical sciences, namely the structure of scientific knowledge, the nature of knowing and learning, real-life applicability, evolving knowledge, and the source of the ability to learn. A number of dimensions of epistemological beliefs have been proposed by Stathopoulou and Vosniadou (2007), namely the structure of physics knowledge, the stability of physics knowledge, the source of physics knowledge, and the justification of physics knowledge.

The dimensions of epistemological beliefs proposed by these researchers could represent and distinguish the level of sophistication of students' epistemological beliefs as classified as either naive or sophisticated epistemology. The development of students' epistemological beliefs from lower to higher levels, or from naive to more sophisticated levels, could influence how they perceive the knowledge taught in the classroom. For instance, students with naive epistemological beliefs commonly believe that knowledge is simple, absolute, and acquired quickly. Knowledge is also believed to be formed from a collection of facts or formulae and then handed down by the authorities or experts, and that the learning process is rapid and does not occur gradually. Students who have naive epistemological beliefs tend to memorise knowledge or facts rather than understand them. When they face challenging situations in their learning environment, these students prefer to avoid obstacles, and use ineffective learning strategies or adopt surface learning approaches (Aslan, 2017; Hasene & Şekercioğlu, 2018; Kirmizigul & Bektas, 2019; Schommer, 1994a; Winberg, Hofverberg, & Lindfors, 2019).

In contrast, students with sophisticated epistemological beliefs believe that knowledge is uncertain, coherent, complex, and is acquired gradually through experience. They believe that knowledge is changeable, tentative and constantly evolving, and subjective. Furthermore, they tend to construct knowledge through experience to gain a deeper understanding, actively engage in learning practices, and adopt deep learning approaches (Aslan, 2017; Hasene & Şekercioğlu, 2018; Kirmizigul & Bektas, 2019; Noroozi & Hatami, 2018; Pamuk et al., 2017; Schommer, 1994a; Winberg et al., 2019). Therefore, students who have sophisticated epistemological beliefs tend to perform better in their learning and are more successful in learning science than those who have naive epistemological beliefs. Kirmizigul and Bektas (2019) concluded that students who have sophisticated epistemological beliefs are likely to be able to explain and apply scientific knowledge in their everyday life and be able to think like a scientist who can implement scientific methods to verify theories or facts.

Further to this, the previous literature has also indicated that sophisticated epistemological beliefs are closely related to constructivist teaching views and adopting more student-centred teaching orientations, while naive epistemological beliefs are more likely to be associated with traditional pedagogical views and adopting more teacher-centred teaching orientations. Pre- and in-service teachers should have sophisticated epistemological beliefs because they have a responsibility to help their students enhance their epistemological beliefs (Kirmizigul & Bektas, 2019). Therefore, it is important to investigate prospective teacher's beliefs about knowledge and knowing because these beliefs will likely affect their future teaching approaches.

A number of surveys have been developed and widely used by previous researchers and educators to assess students' epistemological beliefs. These instruments include the *Views about Science Survey* (VASS) designed by Halloun (1997), the *Maryland Physics Expectation* (MPEX) developed by Redish, Saul, and Steinberg (1998), the *Epistemological Beliefs Assessment about Physical Science* (EBAPS) developed by Elby, Frederiksen, Schwarz, and White (1999), the *Colorado Learning Attitudes about Science Survey* (CLASS) designed by Adams, Perkins, Dubson, Finkelstein, and Wieman (2005), and the *Greek Epistemological Beliefs Evaluation Instrument for Physics* (GEBEP) developed by Stathopoulou and Vosniadou (2007). Among the various surveys mentioned, the EBAPS survey is used more widely by researchers or educators in the field of physics or science education (Kortemeyer, 2007). In the present study, the EBAPS survey has been employed to probe pre-service physics teachers' epistemological beliefs, as described in Chapter 3.

A number of previous studies have assessed students' epistemological beliefs in relation to demographic factors and other learning variables. In terms of the role of gender differences in students’ epistemological beliefs, the previous research has shown mixed results. For instance, several studies have indicated that gender differences had no impact on students' epistemological beliefs (Chen et al., 2019; Efilti & Çoklar, 2016; Tumkaya, 2012; Yalcin & Yalcin, 2017; Yenice, 2015). In contrast, other research has found significant differences between male and female students in relation to epistemological beliefs (Aslan, 2017; Kanadlı & Akay, 2019; Langcay, Gutierrez, Valencia, & Tindowen, 2019; Terzi, Çetin, & Eser, 2012). Furthermore, the relationship between year level and students' epistemological beliefs have also shown mixed results. For example, research at the tertiary level in relation to students' epistemological beliefs did not show significant improvement in beliefs over the four year levels (Ding & Zhang, 2016; Yalcin & Yalcin, 2017; Yenice, 2015). In constrast, other research showed that university students' epistemological beliefs had changed and developed over the time they had spent at university (Aslan, 2017; Belet & Guven, 2011; Schommer-Aikins & Duell, 2013). Ding and Mollohan (2015) pointed out that educators certainly expect their students' epistemological beliefs to move from novice views towards more expert understandings as they develop through their university course. The level of students' epistemological beliefs must be considered to be a factor that affects their learning. Hence, it is important to understand the levels of students' epistemological beliefs and their progression over time in order to move students' beliefs toward a more sophisticated orientation.

Further to this, a number of studies have been carried out to identify the relationships between epistemological beliefs and other cognitive variables. For example, Stathopoulou and Vosniadou (2007) examined the relationship between students' epistemological beliefs and their conceptual understanding of physics at secondary school. The researchers administered the *Greek Epistemological Beliefs Evaluation Instrument* (GEBEP) to collect data on students' epistemological beliefs, and the *Force and Motion Conceptual Evaluation Instrument* (FMCE) to measure students' conceptual understanding of physics. Their research indicated that students holding more sophisticated epistemological beliefs reflected a greater depth of physics understanding compared to students holding less sophisticated beliefs. Stathopoulou and Vosniadou also asserted the critical importance of epistemological beliefs on physics conceptual understanding, and that these beliefs should be considered in the teaching of physics. In the same vein, Franco et al. (2012) also found a relationship between students' epistemological beliefs and their conceptual understanding of physics; basically that students with more sophisticated epistemological beliefs had a deeper understanding of physics concepts than students with naïve epistemological beliefs.

Other researchers also found that epistemological beliefs have a positive impact on students' academic achievement. The studies showed that more sophisticated epistemological beliefs contributed to better science learning outcomes of students (Kaymak & Ogan-Bekiroglu, 2013; Kizilgunes, Tekkaya, & Sungur, 2009; Pamuk et al., 2017). Apart from this, a number of previous studies have also indicated a relationship between students' epistemological beliefs and their argumentation skills (Ku, Lai, & Hau, 2014; Noroozi & Hatami, 2018). Students with sophisticated epistemological beliefs tend to be able to generate complex arguments or express ideas and defend their opinions more than those who have naive epistemological beliefs. Furthermore, several previous studies have also demonstrated that students who have sophisticated epistemological beliefs tend to have better scientific reasoning skills (Hotulainen & Telivuo, 2014; Zeineddin & Abd-El-Khalick, 2010).

There is no denying that the development of epistemological beliefs in the classroom is crucial considering that these beliefs contribute to students' academic performance in science. Teachers should help students to develop more sophisticated epistemological beliefs. Specifically, instructors in the teacher education institutions need to teach and provide an opportunity for pre-service teachers to enhance their epistemological beliefs during their study in higher education. This is important considering that both in- and pre-service teachers have a major responsibility to manage and organise teaching and learning practices in their classrooms. Teachers with sophisticated epistemological beliefs are more likely to implement more effective teaching methods, use appropriate teaching materials and assessment or evaluation techniques in the classroom, and create more supportive learning environments to provide meaningful and lifelong learning for their students (Kirmizigul & Bektas, 2019; Pamuk et al., 2017). Likewise, Bigozzi et al. (2018) and Aslan (2017) noted that teachers' epistemological beliefs play a crucial role in the teaching practices they develop and implement in the classroom.

Much research has been done to investigate school students' epistemological beliefs; however, the literature indicates a lack of attention to pre-service physics teachers' epistemological beliefs, especially in the Indonesian context. In addition, there has been no research investigating the relationship between epistemological beliefs and other learning variables such as argumentation and scientific reasoning skills in an integrated manner. To fill the gap in this research area, the present study investigates pre-service physics teachers' epistemological beliefs and examines the relationships arising between epistemological beliefs and the other identified cognitive variables. This is important because pre-service teachers are the teachers of tomorrow who will affect their future students' academic outcomes.

2.5 The Conceptual Model Proposed for the Study

This section describes the conceptual model proposed in this study. The basis of the proposed conceptual model is grounded in the theories and research findings described above. In fact, prior research is still limited in providing evidence on how demographic factors (e.g., gender, year level, and type of university) and aspects of scientific thinking affect students' conceptual understanding of physics. To fill the gap in the literature, this study examines the relationships between these variables through employing Structural Equation Modeling (SEM) analysis.

Generally speaking, four main components are focused on in this study, consisting of epistemological beliefs, argumentation, scientific reasoning, and physics conceptual understanding (PCU). As mentioned earlier, these aspects of students' scientific thinking have a relationship with their conceptual understanding of physics. In addition, the epistemological beliefs and argumentation skills of students have been found to influence their scientific reasoning skills. Students' epistemological beliefs have also been found to have an impact on their argumentation skills. In previous studies, the relationship between these variables has primarily been investigated and analysed only through bivariate relationships that can be misleading when drawing conclusions, because bivariate relationships only examine the relationship between two variables without considering other variables that might have a greater effect.

In the present study, the variable of epistemological beliefs is predicted to have a direct effect on argumentation, scientific reasoning, and physics conceptual understanding (PCU). In other words, epistemological beliefs are hypothesised to have a direct effect on physics conceptual understanding and indirect effects mediated through argumentation and/or scientific reasoning. In addition, argumentation is hypothesised to have a direct effect both on scientific reasoning and physics conceptual understanding. The variable of scientific reasoning is hypothesised to be a variable that mediates the relationship between argumentation and physics conceptual understanding. Finally, scientific reasoning is predicted to have a direct effect on physics conceptual understanding. Thus, the proposed research model shows the relationship between these four main variables and their predicted direct and indirect effects, as presented in Figure 2. 3.

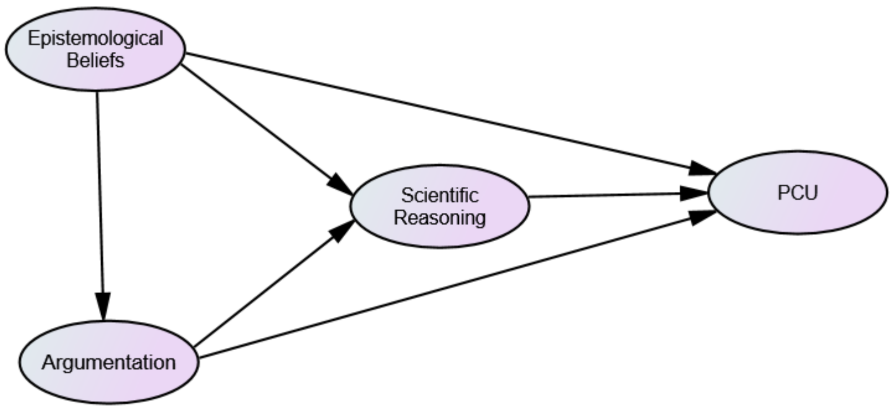


Figure 2. 3 The proposed research model of the relationships between research variables

In the proposed model, epistemological beliefs are the variable that must be addressed and developed first, as these are hypothesised to be a predictor for other constructs in the model. Students might construct knowledge based on their epistemological beliefs which would affect the way they generate arguments or critique other opinions. In addition, when conducting the investigations, students' epistemological beliefs are hypothesised to affect the way they coordinate their existing theories and evidence obtained from experiments or observations. In addition, to understand phenomena that occur in everyday life, students might also depend on existing epistemological beliefs that might be related to their previous experiences. Students' epistemological beliefs might evolve as they acquire more knowledge and experiences from their learning environment, so their level of epistemological beliefs is likely to change. The more sophisticated the level of their epistemological beliefs, the higher the level of their argumentation and scientific reasoning skills, as well as the more effective they are likely to be in understanding the scientific phenomena being studied.

To the best of the researcher's knowledge, studies on such aspects of scientific thinking (i.e., epistemological beliefs, argumentation, and scientific reasoning) for pre-service physics teachers in Indonesia remain limited. Therefore, there is a need to conduct research among pre-service physics teachers to understand the extent to which their skills of scientific thinking influence and possibly enhance their understanding of physics concepts. The model proposed in this study provides insights for instructors and researchers to comprehensively understand how students' cognitive variables relate to one another. Further to this, the existing research offers little explanation of how aspects of teaching and learning practices contribute to Indonesian students' scientific thinking and their understanding of physics concepts. The following section briefly describes the prior research investigating several aspects of teaching and learning that affect students' academic performance.

2.6 Aspects of Teaching and Learning Practices

According to Buber and Coban (2017), many researchers agree that it is crucial to develop students' conceptual understanding and scientific thinking. Furthermore, it is known that these learning variables are affected by a variety of aspects related to teaching and learning practices, demographic factors (e.g., age, gender, school type, major, year level, and students' culture), as well as other learning factors. A number of studies have examined various aspects of teaching and learning that affect students' academic outcomes. Understanding more about learning variables that affect students' academic achievement is important for educators in order to prepare effective teaching and learning practices in the classroom, and to overcome the barriers experienced by students. However, the factors that contribute most to students' learning outcomes remain unclear. This research attempts to fill this gap in the literature.

To improve students' learning outcomes in a higher education context in a gradual way, Biggs and colleagues developed the 3P model of teaching and learning over time (i.e., 1978, 1987, 1989, 1993, 1999, and 2003). The three critical components are presage, process, and product of teaching and learning. Figure 2. 4 presents the 3P model adopted from Biggs (2003). In the 3P model, Biggs (2003) described the presage component as being influenced by both student and teaching contextual factors. This component represents the attributes of students that exist before they enter the classroom. These include cognitive and non-cognitive factors, such as prior knowledge, intelligence or academic abilities, students' beliefs, motivation, and preferred approach to learning. Meanwhile, the teaching context includes the curriculum, teaching methods, assessment strategy, and classroom climate. These factors interact with one another to determine students' ongoing learning approach which, in turn, affects their learning outcomes. As shown in Figure 2. 4, the process component is influenced by the nature of the learning activities.

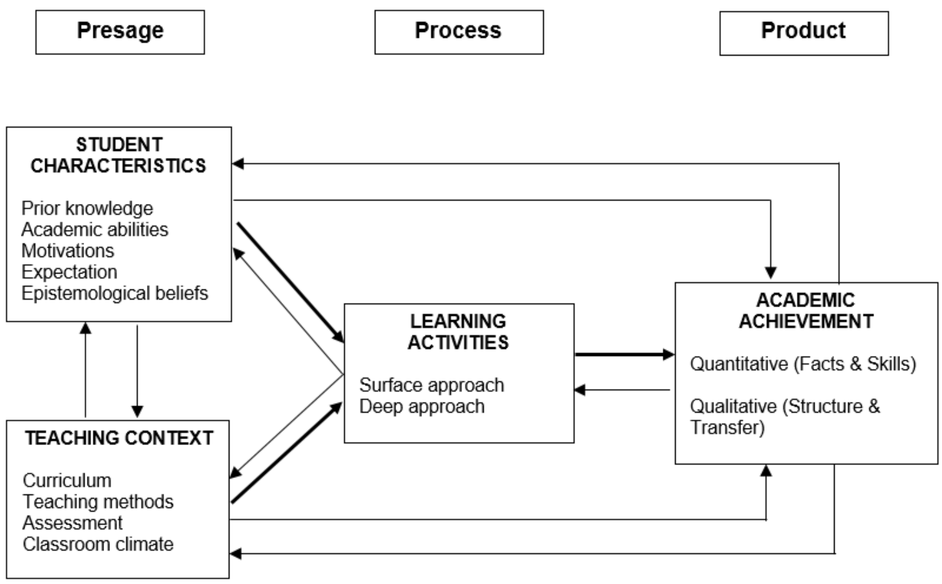


Figure 2. 4 An adapted version of Biggs' 3P model of teaching and learning (Biggs, 2003)

Biggs has a focus on two kinds of learning approaches, namely the surface learning approach and the deep learning approach. Historically, the concept of the learning approach was first conceptualised by Marton and Säljö (1976, as cited in Ozsevgec & Azakli, 2018), and can be described as a learning strategy that is adopted by students according to their perceptions in understanding the learning environment. While a surface approach involves the intention to reproduce the information in compliance with externally imposed task demands, a deep approach involves the intention to understand particular information. The process component represents an ongoing approach to learning, which relates to how students engage in the academic learning environment.

According to this model, students using a surface approach tend to study superficially, use rote learning and view the task as a demand to achieve a goal. On the other hand, students using a deep approach to learning tend to have good learning motivation, link prior information or knowledge with present information or new knowledge, and attempt to acquire a meaningful understanding. In order to fully achieve scientific conceptual understanding, students are encouraged to use deep learning strategies that engage them in meaningful learning (Cavallo et al., 2003). Finally, the product component represents students' academic achievement or other related learning outcome variables. Furthermore, Biggs et al. (2001) pointed out that the most effective way to ensure better teaching and learning practices is for teachers to take responsibility to ensure that contextual elements in the teaching and learning process are constructively aligned to promote deep learning approaches, which allow students to enhance their scientific thinking skills and conceptual understanding of the subject matter. Generally, the 3P model has been empirically examined and validated for almost all academic disciplines (Biggs, 2003); hence, this model can provide a valid and reliable framework to investigate teaching and learning practices in higher education, as well as to understand student learning. Since this study aims to investigate undergraduate students' academic performance in relation to particular aspects of scientific thinking (i.e., epistemological beliefs, argumentation, and scientific reasoning) and their conceptual understanding of physics, the 3P model provides the basis for conducting the qualitative part of the research.

Aspects of Teaching and Learning Influencing Students' Thinking Skills and Conceptual Understanding of Physics

Facilities and Learning Resources

Previous research has investigated various aspects of teaching and learning that are believed to positively influence students' thinking skills and conceptual understanding of physics. For instance, a number of researchers have pointed out various learning resources and facilities that are considered to play an important role in helping students to understand natural phenomena and conduct scientific inquiry, such as the availability of laboratory equipment (Chen, Chang, Lai, & Tsai, 2014; Galarpe, 2017; Husnaini & Chen, 2019). In addition, a conducive learning environment and classroom climate can have a positive impact on the effectiveness of teaching and learning practices in the classroom (Adeyemo, 2012; Mbunde, 2017). Other researchers have also revealed that the availability of textbooks and libraries plays an important role in supporting student learning (Ayaz, Ali, Khan, Ullah, & Ullah, 2017). Lack of facilities and learning resources, as well as inadequate infrastructure, could be a major challenge faced by teachers in providing effective teaching and meaningful learning for their students. As noted by Putri and Rusdiana (2017) and Lee and Sulaiman (2018), a lack of school facilities and inadequate laboratory equipment compels teachers to revert back to traditional and demonstration methods in the classroom which, in turn, have little impact in enhancing students' understanding and skills in learning science.

Teaching Methods

Teaching strategies or methods are another aspect of teaching and learning that contribute to students' academic outcomes. Several previous studies have demonstrated the improvement of students' conceptual understanding in physics or science subjects and their thinking skills through various teaching methods such as the guided constructivism teaching approach (Bigozzi et al., 2018), the interactive-engagement method (Kola, 2017; Von Korff et al., 2016), and inquiry-based instruction (Piraksa et al., 2014; Rusmansyah et al., 2019; Taasoobshirazi & Sinatra, 2011). More specifically, previous research has demonstrated that inquiry-based teaching methods are more effective than traditional teaching methods in improving students' scientific reasoning skills (Benford & Lawson, 2001). These researchers seem to agree that constructivist instructional methods or student-centered approaches that fully engage students in learning activities are more effective in improving their conceptual understanding and thinking skills compared to traditional teaching methods or teacher-centered approaches.

In traditional teaching approaches or direct instruction, teachers generally deliver subject matter according to textbooks, with students listening passively to the teachers' explanations while taking notes (Hairan et al., 2019; Lee & Sulaiman, 2018). Previous researchers have shown that traditional teaching methods do not promote students' conceptual understanding of physics at both school and university level, help them to overcome misconceptions about physics and learn adequate fundamental concepts of physics, and they fail to enhance students' thinking skills (Adolphus & Omeodu, 2016; Hairan et al., 2019; Sobremisana, 2017; Usmeldi, 2016). Bao, Fang, et al. (2009) also stated that traditional teaching instruction failed to promote students' scientific reasoning. Although it has been found that a student-centered approach is more effective than a teacher-centered approach in helping students to improve their academic performance (Dervic et al., 2018; Taasoobshirazi & Sinatra, 2011), teachers prefer to use traditional teaching methods (Kola, 2017; Lee & Sulaiman, 2018). This is usually due to time constraints and issues related to overloaded curricula, and to the practical consideration that it is easier to design lesson plans and organise the class in a teacher-directed way.

Learning Media

There is no denying that some physics phenomena found in everyday life seem to be abstract and difficult to understand for students, and sometimes teachers find it difficult to overcome this issue if they use traditional teaching practices. However, teaching and learning environments that are equipped with appropriate learning media may help students to visualise and understand scientific phenomena more easily. Referring to previous studies, learning media was also found to play an important role in enhancing students' learning outcomes.

For instance, Sobremisana (2017) found that learning using innovative physics devices can enhance students' understanding of physics concepts and their motivation to learn, compared with traditional teaching methods. However, the use of technology in teaching and learning practices, such as employing a virtual laboratory or interactive simulation package (e.g., the physics education and technology – PhET), has been found to be effective in improving students' conceptual understanding of physics and to promote learning independence in comparison with conventional methods (Arista & Kuswanto, 2018; Dervic et al., 2018; Eveline, Wilujeng, & Kuswanto, 2019; Faour & Ayoubi, 2018; Gunawan, Nisrina, Suranti, Herayanti, & Rahmatiah, 2018; Husnaini & Chen, 2019; Phanphech et al., 2019). Interactive learning media also facilitated students to visualise several abstract concepts in physics by providing opportunities for them to use computer animation in physics learning in areas where teachers find it difficult to describe certain phenomena (Gunawan et al., 2018). This approach attracted students' interest and helped them to explore scientific phenomena in a way that could not be carried out through experiments in traditional laboratories

(Faour & Ayoubi, 2018). The effectiveness of the use of a virtual laboratory or interactive simulation in supporting student learning might depend on the teachers' ability to assist students to explore and understand phenomena, as well as to encourage interaction among students to discuss the subject matter being studied. Furthermore, Putra (2019) demonstrated that teaching materials, such as student worksheets, were effective in deepening students' understanding of physics concepts and developing critical thinking skills.

Learning Activities in the Classroom

The learning activities that take place in class are also considered to be a factor that can affect students' academic performance. For instance, Lee and Sulaiman (2018) showed that students' understanding of physics concepts improved through the implementation of practical work or hands-on learning. The hands-on learning activities helped students to construct their knowledge, develop their thinking skills, and to conduct independent investigations which, in turn, enabled them to better understand physics concepts. Consequently, students have more opportunity to learn physics materials deeply when they actively engage in group discussions in the classroom (Adolphus & Omeodu, 2016; Von Korff et al., 2016).

However, Elby (1999, p. S56) indicated that "students spend more time focusing on quantitative activities involving formulae and practice problems, and less time focusing on qualitative activities involving concepts and real-life examples." Such learning activities are closely related to traditional teaching methods, which encourage students to receive lectures passively, and practice and solve a large number of traditional physics problems from the textbook. Learning practices that are more focused on the receiving and absorbing of content knowledge transferred by teachers in the classroom provide only limited opportunities for students to practice their scientific thinking skills (Heng et al., 2014).

Homework

In addition, the existing literature has found that homework assigned to students is another aspect that plays an important role in increasing students' academic achievement (Buijs & Admiraal, 2013; Grodner & Rupp, 2013; Gu & Kristoffersson, 2015; Suárez et al., 2016). These researchers agreed that homework is beneficial for students, giving them the opportunity to consolidate their understanding of the learning material that has been covered during class, and involving them in thinking deeply about the content knowledge being studied. In contrast, Bas, Senturk, and Cigerci (2017) revealed that homework did not benefit students' learning outcomes. Furthermore, homework may have negative effects on students due to promoting anxiety and excessive pressure, which creates a negative attitude towards the subject matter offered in the curriculum.

However, effectiveness in doing homework might also have a positive effect on student learning if teachers provide valuable feedback and consider the duration of the homework. Through critical feedback provided by their teachers, students have the opportunity to review their work, learn from their mistakes, and acquire a better understanding of the content knowledge being studied.

The Approach to Learning and Learning Assessment

As indicated in Biggs' 3P model (see Figure 2. 4), the approach to learning, whether it is surface or deep, is associated with students' academic outcomes in higher education. Adopting a surface learning approach is associated with rote learning or memorising ideas or information received passively without students making deep connections or interacting with the learning material. Broadly speaking, learners using the surface approach only focus on what is needed for examination or assessment and complete the assignments given by instructors by spending little time and effort, to simply get high grades or to pass exams (Alt & Boniel-Nissim, 2018; Efe & Aslan-Efe, 2018; Obura, 2019; Öhrstedt & Lindfors, 2019). Some learners using the surface approach may get high academic scores without mastering in-depth content knowledge (Obura, 2019). They are likely to have fewer opportunities to construct knowledge and actively engage in learning, or to practice and enhance their thinking skills. In contrast, students adopting a deep learning approach tend to focus on understanding content knowledge, developing and connecting ideas, associating prior knowledge with new knowledge, actively participating in meaningful learning, actively constructing knowledge, and thinking critically (Efe & Aslan-Efe, 2018; Ozsevgec & Azakli, 2018; Rozgonjuk, Saal, & Täht, 2018).

As mentioned earlier, the learning approach adopted by students correlates with their learning outcomes. Previous studies have revealed that a deep learning approach is positively correlated with academic achievement and success in learning, while the surface learning approach indicates a negative association with academic achievement and low success in learning (Jeong, González-

Gómez, Conde-Núñez, & Gallego-Picó, 2019; Karaman, Demirci, & Özdemir, 2019; Obura, 2019; Öhrstedt & Lindfors, 2019; Ozsevgec & Azakli, 2018; Rozgonjuk et al., 2018).

Students might adopt either a deep or a surface learning approach (or both) depending on the learning material delivered by the teachers in the classroom and the demands of assessment (Öhrstedt & Lindfors, 2019; Rubin et al., 2018). Learning assessments given to students can lead them to adopt certain learning approaches. For instance, if a learning assessment requires a higher level of cognitive processing, students could adopt a deep learning approach. However, if the type of learning assessment only requires the reproduction of facts, students may adopt a surface learning approach (Öhrstedt & Lindfors, 2019). It is worth saying that the type of learning assessment implemented by instructors contributes to students' academic outcomes. Therefore, it is important to understand the learning approach being adopted by students in order to develop an appropriate learning environment and type of learning assessment that might promote deep learning approaches.

Internal and External Aspects

The existing literature has also demonstrated that misconceptions about natural phenomena, lack of motivation and interest, and parental expectations affect students' academic outcomes (Emerson, Fear, Fox, & Sanders, 2012; Guido, 2018; Guo, Klein, & Ro, 2019; Kola, 2017; Sobremisana, 2017; Widiyatmoko & Shimizu, 2018). Certainly, there are other aspects of teaching and learning that contribute to students' scientific thinking and their conceptual understanding which have not been explored much in previous studies, especially in the Indonesian context in higher education. The aspects that have the most influence on students' academic outcomes are still unclear and need to be explored more deeply. Hence, this information is crucial for instructors when they are considering how to design an appropriate learning environment that will facilitate their students in enhancing their scientific thinking and conceptual understanding of physics.

2.7 Summary

In this chapter, the theories and previous research findings related to a number of variables investigated in this study have been reviewed. The chapter began with a brief description of four learning theories, comprising behaviourism, cognitivism, constructivism, and humanism, with a view to understanding how students learn to think scientifically and understand physics concepts.

These theories form a framework upon which this research has been conducted. Specifically, several aspects of scientific thinking (comprising scientific reasoning, argumentation, and epistemological beliefs) and conceptual understanding of physics have been identified and described, which include the definitions, the research instruments to be used to measure the four main variables, and some findings from the previous research.

A review of the literature demonstrated that the identified aspects of scientific thinking have a positive impact on students' conceptual understanding of science, and physics in particular. In addition, the role of students' demographic factors such as gender and grade level on scientific thinking and conceptual understanding have been explored in the previous research with mixed results. Furthermore, the bivariate relationships that emerged between different aspects of scientific thinking that had been examined in previous studies provided the theoretical background needed to propose a research model for the study undertaken here. The proposed model indicates the direct and indirect hypothesised relationships between the four main variables, in which epistemological beliefs are the first variable to be addressed and developed, which in turn predicts the other constructs in the model. Furthermore, in light of the importance of designing and developing high-quality learning environments that support students' scientific thinking and their conceptual understanding, several aspects related to teaching and learning practices such as facilities, teaching methods, learning media, learning activities, homework, learning approach adopted and the type of assessment used, and other external factors have also been presented. The description regarding the research methods and instruments employed to examine the multiple variables involved in this study will be presented in the following chapter.

CHAPTER 3
RESEARCH DESIGN AND DATA COLLECTION

3.1 Overview

The present study seeks to investigate Indonesian pre-service physics teachers' scientific thinking (comprising epistemological beliefs, argumentation, and scientific reasoning) and conceptual understanding of physics. In this chapter, an explanation of the research design and methods used to achieve this aim are described. The chapter begins with a presentation of the research design, covering the mixed methods design employed to collect the data, and an overview of the research sites and the participants involved in this study. This is followed by a description of the data collection methods for both the quantitative and qualitative phases of the study, including an exploration of a number of research instruments used to answer the research questions. Finally, the ethical considerations of the research will be addressed before concluding the chapter.

3.2 Research Design

3.2.1 Mixed Methods Design

Research is concerned with ways of knowing and understanding the world, which can be carried out through various investigations, such as collecting and analysing data to obtain answers to research questions and drawing conclusions based on evidence. Prior to conducting research, an appropriate design must be selected. Knowing the research design allows the researcher to organise the data collection and analyses as well as to interpret the findings. In order to achieve the aims of this study, a mixed methods design was used to integrate both quantitative and qualitative approaches. In a mixed method design, both quantitative and qualitative data are collected in a single study or series of studies (Creswell, 2014b; Creswell & Clark, 2007). Combining statistical measurement (for the quantitative data) and the personal experiences of participants (for the qualitative data) provides a better understanding of the research problem under investigation than employing only one approach (Creswell, 2014a). In a nutshell, mixed methods design allows the researcher to use various data collection tools or multiple approaches in order to identify research problems comprehensively, and to answer research questions that cannot be addressed through a single approach. Two or more methods in the design might complement each other in relation to the strengths and weaknesses of each method and allow for a complete analysis of the research problem. For instance, the weaknesses of the quantitative method might be complemented by the strengths of the qualitative method and vice versa.

In applying a mixed methods research design, there are a number of steps that need to be taken into consideration by the researcher. The first step is to decide upon the philosophical basis for the investigation. The next step is to decide whether data collection would be implemented concurrently or subsequently, and whether the priority of the data collection would have an equal weighting between the quantitative and the qualitative phases of the study. Figure 3. 1 provides many of the options in relation to this step. The final step is to determine the data analysis and integration procedures (Migiro & Magangi, 2011). For example, the researcher might analyse the data separately and then compare the quantitative and qualitative findings.

The conceptual framework for this study adopted a pragmatist perspective. Pragmatism can be employed as a philosophical foundation to support a mixed methods study which has the potential to provide a comprehensive and deep understanding of the research findings. The pragmatist perspective is the best paradigm or philosophical basis for mixed methods studies (Kaushik & Walsh, 2019; Teddlie & Tashakkori, 2009). It allows the researcher to consider multiple perspectives and various theories, and to employ different research methods to address research problems in a single study (Cameron, 2011; Johnson, Onwuegbuzie, & Turner, 2007). Each method has its limitations, so the different methods should be complementary. According to Rossman and Wilson (1985, as cited in Johnson et al., 2007, p. 115), there are three reasons why a researcher would combine quantitative and qualitative methods. Firstly, a mixed methods design allows the researcher to triangulate the research data. Triangulation is considered to be the main advantage of mixed methods design and refers to the use of at least two different data collection methods to study a phenomenon in order to ensure the validity of the construct under consideration and the research findings (Creswell, 2013; Greene, Caracelli, & Graham, 1989). Secondly, mixed methods research enables the researcher to develop analyses to enrich the depth of the data obtained. Finally, mixed methods research allows the researcher to ‘initiate new modes of thinking’ in relation to the paradox that comes from having two data sources. Hence, a mixed methods design allows the researcher to identify aspects of a phenomenon more accurately using a variety of data sources by analysing them from the different perspectives of multiple methods.

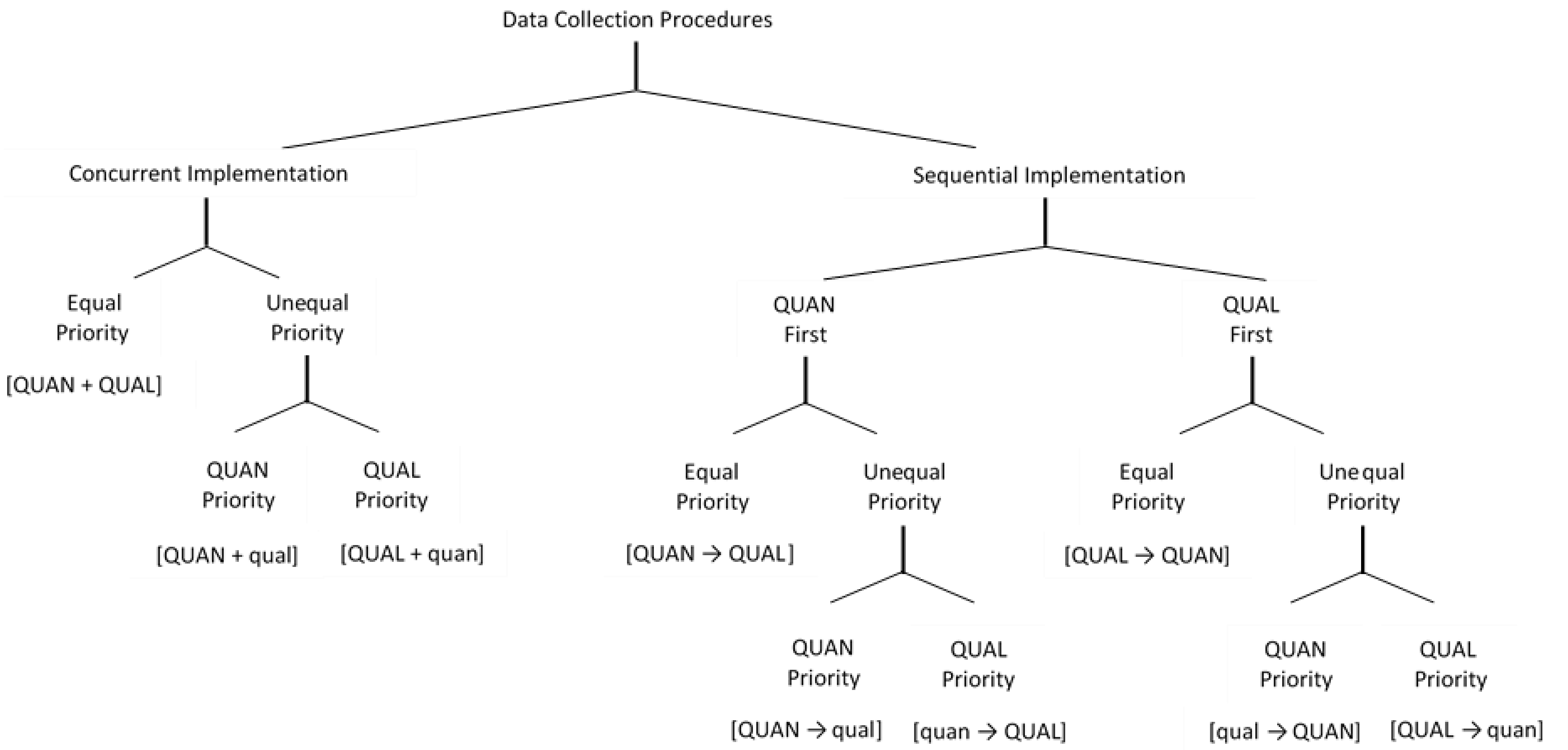


Figure 3. 1 Some options regarding mixed methods data collection procedures (Hanson, Creswell, Clark, Petska, & Creswell, 2005, p. 227)

[QUAN = quantitative data was prioritised; QUAL = qualitative data was prioritised; qual = lower priority given to the qualitative data; quan = lower priority given to the quantitative data].

Figure 3. 1 shows several options for the data collection procedures in mixed methods design that are commonly used by researchers. Procedures for data collection consist of concurrent and sequential implementation. Hanson et al. (2005, p. 227) stated that “implementation refers to the order in which the quantitative and qualitative data are collected, concurrently or sequentially, and priority refers to the weight, or relative emphasis, given to the two types of data, equal or unequal.” Hence, mixed methods design is concerned with the sequencing of and the priorities within, the research. The priority can be equal or can lean towards one component over the other. For instance, unequal data priority indicates that one form of data is emphasised more than the other by collecting one such form of data in more detail than the other. A plus sign in Figure 3. 1 indicates that the quantitative and qualitative data are collected at the same time (concurrently), while a single arrow shows that the quantitative and qualitative data are collected sequentially, or as one dataset followed by the other. A capital letter indicates a higher priority for a certain data collection method, while lowercase letters indicate a lower priority.

Broadly speaking, there are three core designs in the study of mixed methods, namely convergent design, sequential explanatory, and sequential exploratory (Creswell, 2014a). In convergent design, the research process combines concurrent quantitative and qualitative data to address the research problems. The results of the quantitative and qualitative data analyses are merged to create more data from the two sets of data to obtain a more complete understanding. As pointed out by Creswell (2014a, p. 37), “it enables one to gain multiple pictures of a problem from several angles.” In the sequential explanatory design, the research begins with the quantitative phase, in which the data are collected and analysed, followed by the qualitative phase. Mostly, priority is given to the quantitative data, while the qualitative data is used to augment the quantitative data. In this design, data analysis is connected, and integration in most cases occurs at the stage of data interpretation and discussion. This design allows the researcher to draw inferences about how the qualitative findings help to explain the quantitative findings. In addition, the qualitative findings can be used to corroborate, refute, or augment the results from the quantitative study. According to Morse (1991, as cited in Creswell, 2013), if unexpected findings arise from the quantitative data collection, the qualitative data can be useful to verify or explain non-significant or surprising results. Creswell (2013) further argued that this model is easy to implement as the steps are clear with separate phases, and the researcher can easily explain and report the findings. Combining both quantitative and qualitative findings makes it possible to generate more evidence, provide broader interpretations, and give an in-depth understanding of some of the quantitative findings. However, this design is time-consuming and expensive due to employing two separate data collection phases. Finally, in sequential exploratory design, the research begins with the qualitative data collection and analysis. After this stage, the second phase involves the collection of the quantitative data, in which the instruments are administered to a large sample to consider whether the researcher would be able to generalise the research findings (Creswell, 2014a). In this design, the data analysis is connected, and integration in most cases occurs at the stage of data interpretation and discussion. This design allows the researcher to develop assessment instruments based on the qualitative analyses as well as to generalise the qualitative results to a certain population (Hanson et al., 2005).

The present study employed a sequential explanatory design for the data collection. The implementation of quantitative data collection (using surveys) was carried out in the first stage of the study followed by the collection of the qualitative data (through individual interviews) in the second stage. As noted by Creswell and Clark (2007), the primary advantage of the sequential explanatory design is its ability to identify participants' characteristics in the quantitative study to guide sampling for the qualitative study in the second phase. In this study, a preliminary analysis of the quantitative data was carried out in order to select the participants to be involved in the interview phase. Subsequently, the complete analysis of the quantitative data was carried out after the interview data collection had been completed. The data were collected from April to July 2017, and the quantitative data were prioritised more than the qualitative. In the quantitative phase, several software packages were employed to analyse the datasets collected from the participants' responses to the survey instruments to answer research questions (i.e., RQ1 and RQ2, as presented in Chapter 1). Meanwhile, in the qualitative phase, individual semi-structured interviews of a sub-sample of the participants from the quantitative phase were undertaken in order to gather their perceptions, as expressed in their own words, in order to answer research questions RQ3 and RQ4 (as mentioned in Chapter 1). The qualitative findings were used to complement or enrich the findings of the quantitative investigation. Both the quantitative and qualitative findings were combined to generate a better understanding, and a more comprehensive picture, of physics teachers' scientific thinking and conceptual understanding of physics.

This study employed a cross-sectional research design rather than a longitudinal design due to time constraints and resource limitations. According to Creswell (2013), cross-sectional mixed methods design is used to examine studies at a certain point in time, while longitudinal mixed methods design is used to investigate phenomena that change over a certain period of time. In cross-sectional mixed methods, multiple different variables in different population groups can be compared at the same time. This research design enables the researcher to collect data from a large sample, which is comparatively fast to conduct and cheaper to administer (Cohen, Manion, & Morrison, 2011).

In addition, Tashakkori and Teddlie (2010) pointed out that a mixed methods design must integrate the data at one or more stages within the research. In other words, the researcher needs to consider the interconnections between all the research findings from both the quantitative and the qualitative studies. Integration can occur during data collection, data analysis, the data interpretation stages, or in the discussion chapter of the thesis. In this study, both the quantitative and qualitative data were collected and analysed separately, with the integration being provided in the discussion section. Overall, Figure 3. 2 indicates the research procedure employed in this study consisting of research preparation, the process of collecting and analysing the data, and interpreting and reporting on the research findings.

As shown in Figure 3. 2, in the first phase of the data collection, five surveys were distributed to pre-service physics teachers from public and private universities across different year levels (i.e., Year 1 to Year 4). The surveys were used to measure their epistemological beliefs, argumentation, scientific reasoning, and conceptual understanding of physics. Meanwhile, in the second phase, the qualitative data were collected by conducting audiotaped semi-structured interviews with pre-service physics teachers. Further to this, several statistical software packages were used to analyse the quantitative data. Rasch model analysis was conducted using ACER ConQuest v.4. The analyses of the descriptive statistics, the t-test, and the one-way Analysis of Variance (ANOVA) were carried out using IBM SPSS v.25. Meanwhile, the Confirmatory Factor Analysis (CFA) and Structural Equation Modelling (SEM) approaches were undertaken using IBM AMOS v.25. Regarding the qualitative data analysis, NVivo 11 software was used to arrange and code the data. This software allows the researcher to analyse large amounts of data as well as to uncover connections that might be difficult to identify if the analysis was undertaken manually (QSR International Pty Ltd, 2012). Using NVivo 11 allows the researcher to easily identify the source and references which are related to the codes created in the software. A detailed description of the data analysis methods for both the quantitative and qualitative phases is presented in Chapter 4.

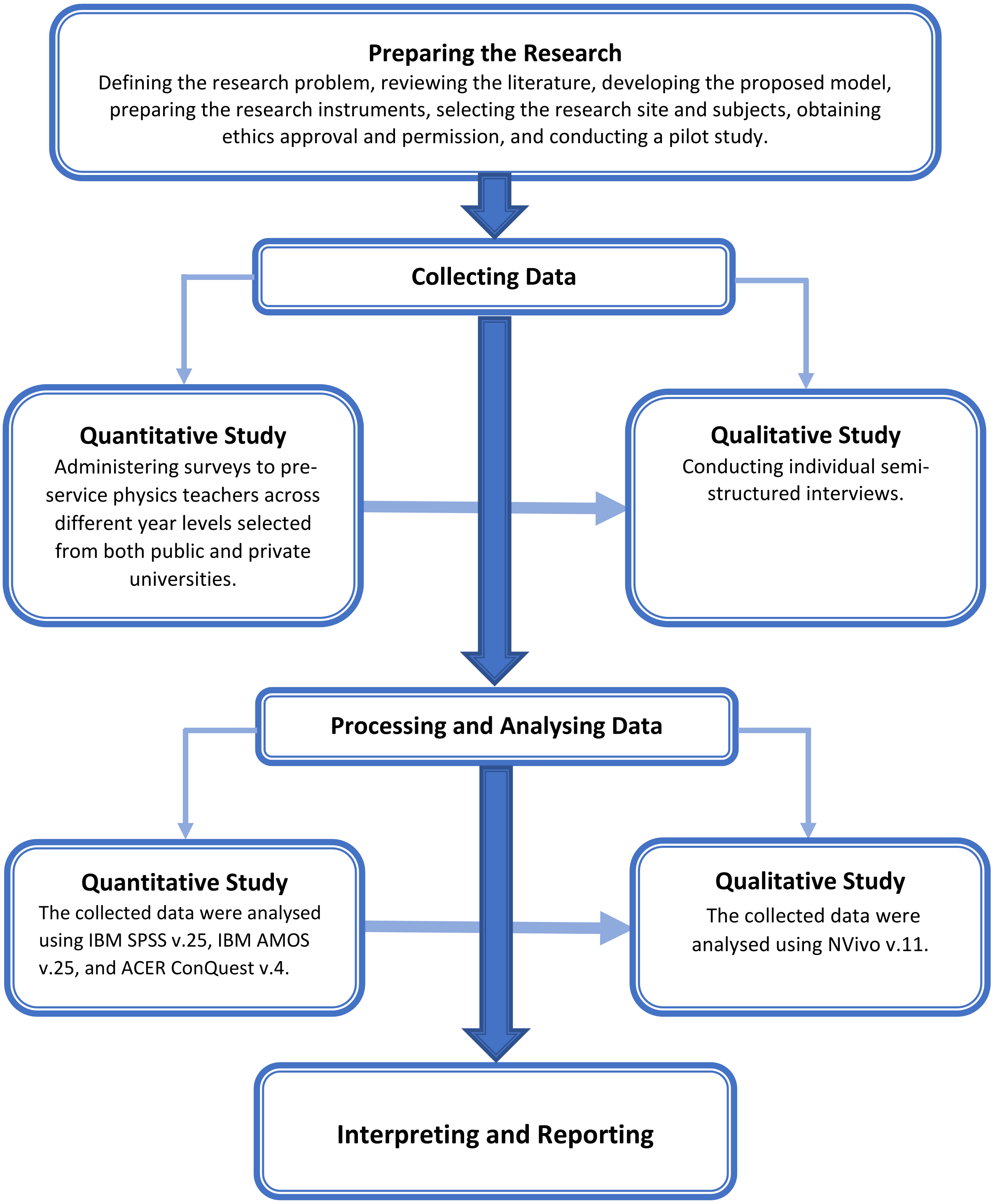


Figure 3. 2 The research procedure of the study

3.2.2 Research Site and Participants

As mentioned earlier, a cross-sectional design was carried out for both the quantitative and qualitative phases of this study. The researcher employed a snapshot of the different years of the participants, rather than tracking the same participant cohort over multiple years. In addition, due to time constraints and resource limitations, the researcher only selected a specific area in Indonesia which has many teacher preparation programs. These institutions are responsible for training future physics teachers. To the best of the researcher's knowledge, Yogyakarta has more pre-service physics teacher programs in higher education institutions than any other region of Indonesia. Yogyakarta is located in Central Java and is famous as a centre for higher education in Indonesia. According to the report of the Ministry of Research, Technology and Higher Education in 2015, one of the teacher education institutions in Yogyakarta was selected as one of Indonesia's best institutions or universities in terms of the quality of human resources and of student activities (http://www.antaranews.com/berita/513728/kemenristek-dikti-umumkan-peringkat-perguruan-tinggi-2015). Hence, based on these considerations, the present study was conducted in Yogyakarta.

In addition, the researcher purposefully selected four out of five universities that had a Physics Education Department who were willing to be involved in the study. The four universities comprised two private and two public universities. Data collection was conducted between April and July 2017. The research population of the study included all pre-service physics teachers in Year 1 to Year 4 from these four selected universities who were registered from 2013/2014 through to the 2016/2017 academic year in the Physics Education Department. All pre-service physics teachers were encouraged to be involved in this study. The total number of pre-service teachers from the Physics Education Department of each university who were recorded as active, non-active, or already graduated at the time of the data collection is provided in Table 3. 1.

As presented in Table 3. 1, the grand total of active pre-service physics teachers represented by year level from across the four universities was 790, 123 pre-service physics teachers were recorded as non-active, and 39 had already graduated. This means that the total pre-service physics teachers registered at these universities were 952. In this study, there were 74 active pre-service physics teachers who could not participate for unspecified reasons, and most of the final year of study level participants were busy with their final project. The number of pre-service physics teachers participating in the surveys was 716. However, of the 716 participants who responded to the survey instruments, 706 completed all the surveys, while 10 decided not to continue their involvement in the study. In conclusion, a total of 706 participants were involved in the quantitative component of the research. This meant that almost 90% of the potential participants from the total population were involved in this study.

Table 3. 1 The number of physics education student teachers by year level from across four universities in Yogyakarta, Indonesia

	Year Level				TOTAL
	Year 1	Year 2	Year 3	Year 4	
Public University 1					
active	66	46	59	79	250
nonactive	-	13	11	9	33
graduated	-	-	-	27	27
Total students	66	59	70	115	310
Public University 2					
active	50	46	39	50	185
nonactive	1	4	5	6	16
graduated	-	-	-	-	-
Total students	51	50	44	56	201
Private University 1					
active	44	59	64	50	217
nonactive	-	14	16	19	49
graduated	-	-	-	4	4
Total students	44	73	80	73	270
Private University 2					
active	48	36	26	28	138
nonactive	-	7	11	7	25
graduated	-	-	-	8	8
Total students	48	43	37	43	171
The grand total of active students across all four universities					790

In the qualitative phase of the study, the participants who were involved in the first phase were selected to be interviewed using a purposive sampling technique (Clark & Creswell, 2011). These included participants who obtained high, medium, and low average scores for each of the surveys. The number of participants involved in the qualitative study was justified using data saturation. Glaser and Strauss (1967, as cited in Guest, Bunce, & Johnson, 2006, p. 64) pointed out that data saturation arises when “no additional data are being found whereby the (researcher) can develop properties of the category. As he sees similar instances over and over again, the researcher becomes empirically confident that a category is saturated ... when one category is saturated, nothing remains but to go on to new groups for data on other categories and attempt to saturate these categories also” (pp. 64-65). In the present study, there were 25 pre-service physics teachers who were involved in the individual semi-structured interviews from both public and private universities across the four-year levels (i.e., Year 1 to Year 4). This number of interviewees to be analysed was considered to meet the requirements of data saturation. The distribution of the interviewees is provided in Table 3. 2.

Table 3. 2 The interviewee distribution in the qualitative phase of the study

Characteristics	Public University		Private University	
	Male	Female	Male	Female
Year Level				
Year 1	2	2	1	1
Year 2	1	-	-	2
Year 3	3	4	1	1
Year 4	2	4	1	-
Total	8	10	3	4

In relation to the demographic characteristics of the participants involved in the study, Table 3. 3 presents the distribution of participants from the public and private universities in terms of year level, gender, and age.

Table 3. 3 Demographic characteristics of the research participants

Characteristics	Public University		Private University		Total (n)
	Frequency (n)	Percentage (%)	Frequency (n)	Percentage (%)	
Year Level					
Year 1	116	29.4	90	28.9	206
Year 2	88	22.3	90	28.9	178
Year 3	96	24.3	84	27.0	180
Year 4	95	24.1	47	15.1	142
Gender					
Male	87	22.0	78	25.1	165
Female	308	78.0	233	74.9	541
Age					
16	-	-	1	0.3	1
17	2	0.5	2	0.6	4
18	58	14.7	37	11.9	95
19	97	24.6	82	26.4	179
20	98	24.8	90	28.9	188
21	81	20.5	64	20.6	145
22	52	13.2	25	8.0	77
23	7	1.8	7	2.3	14
24	-	-	2	0.6	2
25	-	-	1	0.3	1

The number of participants was not the same for each year level in both the public and private universities. The highest number of participants were pre-service physics teachers at year level 1 from the public universities (29.4%). On the other hand, the smallest number of participants involved in the study were pre-service physics teacher at year level 4 from the private universities (15.1%). There were no significant differences in the number of participants across other year levels of study.

As shown in Table 3. 3, the number of males in this study was 165 (23.4 %), while 541 participants (76.6%) were female. The number of females was more than three times the number of males suggesting that, in the present study, females may be more interested in becoming future physics teachers than males. However, this is not the specific area of interest of this study. Table 3. 3 also showed that in both the public and the private universities, the age of the participants ranged primarily from 18 to 22 years (96.88%).

3.3 Methods of Data Collection

Data collection is the collecting of research data empirically to obtain information from participants. The collection of appropriate and valid data is crucial for a study. “Questionnaires, interviews, focus groups, tests, observations, and secondary data” are methods of data collection that can be used to gather data (Johnson & Turner, 2003, p. 298). Questionnaires are mostly used in quantitative studies, while interviews, focus group discussions and case studies are commonly used in qualitative studies (Neuman, 2005).

In this study, survey instruments were used for collecting the quantitative data. The survey method is highly recommended when investigating large populations with only a "snapshot" of the situation at a certain point in time, and is generally used for verification and validation purposes (Gable, 1994). As also noted by Shieh (2003, p. 27), “survey data collection is a very efficient way to gather information for research of interest.” Meanwhile, interviews were conducted to obtain the in-depth views of the participants in the qualitative phase.

There is no denying that it is more convenient for the researcher to use available existing instruments that have been developed and validated, due to the difficulty in developing new instruments and the considerable amount of time and skills required to do this (Fraenkel & Wallen, 2003). Furthermore, delivering paper and pencil tests as well as conducting manual correction are time-consuming activities compared to online tests, which might save time and provide instant summaries of results. Considering the time constraints and the skills needed to develop new instruments, the present study adopted existing survey instruments to evaluate the participants' epistemological beliefs, argumentation, scientific reasoning, and conceptual understanding of physics, particularly on the topics of 'Mechanics' and 'Electricity and Magnetism'. An overview of all the survey instruments used in this study is provided in Appendix 1 to 5. Further to this, due to constraints in the available facilities at the research sites, the study was conducted using a paper and pencil format. This also ensured that the participants responded to the surveys according to their abilities, and that there was no chance for them to ask others or find related sources (e.g., from the Internet or books). The quality of the data collected was determined by the participants' responses to the survey instruments as well as their responses during the interviews. Confidentiality and anonymity were preserved in order to promote honesty and openness in the participants' responses.

In this study, all the existing survey instruments were translated from English to Bahasa Indonesia. Translation of the surveys used in the study was needed because the existing research instruments were from countries whose language and culture were different from that of the participants. Schoua-Glusberg (1998) stated that the purpose of translating survey instruments is to prepare surveys in different languages. However, in some cases, translators may be forced to translate the survey items based on their interpretations. According to Hunt and Bhopal (2004, p. 618), the quality of the translation might suffer due to "inadequate translation procedures, inappropriate content, insensitivity of items, and the failure of researchers to make themselves familiar with cultural norms and beliefs." Therefore, the translation undertaken in the present study might affect the validity of the research and the quality of the research findings.

However, to minimise the translation issues in this study, several techniques were used such as back-translation carried out independetly by two bilingual professional translators in order to compare the accuracy of the two versions for further revision and refinement. As noted by Brislin (1986, as cited in Chen et al., 2019), the back-translation process is carried out by two translators in order to obtain valid and reliable research instruments, especially for cross-cultural studies. Back-translation is one procedure used to improve the translation of interviews (Clark, Birkhead, Fernandez, & Egger, 2017). In addition, before the pilot study was conducted, the translated copy was also reviewed by two Indonesian PhD students at Flinders University. This was done to identify translation errors and to reduce the need for revisions on the survey items. These Indonesian PhD students did not find any incoherence in the Bahasa Indonesia translation of the survey instruments, and the back-translation results showed that the items retained the same meaning as the original items in the surveys. In addition, as the researcher is a native Indonesian, and educated as a lecturer in physics education in Indonesia, the language barrier and the background of Indonesian physics teacher's knowledge was not a disadvantage. However, this issue may also lead to the risk of bias and omissions of important facts that are taken for granted.

A pilot study to examine the clarity of the words used and to identify any language problems in the translated text and the meaning of the survey items was carried out with 30 Indonesian pre-service teachers who came from one of the universities involved in this study before the main study was undertaken. This was undertaken to identify any language problems and to make final changes to the surveys. Based on the pilot study, some necessary changes were made before the final format of each survey was settled upon. The pilot study results indicated that the research participants should have no problems in understanding the survey questions due to translation issues.

3.3.1 Phase 1: Quantitative Study

For the present study, five paper-based survey instruments were distributed to the participants for the quantitative data collection. The survey method is very helpful for the researcher when collecting data from a large group of respondents so that the research data can be collected quickly and efficiently. In the subsequent section, the five surveys used for the data collection are described in more detail.

3.3.1.1 Epistemological Beliefs Assessment for Physical Science (EBAPS)

Epistemological beliefs in science refer to an understanding of how scientific ideas are constructed, including the nature of knowledge and the process of knowing about scientific knowledge (Hofer, 2006). As pointed out by Hofer and Pintrich (2012, p. 52), “the psychological construct of personal epistemology is used to describe how personal beliefs convey what knowledge is, how it is obtained, what it is used for, and how useful it is in any context.”

To probe students' epistemological beliefs particularly in the physical sciences area, there are several well-known surveys that are commonly used by researchers. The surveys are the *Maryland Physics Expectation* (MPEX), the *Views About Science Survey* (VASS), *Colorado Learning Attitudes* *about Science Survey* (CLASS), and the *Epistemological Beliefs Assessment for Physical Science* (EBAPS) (Adams et al., 2006). In this study, to probe Indonesian pre-service physics teachers' epistemological beliefs, the *Epistemological Beliefs Assessment for Physical Science* (EBAPS) survey was employed. According to Elby (2001), the EBAPS survey was developed to probe the epistemological beliefs of undergraduate students studying introductory physics, chemistry or physical science. The EBAPS survey comprises items that deal with science and science learning. This survey is one of the most accessible and frequently used surveys to measure students' epistemological beliefs (Kortemeyer, 2007), particularly by physics education researchers. The EBAPS survey is available at http://www2.physics.umd.edu/~elby/EBAPS/home.htm, and was originally developed and validated by Elby, Frederiksen, Schwarz, and White at the University of California, Berkeley (The Idea Behind EBAPS, 2002). The authors validated this instrument through a pilot study and informal feedback from one hundred local community college students. More specifically, the EBAPS survey was used to probe students' epistemological beliefs in the physics and chemistry fields.

There has been little research reporting on the statistical measurement of the EBAPS survey. Mostly, previous studies have reported the reliability of instruments using Cronbach's alpha. The reliability indices range from 0 to 1, a higher level of reliability indicating a more reliable scale. The rule of thumb for the Cronbach's alpha coefficient is "_ > 0.9 – Excellent, _ > 0.8 – Good, _ > 0.7 – Acceptable, _ > 0.6 – Questionable, _ > 0.5 – Poor, and_ < 0.5 – Unacceptable" (George & Mallery, 2003, p. 231). Previous studies have also reported on the reliability of the EBAPS survey. For example, Muis and Gierus (2014) showed that the Cronbach alpha coefficients for the EBAPS survey for the five sub-scales was found to range from 0.63 to 0.90. Furthermore, there has been no published statistical analysis of the validity of the categories or sub-scales of the EBAPS survey (Adams et al., 2006; Özmen & Özdemir, 2019). Hence, the present study provides a presentation of the statistical measurements of the validity of the sub-scales for the EBAPS survey. Rasch model analysis and Confirmatory Factor Analysis (CFA) were employed to validate this instrument using the Weighted Likelihood Estimate (WLE) scores.

Originally, the EBAPS instrument consisted of 30 items, combining three different item types, namely 17 items on a five-point Likert scale to be rated from "strongly disagree" to "strongly agree" in section one, six multiple-choice questions in section two, and seven debate questions in a multiple-choice format in section three. The EBAPS examines participants' epistemological beliefs along five different sub-scales as provided in Table 3. 4. This table presents five EBAPS sub-scales and the descriptions of the ideas covered by each sub-scale. For instance, a question from the EBAPS survey is “to understand chemistry and physics, the formulas (equations) are really the main thing; the other material is mostly to help you decide which equations to use in which situations." This question examines participants' beliefs about the structure of knowledge (Elby et al., 1999).

Table 3. 4 The five sub-scales of the EBAPS survey (Elby et al., 1999)

EBAPS Sub-scales	Description
Structure of scientific knowledge	Is physics and chemistry knowledge a bunch of weakly connected pieces without much structure and consisting mainly of facts and formulas? Or is it a coherent, conceptual, highly-structured, unified whole?
Nature of knowing and learning	Does learning science consist mainly of absorbing information? Or, does it rely crucially on constructing one's understanding by working through the material actively, by relating new material to prior experiences, intuitions, and knowledge, and by reflecting upon and monitoring one's understanding?
Real-life applicability	Are scientific knowledge and scientific ways of thinking applicable only in restricted spheres, such as a classroom or laboratory? Or, does science apply more generally to real life? These items tease out students' views of the applicability of scientific knowledge *as distinct from* the student's own desire to apply science to real life, which depends on the student's interests, goals, and other non-epistemological factors.
Evolving knowledge	This dimension probes the extent to which students navigate between the twin perils of absolutism (thinking all scientific knowledge is set in stone) and extreme relativism (making no distinctions between evidence-based reasoning and mere opinion).
Source of ability to learn	Is being good at science mostly a matter of fixed natural ability? Or, can most people become better at learning (and doing) science? As much as possible, these items probe students' epistemological views about the efficacy of hard work and good study strategies, *as distinct from* their self-confidence and other beliefs about themselves.

According to Elby et al. (1999), each item is scored between 0 and 4 in a non-linear scheme. A score of 0 represents the least sophisticated, with 4 representing the most sophisticated. In addition, the distribution of items on each sub-scale is provided in Table 3. 5. As shown in this table, there are questions in the EBAPS survey that are included in multiple sub-scales (i.e., 19 and 28), and a few are not included in any of these sub-scales (e.g., 4 and 21). Since a multidimensional model analysis was employed to examine the five dimensions of the Epistemological Beliefs scale in this study, questions 4, 19, 21, and 28 were not included in the subsequent analysis. In other words, item numbers 19 and 28 were excluded from the subsequent analysis because they were across the sub-scales, while item number 4 and 21 were excluded because they did not belong to any sub-scale. As noted by Adams and Wu (2010, p. 14), "… there are no items in common across the subscales." The validity results of the EBAPS survey are presented in Chapter 5.

Table 3. 5 Distribution of items of the EBAPS

Sub-scale	Item Numbers
Structure of scientific knowledge	2, 8, 10, 15, 17, 19, 20, 23, 24, 28
Nature of knowing and learning	1, 7, 11, 12, 13, 18, 26, 30
Real-life applicability	3, 14, 19, 27
Evolving knowledge	6, 28, 29
Source of ability to learn	5, 9, 16, 22, 25
No sub-scale	4, 21

The following is an example of one of the debate questions (item no. 25) in the EBAPS survey. A copy of the EBAPS survey is provided in Appendix 1.

25. Anna: I just read about Kay Kinoshita, the physicist. She sounds naturally brilliant.
Emily: Maybe she is. But when it comes to being good at science, hard work is more important than "natural ability". I bet Dr. Kinoshita does well because she worked really hard.
Anna: Well, maybe she did. But let's face it, some people are just smarter at science than other people. Without natural ability, hard work won't get you anywhere in science!

(a) I agree almost entirely with Anna.
(b) Although I agree more with Anna, I think Emily makes some good points.
(c) I agree (or disagree) equally with Anna and Emily
(d) Although I agree more with Emily, I think Anna makes some good points.
(e) I agree almost entirely with Emily

3.3.1.2 Argumentation Test

Erduran and Jiménez-Aleixandre (2008) defined argumentation as the connections made between claims and data through the justification and evaluation of knowledge. In order to probe Indonesian pre-service physics teachers' argumentation, the *Argumentation Test* developed by Sampson and Clark (2006) was used in the present study. Previous studies reported the reliability of instruments using Cronbach's alpha. The reliability indices range from 0 to 1, with a higher level of reliability indicating a more reliable scale. For instance, Cetin, Erduran, and Kaya (2010) reported that the

Cronbach's alpha coefficient of the *Argumentation Test* for pre-service teachers was found to be 0.68, which is still considered acceptable. As well, Kaya, Cetin, and Erduran (2014) reported the reliability analysis of the *Argumentation Test* using Cronbach's alpha coefficient to be 0.70, which is considered acceptable. Since there have been few discussions on the statistical measurements of this instrument, the present study provides an in-depth discussion of multivariate statistical analysis techniques, such as Rasch analysis and Confirmatory Factor Analysis (CFA) to validate the *Argumentation Test* using the Weighted Likelihood Estimate (WLE) scores, which are presented in Chapter 5.

The *Argumentation Test* consists of six questions with two parts. Each part comprises three questions. In the first part, the participants were given a claim and six different arguments related to this claim, in which “1” meant the most convincing argument and “6” meant the least convincing argument. The participants were asked to rank these six different arguments as “a good scientific argument” hierarchically (Kaya et al., 2012). In rank 1, the argument should cover “data, explanation, and rebuttal”; in rank 2, “explanation and evidence”; in rank 3, “evidence only”; in rank 4, “warrant only”; in rank 5, “appeal to authority”, and in rank 6, the argument would be considered “contradictory”. The scale of this ranking and a sample question in the first part of the *Argumentation Test* is provided in the following:

1 = the most convincing argument (data, explanation, rebuttal)
2 = the 2nd most convincing argument (explanation and evidence)
3 = the 3rd most convincing argument (evidence only)
4 = the 4th most convincing argument (warrant only)
5 = the 5th most convincing argument (appeal to authority)
6 = the least convincing argument (contradictory)

Question 1. Your task is to rank these 6 different arguments in terms of how convincing you think they are. Remember that you can only rank one claim as 1, one claim as 2, one claim as 3, and so on.

Claim: Objects that are in the same room are the same temperature even though they feel different because …	Your Ranking
…… when we measured the temperature of the table, it was 23.4^{O}C, the metal chair leg was 23.1^{O}C, and the computer keyboard was 23.6^{O}C.	_ _ _ _ _ _ _
…… good conductors feel different than poor conductors even though they are the same temperature.	_ _ _ _ _ _ _
…… objects that are in the same environment gain or lose heat energy until everything is the same temperature. Our data from the lab proves that point: the mouse pad and plastic desk were both 23^{O}C.	_ _ _ _ _ _ _
…… objects will release and hold different amounts of heat energy depending on how good an insulator or conductor it is.	_ _ _ _ _ _ _
…… our textbook says that all objects in the same room will eventually reach the same temperature.	_ _ _ _ _ _ _
…… we measured the temperature of the wooden table and the chair leg and they were both 23^{O}C, even though the metal chair leg feels colder. If the metal chair leg was actually colder, it would have been a lower temperature when we compared it to the temperature of the table.	_ _ _ _ _ _ _

In the second part of the *Argumentation Test*, the participants were given a claim supported by an argument for each question. Following the claim, there was a challenge and six different arguments, in which "1" meant the strongest challenge to the argument, and "6" meant the weakest challenge. For each question, the participants were asked to rank these different arguments as "a good challenge to a scientific argument" hierarchically, from the strongest challenge to the argument to the weakest challenge (Kaya et al., 2012). A rank of 1 was characterised as an "argument with backing"; 2 as an "argument with a warrant"; 3 as an "argument with data"; 4 as an "argument with the claim"; 5 as "a counter claim only", and 6 was characterised as an "emotive argument". The scale of this ranking, and an example question in the second part of the *Argumentation Test*, are given in the following:

1 = This comment is the strongest challenge to this argument
(Argument with backing: rebuttal against grounds with grounds)
2 = This comment is the 2nd strongest challenge to this argument
(Argument with a warrant: rebuttal against grounds no grounds)
3 = This comment is the 3rd strongest challenge to this argument
(Argument with data: rebuttal against thesis with grounds)
4 = This comment is the 4th strongest challenge to this argument
(Argument with the claim: rebuttal against thesis with no grounds)
5 = This comment is the 5th strongest challenge to this argument (a counter claim only)
6 = This comment is the weakest challenge to this argument (emotive argument)

Question 4. Jason, Angela, Sarah, and Tim are in physics class together. Their teacher asked them to design an experiment to determine if all objects in the same room have the same temperature, even though they feel different. After they designed and carried out an experiment to answer this question on their own, they met in a small group to discuss what they had found. Suppose Jason suggests that:

> *"I think that all objects in the same room always have different temperatures because they feel different, and when we measured the temperature of the table, it was 23.4^{0}C, the metal chair leg was 23.1^{0}C, and the computer keyboard was 23.6^{0}C."*

Angela disagrees with Jason. Your task is to rank these 6 different challenges in terms of how strong you think they are. *Remember that you can only rank one challenge as 1, one challenge as 2, one challenge as 3, and so on.*

Angela: I disagree …	Your Ranking
…… because your evidence does not support your claim. All of the objects that you measured were within one degree of each other. That small of difference is just measurement error.	_ _ _ _ _ _ _
…… I think that all objects in the same room are the same temperature even though they feel different	_ _ _ _ _ _ _
…… if those objects were really different temperatures, their temperature would have been much different. For example, when I measured the temperature of my arm, it was 37^{0}C, while the temperature of the table was 23^{O}C – that is a difference of 14 degrees. Everything else was right around 23^{O}C.	_ _ _ _ _ _ _
…… I think all objects become the same temperature even though they feel different, because objects that are good conductors feel colder than objects that are poor conductors, because heat transfers through good conductors faster.	_ _ _ _ _ _ _
…… because I know you always rush through labs and never get the right answer.	_ _ _ _ _ _ _
…… I think all objects become the same temperature because the temperature of all those objects you measured were within 1 degree.	_ _ _ _ _ _ _

Regarding the scoring method for each item of the *Argumentation Test*, the correct answers for each item were coded as "1" and the incorrect answers as "0". Thus, the maximum total score which students could get in this *Argumentation Test* was 36. The distribution of items in the *Argumentation Test* is presented in Table 3. 6. A copy of the *Argumentation Test* is provided in Appendix 2.

Table 3. 6 Distribution of items of the Argumentation Test

Part	Rank of Argument	Item Numbers
Part I Making Scientific Argumentation	Contradictory	4, 9, 13,
	Appeal to authority	5, 11, 15,
	Warrant only	2, 7, 17,
	Evidence only	1, 10, 18,
	Explanation and evidence	3, 8, 16,
	Data, explanation, rebuttal	6, 12, 14
Part II Challenging Argumentation	Emotive argument	23, 25, 34
	Counter claim only	20, 30, 32
	Argument with claim: Rebuttal against thesis with no grounds	24, 27, 36
	Argument with data: Rebuttal against thesis with grounds	22, 28, 31
	Argument with warrant: Rebuttal against grounds no grounds	19, 26, 33
	Argument with backing: Rebuttal against grounds with grounds	21, 29, 35

3.3.1.3 Lawson's Classroom Test of Scientific Reasoning (LCTSR)

There has been increasing interest among researchers and educators, particularly in the STEM areas, to investigate students' skills in scientific reasoning. Scientific reasoning represents "the abilities to systematically explore a problem, to formulate and test hypotheses, to manipulate and isolate variables, and to observe and evaluate the consequences" (Bao, Cai, et al., 2009, p. 586). To probe students' scientific reasoning, various instruments have been used by researchers. Historically, the Piagetian clinical interview was used to measure students' formal reasoning abilities. However, this clinical interview method requires experienced interviewers, special materials and equipment, and also sufficient time (Lawson, 1978). A number of researchers have used this Piagetian method to develop their own assessment tools to assess students' scientific reasoning such as the *Test of Logical Thinking* (TOLT) (Tobin & Capie, 1981), the *Group Assessment of Logical Thinking Test* (GALT) (Roadrangka et al., 1982), and the *Lawson's Classroom Test of Scientific Reasoning*-LCTSR (Lawson, 1978, ver. 2000). The *Lawson's Test* is one of the most widely-used instruments among STEM educators and science education communities to investigate students' scientific reasoning (Bao, Cai, et al., 2009; Ding, 2014c). According to Bao and colleagues (2009), the *Lawson's Test* is the only readily available quantitative instrument to measure students' scientific reasoning and, for this reason, has been used in science education studies. The present study employs the *Lawson's Test* to probe Indonesian pre-service physics teachers' scientific reasoning.

Originally, the LCTSR test, or *Lawson's Test*, was first developed in 1978 to investigate students' developmental level specifically for formal-level reasoning. Lawson, Banks and Lovgin (2007) reported that the KR-20 reliability coefficient of the original version of the *Lawson's Test* was 0.79, while the validity of the test has been established in previous studies. The *Lawson's Test* has undergone several revisions and after more than 20 years, the most recent version was released in 2000. Lawson published the latest version of the *Lawson's Test* with multiple-choice questions, which can be scored quickly and objectively. The test items required participants to select the best answer from the choices provided. This most recent version of the instrument comprises 12 pairs of questions (two-tier questions) or 24 items in total. Treagust (1995) explained that the two-tier structure of the instrument is comprised of two related questions in which the first-tier comprises a question with some possible answers, followed by a second-tier question with some possible reasons for the response to the first question. The two-tier questions were designed to examine the process of students' scientific understanding to identify their misconceptions (Treagust, 1995). In addition, *Lawson's Classroom Test of Scientific Reasoning* (LCTSR) measures a number of different dimensions, namely: conservation of weight and volume, proportional reasoning, control of variables, probability and correlational reasoning, and hypothetical-deductive reasoning. The distribution of items on each scientific reasoning dimension is provided in Table 3. 7.

Table 3. 7 Distribution of items of Lawson's Test

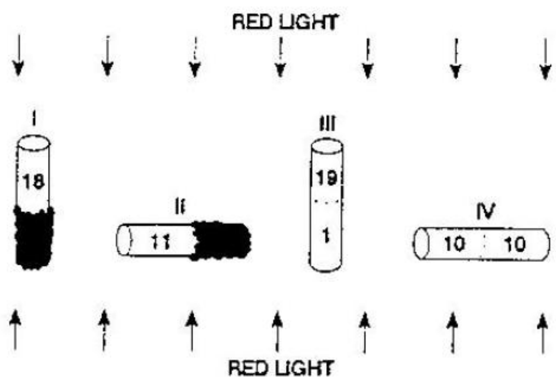
Dimension	Item Numbers
Conservation of weight and volume	1, 2, 3, 4
Proportional reasoning	5, 6, 7, 8
Control of variables	9, 10, 11, 12, 13, 14
Probability and correlational reasoning	15, 16, 17, 18, 19, 20
Hypothetical-deductive reasoning	21, 22, 23, 24

In the first 10 pairs of questions (i.e., item numbers 1 to 20), each pair begins with one question for correctly predicting the outcome of a particular situation, followed by a second question to find the correct reasoning behind the selection in the first question. Meanwhile, items numbered 21 to 24 are proposed to measure the participants' hypothetical-deductive reasoning (Lawson, 2000a). Particularly, in the item pair numbered 21-22, the lead question provides some choices in experimental design for testing a set of hypotheses. The follow-up question asks students to identify the data pattern that would help conclusions to be made regarding the hypotheses. In the item pair numbered 23-24, both questions ask students to identify the data pattern that would support conclusions about the given hypotheses.

In terms of the statistical measurement of the instrument, Han (2013) pointed out that the validity of this test was grounded in large-scale assessment data. Reliability of the *Lawson's Test* has been evaluated by previous researchers. For instance, Lee and She (2010) demonstrated that the Cronbach alpha coefficient for the *Lawson's Test* was 0.71 for the pre-test, 0.61 for the post-test, and 0.76 for the retention-test, which is considered acceptable. Piraksa and colleagues (2014) reported that the reliability analysis of the *Lawson's Test* using Cronbach's alpha coefficient was 0.71, which is considered acceptable. Meanwhile, Ding and colleagues (2016) reported that the overall Kuder-Richardson-20 reliability (KR-20) of the *Lawson's Test* was 0.76, which is considered acceptable. The KR-20 reliability analysis is equivalent to the Cronbach's alpha coefficient (Traxler et al., 2018). Furthermore, Ding and colleagues (2016) established the validity and reliability of the *Lawson's Test* using Rasch analysis. These researchers demonstrated that the person and item reliability of the LCTSR are 0.76 and 0.98, respectively, indicating a sufficient consistency. In this study, in-depth description of the multivariate statistical analyses comprising Rasch analysis and Confirmatory Factor Analysis (CFA) to validate the LCTSR instrument using the Weighted Likelihood Estimate (WLE) scores, is presented in Chapter 5.

The following is an example of item number 11 of the *Lawson's Test*. A copy of the *Lawson's Test* is provided in Appendix 3.

11. Twenty fruit flies are placed in each of four glass tubes. The tubes are sealed. Tubes I and II are partially covered with black paper; Tubes III and IV are not covered. The tubes are placed as shown. Then they are exposed to red light for five minutes. The number of flies in the uncovered part of each tube is shown in the drawing. This experiment shows that flies respond to (respond means move to or away from):



a. red light but not gravity
b. gravity but not red light
c. both red light and gravity
d. neither red light nor gravity

Considering the scoring of the two-tier items in the *Lawson's Test*, educators and researchers (i.e., Coletta & Phillips, 2005; Lee & She, 2010; Treagust, 1995) used the pair-scoring schema in which the correct items are scored with "1" if the students choose the right answers for pair questions (i.e., the answer and the reason). In the two-tier items, students should choose the right answer for both the question and the corresponding correct reason to get a score of "1" but will get a score of "0" if they choose the right answer for the wrong reason or chose the wrong answer for the right reason. This means that the correct total score would be 12.

3.3.1.4 Force Concept Inventory (FCI)

The *Force Concept Inventory* (FCI) test was developed and published in 1992 by Hestenes, Wells, and Swackhamer. As a diagnostic tool, this instrument is useful to identify students' misconceptions specifically in the conceptual domain of Newton's Motion Law and basic Kinematics that is essential for understanding mechanics concepts (Halloun & Hestenes, 1985; Hestenes et al., 1992; Seyranian et al., 2018). Thus, it can be used to examine the extent to which students hold misconceptions in physics, especially in the topic of mechanics. In addition, the FCI instrument has been widely used to evaluate the effectiveness of introductory physics instruction at high school and university level. Historically, the FCI test was constructed as an improvement on the *Mechanics Diagnostic Test* (MDT) designed by Ibrahim Abou Halloun and David Hestenes at Arizona State University (Hestenes et al., 1992). The MDT instrument was developed and published in 1985 to evaluate students' basic qualitative conceptions of mechanics and to identify common misconceptions (Halloun & Hestenes, 1985). Around half the FCI questions were taken directly from the MDT instrument, while the remaining questions were developed by Halloun and Hestenes who argued that "the Inventory (FCI) has the advantage of supplying a more systematic and complete profile of the various misconceptions" (Hestenes et al., 1992). Further, Alwan (2011) pointed out that misconceptions can be identified when "something a person knows and believes does not match what is known to be scientifically correct, also most people who hold misconceptions are not aware of their ideas." A diagnostic test such as the FCI was not only useful for measuring students' understanding of basic mechanics concepts, but also to raise awareness of students' conceptual difficulties. Savinainen and Scott (2002) also argued that the FCI instrument can be used as a tool to enhance the teaching and learning of mechanics. For these reasons, the FCI instrument is a useful physics test for physics teachers and researchers in the field of physics education.

In order to probe Indonesian pre-service physics teachers' conceptual understanding of Mechanics, the *Force Concept Inventory* (FCI) was used in the present study because of its many advantages and its known reliability (Hake, 1998). The Force Concept Inventory is a standardised instrument to measure students' conceptual understanding of basic mechanics topics with minimal use of mathematics (McDermott & Redish, 1999). In addition, Hake (2007, p. 25) pointed out other advantages of using the FCI instrument, namely that the questions in the instrument use a multiple-choice format that makes it easy for the researcher to administer this test to a large number of participants. In addition, the test questions are aimed at investigating the conceptual understanding of the basic concepts of Newtonian mechanics that are developed in such a way that they are easy to understand, even if given to a beginner who has never taken a physics course. As noted by Khairani and Razak (2015), multiple-choice is a test format that is widely used by educators or researchers because it provides several benefits such as enabling research to involve large numbers of samples, getting participants' responses easily and quickly, as well as being more objective in scoring compared to other forms of tests (e.g., essay questions or presentations).

Originally, the FCI consisted of 29 multiple-choice questions when it was first published in *The Physics Teacher* (Hestenes et al., 1992). The revised version of the FCI survey was published in 1995 consisting of 30 items. The scoring method for each item of the FCI test was that the correct answers for each item were scored as "1" and the incorrect answers were scored as "0". Hake (1998, p. 2) asserted that questions in the FCI probe students' conceptual understanding of basic concepts of Newtonian mechanics which can be used to evaluate the effectiveness of instruction in this physics subject. Furthermore, the FCI instrument classifies six Newtonian concepts that are essential to the Newtonian concept of force, as given in Table 3. 8 along with a taxonomy of misconceptions. Hestenes et al. (1992) asserted that all six Newtonian concepts are needed for the complete concept. For instance, the kinematics concepts are important, as Newton's Second Law requires an understanding of the concept of acceleration. Table 3. 8 also shows 28 different misconceptions that correspond to the six Newtonian concepts. Hestenes et al. (1992, p. 4) exemplified that "... terms like 'force', 'energy', and 'power' are often incorrectly used interchangeably, as are the terms 'velocity' and 'acceleration'."

As mentioned previously, half of the FCI questions are the same as those in the MDT. The validity and reliability of the MDT instrument has been tested through statistical analysis and interviews. The Kuder-Richardson reliability coefficients of MDT were 0.86 and 0.89 for the post-test, which are indicative of a highly reliable test (Halloun & Hestenes, 1985). Considering the validity and reliability of the FCI instrument, the author did not repeat the formal procedure on the grounds that the test design is similar to the *Mechanical Diagnostic Test* – MDT (Hestenes et al., 1992). According to Lasry and colleagues (2011, p. 909), "although the FCI has been given more than one hundred thousand times at several hundred institutions worldwide, little data exists on its reliability."

Table 3. 8 The six Newtonian concepts and the taxonomy of misconceptions (Hestenes et al., 1992)

The six Newtonian Concepts	The taxonomy of misconceptions
0. Kinematics	1. Kinematics
Velocity discriminated from position	K1. Position-velocity undiscriminated
Acceleration discriminated from velocity	K2. Velocity-acceleration undiscriminated
Constant acceleration entails	K3. Non-vectorial velocity composition
parabolic orbit	2. Impetus
changing speed	I1. Impetus supplied by the 'hit'
Vector addition of velocities	I2. Loss/recovery of original impetus
1. Newton's First Law	I3. Impetus dissipation
With no force	I4. Gradual/delayed impetus build-up
velocity direction constant	I5. Circular impetus
speed constant	3. Active Force
With cancelling forces	AF1. Only active agents exert force
2. Newton's Second Law	AF2. Motion implies active force
Impulsive force	AF3. No motion implies no force
Constant force implies	AF4. Velocity proportional to applied force
constant acceleration	AF5. Acceleration implies increasing force
3. Newton's Third Law	AF6. Force causes acceleration to terminal velocity
Impulsive forces	AF7. Active force wears out
Continuous forces	4. Action/Reaction Pairs
4. Superposition principle	AR1. Greater mass implies greater force
Vector sum	AR2. Most active agent produces greatest force
Cancelling forces	5. Concatenation of Influences
5. Kinds of forces	CI1. Largest force determines motion
Solid contact	CI2. Force compromise determines motion
Passive	CI3. Last force to act determines motion
Impulsive	6. Other Influences on Motion
Friction opposes motion	CF. Centrifugal force
Fluid contact	Ob. Obstacles exert no force
Air resistance	Resistance
Buoyant (air pressure)	R1. Mass makes things stop
Gravitation	R2. Motion when force overcomes resistance
Acceleration independent of weight	R3. Resistance opposes force/impetus
Parabolic trajectory	Gravity
	G1. Air pressure-assisted gravity
	G2. Gravity intrinsic to mass
	G3. Heavier objects fall faster
	G4. Gravity increases as objects fall
	G5. Gravity acts after impetus wears down

In the study conducted by Lasry et al. (2011), the reliability of the FCI was measured by both internal consistency and test-retest performance. The internal consistency of the FCI instrument using Kuder-Richardson 20 (KR-20) was 0.90 and 0.865 for test-retest combined, which is considered satisfactory. Another researcher who reported on the reliability of the FCI instrument was Demirci (2005). His study established the internal reliabilities using Kuder-Richardson 21, namely 0.67 for the pre-test and 0.69 for post-test. Meanwhile, Kiong and Sulaiman (2010) reported that the FCI test score has a reliability of 0.65 which is still acceptable. An in-depth explanation of statistical analysis for the FCI instrument in this study is presented in Chapter 5. The Rasch model analysis was conducted to transform all raw scores obtained from the participants' responses into an interval scale to generate Weighted Likelihood Estimate (WLE) scores used for further analysis. The following is an example of an FCI question for item no 13. A copy of the FCI survey is given in Appendix 4.

13. A boy throws a steel ball straight up. Consider the motion of the ball only after it has left the boy's hand, but before it touches the ground, and assume that forces exerted by the air are negligible. For these conditions, the force(s) acting on the ball is (are):
    (A) a downward force of gravity along with a steadily decreasing upward force.
    (B) a steadily decreasing upward force from the moment it leaves the boy's hand until it reaches its highest point; on the way down, there is a steadily increasing downward force of gravity as the object gets closer to the earth.
    (C) an almost constant downward force of gravity along with an upward force that steadily decreases until the ball reaches its highest point; on the way down, there is only a constant downward force of gravity.
    (D) an almost constant downward force of gravity only.
    (E) none of the above. The ball falls back to ground because of its natural tendency to rest on the surface of the earth.

3.3.1.5 Brief Electricity and Magnetism Assessment (BEMA)

The present study adopted the *Brief Electricity and Magnetism Assessment* (BEMA) instrument to probe Indonesian pre-service physics teachers' conceptual understanding of basic Electricity and Magnetism topics. BEMA was developed and published by Chabay and Sherwood (1997), and has been widely used at college and university levels. The BEMA instrument comprises mostly qualitative questions and a small number of semi-quantitative questions which only require simple mathematical calculations (Ding et al., 2006). The BEMA test consists of 31 multiple-choice items with up to nine answer choices on some questions that cover a broad range of concepts in the electricity and magnetism domain. The first 19 items of the BEMA survey focus on electricity concepts, and the remaining items focus on magnetism or electromagnetic induction concepts (Ding, 2014b). Each item of the BEMA test is scored "1" for correct answers for each item and incorrect answers are scored as "0". The following is an example of a BEMA question for item no 19. A copy of the BEMA survey is presented in Appendix 5.

19. In static equilibrium, the potential difference between two points inside a solid piece of metal

(a) is zero because metals block electric interactions.
(b) is zero because the electric field is zero inside the metal.
(c) is non-zero if the piece of metal is not spherical.
(d) is non-zero if there are charges on the surface of the metal.
(e) is non-zero for reasons not given above.
(f) is zero for reasons not given above.

Statistical tests of the BEMA instrument have been established by Ding and colleagues consisting of individual item analyses (i.e., item difficulty index, item discrimination index, and item point biserial coefficient) and the whole test (i.e., test reliability and Ferguson's delta) using data from a sample of undergraduate students at Carnegie Mellon University and North Carolina State University (Ding et al., 2006). Meanwhile, the validity of the BEMA instrument was examined by eight faculty members at Carnegie Mellon University (CMU) involved in teaching undergraduate students in Electrical and Magnetics courses at various levels. Ding et al. (2006) indicated that the difficulty index values of BEMA items ranged from slightly below 0.2 to slightly above 0.8, demonstrating the desired range. In addition, the discrimination index values of BEMA items vary between 0.2 and 0.6, indicating that BEMA items have quite satisfactory discriminatory power. The average point biserial coefficient for the BEMA instrument was 0.43, showing that BEMA items have fairly high correlations with the whole test. In their study, the reliability index of the BEMA instrument using Kuder-Richardson 21 (KR-21) was 0.85, suggesting a satisfactory level of reliability. Meanwhile, Ferguson's delta of BEMA instrument was 0.98, which is considered to offer good discrimination. According to these statistical tests, BEMA is a reliable assessment tool with sufficient discriminatory power to probe students' conceptual understanding of the Electricity and Magnetism topics (Ding et al., 2006).

In addition, Ding (2014b) carried out a Rasch model analysis to measure person and item reliability, item and person estimates, and item fit. Person reliability in Rasch analysis is equivalent to the conventional Kuder-Richardson 20 (KR-20) or Cronbach's alpha coefficient. Based on the Rasch analysis results, the person reliability of BEMA was 0.78 and the item reliability was 0.96, indicating a satisfactory level of reliability. Meanwhile, the person–item map suggested that although BEMA may be challenging to some students, the Rasch model fitted BEMA items with student participants well (Ding, 2014b). In the Rasch model, the fit statistics that are commonly reported are the mean squares (the average of squared residuals) of each item. There are two types of mean squares, namely infit mean squares and outfit mean squares. The infit statistics give more weight to persons whose ability levels are close to the item difficulties, whereas outfit statistics gives equal weight to all persons, including outliers (Bond & Fox, 2015). Ding (2014a) reported that the majority of the BEMA items have both satisfactory infit and outfit mean squares. However, he suggested that a few items need to be modified or revised in future studies. According to this Rasch model, the results demonstrated the existence of a unidimensional construct among the BEMA items covering a broad range of Electricity and Magnetism topics.

As the empirical investigation of the construct of the BEMA remains poorly examined in previous studies, further statistical measurements are needed to investigate the construct validity of the BEMA instrument (Ding, 2014b). Chapter 5 provides an in-depth explanation of the statistical analysis for the BEMA instrument used in this study. Through the Rasch model analysis, all raw scores obtained from the participants' responses were first transformed into an interval scale to generate Weighted Likelihood Estimate (WLE) scores prior to conducting any analysis.

3.3.2 Phase 2: Qualitative Study

Data collection, analysis, interpretation, and report writing in qualitative studies are very different from those employed in quantitative studies. In general, qualitative procedures concern "purposeful sampling, a collection of open-ended data, analysis of text or pictures, representation of information in figures and tables, and personal interpretation of the findings" (Creswell, 2013, p. 22). Data collection techniques such as interviews, focus group discussions, and case studies are used in qualitative research (Adhabi & Anozie, 2017; Punch, 2013). More specifically, the interview method is a widely-used and well-accepted approach for data collection in qualitative research (Creswell, 2012; McGrath, Palmgren, & Liljedahl, 2019), that can be used to address a number of research questions. An interview is conducted to explore the meanings and perceptions of respondents to elicit more in-depth and rich information to answer the research questions (Adhabi & Anozie, 2017). Open-ended questions are asked verbally which might otherwise be difficult to carry out through quantitative research. Using interviews allows the researcher to gain a deeper understanding of the data in relation to the complex issues which cannot be obtained simply by responding to survey instruments.

In general, there are three types of interviews, structured, unstructured, and semi-structured (Adhabi & Anozie, 2017; Mueller & Segal, 2014). According to Adhabi and Anozie (2017, p. 89), the structured interview is mostly controlled by the interviewer, and provides a less flexible and casual interview environment for the participants due to its rigid format. The types of questions asked by the interviewer are commonly very short, and participants are expected to respond to these questions with short and straightforward answers. Structured interviews are most suitable for quantitative data. In addition, unstructured interviews are qualitative data collection tools that commonly originate from the ethnographic tradition of anthropology where this type of interview is conceptualised as a narrative interview (Adhabi & Anozie, 2017; DiCicco-Bloom & Crabtree, 2006). Meanwhile, semi-structured interviews provide more flexibility in collecting data or particular information from participants (Adams, 2015; Adhabi & Anozie, 2017; Newton, 2010; Schultze & Avital, 2011). Although there are a set of questions that have been developed and predetermined in the interview guide, the participants' responses provide opportunities for the researcher to ask further questions to clarify the responses, and the participants also have the freedom to express their views. In addition, semi-structured interviews are the data collection techniques or interview methods that are most widely used among qualitative researchers (Alshenqeeti, 2014). Semi-structured interviews can be conducted with individuals or groups. Creswell (2012) pointed out that interviews serve the purpose of providing participants with a more comfortable environment in which to speak or share ideas, and this enables the researcher to gain a deeper understanding of a range of topics that reflect the problems being investigated.

In this study, individual semi-structured interviews were carried out to achieve an in-depth understanding of the participants' ideas and perceptions in order to answer the research questions (i.e., RQ3 and RQ4) presented in Chapter 1. The interview sessions were conducted after all participants had completed all the survey instruments, and the quantitative scores had been obtained in the first phase of the study. The results of this preliminary data analysis were intended to select participants who could be involved in the interview sessions, as identified by obtaining high, medium, and low average scores. Before the main qualitative research was undertaken, a pilot interview was conducted with five volunteers who participated in the quantitative stage. In this study, trialing the interviews allowed the researcher to practice the interview techniques, to examine the feasibility of the interview questions, and to monitor the timing of the interview which, in turn, made it possible to modify the interview questions. Based on feedback from the participants, some minor corrections were made to the interview questions and minor adjustments were made to the order of the questions.

Further to this, interviews can be conducted in many ways, such as face-to-face, by telephone, or email, or in a focus group (Creswell, 2014a). Face-to-face interviews were preferred in this study. Before interviewing the participants, the researcher made appointments by phone and asked about their availability to be interviewed, as well as arranging the time and location of the interview. Then, the researcher organised a schedule for all the interview sessions for time efficiency, so that they were more manageable. The face-to-face interviews were held in a quiet room at the participant's university, or at a place requested by the interviewees. This was done to create a conducive environment during the interviews. As suggested by McGrath et al. (2019), interviews should be conducted at a time and place that is suitable for the participants, and it is very important to ensure a comfortable atmosphere that is free from distractions or noise. In this study, to maintain consistency, all the interviews were conducted by the researcher only. In total, 25 participants were selected for the individual semi-structured interviews, with each lasting between 60 and 90 minutes on average. All interviews were conducted in Bahasa Indonesia and recorded with the permission of the interviewees. Furthermore, all participants were informed that participation in this study was voluntary and that they could refuse or withdraw their consent at any time.

In sum, the qualitative data were collected through individual semi-structured interviews with 25 pre-service physics teachers who were purposefully selected, until the qualitative data had been considered saturated. As noted by Weller et al. (2018, p. 2), 'saturated' is defined as "the point during a series of interviews where few or no new ideas, themes, or codes appear." The interview data complemented and enriched the quantitative findings, probing the participants' perceptions of the relationship between scientific thinking (i.e., epistemological beliefs, argumentation, scientific reasoning) and their conceptual understanding of physics. The participants' perceptions of current physics teaching and learning practices in relation to the opportunities and barriers they experienced in enhancing their skills to think scientifically (i.e., epistemological beliefs, argumentation, scientific reasoning), and understanding of physics concepts during their higher education studies were also explored. A set of open-ended questions for the semi-structured interviews was developed as a guide to maintaining focus and interaction during the collection of the qualitative data. The interview protocols are presented in Appendix 6. All the interviews were audiotaped and later transcribed verbatim. These transcripts provided a word-for-word copy of the interview data for further analysis (Clark et al., 2017; McGrath et al., 2019). In this study, the transcripts were imported into the qualitative data analysis software package NVivo v11 (QSR International Pty Ltd, 2016) for data analysis. NVivo software allows users to manage, explore, and find patterns in datasets used in research for further analysis. As pointed out by Bazeley and Jackson (2013), NVivo provides a systematic and efficient procedure for storing, managing, organizing, and coding large amounts of data. The qualitative data analyses employed in the study are described in Chapter 4.

3.4 Ethical Considerations

Prior to the data collection, permission and ethics approvals to conduct the study were submitted to the relevant parties. The researcher asked permission from the Dean and Head of the Physics Education Program to distribute surveys and conduct interviews with physics education undergraduate students at four universities in Yogyakarta. The ethics approval to carry out the primary data collection was granted by the Social and Behavioural Research Ethics Committee (SBREC) at Flinders University on 10 April 2017 with project number 7606 (see Appendix 7).

During the data collection process, the researcher introduced herself as a PhD student at Flinders University and as a lecturer at the Indonesia University of Education. In order to minimise the risks of the research or to mitigate the ethical issues, the participants were informed about the purpose of the research, any benefits that might derive from the study, as well as the stages in the data collection process. Their participation was voluntary, and all the study data were treated as confidential and completed anonymously. The participants understood the nature of the research and were guaranteed certain rights. They agreed to be involved in the study and acknowledged that their rights were protected. The participants were free to withdraw their involvement from the study at any time, and were informed that they would not be penalised if they chose to do so. In the quantitative phase of the study, the researcher did not provide a consent letter to the participants regarding their willingness to participate in the study. Those who participated by completing the surveys were taken as having consented. Meanwhile, in the qualitative phase, the researcher sought the participants' consent to record all conversations during the interviews. The participants were offered an informed consent form to sign indicating their voluntary willingness to participate in the study.

3.5 Summary

This chapter has presented the research design and methods that underpin the data collection process, including a description of the instruments employed in the study. Both quantitative and qualitative data sources are used to achieve more comprehensive research findings. The procedures, sample, and research site, including the processes of recruitment of the participants both in the quantitative and qualitative phases of the study have been documented to provide a basis for conducting the mixed methods design in the current study. Finally, the ethics considerations have been explored.

A sequential explanatory design was used to undertake this research in which the quantitative data collection was conducted in the first phase of the study, followed by the qualitative study of a sub-sample of the participants to gain a deeper understanding and to enrich the quantitative findings. This study is also weighted unequally; the quantitative data were prioritised over the qualitative data. To collect data in the quantitative phase, five survey instruments were administered to 706 Indonesian pre-service physics teachers. The focus of the quantitative phase was to probe aspects of scientific thinking in relation to epistemological beliefs, argumentation, and scientific reasoning of Indonesian pre-service physics teachers, as well as their conceptual understanding of physics. In the quantitative study, Rasch analysis and Confirmatory Factor Analysis (CFA) were employed to validate the survey instruments. In addition, the relationships between the research variables were examined using the Structural Equation Modelling (SEM) approach. Meanwhile, individual semi-structured interviews were undertaken with 25 pre-service physics teachers to explore their perceptions about the context of the research. The results of both the quantitative and qualitative phases of the study were analysed separately and integrated at the discussion stage. The data analysis methods for both the quantitative and qualitative phases of the study will be explored in the following chapter.

CHAPTER 4
METHODS OF DATA ANALYSIS

4.1 Overview

In order to investigate several aspects of scientific thinking (specifically relating to epistemological beliefs, argumentation, and scientific reasoning), and the understanding of physics concepts among pre-service physics teachers, as well as to examine the causal relationships between these research variables, it is necessary to consider the appropriate analytical techniques to use in this study. This chapter, therefore, provides a detailed description of the data analysis methods for both the quantitative and qualitative data collected in this research. Considering the strengths and weaknesses of both data sources, a triangulated design was employed to obtain a more comprehensive perspective on the research questions. Nevertheless, the process of data collection, as well as data analysis, was more focused on the quantitative data rather than the qualitative in order to achieve the goals of the research. The qualitative data collected in this study aimed to complement the data collected through the quantitative methods. The qualitative study also functioned as a strategy to enrich and give further explanation to the quantitative findings.

The analysis of the quantitative and qualitative data was carried out to achieve the research objectives by addressing the research questions and was supported by appropriate statistical techniques and procedures. The following are the research questions formulated for this study:

1. Are there any differences in demographic factors with regard to pre-service physics teachers' scientific thinking and conceptual understanding of physics?
    a. Are there any gender differences in pre-service physics teachers' epistemological beliefs, argumentation, scientific reasoning, and conceptual understanding of physics?
    b. Are there differences between year level with regard to pre-service physics teachers' epistemological beliefs, argumentation, scientific reasoning, and conceptual understanding of physics?
    c. Are there differences between university type with regard to pre-service physics teachers' epistemological beliefs, argumentation, scientific reasoning, and conceptual understanding of physics?

2. What are the relationships between pre-service physics teachers' scientific thinking (i.e., epistemological beliefs, argumentation, and scientific reasoning), conceptual understanding of physics, and their demographic factors?
3. What are pre-service physics teachers' perceptions of the relationships between scientific thinking (i.e., epistemological beliefs, argumentation, and scientific reasoning) and their conceptual understanding of physics?
4. What are pre-service physics teachers' perceptions of the opportunities and barriers in enhancing their scientific thinking and conceptual understanding of physics?

RQ1 and RQ2 were addressed using quantitative analysis, while RQ3 and RQ4 were addressed through qualitative analysis. Hence, several software packages were employed to analyse these complex data sets. The data sets in this study only contained information on the constructs gathered from undergraduate students studying physics education as a major and did not include information from academic staff or stakeholders within the universities. This chapter begins with a description of the general methodological considerations, which includes the missing values and tests for normality and multicollinearity. This will be followed by an overview of the quantitative data analysis covering reliability and validity tests of the research instruments, descriptive and inferential statistical analysis, as well as the Rasch model for item analysis. The Confirmatory Factor Analysis (CFA) for testing construct validity and Structural Equation Modelling (SEM) analysis for examining the complex set of relationships arising between variables will also be illustrated. Finally, the qualitative data analysis methods will be provided to deepen the understanding of the quantitative findings.

4.2 General Methodological Considerations

The research instruments play a vital role in the present study as they underpin the empirical investigation, e.g., the relationships between the pre-service physics teachers' epistemological beliefs, argumentation, scientific reasoning, and conceptual understanding of physics. As the quality of the data analysis is crucial in generating credible and meaningful findings, ensuring a proper technical analysis is fundamental. In order to determine the most appropriate technique for analysing the data sets in this study, there are a number of general methodological issues that should be considered. These methodological issues which relate to missing values, and tests for normality and multicollinearity, are briefly described in the following sections.

4.2.1 Missing Values

Missing data is almost unavoidable for researchers, especially for fairly large-scale studies using survey instruments. By and large, missing data occurs in the process of filling out research questionnaires by participants. Some participants may provide an incomplete response to an entire research questionnaire (Pituch & Stevens, 2016), or there may be the problem of respondents being reluctant to answer all questionnaire items because the questions might not be relevant to their situation (Cheema, 2014). This could be due to personal reasons. As noted by Pituch and Stevens (2016, p. 18), some participants do not provide complete responses because they are not motivated to answer all the questions and eventually stop responding completely. In the same vein, Pigott (2001) revealed that missing data can be caused by a number of issues, such as the participants may be reluctant, or might forget, to respond to some questions, files being lost, or mistakes being made in data entry. Cases of missing data are beyond the control of the researcher and are certainly not expected. Consequently, this matter might lead to poor research findings and conclusions if not handled properly. This study certainly has missing data that needs to be considered before conducting subsequent data analysis. The presence of missing data can reduce the sample size, whereas the data analysis process may generate biased results if the data contains non-random missing data (Hair, Black, Babin, & Anderson, 2014). Hence, it is important to consider an appropriate statistical technique for the specific circumstances of a researcher dealing with such missing values before carrying out further statistical analysis. These circumstances can be related to "…sample size, proportion of missing data, method of analysis …" (Cheema, 2014, p. 54).

Many different approaches have been developed with standard statistical techniques for handling data sets with missing values in multivariate data. Little and Rubin (1987, 1990, as cited in Darmawan, 2002, p. 51) outlined three approaches for dealing with missing values. These techniques include complete case analysis (listwise deletion), available case methods (pairwise deletion), and filling in the missing values with estimated scores (imputation). Listwise deletion is the easiest and simplest approach to missing data in a multivariate data set. In this approach, the cases that have missing data on one or more variables will be excluded from the statistical analysis. Eliminating all cases containing missing data on the observed variables can potentially lead to loss of information when incomplete cases are large in number (Allison, 2003; Darmawan, 2002). According to Graham (2012), listwise deletion is useful for handling multiple regression analysis and structural equation modelling (SEM). Meanwhile, pairwise deletion can be a simple alternative to listwise deletion that serves to retain more data. As revealed by Allison (2003), the pairwise deletion method uses all cases that have complete values for each variable, or each pair of variables, based on each analysis. Although pairwise deletion does not seem to cause much loss in the sample size, it makes data incomparable because the sample base changes on an analysis-by-analysis basis depending on the missing value patterns (Darmawan, 2002). Cases of missing data can also be resolved by employing the imputation approach. Graham (2012, p. 51) stated that "mean substitution is a strategy in which the mean is calculated for the variable based on all cases that have data for that variable. This mean is then used in place of any missing value on that variable." The practice of inserting the mean in the missing value with the mean of the variable could apparently solve the problem of incomplete data at the beginning of the analysis. However, this method of imputation distorts the covariance structure which leads to the underestimation of variances and produces estimates of covariance biased towards zero (Darmawan, 2002). Hair, Black, Babin, and Anderson (2010) further argued that using the mean for missing data as the imputation method is best applied when the missing value is less than 10% of the entire data set and the relationship between variables tends to be strong.

In the present study, each questionnaire responded to by the participants was checked in detail by the researcher to ensure that all items from each survey question were answered completely. If an item was found to have a missing response from the participants, then those concerned were contacted and asked to complete it. Empirically, there were ten participants who did not complete the five questionnaires used in this study. The researcher attempted to contact them and ask for their willingness to complete these five surveys. However, since these participants did not give a positive response to these requests, and seemed reluctant to complete the whole survey, the researcher considered that these participants had withdrawn their involvement from the research. Therefore, for the aforementioned ten participants who did not complete the whole survey, the researcher decided not to involve them in the subsequent stage of the research. Because the proportion of missing data was 5% or less, the strategy suitable to be applied in this study relating to the issue of missing data was the listwise method. In other words, the researcher decided to discard those participants who failed to complete the surveys from the sample of this study and only carried out statistical analysis on the participants who had complete data for all of the research variables. As highlighted by Darmawan (2003), applying the listwise deletion approach is seen as a reasonable strategy and a simple way to handle the problem of missing data when the missing data in the study is less than 5% of total cases. This approach was used in this study, as the missing data were very minimal, thus ensuring that each of the analyses performed contained the same number of cases (Kline, 2011).

4.2.2 Tests for Normality and Multicollinearity

As multivariate data analyses were undertaken in this study, it was important to examine the normality of the variable distributions as well as multicollinearity before moving into further analysis. According to Hair et al. (2014), normality refers to the formed data distribution of the research sample and its correspondence to a normal distribution. To identify the normality of data distribution, researchers can use the absolute values of skewness and kurtosis of the sample responses. Skewness indicates the symmetry of a distribution, while kurtosis indicates the peakedness or flatness of a distribution in comparison to a normal distribution (Pallant, 2011; Tabachnick & Fidell, 2001). A simple way to check the normality of the data is by means of a histogram generated by the SPSS (Statistical Package for Social Sciences) software. Graphically, the shape of the curve in normally distributed data looks like a bell, where the spread of sample responses around the mean is symmetrical. As documented by Kline (2011), in normally distributed sample scores, the values of skewness and kurtosis are close to zero. Kline (2011, p. 60) further pointed out that a positive skew distribution shows the majority of scores are distributed below the mean, while a negative skew shows that most of the scores are distributed above the mean. In addition, a positive kurtosis depicts heavier tails and a higher peak than the normal distribution, while a negative kurtosis shows the opposite distribution. For the purpose of this study, the accepted absolute value for skewness was less than 3, while it was less than 8 for kurtosis, as suggested by Kline (2011).

Prior to conducting the final analysis to examine the relationships among the variables, it is also essential to test the multicollinearity among the independent variables. Multicollinearity is the statistical term used to describe the problem that exists when two or more independent variables become highly related to each other (Ben, 2010; Khine, 2013). As stated by Kline (2011, p. 51), "extreme collinearity can occur because what appear to be separate variables actually measure the same thing." When multicollinearity occurs, the results of certain statistical tests could be biased. For instance, multicollinearity can have an adverse effect on the regression, multiple correlation coefficients, and standard errors of the regression coefficients, as well as on the accuracy of computations (Darmawan & Keeves, 2006). To detect any potential multicollinearity issues between the independent variables, the Variance Inflation Factor (VIF) and Tolerance were employed in this study. Variance Inflation Factor (VIF) analysis, is a statistical technique that involves calculating the regression coefficients to detect the correlation between the independent variables (Field, 2013; O'Brien, 2007). VIF analysis can be easily carried out on regression diagnosis procedures through programs for general statistical analysis, such as the SPSS software program. Meanwhile, related to the VIF is the tolerance statistic, which is its reciprocal (1/VIF)(Field, 2013, p. 990). In addition, the rule of thumb in this study was a VIF value of more than 10 and a Tolerance value below 0.1, which indicate serious multicollinearity, and so, such variables were considered to be redundant. According to Field (2013), if multicollinearity is identified, removing highly correlated variables is suggested in order to reduce the effects of multicollinearity, and multicollinearity is then re-reviewed. However, O'Brien (2007) stressed that caution must be exercised if excluding such a variable, because there may be other factors that influence the results of the VIF or the Tolerance analysis.

4.3 Quantitative Data Analysis

This research includes multiple variables and scales that require multiple statistical analysis techniques as well as several software programs to analyse the complex data sets. All data collected from the participants' responses on the questionnaires were entered into a spreadsheet using Microsoft Excel software. Each item entered into the Microsoft Excel software was assigned a code and then exported as an SPSS file format for data tidying and carrying out descriptive statistical analysis for descriptive information on the samples. In addition, of the five questionnaires used in this study, there was one questionnaire that required reverse scoring, namely the *Epistemological Beliefs Assessment for Physics Science* (EBAPS) questionnaire that was used to measure participants' epistemological beliefs. So, these items were re-coded and saved as a different file with the addition of an 'R' suffix. The SPSS data files were also converted to the ASCII format for use in other software applications employed in this study.

The data saved in the SPSS format became the raw data that were then evaluated for the accuracy of data entry using a data screening and cleaning procedure through the examination of the frequency analysis and the basic descriptive statistics. According to Creswell (2012, p. 181), "cleaning the data is the process of inspecting the data for scores (or values) that are outside the accepted range." Any errors detected through such data screening and cleaning processes were then corrected and repeated until all the data were free of errors. The data were subsequently classified to appropriate numerical form for use in some of the specialised software packages employed in this study. For instance, numbers were assigned to items such as gender (i.e., 0 for female and 1 for male), university type (i.e., 0 for private university and 1 for public university), and year level (i.e., 1 for year level 1, 2 for year level 2, 3 for year level 3, and 4 for year level 4). In addition, the data in SPSS file format were also converted to the tab-delimited i.e., dat file format for data processing. This was to generate the Weighted Likelihood Estimates (WLE) score by using the Rasch model for further analysis, such as descriptive statistics analysis, Confirmatory Factor Analysis (CFA), and Structural Equation Modelling (SEM), which are described in the following sections.

4.3.1 Reliability and Validity

Before carrying out further data analysis, it was necessary to establish the reliability and validity of the research instruments, as they are considered to be complementary to each other in research (Ben, 2010). In this study, the instruments used were adopted from various existing parts of the literature. Although it is implicit that the adopted instruments had already been validated and reported by the publishers or authors, it was important to re-validate the instruments used to ensure that they measured the factors investigated in this study. In the case of this study, the instruments were adopted for Indonesian pre-service physics teachers, which clearly differed from the settings or context (e.g., cultural background and cohort) in which they were originally used. Thus, it was necessary to establish both the reliability and validity of the research instruments used in order to ensure their suitability for the Indonesian context to ensure meaningful data interpretation and valid research findings. Establishing the validity and reliability of scores on intruments has been advocated by researchers, especially for quantitative research studies (Ben, 2010; White, 2011), with the argument being that empirical research should be valid and reliable.

The concept of reliability in quantitative studies refers to the measurement of consistency and stability of an instrument (Creswell, 2012). In a broader sense, the aim of the reliability test is to measure what it was designed to measure with only minimal errors in the scores, or scores being free of random error (Ben, 2010). The procedures frequently performed to examine an instrument's classical reliability are test-retest reliability and internal consistency. According to Field

(2013), test-retest reliability is determined by administering instruments to the same group of people at two different times and calculating the correlation between the two sets of scores obtained. In other words, each respondent completes the instruments twice at different time intervals. Meanwhile, internal consistency reliability refers to how consistent the individuals' responses are across items within a single test form in measuring the concept (Creswell, 2012). In addition, internal consistency reliability aims to measure "… the equivalence of sets of items from the same test" (Kimberlin & Winterstein, 2008, p. 2277). Internal consistency is a form of reliability test that is more commonly conducted than the reliability test-retest. To measure internal consistency reliability, Cronbach's coefficient alpha is one of the most commonly used indicators of scale reliability. The reliability coefficients or Cronbach's coefficient alpha in general ranges from 0 to 1. This implies that a higher value of Cronbach's coefficient alpha indicates a higher level of reliability. For estimates of internal consistency, Hair et al. (2010) suggested that a Cronbach's alpha coefficient of 0.70 and above is the acceptable reliability coefficient level to indicate consistent responses across the items on the instruments. However, van Griethuijsen et al. (2015) argued that a Cronbach's alpha coefficient of 0.70 or 0.60 is the acceptable reliability value.

Establishing classical reliability in research, however, has several shortcomings. As pointed out by Alagumalai and Curtis (2005), classical reliability estimates are affected by a number of factors such as measurement precision, group heterogeneity, and test administration settings (e.g., length of the questionnaire, and the time limit given to the test takers), which should be taken into account in designing or adopting the instruments for the study. Nevertheless, these shortcomings can be addressed by using the Rasch model analysis approach which is based on the logits scale (i.e., interval scale) rather than on the nonlinear raw score metric approach of classical reliabilities that might raise concerns relating to appropriateness. An explanation of the Rasch model is provided in more detail in the following sections. In this study, classical reliability was still analysed as additional information. Thus, to assess item performance in this study, the reliability of the questionnaires was examined using Cronbach's alpha, while item separation reliability was generated through the Rasch model analysis using the ConQuest software. As noted by Bond and Fox (2015), the reliability of an instrument can be assessed using the Rasch model. Low reliability of the instruments indicates large errors or imprecise estimates of the persons or items. However, other measures of validity can be carried out to determine the effectiveness of the instruments used in the study (Bond & Fox, 2015). According to Kane (2013), reliability is a necessary, though insufficient, condition for validity.

In a general sense, the concept of validity in research can be explained as "… the extent to which an instrument measures what it purports to measure" (Kimberlin & Winterstein, 2008, p. 2278). Validity has come to mean whether an instrument can measure what it is supposed to measure. In addition, establishing the validity of the respondent's individual score from an instrument helps to identify whether the instrument is good and appropriate to use in the proposed research (Creswell, 2013), and to examine the quality and effectiveness of the instrument (Alagumalai & Curtis, 2005). As highlighted by Hair et al. (2014, p. 600), "no valid conclusions exist without valid measurement."

As noted earlier, this study has adopted several existing questionnaires that have been validated by the authors and widely used by researchers. The instruments were given to the research supervisors to be checked in relation to their suitability to the research context before being administered to the potential participants. Subsequently, a pilot study was carried out involving 30 prospective science/physics teachers from one of the universities involved in this study before the main study was undertaken. The participants were encouraged to give comments on the questionnaires, check the clarity of the words and identify language problems in the translated text, and to check the meaning of the survey items. This phase also enabled the researcher to monitor the time required by the participants to complete the entire survey. Based on the feedback provided by the participants, only minor corrections were made to the survey questions. Thus, all the items in the entire survey were then used in the main study.

In the context of this study, the validation of the instruments was achieved by establishing construct validity. According to Kimberlin and Winterstein (2008, p. 2279), construct validity refers to "… a judgment based on the accumulation of evidence from numerous studies using a specific measuring instrument. Evaluation of construct validity requires examining the relationship of the measure being evaluated with variables known to be related or theoretically related to the construct measured by the instrument." Broadly, the construct validity of the instruments in this study was confirmed through the Rasch Model analysis and Confirmatory Factor Analysis (CFA) approaches. Rasch analysis was used to verify if the items fitted well with the Rasch model, while CFA was employed to examine the underlying structure of the scales. The use of both forms of measurement modelling can be considered as validation techniques or types of test validation that are complementary to each other (Ben, 2010). Hailaya et al. (2014) further advocated the use of both models of the approach to establish the validity of an instrument. The explanations of the Rasch analysis and the CFA model are presented in the following sections in further detail.

4.3.2 Descriptive and Inferential Statistics Analysis

Conducting a descriptive and inferential statistics analysis of the entire data set of the study is important for the preparation of the subsequent analyses. This step was carried out to obtain information about the nature of the data from the sample demographics and the research variables used in the analysis, to test the normality of the distribution of the data and multicollinearity, as well as to obtain some comparisons among the research variables tested in this study.

In the present study, frequencies and percentages were calculated for the nominal data, such as gender, year level, and university type. Meanwhile, means and standard deviations were calculated for continuous data in the observed cases. The descriptive statistics, including skewness and kurtosis, were also calculated to identify whether the data were normally distributed. Further, to compare the two independent groups regarding significant differences between the means, an independent t-test was used. Meanwhile, a one-way ANOVA approach can be employed when a comparison involves at least three groups (Field, 2013). In this study, a one-way Analysis of Variance (ANOVA) and a t-test of independent samples were employed to determine the significant differences in the means of variables among groups, i.e., gender, year level, and university type. In other words, the comparison was in terms of the significant differences between the means of the compared groups. The descriptive statistical, ANOVA, and t-test analyses were carried out using the SPSS 25 software program. Through the use of this software, the results of the analysis were generated in the tabulation, chart, and diagrammatic formats.

For inferential statistics, it is also important to calculate the strength of the difference between two means or two variables in order to determine whether it can be practically meaningful (Creswell, 2012); this can be carried out by calculating the effect size. Creswell (2012, p. 195) revealed that "effect size identifies the strength of the conclusions about group differences or about the relationship among variables in a quantitative study." In the present study, effect size between groups was measured using Cohen's *d* that can be easily calculated through the following equation (Field, 2013, p. 299):

$$d = \frac{\overline{x_1} - \overline{x_2}}{s} \quad (4.1)$$

Formula 4.1 shows that Cohen's *d* is simply the difference between two means (i.e., $\overline{x_1}$ and $\overline{x_2}$) divided by the standard deviation (*s*) of both groups. As highlighted by Cohen (1988, 1992, as cited in Field, 2013), the absolute value for effect size (*d*) was interpreted as: small (0.20), medium (0.50), and large (0.80). Calculating the effect size makes it possible to very clearly see the magnitude of the effect that has been observed, which can then be compared to other studies (Field, 2013).

4.3.3 The Rasch Model for Items Analysis

For the purpose of this study, the Rasch model was employed to undertake scale item analysis on questionnaires using multiple choice and Likert-type items. The Rasch measurement model was introduced by Georg Rasch, a Danish mathematician in the 1960s (Baker, 2001). Wright and Mok (2000, p. 84) revealed that "the Rasch model is a way to make sense of the world." Wright and Mok (2000) further asserted that there are a number of advantages when applying Rasch measurement models because it: (a) produces linear measures, (b) overcomes missing data, (c) gives estimates of precision, (d) has devices for detecting misfit, and (e) the parameters of the object being measured and of the measurement instrument must be separable (pp. 86-87). Rasch measurement also offers the benefit of designing or revising a measurement instrument through parametric statistical tests (Boone, Staver, & Yale, 2013).

In addition, the Rasch model which is classified under the family of Item Response Theory (IRT) models, can be applied as a complementary analysis in overcoming several shortcomings of the traditional measurement theory (or Classical Test Theory, CTT) methods in research instrument measurement (Alagumalai & Curtis, 2005). As noted by Kline (2005, p. 167), the basic concept of CTT is formulated as "… the raw score (*X*) on a test is made up of a true component (*T*) and a random error (*E*) component ($X = T + E$). The less random the error, the more the raw score represents the true score." The CTT method assumes that errors of measurement for the entire sample of respondents to whom the questions were administered are equal; whereas, in the real world, error estimates could be different according to the high and low abilities of the respondents. Furthermore, in CTT methods, item difficulties (i.e., correct response proportions), item discrimination (i.e., corrected item-total correlations), and reliability depend on the respondents' sample, while IRT does not depend on the respondents' sample to be used to generate the parameters.

According to Piquero et al. (2000), IRT is a model-based measurement that is used to model the relationships between a persons' responses and the properties of the survey items administered. More specifically, there are three different item response models that can be identified through the IRT model, namely the one-parameter, two-parameter, and three-parameter IRT models. In the one-parameter (1-PL) IRT model, the model allows items to vary in their difficulty level, does not contain pseudo-guessing, and assumes fixed item discrimination. Meanwhile, the two-parameter (2-PL) IRT model comprises of the item discrimination and item difficulty parameters; and the three-parameter (3-PL) IRT model caters for pseudo-guessing, the item discrimination and item difficulty parameters (Bond & Fox, 2015; Hays, Morales, & Reise, 2000). The Rasch model is referred to as the one-parameter (1-PL) IRT model, which is particularly used to model the data. Alagumalai and Curtis (2005) further stated that it is appropriate to employ the Rasch models to measure a scale, such as the Rating Scale Model (RSM), the Partial Credit Model (PCM), the multidimensional model, and the many facets model.

Furthermore, the role of Rasch measurement is emphasised in transforming raw scores into an interval scale in existing data sets. Raw data is categorised as the ordinal data that shows that "... the relative differences among values composing the scale are unequal in terms of what is being measured, permitting only a rank ordering of scores" (Harwell & Gatti, 2001, p. 105). Data containing ordinal scales will not fulfill the normality assumptions needed in many statistical procedures and can generate bias, therefore, ordinal data needs to be rescaled to interval data for statistical analysis purposes. As noted by Alagumalai, Curtis, and Hungi (2005), the Rasch measurement models are able to transform ordinal data into an interval scale. Using raw scores to express the results of an investigation in an analysis can lead to bias; e.g., misinterpretation in measuring the quality of the test as well as the achievement of the respondent. In addition, using the CTT models may also be problematic due to the assumption of equal errors of measurement for all test-takers, whereas error estimates can differ between persons with high and low ability. The presence of the Rasch model approach can help researchers to counter the challenges of scoring survey responses and test items in order to obtain a reliable numerical score.

Key Features of the Rasch Model

According to Wu and Adams (2007), in the Rasch measurement model, the probability of success on an item depends on two key features, namely the person's (or test-takers) ability and the difficulty of the test items (or survey items). "Rasch measures represent a person's ability as independent of the specific test items and item difficulty as independent of specific samples within standard error estimates" (Bond & Fox, 2015, p. 349). Each estimation of a person's abilities and item difficulties has their respective precision measures, namely the estimation of errors for each ability score. The estimated item characteristics do not depend on any particular cohort and estimates of a persons' abilities do not depend on the particular instrument administered. This means that the Rasch measurement model brings persons and items to a common scale that is independent, in which the data were collected, and the estimation of errors done for each individual person's ability and item difficulty, rather than for the instrument as a whole. "Then it is possible to estimate probability of success for persons of any ability on items of any difficulty" (Bond & Fox, 2015, p. 302). Furthermore, Bond and Fox (2015, p. 11) pointed out that the basic principle of the Rasch model approach is:

> a person having a greater ability than another person should have the greater probability of solving any item of the type in question, and similarly, one item being more difficult than another means that for any person the probability of solving the second item is the greater one (p. 117).

In other words, if a question is considered easy to answer by test-takers, then those who have higher abilities are more likely to be able to answer such a question correctly.

The Rasch model was initially introduced to analyse dichotomous item responses and was further developed so that it could also be used to analyse items of polytomous response categories (Ben, 2010), such as Likert-type questions. The Rasch model with dichotomous data is the simplest model of the Rasch family models. In the dichotomous model, the person's ability and item difficulty are estimated from the proportion of correct or incorrect responses for each item and person. This model requires binary responses to analyse items such as multiple-choice questions, where a score of "1" is given for the correct response and a score of "0" for the wrong response. Estimates of item difficulties are calculated from the proportion of persons who succeed on each item, while a person's abilities are calculated from the proportion of items that each person succeeds on. In the Rasch modelling approach, the probability of success in getting the item right is modelled as a logistic function that puts a person's ability and test item difficulty on a common scale known as the logit (i.e., log-odds) scale, and both parameters are sample independent (Bond & Fox, 2015). As noted by Jackson et al. (2002, p. 235), “a logit is a unit of measurement used in Rasch analysis for calibrating items and measuring persons, based on the natural logarithmic odds of the probability of a response.”

In some cases, test item responses can reflect a degree of correctness in answering a question, instead of just simply being correct or incorrect. In other words, instruments may include three or more response categories (e.g., Likert type questions). To model these item responses, the partial credit model (PCM) and rating scale model (RSM) can be employed to perform item analysis (Wu & Adams, 2007). The partial credit model (PCM) is a model for constructing measures using items with two or more ordered response categories; namely, polytomously scored items (Masters, 2016). As documented by Bond and Fox (2015), Geoff Masters is recognised as the developer of the partial credit Rasch (PCM) models for polytomous data. PCM is a simple adaptation of the Rasch model, in which each item has its own threshold that is independent of the threshold of other items on the same test. In other words, the partial credit model allows a number of ordered response categories where the threshold values may vary for different items. Bond and Fox (2015, p. 141) asserted that “the partial credit model is highly applicable in educational and other testing situations in which ‘part marks’ are awarded for ‘partial success’, that is, for progress between complete failure and complete success on that item.” The PCM response categories must be ordered to reflect increasing competence of some of the person’s abilities. For instance, the ordered values 0, 1, and 2 might be applied to a test item such as 0 for totally wrong, 1 for partially correct, and 2 for completely correct. This indicates that higher scores are closer to total success on an item test. In other words, a person with higher abilities is more likely to score higher for the items than a low ability person (Wu & Adams, 2007).

Meanwhile, RSM was developed by David Andrich and Earling Andersen (Bond & Fox, 2015). RSM is similar to PCM in treating polytomous data. As stated by Masters (2016), RSM can be used to analyse a questionnaire that uses a fixed set of response categories for each item, such as "strongly disagree", "disagree", "agree", and "strongly agree", although this type of questionnaire can also be analysed by using the PCM. According to Bond and Fox (2015, p. 370), “the RSM constraint requires that every item in a test has the same number of response options and applies the one set of response threshold values to all items on the test.”

As mentioned previously, because this study involves multiple choice and Likert-type items, the Rasch model, which is appropriate for both dichotomous and polytomous data, was employed. For the purposes of this study, the Rasch analyses adopted PCM instead of RSM for the polytomous item responses, specifically to analyse the data collected from the Epistemological Belief questionnaire. As noted by Martin, Mullis, Foy, and Arora (2011, p. 2), "partial credit IRT scaling is based on a statistical model that relates the probability that a person will choose a particular response to an item to that person's location on the underlying construct." Furthermore, the justification for the use of PCM can be taken from the following assertions: (a) credits are given for partially correct answers, (b) there is a hierarchy of cognitive demand on respondents in each item, (c) each item requires a sequence of tasks to be completed, and (d) there is a batch of ordered response items with individual thresholds for each item (Wright & Mok, 2004, pp. 22-23).

According to Bond and Fox (2015), a unidimensional model is required when the data sets are analysed using the Rasch measurement model, in which each test item contributes in a meaningful way to the measurement of a single underlying trait. Furthermore, analysis of multidimensional models may also be useful if the scale contains a hierarchical structure or several sub-scales. As documented by Adams and Wu (2010, p. 15), previous researchers have tended to analyse the scales that contained several sub-scales "… by either applying a unidimensional model to each of the scales separately or by ignoring the multidimensionality and treating the test as unidimensional." Multidimensional calibration may be more desirable because calibration methods that apply unidimensional models to each scale separately have weaknesses. The multidimensional model was employed in this study without violating the requirement of unidimensionality of the scale in which this model simultaneously calibrates all sub-scales. Adams, Wilson, and Wang (1997) pointed out that in the analysis of a multidimensional measurement structure, each sub-scale has unidimensional characteristics.

In addition, the multidimensional model approach is useful for estimating correlations between sub-scales in order to improve measurement precision, in which "the greater the correlations, the greater the measurement precision" (Wang, Yao, Tsai, Wang, & Hsieh, 2006, p. 608). In this study, the estimated correlation between the dimensions or sub-scales were also examined to investigate associations among dimensions, as well as to validate the dimensionality of the instruments. To interpret a correlation coefficient, rules of thumb are employed as suggested by Cohen (1988, as cited in Berben, Sereika, & Engberg, 2012), namely small (0.10), medium (0.30) and large (0.50).

The use of ConQuest 4 Software for Rasch Analysis

In this study, the ACER ConQuest 4 computer software package was used to carry out item analysis for multidimensional item response models, which was based on Rasch measurement techniques. As noted by Bond and Fox (2015), the scale containing dichotomous or polytomous data can be constructed and validated using the ConQuest software. ConQuest is a modeling program which is able to fit a range of item-response models such as both unidimensional and multidimensional item-response models, as well as latent regression models using a special keyword syntax.

In the data entry section, it is important to note that the ConQuest software program requires the data to be in text file format (ASCII format). Hence, the raw data were prepared using the Microsoft Excel spreadsheet software then converted to the SPSS file format. Next, the SPSS files were also converted to tab-delimited data files in ASCII (or text) format in which each variable is entered on fixed columns in a file. In this study, the ACER ConQuest 4 default method of parameter estimation used the Marginal Maximum Likelihood (MML) estimator (Wu, Adams, Wilson, & Haldane, 2007). Furthermore, the outputs of ConQuest provide item response model parameter estimates as well as graphics for items, allowing the examination of the properties of assessments, traditional assessments, and rating scales. For instance, the ConQuest software can generate a pattern of infit and outfit mean square values for items, t values, item difficulty, and discrimination indices.

Fischer and Molenaar (1995 as cited in Bond & Fox, 2015, p. 45) revealed that “the Rasch analysis software programs perform a logarithmic transformation of the item and person data to convert those ordinal data to yield interval data.” Bond and Fox (2015) also noted that the logistical functions place a person’s abilities and test item difficulties on common scales called the logit scale (i.e., log-odds unit). The logarithmic transformation in Rasch measurement enables the researcher to compare a person’s level of ability and the difficulty level of the item, because they share the same unit (i.e., logit scale), and how much more able one test-taker is than another. In the Rasch measurement model, the unit of measurement on the scale is one logit (the unit of logarithm chances). The logit scale does not have a limited range of values. Theoretically, the log odds scale of either a person’s ability or item difficulty can range from minus infinity ($-\propto$) to plus infinity ($+\propto$). According to Wright and Mok (2004), the probability (or odds) of a random person succeeding on a given item is mathematically expressed as a logistic function as follows:

$$P_{ni} = \frac{\exp(B_n - D_i)}{1+\exp(B_n - D_i)} \qquad (4.2)$$

where P_{ni} is the probability of person 'n' in responding item 'i' correctly, while B_n and D_i represent a person's ability and item difficulty respectively (p. 11). This revealed that the probability of a person getting a correct response is dependent upon the person's ability and item difficulty. Thus, the person's ability and item difficulty can be placed on the same measurement scale for interpretation.

Bond and Fox (2015) further pointed out that a person's ability and an item difficulty that have negative logit estimates indicate the low level of a person's ability and an item difficulty. Conversely, a person's ability and an item difficulty that have positive logit estimates indicate the high level of a person's ability and an item difficulty. In addition, the average logit value is arbitrarily set at zero (0) to represent item difficulty. For instance, a person with an ability level of zero logits is indicated to be more able than a person with the ability of -1 logit; however, he/she is indicated to be less able than a person with the ability level of 1 logit.

Thus, the ConQuest software can also be used to visually observe the item–person variable map or known as the Wright map. This map visualises the relationship between person ability and item difficulty estimate on one scale which is located on the right and left along a linear vertical scale on the map respectively and measured in logits. In addition, the most able respondents and most difficult items are placed at the top (Bond & Fox, 2015). As noted by Bond and Fox (2015, p. 69), "the mean of the item difficulties is adopted by default as the 0 point." It can be illustrated that persons with ability levels above the average of the test items are placed in the positive part along the scale, while persons with ability levels below the average of the test items are placed in the negative part along the scale (Maley & Bond, 2007). Liu et al. (2008) further pointed out that the item difficulty distribution should cover the distribution range of the respondents' ability, thus providing an accurate measure of respondents' ability across the scale.

Person Fit and Item Fit Analysis

Further, ConQuest output also provides a goodness-of-fit statistic to determine how well the observed responses fit the Rasch measurement model, i.e. person and item fit (Arnadóttir & Fisher, 2008). The examination of the fit index is conducted not only for controlling the quality, but also to measure improvement for expected item function (Bond & Fox, 2015; Jackson et al., 2002). Based on the Rasch model analysis, the person's fit and item analysis were employed in this study to detect if there were any misfitting persons or items. This misfit can occur when the collected data does not adequately fit the Rasch model estimates. To detect the misfitting person or item, the infit MNSQ (weighted fit mean square) and outfit MNSQ (unweighted fit mean square) can be employed. As noted by Bond and Fox (2015, pp. 65-66), "the 'mean squares' is the unstandardized form of the fit statistic, and is merely the mean, or average value, of the squared residuals for any item." This value shows how much misfits are revealed in the existing data. Furthermore, Bond and Fox (2015, p. 66) stated that "the residuals represent the differences (i.e., the amount left over) between the Rasch model's theoretical expectation of item performance and the performance actually encountered for that item in the data matrix." Larger residuals may happen if an item has a greater difference between how the item should have performed according to the Rasch model and how it was actually carried out in the test.

Bond and Fox (2015) stated that infit statistics are a weighted indicator of misfit that give a relatively greater weight to responses (i.e., a person's performances/abilities) which are located closer to the value of the item difficulty. In addition, "an infit statistic indicates the degree to which the observations for a particular item meets the model expectations" (Jackson et al., 2002, p. 235). Meanwhile, the outfit statistic is unweighted in which the performance of persons is relatively influenced by outlying scores (e.g., unexpected responses) from the item's location (Bond & Fox, 2015). In other words, infit statistics focus less on outliers, while outfit statistics are more sensitive to outliers. As noted by Luo et al. (2009), the value of both infit and outfit statistics is expected to be close to one. Bond and Fox (2015) further argued that infit statistics are more of a concern than outfit statistics. Since outliers are not a concern in the proposed study, only infit statistics (or weighted fit mean square) were tested and reported.

In addition, a person fit analysis was first conducted to detect misfitting persons and items. The person fit analysis allows the researcher to detect participants who were misfits to the Rasch model. This misfit may be caused by a lack of understanding of the questions contained in the questionnaires and limited knowledge in providing the correct answers, or guesses in answering the questions in the questionnaires. In practical terms, respondents might guess the answer to a difficult item correctly, even though they have very low abilities, but there is still a greater probability of answering an item correctly (Crocker & Algina, 2006). In the next step, any misfitting persons were then temporarily removed in order to examine the infit MNSQ for each item as a basis for model fitting or misfitting items. The misfitting items, or items that were considered to not fit the model, were then deleted from the subsequent analysis. A misfitting item can occur if a difficult item can be answered correctly by participants who have low abilities (e.g., through guessing or thoughtless errors), but not all low-ability participants could answer the item correctly, although a number of them did. Conversely, an easy item could not be correctly answered by participants who have high abilities (Boone et al., 2013). Hence, it is important to note that examining fit statistics should be employed to detect misfitting items or a person's performance, not merely to determine which items should be removed from a test or questionnaire (Bond & Fox, 2015). However, extra caution needs to be exercised in removing any misfitting items, because the items might be valuable in providing important information or findings that might arise from the study which were not even considered as part of the study.

In fact, there are no definitive rules regarding the acceptable range for MNSQ fit values. The range can vary depending on the type of test and its purpose. However, it is necessary to define the acceptable range in order to indicate acceptable values or indices, or whether it fits or does not fit the model. In the present study, respondents with weighted mean square (INFIT MNSQ) values that fall outside of 2.00 are considered as misfitting (Wright, 1994), and were thus considered for removal. As noted by Wright (1994), an INFIT MSNQ of less than 2.00 indicates that the psychometric properties are not degraded.

Meanwhile, the threshold values for examining item fit refer to the threshold values proposed by Bond and Fox (2015), as shown in Table 4. 1, which are based on the type of test administered. In this study, the acceptable range for infit MNSQ values falls in the range of 0.70 – 1.30 and was employed for all items, which suggests a reasonable fit of the data to the model. Using the fit criteria suggested by Bond and Fox (2015), items that conformed to the measurement requirements were retained, while items with infit mean square values that fell outside the accepted range were then considered misfitting and removed with caution.

Table 4. 1 Reasonable ranges for item mean square infit and outfit (Bond & Fox, 2015, p. 273)

Type of Test	Range
Multiple-choice test (high-stakes)	0.80 – 1.20
Multiple-choice test (run of the mill)	0.70 – 1.30
Rating scale (Likert/survey)	0.60 – 1.40
Clinical observation	0.50 – 1.70
Judged (where agreement is encouraged)	0.40 – 1.20

Other than the fit statistics, ConQuest software can also be used to generate t-values or standardised fit statistics. The t-values indicate how well the data fits the model, which may be positive or negative. As indicated by Wu and Adams (2007), t-values that fall outside the range of -2.00 to 2.00 (or - 1.96 to 1.96) are considered to be an indication of a misfit at the 95% confidence level. According to Schumacker (2004), a t-value greater than 2.00 demonstrates unexpected response patterns on all items, while a t-value lower than -2.00 indicates the possibility of redundancy in item response. However, it is important to note that t-values are affected by the number of samples in the studies. For instance, collected data from large sample sizes tend to generate large t-values (Wilson, 2005); consequently, many misfitting items will be detected. Because the t-value is influenced by sample size, it was not a great concern in the current study. Hence, persons and items are regarded as misfitting only when they are indicated as misfitting according to the INFIT MNSQ statistics. However, Wu and Adams (2007) suggested that fit statistics should serve as an indication for detecting items that are problematic, and not arbitrarily for setting strict rules about whether an item can be accepted or rejected, which may lead to removing the best items in the test or questionnaire.

Furthermore, the discrimination index was also employed in this study to perform item analysis that is classified under item response models. As pointed out by Wu and Adams (2007), the item discrimination index should also be considered in inferring the overall assessment of an instrument. This discrimination index provides a correlation between a person's score on an item and his or her total score (i.e., from all items) on the test used, "… which reflects more about the amount of 'noise' in the data than fit statistics do" (Wu & Adams, 2007, p. 80). The discrimination index ranges from zero to one, and may be positive or negative. Items with zero discrimination values indicate no relationship between item scores and total scores, while items with positive discrimination values show a positive relationship between them (Wu & Adams, 2007). In addition, the items with negative discrimination values indicate that test takers with high ability tend to get these items wrong (incorrect), while test takers with low ability get them right (correct). Following Ding et al. (2006), it is important to revise or eliminate any items that have a negative discrimination index as there is no correlation between the item score and the total score. Wu and Adam (2007, p. 64) stated that "… the higher the discrimination index, the better the item is able to discriminate between people ...". In the same vein, Alagumalai and Curtis (2005) noted that discrimination indices indicate to what extent an item on the test distinguishes between high-ability and low-ability respondents. In the present study, any items which have a discrimination index value of less than 0.15 were eliminated from the data sets with deep consideration, while others were retained. As recommended by Rush, Rankin, and White (2016), these items were interpreted as poor items, and therefore, should be removed with extra caution as they may yet contain valuable findings.

Scoring Procedures

For the purpose of this study, the initial data collected through the participants' responses to items on the questionnaires were obtained in the form of raw scores, especially in the quantitative research. It is common for the estimate of a person's ability on a survey item to be expressed as a total raw score by summing the correct or incorrect responses (Bond & Fox, 2015), which are then treated as measures. However, Wright and Mok (2004) asserted that raw scores or counts cannot be treated as measures because they are only indications of possible measures. Using raw scores as measures may introduce bias and raise concerns about the usefulness of the conclusions made, which also becomes more problematic when advanced statistical models with multiple indicators such as Structural Equation Modelling (SEM) are applied (Bond & Fox, 2015). As highlighted by Alagumalai and Curtis (2005, p. 10), "raw scores add further ambiguity to measurement as student abilities, which are based on the total score obtained on a test, cannot be compared." Wright and Mok (2004, p. 2) further revealed that "raw counts cannot be the measures sought because in their raw state, they have little inferential value. To develop metric meaning, the counts must be incorporated into a stochastic process which constructs inferential stability." This implies that the raw scores should not be used in the analysis. Furthermore, Ben (2010) suggested that the raw scores need to be transformed into measures before proceeding to further stages of data analysis in order to achieve uniformity as well as to examine the psychometric properties of the scales.

As documented by Ben (2010), there are several estimation methods that can be used to transform scores into measures such as the Maximum Likelihood Estimation (MLE), Bayes Modal Estimation (BME), Expected A-Posteriori (EAP), Marginal Maximum Likelihood Estimation (MMLE), and Weighted Likelihood Estimation (WLE). However, Warm (1989 as cited in Ben, 2010) argued that WLE is a better estimation method compared to other transformation techniques which are more biased, because the WLE method can reduce the bias as well as the standard error of the estimates. Wang and Wang (2001, p. 318) stated that "bias can systematically affect the precision of the cut score and, consequently, the validity of the classification decisions." Hence, transforming raw scores to measures by using the WLE method could provide better estimates of students' scores.

The WLE score is presented in the form of a logit (log odds unit) scale as an interval scale in which "the distances between scale units are made equal and meaningful" to enable estimates of a person's ability (Bond & Fox, 2015, p. 22). As noted by Bond and Fox (2015, p. 30), "… a log odds scale avoids the problem of compression at the ends of the raw score scale due to its restricted range, leading to floor and ceiling effects." Zhang and Lu (2007, p. 1) further stated that employing the WLE method is an effective way to reduce bias. Furthermore, the WLE method has been used as part of data analysis techniques in large-scale international studies such as PISA (or the Programme for International Student Assessment) and TIMSS (or Trends in International Mathematics and Science Study) (Adams & Wu, 2003; Martin et al., 2011). This implies that using WLE scores instead of the raw scores is recommended in order for measurements to be more useful for inference. Hence, the WLE method was employed in this study with the consideration that such a method can minimise estimation bias (or lessen the bias), make interpretations more meaningful, and be consistent with what has been used in large-scale studies such as PISA and TIMSS. The ACER ConQuest 4 statistical software package was employed to obtain the WLE scores using the anchoring approach. In the case of this study, WLE scores were obtained once the removal of misfitting items was carried out, where the ability estimates of the cases (or respondents) were then anchored to the entire sample of respondents through the Rasch analysis approach.

4.3.4 Confirmatory Factor Analysis (CFA) for Testing Construct Validity

In this study, the construct variables involved epistemological beliefs, argumentation, scientific reasoning, and conceptual understanding of physics. These latent variables were defined by a number of observed or manifest indicators which can be tested with a statistical analysis technique called factor analysis. "Latent variables formed from multiple variates have been found to have greater reliability and validity than variables singly observed" (Sellin & Keeves, 1997, as cited in Aldous, 2017, p. 1866).

Factor analysis of the model constructs can be either based on Exploratory Factor Analysis (EFA) or Confirmatory Factor Analysis (CFA). EFA is used by researchers to identify the relationship between the observed variables and the underlying factors of a set of observed measures. Conducting EFA enables researchers to obtain information about how many constructs are required to best represent the data deriving from statistical results. Kline (2011) asserted that EFA does not require a specific hypothesis or prior knowledge about the possibility of a data structure or even the number of factors. Kline (2011, p. 116) further argued that "… EFA tests unrestricted factor models." In a nutshell, EFA is conducted without previously having knowledge or theory about how many factors really exist, or which observed variables load on specific factors. By using appropriate statistical software, EFA can be applied to allow the underlying pattern of the data to determine the factor structure and provide information about how many factors represent the data properly.

Meanwhile, CFA is used by researchers to verify the extent to which the observed variables represent a latent factor or construct which is not directly observed. Unlike the EFA that allows statistical techniques to determine the number of factors and which indicators (or observed variables) load on those factors, the CFA provides more information about how well the established theories of the factors correspond to the actual data. As described by Kline (2011, p. 112), the CFA is a useful statistical analytical technique to examine the structure of the factors, where the number of factors and their correspondence with observed variables is explicitly determined. In other words, the CFA technique is used to test a hypothetical factor model or the proposed factor structure model. In the CFA measurement model, the researchers are assumed to have some prior hypothesis or theoretical knowledge of the underlying factor structure of the scales based on previous research or established theories (Byrne, 2016). In the same vein, Schreiber et al. (2006) pointed out that CFA techniques are governed by the theoretical relationship between observed and latent variables that are empirically tested and confirmed by a set of actual data in a study. In turn, results from the CFA model enable researchers to determine whether to confirm or reject the previous studies or theories (Hair et al., 2014, p. 603). Thus, conducting the CFA measurement technique enables researchers to test the conceptual theories that explain how observed variables represent factors or constructs; they might also gain a better understanding of the quality of their measures. According to Pituch and Stevens (2016, p. 383), "CFA is part of a broader modelling framework known as Structural Equation Modelling (SEM), which allows for the estimation of more sophisticated models." A description of the SEM model is provided in section 4.3.5.

The CFA model is mostly used to examine the measurement model which is aimed to ensure the construct validity analysis of an instrument (Khine, 2013). Thompson (2004) further pointed out that the CFA measurement model is more useful than EFA for a number of reasons, such as the CFA method directly tests the theory and can assess the model fit in several ways. Since this study used questionnaires adopted from a number of existing instruments based on previous studies, the researcher had already hypothesised that the observed variables have correlations to their underlying latent constructs. Hence, the CFA model was the appropriate method used in this study to examine the construct validity of the factors.

According to Hair et al. (2014), it is essential that researchers consider certain issues related to the use of the measurement model such as unidimensionality, the congeneric measurement model, and the number of items per construct. Unidimensionality refers to a set of indicators assigned to only one particular factor, with no cross-loadings (loading on more than one single construct or scale) being assigned. In other words, each observed variable can be explained by only a single factor and no observed variable is determined by more than one factor or latent variable when the model is identified as unidimensional. As asserted by Hair et al. (2014), "the existence of significant cross-loadings is evidence of a lack of construct validity." Additionally, the measurement model can be considered congeneric if the model comprises of several unidimensional constructs with all cross-loadings fixed at zero. Another issue that needs to be considered regarding the use of the measurement model is the number of items in a factor or construct. Practically, more indicators (or observed variables) per construct are not necessarily better. Hair et al. (2014) advocated a minimum of three items per factor to assess construct validity. This is to ensure that these items sufficiently represent the latent constructs.

For the purpose of this study, the IBM SPSS AMOS (Analysis of Moment Structures) 25 software was employed to examine the construct validity of the factors (i.e., in the measurement model), as well as to explore the multiple relationships between unobserved variables or latent variables (i.e., in the structural model). The IBM SPSS AMOS's default method of parameter estimation was Maximum Likelihood (ML), which is the most widely used method for estimating the SEM model (Allison, 2003). The SPSS AMOS program provides a graphical interface instead of syntax commands which can assist researchers to more easily operate the software to determine or modify a model through an interactive path diagram.

As documented by Hair et al. (2014), there are five elements involved when the CFA method is applied, namely the latent constructs (or unobserved variables), the observed variables, the item loadings on specific constructs, the relationships among the constructs, and the error terms for each observed variable. Graphically, latent constructs are represented by ellipses or circles, while observed variables (sometimes called manifest variables) are shown by rectangles or squares. The item loadings on specific constructs are indicated by a single-headed arrow from the construct to the observed variable. This represents the relationships between the latent constructs and the respective observed items (i.e., factor loadings). It is essential to note that one observed variable of each construct must be assigned a value of 1 (i.e., on its factor loading), and the value of the variance of the construct must be fixed to 1 in the testing of a measurement model. The IBM SPSS AMOS program can automatically assign these values by default. Schreiber et al. (2006) pointed out that researchers can select values other than 1 that will not affect the overall fit of the model but might change the variance of the error. In addition, a correlational relationship between constructs is represented by a two-headed curved arrow, in which all constructs are considered as exogenous variables. The last element reveals that each observed variable has an error term (i.e., shown as an 'e' in the path diagram, see Figure 4. 1). The error terms refer to "the extent to which the latent factor does not explain the measured variable" (Hair et al., 2014, p. 604). In other words, measurement error indicates that the observed variable might measure something different from the latent factor. Each dependent variable in the path model is contributed by a circle or an oval around the error term that leads to the dependent variable (Schumacker & Lomax, 2016). Figure 4. 1 shows an example of a CFA model which consists of two constructs where each latent variable is measured by three observed variables.

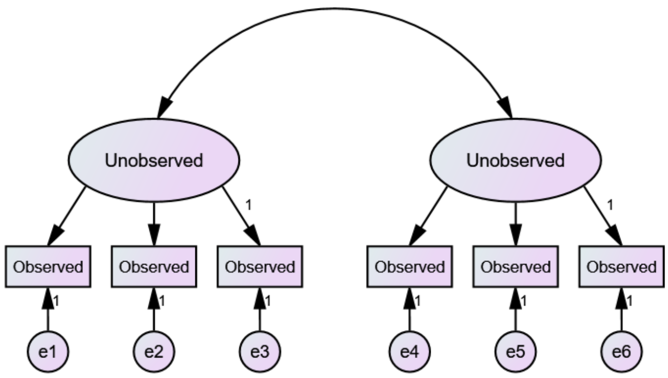


Figure 4. 1 An example of the CFA model (adapted from Schreiber et al., 2006)

Assessing the measurement model fit

As mentioned earlier, CFA is used to provide a confirmatory test of the measurement theory. The CFA measurement analysis results provide information regarding the hypothesised measurement model such as factor loadings and model fit indices. Examining the factor loadings of each observed variable in the models is considered important. As noted by Hair et al. (2010), the value of factor loadings in a model represents the correlation of each variable and the factor. Hair et al. (2014, p. 115) further suggested the rule of thumb in determining factor loading values as follows:

- Factor loadings in the range of ± 0.30 to ± 0.40 are considered to have met the minimal level for interpretation of structure.
- Loadings of ± 0.50 or greater are considered to be practically significant.
- Loadings exceeding 0.70 are considered to be indicative of well-defined structure.

In addition, it is also important to assess the statistical significance of the factor loadings for differing sample sizes. Hair et al. (2014, p. 115) proposed a guideline for identifying statistical significance of factor loading values according to sample size, as presented in Table 4. 2. Based on the guidelines provided in Table 4. 2, a minimum factor loading value of ± 0.30 is used as the cut-off value to exhibit statistical significance, because the sample size in this study was more than 350 respondents (i.e., 706 respondents). Thus, observed variables with loadings of ± 0.30 or greater were considered acceptable as the minimum value for the measurement model to be interpretable. As highlighted by Hair et al. (2014), the minimum standard for the factor loading value is 0.30 to indicate a good fit for the proposed model.

Table 4. 2 Guidelines for identifying significant factor loadings based on sample size

Factor Loadings	Sample Size Needed for Significance
0.30	350
0.35	250
0.40	200
0.45	150
0.50	120
0.55	100
0.60	85
0.65	70
0.70	60
0.75	50

Apart from factor loading values, it is also essential to examine the model fit indices of each construct. By and large, researchers use a variety of fit indices to identify the best model that fits the data well (Schreiber et al., 2006). As asserted by Kline (2011), each fit index has its own limitations for a number of reasons. Hence, it is strongly advocated that researchers use multiple indices in deciding whether a good fit exists, which appears to work well in accordance with the purpose of the study. The examination of the proposed CFA model in this study was assessed using a number of fit indices, including ratio of chi-square to its degrees of freedom (χ^2/DF ratio), the goodness-of-fit index (GFI), the adjusted goodness-of-fit index (AGFI), the Tucker-Lewis index (TLI), the comparative fit index (CFI), and the root mean square error of approximation (RMSEA).

In terms of the χ^2/DF ratio value or chi-square divided by the number of degrees of freedom, Kline (2011) pointed out that there is no specific threshold with regard to the model to be accepted; however, a smaller value of the χ^2/DF ratio is more expected. In this study, a value of χ^2/DF ratio of less than 5 would indicate a good fit, or is demonstrated as an acceptable fit in the large sample size of the study (Darmawan, 2003). Another fit index is GFI and AGFI. According to Khine (2013), GFI and AGFI indicate the relevant amount of the observed variances and covariances that is explained by the hypothesised model. Differing from GFI, AGFI adjusts to the number of degrees of freedom in the specified model. The GFI and AGFI values range between 0 and 1, where higher values indicate a better fit. In addition, Schermelleh-Engel, Moosbrugger, and Müller (2003) summarised that the cut-off value for the GFI index is 0.95, which indicates a good fit; while values above 0.90 are considered as an acceptable fit. Meanwhile, threshold values for an AGFI index of 0.90 indicate a good fit, while a value greater than 0.85 is considered as an acceptable fit. Schermelleh-Engel et al. (2003) also noted that both the GFI and AGFI indices are influenced by sample size. More specifically for smaller sample sizes, these indices may decrease with increasing model complexity. For the purpose of this study, a model with GFI and AGFI values of close to, or on, 0.90 were considered an acceptable fit. Turning to the TLI index, Khine (2013, p. 15) stated that the TLI index is employed "… to compare a proposed model to the null model." As noted by Hair et al. (2014), TLI indices can range below 0 or above 1, in which a value close to 1 is considered a good fit. Generally, the higher the value of the TLI index, the better the model fits the data. Meanwhile, the value of the CFI indices falls between 0 and 1. According to Kline (2011, p. 208), the CFI index is "an incremental fit index that measures the relative improvement in the fit of the researcher's model over that of a baseline model, which typically is the independence model." As documented by Hair et al (2014), CFI values greater than 0.90 indicate a good model fit. In the case of this study, a model with the TLI and CFI indexes of close to, or on, 0.90 was considered as an acceptable fit. According to Byrne (2016), the RMSEA is considered to be one of the most informative fit indexes which represents the error of approximation in the population. This index is relatively independent of sample size, even though it is affected by the complexity of the model (i.e., degrees of freedom) (Schumacker & Lomax, 2016). In terms of values on the RMSEA, Byrne (2016) further stated that a RMSEA index of less than 0.05 is considered as a good fit, values between 0.05 and 0.08 are considered as a reasonable fit, values between 0.08 and 0.10 are considered as a mediocre fit, and values over 0.10 are considered as a poor fit. In this study, a model with a RMSEA of less than 0.08, or not more than 0.10, was considered as an acceptable fit. The summaries of all the fit indices to indicate good model fit used in this study are presented in Table 4. 3.

Table 4. 3 Summaries of fit indices to indicate good model fit

Model Fit Indices	Values to Indicate Good Fit
χ^2/DF ratio	Close to or < 5
Goodness-of-Fit Index (GFI)	Close to or ≥ 0.90
Adjusted Goodness-of-Fit Index (AGFI)	Close to or ≥ 0.90
Tucker-Lewis Index (TLI)	Close to or ≥ 0.90
Comparative Fit Index (CFI)	Close to or ≥ 0.90
Root Mean Square Error of Approximation (RMSEA)	≤ 0.08 or not more than 0.10 is still acceptable

4.3.5 Structural Equation Modelling (SEM) for Testing the Effect

The Structural Equation Modelling (SEM) model can be categorised into two sub-models, which are the measurement (outer) and structural (inner) models. According to Khine (2013), the measurement model as part of the SEM model is defined as the relationship between observed and unobserved variables. Meanwhile, the structural model specifies the complex and multiple relationships between the unobserved variables (i.e., both independent and dependent variables). The terms unobserved variables, traits, or constructs can be used interchangeably with 'latent factors' or 'latent variables' (Ben, 2010). It is also important to note that the terms factor and construct are used interchangeably in this study. In addition, other terms related to SEM models are exogenous and endogenous variables. Exogenous represents the independent variables, while endogenous represents the dependent or outcome variables. As noted by Schreiber (2006, p. 325), "exogenous and endogenous variables can be observed or unobserved, depending on the model being tested."

In the quantitative analysis, multiple factors including a number of dependent and independent variables are examined, where Structural Equation Modelling (SEM) was used as one of the statistical analysis approaches to address the research question. The SEM analysis is useful to test the extent to which the empirical data support the theoretical model proposed in the study. If the theoretical model is supported by the real data, the hypothesised model can be accepted, and the more complex theoretical models can be examined further. Meanwhile, if the theoretical model is not supported by the empirical data of the study, the hypothesised model needs to be modified or rejected, so a new theoretical model would need to be developed and examined to understand the unique relationships among the variables that are of interest to the researcher (Schumacker & Lomax, 2016). In testing the SEM models, there are five sequential steps, namely model specification, model identification, model estimation, model testing, and model modification (Kline, 2016; Schumacker & Lomax, 2016).

In model specification, the researcher specifies a hypothesis for every relationship among the variables and parameters in the model, referring to the existing theories, prior research findings, and relevant knowledge in order to test a theoretical model (Schumacker & Lomax, 2010). Providing a rationale and purpose for specifying a structural model requires plausible explanations relating to complex phenomena tested prior to any data collection or analysis. Hence, it is important to be clear about the variables and the kind of relationships in the particular model that are of interest to the researcher. In model identification, to test a hypothesised model, the researcher examines whether the model can be estimated using the empirical data of the study. There are three categories of model identification, namely 'under-identified' (unidentified), 'just-identified', and 'over-identified' (Khine, 2013; Kline, 2016). The model is classified as 'under-identified' if the degrees of freedom for the model are negative. This would probably be due to a measurement model having many parameters or insufficient observed variances to be estimated, so it would be impossible to estimate the parameters. A particular model is 'just-identified' if the degrees of freedom equal 0 due to the model having the same number of variances or covariances and parameters. Meanwhile, the model is 'over-identified' when the degrees of freedom have a positive value (or greater than zero) in which a measurement model has fewer parameters than the observed variances and covariances. In addition, Schumacker and Lomax (2010) argued that the parameter estimates can be trusted if the models are classified as 'just-identified' or 'over-identified'. In model estimation, the researchers need to determine an appropriate estimation technique to estimate the parameters in analysing SEM models. "Model estimation involves determining the value of the unknown parameters and the error associated with the estimated value"(Khine, 2013, p. 12). Maximum Likelihood Estimation (MLE) is a common estimation method employed in SEM analysis (Khine, 2013). Schumacker and Lomax (2016, pp. 244-245) further outlined that MLE techniques must "… meet the multivariate normality assumption (acceptable skewness and kurtosis); there are no missing data; no outliers; and continuous variable data." In this study, AMOS 25 was used to estimate the parameters using MLE as a default method. In model testing, the researchers might test to what extent the empirical data fits the theoretical model. The fit is determined by identifying a number of model-fit indices for the statistical significance of each parameter in the model (Schumacker & Lomax, 2016). In the case of the study, the hypothesised model was examined using the critical ratio (C.R.) and the p-value to test how well the model was supported by the real data. A critical ratio exceeding 1.96 is considered to be significant at p-value ≤ 0.05, in which the p-value is used to test the statistical significance of the relationships between the research variables. A critical ratio of below 1.96 or a p-value of more than 0.05, are considered to be not significant. Therefore, the relationships between the latent variables which have no significant paths are then deleted in the model. Khine (2013) highlighted that reporting the various model-fit indices is needed to consider how well the specified model fits the empirical data of the study. Finally, in model modification, the initial model may be modified to obtain the best-fitting model (Schumacker & Lomax, 2016). This can be done by adding or excluding parameters in the model which might improve the fit of the model. As noted by Khine (2013), the AMOS software program provides the modification indices (MI) for each parameter to improve the fit of the model generated by the SEM techniques. In the same vein, Schreiber et al. (2006) asserted that model modification is conducted on the proposed model in order to obtain a better fitting model or a more parsimonious model. However, the researcher needs to consider the results obtained from the modification procedure for the model. Extra caution must be exercised in interpreting the modified model, no matter how well the AMOS program helps the researcher to improve model fit. The results of model modification must be supported by relevant theories or existing theories to make theoretical sense, so that the best theoretical models in research can be obtained. Schreiber et al. (2006, p. 327) also noted that "a model that has been modified, a trimmed model, is termed a nested or hierarchal model."

As mentioned earlier, the structural model refers to the complex theoretical network of relationships (i.e., directly or indirectly) between unobserved variables (or latent constructs), including both independent and dependent variables. To investigate these relationships in the proposed study, the Structural Equation Modelling (SEM) technique was applied using the IBM SPSS AMOS software package (version 25). Maximum Likelihood (ML) was chosen as the IBM SPSS AMOS default method to compute parameter estimation. For the purpose of this study, SEM techniques were used to test the hypothesised model, referring to previous studies and the literature, which in turn enabled the researcher to establish new theoretical frameworks, with great caution. Hence, the theoretical underpinnings used to test the proposed model (i.e., involving both measurement and structural) were essential to consider while identifying the best fitting model.

As asserted by Byrne (2010, p. 3), the SEM method is essential in representing the nature of "statistical methodology that takes a confirmatory (i.e., hypothesis-testing) approach to the analysis of the structural theory bearing on some phenomenon." Technically, in structural modelling, exogenous variables represent latent variables which might influence other latent variables under consideration, but such exogenous variables are not influenced by other latent variables in the hypothesised model. The latent variables that are influenced by these exogenous variables are identified as endogenous variables which are also influenced by other endogenous variables in one overall model (Schreiber et al., 2006). The SEM model technique enables “one dependent variable to become an independent variable in subsequent dependence relationships” (Darmawan, 2003, p. 82).

According to Hair et al. (2014, p. 585), ”structural model specification focuses on adding single-headed, directional arrows to represent structural hypotheses in the researcher’s model.” In other words, the SEM model can be illustrated by the arrows pointing from one construct (or latent variable) to another construct variable. In the SEM model, the direct, indirect, and total effects between construct variables are determined by established theory or empirical assumptions (Schreiber et al., 2006). The presence of a direct effect is evident when an independent variable influences a dependent variable in which the relationships between such construct variables are linked by a single arrow. Meanwhile, an indirect effect refers to the effect of an independent variable on a dependent variable which is mediated by one or more construct variables. In addition, the total effect can be obtained through the summation of the direct and indirect effect coefficients (Hair et al., 2010; Schreiber et al., 2006). The construct variables in a structural model are ordered from left to right and are connected by arrows showing the direction of the influences among these construct variables. In the present study, the models were depicted by multiple single-headed arrows, indicating both multiple direct and indirect effects among the construct variables. An example of an overall model or path diagram is provided in Figure 4. 2. It not only shows the complete set of constructs and observed variables in the measurement model, but also the structural relationships among the construct variables.

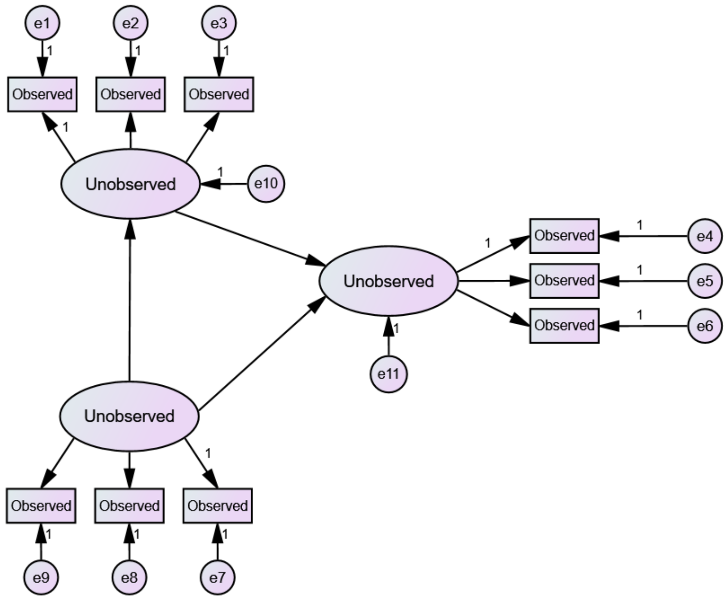


Figure 4. 2 An example of the SEM model (adapted from Khine, 2013)

Assessing the Model Fit for SEM

Similar to the CFA analysis, the SEM model was assessed by using a number of fit indices to examine whether it fitted well with the data. The fit indices used in the analysis included the ratio of chi-square to its degrees of freedom (χ^2/DF ratio), the goodness-of-fit index (GFI), the adjusted goodness-of-fit index (AGFI), the Tucker-Lewis index (TLI), the comparative fit index (CFI), and the root mean square error of approximation (RMSEA). In this study, the acceptable cut-off values for fit indices to examine model fit is provided in Table 4. 3. In addition, a number of estimates was also generated and examined, which include the unstandardised parameter estimates (B), the standard errors (S.E.), the critical ratio (C.R.), the p-value, the standardised estimates (β), the standardised indirect effect (ie), and the total effect (te).

Unstandardised estimates (B) represent the strength of the relationship between observed and unobserved variables. These parameter estimates are employed "… to judge statistical significance of parameters along with standard errors" (Khine, 2013, p. 29). Meanwhile, the standard errors (SE) represent the variability of the estimates. Khine (2013) further stated that the critical ratio (C.R.) can be obtained by dividing the unstandardised parameter estimates by the standard error (B/SE) to determine statistical significance. A critical ratio exceeding 1.96 is considered significant at p-value ≤ 0.05, in which the p-value is used to test the statistical significance of the relationships among the variables. Next, standardised estimates (β) allow a direct comparison of the relationship among the variables to be carried out (Hair et al., 2014). In addition, "standardized coefficients are model parameter estimates based on the analysis of standardized data, in the sense that all variables are supposed to have unit variance" (Kwan & Chan, 2011, p. 730). These coefficients may be used to identify which variables have a greater effect on the dependent variables in the hypothesised model. To interpret the results of the path analysis, Cohen suggested that the effect sizes of the path coefficients (i.e., relationships between latent variables and other latent variables) should be categorised as small (0.02), medium (0.15), and large (0.35) (Kock, 2014). Effect sizes below 0.02 indicate effects that are too weak to be considered relevant for interpretation. However, as recommended by Sellin and Keeves (1977 as cited in Aldous, 2014), the standardised estimate values should be greater than 0.10. Thus, path coefficient values below 0.10 are removed from the model because these values show inadequate effects in estimating the relationship between latent variables.

Since there is interest in understanding more about mediating effects and direct effects in the SEM model, this study has also examined the indirect effects (ie) and total effects (te) of the final model. The standardised indirect effects (ie) represent the relationship between an independent variable and a dependent variable that is mediated by one or more latent variables in the model (Khine, 2013). The size of an indirect effect can be calculated by multiplying the path coefficients of the entire association between the variables involved, in which these variables are ordered from left to right in a model (Cramer, 2003). The total of indirect effects is provided by AMOS output. As stated by Kline (2011, p. 166), "total effects are the sum of all direct and indirect effects of one variable on another." In other words, the standardised total effect (te) refers to the combined direct and indirect effects of a latent variable on a dependent latent variable. Lack of information about this relationship might lead to biased research results and inaccurate conclusions in the study.

In addition, after the proposed model was built using the AMOS 25 software, model trimming was carried out as part of the path analysis process in order to obtain the best final model. As noted by Kline (2011, p. 214), trimming a path model is carried out to "find the model with the properly specified covariance structure that fits the data and is theoretically justifiable." In this process, the manifest or latent variables which showed non-significant paths were then simplified by removing pathways from the proposed model. For the purpose of this study, in determining whether or not the paths are significant, model trimming was carried out by examining the critical ratio for significance, as highlighted by Darmawan (2003). As noted earlier, any critical ratio that exceeds 1.96 is considered significant at p-value ≤ 0.05. In other words, any path with a critical ratio value of less than 1.96, and p-values of more than 0.05, were considered not significant and were removed from the model. The process of trimming the model was carried out by eliminating non-significant paths one at a time. Once the non-significant paths were removed, the process of modelling was repeated, and the results re-examined until all the remaining paths showed statistical significance.

4.4 Qualitative Data Analysis

As presented earlier in Chapter 3, the aim of conducting the qualitative study was to complement or provide an in-depth understanding of the findings gained from the statistical analysis; as a consequence, not all variables were investigated in the qualitative study. In other words, this qualitative study was conducted to gain a deeper understanding of specific results obtained from the quantitative study. In the qualitative stage, data were gathered through a qualitative inquiry using individual in-depth interviews involving 25 participants from the four universities (i.e., both private and public) engaged in this study. The qualitative study was carried out after all the participants involved had completed the overall instrument tests, and raw scores were obtained in the first phase of the study (i.e., the quantitative stage). The participants of the qualitative study were then drawn from the pool of pre-service physics teachers who participated in the quantitative data collection phase of the study. Furthermore, the purpose of collecting the qualitative data was to probe the participants' perceptions of existing physics teaching and learning practices (i.e., opportunities and barriers) with regard to the skill to think scientifically (i.e., epistemological beliefs, argumentation, and scientific reasoning) and their understanding of physics concepts during their higher education studies. The participants' perceptions of the relationship between scientific thinking (i.e., epistemological beliefs, argumentation, and scientific reasoning) and their conceptual understanding of physics were also explored.

Prior to the data analysis, all the interviews were audio-recorded and then transcribed verbatim. The interview transcripts were imported into the NVivo (version 11) software program to organise the raw qualitative data into more manageable form. In analysing and interpreting qualitative data, Creswell (2012, p. 236) proposed six steps, including "preparing and organizing the data, exploring and coding the database, describing findings and forming themes, representing and reporting findings, interpreting the meaning of the findings, and validating the accuracy of the findings." Furthermore, Creswell (2012) suggested the process of data analysis in a qualitative study as visualised in Figure 4. 3.

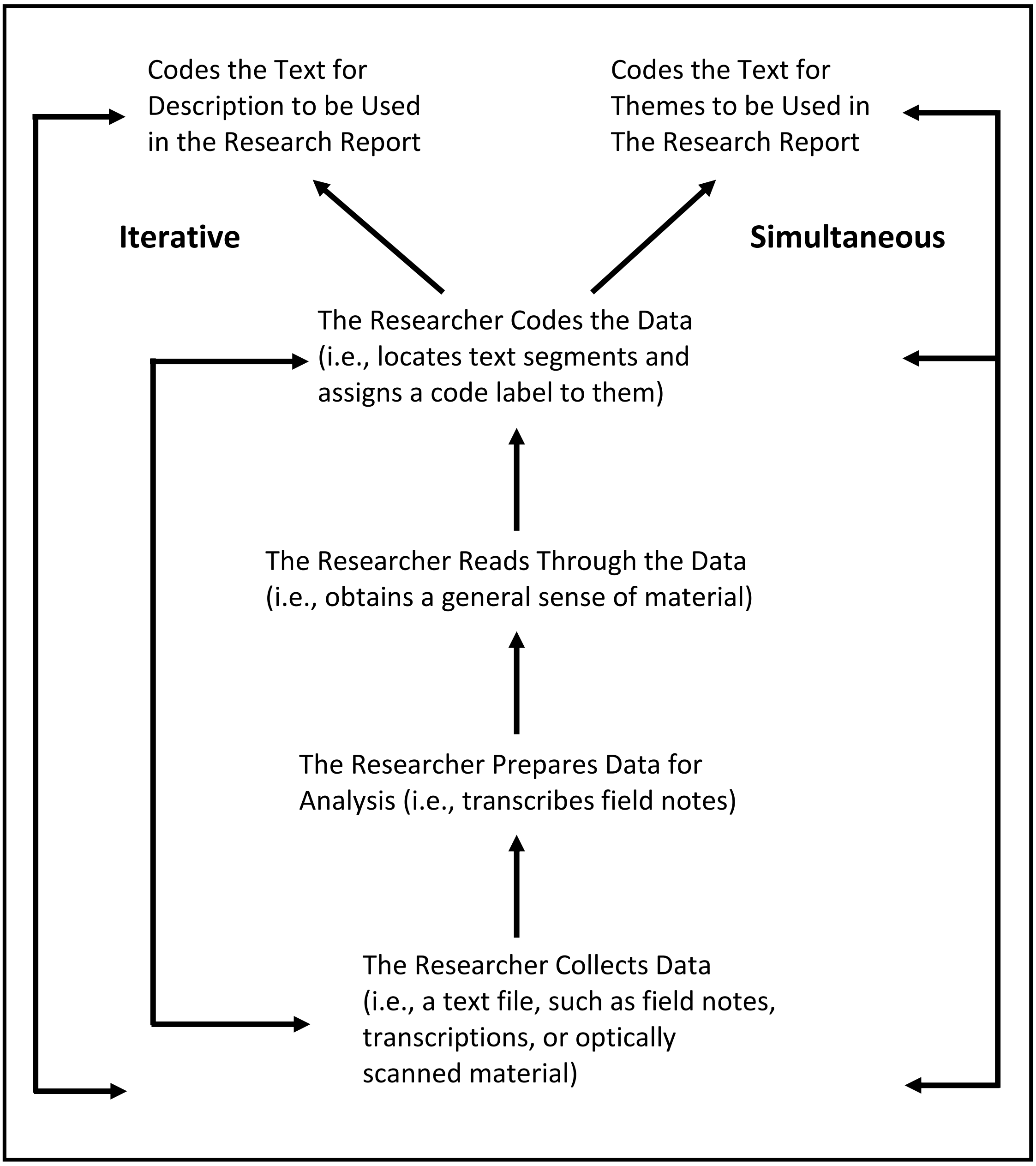


Figure 4. 3 Qualitative data analysis process (Creswell, 2012, p. 237)

Referring to Figure 4. 3 as proposed by Creswell (2012), researchers generally collect data in the first step and then prepare it to carry out the data analysis, which consists of obtaining a general sense of the data as well as generating codes and themes in relation to the context of the study. In qualitative research, data collection and analysis can include both simultaneous and interactive activities that may differ from traditional approaches in quantitative research. Furthermore, Creswell (2012) stated that there is no single approach that is fully acceptable in analysing qualitative data; however, some guidelines might be useful to apply in the process of analysing qualitative data. For the purpose of this study, some of the steps for analysing the qualitative data are explained in the following sections.

Data Preparation

The individual semi-structured interviews in this study were recorded using a digital audio recorder. If there were responses that were considered unclear during the interviews, the researcher would immediately follow up by repeating the questions and asking for more detailed information from the interviewee. All of the interview data gathered were then organised and exported to different file folders or computer files, which were categorised based on gender, year level, and the name of the university for ease of data organisation. The verbatim transcripts of the interviews were in Bahasa Indonesia, i.e., the language used in the interviews. As the researcher is from Indonesia, and is very familiar with the Indonesian language, it was considered more convenient to analyse the interview data in Indonesian to avoid misinterpretation of the responses of the interviewees. To carry out the data analysis, all the interview transcripts were kept in Bahasa Indonesia and exported to the NVivo software program (QSR International Pty Ltd, 2012). This software enables the researcher to store and organise interview data, to easily identify the source and references related to the codes, as well as to search for certain texts or words.

Coding

Code development was the next step in the data analysis. Initially, the researcher carefully read each of the interview transcripts several times to acquire a general sense of the material. After familiarisation with the interview data, the researcher generated preliminary codes for several ideas related to the research context (e.g., variables in the research model). As more information about the topic was discovered, the researcher was able to add a number of codes or refine or re-code the preliminary codes when needed. Furthermore, if very similar codes were found, they were then merged as part of the process of eliminating redundancies. As noted by Creswell (2012, p. 243), coding is the step in data analysis that aims "… to make sense out of text data, divide it into text or image segments, label the segments with codes, examine codes for overlap and redundancy, and collapse these codes into broad themes." Through this process, the researcher was able to select specific data used in the study and disregard other data that did not specifically provide evidence for generating themes. In the case of this study, the codes were generated based on the guidelines of the interview protocol that explored specific topics intended to complement and gain a deeper understanding of the quantitative findings. A priori concepts can also be used as codes referring to the existing literature. The codes were then generated under these concepts from the interview data (Lacey & Luff, 2001).

Theme Generation

After generating a number of codes, the next step was to combine related codes to develop common themes from the data that reflected the questions provided in the interview guide. As stated by Creswell (2012), themes can be built from a collection of codes that have similar responses to form a major idea. These themes should relate to the theoretical concepts associated with the study. Checking and revising the initial themes was needed to ensure that the codes matched the data set entirely. Creswell (2012, p. 251) also asserted that the researcher may "reach a point where themes are fully developed, and new evidence will not provide additional themes." At this stage, the researcher has thus obtained a number of themes which can then be reported as part of the research findings or evidence, to enrich and provide deeper understanding of the quantitative findings collected through questionnaires. In addition, quotes from the interviews used in this study may provide narrative evidence for each theme. Although this qualitative study was designed to complement the quantitative findings, it also enables a deeper understanding to be gained of research topics beyond the hypothesised variables and models. Doing so, allows the researcher to obtain a more comprehensive understanding of the phenomena under study.

In the case of this study, four samples of non-English transcriptions and a number of quotations used in this thesis were translated from Bahasa Indonesia to English by a bilingual professional translator who was also very fluent in Indonesian. The translation results were then checked by the researcher as well as an Indonesian speaking colleague at Flinders University (i.e., a PhD student) to ensure that the translated interview results had the same meaning as the original versions. As noted by Birbili (2000), researchers need to be aware that translation from one language to another should be carried out with great caution. Considering this process has a direct impact on the validity of the research and its reporting which involves matters of conceptual equivalence or comparability of meaning. To ensure a trustworthy qualitative component, four transcript samples were checked by the research supervisors. Furthermore, parts of the transcripts relating to the themes or quotes from the interviews used in the thesis were also examined by the supervisors to ensure that the quotations used supported the themes specified in the study.

4.5 Summary

This chapter has highlighted the methods of data analysis used in this thesis for both the quantitative and qualitative study to address the research questions. Most specifically for the quantitative study, the general methodological considerations such as missing values, normality, and multicollinearity are described, as they influence the selection of appropriate analysis techniques.

Since the study used a number of questionnaires, reliability and validity of the instruments were established to produce useful and meaningful inferences through Rasch analysis and Confirmatory Factor Analysis (CFA) measurement models. In the Rasch model analysis, the Partial Credit Model (PCM) was selected to examine a person’s ability and item difficulty, which was expressed on a scale of log-odd ratios, or logits. A logit value of zero was set as the mean; hence, items of above average difficulty were plotted as positive, while items of below average difficulty were plotted as negative. In addition, fit indices such as the infit MNSQ (weighted fit mean square) and item discrimination index used to identify whether the Rasch model fits the data well were also defined. Meanwhile, the CFA model was conducted to examine the construct validity analysis of the research instruments. Subsequently, the Structural Equation Modelling (SEM) technique was carried out to examine the relationships among the research variables or scales. In analysing complex data sets, a number of statistical procedures were carried out in this study that used software packages such as ACER ConQuest 4 for the Rasch analysis and IBM SPSS AMOS 25 for the CFA and SEM analyses.

Finally, qualitative data analysis techniques were used to complement and enrich the information obtained from the quantitative findings. The qualitative analysis begins with transcribing the interviews, reading the interview transcripts, and coding them by referring to the transcribed interview data. This was followed by the generation of themes from a collection of codes with similar responses to form a major idea related to the theoretical concepts associated with the study. The results of the quantitative and qualitative analyses are presented in the following chapter.

CHAPTER 5
PRELIMINARY ANALYSIS AND SCALE VALIDATION

5.1 Overview

This study focuses on investigating scientific thinking (specifically epistemological beliefs, argumentation, and scientific reasoning) and the conceptual understanding of physics of pre-service teachers in Yogyakarta, Indonesia. The previous chapter described the quantitative data analysis techniques employed in this study. This chapter reports on the results of the preliminary analysis and validation of the scales carried out in accordance with the procedures for analysing the quantitative data, as described in Chapter 4. This chapter begins with demographic information on the 706 Indonesian pre-service physics teachers involved in this study, consisting of descriptive data about gender, year level, and type of university attended. As proposed in the previous chapter, it is necessary to transform raw scores into an interval scale in the existing data sets by employing ConQuest 4 software. Furthermore, through a multidimensional Rasch model analysis, the person fit and item fit mean square, item discrimination, t-value, and an item–person map were generated. The reliability of the instruments was examined by employing a Cronbach alpha index and item separation reliability analysis. In addition, the analysis of the scale validation of the questionnaires using Confirmatory Factor Analysis (CFA) is presented. This is followed by an analysis of the descriptive and inferential statistics, including the mean, standard deviation, testing the normality and multicollinearity of the data, t-tests, as well as a one-way Analysis of Variance (ANOVA). The results of the analysis reported in this chapter are crucial for specifying the structural model using the Structural Equation Modelling (SEM) procedures described in Chapter 6.

5.2 Demographic Information of the Participants

A general picture of the data sets in the present study, including demographic information about the participants such as gender, year level, and type of university attended, was generated using the IBM SPSS 25 software program. The study involved 706 pre-service physics teachers from four universities located in Yogyakarta, Indonesia, also known as teacher preparation programs at higher education institutions. As highlighted in Chapter 4, the researcher checked the completeness of the data obtained from the participants by detecting whether there were any missing data. All the participants completed all items on each of the research instruments employed in this study, which means that there were no missing values. All participants came from one of four universities, which comprised two public universities and two private universities. Figure 5. 1 shows that 44.1% of respondents (311 pre-service physics teachers) were from private universities, while 55.9% (395 pre-service physics teachers) were from public universities. The diagram below illustrates that the number of participants from public universities involved in this study was higher than the number of participants from the private universities.

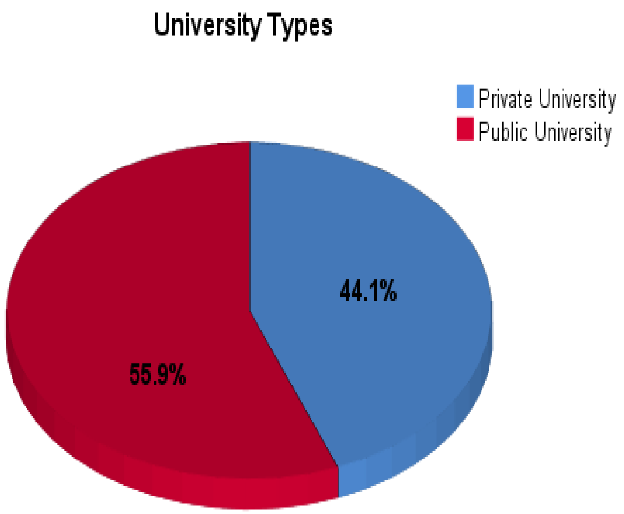


Figure 5. 1 Distribution of pre-service physics teachers by university type

Of the 706 participants involved in this study, 206 (29.2%) pre-service physics teachers were in Year 1, 178 (25.2%) in Year 2, 180 (25.5%) in Year 3, and 142 (20.1%) were in their fourth year of study. The number of participants in Year 1 was larger than those in the other year levels, as presented in Table 5. 1. The distribution of pre-service physics teachers by year level for each university type is illustrated in Figure 5. 2.

Table 5. 1 Distribution of pre-service physics teachers by year level group

Year Level	Frequency	Percent
Year 1	206	29.2
Year 2	178	25.2
Year 3	180	25.5
Year 4	142	20.1
Total	706	100.0

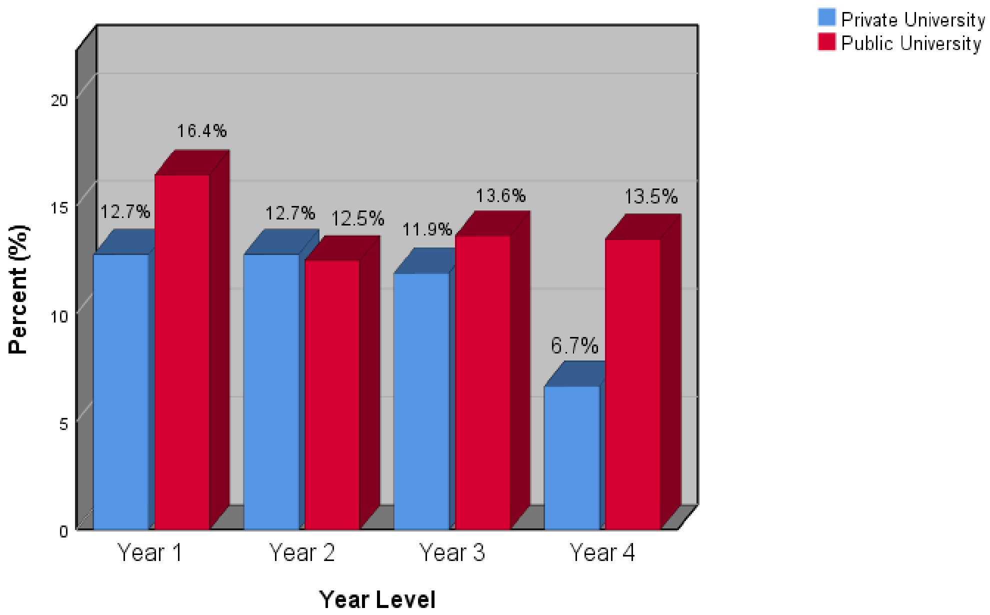


Figure 5. 2 Distribution of pre-service physics teachers by year level groups for each university type

Of the total number of participants who completed the questionnaires in this study, 541 were female (76.6%) and 165 were male (23.4%). The distribution of pre-service physics teachers by gender is shown graphically in Figure 5. 3. As the figure indicates, the number of male participants was lower than the number of female participants in each academic year. Referring to this trend, it is possible to infer that the number of female physics teachers will be considerably larger than the number of male physics teachers in the next few years. In fact, this is consistent with the trend of women's participation in the Indonesian education sector (MOEC, 2016).

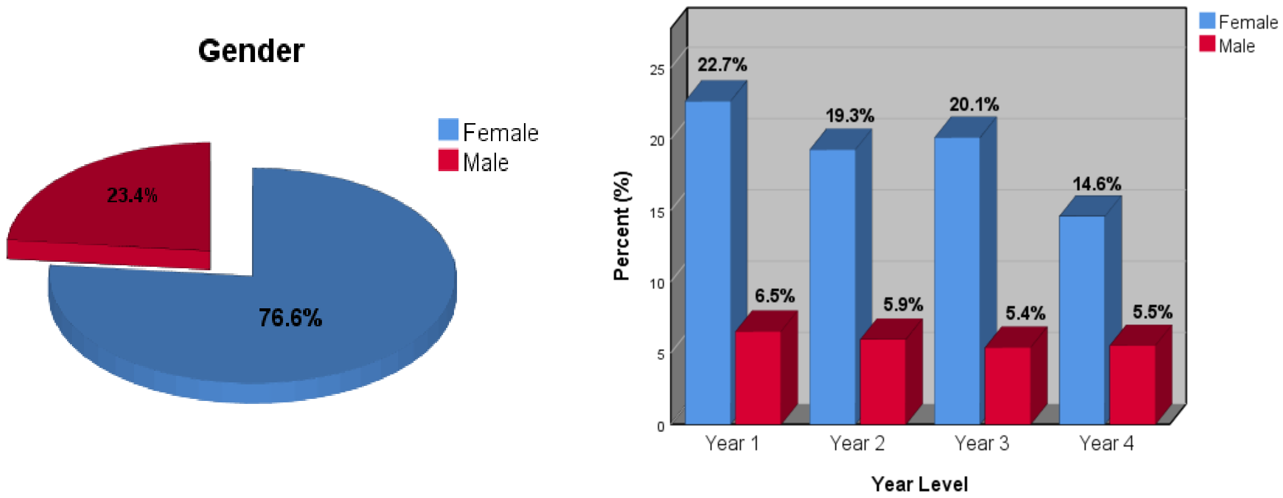


Figure 5. 3 Distribution of pre-service physics teachers by gender group

5.3 Rasch Model Analysis

Prior to carrying out further analysis, it is essential to establish the validity and reliability of the research instruments employed in order to provide accurate estimates of the structure of the scales. In the case of this study, the validity and reliability of the scales were examined using multidimensional Rasch model analysis and Confirmatory Factor Analysis (CFA). These two analyses were carried out to investigate the componential structure of the instruments and to provide information about whether or not the models reported in previous studies or theories fit the real data.

Fit statistics in Rasch model analysis were used to identify misfitting items and the pattern of responses for each person to assess how well an item or a person fitted the measurement model in this study. In addition, all raw scores obtained from the participants' responses to each questionnaire were transformed into an interval scale using the anchoring approach to generate Weighted Likelihood Estimate (WLE) scores. The WLE scores were then used in subsequent analyses such as Confirmatory Factor Analysis (CFA), descriptive and inferential statistical analysis, as well as in the Structural Equation Modelling (SEM) approach. In the multidimensional Rasch model analysis, the data were analysed using the ACER ConQuest 4 statistical software package in which the interpretation of fit statistics were reported as log odd units (logits). For the purpose of this study, ConQuest 4 software was employed to generate the mean square fit statistic (INFIT MNSQ), the t-value, the item discrimination index, the Wright map (or the item–person map), the Cronbach alpha, and the item separation reliability, which are provided in the following section.

5.3.1 Person and Item Fit Analysis

As stated previously, Rasch model analysis was conducted to transform raw scores to an interval scale and to validate the research instruments. In this study, examining the person and item fit mean square was carried out to check how well a person or an item fitted the Rasch model. The person fit analysis was conducted to detect any misfitting persons before examining for any misfitting items through the multidimensional Rasch model analysis. As highlighted in Chapter 4, persons with weighted fit mean square (INFIT MNSQ) values above 2.00 were excluded in the subsequent item analysis, as this value indicated degraded psychometric properties. Table 5. 2 presents the number of misfitting persons detected for each latent variable.

Table 5. 2 Misfitting persons for each latent variable

Latent Variable	Number of Misfitting Persons
Epistemological Beliefs	8
Argumentation	0
Scientific Reasoning	0
Physics Conceptual Understanding	0

Note: It is worth noting that the latent variables or scales described in the CFA and Rasch model have been 'bolded' to distinguish them from concepts related to these latent variables that are being explained.

As can be seen in Table 5. 2, the number of misfitting persons for the **Epistemological Beliefs** scale was 8 persons. Meanwhile, there were no misfitting persons for the **Argumentation**, **Scientific Reasoning**, and **Physics Conceptual Understanding** scales. The misfitting persons identified were then temporarily removed from the data sets to identify any misfitting items in the next step. Removing the misfitting persons, while conducting item fit analysis, is intended to ensure that items identified as misfitting, are truly problematic items. Once the misfitting items were excluded from the data sets, the misfitting persons were incorporated back into the data sets to generate the WLE scores using the anchoring approach.

As outlined in Chapter 4, the infit (weighted) statistics are more of a concern for this study than the outfit (unweighted) statistics. Bond and Fox (2015) stated that for the infit statistics, the performance of persons located closer to the item's difficulty value is given more weight, which should likely provide more information regarding the item's performance. Meanwhile, the outfit statistics are an unweighted estimate that is more sensitive to the performance of persons who are some distance from the location of the item and who may be influenced by unexpected responses from test takers (or examinees) on items that are either relatively very difficult or very easy for them. Hence, since the outfit mean square (or outlier-sensitive fit statistic) was not a concern in this study, only infit statistics are reported here. Items with a weighted fit mean square Standardized Weighted Mean Square (INFIT MNSQ) value falling outside the range 0.70 - 1.30 were considered as not fitting the model and were then removed with caution. In addition, the ConQuest software generates t-values or a standardised fit statistic (Z values) that may be positive or negative. The Z or t-values show 'how likely is the misfit' (Bond & Fox, 2015, p. 67). Further, a t-value greater than +2 or less than −2 was considered as not fitting the model at the 95% confidence level ($p < 0.05$) (Bond & Fox, 2015). However, as the t-value is influenced by the number of samples, it is worth noting that this value was not a great concern in this study. As highlighted by Wu and Adams (2007), the t-value commonly used to indicate statistical significance is sensitive to sample size. The higher the number of samples, the higher the t-value obtained (Wilson, 2005). Wu and Adams (2007) also asserted that fit statistics should serve as an indicator for detecting problematic items, and not to solely stick to the rules of the threshold or cut-off values when deciding whether a misfitting item should be retained or deleted from the data sets. Therefore, the t-value is presented only as additional information.

In addition, it is also important to examine the item discrimination index when assessing item performance for the model. As noted by Tiruneh et al. (2017), the item discrimination index describes how well an item discriminates between examinees (e.g., students or test takers) with different levels of ability, having either high ability or low ability. The item discrimination index also indicates the correlation between the item score and the total score on a questionnaire in which the index ranges from zero to one, that may be positive or negative. An item that has a discrimination value of zero indicates that there is no relationship between the item score and the total score, while an item that has a positive discrimination value indicates that there is a positive relationship (Wu & Adams, 2007). Furthermore, the items with negative discrimination values are considered undesirable. The items indicate that respondents with high ability tend to get these items wrong (incorrect), while respondents with low ability get them right (correct). In the case of this study, any items which had a discrimination index less than 0.15 were dropped from the data sets with extra caution. These items were interpreted as poor items and should be removed (Khan, Ishrat, & Khan, 2017; Rush et al., 2016).

To assess item performance in this study, test reliability was examined using Cronbach's alpha and item separation reliability, which was generated by the ConQuest software program. As noted in Chapter 4, Cronbach's alpha is widely used to represent internal consistency reliability for studies, and it ranges from 0 to 1. The higher the value of the Cronbach's alpha coefficient, the better the reliability of a scale. A Cronbach's alpha coefficient of around 0.70 or greater is considered desirable and indicates higher reliability (Taber, 2017). However, van Griethuijsen et al. (2015) mentioned that a Cronbach's alpha coefficient of 0.70 or 0.60 is an acceptable reliability value. In this study, a Cronbach's alpha coefficient of 0.60 was considered as an acceptable reliability value. Furthermore, the item separation reliability was also used to identify the Rasch model fit of the data. As highlighted by Bond and Fox (2015, p. 49), the item separation reliability points out "the replicability of item placements along the pathway if these same items were given to another same-sized sample of persons who behaved in the same way.” The higher the values of the item separation reliability, the greater the reliability that the placement of items in other sample groups will be replicated. The lower the item separation reliability index, the larger the measurement error of item parameter estimates in the model or imprecise estimates of the items (Ben, 2010; Bond & Fox, 2015).

The ConQuest 4 software was also employed to generate the item–person variable map or Wright map. This map provides a visual estimate of the relationship between person ability and item difficulty on a single scale and are located respectively to the left and to the right along a linear vertical scale on the map. The person ability and item difficulty estimates are displayed on a logit scale running down the middle of the map where the unit intervals between the locations on the Wright map have equal values or consistent meanings (Bond & Fox, 2015). Bond and Fox (2015) also asserted that the highest positive values are located at the top of the map for the most difficult items, while the lowest values (on a negative scale) are located at the bottom of the map, indicating the items that are most easily endorsed by the examinees. In other words, the higher the positive number scale, the higher the difficulty level of items and vice versa. Furthermore, “the mean of the item difficulties is adopted by default as the 0 point” (p. 79). It can be illustrated that items that are above average in difficulty, as well as persons above average in ability, are located on the positive end of the scale, while items below average in difficulty, as well as persons below average in ability, are located on the negative end of the scale. The item–person map is useful to assess if the test is well-targeted to the ability distribution of the examinees or test takers, and whether all regions of the person ability distribution are covered by the items. Ideally, the item difficulty distribution covers the distribution range of the participants' ability, thus providing an accurate measure of participants' ability across the scale (Liu et al., 2008).

Last but not least, the estimated correlation between the dimensions or sub-scales was also examined to investigate associations among the dimensions as well as to validate the dimensionality of the instruments used in this study. As noted in Chapter 4, the rules of thumb for interpreting correlation are small (0.10), medium (0.30), and large (0.50). The results of all item fit analyses for each latent variable (or scale) are provided in the following section for each latent variable (or scale).

The Epistemological Beliefs (EB) Scale

The EBAPS (*Epistemological Beliefs Assessment for Physical Science*) questionnaire (Elby, 2001) was adopted to examine the epistemological beliefs of 706 Indonesian pre-service physics teachers. As stated in Chapter 3, the EBAPS questionnaire consists of 30 items (i.e., EB1-EB30) that assess the participants' epistemological beliefs along five sub-scales. It was found that questions number EB19 and EB28 measured simultaneously more than one dimension or sub-scale. Furthermore, items EB4 and EB21 did not belong to any sub-scales. Since a multidimensional between-item model analysis was employed to examine the five dimensions of the **Epistemological Beliefs** scale, questions EB4, EB19, EB21, and EB28 were not included in the subsequent analysis.

The multidimensional form of the partial credit model (PCM) was used to examine the fit of these sub-scales. Using a multidimensional Rasch model approach may give an advantage in measuring person ability separately on several dimensions involved. In the case of this study, each item was expected to load on only one particular dimension measuring a single latent trait, meaning that there were to be no items loading on more than one dimension. According to Adams and Wu (2010), this method is known as the multidimensional between-item model. It is worth noting that there are various types of multidimensional models, namely between-item and within-item multidimensionality. A test is considered as multidimensional between-item if each item loads on only one sub-scale or dimension, while if any of the items measures simultaneously on more than one sub-scale, then the test is regarded as within-item multidimensional (Adams & Wu, 2010). The multidimensional between-item model procedure worked better with the data in this study than the within-item multidimensionality model. Hence, all data in this quantitative study were analysed using the multidimensional between-item model.

Further to this, the initial item analysis indicated that there were five items that have a discrimination index below 0.15, namely items EB5, EB11, EB12, EB16, and EB24. These items did not contribute information to the test needed in this study; hence, removal of these items was necessary. In other words, these items were also not included in further analysis because they potentially affected the results of the analysis obtained from this scale, or the accuracy in interpreting the structure. In short, there were 9 items from the epistemological beliefs questionnaire that were deleted and not processed in the data analysis in this study, which are EB4,

EB5, EB11, EB12, EB16, EB19, EB21, EB24, and EB28. The same procedure for item analysis was applied again to ensure that all items fitted the model.

As mentioned earlier, the weighted mean square (INFIT MNSQ) ranged from 0.70 to 1.30, which was used as an acceptable range to determine the model fit for all items. The items with infit mean squares falling outside the acceptable range were considered to be an unsatisfactory model-data fit, or misfitting and were therefore dropped from the data sets. The item properties of the epistemological beliefs questionnaire (EBAPS), after removing misfitting persons and items with a discrimination index below 0.15 in each sub-scale are presented in Table 5. 3. The initial item number is the encoding of the item number used for each item from the EBAPS questionnaire consisting of 30 items (i.e., no. EB1-EB30), while the final item number (i.e., no. 1 - 21) is the encoding of the item number used for each item from the EBAPS questionnaire after a number of items were removed from the data analysis due to the considerations explained previously.

Table 5. 3 Item parameter estimates for the five-dimensional Epistemological Beliefs model (n = 698)

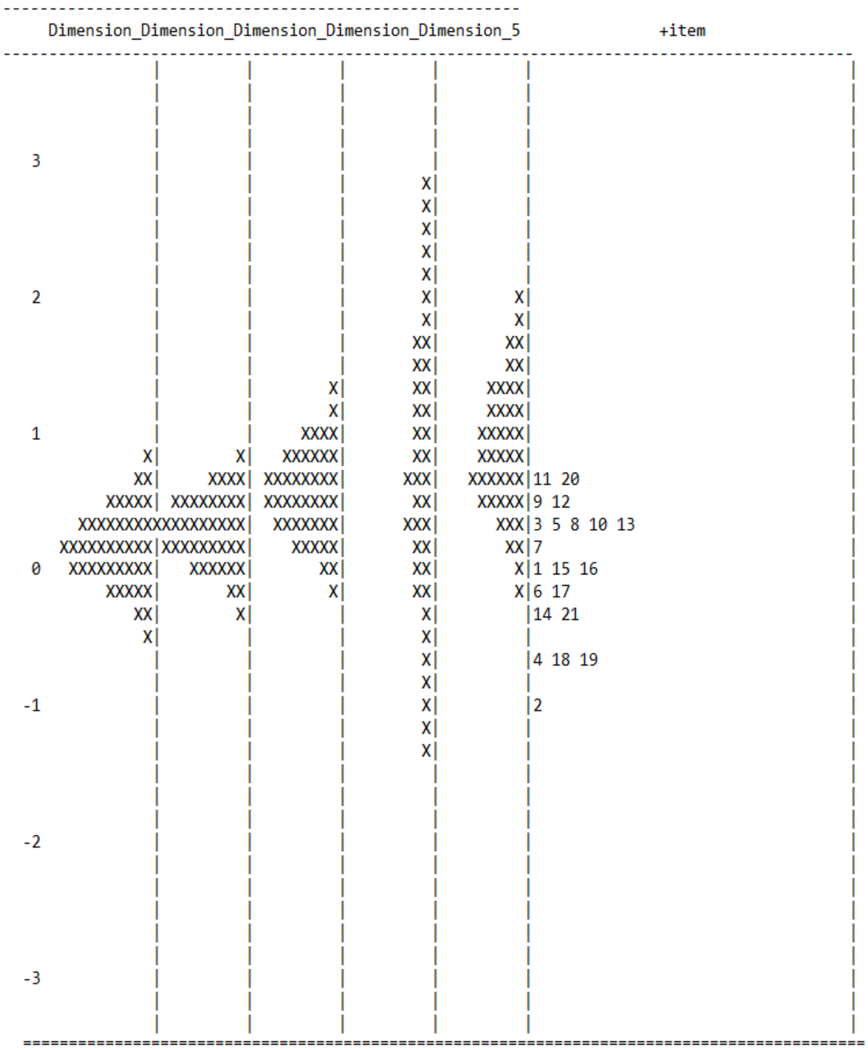
Initial Item Label	Final Item Label	Estimates (Logits)	Standard Errors (SE)	Weighted Fit (INFIT)			Item Discrimination
				MNSQ	CI	t	
Dimension 1: Structure of Scientific Knowledge (SSK)							
EB2	2	-0.96	0.07	0.97	(0.82, 1.18)	-0.30	0.24
EB8	6	-0.12	0.04	1.02	(0.91, 1.09)	0.40	0.31
EB10	8	0.29	0.03	0.99	(0.92, 1.08)	-0.20	0.41
EB15	11	0.64	0.06	0.99	(0.91, 1.09)	-0.30	0.31
EB17	12	0.56	0.04	0.98	(0.92, 1.08)	-0.60	0.47
EB20	14	-0.42	0.04	1.03	(0.89, 1.11)	0.60	0.26
EB23	16	0.00	0.04	1.01	(0.91, 1.09)	0.20	0.32
Dimension 2: Nature of Knowing and Learning (NKL)							
EB1	1	-0.04	0.04	0.95	(0.91, 1.09)	-1.10	0.45
EB7	5	0.28	0.03	1.01	(0.93, 1.07)	0.20	0.38
EB13	9	0.50	0.04	1.02	(0.93, 1.07)	0.70	0.37
EB18	13	0.25	0.04	1.08	(0.91, 1.09)	1.60	0.15
EB26	18	-0.68	0.05	0.98	(0.88, 1.12)	-0.20	0.34
EB30	21	-0.31	0.04	1.02	(0.89, 1.11)	0.50	0.33
Dimension 3: Real-life Applicability (RLA)							
EB3	3	0.36	0.04	1.06	(0.91, 1.09)	1.30	0.17
EB14	10	0.29	0.04	1.00	(0.91, 1.09)	0.10	0.43
EB27	19	-0.65	0.05	1.02	(0.85, 1.15)	0.30	0.27
Dimension 4: Evolving Knowledge (EK)							
EB6	4	-0.63	0.06	1.02	(0.89, 1.11)	0.30	0.38
EB29	20	0.63	0.06	0.98	(0.91, 1.09)	-0.50	0.46
Dimension 5: Source of Ability to Learn (SAL)							
EB9	7	0.19	0.05	0.95	(0.91, 1.09)	-1.20	0.51
EB22	15	0.04	0.05	1.07	(0.91, 1.09)	1.50	0.32
EB25	17	-0.24	0.05	1.02	(0.89, 1.11)	0.40	0.26

Item Separation Reliability = 0.991

As shown in Table 5. 3, the infit mean square (INFIT MNSQ) values for 21 items in the EBAPS questionnaire were within the acceptable range (0.70 - 1.30), ranging from 0.95 to 1.08. These values indicate a sufficient fit to the Rasch model for practical measurement purposes. In addition, all items (21 items) on the **Epistemological Beliefs** scale had a discrimination index ≥ 0.15, ranging from 0.15 to 0.51. This index shows significant relationships arising between each item and the total score in the sub-scales within the EBAPS questionnaire.

Furthermore, the reliability test of the research instrument was also conducted in this study. Internal consistency using Cronbach's alpha coefficient was examined to test the reliability of the instruments. The Cronbach's alpha coefficient of the **Epistemological Beliefs** scale was 0.61, which indicates acceptable reliability values (van Griethuijsen et al., 2015). In addition, the item separation reliability index for this scale was 0.991 indicating high reliability as well as representing the consistency of item estimates for replication of another sample within a similar population. The Rasch analysis continued with the examination of the person–item map (Wright map) to identify the distribution pattern of person ability and item difficulty estimates which were calibrated to be on the same logit scale. In doing so, all model parameter estimates can be easily compared on the same scale in terms of targeting of the test. The item–person map in Figure 5. 4 displays a five-dimensional model comprising 21 items to examine the epistemological beliefs of Indonesian pre-service physics teachers. Person ability and item difficulty were placed on the left side and right side of the logit scale, respectively. Respondents with the highest level of epistemological beliefs and items with the highest difficulty levels are located at the top of the map, while respondents with lowest level of epistemological beliefs and items with the lowest difficulty levels are located at the bottom of the map, with a logit scale of 0 as the mean of the item difficulty level.

As shown in Figure 5. 4, the numbers on the left side represent the logit scale, which is a metric for both the person ability estimates and the item difficulty estimates. As noted earlier, the greater the logit value, the higher the level of the person's ability, and the less likely the items will be endorsed. In addition, each 'X' represents a number of cases; this number is dependent on the sample size of the study. The Wright map for the **Epistemological Beliefs** scale indicates that each 'X' represents about 16.6 cases. Meanwhile, the numerals in the far-right column represent the item numbers. It can be seen that the item difficulty was distributed between logit scale 0.6 and -1 along the five dimensions. Respondents performed better in dimensions three and five because most of the respondents were clustered more towards the upper part of the logit scale. In dimension four, the distribution of respondents' ability covered a wider range of the scales.



Note: Each 'X' represents 16.6 cases and the column of numbers to the left is a logit scale

Figure 5. 4 Item–person map of the five-dimensional Epistemological Beliefs model

Generally speaking, the item difficulty of the **Epistemological Beliefs** variable covers the span of the pre-service physics teachers' ability distributions quite well in all five dimensions. In other words, the distributions of item difficulty in all sub-scales were well-targeted to the **Epistemological Beliefs** traits, even though they were not an ideal distribution pattern of person ability and item difficulty estimates. There are a number of cases which are above the most difficult items, which means that the epistemological beliefs questionnaire (EBAPS) was slightly easier to answer correctly by the respondents in this study, especially for dimensions four and five. The estimated correlation between the five dimensions of the **Epistemological Beliefs** variables produced by the ConQuest software program can be seen in Table 5. 4.

Table 5. 4 The estimated correlations between the dimensions of Epistemological Beliefs variable

Dimension	SSK	NKL	RLA	EK	SAL
Dimension 1: SSK	1.00				
Dimension 2: NKL	0.60	1.00			
Dimension 3: RLA	0.47	0.54	1.00		
Dimension 4: EK	0.42	0.58	0.33	1.00	
Dimension 5: SAL	0.56	0.51	0.36	0.67	1.00

Note: SSK = Structure of Scientific Knowledge; NKL = Nature of Knowing and Learning; RLA = Real-life Applicability; EK = Evolving Knowledge; and SAL = Source of Ability to Learn

Table 5. 4 summarises the correlations (*r*) among the dimensions under the multidimensional model, with the lowest correlation ($r = 0.33$) being between the dimensions of evolving knowledge and real-life applicability, and the highest correlation ($r = 0.67$) being between the source of ability to learn and evolving knowledge dimensions. The correlation value between the **Epistemological Beliefs** sub-scales ranged from medium to large positive correlations. As expected, the pattern of correlations between the five dimensions was quite satisfactory, given the fact that all the dimensions were designed to measure the concept of epistemological beliefs in accordance with the theory proposed by the authors of this instrument (EBAPS).

The Argumentation (AG) Scale

The *Argumentation Test* (Sampson & Clark, 2006), adopted to assess the participants' skills in argumentation, consists of 36 items separated by two dimensions. The initial item analysis revealed that there were eight items detected as having a discrimination index below 0.15, namely items AG1, AG4, AG5, AG6, AG16, AG22, AG26, and AG30. These items were indicated as poor functioning items and were considered as not providing sufficient information for the tests needed in this study; hence, removal of these items was necessary. In other words, these items were not included in the subsequent analysis because they could have affected the results of the analysis obtained on this scale. The same procedure of item analysis was applied a second time to ensure that all items fitted the model. The item properties of the **Argumentation** latent variable, after dropping the items that had a discrimination index below 0.15 for each sub-scale, are presented in Table 5. 4. The initial item number is the encoding of the item number used for each item from the *Argumentation Test* consisting of 36 items (i.e., no. AG1-AG36), while the final item number (i.e., no. 1 - 28) is the encoding of the item number used for each item from the *Argumentation Test* after a number of items were removed from the data sets.

As shown in Table 5. 5, none of the items with the weighted mean square value falls outside the acceptable range. The INFIT MNSQ value for 28 items ranged from 0.84 to 1.17, which exhibited good model fit, where 0.70 to 1.30 was the cut-off value. The discrimination index for these items ranged from 0.16 to 0.58. These indices show significant relationships between each item and the total score in the sub-scales within the *Argumentation Test*.

Table 5. 5 Item parameter estimates for the two-dimensional Argumentation model (n = 706)

Initial Item Label	Final Item Label	Estimate (Logits)	Standard Errors (SE)	Weighted Fit (INFIT)			Item Discrimination
				MNSQ	CI	t	
Dimension 1: Making Scientific Argumentation (MSA)							
AG2	1	0.44	0.09	1.03	(0.89, 1.11)	0.60	0.20
AG3	2	0.44	0.09	1.03	(0.89, 1.11)	0.60	0.17
AG7	3	1.48	0.13	0.99	(0.78, 1.22)	-0.10	0.17
AG8	4	-0.40	0.08	0.99	(0.94, 1.06)	-0.20	0.32
AG9	5	-0.21	0.08	1.00	(0.93, 1.07)	0.00	0.28
AG10	6	0.06	0.09	1.02	(0.92, 1.08)	0.40	0.17
AG11	7	0.44	0.09	1.01	(0.89, 1.11)	0.20	0.22
AG12	8	-0.68	0.08	0.95	(0.95, 1.05)	-2.10	0.46
AG13	9	-2.14	0.09	0.95	(0.92, 1.08)	-1.40	0.39
AG14	10	-0.21	0.08	1.04	(0.93, 1.07)	1.30	0.20
AG15	11	-0.38	0.08	0.99	(0.94, 1.06)	-0.50	0.35
AG17	12	0.35	0.09	1.03	(0.89, 1.11)	0.60	0.16
AG18	13	0.83	0.10	1.03	(0.85, 1.15)	0.40	0.16
Dimension 2: Challenging Argumentation (CA)							
AG19	14	1.63	0.10	1.10	(0.88, 1.12)	1.60	0.19
AG20	15	1.08	0.09	1.04	(0.91, 1.09)	0.90	0.32
AG21	16	0.70	0.09	1.09	(0.93, 1.07)	2.40	0.32
AG23	17	-1.59	0.09	0.95	(0.90, 1.10)	-1.00	0.41
AG24	18	1.02	0.09	1.07	(0.91, 1.09)	1.60	0.27
AG25	19	-2.70	0.13	0.87	(0.82, 1.18)	-1.50	0.41
AG27	20	0.53	0.08	1.09	(0.93, 1.07)	2.60	0.27
AG28	21	0.56	0.08	1.17	(0.93, 1.07)	4.60	0.17
AG29	22	0.05	0.08	0.98	(0.94, 1.06)	-0.80	0.47
AG31	23	0.50	0.08	0.98	(0.93, 1.07)	-0.50	0.40
AG32	24	-0.22	0.08	0.93	(0.94, 1.06)	-2.50	0.51
AG33	25	0.62	0.08	0.95	(0.93, 1.07)	-1.50	0.47
AG34	26	-1.91	0.10	0.89	(0.88, 1.12)	-1.80	0.47
AG35	27	-1.09	0.09	0.84	(0.92, 1.08)	-4.40	0.58
AG36	28	0.83	0.09	0.98	(0.92, 1.08)	-0.60	0.44

Item Separation Reliability = 0.993

In addition, the Cronbach's alpha coefficient of the **Argumentation** scale was 0.68, indicating acceptable reliability values (van Griethuijsen et al., 2015). Meanwhile, the item separation reliability index for the **Argumentation** scale was 0.993, demonstrating high reliability and indicating the consistency of item estimates for replication of another sample within a similar population. The next Rasch measurement analysis was carried out to identify the distribution pattern of person ability and item difficulty estimates for Indonesian pre-service physics teachers, as presented in Figure 5. 5.

```
----------------------------------------------
          Dimension_1         Dimension_2                      +item
---------------------------------------------------------------------------------------
                          |                   |                                       |
                          |                   |                                       |
                          |                   |                                       |
                          |                   |                                       |
   3                      |                   |                                       |
                          |                   |                                       |
                          |                   |                                       |
                          |                   |                                       |
                          |                   |                                       |
                          |                   |                                       |
                          |                   |                                       |
                          |                   |                                       |
   2                      |                  X|                                       |
                          |                  X|                                       |
                          |                  X|                                       |
                          |                 XX|14                                     |
                          |                 XX|3                                      |
                          |                XXX|                                       |
                          |              XXXXX|                                       |
                          |              XXXXX|                                       |
   1                      |          XXXXXXXXX|15 18                                  |
                          |           XXXXXXXX|28                                     |
                          |             XXXXXX|13                                     |
                          |           XXXXXXXX|16 25                                  |
                          |           XXXXXXXX|20 21 23                               |
                         X|       XXXXXXXXXXXX|1 2 7 12                               |
                        XX|      XXXXXXXXXXXXX|                                       |
                        XX|     XXXXXXXXXXXXXX|                                       |
   0                  XXXX|     XXXXXXXXXXXXXX|6 22                                   |
                      XXXX|      XXXXXXXXXXXXX|                                       |
                  XXXXXXXX|         XXXXXXXXXX|5 10 24                                |
               XXXXXXXXXXX|       XXXXXXXXXXXX|4 11                                   |
            XXXXXXXXXXXXXX|         XXXXXXXXXX|                                       |
            XXXXXXXXXXXXXX|        XXXXXXXXXXX|                                       |
          XXXXXXXXXXXXXXXX|            XXXXXXX|8                                      |
         XXXXXXXXXXXXXXXXX|          XXXXXXXXX|                                       |
  -1  XXXXXXXXXXXXXXXXXXXX|             XXXXXX|                                       |
      XXXXXXXXXXXXXXXXXXXX|            XXXXXXX|27                                     |
       XXXXXXXXXXXXXXXXXXX|           XXXXXXXX|                                       |
        XXXXXXXXXXXXXXXXXX|            XXXXXXX|                                       |
        XXXXXXXXXXXXXXXXXX|              XXXXX|                                       |
             XXXXXXXXXXXXX|               XXXX|17                                     |
               XXXXXXXXXXX|                XXX|                                       |
                  XXXXXXXX|               XXXX|26                                     |
  -2                XXXXXX|                XXX|                                       |
                     XXXXX|                  X|9                                      |
                       XXX|                 XX|                                       |
                        XX|                  X|                                       |
                          |                   |                                       |
                         X|                  X|                                       |
                         X|                   |19                                     |
                         X|                  X|                                       |
  -3                      |                   |                                       |
                          |                   |                                       |
                          |                   |                                       |
                          |                   |                                       |
                          |                   |                                       |
                          |                   |                                       |
                          |                   |                                       |
                          |                   |                                       |
=======================================================================================
```

Note: Each 'X' represents 3.0 cases and the column of numbers to the left is a logit scale.

Figure 5. 5 Item–person map of the two-dimensional Argumentation model

The respondents are placed on the left side of the scale according to their argumentation skills while the items are shown on the right side of the scale. Respondents with the highest argumentation skills and items with the highest difficulty levels are located at the top of the map, while respondents with the lowest argumentation skills and the items with the lowest difficulty levels are located at the bottom of the map, where a logit scale of 0 represents the mean item difficulty level. The item difficulty ranged between 1.63 and -2.70 logits along each of the two dimensions. The item–person map for the **Argumentation** variable indicated that each 'X' represents about 3.0 cases. Figure 5. 5 shows that the respondents performed better in dimension two, while most of the respondents were clustered more towards the lower part of the logit scale with 0 as the mean of the item difficulty level in dimension one. Notably, items 3 and 13 were somewhat difficult for the participants to answer correctly. Overall, the item difficulty of the **Argumentation** variable covered the span of the pre-service physics teachers' ability distributions quite well in both dimensions. In other words, the distribution of item difficulty in both sub-scales was well-aligned with the **Argumentation** scale, although the distribution pattern for person ability was below the mean for item difficulty. In addition, the estimated correlation between the two dimensions of the **Argumentation** variable (i.e., Making Scientific Argumentation and Challenging Argumentation) was 0.51, which indicated large positive correlations. This value reflected the fact that both dimensions were measuring the concept of argumentation skills among the test takers.

The Scientific Reasoning (SR) Scale

The LCTSR (*Lawson's Classroom Test of Scientific Reasoning*) questionnaire (Lawson, 1978, ver. 2000), which is commonly known as the *Lawson's Test*, was adopted in this study to examine the scientific reasoning skills of 706 Indonesian pre-service physics teachers. In the case of this study, the **Scientific Reasoning** scale involved five dimensions as described in Chapter 3. The Rasch measurement analysis indicated that there were no items with a discrimination index below 0.15. This means that no items were deleted from the data sets, and all items were included in the subsequent analysis. The item analysis of the **Scientific Reasoning** latent variable for each dimension is presented in Table 5. 6.

As can be seen in Table 5. 6, all items were within the threshold values of 0.70 to 1.30, ranging from 0.93 to 1.06. These INFIT MNSQ values indicate a good model fit. In addition, the discrimination index for these items ranged from 0.26 to 0.63. These indices show significant relationships between each item and the total score in the sub-scales within the *Lawson's Test*. With regard to the reliability test, the Cronbach's alpha coefficient of the **Scientific Reasoning** scale was 0.66, showing acceptable reliability values (van Griethuijsen et al., 2015). Meanwhile, the item separation reliability index for the **Scientific Reasoning** scale was 0.977, demonstrating high reliability and indicating consistency in item estimates for replication in another sample within a similar population.

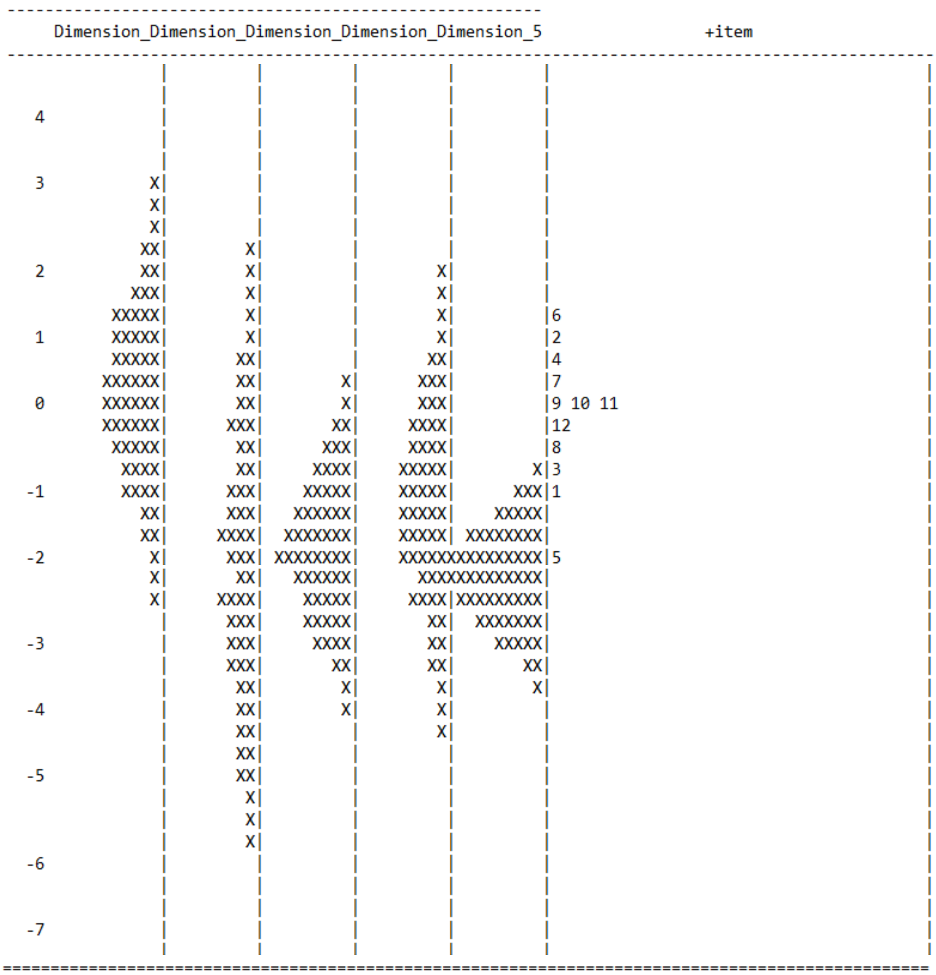
Table 5. 6 Item parameter estimates for the five-dimensional Scientific Reasoning model (n = 706)

Initial Item Label	Final Item Label	Estimate (Logits)	Standard Errors (SE)	Weighted Fit (INFIT)			Item Discrimination
				MNSQ	CI	t	
Dimension 1: Conservation of Weight and Volume (CWV)							
SR1	1	-1.16	0.35	1.00	(0.91, 1.09)	0.00	0.42
SR2	2	1.16	0.35	1.01	(0.92, 1.08)	0.30	0.49
Dimension 2: Proportional Reasoning (PR)							
SR3	3	-0.87	0.09	1.00	(0.89, 1.11)	0.00	0.62
SR4	4	0.87	0.09	1.04	(0.84, 1.16)	0.50	0.49
Dimension 3: Control of Variables (COV)							
SR5	5	-1.84	0.09	1.01	(0.94, 1.06)	0.30	0.48
SR6	6	1.34	0.11	0.95	(0.75, 1.25)	-0.30	0.36
SR7	7	0.50	0.10	1.01	(0.84, 1.16)	0.20	0.36
Dimension 4: Probability and Correlation Reasoning (PCR)							
SR8	8	-0.31	0.08	0.93	(0.92, 1.08)	-1.70	0.63
SR9	9	0.13	0.08	1.06	(0.90, 1.10)	1.30	0.47
SR10	10	0.18	0.08	1.06	(0.90, 1.10)	1.10	0.49
Dimension 5: Hypothetical-Deductive Reasoning (HDR)							
SR11	11	0.16	0.09	0.99	(0.82, 1.18)	-0.10	0.34
SR12	12	-0.16	0.09	1.02	(0.85, 1.15)	0.30	0.26

Item Separation Reliability = 0.977

The next Rasch analysis was to examine the person–item map to identify the distribution pattern of person ability and item difficulty estimates, as displayed in Figure 5. 6. Respondents with the highest scientific reasoning skills and items with the highest difficulty levels were located at the top of the Wright map, while respondents with the lowest scientific reasoning skills and items with the lowest difficulty levels were located at the bottom of the map, with a logit scale value of 0 as the mean of the item difficulty level.

Item difficulty was distributed between 1.34 and -1.84 logits along the five dimensions. In addition, the Wright map for the **Scientific Reasoning** scale indicated that each 'X' represents about 11.4 cases. According to Figure 5. 6, the respondents performed better in dimensions one, two, and four. However, most of the respondents were clustered towards the lower part of the logit scale, with 0 as the mean of the item difficulty level in dimensions three and five. In particular, items 6, 11, and 12 were somewhat difficult for the respondents to answer correctly.



Note: Each 'X' represents 11.4 cases and the column of numbers to the left is a logit scale.

Figure 5. 6 Item–person map of the five-dimensional Scientific Reasoning model

Generally speaking, the distributions of item difficulty were well targeted to the **Scientific Reasoning** scale, especially for dimensions one, two, three, and four. The item difficulty level for dimension five does not seem to cover well the span of Indonesian pre-service physics teachers' ability distribution. The estimated correlation between the five dimensions of the **Scientific Reasoning** variable is presented in Table 5. 7.

Table 5. 7 The estimated correlation between the dimensions of Scientific Reasoning variable

Dimension	CWV	PR	COV	PCR	HDR
Dimension 1: CWV	1.00				
Dimension 2: PR	0.72	1.00			
Dimension 3: COV	0.71	0.86	1.00		
Dimension 4: PCR	0.58	0.69	0.79	1.00	
Dimension 5: HDR	0.54	0.63	0.73	0.85	1.00

Note: CWV = Conservation of Weight and Volume; PR = Proportional Reasoning; COV = Control of Variables; PCR = Probability and Correlation Reasoning; and HDR = Hypothetical-Deductive Reasoning

Table 5. 7 summarises the correlations (*r*) among dimensions under the multidimensional model, with the lowest correlation (*r* = 0.54) being between the CWV (Conservation of Weight and Volume) and the HDR (Hypothetical-Deductive Reasoning) dimensions, while the highest correlation (*r* = 0.86) was between the PR (Proportional Reasoning) and COV (Control of Variables) dimensions. The correlation value between the **Scientific Reasoning** sub-scales indicates large positive correlations. As expected, the pattern of correlation between the five dimensions was quite satisfactory, given the fact that all the dimensions were designed to measure the concept of scientific reasoning in accordance with the theory proposed by the authors of the *Lawson's Test*.

The Physics Conceptual Understanding (PCU) Scale

In assessing the physics conceptual understanding of 706 Indonesian pre-service physics teachers, the *Force Concept Inventory* (FCI) and the *Brief Electricity and Magnetism Assessment* (BEMA) tests were adopted in this study. These tests have been proposed respectively as diagnostic assessment tools in basic Mechanics and Electricity & Magnetism topics. The FCI test consists of 30 multiple-choice items (Hestenes et al., 1992), while the BEMA test consists of 31 multiple-choice items (Chabay & Sherwood, 1997). In other words, the total items used to examine the conceptual understanding of physics of the participants was 61 items grouped into two dimensions.

According to Rasch measurement analysis, there were 21 items that had a discrimination index below 0.15, namely items CU9, CU11, CU13, CU17, CU18, CU25, CU26, CU30, CU39, CU41, CU42, CU44, CU46, CU47, CU48, CU49, CU53, CU57, CU58, CU59, and CU61. Consequently, these items were detected to be poor functioning items and considered as not providing sufficient information for the tests used in this study; they were removed from subsequent analyses. In other words, these items were not included in the subsequent analysis because they potentially affected the results of the analysis obtained from this scale, or the accuracy in interpreting the structure. As outlined by Wu and Adams (2007), the item discrimination index indicates the correlation between an individual's score on the item and their total score on the test. Therefore, the low discrimination index indicated that the items did not show a significant relationship to the total score. After removing these items, the same procedure of item analysis using the multidimensional Rasch model was re-applied to ensure that all items involved in this study fitted the model. The item properties of the **Physics Conceptual Understanding** variable, after removing the items with a discrimination index below 0.15 for each sub-scale, are presented in Table 5. 8.

Table 5. 8 Item parameter estimates for the two-dimensional PCU model (n = 706)

Initial Item Label	Final Item Label	Estimate (Logits)	Standard Errors (SE)	Weighted Fit (INFIT)			Item Discrimination
				MNSQ	CI	t	
Mechanics (MECH)							
CU1	1	-0.94	0.08	0.97	(0.96, 1.04)	-1.50	0.38
CU2	2	0.27	0.09	0.98	(0.89, 1.11)	-0.40	0.29
CU3	3	0.36	0.10	1.02	(0.88, 1.12)	0.30	0.23
CU4	4	0.90	0.11	0.99	(0.83, 1.17)	-0.10	0.26
CU5	5	0.54	0.10	1.03	(0.87, 1.13)	0.40	0.18
CU6	6	-0.97	0.08	1.08	(0.96, 1.04)	3.40	0.15
CU7	7	-0.74	0.08	1.03	(0.95, 1.05)	1.10	0.26
CU8	8	0.24	0.09	1.02	(0.89, 1.11)	0.30	0.22
CU10	9	0.26	0.09	1.00	(0.89, 1.11)	0.00	0.26
CU12	10	-1.28	0.08	0.96	(0.96, 1.04)	-2.20	0.40
CU14	11	0.36	0.10	0.96	(0.88, 1.12)	-0.70	0.38
CU15	12	0.55	0.10	1.00	(0.87, 1.13)	0.00	0.24
CU16	13	-0.54	0.08	1.04	(0.94, 1.06)	1.20	0.22
CU19	14	0.15	0.09	0.97	(0.90, 1.10)	-0.50	0.36
CU20	15	-0.02	0.09	0.94	(0.91, 1.09)	-1.50	0.40
CU21	16	0.40	0.10	1.01	(0.88, 1.12)	0.10	0.24
CU22	17	0.24	0.09	1.06	(0.89, 1.11)	1.00	0.16
CU23	18	1.38	0.14	1.02	(0.78, 1.22)	0.20	0.17
CU24	19	-0.77	0.08	1.02	(0.95, 1.05)	0.90	0.28
CU27	20	-0.25	0.08	1.02	(0.93, 1.07)	0.70	0.23
CU28	21	0.77	0.11	0.95	(0.85, 1.15)	-0.70	0.37
CU29	22	-0.92	0.08	0.97	(0.96, 1.04)	-1.30	0.35
Electricity and Magnetism (EM)							
CU31	23	-1.05	0.08	0.98	(0.95, 1.05)	-1.00	0.38
CU32	24	-1.95	0.08	1.02	(0.94, 1.06)	0.50	0.29
CU33	25	-1.18	0.08	0.98	(0.96, 1.04)	-1.10	0.38
CU34	26	-0.25	0.08	0.99	(0.93, 1.07)	-0.30	0.37
CU35	27	0.59	0.10	1.02	(0.87, 1.13)	0.30	0.26
CU36	28	0.61	0.10	1.01	(0.87, 1.13)	0.20	0.24
CU37	29	0.01	0.09	1.00	(0.91, 1.09)	0.00	0.33
CU38	30	-0.38	0.08	1.01	(0.93, 1.07)	0.40	0.29
CU40	31	-0.56	0.08	1.03	(0.94, 1.06)	1.00	0.19
CU43	32	-0.05	0.09	1.00	(0.92, 1.08)	0.00	0.30
CU45	33	0.52	0.10	0.97	(0.87, 1.13)	-0.40	0.34
CU50	34	0.41	0.10	1.00	(0.88, 1.12)	0.00	0.24
CU51	35	-0.05	0.09	1.02	(0.92, 1.08)	0.60	0.25
CU52	36	1.19	0.12	1.03	(0.81, 1.19)	0.30	0.19
CU54	37	0.32	0.10	1.03	(0.89, 1.11)	0.60	0.24
CU55	38	0.99	0.12	1.01	(0.83, 1.17)	0.10	0.24
CU56	39	0.17	0.09	0.99	(0.90, 1.10)	-0.10	0.28
CU60	40	0.69	0.11	0.97	(0.86, 1.14)	-0.50	0.34

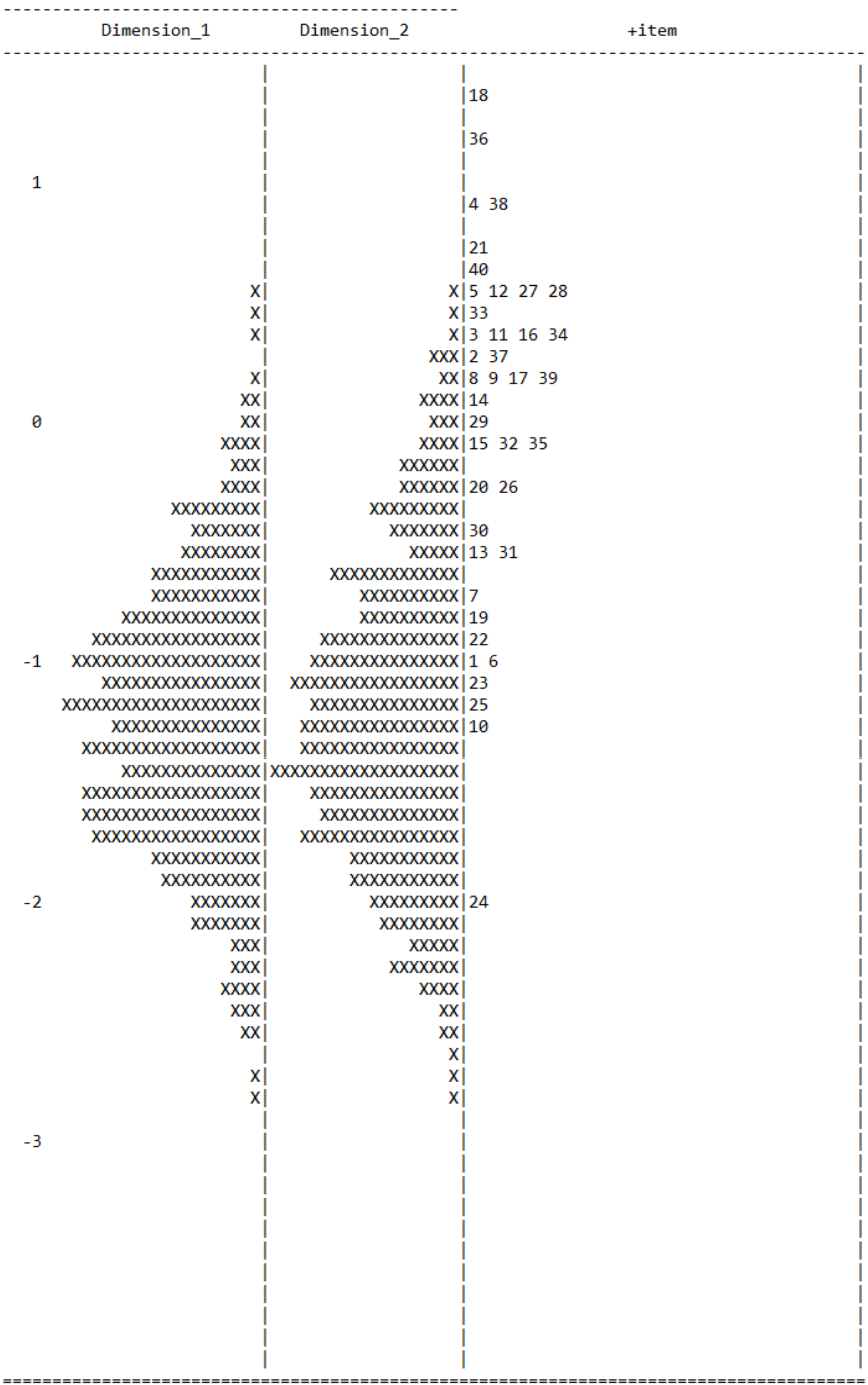
Item Separation Reliability = 0.984

The initial item number is the encoding of the item number used for each item from the physics conceptual understanding instruments consisting of 61 items (i.e., no. CU1-CU61), while the final item number (i.e., no. 1- 40) is the encoding of the item number used for each item from the physics conceptual understanding instruments after a number of items were removed from the data sets. As can be seen in Table 5. 8, the item fit analysis indicates that all items were within the threshold range of 0.70 to 1.30, indicating good item fit, with values ranging from 0.94 to 1.08. In addition, the item discrimination index ranged from 0.15 to 0.40. The Cronbach's alpha coefficient of the scale for **Physics Conceptual Understanding** was 0.70, demonstrating acceptable reliability values (van Griethuijsen et al., 2015). Meanwhile, the item separation reliability index for the **Physics Conceptual Understanding** scale was 0.984, indicating high reliability. As asserted by Bond and Fox (2015), the higher the values of the item separation reliability, the better the reliability of item location replicability on the logit scale, should these items be given to other similar sample groups for whom it were appropriate, as well as for there to be a smaller measurement error (Ben, 2010).

The following Rasch measurement analysis was used to identify the distribution pattern of person ability and item difficulty estimates for Indonesian pre-service physics teachers, as presented in Figure 5. 7. Respondents with the highest ability to understand physics concepts and items with the highest difficulty level were located at the top of the map, while respondents with the lowest understanding of physics concepts and the lowest difficulty items level were located at the bottom of the map, where the logit scale of 0 represented the average item difficulty. The item difficulty was distributed between 1.38 and -1.95 logits for each of the two dimensions. The item–person map for the **Physics Conceptual Understanding** variable indicated that each 'X' represented about 2.3 cases.

Figure 5. 7 shows that most of the respondents were clustered more towards the lower part of the mean value of the item difficulty level in both dimensions. It seems that respondents found some items from both instruments difficult to answer correctly, particularly items 4, 18, 21, 36, 38, and 40. Nevertheless, the other items in the physics conceptual understanding test spanned the pre-service physics teachers' ability distributions quite well in each dimension. In other words, the distributions of item difficulty in both sub-scales were appropriately targeted to the **Physics Conceptual Understanding** scale, although the distribution pattern of person ability estimates was below that of the mean for item difficulty. In addition, the estimated correlation between each

dimension of the **Physics Conceptual Understanding** variable (i.e., Mechanics and Electricity & Magnetism dimension) was 0.60, indicating large positive correlations. This value suggests that the two dimensions are measuring the test takers understanding of physics concepts.



Note: Each 'X' represents 2.3 cases and the column of numbers to the left is a logit scale.

Figure 5. 7 Item–person map of the two-dimensional Physics Conceptual Understanding model

To sum up, according to the person and item fit analysis using the ConQuest 4 statistical software package, the number of items fitting the Rasch model for each latent variable after removing the misfitting persons was 21 items for the **Epistemological Beliefs** scale, 28 items for the **Argumentation** scale, 12 items for the **Scientific Reasoning** scale, and 40 items for the **Physics Conceptual Understanding** scale. These items were considered as fitting the model and to be good quality measures of what is needed in this study. Hence, they were used for the subsequent analysis. Furthermore, the identified misfitting persons were then re-incorporated back into the data sets for scoring purposes. As highlighted earlier, this study employed the WLE scores instead of raw scores. These WLE scores were used for subsequent analysis such as Confirmatory Factor Analysis (CFA) as well as descriptive and inferential analyses, as presented in the following section.

5.4 Confirmatory Factor Analysis (CFA) Model

As noted earlier, the present study adopted five existing questionnaires or tests to investigate scientific thinking (consisting of epistemological beliefs, argumentation, and scientific reasoning) and conceptual understanding in physics of Indonesian pre-service physics teachers. The raw data collected from the participants' responses to each questionnaire was transformed into interval scales, and given a Weighted Likelihood Estimation (WLE) score. The WLE score was presented as the unit of measurement in the Rasch analysis, a logit (log odds unit) scale. As suggested by Ben (2010), transforming raw scores to measures using the WLE method could provide better estimates of respondents' scores. In the case of this study, the WLE score was obtained using the anchor method in the Rasch measurement analysis, where the score was then used for subsequent analysis such as the Confirmatory Factor Analysis (CFA).

CFA analysis is carried out to validate the relationship between observed and unobserved variables based on previous theories. In addition, a confirmatory analysis was also used in this study to test if the hypothesised factor structure of the model was supported by the real data. Description of the CFA analysis results for each scale or construct (factor) are presented in the following section.

5.4.1 CFA Model Fit Indices

The CFA procedure was used to examine construct validity using AMOS software version 25. In the present study, the scale consisted of **Epistemological Beliefs**, **Argumentation**, **Scientific Reasoning**, and **Physics Conceptual Understanding**. Meanwhile, the other variables such as year level, gender, and type of university were not examined through the CFA procedure because these variables have simple and fixed structures in which the data sets obtained from the official documents were provided by the universities involved in the study. The CFA model was carried out by assessing the factor loading of each observed variable. In this study, the value of the factor loading considered as a good fit for the hypothesised model was 0.30 (Hair et al., 2014).

Multiple fit indices were used to examine the CFA model fit comprising the χ^2/DF ratio (chi-square divided by the number of degrees of freedom), Goodness-of-Fit Index (GFI), Adjusted Goodness-of-Fit Index (AGFI), Tucker-Lewis Index (TLI), Comparative Fit Index (CFI), and Root Mean Square Error of Approximation (RMSEA). The CFA model performed good fit if the value of the χ^2/DF ratio was small. Generally speaking, there is no specific threshold for the χ^2/DF ratio, which indicates the maximum acceptable value for a model (Kline, 2016). Meanwhile, the index of the GFI, AGFI, TLI, and CFI indicates a good model fit if the value is close to or at ≥ 0.90. In addition, the CFA model performs good fit and is acceptable if the value of the RMSEA is less than 0.08, or not more than 0.10 (Byrne, 2016). As highlighted by Kline (2011), a model that has an RMSEA value greater than 0.10 would not be accepted as a good model, and would be regarded as a problematic model fit.

The Epistemological Beliefs (EB) Scale

The CFA model for **Epistemological Beliefs** scale consists of five sub-scales or dimensions namely: Structure of Scientific Knowledge (SSK), Nature of Knowing and Learning (NKL), Real-Life Applicability (RLA), Evolving Knowledge (EK), and Source of Ability to Learn (SAL). The structure of the CFA model for the **Epistemological Beliefs** scale is presented in Figure 5. 8.

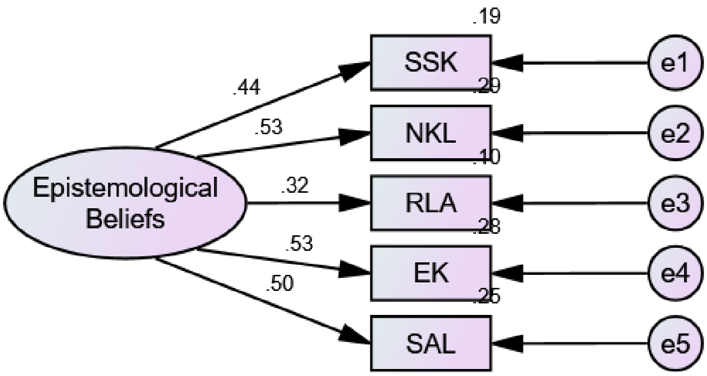


Figure 5. 8 The CFA model of Epistemological Beliefs scale

Note: SSK = Structure of Scientific Knowledge; NKL = Nature of Knowing and Learning; RLA = Real-Life Applicability; EK = Evolving Knowledge; and SAL = Source of Ability to Learn

As noted previously, CFA analysis concerns the extent to which the observed variables represent the underlying latent variables. As shown in Figure 5. 9, the results of CFA analysis simultaneously estimate the strength of the regression path from the latent variable (or unobserved variable) to the observed variables or sub-scales. The loading value for each sub-scale is presented in Table 5. 9.

Table 5. 9 Factor loadings of Epistemological Beliefs scale

Scale	Sub-scale	Loading
	SSK	0.44
	NKL	0.53
Epistemological Beliefs (EB)	RLA	0.32
	EK	0.53
	SAL	0.50

Note: SSK = Structure of Scientific Knowledge; NKL = Nature of Knowing and Learning; RLA = Real-Life Applicability; EK = Evolving Knowledge; and SAL = Source of Ability to Learn

Table 5. 9 shows that the values of the factor loadings ranged from 0.32 to 0.53. The RLA sub-scale had the lowest factor loading in this questionnaire, namely 0.32. However, this loading value was still within the minimum cut-off level of 0.30 applied in this study (Hair et al., 2014). Overall, all sub-scales were loaded above 0.30, indicating a good model fit. These dimensions represented the latent variables that were intended to be measured. The summary of the fit indices for the **Epistemological Beliefs** scale is presented in Table 5. 10.

Table 5. 10 Summaries of fit indices of the Epistemological Beliefs scale

Model Fit Indices	Values
χ^2/DF ratio	6.68
Goodness-of-Fit Index (GFI)	0.98
Adjusted Goodness-of-Fit Index (AGFI)	0.95
Tucker-Lewis Index (TLI)	0.80
Comparative Fit Index (CFI)	0.90
Root Mean Square Error of Approximation (RMSEA)	0.09

The summary findings of fit indices of this model in Table 5. 10 shows that the value of χ^2/DF was 6.68, indicating an acceptable value, even though it was higher than the acceptable value of the ratio of a chi-square to the number of degrees of freedom. As suggested by Darmawan (2003), a χ^2/DF ratio less than 5 indicates a good fit to the data. However, the value of the chi-square is very sensitive to sample size in which a large sample size contributes to the high value of the chi-square. Therefore, it is important to check other fit indices to ensure that the model provides a good fit to the data of this study. The other model fit indices showed satisfactory values. For instance, the GFI, AGFI, and CFI indicated a good model fit to the data, namely 0.98, 0.95, and 0.90, respectively.

Meanwhile, the TLI value shown was slightly lower than 0.9 and the value of the RMSEA was higher than 0.08, but still less than 0.10. These fit indices indicated acceptable values and a good model fit for data correspondence.

Overall, the results of the analyses using the CFA model indicate that the factor loadings of each sub-scale effectively contribute to the latent variables in the model. In addition, the goodness of fit statistics indicate that the structure of the model performed a good fit with the data. This CFA model for the **Epistemological Beliefs** scale were then used for the subsequent analyses, to form a structural model, by using the SEM approach to test the multiple relationships between all of the unobserved variables (or latent variables) in the hypothesised model.

The Argumentation (AG) Scale

The CFA model for the **Argumentation** scale consisted of two sub-scales, namely Making Scientific Argumentation (MSA) and Challenging Argumentation (CA) with factor loadings of 0.56 and 0.45 respectively. These loadings indicated a good model fit. As the CFA model for this scale has the same number of parameters and observations, the findings of the CFA model were classified as just-identified. In other words, the model's degrees of freedom for the **Argumentation** scale was zero (DF = 0). According to Kline (2011), models with zero degrees of freedom do not test a particular hypothesis. Therefore, there was no calculation of goodness of fit statistics provided for this scale.

The Scientific Reasoning (SR) Scale

The CFA model for **Scientific Reasoning** scale comprised five dimensions, namely Conservation of Weight and Volume (CWV), Proportional Reasoning (PR), Control of Variables (COV), Probability and Correlation Reasoning (PCR), and Hypothetical-Deductive Reasoning (HDR). The structure of the model for the **Scientific Reasoning** scale is presented in Figure 5. 9.

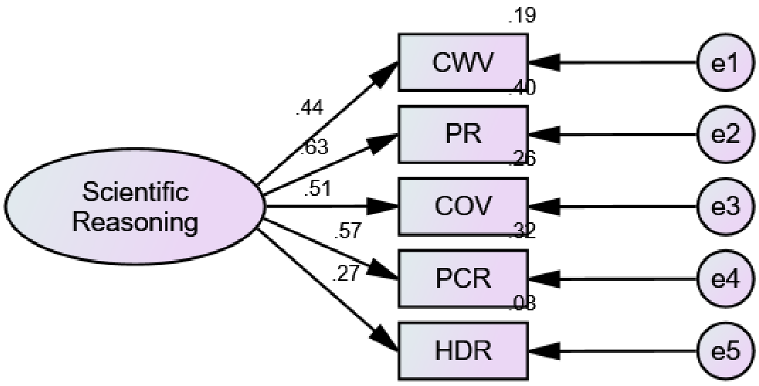


Figure 5. 9 The CFA model of Scientific Reasoning scale

Note: CWV = Conservation of Weight and Volume; PR = Proportional Reasoning; COV = Control of Variables; PCR = Probability and Correlation Reasoning; and HDR = Hypothetical-Deductive Reasoning

Figure 5. 9 shows the results of the CFA analyses that simultaneously estimated the strength of the regression path from the latent variable to each of the sub-scales by employing the same procedure. As noted by Hair et al. (2014), the minimal level of factor loadings for the interpretation of structure was 0.3, which indicates a good fit for the hypothesised model. The loading values for each dimension are presented in Table 5. 11.

Table 5. 11 Factor loadings of Scientific Reasoning scale

Scale	Sub-scale	Loading
	CWV	0.44
	PR	0.63
Scientific Reasoning (SR)	COV	0.51
	PCR	0.57
	HDR	0.27

Note: CWV = Conservation of Weight and Volume; PR = Proportional Reasoning; COV = Control of Variables; PCR = Probability and Correlation Reasoning; and HDR = Hypothetical-Deductive Reasoning

Table 5. 11 demonstrates that the value of the factor loadings ranged from 0.27 to 0.63. The HDR sub-scale had the lowest factor loading, namely 0.27. However, this loading is still considered close to the acceptable value of 0.30 applied in this study (Hair et al., 2014). Therefore, this sub-scale was retained with high caution because it might potentially affect the accuracy of interpreting results in the subsequent analysis.

Overall, all the sub-scales established a good model fit, where these dimensions represented the latent variable that was intended to be measured. In addition, to assess the goodness of fit of the model, a number of fit indices were used in this study. The summary findings of the fit indices for the **Scientific Reasoning** scale are presented in Table 5. 12.

Table 5. 12 Summaries of fit indices of Scientific Reasoning scale

Model Fit Indices	Values
χ^2/DF ratio	0.94
Goodness-of-Fit Index (GFI)	1.00
Adjusted Goodness-of-Fit Index (AGFI)	0.99
Tucker-Lewis Index (TLI)	1.00
Comparative Fit Index (CFI)	1.00
Root Mean Square Error of Approximation (RMSEA)	0.00

As shown in Table 5. 12, the ratio of the chi-square (χ^2/DF) value of this model was 0.94, which is less than 5 (the acceptable χ^2/DF value), indicating that the model provided a good fit to the data for the **Scientific Reasoning** scale. Turning to the other fit indices, the model obtained GFI, AGFI, TLI, and CFI values ranging from 0.99 to 1.00, which were greater than 0.90. These fit indices revealed that the CFA model performed a good fit. In addition, the RMSEA value was zero, indicating a good fit for the hypothesised model.

Generally speaking, the results of the analysis using the CFA model demonstrated that the factor loadings successfully contributed to each distinct dimension in the model. In addition, all goodness of fit statistics indicated a good fit to the data, which implies that these findings are relevant to the theory underlying the model in this study. The CFA model for the **Scientific Reasoning** scale were then used in the structural model to be tested with the SEM approach to investigate the multiple relationships between all the latent variables in the hypothesised model.

The Physics Conceptual Understanding (PCU) Scale

The CFA model for **Physics Conceptual Understanding** scale consisted of two sub-scales, namely the Mechanics (MECH) and the Electricity and Magnetism (EM) dimensions, each with acceptable factor loadings of 0.59 and 0.55, respectively. Since the CFA model for this scale has the same number of parameters as observations, the result of the model was classified as just-identified. In other words, the model's degrees of freedom for the **Physics Conceptual Understanding** scale was zero (DF = 0). As pointed out by Kline (2011), models with zero degrees of freedom do not test a particular hypothesis. Therefore, there was no calculation of goodness of fit statistics provided for this scale.

5.5 Descriptive and Inferential Statistics Analysis

As presented earlier, the results of the analyses using the multidimensional Rasch scaling and CFA approach were used to investigate the associations among each of the variables in the structural model using the Structural Equation Modelling (SEM) method. However, it was important to conduct a descriptive analysis prior to testing the hypothesised research model. In order to illustrate the nature of the data obtained in this quantitative study, a number of statistical calculations were performed using the IBM SPSS 25 software. The results of the descriptive analysis are presented in Table 5. 13.

It is important to note that the values listed in Table 5. 13 are not raw scores, but WLE scores (in log odd units or logits). Transforming raw scores to WLE scores was carried out using the ConQuest 4 software as described previously. In addition, the values obtained from the results of descriptive analyses for each scale cannot be compared arbitrarily because each scale has a different logit scale 0 (or zero point) as the mean of the item difficulty estimates. As highlighted by Bond and Fox (2015, p. 69), "… the mean of the item difficulties is usually adopted as the zero scale origin." This indicates that every scale of latent variables has its own mean of item difficulty. Therefore, in the case of this study, the achievements obtained by the respondents for each scale were not intended to be compared with one another.

Table 5. 13 The descriptive information for each scale

Descriptive Information	Scale (WLE Score)			
	Epistemological Beliefs	Argumentation	Scientific Reasoning	Physics Conceptual Understanding
Mean	0.140	-0.318	-0.619	-0.590
Median	0.134	-0.290	-0.717	-0.598
Mode	0.067	-0.242	-1.218	-0.805
Std. Deviation	0.142	0.337	0.555	0.331
Skewness	0.189	-0.356	0.762	0.400
Kurtosis	0.110	-0.117	0.028	0.750
Minimum Statistics	-0.274	-1.341	-1.409	-1.588
Maximum Statistics	0.623	0.482	1.372	0.925

As shown in Table 5. 13, the mean (*M*) and the standard deviation (*SD*) of the **Epistemological Beliefs** scale were 0.140 and 0.142 respectively, with a minimum score of -0.274 and a maximum score of 0.623. For the **Argumentation** scale, the mean was -0.318 (*SD* = 0.337), with a minimum score of -1.341 and a maximum score of 0.482. The mean of the **Scientific Reasoning** scale was -

0.619 (*SD* = 0.331), with a minimum score of -1.409 and a maximum score of 1.372. Meanwhile, the mean and the standard deviation of the **Physics Conceptual Understanding** scale were -0.590 and 0.331 respectively, with a minimum score of -1.588 and a maximum score of 0.925. The results showed that the mean score of the **Epistemological Beliefs** scale was above the mean of the item difficulty level (the logit scale 0), while the mean score of the **Argumentation**, **Scientific Reasoning**, and **Physics Conceptual Understanding** scales was below the mean of the item difficulty level.

As illustrated previously, the Rasch multidimensional model was employed to scale person ability and item difficulty estimates in a questionnaire on the same continuum with the average item logit centred to zero. If a respondent finds a questionnaire item is easy to answer, then a respondent who has higher ability is likely to be able to answer this item correctly. In addition, the scoring for each scale that is applied in the analyses using the Weighted Likelihood Estimates (WLE) scores, was represented in the form of logits. The WLE score may also be positive or negative, where a positive WLE score indicates that the score is above the mean and a negative WLE score indicates that it is below the mean. Hence, it can be stated that the achievement score of the Indonesian pre-service physics teachers in relation to the **Epistemological Beliefs** scale was slightly above the mean, while their achievement scores in terms of **Argumentation**, **Scientific Reasoning**, and **Physics Conceptual Understanding** scale were below the mean.

In addition, testing the normality of data distribution was needed before conducting further analyses. In the present study, the values of skewness and kurtosis were employed to check whether or not the data were normally distributed. As presented in Table 5. 13, the values of skewness for each scale, **Epistemological Beliefs**, **Argumentation**, **Scientific Reasoning**, and **Physics Conceptual Understanding** were close to zero, namely 0.189, -0.356, 0.762, and 0.400 respectively. Meanwhile, the values of kurtosis for each scale were also close to zero, namely 0.110, -0.117, 0.028, and 0.750 respectively. These values indicated that the distribution of sample scores for all scales, or latent variables, in this study was normal and could therefore be used for subsequent analyses. As proposed by Kline (2011), in normally distributed sample scores, the values of skewness and kurtosis are close to zero.

Examining the multicollinearity between the independent variables was also essential before moving to the analyses of the structural models. Multicollinearity is identified when two or more independent variables are highly correlated (Field, 2013). As documented in Chapter 4, the

Variance Inflation Factor (VIF) and Tolerance were employed in this study to detect for multicollinearity of variables by using the IBM SPSS 25 software program. Large values of VIF demonstrated a high level of multicollinearity between the independent variables (Hair et al., 2014). According to Field (2013), a VIF value of more than 10 (> 10) and a Tolerance value below 0.1 (< 0.1) indicated a serious problem with multicollinearity, and deleting the variables that were highly correlated with other independent variables was recommended. The results of the multicollinearity test using SPSS software demonstrated that none of the independent variables were detected as having a multicollinearity issue. VIF values were all less than 10, and the Tolerance values were all more than 0.1; thus, all the variables were retained (or no variables were dropped) and used in the subsequent analysis. Furthermore, inferential statistics analyses comprising the Analysis of Variance (ANOVA) and the t-test of independent samples were conducted to examine the mean differences among groups (i.e., year level, gender, and university type) for each latent variable, statistically significant at the 5% level. The ANOVA approach and t-tests were carried out using the same software, namely IBM SPSS 25.

In the present study, one-way ANOVA was employed to examine the significant differences on the epistemological beliefs, argumentation, scientific reasoning, and physics conceptual understanding of the participants between year level. The results of the one-way ANOVA, by year level for each questionnaire, is presented in Table 5. 14.

Table 5. 14 One-way analysis of variance results of significant difference on all scales by year level

Variables	Comparison	Sum of Squares	Degrees of Freedom (df)	Mean Squares	F	p-level
Epistemological Beliefs	Between Groups	0.024	3	0.008	0.402	0.752
	Within Groups	14.203	702	0.020		
	Total	14.227	705			
Argumentation	Between Groups	0.653	3	0.218	1.928	0.124
	Within Groups	79.276	702	0.113		
	Total	79.929	705			
Scientific Reasoning	Between Groups	1.533	3	0.511	1.663	0.174
	Within Groups	215.703	702	0.307		
	Total	217.236	705			
Physics Conceptual Understanding	Between Groups	3.012	3	1.004	9.520	0.000
	Within Groups	74.036	702	0.105		
	Total	77.048	705			

Note: significant at $p < 0.05$

As can be seen in Table 5. 14, there was no significant difference in the **Epistemological Beliefs** of the Indonesian pre-service physics teachers across year level, $F(3, 702) = 0.40$, $p=0.75$, the **Argumentation** variable $F(3, 702) = 1.93$, $p=0.12$, and the scale of **Scientific Reasoning**, $F(3, 702) = 1.66$, $p=0.17$. Meanwhile, there was a significant difference in the **Physics Conceptual Understanding** of the Indonesian pre-service physics teachers across year level, $F(3, 702) = 9.52$, $p=0.00$. To detect the differences which occurred between groups, a post hoc test was carried out. In addition, to calculate the strength of group difference, the effect size between groups was measured using the Cohen's d formula as described in the previous chapter. The results of the post hoc test and the Cohen's d are presented in Table 5. 15.

As shown in Table 5. 15, the post hoc analyses using the Tukey criterion, and the results of Cohen's d analysis, indicated that there was no significant difference between participants in Year 1 and Year 2 as well as participants in Year 3 and Year 4. On the other hand, there was a significant difference in the **Physics Conceptual Understanding** of pre-service teachers between Year 1 and Year 3, Year 2 and Year 3, as well as between Year 2 and Year 4 with the effect size in the small category.

Table 5. 15 Post Hoc Test (Tukey) results and Cohen's d on the Physics Conceptual Understanding variable by year level

Comparison	Mean Difference	Std. Error	p-level	Cohen's d (criteria)
Year 1 vs Year 2	-0.027	0.033	0.846	-0.08 (non-significant)
Year 1 vs Year 3	-0.139	0.033	0.000	-0.42 (small)
Year 1 vs Year 4	-0.147	0.035	0.000	-0.48 (close to medium)
Year 2 vs Year 3	-0.111	0.034	0.007	-0.33 (small)
Year 2 vs Year 4	-0.119	0.037	0.006	-0.38 (small)
Year 3 vs Year 4	0.008	0.036	0.996	-0.02 (non-significant)

Note: significant at $p < 0.05$

Meanwhile, the significant difference between the participants in Year 1 and Year 4 showed the highest effect size compared to the other groups which were in the medium category. Graphically, the mean score for each scale by year level is displayed in Figure 5. 10.

In addition, to examine the significant differences between gender (i.e., males and females) for all latent variables, the independent samples t-test was carried out. The results of the t-test and Cohen's d are presented in Table 5. 16.

Table 5. 16 t-Test results of significant differences among gender on all scales and the Cohen's d

Variables	t-value	DF	p-level (2-tailed)	Mean Difference	Std. Error Difference	Cohen's d (criteria)
Epistemological Beliefs	-0.159	704	0.874	-0.002	0.013	-0.01 (non-significant)
Argumentation	-1.941	704	0.053	-0.058	0.030	-0.17 (close to small)
Scientific Reasoning	-4.439	704	0.000	-0.216	0.049	-0.38 (small)
Physics Conceptual Understanding	-7.276	704	0.000	-0.206	0.028	-0.60 (medium)

Note: DF = Degrees of Freedom, significant at p < 0.05

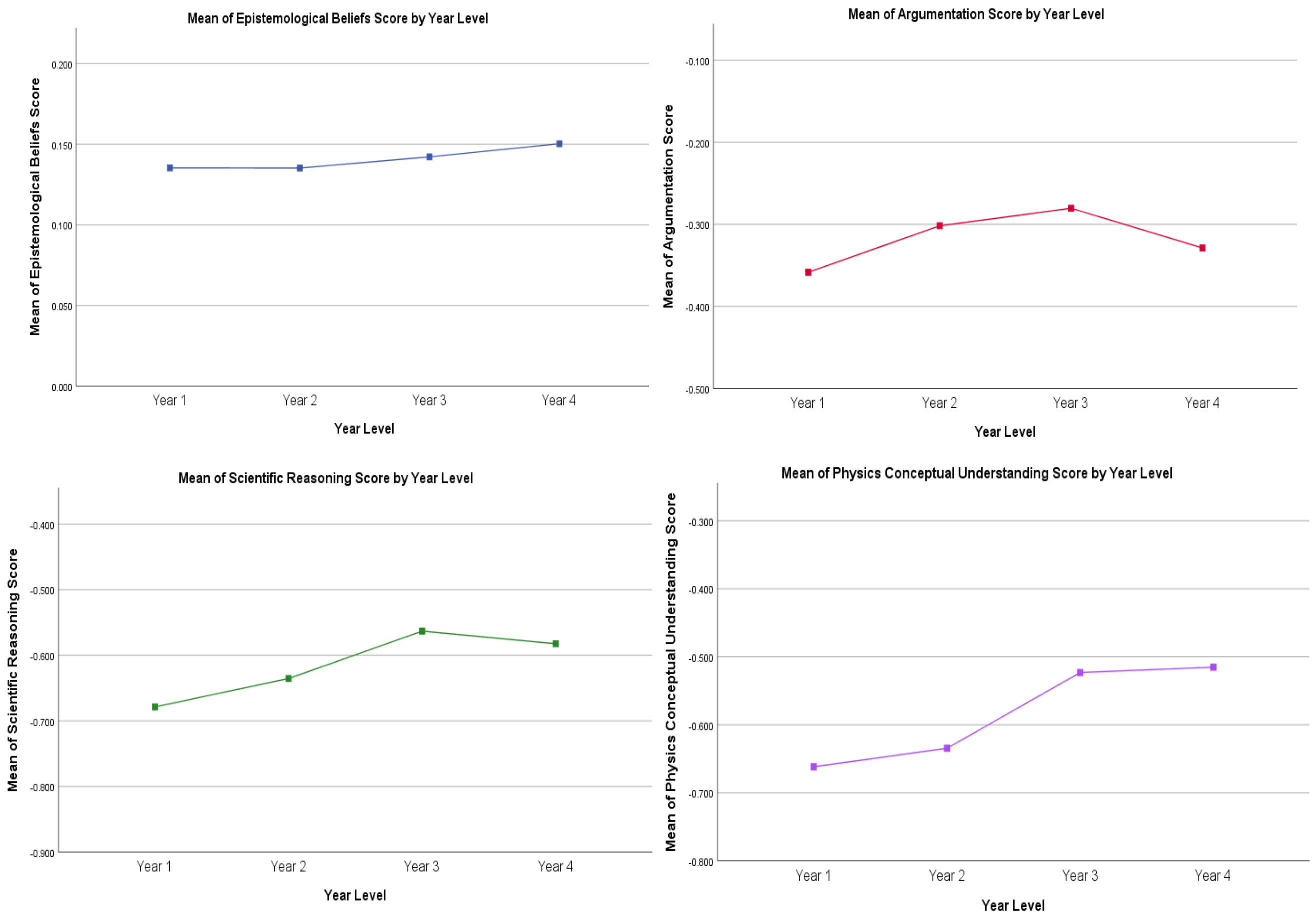


Figure 5. 10 The mean of each scale score across year level

As can be seen in Table 5. 16, the results of the independent samples t-test showed that there were no significant differences between male and female participants in terms of **Epistemological Beliefs**. A significant difference was found in terms of the **Argumentation** and **Scientific Reasoning** variables with the effect size in the small category. Meanwhile, the significant difference between the male and female participants in terms of **Physics Conceptual Understanding** showed the highest effect size compared to the other groups being in the medium category. Graphically, the mean score for each scale by gender is presented in Figure 5. 11

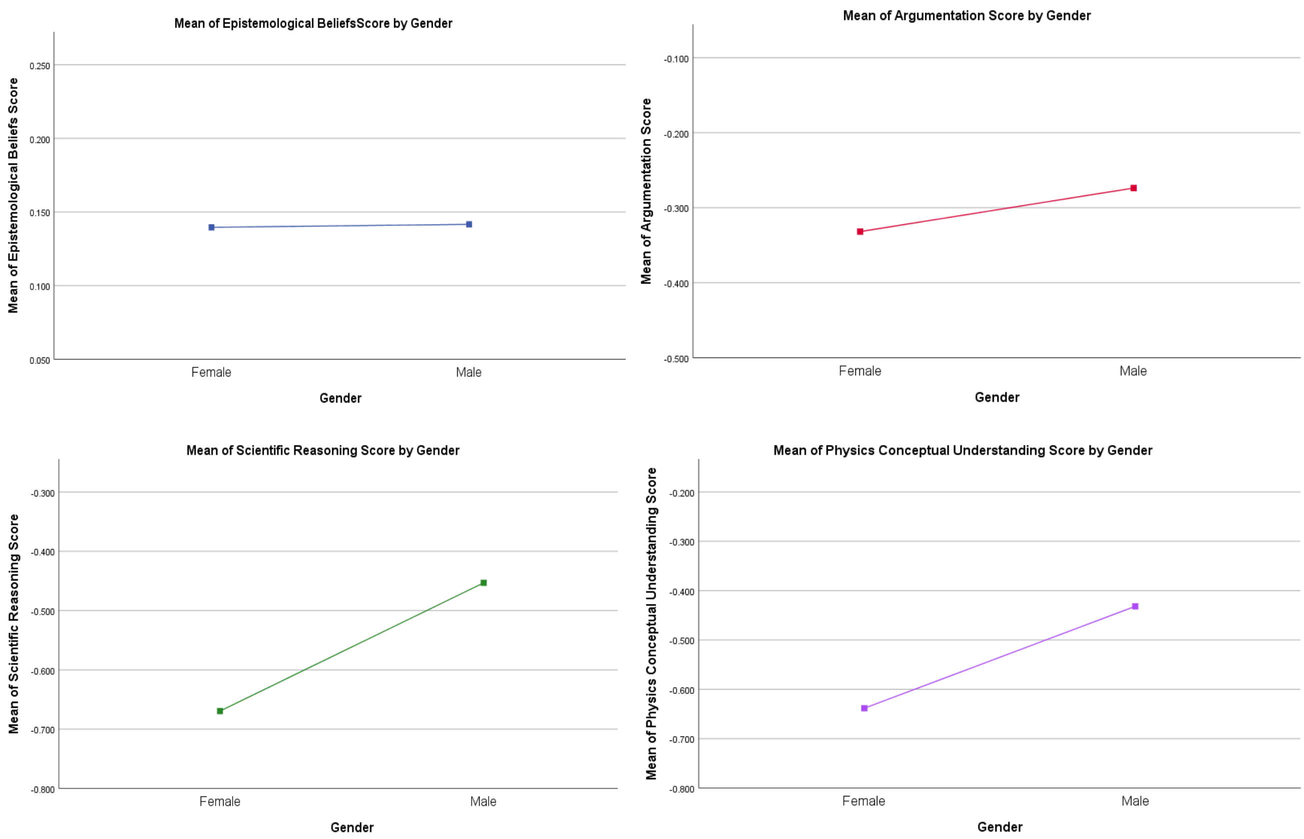


Figure 5. 11 The mean of each scale score between gender

Turning to the other group which examines significant differences between university type (i.e., private and public) for all scales in this study, the independent samples t-test was conducted. The results of the t-test and Cohen's d are given in Table 5. 17.

Table 5. 17 t-Test results of significant differences among university type on all scales and the Cohen's d

Variables	t-value	DF	p-level (2-tailed)	Mean Difference	Std. Error Difference	Cohen's d (criteria)
Epistemological Beliefs	-9.774	704	0.000	-0.099	0.010	-0.74 (medium)
Argumentation	-5.611	704	0.000	-0.140	0.025	-0.42 (small)
Scientific Reasoning	-11.058	704	0.000	-0.430	0.039	-0.85 (large)
Physics Conceptual Understanding	-10.299	704	0.000	-0.241	0.023	-0.78 (close to large)

Note: DF = Degrees of Freedom, significant at $p < 0.05$

The results of the independent samples t-test in Table 5. 17 revealed that there were significant differences between participants from private and public universities on all scales. A significant difference was found for the **Epistemological Beliefs** and **Argumentation** variables with an effect size in the medium and small category, respectively. Meanwhile, t-test results of significant differences between the type of university in terms of **Scientific Reasoning** and **Physics Conceptual Understanding** showed an effect size in the large category. Graphically, the mean score for each scale by university type is presented in Figure 5. 12.

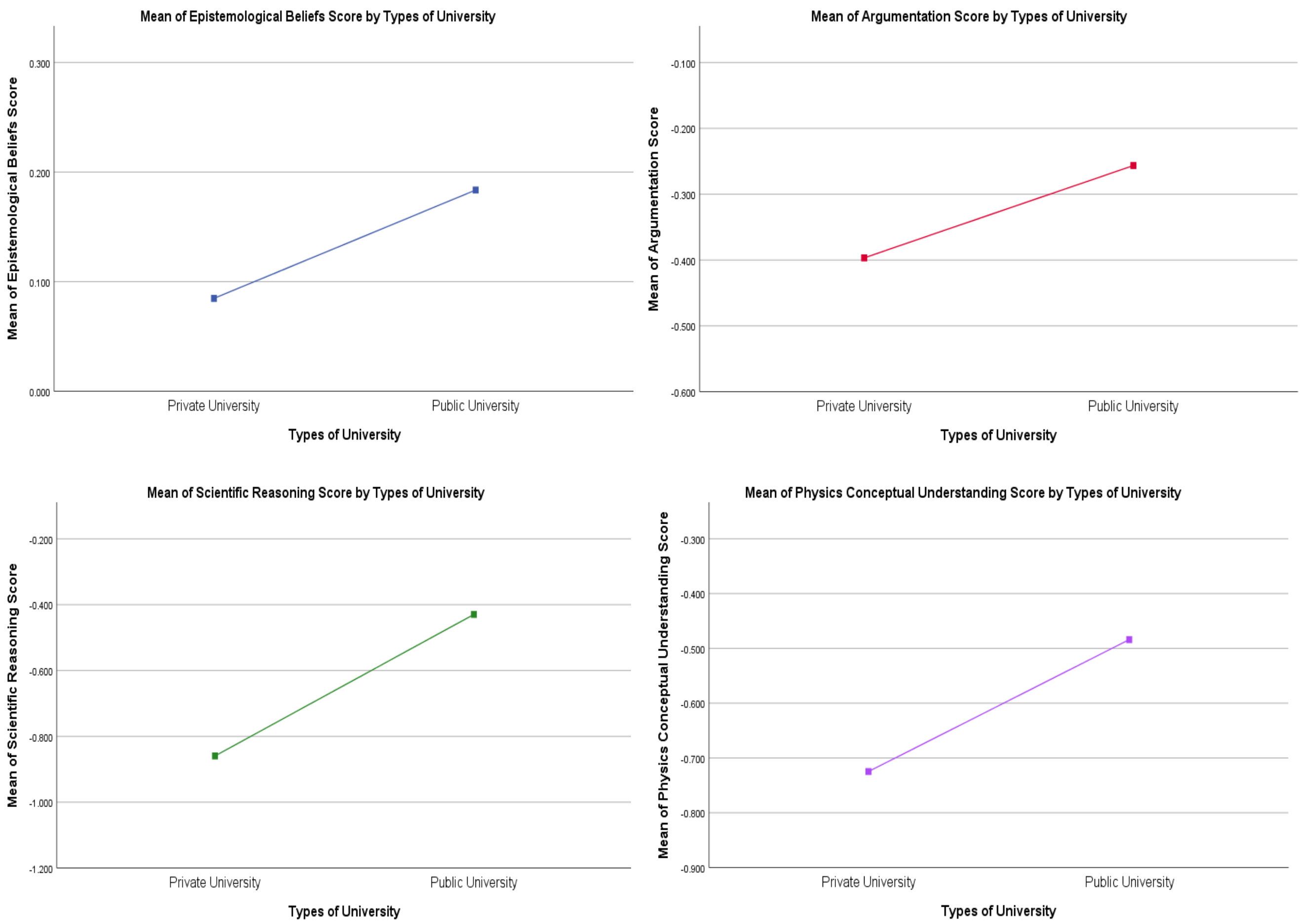


Figure 5. 12 The mean of each scale score between types of university

5.6 Summary

In this chapter, preliminary analyses and validation of instruments were carried out through the multidimensional Rasch analysis and the CFA modelling approach for each of four scales. These included the scales for measuring **Epistemological Beliefs**, **Argumentation**, **Scientific Reasoning**, and **Physics Conceptual Understanding**. Demographic information about the participants was also described prior to the analyses of the research findings. This study involved 706 Indonesian pre-service teachers coming from two private and two public universities. The number of female participants was relatively larger than the number of male participants in each academic year. All of the 706 participants involved in this study completed the five questionnaires. In addition, there were no missing values, and all data sets were used in the analyses involving several statistical software packages.

To examine the validity and reliability of the research variables, the multidimensional Rasch model analysis and Confirmatory Factor Analysis (CFA) procedures were employed by using the WLE scores instead of raw scores. The results of the Rasch analyses revealed that after removing misfitting persons and items with a discrimination index below 0.15, the person and item for each scale showed a good fit to the Rasch model. In addition, the results of the CFA procedure for each scale indicated that factor loadings successfully contributed to each distinct dimension in the model, and all goodness of fit statistics indicated a good fit to the data. This implies that the structure of the scales was relevant to the theory underlying the model examined in this study.

Furthermore, the results of the test for normality and multicollinearity revealed that the data were normally distributed and none of the independent variables were multicollinear. These steps were crucial to being tested before investigating the multiple relationships between the scales in the hypothesised model using the Structural Equation Modelling (SEM) approach, which is described in the following chapter.

CHAPTER 6
STRUCTURAL EQUATION MODELLING (SEM)

6.1 Overview

This chapter highlights the results of the proposed model analysis which was used to examine complex sets of relationships between research variables specified in the model by using the Structural Equation Modelling (SEM) approach. More specifically, this study investigates the relationships between demographic factors, the aspects of scientific thinking comprising epistemological beliefs, argumentation, and scientific reasoning, and their influence on the conceptual understanding of physics of Indonesian pre-service physics teachers. To examine the causal relationships between these variables, the IBM SPSS AMOS (Analysis of Moment Structures) software program version 25 was employed. The analysis of the SEM model was carried out based on the findings of the multidimensional Rasch analysis and scale validation using the Confirmatory Factor Analysis (CFA) procedure outlined in Chapter 5. In the current chapter, a number of variables used in the research model are presented first. This is followed by the results of the measurement models and the analyses of the structural model obtained from the hypothesised model along with an examination of the paths trimmed from the model. Last but not least, there is an outline of the final model resulting from the modelling approach, as well as a summary of the fit indices for the model. The path diagram identifies both direct and indirect effects among the research variables in the final model.

6.2 Variables Used in the SEM Model

Prior to investigating the causal relationships among variables using the modelling approach, a number of variables involved in this study are described. The variables include **Gender**, **Year Level**, **University Type**, the aspects of scientific thinking (namely **Epistemological Beliefs**, **Argumentation**, and **Scientific Reasoning**), and **Physics Conceptual Understanding** (**PCU**). A description of all the research variables employed in the model is presented in Table 6. 1. It is worth noting that the latent variables described in the SEM model have been **'bolded'** to distinguish them from the concepts related to these latent variables that are being explained.

As presented in Table 6. 1, there are four latent variables and 17 observed variables used in this study. Three of these manifest variables i.e. **Gender**, **Year Level**, and **University Type** were treated as exogenous variables, because these variables were not influenced by other variables in the model. **Gender** was coded 0 for female and 1 for male participants. With regard to the **Year Level** of the participants, coding 1 indicated participants in Year 1, coding 2 and 3 indicated participants in Years 2 and 3 respectively, and 4 indicates participants in Year 4. Meanwhile, the coding used for **University Type** was 0 for private universities and 1 for public universities.

Table 6. 1 Variables used in the research model

Latent Variables	Manifest Variables	Description	Coding
	Gender	Gender of the pre-service physics teachers	0 = Female 1 = Male
	Year Level	Year level of the pre-service physics teachers	1 = Year 1 2 = Year 2 3 = Year 3 4 = Year 4
	University Type	Types of university attended of the pre-service physics teachers	0 = Private 1 = Public
Epistemological Beliefs	SSK	Structure of Scientific Knowledge	WLE scores
	NKL	Nature of Knowing and Learning	WLE scores
	RLA	Real-Life Applicability	WLE scores
	EK	Evolving Knowledge	WLE scores
	SAL	Source of Ability to Learn	WLE scores
Argumentation	MSA	Making Scientific Argumentation	WLE scores
	CA	Challenging Argumentation	WLE scores
Scientific Reasoning	CWV	Conservation of Weight and Volume	WLE scores
	PR	Proportional Reasoning	WLE scores
	COV	Control of Variables	WLE scores
	PCR	Probability and Correlation Reasoning	WLE scores
	HDR	Hypothetical-Deductive Reasoning	WLE scores
Physics Conceptual Understanding (PCU)	MECH	Mechanics	WLE scores
	EM	Electricity and Magnetism	WLE scores

The **Epistemological Beliefs** latent variable consists of five manifest variates or dimensions, namely structure of scientific knowledge (SSK), nature of knowing and learning (NKL), real-life applicability (RLA), evolving knowledge (EK), and source of ability to learn (SAL). The **Argumentation** latent variable comprises two dimensions, namely making scientific argumentation (MSA) and challenging argumentation (CA). The **Scientific Reasoning** factor comprises five dimensions, namely conservation of weight and volume (CWV), proportional reasoning (PR), control of variables (COV), probability and correlation reasoning (PCR), and hypothetical-deductive reasoning (HDR). Meanwhile, the **Physics Conceptual Understanding** scale comprises two dimensions, which are

Mechanics (MECH) and Electricity Magnetism (EM). All of these observed variables in the model were represented by the Weighted Likelihood Estimates (WLE) scores. The WLE scores are outlined in Chapter 4. The proposed model used to investigate the relationships between each of these research variables, is described in the following section.

6.3 The Hypothesised Model

The SEM approach is useful to test a theoretical model proposed by a researcher regarding the relationships between variables which are related to some complex phenomena being studied (Khine, 2013; Schumacker & Lomax, 2010). AMOS (Analysis of Moment Structure) version 25 is a software program used in this study to test the hypotheses about the causal relationships between the research variables simultaneously by implementing the SEM procedure. As Arbuckle (2017, p. 1) stated, "IBM SPSS AMOS implements the general approach to data analysis known as Structural Equation Modelling (SEM), also known as analysis of covariance structures, or causal modelling." AMOS is an easy-to-operate program for visualising the research modelling approach by permitting the drawing of path diagrams of the analyses to be carried out. This software offers the benefit of conducting multivariate procedures and provides standardised and unstandardised parameter estimates in the output (Khine, 2013). Referring to the AMOS output, the model can be specified, and modifications could be made to improve the fit of the model.

When implementing the SEM procedure, it is important to conduct the tests for normality and multicollinearity of the empirical data prior to investigating the relationships among the research variables in the proposed model, especially when using maximum likelihood as a method of estimation in the SEM model (Schumacker & Lomax, 2010; Kline, 2011). More specifically, if the normality assumption is not fulfilled properly, it can affect the accuracy of statistical tests in the SEM model. In other words, examining a model with data not normally distributed can lead to misinterpretation of whether or not the model fits the observed data (Khine, 2013). As presented in Chapter 4, the absolute values of skewness and kurtosis of the sample responses were employed to identify the normality of the data distribution. The accepted absolute value for skewness was less than 3, while it was less than 8 for kurtosis (Kline, 2011). In addition, it is also necessary to check whether multicollinearity among the variables involved in the research model is non-existent. Multicollinearity exists when two or more variables are too highly related, resulting in statistical tests that might be biased (Khine, 2013). In this study, the Variance Inflation Factor (VIF) and

Tolerance were employed to detect for any multicollinearity issues between the independent variables from the results of the SPSS data analysis. Field (2013) argued that VIF values of more than 10 and Tolerance values below 0.1 indicate serious multicollinearity and, as such, these variables are essentially redundant, leading to problematic interpretation of the model findings. If multicollinearity exists, then one of the redundant variables must be removed from the model. As outlined in Chapter 5, none of the variables tested in this study showed skewness greater than 3 or kurtosis greater than 8. The results revealed that the data obtained were normally distributed. Hence, the maximum likelihood parameter estimation (ML) procedure was chosen as the default method in this study because the data were normally distributed (Kline, 2011). In addition, the resulting VIF values were all below 10 (none of the variables had VIF values greater than 10) and the Tolerance values were below 0.1, indicating that there were no multicollinearity issues among the variables involved in this study. Thus, all the research variables were retained and used for the subsequent analyses.

As highlighted by Khine (2013), it is also crucial to ensure that the minimum sample size requirement is met for conducting the SEM analysis. If the sample size is too small, then there will not be adequate information for estimating the parameters, especially for models consisting of a large number of research variables. Khine (2013) also stated that the minimum sample size should be "… ten participants per parameter estimated" (p.10). The more complex the model, the larger the sample size needed. In the present study, the hypothesised model was used to test the causal relationships among a number of variables by using the SEM techniques, as displayed in Figure 6. 1.

In general, the relationships between latent variables in SEM consist of direct effects, indirect or mediated effects, and covariances. The proposed model comprised 18 path coefficients (i.e., the number of relationships arising between latent variables and other latent variables), 14 factor loadings (i.e., directional effects arising between the observed variables and the latent variables), 18 error variances, and no covariances (non-directional relationships between exogenous variables). As stated by Weston and Gore Jr (2006), covariances represent correlations that are defined as non-directional relationships among exogeneous variables and are indicated by curved lines with double-headed arrows. Since the researcher did not anticipate any non-directional relationships between the latent variables, no covariance relationships were tested in the proposed model. In addition, the default setting in the AMOS program specifies one factor loading at 1 for each latent variable. To summarise, there were 50 parameters specified in the hypothesised model for estimation. Based on this, a minimum of 500 respondents were needed to test the SEM model. In this study, 706 participants were involved and there were no missing data. This means that the available sample of 706 participants was acceptable and more than met the minimum number of the required sample size for testing the SEM model.

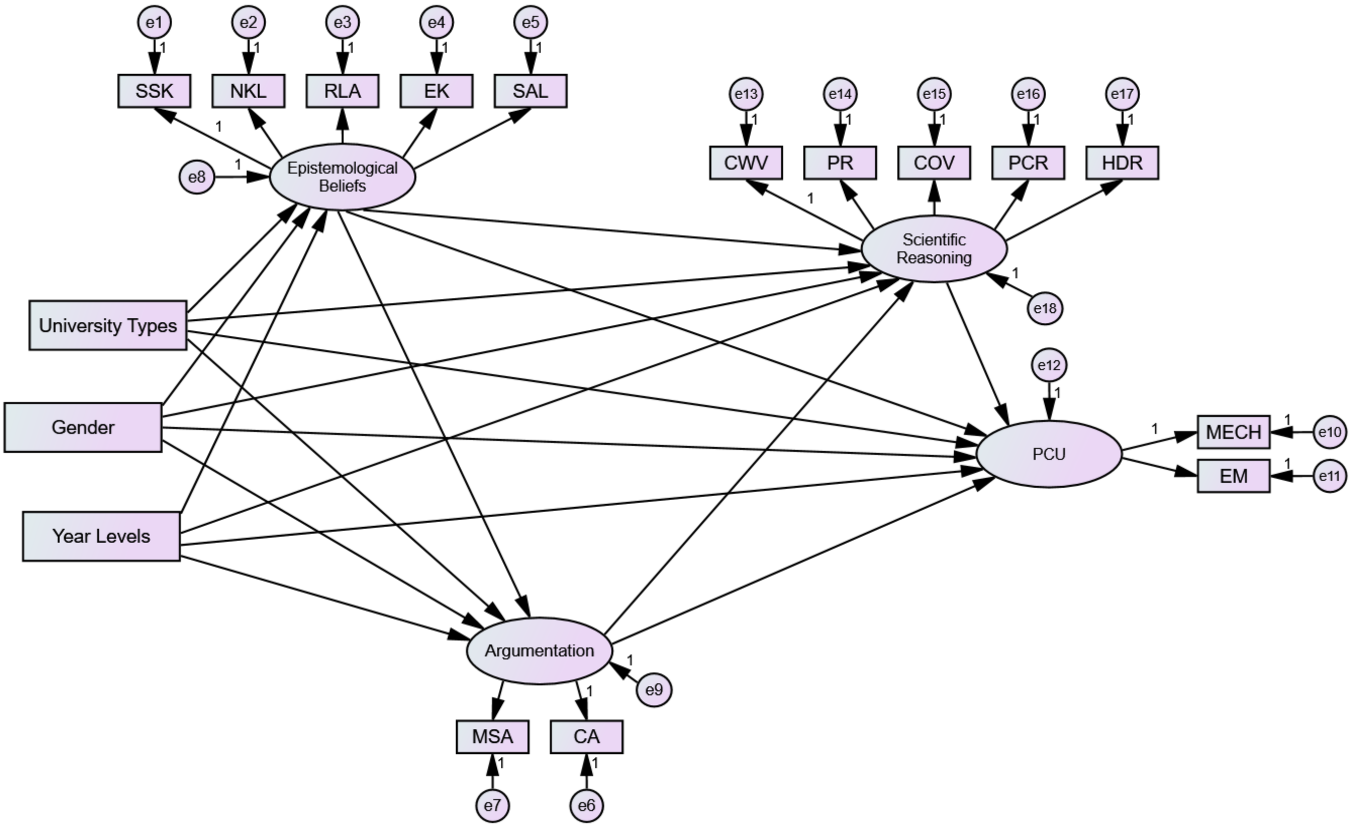


Figure 6. 1 The hypothesised model

Note:
Epistemological Beliefs scale consisted of five sub-scales, namely Structure of Scientific Knowledge (SSK), Nature of Knowing and Learning (NKL), Real-Life Applicability (RLA), Evolving Knowledge (EK), and Source of Ability to Learn (SAL).
Argumentation scale consisted of two sub-scales, namely Making Scientific Argumentation (MSA) and Challenging Argumentation (CA).
Scientific Reasoning scale consisted of five sub-scales, namely Conservation of Weight and Volume (CWV), Proportional Reasoning (PR), Control of Variables (COV), Probability and Correlation Reasoning (PCR), and Hypothetical-Deductive Reasoning (HDR).
Physics Conceptual Understanding (PCU) scale consisted of two sub-scales, namely Mechanics (MECH) and Electricity & Magnetism (EM).

As noted in Chapter 4, the SEM model is classified into two sub-models, which are the measurement model and the structural model. A structural model consists of a number of measurement models. The measurement models are used to test the strength of the relationship between observed (manifest) and unobserved (latent) variables. To examine how well the latent constructs are reflected by the manifest variates, an examination of the factor loadings is carried out for each manifest variable. As outlined in Chapter 4, factor loadings are considered significant if their values are equal to or greater than 0.3 (Hair et al., 2014). In this study, any manifest variates with loadings below 0.3 indicates that they poorly reflected the latent variables and were removed from the model. The relationships among latent variables in the model can be better interpreted if each latent variable is well reflected by its manifest variates, which also indicates how well the latent variable measures the concept (Hair et al., 2014) under consideration. Further to this, the structural model is the path model used to examine the relationships among exogenous and endogenous variables, including direct and indirect effects (Hair et al., 2014). In other words, the structural model identifies how certain variables directly (exogenous variables) or indirectly influence other certain variables (endogenous variables) in the proposed model, or how the path model connects the one latent variable to other latent variables.

This is worth noting that the causal relationships between research variables in the hypothesised model in this study were specified based on theoretical assumptions and empirical evidence. As can be seen from the model in Figure 6. 1, and as outlined in Chapter 2, the researcher hypothesised that the variable of **Epistemological Beliefs** would have a direct effect on **Argumentation**, **Scientific Reasoning**, and **Physics Conceptual Understanding**. In fact, **Epistemological Beliefs** not only has a direct effect on **Physics Conceptual Understanding**, but also an indirect effect mediated through **Argumentation** and/or **Scientific Reasoning**. In addition, the latent variable of **Argumentation** has a direct effect on both **Scientific Reasoning** and **Physics Conceptual Understanding**. In this case, **Scientific Reasoning** was specified as a mediating variable on the relationship between **Argumentation** and **Physics Conceptual Understanding**. Meanwhile, the **Scientific Reasoning** construct is predicted to have a direct effect on **Physics Conceptual Understanding**. All relationships between the research variables which were hypothesised are presented in the path diagram in Figure 6. 1.

The proposed model is used to address the research question 2 (RQ2), which is "*What are the relationships between pre-service physics teachers' scientific thinking (i.e., epistemological beliefs, argumentation, and scientific reasoning), conceptual understanding of physics, and their demographic factors?*" As noted previously, the model to be tested in this study comprised the demographic factors of the participants i.e. **Gender**, **Year Level**, and **University Type,** which were assigned as the exogenous variables. In terms of the aspects of participants' scientific thinking comprising **Epistemological Beliefs**, **Argumentation**, and **Scientific Reasoning**, they were treated as the endogenous variables in the model. These variables might influence one another or mediate the effects between variables. Because this study aims to investigate how aspects of scientific thinking as well as demographic factors influence participants' physics conceptual understanding (PCU), the **PCU** scale was also assigned as an endogenous variable (dependent latent variable) in the model. In the path diagram, "the variables that release one-way arrows are independent variables (also called exogenous variables), and those that receive arrows are dependent variables (also called endogenous variables)" (Khine, 2013, p. 25). In other words, the exogenous variables are illustrated as variables where there are no straight one-headed arrows pointing to them, but where there are one-headed arrows departing from them to the other latent variables. Meanwhile, the endogenous variables are variables that are pointed to by at least one single-headed arrow.

Through the analysis of the hypothesised SEM model, it is possible to estimate the magnitude and significance of interactions between multiple variables. In the context of this study, the extent to which the proposed model fits the observed data was tested using the AMOS program. The relationships between latent variables which were detected as having no significant paths were deleted in the model (Darmawan, 2003). As recommended by Sellin and Keeves (1977, as cited in Aldous, 2014), the standardised estimate values should be greater than 0.10. The paths with coefficients having $\beta < 0.10$ should be removed from the model because these values indicate only minimal effect in estimating the relationship between latent variables. In other words, the larger the β value (the maximum beta = 1.00), the larger the effect in the model.

Measurement Model Results

As noted previously, the measurement model as part of the SEM model is defined as the relationship between observed and unobserved variables. In the measurement model, the relationships between observed and unobserved variables were assessed using AMOS version 25. The results of the measurement model for the hypothesised model are presented in Table 6. 2.

Table 6. 2 presents a number of parameter estimates or indices used to interpret the model. The indices include the unstandardised parameter estimates (UnstdEst.), indicating the strength of the relationships between the observed and unobserved variables (latent variables) (Arbuckle, 2009). The standard error (S.E.) indicates the variability of the estimates. In addition, the critical ratio value (C.R.) is obtained by dividing the unstandardised parameter estimates by its standard error (B/S.E.),

and any critical ratio exceeding 1.96 is considered significant for p ≤ 0.05 (Khine, 2013). The p value is used to present the statistical significance of the relationships among the variables. Three asterisks (***) in the p-value column indicate significance smaller than 0.001. Meanwhile, the standardised estimates (StdEst.) or loadings in the measurement model indicate the strength of the relationship between latent variables and manifest variables. To exhibit statistical significance or good fit, the factor loadings (or the standardised estimates) of equal to or greater than 0.30 are considered practically significant at the 0.05 level (Hair et al., 2014), which provides a better indication of how well the latent variable is reflected by the manifest variables. Referring to Table 6. 2, there were no critical ratio values detected below 1.96 and a p-value more than 0.05, which means that all manifest variables were considered significant in reflecting their latent variables.

Table 6. 2 Results of measurement model in the hypothesised model

Latent Variables	Manifest Variables	UnstdEst.	S.E.	C.R.	p	StdEst. (loadings)
Epistemological Beliefs	SSK	1.00	0.00	0.00	***	0.52
	NKL	1.73	0.20	8.83	***	0.58
	RLA	0.83	0.12	6.80	***	0.37
	EK	2.74	0.36	7.59	***	0.43
	SAL	1.12	0.16	7.12	***	0.39
Argumentation	MSA	0.65	0.11	5.95	***	0.45
	CA	1.00	0.00	0.00	***	0.56
Scientific Reasoning	CWV	1.00	0.00	0.00	***	0.47
	PR	1.10	0.11	9.68	***	0.59
	COV	0.96	0.11	8.94	***	0.50
	PCR	0.99	0.10	9.66	***	0.58
	HDR	0.32	0.05	6.45	***	0.31
Physics Conceptual Understanding	MECH	1.00	0.00	0.00	***	0.60
	EM	0.97	0.08	11.65	***	0.54

Note. The symbol of three asterisks (***) indicates a p-value of < 0.001; UnstdEst. = unstandardised parameter estimates; S.E. = standard errors; C.R. = critical ratio, p-value = probability, StdEst. = standardised estimates

In addition, the **Epistemological Beliefs** scale was reflected by five manifest variables with the corresponding factor loading ranging from 0.37 to 0.58. All these loading values were greater than 0.3, indicating that all of the manifest variables contributed well to the **Epistemological Beliefs** scale. Likewise, the **Argumentation** scale was reflected by two manifest variables with the corresponding factor loading ranging from 0.45 to 0.56, indicating that all of the manifest variables contributed well to the **Argumentation** scale with the loading values being greater than 0.3. The **Scientific Reasoning** latent variable was reflected by five manifest variables with corresponding factor loadings ranging from 0.31 to 0.59. Based on the results of the CFA analysis mentioned in

Chapter 5, the HDR sub-scale in the **Scientific Reasoning** latent variable had the lowest factor loading (i.e., 0.27) when this construct was treated as an independent construct. However, this loading was close to the acceptable level for a good-fitting model (0.3). Consequently, this manifest variable was retained for subsequent analysis. When the **Scientific Reasoning** construct was included in the SEM model, as shown in Figure 6. 1, the loading of the HDR sub-scale increased slightly to 0.31. Since the minimum standard for the factor loading was 0.3, the HDR sub-scale indicated a good fit. Thus, each of the manifest variables can be considered as strong reflectors of the **Scientific Reasoning** scale. Meanwhile, the **Physics Conceptual Understanding** scale was reflected by two manifest variables with the corresponding factor loading ranging from 0.54 to 0.60, indicating that all of the manifest variables contributed well to the **Physics Conceptual Understanding** scale with the loading values being greater than 0.3. In summary, all manifest variables loaded at above 0.30, indicating a good fit for the model and that they contributed to the latent variable they were intended to measure.

Structural Model Results

In the structural model, the strength of relationships between the unobserved variables (or latent constructs), including both exogenous and endogenous variables, were assessed using the AMOS software. The results of the structural model for the hypothesised model are showed in Table 6. 3.

Table 6. 3 Results of structural model in the hypothesised model

Outcome	Predictor	B	S.E.	C.R.	p	β
Epistemological Beliefs	University Type	0.22	0.03	8.52	***	0.49
	Gender	0.02	0.02	1.01	0.31	0.05
	Year Level	0.00	0.01	0.35	0.73	0.02
Argumentation	Epistemological Beliefs	1.08	0.29	3.74	***	0.37
	Gender	0.13	0.09	1.54	0.12	0.09
	Year Level	0.03	0.03	0.93	0.35	0.05
	University Type	0.19	0.10	2.03	0.04	0.15
Scientific Reasoning	Epistemological Beliefs	0.77	0.29	2.71	0.01	0.23
	University Type	0.34	0.09	3.87	***	0.23
	Gender	0.29	0.08	3.66	***	0.17
	Argumentation	0.51	0.13	3.86	***	0.45
	Year Level	0.02	0.03	0.65	0.52	0.03
Physics Conceptual Understanding	Epistemological Beliefs	-0.14	0.21	-0.68	0.50	-0.07
	Gender	0.19	0.06	3.03	0.00	0.17
	Scientific Reasoning	0.72	0.12	6.12	***	1.12
	Argumentation	-0.20	0.11	-1.80	0.07	-0.27
	Year Level	0.07	0.02	3.43	***	0.17
	University Type	0.08	0.07	1.27	0.21	0.09

Note. The symbol of three asterisks (***) indicates a p-value of < 0.001; B = unstandardised parameter estimates; S.E. = standard errors; C.R. = critical ratio, the p-value = probability, β = standardised estimates

To interpret the structural model, a number of estimates were examined comprising unstandardised parameter estimates (B), standard errors (S.E.), critical ratio (C.R.), p-value, and standardised estimates (β). In the structural model, the standardised estimates (β) indicate the direct effect of a given latent variable on another latent variable, which is also referred to as the path coefficient (Khine, 2013). In other words, a path coefficient represents the relationship between one latent variable and another latent variable. In determining the statistical significance of the structural model, any critical ratio that exceeded 1.96 was considered significant at the p-value ≤ 0.05 level.

As highlighted in Chapter 4, the effect sizes of the path coefficients suggested by Cohen (1988, as cited in Kock, 2014) are small, medium, or large. The value of the standardised path coefficient is usually recommended as 0.02 representing a small effect, 0.15 representing a medium effect, and values larger than 0.35 being considered a large effect. However, in interpreting the results of the structural model analysis in this study, the path coefficient (the strength of relationships between the latent variables) was considered adequate if the value was equal to or greater than 0.10 (≥ 0.10), as suggested by Aldous (2014), particularly if the path was considered to be of theoretical interest. The value of the path coefficient enables the identification of latent variables which have a greater effect on the endogenous variables in the proposed model. Thus, the relationships between the latent variables detected as having non-significant paths were deleted from the model or re-specified to improve the fit of the model to the data.

The hypothesised model depicted in Figure 6. 1 presents the **Epistemological Beliefs** scale as being influenced by three exogenous variables, namely **Gender**, **Year Level,** and **University Type**. In addition, four predictor variables namely **Gender**, **Year Level**, **University Type**, and **Epistemological Beliefs** are hypothesised to influence the **Argumentation** latent variable. Meanwhile, the **Scientific Reasoning** scale is hypothesised to be directly influenced by **Gender**, **Year Level**, **University Type**, **Epistemological Beliefs**, and **Argumentation**. In the model formulated for testing, the **Physics Conceptual Understanding** scale is influenced by six latent variables, namely **Gender**, **Year Level**, **University Type**, **Epistemological Beliefs**, **Argumentation**, and **Scientific Reasoning**.

Referring to Table 6. 3, there were several paths that were found to have a critical ratio value below 1.96 and a p-value of more than 0.05, as well as a beta (β) value below 0.10 in the path diagram. The paths whose coefficients did not meet the thresholds for significance were then removed from the model. More specifically, the paths whose coefficients were found not to be significant were the relationships between **Gender** and the **Epistemological Beliefs** scale, **Year Level** and the **Epistemological Beliefs** scale, **Gender** and the **Argumentation** scale, **Year Level** and the **Argumentation** scale, **Year Level** and the **Scientific Reasoning** scale were deleted from the model. In addition, other paths deleted from the model included the relationship between **Epistemological Beliefs** and the **Physics Conceptual Understanding**, **Argumentation** and the **Physics Conceptual Understanding**, as well as types of universities attended and **Physics Conceptual Understanding**. The results of the final model analysis are described in the following section.

6.4 Final Model

The final SEM model was established to test the causal relationships between the research variables or constructs under investigation. In the hypothesised model, model trimming was carried out by removing any paths that had insignificant relationships in order to obtain the best fitting model. The final model results obtained through SEM procedures included both a measurement model and a structural model, which were generated simultaneously by AMOS version 25 as displayed in Figure 6. 2.

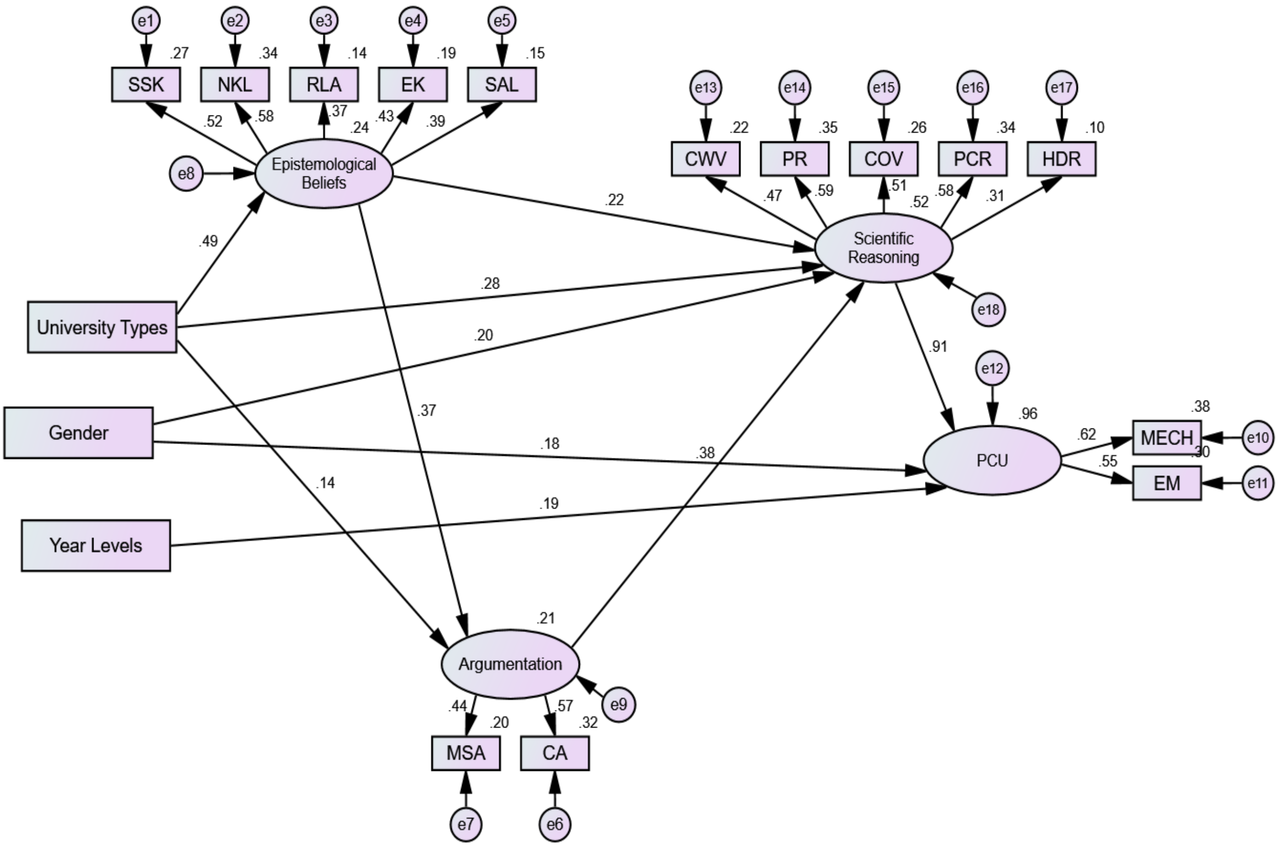


Figure 6. 2 The final model (standardised parameter estimates)

Note:
Epistemological Beliefs scale consisted of five sub-scales, namely Structure of Scientific Knowledge (SSK), Nature of Knowing and Learning (NKL), Real-Life Applicability (RLA), Evolving Knowledge (EK), and Source of Ability to Learn (SAL).
Argumentation scale consisted of two sub-scales, namely Making Scientific Argumentation (MSA) and Challenging Argumentation (CA).
Scientific Reasoning scale consisted of five sub-scales, namely Conservation of Weight and Volume (CWV), Proportional Reasoning (PR), Control of Variables (COV), Probability and Correlation Reasoning (PCR), and Hypothetical-Deductive Reasoning (HDR).
Physics Conceptual Understanding (PCU) scale consisted of two sub-scales, namely Mechanics (MECH) and Electricity & Magnetism (EM).

The path coefficient results from the SEM model can be presented as either unstandardised or standardised estimates. Unstandardised estimates are used to specify the significance of the path coefficients in SEM. This parameter estimate (B) is also preferred when comparing results of statistical tests in SEM for the same predictors across different samples or groups (Kline, 2016). However, the unstandardised path coefficients are not directly comparable across variables within the same model because they are affected by identification constraints and their variances. Meanwhile, the value of standardised estimates provides useful information regarding the strength of relationships among variables (i.e., small, medium, and large) (Yu & Shek, 2014). In other words, the value of standardised estimates is useful for making within-model comparisons. Comparing standardised estimates allows the specification of variables that have the greatest effect in the model because the unit for each variable is similar (Weston & Gore Jr, 2006). This means that the standardised coefficients place the variables on the same scale of measurement so that they are more easily interpreted (Schumacker & Lomax, 2016). Therefore, the results for standardised parameter estimates for path coefficients in the SEM model are outlined in detail in this study. In the same vein, Kline (2016) further stated that researchers should not associate the results of statistical tests in the SEM model for unstandardised estimates, but instead, with the corresponding standardized estimates.

Measurement Model Results

As highlighted by Byrne (2016), the measurement model of SEM is concerned with how well the manifest variables (observed variables) represent the underlying latent variable. Similar to the analysis of the hypothesised model described earlier, the measurement model that examines the strength of the relationship between the manifest and the latent variables was assessed based on a number of indices to interpret the final model. Five different types of indices were used to examine these relationships including the unstandardised parameter estimates (UnstdEst.), the standard errors (S.E.), the critical ratio (C.R.), the p-value, and the standardised estimates (StdEst.) or loadings. The results of the measurement model are presented in Table 6. 4.

In the measurement model, factor loadings of the manifest variables onto their latent variable in the model that are equal to or more than 0.30 and statistically significant at the 5% level are considered acceptable and interpreted as a good fit (Hair et al., 2014). Referring to Table 6. 4, there were no critical ratio values detected below 1.96 and having a p-value of more than 0.05 in the model, which means that all observed variables were considered significant in reflecting their latent variable.

Table 6. 4 Results of measurement model in the final model

Latent Variables	Manifest Variables	UnstdEst.	S.E.	C.R.	p	StdEst. (loadings)
Epistemological Beliefs	SSK	1.00	0.00	0.00	***	0.52
	NKL	1.73	0.20	8.80	***	0.58
	RLA	0.83	0.12	6.80	***	0.37
	EK	2.76	0.36	7.60	***	0.44
	SAL	1.13	0.16	7.11	***	0.39
Argumentation	MSA	0.63	0.11	5.66	***	0.44
	CA	1.00	0.00	0.00	***	0.57
Scientific Reasoning	CWV	1.00	0.00	0.00	***	0.47
	PR	1.11	0.12	9.61	***	0.59
	COV	0.97	0.11	8.88	***	0.51
	PCR	1.00	0.10	9.57	***	0.59
	HDR	0.31	0.05	6.34	***	0.31
Physics Conceptual Understanding	MECH	1.00	0.00	0.00	***	0.62
	EM	0.96	0.08	11.41	***	0.55

Note. The symbol of three asterisks (***) indicates a p-value of < 0.001; UnstdEst. = unstandardised parameter estimates; S.E. = standard errors; C.R. = critical ratio, p-value = probability, StdEst. = standardised estimates

Epistemological Beliefs

The **Epistemological Beliefs** scale was reflected by five manifest variables with the corresponding factor loadings ranging from 0.37 to 0.58. These manifest variables included the structure of scientific knowledge (SSK), nature of knowing and learning (NKL), real-life applicability (RLA), evolving knowledge (EK), and source of ability to learn (SAL), with the loading values obtained being 0.52, 0.58, 0.37, 0.44, and 0.39, respectively. All these loading values were more than 0.3, indicating that all manifest variables contributed to the **Epistemological Beliefs** latent variable.

Argumentation

The **Argumentation** scale was indicated by two manifest variables namely making scientific argumentation (MSA) and challenging argumentation (CA), with corresponding loadings of 0.44 and 0.57, respectively. These loadings also indicated that the manifest variables were strong reflectors of the **Argumentation** variable, with loading values greater than 0.3.

Scientific Reasoning

The **Scientific Reasoning** latent variable was reflected by five manifest variables, with corresponding factor loadings ranging from 0.31 to 0.59. These manifest variables included the conservation of weight and volume (CWV), proportional reasoning (PR), control of variables (COV), probability and correlation reasoning (PCR), and hypothetical-deductive reasoning (HDR). The loadings for each manifest variable were 0.47, 0.59, 0.51, 0.59, and 0.31, respectively.

Physics Conceptual Understanding (PCU)

The factor loadings of each manifest variable for the **Physics Conceptual Understanding** scale was 0.62 for the Mechanics and 0.55 for the Electricity and Magnetism dimensions. All manifest variables loaded above 0.30, indicating a good fit for the model and that they reflected the latent variable they intended to measure.

Structural Model Results

In the structural model, seven different types of indices were used to estimate the strength of the relationships between one latent variable and other latent variables, including unstandardised parameter estimates (B), standard errors (S.E.), critical ratio (C.R.), p-value, standardised estimates (β), indirect effects (ie), and total effects (te). The first five indices have been described earlier. Meanwhile, the standardised indirect effects (ie) reflect the relationship between an independent variable and a dependent variable that is mediated by one or more latent variables in the model (Khine, 2013). The size of an indirect effect can be calculated by multiplying the path coefficients of the entire association between the latent variables involved, in which these latent variables are ordered from left to right in a SEM model (Cramer, 2003). The total of indirect effects can also be found in the AMOS output. As stated by Kline (2011, p. 166) , “total effects are the sum of all direct and indirect effects of one variable on another.” In other words, the standardised total effect (te) refers to the combined direct and indirect effects of a latent variable on a dependent latent variable (endogenous variable). The results of the structural model in the final SEM model are presented in Table 6. 5.

Table 6. 5 Results of structural model in the final model

Variables		Direct Effect					Indirect Effect	Total effect
Outcome	**Predictor**	**B**	**S.E.**	**C.R.**	**p**	**β**	**ie**	**te (β+ie)**
Epistemological Beliefs	University Type	0.22	0.03	8.48	***	0.49	0.00	0.49
Argumentation	Epistemological Beliefs	1.12	0.29	3.82	***	0.37	0.00	0.37
	University Type	0.19	0.10	1.96	0.05	0.14	0.18	0.32
Scientific Reasoning	Epistemological Beliefs	0.75	0.26	2.86	0.00	0.22	0.14	0.36
	University Type	0.41	0.08	4.98	***	0.28	0.23	0.51
	Gender	0.35	0.08	4.70	***	0.20	0.00	0.20
	Argumentation	0.42	0.12	3.69	***	0.38	0.00	0.38
Physics Conceptual Understanding	Scientific Reasoning	0.60	0.06	9.50	***	0.91	0.00	0.91
	Year Level	0.08	0.02	4.45	***	0.19	0.00	0.19
	Gender	0.20	0.05	3.75	***	0.18	0.18	0.36
	University Type	-	-	-	-	-	0.46	0.46
	Epistemological Beliefs	-	-	-	-	-	0.33	0.33
	Argumentation	-	-	-	-	-	0.35	0.35

Note. The symbol of three asterisks (***) indicates a p-value of < 0.001; B = unstandardised parameter estimates; S.E. = standard errors; C.R. = critical ratio, the p-value = probability, β = standardised estimates; ie = standardised indirect effect; and te = total effects.

Similar to the hypothesised model described earlier, the demographic factors of the participants, comprising **Gender**, **Year Level**, and **University Type,** were treated as exogenous variables in the final model. In addition, the aspects of participants' scientific thinking, consisting of **Epistemological Beliefs**, **Argumentation**, and **Scientific Reasoning**, were treated as the endogenous variables in which these variables may influence another variable or mediate the effects between variables. These latent variables were reflected by different manifest variables, as mentioned earlier in Table 6. 1. All of the manifest variables are endogenous as they are predicted by their respective latent variables (Weston & Gore Jr, 2006). Meanwhile, the **Physics Conceptual Understanding** scale was also assigned as an endogenous variable (dependent latent variable) which was reflected by its manifest variables. The inclusion of all of the research variables in the final model indicates that none of the variables were removed from the initial model after model trimming was carried out by examining the critical ratios and p-values for significance, as well as the effect size of the path coefficients. In addition, all path coefficients (straight arrows) were detected as having a critical ratio values exceeding 1.96 with p-values equal to or less than 0.05 (≤ 0.05), referring to Table 6. 5. All beta (β) values were also above 0.10. This indicates that all path coefficients were considered significant, and then used to interpret the trend of the relationships among variables in the final model in order to answer the research questions.

The standardised estimate results for the final SEM model in Figure 6. 2 show that the antecedent variables of **Gender**, **Year Level**, and **University Type** were not influenced by other latent variables within the model. Therefore, the following explanation centres on the four latent variables in the model, namely: **Epistemological Beliefs**, **Argumentation**, **Scientific Reasoning**, and **Physics Conceptual Understanding** (**PCU**). As mentioned earlier, the coding used for **Gender** was 0 representing female participants and 1 representing male participants. Hence, a positive value of the path coefficient for **Gender** indicates that males are performing better than females on a given measure, while a negative value indicates the reverse. Similarly, the coding used for the **University Type** was 0 representing private universities and 1 representing public universities. In a similar vein, a positive value path coefficient for **University Type** indicates that participants from public universities are performing higher than participants from private universities on a particular measure, while a negative value path coefficient indicates that participants from private universities are performing higher than participants from public universities. With regard to the **Year Level**, coding 1 represents participants in Year 1, coding 2 and 3 is used for participants in Years 2 and 3 respectively, and the code of 4 is used for participants in Year 4. Therefore, a positive value of path coefficient for **Year Level** indicates that participants in the higher **Year Level** are performing better than the participants in the lower **Year Level** and vice versa.

Epistemological Beliefs

The latent variable of **Epistemological Beliefs** was hypothesised to be influenced by three exogenous variables (predictor variables), namely **Gender**, **Year Level**, and **University Type**. However, the results of the SEM analysis presented in Figure 6. 2 and Table 6. 5 showed that only the variable for **Epistemological Beliefs** is directly influenced by the **University Type** having a path coefficient (β) of 0.49 (University Type → Epistemological Beliefs), while **Gender** and **Year Level** did not have an influence or contribute to **Epistemological Beliefs**. The significance of this path coefficient can be examined using the unstandardised output, which indicated that the unstandardised coefficient (the B weight) was 0.22 with a standard error of 0.03. To determine whether the coefficient was significant (i.e., C.R. $\geq$ 1.96 for $p \leq 0.05$), the unstandardised coefficient was divided by its standard error. However, the critical ratio (C.R.) score was automatically calculated and provided with output in the AMOS software program. As can be seen in Table 6. 5, the C.R. value was 8.48 which is greater than 1.96 at $p \leq 0.05$, indicating that the parameter was significant. The resulting path coefficients showed a strong positive relationship between the **University Type** and the **Epistemological Beliefs** scale. Since **University Type** is coded as 1 for public universities and 0 for private universities, the positive sign could be interpreted to mean that participants from the public universities were more likely to score better than participants from the private universities on the scale of **Epistemological Beliefs**, namely in terms of the structure of scientific knowledge (SSK), nature of knowing and learning (NKL), real-life applicability (RLA), evolving knowledge (EK), and source of ability to learn (SAL). Thus, participants from public universities tend to have more sophisticated levels of epistemological beliefs than participants from private universities.

Argumentation

In the model formulated for testing, four predictor variables namely **Gender**, **Year Level**, **University Type**, and **Epistemological Beliefs** were hypothesised to influence the **Argumentation** construct. The results of the model analysis presented in Figure 6. 2 and Table 6. 5 showed that there were only two factors that were found to have effects on the **Argumentation** scale. These included **Epistemological Beliefs** that was found to have a significant direct effect on **Argumentation**. Meanwhile, **University Type** was found to have both a direct and an indirect effect on the **Argumentation** construct. In addition, no significant effect was found either directly or indirectly on the **Argumentation** scale with respect to **Gender** or **Year Level** of participants.

The direct effect of **Epistemological Beliefs** on **Argumentation** was $\beta = 0.37$ (Epistemological Beliefs → Argumentation). This positive path coefficient indicated that the higher the level of participants' epistemological beliefs, the more likely they were to become more skilled in argumentation, namely in terms of making scientific argumentation (MSA) and challenging argumentation (CA). In addition, the direct effect of **University Type** on **Argumentation** was $\beta = 0.14$ (University Type →

Argumentation). The type of university attended as an exogenous variable in the model was also found to have an indirect effect (ie = 0.18) on the **Argumentation** scale, which was mediated by **Epistemological Beliefs** (University Type → Epistemological Beliefs → Argumentation). Combining both direct and indirect effects of the relationship between **University Type** and **Argumentation** generated a total effect (β+ie = 0.14+0.18) of 0.32. The positive sign indicated that participants who are currently attending public universities were more likely to perform better than participants who were studying at private universities with respect to the measure of **Argumentation**.

Table 6. 5, furthermore, indicates that the values of critical ratio (C.R.) for the direct effect between **University Type** and **Argumentation** as well as between **Epistemological Beliefs** and **Argumentation** are 1.96 and 3.82 respectively, which are equal to or greater than 1.96 ($p \leq 0.05$), indicating that these parameter estimates are significant. Thus, it could be concluded that participants from public universities tend to have a higher level of epistemological beliefs which enables them to be more skilled in argumentation compared with participants from private universities with more naive levels of epistemological beliefs. In turn, the epistemological beliefs of participants in private universities have less impact on increasing their argumentation skills.

Scientific Reasoning

The **Scientific Reasoning** scale was hypothesised to be directly influenced by five factors, namely **Gender**, **Year Level**, **University Type**, **Epistemological Beliefs**, and **Argumentation**. The results of the model analysis presented in Figure 6. 2 and Table 6. 5 showed that four factors were found to have effects on the **Scientific Reasoning** latent variable. These included **Gender** and **Argumentation** that were found to have a direct effect on **Scientific Reasoning**. Meanwhile, **Epistemological Beliefs** and **University Type** were found to have both a direct and an indirect effect on the **Scientific Reasoning** construct. In addition, the direct effect of **Year Level** on the measure of **Scientific Reasoning** was shown to be negligible.

Scientific Reasoning was directly influenced by **Gender** with a path coefficient (β) of 0.20 (Gender → Scientific Reasoning) and there were no indirect effects (ie = 0) to this construct. Since **Gender** was coded 1 for male participants and 0 for female participants, the positive path coefficient indicated that there were significant differences between male and female participants with regard to their **Scientific Reasoning**. Male participants tended to perform better compared to female participants in **Scientific Reasoning**. As shown in Figure 6. 2, **Argumentation** was also found to have a strong direct effect (β = 0.38) and there were no indirect effects (ie = 0) in operation on **Scientific Reasoning** (Argumentation → Scientific Reasoning). The results indicated a strong relationship between the constructs of **Argumentation** and **Scientific Reasoning**. The positive path coefficient showed that participants who were more skilled in argumentation were more likely to be more skilled in reasoning scientifically than the reverse situation.

In addition, as depicted in the path diagram in Figure 6. 2, **Epistemological Beliefs** had a direct effect (β = 0.22) on **Scientific Reasoning** (Epistemological Beliefs → Scientific Reasoning). The positive path coefficient indicated that participants with more sophisticated epistemological beliefs tended to have higher scientific reasoning skills than participants with lower levels of epistemological beliefs. This implied that the lower levels of epistemological beliefs of the participants tended to contribute less to developing their skills in scientific reasoning. **Epistemological Beliefs** were also found to have an indirect effect (ie = 0.14) on **Scientific Reasoning**, which was partially mediated by the **Argumentation** latent variable (Epistemological Beliefs → Argumentation → Scientific Reasoning). Combining both direct and indirect effects generated a total effect (β+ie = 0.22+0.14) of 0.36. The positive path coefficient indicated that participants with more sophisticated levels of epistemological beliefs were more likely to be more skilled in argumentation, which in turn influenced their scientific reasoning more positively than participants with lower levels of epistemological beliefs. Naive epistemological beliefs tend to contribute less to the improvement of participants' skills in argumentation which, in turn, has little impact on improving their scientific reasoning skills.

With regard to the **University Type** factor, there are four paths that can be outlined in the final model from the **University Type** to **Scientific Reasoning**. First, **Scientific Reasoning** was directly influenced by **University Type** (University Type → Scientific Reasoning) with a path coefficient (β) of 0.28. The positive sign indicated that participants from public universities tended to have higher scientific reasoning skills than participants from private universities, namely in terms of conservation of weight and volume (CWV), proportional reasoning (PR), control of variables (COV), probability and correlation reasoning (PCR), and hypothetical-deductive reasoning (HDR). Second, **University Type** as an exogenous variable was also found to have an indirect effect on **Scientific Reasoning**, which was mediated by the **Epistemological Beliefs** construct (University Type →

Epistemological Beliefs → Scientific Reasoning). As described previously, the size of the indirect effect was calculated by multiplying the path coefficient between **University Type** and **Epistemological Beliefs** by the path coefficient between **Epistemological Beliefs** and **Scientific Reasoning**. This gave the size of the indirect effect as 0.11 (0.49 x 0.22 = 0.11). Thus, based on the results, it can be concluded that participants who currently attend public universities tend to have sophisticated epistemological beliefs which enables them to be more skilled in scientific reasoning compared to participants studying at private universities that tend to have naive epistemological beliefs. In turn, their epistemological beliefs had less impact on increasing their scientific reasoning skills. In the third path, **University Type** was found to have an indirect effect on **Scientific Reasoning**, which was mediated by **Argumentation** (University Type → Argumentation → Scientific Reasoning). The size of the indirect effect was calculated by multiplying the path coefficient between **University Type** and **Argumentation** by the path coefficient between **Argumentation** and **Scientific Reasoning**. This gave the size of the indirect effect at 0.05 (0.14 x 0.38 = 0.05). Thus, from the results detailed in Table 6. 5 and Figure 6. 2, it can be concluded that participants attending public universities tended to be more skilled in argumentation than participants studying at private universities. In turn, participants from public universities tended to have more skills to reason scientifically. Lastly, **University Type** was also found to have an indirect effect on **Scientific Reasoning**, mediated by two constructs i.e., **Epistemological Beliefs** and **Argumentation** (University Type → Epistemological Beliefs → Argumentation → Scientific Reasoning). The size of the indirect effect was calculated by multiplying the path coefficient between **University Type** and **Epistemological Beliefs** by the path coefficient between **Epistemological Beliefs** and **Argumentation**, and by the path coefficient between **Argumentation** and **Scientific Reasoning**. This gave the size of the indirect effect at 0.07 (0.49 x 0.37 x 0.38 = 0.07). Thus, the total size of the indirect association between **University Type** and **Scientific Reasoning** was 0.23 (0.11 + 0.05 + 0.07 = 0.23), as can be seen in Table 6. 5.

It can be stated that both direct and indirect effects of the relationship between **University Type** and **Scientific Reasoning** generated a large effect size of the total effect (β+ie = 0.28+0.23) of 0.51. Furthermore, Table 6. 5, indicated that the values of the critical ratio (C.R.) for the direct effect between **Gender**, **University Type**, **Epistemological Beliefs**, and **Argumentation** on the measure of **Scientific Reasoning** were 4.70, 4.98, 2.86, and 3.69 respectively, which were all greater than 1.96 ($p \leq 0.05$), indicating that these parameters were significant. Based on the results, it can be concluded that participants attending public universities tended to have more sophisticated epistemological beliefs which enabled them to be more skilled in argumentation. Therefore, they might reasonably be expected to perform better on scientific reasoning than participants studying at private universities with naive epistemological beliefs. As mentioned previously, participants with low levels of epistemological beliefs tended to contribute less to their argumentation skills which, in turn, had little impact on increasing their scientific reasoning skills.

Physics Conceptual Understanding (PCU)

The final criterion in the SEM model was **Physics Conceptual Understanding** (**PCU**). The latent variable of **PCU** was hypothesised to be directly influenced by six factors in the model, namely **Gender**, **Year Level**, **University Type**, **Epistemological Beliefs**, **Argumentation**, and **Scientific Reasoning**. However, the results of the final SEM model analysis presented in Figure 6. 2 and Table 6. 5 showed that **Year Level** and **Scientific Reasoning** were found to only have a direct effect on **PCU**. Meanwhile, **Gender** was found to have both a direct and indirect effect on the **PCU** construct. In addition, **Argumentation**, **Epistemological Beliefs**, and **University Type** were found to only have an indirect effect on **PCU**.

As presented in the final SEM model and Table 6. 5, **PCU** was directly influenced by **Year Level** (Year Level → PCU) with a path coefficient (β) of 0.19, and there were no indirect effects (ie = 0) in operation on **PCU**. The positive sign indicated that participants at the higher year level were more likely to perform better compared to participants at the lower year level on the **PCU** construct. In addition, the **Scientific Reasoning** factor was found to strongly influence the **Physics Conceptual Understanding** of the participants (Scientific Reasoning → PCU) with a path coefficient (β) of 0.91 and there was no indirect effect (ie = 0), as can be seen in Table 6. 5. The resulting path coefficient showed a strong positive relationship between the **Scientific Reasoning** and the **PCU** factors. This can be interpreted to mean that the participants tended to perform better in mastering the conceptual understanding of physics (in terms of Mechanics and Electricity & Magnetism topics) if they had a higher level of scientific reasoning skills than the reverse situation. In other words, the higher the participants' skills in being able to reason scientifically, the more likely they were to become better in understanding physics concepts.

Figure 6. 2 and Table 6. 5 also indicated that **Gender** had both a direct effect (β = 0.18) and an indirect effect (ie = 0.18) on **PCU**, which was partially mediated by the **Scientific Reasoning** latent variable (Gender → Scientific Reasoning → PCU). Combining both direct and indirect effects generated a total effect (β+ie = 0.18+0.18) of 0.36. The positive sign indicated that male participants were more likely to perform better on measures of **PCU** than female participants, namely in terms of the Mechanics and Electricity & Magnetism topics. Furthermore, this might be explained by the fact that male participants with high levels of scientific reasoning tended to understand the physics concepts better than female participants who were more likely to have lower levels of scientific reasoning, and therefore, might reasonably have low levels of conceptual understanding of physics as well.

Referring to the variables presented in Table 6. 5, it is interesting to note that **Argumentation**, **Epistemological Beliefs**, and **University Type** were found to have indirect effects on the **PCU** factor. The **Argumentation** construct was found to have an indirect effect (ie = 0.35) on **PCU**, which was mediated by **Scientific Reasoning** (Argumentation → Scientific Reasoning → PCU). Meanwhile, **Epistemological Beliefs** was found to have an indirect effect on **PCU**, which was mediated by **Scientific Reasoning** (Epistemological Beliefs → Scientific Reasoning → PCU). The size of the indirect effect was calculated by multiplying the path coefficient between **Epistemological Beliefs** and **Scientific Reasoning** by the path coefficient between **Scientific Reasoning** and **PCU**. This gave the size of the indirect effect at 0.20 (0.22 x 0.91 = 0.20). As depicted in the path diagram shown in Figure 6. 2, **Epistemological Beliefs** was also found to have an indirect effect on **PCU**, which was mediated by two constructs, namely **Argumentation** and **Scientific Reasoning** (Epistemological Beliefs → Argumentation → Scientific Reasoning → PCU). The size of the indirect effect was calculated by multiplying the path coefficient between **Epistemological Beliefs** and **Argumentation** by the path coefficient between **Argumentation** and **Scientific Reasoning**, and by the path coefficient between **Scientific Reasoning** and **PCU**. This gave the size of the indirect effect at 0.13 (0.37 x 0.38 x 0.91 = 0.13). Thus, the total size of the indirect association between **Epistemological Beliefs** and **PCU** was 0.33 (0.20 + 0.13 = 0.33), as can also be seen in Table 6. 5.

In terms of **University Type**, there are four paths that can be traced in the final SEM model from **University Type** to the **PCU** scale. Firstly, **University Type** was found to have an indirect effect on **PCU**, which was mediated by **Scientific Reasoning** (University Type → Scientific Reasoning → PCU).

The size of the indirect effect was calculated by multiplying the path coefficient between **University Type** and **Scientific Reasoning** by the path coefficient between **Scientific Reasoning** and **PCU**. This gave the size of the indirect effect at 0.25 (0.28 x 0.91 = 0.25). Thus, it could be interpreted that participants from public universities tended to have higher scientific reasoning skills. Therefore, they might reasonably be expected to score more highly in conceptual understanding of physics than participants from private universities who were more likely to have fewer skills in scientific reasoning. Secondly, **University Type** was found to have an indirect effect on **PCU**, which was mediated by two constructs (i.e., **Epistemological Beliefs** and **Scientific Reasoning**), where the indirect effect of **Epistemological Beliefs** on **PCU** was mediated by **Scientific Reasoning** (University Type → Epistemological Beliefs → Scientific Reasoning → PCU). The size of the indirect effect was calculated by multiplying the path coefficient between **University Type** and **Epistemological Beliefs** by the path coefficient between **Epistemological Beliefs** and **Scientific Reasoning**, and by the path coefficient between **Scientific Reasoning** and **PCU**. This gave the size of the indirect effect at 0.10 (0.49 x 0.22 x 0.91 = 0.10). Thus, it can be concluded that participants from public universities tend to have more sophisticated epistemological beliefs which strongly influence their scientific reasoning. Therefore, they might be expected to be able to understand physics concepts better than participants from private universities who are more likely to have naive epistemological beliefs that might not impact on increasing their scientific reasoning. Thirdly, the impact of **University Type** on **Physics Conceptual Understanding** was also mediated by two constructs i.e., **Argumentation** and **Scientific Reasoning**, where the impact of **Argumentation** on **PCU** was fully mediated by **Scientific Reasoning** because only an indirect effect was specified (University Type → Argumentation → Scientific Reasoning → PCU). The size of the indirect effect was calculated by multiplying the path coefficient between **University Type** and **Argumentation** by the path coefficient between **Argumentation** and **Scientific Reasoning**, and by the path coefficient between **Scientific Reasoning** and **PCU**. This gave the size of the indirect effect at 0.05 (0.14 x 0.38 x 0.91 = 0.05). Thus, from the results detailed in Table 6. 5 and Figure 6. 2, it can be concluded that participants who attend public universities tend to be more skilled in argumentation which is more likely to significantly influence their scientific reasoning. This leads to participants having a better understanding of physics concepts compared to participants from private universities who seem to have lower argumentation skills which, in turn, have less impact on increasing their skills in reasoning scientifically. Finally, **University Type** was also found to have an indirect effect on **PCU**, which was mediated by three constructs, namely **Epistemological Beliefs**, **Argumentation**, and

Scientific Reasoning (University Type → Epistemological Beliefs → Argumentation → Scientific Reasoning → PCU). The size of the indirect effect was calculated by multiplying the path coefficient between **University Type** and **Epistemological Beliefs** by the path coefficient between **Epistemological Beliefs** and **Argumentation**, and by the path coefficient between **Argumentation** and **Scientific Reasoning**, as well as by the path coefficient between **Scientific Reasoning** and **PCU**. This gave the size of the indirect effect at 0.06 (0.49 x 0.37 x 0.38 x 0.91 = 0.06). Thus, the total size of the indirect association between **University Type** and **PCU** was 0.46 (0.25 + 0.10+ 0.05 + 0.06 = 0.46), as can also be seen in Table 6. 5.

Furthermore, Table 6. 5, indicates that the value of the critical ratio (C.R.) for the direct effect between **Gender**, **Year Level,** and **Scientific Reasoning** on the **PCU** scale are 3.75, 4.45, and 9.50 respectively, which are greater than 1.96 ($p \leq 0.05$), indicating that these parameters are significant. Thus, these results seem to indicate that the participants currently attending public universities are more likely to have higher levels of epistemological beliefs, which enables them to have better argumentation skills. In turn, this influences the way they reason scientifically. Therefore, they might reasonably be expected to become better in understanding physics concepts compared to the reverse situation. In short, participants' **Epistemological Beliefs**, **Argumentation**, and **Scientific Reasoning** are found to have significant effects on the participants' **Physics Conceptual Understanding** either indirectly or directly. Based on Table 6. 5, it can be concluded that the type of university attended by the participants has a great impact on all aspects of their scientific thinking (consisting of epistemological beliefs, argumentation, and scientific reasoning). This, in turn, strongly influenced their ability to understand physics concepts.

Overall, it can be stated that the higher the levels of participants' scientific thinking (in terms of epistemological beliefs, argumentation, and scientific reasoning), the more likely they were to become better in understanding physics concepts. However, the results of the SEM model analysis in this study showed that being in a higher year level did not indicate that the participants would have a higher level of scientific thinking skills. Hence, as described previously, the year level of the participants only had a small effect on their conceptual understanding of physics.

6.5 Model Fit Indices Summary

A number of fit indices were employed to assess the model fit in this study in order to examine how well the final model fitted the observed data. A selection of fit indices produced by the AMOS program are presented in Table 6. 6. AMOS output in Table 6. 6 shows that the value of the ratio of chi-square (χ^2/DF), calculated by dividing the chi-square by its degrees of freedom was 1.65 (DF = 113), with a corresponding p-value of $p < 0.01$. The results of the χ^2/DF value was less than 5 (the acceptable χ^2/DF value), indicating that the structure of the final SEM model provided a good fit to the observed data, as suggested by Kline (2011).

Table 6. 6 Summaries of fit indices of final model

Model Fit Indices	Values
χ^2/DF ratio	1.65
Goodness-of-Fit Index (GFI)	0.97
Adjusted Goodness-of-Fit Index (AGFI)	0.96
Tucker-Lewis Index (TLI)	0.94
Comparative Fit Index (CFI)	0.95
Root Mean Square Error of Approximation (RMSEA)	0.03

With regard to the other model fit indices, the model obtained GFI, AGFI, TLI, and CFI values ranging from 0.94 to 0.97 which are greater than 0.90. As noted by Hair et al. (2014), the values for these fit indices (i.e., the GFI, AGFI, TLI, and CFI) theoretically range from 0 (poor fit) to 1 (good fit). The index of the GFI (Goodness-of-Fit Index) was 0.97, AGFI (Adjusted Goodness-of-Fit Index) was 0.96, TLI (Tucker-Lewis Index) was 0.94, and CFI (Comparative Fit Index) was 0.95. As a value close to 0.90 is considered as a minimum cut-off level for model acceptance, all fit indices in the current study provided satisfactory values. Thus, the final SEM model in this study fits the data well. To reflect the fit of the final proposed model, the index of the RMSEA was also examined using the AMOS program in which the cut-off criterion for RMSEA value is ≤ 0.08 or not more than 0.10. Byrne (2010) noted that the RMSEA is a fit index which is highly informative in measuring the model specification. The RMSEA is relatively independent of sample size, even though it is affected by the complexity of the model (i.e., degrees of freedom) (Schumacker & Lomax, 2016). In the final model, the RMSEA value obtained was 0.03, indicating a good model fit. Referring to the fit indices that were provided by the AMOS output, it can be concluded that all goodness of fit statistics values indicated a good fit between the final model and the observed data.

6.6 Summary

This chapter presents the results of the model analysis examining the possible relationships between the research variables specified in this study. The variables consisted of **Gender**, **Year Level**, and **University Type** which were treated as exogenous variables. The other variables were the aspects of scientific thinking including **Epistemological Beliefs**, **Argumentation**, and **Scientific Reasoning** which were treated as endogenous variables. The last variable was **Physics Conceptual Understanding** (**PCU**) which was also treated as an endogenous or dependent latent variable in the model. The relationships between the variables were explored based on research question 2 (RQ2) presented in Chapter 1, namely: "What are the relationships between pre-service physics teachers' scientific thinking (i.e., epistemological beliefs, argumentation, and scientific reasoning), conceptual understanding of physics, and their demographic factors?" In addition, the test for normality and multicollinearity of the observed data was carried out prior to investigating these causal relationships. The results revealed that the data obtained in this study were normally distributed and there were no multicollinearity issues among the variables. Therefore, the maximum likelihood parameter estimation (ML) was chosen as a default method in the data analysis using the AMOS program version 25.

The Structural Equation Modelling (SEM) procedure, characterised as the measurement model and structural model, was used in this study to examine the strength of the relationships between manifest variables and their corresponding latent variables as well as the relationships between one latent variable and other latent variables in the proposed model. To evaluate how well the overall model fits the observed data, a variety of fit indices were used. The multiple criteria involved the chi-square ratio test, Goodness-of-Fit Index (GFI), Adjusted Goodness-of-Fit Index (AGFI), Tucker-Lewis Index (TLI), Comparative-Fit Index (CFI), and Root Mean Square Error of Approximation (RMSEA). The results of the hypothesised model analysis indicated that all manifest variables reflected the related latent variable that they intended to measure. In addition, model trimming was carried out by removing any paths in the hypothesised model that had insignificant relationships in order to generate the best final model.

Similar to the hypothesised model, all manifest variables obtained acceptable loadings (loaded above 0.30) in the final model, indicating that the manifest variables contributed to each corresponding latent variable. In addition, the results generated from the AMOS output showed how the demographic factors of the participants and the aspects of scientific thinking were related to their conceptual understanding of physics. With regards to the demographic factors, the results indicated that the male participants tended to have better scientific reasoning than the female participants which, in turn, was more likely to have a strong impact on their understanding of physics concepts. In other words, the male participants were more likely to have performed better than the female participants, not only in scientific reasoning, but also in physics conceptual understanding (PCU). In terms of the **Year Level** factor, it was found that the participants at the higher year levels were more likely to score better on the conceptual understanding of physics than the reverse situation.

Epistemological beliefs, argumentation, and scientific reasoning are all aspects of scientific thinking. Based on the final model, the participants from public universities were more likely to perform better in scientific thinking and conceptual understanding of physics than the participants from private universities. It is worth noting that these relationships were found in direct and indirect effects (including the mediator variables). In summary, the results of the final SEM model suggest important insights about the participants' demographic factors and aspects of scientific thinking that have significant relationships with their conceptual understanding of physics. To reveal the participants' perceptions regarding the relationship between aspects of scientific thinking and their conceptual understanding of physics, as well as the opportunities and barriers to enhancing their scientific thinking and conceptual understanding of physics, a qualitative study was conducted. The findings of the qualitative data analysis are described in the next chapter.

CHAPTER 7
THE INTERVIEW RESULTS:
PERCEPTIONS OF PRE-SERVICE PHYSICS TEACHERS

7.1 Overview

This chapter reports upon the results of interviews that explored pre-service physics teachers' perceptions about the relationships between scientific thinking (i.e., epistemological beliefs, argumentation, and scientific reasoning) and the conceptual understanding of physics. In addition, the interviews were used to explore participants' perceptions of physics teaching and learning they had received in their classrooms and how it helped to shape their scientific thinking and conceptual understanding of physics. This information was then used to illuminate the opportunities and barriers experienced by participants in fostering their scientific thinking and conceptual understanding of physics at their university. A description of the participants' demographic information is provided at the beginning of the chapter, which is followed by the findings from the qualitative study. The descriptions of the interview responses are grouped thematically for each qualitative research question proposed in this study.

7.2 Participants' Demographic Information

In the qualitative phase, data were collected from face-to-face semi-structured individual interviews with 25 selected pre-service physics teachers from four universities who participated in this study. The data were recorded using a digital audio recorder with commentary being transcribed for further analysis. In reporting the qualitative findings, the names of the individual participants and universities were de-identified by code in order to retain anonymity and confidentiality. The four universities have been given the pseudonyms of University A, University B, University C, and University D. The names of the participants were coded as P1, P2, P3, ..., P25, followed by the name of the university they attended. For example, a participant named P18 from University D was coded as P18_D. Interviews were conducted with 11 male (M: 44%) and 14 female (F: 56%) pre-service physics teachers from four universities. The interview participants represented the entire year level, namely Year 1 to Year 4, as follows: six participants (24%) represented the research sample for Year 1, three participants (12%) for Year 2, nine participants (36%) for Year 3, and seven participants (28%) for Year 4. In addition, 18 participants (72%) were from public universities, while seven participants (28%) were from private universities. Table 7. 1 provides the demographic information for the participants in this qualitative study. The results of the interviews are presented in the following section.

Table 7. 1 Demographics of the participants

Participant Code	Gender (F/M)	Year Level	University Code	Participant Code	Gender (F/M)	Year Level	University Code
P1	M	1	A	P14	M	3	A
P2	F	1	A	P15	F	3	B
P3	F	1	B	P16	F	3	B
P4	M	1	B	P17	F	3	C
P5	F	1	C	P18	M	3	D
P6	M	1	C	P19	M	4	A
P7	M	2	B	P20	F	4	A
P8	F	2	C	P21	M	4	A
P9	F	2	D	P22	F	4	A
P10	F	3	A	P23	F	4	B
P11	M	3	A	P24	F	4	B
P12	M	3	A	P25	M	4	C
P13	F	3	A				

7.3 The Relationships Between Scientific Thinking and Physics Conceptual Understanding

This section focuses on the answers to the third research question proposed in this study, namely:

Research Question 3:

What are pre-service physics teachers' perceptions of the relationships between scientific thinking (i.e., epistemological beliefs, argumentation, and scientific reasoning) and their conceptual understanding of physics?

The interview process was guided by three open-ended questions for each participant as provided in the interview protocol attached in Appendix 6. The participants were asked about their perceptions of the relationship between the three aspects of scientific thinking (i.e., epistemological beliefs, argumentation, and scientific reasoning) and their conceptual understanding of physics.

Relationships between Epistemological Beliefs and Physics Conceptual Understanding

The following responses of the participants are highlighted to represent the results of the interviews in order to address the question related to the participants' perceptions about the relationship between epistemological beliefs and physics conceptual understanding. Before expressing their opinions about this relationship, the participants seemed to be trying to conceptualise the description of epistemological beliefs which helped them to better respond to the interview questions. For instance, P17_C from University C revealed that "epistemological beliefs are the beliefs underlying the search for truth of knowledge or the origin of a theory, including concepts and theories in physics. So, I believe that there is a relationship between epistemological beliefs and physics conceptual understanding" (P17_C). However, P17_C did not provide further explanation about how these variables are connected.

P18_D from University D believed that it is not an easy task to understand phenomena occurring in everyday life. He needed to make an effort to search for more information which in turn could have an impact upon his brain and change his way of thinking. As stated by P18_D,

> In understanding the physical and natural phenomena occurring in daily life, reading books alone is not enough. Other supporting references will be needed. Exploring the information provided by such references affects my initial beliefs about certain physical phenomena or knowledge, which eventually also influences the way I think to understand the concepts that underlie such phenomena (P18_D).

This participant did not explicitly state that there was a connection between epistemological beliefs and conceptual understanding of physics. He believed that in order to find out the truth about a concept, one source of information was not enough to change his thinking and help him to understand better.

Meanwhile, P6_C from University C responded to the interview questions showing more concern with differences in the levels of epistemological beliefs of students. He gave more detailed responses to explain how his epistemological beliefs might affect his ability to understand physics concepts. He said, "According to me, epistemological beliefs have a relationship with conceptual understanding of physics, or other scientific concepts, for the reason that epistemological beliefs are the beliefs underlying the search for truth of knowledge." He added:

> If we have low levels of epistemological beliefs, we will assume that the information or knowledge delivered by the teachers in the classroom is true, so we will tend to just accept such information or knowledge without exploring further the information from other sources to find out the truth of the knowledge delivered. Contrarily, if we have high levels of epistemological beliefs about the truth of knowledge, we will be compelled to think further on the truth of the knowledge or information delivered by teachers in the classroom. As a result, we will have better understanding of physics concepts and also be more confident with the results of our thoughts (P6_C).

P6_C illustrated that the students may have varying levels of epistemological beliefs. He considered that epistemological beliefs affected students' ability to process new information they received which, in turn, contributed to their academic performance.

A similar response was also provided by P12_A from University A. He commented,

> There should be a relationship between epistemological beliefs and physics conceptual understanding. If a student just believes in the information provided by their teacher as a truth, they will stumble in improving their understanding of physics concepts for the lack of motivation to find information from other sources. Besides, if they just accept the wrong physics concepts from their teacher and make no effort to find information from other sources, they will likely develop misconceptions. But if they have greater motivation to explore information from various sources, their thinking will be better developed, and they will have correct understanding of concepts in physics (P12_A).

P12_A argued that misconceptions developed by students might be triggered by students' epistemological beliefs. Both P6_C and P12_A's points of view emphasised that students' epistemological beliefs can lead them to be passive receptors or active learners. According to them, a passive learner indicated by a low level of epistemological beliefs tends to believe in information obtained from his or her own instructors as an undeniable truth. On the other hand, an active learner indicated by high levels of epistemological beliefs tends to be curious to find out more information that can stimulate their brain to think critically about new information obtained in the classroom, so they can develop a better understanding of the concepts.

Interestingly, there was one participant (i.e., P19_A) from University A who pursued another point. He stated that epistemological beliefs affect other aspects of learning more than his understanding of physics concepts. He said, "According to my mind, epistemological beliefs do not affect the ability to understand the physics concepts, but rather influence the strategies and motivation to learn concepts and theories of physics" (P19_A). This participant believed that epistemological beliefs can promote important elements that help students, such as their motivation and strategies or approaches to learning. P9_D from University D also expressed another opinion. She considered that epistemological beliefs should influence students' ability to develop and express arguments. She said, "epistemological beliefs are our beliefs about the truth of a theory, supported by experiment-derived evidence, which in turn, will be able to influence our ability to express opinions or refute different opinions of others" (P9_D). P19_A and P9_D clearly recognised that epistemological beliefs affect various aspects of student learning such as motivation, learning approaches, and the ability to carry out argumentation.

In short, the participants in this qualitative study have tried to outline the term epistemological beliefs. They revealed that epistemological beliefs are beliefs that underlie the search for the truth of knowledge or the origin of a theory. Of the 24 participants who responded to this interview question, 22 (91.7%) agreed that there is a relationship between epistemological beliefs and conceptual understanding of physics, in which epistemological beliefs could affect their mastery of understanding physics concepts. Additionally, other participants acknowledged that their epistemological beliefs contributed to their motivation and learning strategy, as well as their ability to share arguments. In the next section, the participants' perceptions about the relationship between argumentation and conceptual understanding of physics is presented.

Relationships between Argumentation and Physics Conceptual Understanding

This section presents the participants' responses to the questions about the relationship between argumentation and conceptual understanding of physics. Before expressing their further opinions about this relationship, the participants tried to outline the term 'argumentation'. For example, P20_A from University A said that "The skill of argumentation is one's skills of conveying and defending opinions or ideas and even refuting opposing opinions of others." She also stated, "If I have a correct understanding of physics concepts, I will have the confidence to convey my opinions or even refute opinions that may differ from mine. But if my opinions turn out to be incorrect, there will be a thinking process to correct my previous understanding of certain concepts, which in turn, will help me understand physics concepts correctly" (P20_A). From this participant's point of view, having a good basic knowledge of physics and self-confidence is an important factor for students when engaging in a class discussion. She also believed that in a discussion, a thought process had taken place which, in turn, could increase her understanding of physics materials. Similar remarks were made by both P16_B from University B and P18_D from University D as follows:

> The ability to express opinions or ideas is strongly influenced by the understanding or knowledge of a phenomenon or a previously mastered scientific concept. It will be hard for me to share my opinions or arguments regarding the topic discussed if I have no adequate understanding or knowledge. My opinions that I present or my ability to get my arguments across will correspond to the limitations of my understanding of a particular concept. Therefore, reading books to obtain as much information as possible and to understand materials related to the topic to be discussed is important. It will foster my confidence and motivation to get actively engaged in discussions where I can express my ideas or defend my opinions (P16_B).

> To have good argumentation skills in discussions, it is important to read various references, so we gain insights and basic knowledge that we can use as our basis to express our ideas or defend our opinions because when we present our opinions, it must be accompanied by a logical and scientific explanation with unquestionable sources for them to be acceptable. So, in my opinion, one's understanding of physics concepts is related to his/her argumentation skills, regardless of how it affects or how they are related to each other. Additionally, we must muster our courage to speak in public. If I have good understanding but no courage or confidence to speak and express my opinions before my friends, my attempt to convey arguments will be impeded. From my perspective, those who often speak in the classroom typically have gained the basic knowledge of certain topics to be discussed, although their opinions sometimes are not entirely true. Thus, discussions should serve as media for exchanging information and thoughts to correct our understanding of certain concepts that might not be right (P18_D).

In terms of engaging in argumentation, the participants highlighted the fact that having relevant knowledge is essential to constructing or defending their arguments based on logical and scientific evidence. They acknowledged that reading books is a way to enrich their insights and knowledge. The participants also emphasised that a class discussion should be a complex scientific practice in order to encourage students in a dialogic exploration of ideas and thoughts which, in turn, can overcome their confusion about the initial knowledge structure they already have, and help them gain a better understanding of physics concepts. Another participant, P15_B from University B, expressed different views as follows:

> From my standpoint, there is a relationship between epistemological beliefs and argumentation ... because when we believe in the truth of a science or certain physics concepts, this belief will help us to defend our opinions or refute others' opinions which might disagree with what we understand regarding such knowledge or concept. I think our low argumentation skills might be attributed to our low beliefs in the truth of certain physics knowledge or concepts. As a consequence, we are bound to be passive learners who do nothing but collect the information coming to us. So, it can be stated that argumentation skills, too, have an influence on our physics conceptual understanding (P15_B).

P15_B further added:

> According to my mind, to have the ability of argumentation, to have a good basic concept understanding, or knowledge is not a priority as we can actually argue based on our own thoughts or daily experiences. It will not matter whether the information or knowledge we present is correct or incorrect. So, it will not hurt to share our opinions or arguments despite our limited understanding or knowledge as long as our opinions and arguments are made by referring to our prior experiences (P15_B).

P15_B clearly asserted that having good basic scientific knowledge is not necessary when engaging in class discussions as long as it refers to prior experiences in daily life. There is a possibility that existing knowledge or preconceptions constructed by students are not the "correct" ones. She also highlighted that the ability to develop arguments is not only related to mastery of physics concepts, but also to epistemological beliefs, namely beliefs in the truth of knowledge. As mentioned in the previous section, passive learners may be caused by the low level of epistemological beliefs they have. Naive epistemological beliefs can influence students' ability to be actively engaged in scientific learning practices such as class discussions that could contribute to their argumentation skills.

Based on the results of the interviews, the participants acknowledged the close connection between their ability to develop and present arguments and the depth of their understanding of physics materials. More than half of the participants stated that mastering basic scientific knowledge or having adequate cognitive resources would help them to be actively engaged in class discussions. They further argued that reading various references could help them to acquire this basic knowledge. Through class discussions, where the thinking process takes place, the participants believed that they gained more opportunities to practice expressing their opinions or refuting others' opinions. This in turn helped them to better understand physics concepts. Additionally, participants in this qualitative study also revealed that self-confidence and courage, as well as motivation, were key factors in being able to actively speak in public in order to practice their argumentation ability by sharing their opinions or arguments.

Relationships between Scientific Reasoning and Physics Conceptual Understanding

Participants' perceptions about the relationship between scientific reasoning and the conceptual understanding of physics are described in this section. The participants in this study had positive views about the relationship between these two factors. More specifically, P12_A from University A stated that "The ability to reason scientifically aids me in understanding the concept of a natural phenomenon, including physical science" (P12_A). A similar response was also made by P20_A from the same university. She expressed her views which began with her description of the term 'scientific reasoning'. She said, "scientific reasoning, to the best of my knowledge, is a process of thinking about a natural phenomenon to make it acceptable and understandable in a logical and reasonable way." She elaborated her view: "making sense of a natural phenomenon helps my way of thinking in understanding the concept underlying it. Thus, to have good scientific reasoning skills is highly helpful for me to understand the physics concepts" (P20_A). This indicates that the participants' understanding of physics concepts would be better when they have good scientific reasoning ability, due to physics learning requiring higher-level thinking skills.

In addition to the importance of having the ability to reason scientifically in understanding a natural phenomenon, P19_A from University A also suggested that understanding physics concepts could not be achieved by simply memorising a concept that could be forgotten later. He stated,

> As far as I am concerned, in understanding the natural phenomena happening around us or in proving the truth of physics concepts or theories, it is critical for one to have good reasoning ability, so they can make sense of their thought. That is why memorization is not the right way to understand the physics concepts. To do so, and correctly, one should go through a thinking process using their scientific reasoning skills. Therefore, learning physics is not through memorization of formulae or words only, but also through a thinking process using reasoning skills. If one learns that way, I am sure that the physics concepts that have been studied will not be easy to forget. I believe that I will be able to understand the physics concept far better if I have high levels of scientific reasoning ability than if my scientific reasoning ability is inadequate (P19_A).

In contrast to the opinion of the participants mentioned previously, one participant had different ideas about the relationship between scientific reasoning and the conceptual understanding of physics. P8_C from University C shared her personal experiences by giving an example. She said,

> From where I stand, physics conceptual understanding affects the ability of scientific reasoning. The reason is that many natural phenomena that are not in accordance with the theories or concepts described in books. For instance: Two objects, a rock and a piece of paper, which are of the same weight are dropped at the same height. I would assume that the object which would hit the ground first would be a rock which is obviously denser than a piece of paper. I believe that paper would float in the air. But in the books, it is stated that these two objects will hit the ground at the same time. It turned out that my reasoning ability was unable to help me in understanding the physics concept related to such falling objects, so that it could make sense and appear logical. Hence, I conclude that having the correct understanding of physics concepts will help me engage in scientific reasoning activity better (P8_C).

Interestingly enough, while many participants argued that having the ability to reason scientifically is needed in order to understand physics concepts, P8_C tended to argue otherwise. She faced a problem along the way of understanding basic physics principles and being able to apply them in real life situations. For instance, she did not succeed in understanding natural phenomena that should be approached with the concept of free-falling objects. The participant's low ability for scientific reasoning might have led to this situation.

The responses of the participants in these interviews inferred that there was a relationship between their ability to reason scientifically and the ability to master physics concepts. This in turn helped them to understand phenomena that occur in everyday life. These participants recognised that better understanding of physics concepts cannot be easily acquired by memorising the physics material or physics formulae. However, the process of thinking using scientific reasoning skills must be present in their learning. They also argued that a higher level of scientific reasoning skills contributed to a better conceptual understanding of physics.

With regard to the third research question in this study, it can be concluded from the evidence that the most dominant view found among the participants was that scientific thinking, comprising epistemological beliefs, argumentation, and scientific reasoning, had a positive effect on their understanding of physics concepts. The responses of these participants are consistent with some of the quantitative findings presented in Chapter 6. The qualitative data have provided important insights and understanding about what the results of the quantitative analysis mean, particularly about such relationships.

The participants' perceptions of the classroom environment and their approaches to learning related to aspects of scientific thinking and understanding of physics concepts are presented in the following section. This section will outline some of the opportunities and barriers experienced by the participants in improving their scientific thinking and conceptual understanding of physics during their university studies.

7.4 Physics Teaching and Learning

In this section, the initial focus of the interviews was to find out the participants' perceptions about physics teaching and learning at their university as they attempted to answer the fourth research question.

Research Question 4:

What are pre-service physics teachers' perceptions of the opportunities and barriers in enhancing their scientific thinking and conceptual understanding of physics?

A list of questions for the interviews is provided in the interview protocol as attached in Appendix 6. The participants were asked about their perceptions of several aspects related to the classroom environment and their approaches to learning. This section outlines the extent to which opportunities and barriers experienced by the participants enhanced or impeded their scientific thinking and conceptual understanding of physics during the course they attended at university.

Facilities and Learning Resources

Participants' perceptions of the facilities and learning resources provided by their universities are illustrated. Based on the interview results, the participants acknowledged that their university had provided adequate facilities to support their independent learning. This included the provision of books in the library, an Internet connection, and laboratory equipment, etc. More specifically, P3_B from University B revealed some of the facilities or learning resources provided by her university and the benefits she obtained from these, as follows:

> Such facilities as books in the library have been provided by the university quite a lot, and Internet connection has also been available. It will be up to how motivated the students are to use the facilities as sources of information to improve their knowledge and understanding of physics material. Sometimes, I learn physics from the books provided in the library, especially the practicals modules which I use as references for conducting experiments in the laboratory (P3_B).

Another participant, P11_A from University A also had a similar point of view, as illustrated by the following comment:

> The facilities made available by the university, such as books in the library, laboratory equipment, and Internet connection, are just sufficient for supporting students' out-of-class learning. I personally often use the laboratory equipment to conduct independent or group experiments because it will be easier for me to understand the physics concepts if I observe the phenomena firsthand (P11_A).

P3_B and P11_A highlighted the fact that, their university had provided opportunities for students to use various learning resources to support them in finding more information and being able to benefit from learning experiences on campus. Additionally, they agreed that a variety of these learning facilities helped them to improve their knowledge and their understanding of physics concepts.

Conversely, other participants were more forthright in stating that the facilities or learning resources provided by the university were limited and did little to support their learning activities in the classroom. This view of a barrier to learning was revealed by P18_D from University D who said, "In my opinion, the books available in the library are very limited in number. This is one of the inhibiting factors for students when trying to find and explore information deeper" (P18_D). Another participant, P15_B from University B, also expressed a similar point by saying,

> Actually, the university's library has provided a wide range of books, but most of the time, I could not find the books I needed. The books provided mostly are the textbooks used in lectures, while to gain more broader insights and knowledge, I need other scientific books such as the latest and advanced Encyclopedia books (P15_B).

P18_D and P15_B acknowledged that books as learning resources can help them to gain more insight and knowledge, as well as supporting their learning. According to them, the limited availability of textbooks, including lecture textbooks and general reference books at their universities, was one factor that inhibited their learning process.

Furthermore, the participants described the limited availability of laboratory equipment and little use in supporting physics teaching and learning, either in the classroom or in the laboratory. These views are highlighted by P3_B from University B:

> To me, the laboratory equipment for conducting experiments is still limited in quantity, and some is even unavailable. This may cause impediment when I try to prove the truth of a physics concept or a phenomenon through an experiment that I believe does not make sense somehow. Consequently, I tend to take for granted the information conveyed by the lecturers or from textbooks (P3_B).

P17_C from University C also made a comment saying:

> I think the laboratory equipment available at my university is not fully utilized in teaching and learning activities in the classroom or in the laboratory. So, we scarcely conduct practicals. Besides, the Internet connection on campus is poor and a bit unreliable to use as it is too slow. This is a hindrance for students to find information or learning resources through the Internet (P17_C).

P17_C expressed her disappointment with the slow Internet connection at her university which she believed did not support her learning. Undoubtedly, the role of reliable Internet access in higher education is crucial for students' learning. The lack of a fast Internet connection could provide a difficult hurdle for some students and this issue might impede their academic practice and performance.

P8_C from University C pursued another point. This is illustrated by the following comment:

> So far, I haven't had high motivation to use the facilities provided by the campus to support my learning. I still have a hard time managing my time to explore information or read books from the various reading sources available at the library or to conduct experiments independently in the laboratory. Perhaps, I am too occupied by campus assignments for physics or other courses, preparation for exams, and other activities out of the campus (P8_C).

The motivation of students to use facilities or learning resources provided by their university is also an important aspect of concern. According to P8_C, she should take advantage of the learning resources provided by her university if she really wants to use it. However, a lack of motivation and limited time impeded her from fully utilising the learning resources available to her on campus.

Furthermore, students' lack of motivation or enthusiasm for learning might also be triggered by the classroom environment. Undoubtedly, instructors have a crucial role in creating a good classroom environment. The classroom environment should be effective and conducive to providing good learning opportunities for students. For instance, if a classroom has high noise levels and inadequate air conditioning, this could have a tremendous impact on student learning. This was revealed by P17_C from University C. She expressed her views as follows:

> In certain subjects, the classroom atmosphere is quite noisy to the extent that it grows less conducive for us to learning. Of course, this condition is a bit irritating and prohibitive for the learning activities to take place, causing us being less focused in the learning process in the classroom. In my personal opinion, if the teaching lecturer earns enough respect from students, the class condition will tend to be quiet and conducive. But on the contrary, if they do not get the due respect from the students, the class tend to be noisy and uncomfortable for learning. Besides, due to insufficient air conditioning in the classroom, the class temperature rises, and this also affects our concentration and comfort while studying in the classroom (P17_C).

This participant's statement indicated that a conducive learning environment as well as maintaining thermal comfort in the classroom contributes greatly to students' ability to learn. Another participant, P24_B from University B pointed out that class size also had an impact on students' learning and academic performance. P24_B revealed her opinion as follows:

> In my opinion, lecturers have tried to teach well. But because there are too many students in the classroom, there are 50 of them, the class atmosphere becomes less conducive for learning because the lecturers will have a hard time giving attention to all students, especially those in the back seats. It is also hard to hear the lecturer's voice clearly. As a result, I have difficulty accepting or understanding the materials delivered in class. Moreover, the ability to absorb information differs between students. I am one of those who are slow at understanding things, so it is hard for me to digest new information or knowledge delivered by the lecturers. Because of the unconducive class condition, I become less motivated to pay attention to all information given by the lecturers in class (P24_B).

The participant's responses in this conversation indicated that large class sizes had an effect on her motivation for learning. According to P24_B, classes that consisted of many students can limit the ability of instructors to supervise students' learning activities that occur during the teaching and learning process in the classroom. In addition, instructors may experience difficulties in meeting the needs of all students, in order for them to be able to learn effectively. Overcrowded classrooms could also prohibit student movement. This implies that large class sizes present many obstacles that hinder the optimal learning of students and this may, in turn, affect their motivation and academic achievement.

To sum up, the interviews above outlined various participants' views regarding the facilities and learning resources provided by their university. Some participants in this study expressed satisfaction with the availability of learning facilities on their campus that supported their learning, such as the book collection in the library, the Internet connection, and laboratory equipment. Meanwhile, other participants expressed disappointment with the limited learning facilities available in their university that affected their motivation and the quality of their learning. For example, they complained about the problems created by high temperatures in class and overcrowded classrooms. However, to confirm these findings, further investigation is needed to obtain a broader picture of the facilities and learning resources provided by the universities involved in this study.

To find out information about the participants' perceptions regarding the instructional methods implemented in the physics classroom, their responses from the interviews are outlined in the following sections.

Teaching Methods

The participants were asked to comment on the types of teaching methods implemented by their instructors in class. The purpose of this question was to understand the extent to which teaching methods implemented in the classroom had an impact upon, or enhanced, participants' scientific thinking and conceptual understanding of physics.

As far as the teaching method is concerned, of the 25 participants, all agreed that they often received lectures as part of traditional classroom instruction where lecturers delivered the content through one-way communication. For example, P11_A from University A expressed his positive views about the teacher-centred method applied in the classroom. He stated,

> The physics learning in the classroom is mostly conducted using teacher-centred or lecture methods. In my personal point of view, the lecture method is not too effective to promote students’ scientific thinking but can improve students’ understanding of physics material and concepts. Because the culture here in Indonesia is that students are lazy to read and find sources

> of information on their own. According to me, the lecture method is appropriate in helping students to acquire knowledge or insights about the physical sciences easily (P11_A).

P11_A further asserted that the strategy of teaching by telling was effective for helping students master physics concepts or physics material, because of the reading culture issue in Indonesia. According to him, traditional teaching methods seem to be more efficient when applied to Indonesian students who have a low interest in reading and are reluctant to look for other sources of information. In the traditional classroom, a traditional teacher-directed approach provides an opportunity for students to ask questions and receive answers immediately if they are having difficulties with their studies.

Other teaching methods which encourage students to focus their attention on learning were also experienced by the participants, as mentioned by P16_B from University B:

> Some lecturers have used inquiry-based teaching methods that encourage students to understand the natural phenomena happening in daily life by applying physics concepts and theories. In this way, students may practice finding solutions in their own way through an investigation to solving the problems about physical phenomena presented by their lecturers. Inquiry-based teaching methods are able to encourage me to use the ability to think scientifically in solving problems and to better understand physics phenomena happening in nature (P16_B).

The responses by P16_B indicated that the inquiry-based approach is a type of teaching method that applies a two-way communication approach. An inquiry-based teaching approach provides a rich learning experience for students, which is certainly different from the situation in traditional classroom. The inquiry-based teaching approach actively engages students and prompts them to use their thinking skills and to develop their curiosity as they attempt to understand real-life phenomena. A similar response was also provided by P10_A from University A. She revealed that demonstrations and experiments were teaching approaches that promoted students' scientific thinking and conceptual understanding of physics. She expressed her views as follows:

> Lecturers generally deliver physics material using the lecture method. Some use props for class demonstrations and simple experiments. Physics experiments are typically conducted in the laboratory. In experimental activities, from designing experiments, collecting data, and analysing data of experiment results, I am expected to use my scientific thinking skills and adequate physics conceptual understanding in order to draw satisfying conclusions. But at times, experiment results are not in accordance with the physics theories explained in the textbooks. For this reason, we are trained to be more critical in conducting an experiment or analysing the results of such experiments (P10_A).

These teaching approaches allow students to actively engage in learning and develop their scientific knowledge. More specifically, experimental activities allow students to confirm a physics phenomenon and to practice being scientists. Conducting ‘hands-on’ learning such as this may increase students' curiosity and enhance their understanding of certain physics concepts.

In addition to this, P1_A from University A indicated that class presentations and discussions are other types of teaching methods implemented in the classroom. He explained:

> The most frequently used physics teaching method applied in the classroom is the conventional method, but in some lectures, presentation method is usually applied, followed by discussions. Through class presentations and discussions, students are encouraged to practice speaking in class to express or defend their opinions. I think, how much the physics teaching methods applied in class affect students’ thinking skills and physics conceptual understanding is indivisible from how much the students are motivated to get actively involved in the learning activity and how big the lecturer’s support is in supervising and giving feedback to the class presentations and discussions (P1_A).

As suggested by P1_A, active participation by both lecturers and students should take place in the classroom in order to create a meaningful learning environment. Based on the responses from the participants above, it seems that the various teaching approaches applied in some university classrooms actually have offered an opportunity for them to gain a better understanding of the physics material and to practice their scientific thinking skills.

In contrast to the participants' experiences mentioned earlier, other participants had different responses to the interview questions regarding the extent to which teaching methods had facilitated their understanding of physics concepts and promoted their scientific thinking. P20_A from University A said, “Overall, some 70% of the teaching method applied in class is the lecture method, while the remaining 30% are simple practicals, mini-projects, group discussions, and class presentations.” She further added:

> To me, the physics learning I am experiencing in the university is not that different from what I was experiencing back in high school, especially when it comes to the teacher-centred methods, which restrict students’ active involvement in the class. I even feel that the physics material given during high school are better retained than those given in the university. Maybe it is because my high school teachers were more attentive to their students than my university lecturers (P20_A).

According to this participant, traditional teaching approaches that transmit knowledge from instructors to students are not only experienced in secondary schools, but also at the university level. In the conventional classroom, the teacher as the provider of information, transmits material rather than facilitating students to actively engage in the learning process. The teaching methods implemented in the classroom can indirectly influence the learning strategies applied by students.

For example, they might focus on memorising rather than developing their higher-level thinking skills. P24_B from University B, for example, stated that:

> Most physics material is delivered by lecture method, so I tend to only listen to and take notes of the materials. What matters most is that I have notes as my reading material for my exam preparation. Sometimes, lecturers also provide PowerPoint slides for us, so we can study the materials delivered in class again at home, although we do not necessarily understand the content of such notes or PowerPoint slides. So, I focus more on taking notes instead of attempting to understand the materials because I cannot afford to miss the information or knowledge delivered by the lecturers in class if I do not write it down right away (P24_B).

Delivering physics material through traditional teaching methods allows students to sit passively, listening to the teacher's explanation, and to make notes. Ideally, the mastery of physics concepts may be achieved by students when different teaching and learning resources are used by instructors. Instructors should encourage their students to read many other books or references, rather than just relying on the notes they take in the physics classroom. The traditional physics teaching methods implemented in the classroom seem to provide little opportunity for students to develop the necessary conceptual understanding of physics, or training in thinking scientifically. In turn, this can affect students' motivation to learn physics, as indicated by P23_B from University B:

> The courses we are taking in one semester are quite a lot, so if the materials are delivered by the lecture method more often than not, they will not be challenging enough for me to think nor attractive enough for me to pay more attention, especially in late hours when my learning motivation is dropping and it is difficult for me to stay concentrated to accept physics information or materials in class (P23_B).

P18_D from University D expressed the same opinion:

> I think that the lecture method gives me only a little chance to ask questions or express opinions. I do not even have the chance to correct each other if my lecturers deliver inaccurate information or material. Therefore, the lecture method which allows no more than information transfer prevents me to get the opportunity to practice my thinking skills (P18_D).

Participants' responses indicated that instructors were expected to be able to help their students become active learners who can construct their knowledge better and practice their thinking skills in order to understand physics concepts through two-way communication. P4_B from University B further highlighted a number of other points. He argued,

> The delivery of physics material in class is predominantly done by the lecture method. After delivering the materials, lecturers give students the opportunity to ask questions if they do not understand some of the concepts. I think lecturers should provide more opportunities for students to present their opinions or comments on the physics material delivered in the classroom, so they will be stimulated to use their thinking or reasoning abilities to express their

> thoughts or opinions. In this way, two-way learning may take place. What might become a problem for lecturers is the small amount of time available for delivering a lot of physics material. If there is a long discussion in class, it will be time-consuming as well, while they are responsible for delivering all of the materials planned. In the end, the lecture method is always a favourite choice for lecturers to deliver materials in the classroom as it is far more effective, and it does not require a long period of time both in the teaching preparation and the material delivery in the classroom, as compared to other teaching methods (P4_B).

The response from P4_B indicates that traditional teaching methods can help students acquire lots of physics material more easily than other teaching methods. In addition, implementing conventional teaching approaches makes it easy for instructors to prepare and deliver large amounts of physics material in class, even in limited time.

Based on the interview responses mentioned earlier, this study concludes that the barriers to physics teaching and learning are closely related to traditional teaching approaches being mostly implemented in the physics classroom. Lecturers tend to adopt conventional teaching approaches in which they play a central role, and their students are placed in a passive state of learning where they are required to simply absorb information rather than understand new concepts. In other words, traditional teaching methods allow no more than information transfer by instructors, while the students tend to only listen to their instructors, take notes about the materials, and practice solving physics problems using formulae and mathematics calculations. According to the participants' responses, this teaching method gave them little chance to ask questions or express their opinions. This in turn restricted their active involvement in the class. Consequently, students had little interaction with their instructors and the learning materials.

Physics Material

Physics material is crucial for understanding the complexity of the real world. The participants were asked to comment on the physics material delivered by instructors during their courses. The participants acknowledged that physics material delivered in class should help them to make sense of the natural phenomena around them. P1_A from University A, for example, mentioned: “The physics material delivered in the classroom have actually encouraged students to think scientifically because physical sciences are exact sciences which are supported by data or facts that cannot be influenced by internal personal life” (P1_A). Another participant, P15_B from University B, also expressed a similar view illustrated by the following comment: “From my point of view, physics is not abstract. Learning physics means learning to understand nature. Thus, learning physics is not by memorizing, but by understanding how natural phenomena happen" (P15_B). P11_A from University A further emphasised that thinking skills are needed to understand what we experience in the real world. He stated: "The physics material delivered by lecturers in the classroom are derived from experiments or the scientific thinking of physicists. Thus, it is necessary to have thinking skills to understand such physics material" (P11_A).

Based on the above interview responses, the participants recognised the close relationship between physics and phenomena in real life. Studying physics allows students to connect to the world around them. The participants also believed that if students are to understand natural phenomena well, just memorising physics material was not enough, and that adequate scientific thinking skills were needed.

However, the participants also revealed that the physics teaching material they received in class only helped them in a small way to understand natural phenomena. They stated that physics material delivered in the classroom was mostly related to the collection of formulae and the derivation of physics formulae, rather than physics concepts or physics applications in the real world. Some examples of such views were:

> After receiving physics material from our lecturers, we practice solving physics problems using mathematical formulae and calculations. We rarely practice solving physics questions that measure our ability in understanding physics concepts or relate the application of the physics material to the phenomena in the universe (P11_A).

Another participant said,

> In general, the learning in the classroom has been more centred around physics formulae derivations and exercises. That is why I find difficulties in relating physics formulae with the natural phenomena happening in everyday life. Besides, at times, the physics material delivered in the classroom different from the material that is experimented in the laboratory. I think experiments in the laboratory should be able to help me understand the physics material delivered by lecturers in the class, or the physics theories presented in the classroom should be able to be tested through experiments in the laboratory. Unfortunately, in reality, the physics material delivered in the classroom and physics topics in experimental activities in the laboratory is not in line (P3_B).

P3_B from University B further argued that the role of scientific experiments in the laboratory is less helpful for gaining an understanding of physics material presented in the classroom. She said that she expected that physics instruction should help students to link theory and practice to motivate them to understand the physics world better. Apparently, participants held the view that the role of laboratory activities was less effective in helping students to understand the physics material in class.

Another point was expressed by P19_A from University A, who explained:

> In my opinion, there are too many physics material given, so I often forget the materials I have ever learned. There are certain physics topics that I have not been able to understand fully, but then I must receive other physics material. The amount of materials given to students is not synchronous with the time provided. Moreover, I also am quite busy with my campus's student organization activities. With that many credits (144 credits), a study period of 4 years is not enough for me to understand all physics material correctly (P19_A).

P19_A asserted that the overload of learning material and the limited availability of time contributed to ineffective student learning. P23_B from University B also commented on the learning material she received in class saying:

> As senior students in the final year level, we no longer receive courses with physics material, but we do receive courses related to education or pedagogy, and I focus more on completing my final project (thesis). I hardly ever read physics books or go deeper into the physics material I have ever received early in my campus life. I think that is the reason why I have forgotten a large portion of the physics material I have learned (P23_B).

The responses of these participants indicated that a large amount of material given over a limited time provided challenges for students in being able to obtain meaningful and long-lasting learning that stayed in their memory. They also had difficulty remembering and seemed to forget much of the physics material they had learned.

Further to this, the interview responses also indicated that the content of the physics learning materials delivered by a lecturer in class should provide an opportunity for students to understand physics concepts. In doing so, they need to have the ability to think scientifically in order to be able to connect these physics concepts with natural phenomena that occur in everyday life. However, there were still some barriers faced by the participants in their efforts to understand these physics learning materials. For example, the participants in this study had the view that physics teaching and learning practices were more focused on the derivation of physics formulae. As a result, they faced difficulties in relating these physics formulae to physics phenomena happening in the real world.

To obtain more associated information, the participants' perceptions of the learning media that were often used in the classroom are described in the following sections.

Learning Media

Learning media are usually used by instructors to enhance their effectiveness in transferring learning material to students in the classroom, as well as to motivate the students and enhance their interest in learning. The participants in this study revealed that their instructors generally used chalk and a blackboard, and/or PowerPoint presentations as their learning media tools when delivering physics material in class. After the lecturers explained certain physics topics, the students would sometimes receive a copy of the PowerPoint slides to read at home. For instance, P7_B from University B said, "The learning media frequently used by lecturers are chalk, blackboard, and PowerPoint presentations, sometimes there are also props for demonstrations in class" (P7_B). A similar response was also provided by P12_A from University A, who said, "The learning media used in class are blackboard and chalk, PowerPoint presentations, learning videos, and props for demonstration in class. But the ones frequently used are blackboard, chalk, and PowerPoint slides." He further expressed his opinion: "To me personally, learning media like props and learning videos are more interesting and are more helpful for me to better understand physics concepts. Unfortunately, they are still rarely used by lecturers as aids for delivering physics material in the classroom" (P12_A).

Illustrating concepts or phenomena through demonstration activities and learning videos seems to attract students' attention more than just listening to lecturers' explanations in the class through the use of a blackboard and/or PowerPoint presentations. The use of props for demonstration and learning videos provided more learning opportunities for students to be actively engaged in using their thinking skills. However, learning media such as PowerPoint slides, and chalk and blackboard, seemed to achieve little to help students to understand physics concepts, or to engage them actively in the learning process during their course. As revealed by P15_B from University B, "According to my mind, learning media such as PowerPoint slides, chalk, and blackboard give little pressure for students to think or participate actively in the learning process in the classroom" (P15_B).

Similarly, P17_C from University C also described her experience by giving an example to justify her point of view:

> I think such learning media as chalk, blackboard, or even PowerPoint slides give little stimulation to students to actively ask questions or express opinions when the lecture is underway. I even

> get bored and am not too motivated to pay attention to the materials being delivered by the lecturer in the classroom because the PowerPoint slides used only display a collection of sentences or physics formulae, which is not so different from when blackboard and chalk are used. Ideally, learning media can help students understand physics material more easily and attract students' attention so they can be more focused on learning. They should also be able to help students understand the many phenomena frequently considered to be abstract, or in other words, making concrete what is considered abstract by students. To me personally, physics is in no way abstract. It is real and provable. It only appears abstract because we cannot see it with our naked eye, for example, we cannot see electric charges. Because of their micro size, electric charges cannot be seen with the naked eye, but their presence can be proved through experiments or can be simulated through animation or learning videos that can better attract students to learn (P17_C).

In summary, the participants in this study acknowledged that instructors had used several types of learning media to support their teaching in their efforts to transfer subject content knowledge to students in the class. This learning media included chalk and blackboard, and PowerPoint presentations. However, such learning media often used by instructors did not provide many opportunities for students to be more motivated about the learning process, to be actively engaged in learning activities, or to use their thinking skills to understand physical phenomena or the materials presented in class. The participants also revealed that learning media such as props and animated videos or learning videos would better attract them to learn physics and help them to better understand physics concepts. However, these learning media are still rarely used by lecturers as the learning tools for delivering physics materials in class.

To find out more information about student learning activities during the learning process in the classroom, the participants were asked for their opinions regarding this matter, as described in the following section.

Learning Activities in the Classroom

Learning activities should essentially arouse students' interest and attention in learning physics, as well as promote their engagement in the learning process in the physics class. A number of participants expressed their positive views on the learning activities they had experienced at their university. They revealed a variety of learning activities that were more student-centred, which gave students more opportunities to develop their ability to think scientifically and to master physics concepts. For example, P10_A from University A expressed her opinion about the class demonstration activities and group discussions she had experienced. She said,

> According to my personal experience, class demonstrations and group discussions are more encouraging for me to use the ability to think scientifically and physics conceptual understanding because when expressing my opinions or arguments, I must be able to provide logical information or knowledge so that it does not deviate from existing concepts or theories (P10_A).

Another participant (P19_A from University A) also shared his experience regarding his experience of class presentations:

> In the final year of my studies, the physics course is mostly delivered through class presentations where my peers present their works. These activities can encourage students' active participation in delivering their opinions or sharing their insights on a particular physics topic (P19_A).

P1_A from University A also provided his personal experience and comments. He very clearly stated:

> From where I stand, class discussions are one of the effective ways to practice scientific thinking skills and improve understanding of physics concepts of students. By actively expressing or defending my opinions in the classroom, I will have the opportunity to get corrections or feedback from my lecturers or my classmates if there is inaccuracy in my understanding. In this way, I will gain a correct understanding of a certain physics concept, so that existing misconceptions can be minimized. However, discussions will be more useful to me if the lecturer also takes an active role in providing feedback or supervising students in this activity. By doing so, students' active role will be better controlled. Students' misconceptions in physics can then be corrected directly during the discussions (P1_A).

Meanwhile, similar remarks were made by P13_A from University A, who expressed positive views about her experiences of actively engaging in experimental laboratory activities.

> In my opinion, the experiments or practicals in the laboratory can provide me with encouragement to practice my scientific thinking skills because to conduct an experiment, I will need to go through a thinking process during the data collection or analysis for the purpose of proving a certain physics theory. If from the experiment I draw a conclusion that is inconsistent with the existing theories, I will have an opportunity to re-collect data and study the factors that might cause the errors in the data I previously collected, which will give me a better understanding of the physics concepts (P13_A).

Clearly, the responses in the interviews at this point indicated that learning activities such as class demonstrations, group discussions, class presentations, and experiments in the laboratory provided more opportunities for students to deeply explore physics materials and to develop higher-level thinking skills that might ultimately improve their learning outcomes.

Other participants had different perspectives on the learning activities in the classes they had experienced. For instance, P19_A from University A argued,

> To increase students' active participation in arguing or exchanging opinions, some courses are delivered through class presentations. Unfortunately, students can only present physics material to the extent of their knowledge or understanding. In class presentations, lecturers as sources of information are not too active in adding information or insights to strengthen the material presented by the presenting groups and do not moderate the flow of discussions between students, so the discussions only go according to the depth of the students' knowledge or insight, which might not necessarily be always correct. I think, discussions will be more beneficial for students if the lecturers can provide feedback or additional insights into the physics material being discussed, especially when the information or material delivered by the presenting groups is incorrect. In this way, students will get a better correct understanding (P19_A).

Student-centred learning activities are sometimes not fully, or well, implemented, so the active role of both the students and the lecturers is crucial. More specifically, class discussions and presentations may run effectively if the instructors also supervise their students, give opportunities for them to speak up in discussions by providing complex or interesting questions, as well as providing critical feedback. Another participant, P7_B from University B, also shared his experience by saying:

> I think there is not much difference in the class activities I experience in the university and the ones I did in high school. There is not much variety in the class activities in which I get engaged. I more often listen to my lecturers and take some notes. Besides, I also do some physics problem exercises and assignments. There hardly ever be group or class discussions in my class. This is probably due to the considerable amount of material that must be delivered by lecturers in the class while the time is very limited. So, having too many discussions in class might prevent a lot of physics material from being delivered in detail (P7_B).

Conventional teaching environments do not seem to provide many opportunities for students to construct knowledge and develop their learning skills. A range of learning activities are needed to help students focus their attention and interests. Passively receiving lectures and assignments is the learning activity most frequently experienced by students. Consequently, they are less motivated to participate in class.

Responses from other participants also indicated a number of barriers to learning in the classroom. In the following, P11_A from University A mentioned:

> I feel that the lecturers have given the students the chance to get actively engaged in physics learning in the classroom, but most students did not make use of such opportunity as they do not have enough courage to speak in the classroom either for expressing their opinions or giving comments on the physics material delivered by the lecturers. Maybe the students do not really understand the materials, so they are afraid of making mistakes when presenting their opinions and comments in the classroom. I think it is also related to our country's culture, which makes it as if one who frequently speak up in the class wants to show off his or her capability in front of lecturers or classmates. This makes us tend to speak and discuss with only those sitting next to us (P11_A).

A similar point was put forward by P8_C from University C, who said,

> I am not too active in speaking or in the learning activities in the classroom because I am worried that my statements or opinions would be too narrow or mistaken, so my friends would comment negatively or judge that my abilities and understanding are too low and eventually take me for a fool. Besides, it is hard for me to muster up my courage and improve my confidence to speak up in the classroom, so I tend to become a passive student (P8_C).

In reality, students' anxiety about making a mistake affects their motivation to be actively involved in classroom learning activities. According to P8_C, the fear of being judged badly by classmates for making mistakes is a significant impediment to encouraging her to be more confident in expressing her opinions and commenting in the classroom.

In short, the interviews discussed the participants' views about various learning activities that took place in the classroom. The participants' responses reflected that the instructors presented different types of learning activities in the classroom to encourage student engagement in practicing their thinking skills in order to master physics concepts. These learning activities included class demonstrations, group discussions, class presentations, experiments, and practicums. However, other participants revealed that such learning activities were rarely carried out in the classrooms; instead, they were more likely to only listen to the lecturers and to take notes. Learning activities that are more focused on the receiving and absorbing of content knowledge are likely to encourage students to spend more time on memorising knowledge and physics formulae and in practicing the solving of traditional physics problems from the textbooks, instead of practicing their thinking skills to understand physics concepts. The limited availability of time was suspected to be one reason why the instructors minimised the application of these more proactive learning activities in the classroom. The participants' responses about their lack of motivation to be actively involved in learning activities, and the lack of lecturer involvement in providing feedback or additional insights related to physics concepts, for example in a class discussion, were some of the barriers that were expressed by the participants in the present study.

Homework or Assignments

Generally speaking, instructors give assignments or homework to students in an effort to encourage them to become independent learners. In this study, the participants were asked to comment on the extent to which assignments or homework contributed to the improvement of their thinking skills and their understanding of physics concepts. The participants in this study agreed that the instructors gave them homework opportunities to complement their learning and the tasks carried out in the classroom. For instance, P24_B from University B expressed her opinion, “To encourage students to study independently and explore deeper into physics material, lecturers have given assignments that are due within certain periods” (P24_B). A similar response was also provided by P2_A from University A, who said, “Because a lot of physics material should be delivered in class, we are given assignments to be completed out of the lecture hours, both individually and in a group. It is hoped that we will be able to better understand the physics material being studied” (P2_A). The responses of the participants indicated that their lecturers had provided an opportunity for them to independently develop their knowledge, and to practice their learning skills through assignments or homework.

Other participants stated that homework provided little academic benefit for them. P15_B from University B, for example, commented about the amount of homework saying, “I feel that the assignments were given by my lecturers are quite a lot, so I have a hard time managing my time to re-read or explore the physics material that has been delivered in class” (P15_B). Additionally, P6_C from University C made a similar point:

> I think that the assignments given by my lecturers do a little in stimulating students to think scientifically. When assignments are given, most students only copy the work of the others who have completed the assignments or work on those assignments together. In my point of view, students do not exert themselves or are not serious enough to do the assignments. I myself haven’t felt the contribution of these assignments in improving my thinking skills or my understanding of the physics concepts (P6_C).

In addition, P3_B from University B revealed her opinion that homework is more centred around traditional physics problems. She stated:

> We usually are given problem-solving questions, most of which calculation problems are mainly rather than physics concepts. More often than not, feedback is not given, so I am not really sure whether what I have done is correct. Sometimes we do the assignment together with our friends. The only thing that matters for us is that we complete and submit the assignment since we already suspect that what we do won’t be reviewed by the lecturers (P3_B).

P7_B from University B further pointed out that homework as a tool for learning contributed less to the development of student learning outcomes. Limitations in providing feedback on errors in homework and the lack of tracking of student progress had an impact on the low level of motivation of students to complete their assignments. In his own words, he mentioned:

> I think the works submitted to the lecturers are rarely reviewed or given feedback, so we simply submit them (for the sake of completing the assignments). I am not too sure whether my works are correct. Maybe this has something to do with my lecturers' activities, so they have no time for giving us feedback (P7_B).

Based on the interview responses above, it can be concluded that assignments or homework overload that has been given by lecturers raises its own challenges for students to re-learn physics material that has been received in class. In addition, the barriers faced by students regarding homework given by the instructors relate to the types of assignments that do not stimulate students' thinking skills and have little impact on their motivation to learn independently outside the classroom. In fact, the participants expected valuable feedback from their instructors to assist them to develop the quality of their learning because feedback on homework is crucial for promoting better understanding. So far, the benefits of homework for the students are still a question mark in helping them to promote their thinking skills and to master physics concepts.

Learning Assessment

Undoubtedly, student learning is influenced by a number of factors, with one being the type of test or assessment used by instructors. This section presents the participants' responses to questions concerning the learning assessments. Generally speaking, tests or examinations such as those used in the mid-semester and final semester exams are designed to reflect what the lecturers expect their students to have mastered. In the following quote, the type of questions or tests used for assessments were revealed by P15_B from University B, who stated,

> I believe that the physics material tested on students is quite varied, but it will be also up to the lecturers. Some questions are meant to measure the ability to understand physics concepts and also to solve problems using physics formulae. The exam questions mostly refer to the physics material or exercises that have been taught in the classroom, but the level of difficulty varies (P15_B).

Of the 22 participants who responded to this interview question, all revealed that the test items provided in the mid-semester and final semester exams usually consisted of problem-solving questions which emphasised the use of formulae and calculations without probing the conceptual understanding of physics of the students. Working on quantitative questions that tend to rely on memorising formulae and practicing calculations appears to be solved more easily by students than qualitative questions that require conceptual understanding. Interestingly, P13_A from University A expressed her opinion and justified her point of view about the reasons why lecturers tended to provide quantitative rather than qualitative problems in examinations. She stated,

> The physics questions in exams are mostly related to problem-solving with the use of physics formulae or physics formulae derivations and mathematical calculations. In my opinion, it is easier for lecturers to develop physics questions with highly complicated mathematics calculations than to develop physics questions that test students' physics conceptual understanding. The reason probably is that there are only a limited number of questions that can test students' physics conceptual understanding. This is why lecturers tend to give physics questions with highly difficult calculations in the form of problem-solving involving complex physics formulae than questions that primarily measure students' physics conceptual understanding (P13_A).

This in turn contributed to the participant's perception that the physics courses she studied were only a collection of formulae used to solve numerical problems that had little relevance to real-life applications and had little connection to students' daily life experiences. In reality, the type of questions in examinations contributed to the students' efforts to prepare for the test if the test items were dominated by quantitative questions without further interpretation of the numerical answers being needed. This leads students to spend more time memorising formulae, imitating algorithmic problem-solving procedures, and solving textbook problems in order to pass the exams. No doubt, passing the examinations was still the main priority of most of the participants in their study journey at university. For instance, P24_B from University B shared her experiences:

> The questions given in the mid-term and final exams are more about problem-solving using physics formulae, so most students prepare for the exams or study the exam materials the day before the exam days. We usually practice with physics questions given to us in the classroom or from textbooks. Sometimes, the questions given are also similar with those given in previous years. But there are also some unpredictable questions we usually have a hard time answering, which causes us to unquestionably get unsatisfactory grades (P24_B).

A similar response was also provided by P16_B from University B, who said,

> In my humble opinion, the assessment instruments provided by lecturers in either midterm or final exams do little to measure students' scientific thinking skills or their physics conceptual understanding or physics application in daily life. At times, the exam questions can be done out of the class hours, so my friends and I are given the chance to answer the exam questions by searching information from the Internet or even copying the work of those who have solved the questions. I am not even sure whether my work is correct. The most important thing for me is that I can answer the exam questions and get good grades. This is especially the case when the questions given by the lecturers can be solved in a group. A group will get the same grade, so my grade will depend on my group's work, not my individual ability (P16_B).

P16_B asserted that the questions given in the examinations did not measure the extent to which students were able to think scientifically. In addition, the examination results obtained by the students did not seem to reflect their real competencies or abilities to master physics concepts. A similar point of view was put forward by P4_B from University B, who stated,

> I think the test scores given by lecturers have not represented or reflected students' real competences or abilities, because maybe the standard is unclear and too low. The test instruments used mostly are centred around problem-solving, so they do not really measure students' physics concept understanding. If students often practice with question examples similar to those tested, the chance is that they will be able to answer the exam questions correctly then get high scores. In my opinion, students who get high grades do not necessarily reflect that they have mastered or understood the physics concepts correctly (P4_B).

Interestingly, P19_A from University A was very impressed with the types of questions given in this study. In the following, P19_A shared his personal experience:

> In all honesty, I had a rather hard time answering the questions in the research survey, especially those that measure my ability in the understanding of the physics concepts. In the classroom, the lecturers hardly ever discuss physics questions that measure the understanding of physics concepts such as those in the survey. Up until now, we are used to working on questions that involve the application or derivation of physics formulae. The physics questions given every semester do not really measure students' physics conceptual understanding. So, I as a prospective physics teacher realize that there are specific questions for measuring to which degree students understand physics concepts and that there are apparently many physics concepts that I have yet to truly understand.

He added to his comment by saying:

> Besides, some midterm or final exam questions tested sometimes are similar to those tested in the previous year, so I focus more on practicing solving the exam questions of last year and memorizing them as preparation for the exam. My main priority is then passing and getting good grades. Because I prioritize getting high grades, I focus less on improving my competence as a prospective physics teacher (P19_A).

The goal of passing the examination and getting high grades is certainly an expectation for almost all students. Ironically, how well they master the physics materials covered during the learning process was not a major concern for them. High competition for better jobs and income nowadays made them perceive that bad grades could affect their careers in the future. Furthermore, employment demands, and career competition also determined the students' goals to study at university. P11_A from University A, for example, expressed his opinion as follows:

> In my personal opinion, it is way easier to gain grade A now than it was in previous years. This is probably caused by the employment demand and competition that require applicants to have a high GPA. In Indonesia, high GPA becomes a requirement for one to be hired. This makes students be more oriented to achieve a high GPA in order to get decent works, regardless of whether the learning process and the outcome have met the expectation (P11_A).

The participants' responses above imply that the examinations given by the instructors have not measured the extent to which participants think scientifically and understand physics concepts. Most of the tests measured the participants' ability to solve problems by applying more formulae and mathematics calculations.

Further to this, the participants in this study were interviewed for their opinions or comments related to the learning approach adopted during their studies at university, as outlined in the following section.

Approaches to Learning

After exploring the information regarding the assessment of learning, the next issue addressed in the interviews was whether approaches to learning implemented by the participants contributed to the improvement of their scientific thinking skills and their understanding of physics concepts. The interview results indicated that the participants already had self-awareness of the importance of independent learning and tried to not simply rely on the teaching and learning delivered by the instructors in the classroom to enhance the quality of their understanding and their skills. For instance, P17_C from University C expressed her experience by saying:

> I often study independently by reading books from several references to having a deeper understanding of the physics material delivered by lecturers in the classroom. If after reading books I still find difficult to understand some materials, I will ask my friends who have a better understanding. I do not ask my lecturers directly because I am worried that my questions are just simple and unimportant ones, which I should have understood (P17_C).

Similarly, P19_A from University A also described his personal experience and how he stimulated his own motivation by saying,

> To learn physics concepts, I prefer using diagrams or pictures because since when I was a child, I find it easier to understand phenomena through pictures or diagrams, so I can imagine how they are in reality. So, as much as possible, the physical phenomena I learned can be visualized to make it easier for me to understand it. Moreover, it is hard for me to memorize the physics theories delivered by my lecturers in the classroom or contained in textbooks. As a motivation to review the physics material I have learned, I give high schoolers private tutoring. In this way, I am more motivated to understand some physics material before teaching them to my students (P19_A).

Another participant, P16_B from University B expressed her opinion about the benefits of collaborative learning, stating that:

> In my personal standpoint, studying in a group or discussing with my friends helps me to understand the physics material or concepts better ... because I will be more confident and less ashamed to ask my friends in my study group than in my class. So, joining a study group gives me and my friends an opportunity to improve our confidence to express our opinions or share information and even motivate each other to be braver in conveying our arguments. Ultimately, it allows me to improve my knowledge and thinking skills (P16_B).

P11_A from University A revealed that he sought to implement various learning strategies to help him improve his ability to think scientifically and master or understand physics concepts. These are different from rote learning which only relies on memorising materials which can be forgotten quickly. He expressed his opinion by saying:

> To have a good understanding of physics material, the learning strategy I apply is conducting independent experiments either at home or in the laboratory. I also often explore learning videos related to physics phenomena. The reason is that conducting physics practicals or watching natural phenomena through the Internet makes it easy for me to remember them than memorization of physics theories from the book does. Over the course of my campus life, I often form study groups and engage in discussions out of class hours. I also help explain the physics material to my friends who do not really understand. In this way, it is easier for me to understand and remember the materials I have learned (P11_A).

Interestingly, P1_A from University A also shared his experience of realising that the teaching and learning carried out in the classroom was not enough to improve his performance and confidence in expressing his opinions or comments in class. He was more interested in being actively involved in learning activities outside the classroom that included attending study clubs on campus which he found helped to increase his confidence in developing arguments and expressing his opinions in public. He clearly mentioned:

> I realize that as a student, I have to study more independently and do not depend on the learning in the classroom. Although I have been given the chance to be actively engaged in learning in the classroom, I feel that this is insignificant in improving my ability to speak and express my opinions. So, to improve my ability to speak in the classroom and to be braver to express my opinions in discussions, I actively engage in the physics study club on my campus, where I get a great deal of opportunity to exchange thoughts and opinions with my friends. Through this club, I gain greater confidence to speak in front of many people and to express my opinions, and this, in turn, can promote my thinking skills and improve my insights or knowledge ... because there, I learn a lot from my seniors who are smarter and more experienced (P1_A).

In contrast to the experiences of the participants mentioned previously, other participants stated that they did not have an adequate learning strategy that was effective for supporting their learning, other than rote learning. P8_C from University C, for example, mentioned:

> I feel that I am short of time to learn and study deeper physics material because there is a lot of physics material given by lecturers, and the courses, I take are also many. Additionally, the out-

> of-class assignments that I have to complete also take most of my time. So, I have not found the right learning strategy that can improve my ability to gain a good understanding of physics material (P8_C).

In addition, learning approaches, such as memorising learning materials to prepare for examinations, are learning strategies that were often used by the participants. In the following quote, P13_A from University A shared her experience:

> According to my mind, I am more serious and focused on learning when exams are approaching. I usually write down a summary of the physics material that will be tested and then memorize it. Because the learning process is fast, I will easily forget the physics material I have learned or memorized (P13_A).

Similarly, P9_D from University D also described her personal experience. She said,

> When studying physics material, I tend to memorize physics formulae instead of trying to understand physics theories. I have applied this sort of learning style since I was in high school until now at the university. Because my priority during high school was passing the National Examination, so I have been used to solving physics questions using formulae and mathematics calculations regardless of whether I understand the physics concepts or not. If I do not memorize the formulae, it is certain that I will not be able to work on the exam questions (P9_D).

Paying attention to the learning approach applied by students should not be ignored given that it contributes to their learning outcomes. In fact, there are many factors that influence students in determining their learning strategies, one of which is based on the situations and barriers experienced by them during the learning process. For instance, lack of motivation and procrastination issues were described by P16_B from University B, who described her experience by saying:

> My laziness, lack of curiosity, and low motivation greatly affect the way I study, especially when I am surrounded by friends who don't really encourage me to be more focused and diligent in studying. As a result, I often procrastinate in doing my assignments or studying independently to get deeper into the materials given on my campus. This presents me a great challenge to improve my competence and ability to understand the physics material or concepts (P16_B).

Based on the interview responses above, it can be concluded that the types of learning approaches applied by the participants were quite varied. These learning strategies include memorising physics formulae, working on physics questions using formulae and mathematic calculations, conducting experiments, discussing the physics material with friends, exploring learning videos, and being actively involved in physics study clubs on campus. To sum up, these learning approaches can be categorised by surface and deep learning (Biggs & Moore, 1993). The participants who used a surface approach to learning, for example, P13_A and P9_D, tended to memorise the content of the physics material or use rote learning in order to prepare for tests and exams. Meanwhile, a participant who used a deep approach to learning; for example, P11_A, tended to explore and try to understand physics concepts or ideas. The participants' responses indicated that of the 23 participants who responded to this interview question, 13 (56.5%) tended to adopt a surface learning approach, while 10 (43.5%) tended to adopt a deep learning approach.

To gain additional information related to other factors that might contribute to the improvement of scientific thinking and the conceptual understanding of physics of the participants, they were asked for their opinion, or to make further comments, on this matter based on their personal experience. This is described in the subsequent section.

Other Factors (Internal Personal Barriers and External Factors)

The participants in this study described various internal personal barriers that they had experienced during their studies at university that may have influenced their academic achievement. For example, P12_A from University A shared both his learning experiences while in high school as well as at university that related to the experience of misconceptions about physics. He said,

> To me personally, the physics learning in the classroom has yet to discuss physics concepts so far, so many students have misconceptions. The misconceptions that have already been there since high school have yet to be corrected in university now. Perhaps, I unquestionably believed that all of the knowledge delivered by my teachers at school was correct. Or perhaps there was no inaccuracy in my teachers' explanations, but it was me who developed misconceptions when accepting those explanations. I think that these misconceptions will always exist unless an improvement is made in the learning system. Especially to me as a prospective physics teacher, it is important to establish a better understanding of physics concepts, the basic ones, in particular (P12_A).

Another participant, P23_B from University B also revealed the learning barriers that contributed to her academic performance and achievement. She described her educational background when studying in secondary school and her prior knowledge as factors that influenced the quality of her learning. She stated,

> I am not too confident in my ability or mastery of physics material. I previously attended vocational high school, not a general high school. In my opinion, the materials delivered in general high school are more wide ranging and detailed than those delivered in vocational high school. Although I took the electronics program at my vocational high school, so I have an adequate understanding of materials related to electricity, I still feel that I am left far behind in other physics topics. When I was in vocational high school, the physics material I learned were by far easier than those in university. Back then, I only learned simple physics calculations, but now, there are many physics formula derivations which I think are very complicated. That is why

> I speak a little in class and become a very passive student. I am afraid my classmates would laugh at me if I ask questions or present wrong information, so I would rather stay quiet and be a listener in the classroom (P23_B).

Meanwhile, P14_A from University A commented on government policies related to the implementation of the National Examination (UN) as a standard evaluation system for students. His point of view suggested that the UN might affect various aspects of students' learning. He argued,

> In high school, the physics material given by my teachers were typically a collection of physics formulae. We have rarely been given questions that can measure the extent of our physics conceptual understanding. Well, there is no helping it when it comes to National Examination questions whose solutions would involve calculations and speed. I also tend to just memorize physics formulae and practice solving problems. Had I only relied on physics concepts, I might have gotten low the national exam scores. It was a difficult choice. I think, this condition was caused by the Government's policy which sets forth that high school students' achievement be assessed through the National Examination and the grades be used as passing parameters (P14_A).

In contrast to the views of the participants mentioned previously, P19_A from University A outlined the difficulty he had organising his study as well as managing his time for learning in class. He also described the difficulty he experienced engaging in activities outside of class, such as his activities for an organisation and providing private tutoring to several secondary school students. Actively engaging in activities outside of class affected his ability to master physics material delivered in class. In his own words, he said:

> My out-of-class activities such as being involved in the organization and giving private tutoring to secondary school students make it difficult for me to manage the time to independently study physics material deeply. Consequently, I do not really get actively engaged in learning activities in the classroom. Due to these very activities, I have also been frequently late in submitting my work to my lecturers. Because of my lack of knowledge of some physics material, I have low confidence, and this makes me less active in speaking or expressing my opinions in the classroom (P19_A).

Apart from this, some of the other participants revealed the external factors that might have an impact on their learning. These participants highlighted that parental influences on students' education might also contribute to their learning outcomes. P16_B from University B shared her experience:

> Studying in the Physics Education major is not my passion. My parents hope that I become a teacher because there is a significantly high opportunity to be a teacher, especially in rural areas like my place of birth. In addition, there has been little interest in the Physics Education major, so the chance of passing or getting admitted in this major is fairly high. Another reason for my parents' choice is that because I am a woman, being a teacher will allow me to spare my time for taking care of my family. But because it is not my interest, I am not too motivated to attend my

> class and learn physics. That is why I am in a great need for support from my friends, lecturers, and especially my parents or family to have my learning motivation elevated (P16_B).

A similar response was also provided by P25_C from University C, who revealed that his parents' expectations influenced his motivation to study a Physics Education major, his learning motives, and his goals in attending university. He stated:

> To be honest, I am not too interested in studying physics fields. But I have no choice but satisfying my parents' wishes as it is they who pay for my studies. So, I am not too motivated to get deeper into the physical sciences, and this prevents me from improving my abilities and performance. Besides, my motivation to study narrows down to only getting a high GPA and earning a bachelor's degree because my parents hope that I get a high GPA. They tend to see the learning outcome more than the learning process. Because applicants with high GPA are prioritized in the employment world, we, who basically are only graduates of private universities, will compete hard to get a job against graduates of public universities or high-profile universities. So, to get a job and to be able to compete with public universities' graduates, I must have high GPA, although high GPA does not necessarily represent that the graduates have better thinking skills and physics material understanding than those with lower GPA (P25_C).

Based on the participants' responses, it can be highlighted that in some cases, students' misconceptions, lack of prior knowledge, and motivations influenced their mastery of content knowledge and improvement of their learning skills. Parental expectations were another factor that affected the students' academic performance. In addition, students who were less interested in studying physics tended to memorise the physics material rather than trying to understand physics concepts. Poor time management, in terms of managing the time to learn and to engage in activities outside of class, was also considered as one of the factors that had an impact on their learning outcomes. Government policies have also had an impact on the practices of the curriculum applied in schools, such as the implementation of the National Examination (UN) as a standard test to evaluate students' learning outcomes.

In short, this research provides valuable insights, indicating that participants' skills in thinking scientifically and their ability to master physics concepts are influenced by various factors as outlined above in a number of themes. This implies that policymakers and instructors need to play a crucial role in considering these factors which contribute to improving the quality of physics education by implementing better physics teaching and learning in the classroom, especially in higher education.

7.5 Summary

The findings in this qualitative section of the research provide important insights and understanding of several aspects or factors that are considered to contribute to participants' scientific thinking and conceptual understanding of physics. In other words, this qualitative research attempts to extend the limited understanding provided by the results of the quantitative analysis.

The interview data showed positive perceptions about the relationship between scientific thinking (comprising epistemological beliefs, argumentation, and scientific reasoning) and the conceptual understanding of physics, which is consistent with the quantitative findings in this study. Furthermore, the interview results indicated that the improvement of scientific thinking skills and the conceptual understanding of physics of the participants was affected by several aspects of physics teaching and learning. These aspects consisted of the facilities and learning resources, the instructional methods, physics materials, learning media, learning activities, homework or assignments, assessment instruments and tests, approaches to learning, and other factors (such as internal personal barriers and external factors). The participants recognised various internal personal barriers, including misconceptions about physics, parental influences, lack of motivation and prior knowledge, and government policies. Overall, this chapter has also described pre-service physics teachers' views about the opportunities obtained and the barriers they have experienced in enhancing their skills for thinking scientifically and improving their conceptual understanding of physics during their course at university.

In this study, the participants' interview responses indicated that the types of teaching methods applied in the classroom and the assessment of learning, or the questions posed by the instructors in tests or examinations, seemed to be the factors that most contributed to the approach to learning chosen by them. These learning approaches, in turn, affected the participants' learning outcomes. The results of the interviews also revealed that the education system standards implemented in the universities, as well as employment demand and career competition oriented the participants towards achieving high grades regardless of the abilities or competencies they should have as future physics teachers. Finally, these qualitative findings in conjunction with the quantitative findings will be elaborated upon and linked to the relevant literature in the next chapter.

CHAPTER 8
DISCUSSION

8.1 Overview

This study has investigated Indonesian pre-service physics teachers' scientific thinking (comprising epistemological beliefs, argumentation, and scientific reasoning) and their understanding of physics concepts. The variable of epistemological beliefs consists of five dimensions, namely the structure of scientific knowledge, the nature of knowing and learning, real-life applicability, evolving knowledge, and the source of the ability to learn. The argumentation variable consists of two dimensions, which are making scientific argumentation and challenging argumentation. Meanwhile, the scientific reasoning variable comprises five dimensions, specifically conservation of weight and volume, proportional reasoning, control of variables, probability and correlational reasoning, and hypothetical-deductive reasoning. Physics conceptual understanding focuses on two content areas, namely mechanics as well as electricity and magnetism (E&M).

Previous studies have revealed the relationship between the following cognitive variables i.e. epistemological beliefs, argumentation, scientific reasoning, and physics conceptual understanding through the analysis of single variate or bivariate relationships (Acar et al., 2015; Coletta & Phillips, 2015; Hotulainen & Telivuo, 2014; Madsen, McKagan, & Sayre, 2015; Sandoval & Millwood, 2007; Venville & Dawson, 2010). However, there has been no research that integrates these multiple variables into a single model. This study has sought to address this gap in the prior research by incorporating all these variables into a single model (see Figure 2. 3 on page 52 in Chapter 2) and implementing multiple statistical analysis techniques. Thus, a comprehensive picture of the relationships arising among these variables, and the demographic factors of the participants, consisting of gender, year level, and university type, was obtained. The participants' perceptions of the relationship between the aspects of scientific thinking and the conceptual understanding of physics were also explored in the interviews. This study also investigated the participants' perceptions of the opportunities and barriers they experienced in terms of the extent to which a range of factors in physics teaching and learning could promote their scientific thinking and their conceptual understanding of physics.

The present study employed a cross-sectional design to collect data from 706 pre-service physics teachers from four universities in Indonesia. A mixed-methods design was used to triangulate quantitative and qualitative data sources. Quantitative data were collected from the participants' responses to several surveys, while qualitative data were collected through semi-structured interviews. Several software packages were employed in the data analysis. IBM SPSS 25, IBM SPSS AMOS 25, and ACER ConQuest 4 for the quantitative data analysis, and NVivo 11 software for the analysis of the qualitative data.

This chapter outlines the findings of the quantitative and qualitative parts of this study presented in the previous three chapters. These findings will be interpreted, and possible explanations made in light of the previous research and the existing literature. This chapter is organised into three main sections, which are the effects of demographic variables (comprising university type, gender, and year level) on pre-service physics teachers' scientific thinking and understanding of physics concepts, the inter-relationships between the aspects of scientific thinking (comprising epistemological beliefs, argumentation, and scientific reasoning) and the conceptual understanding of physics, and pre-service physics teachers' perceptions about the physics teaching and learning factors with respect to the opportunities and barriers to improving their skills in scientific thinking and developing their understanding of physics concepts.

8.2 The Effect of Demographic Variables on Pre-service Physics Teachers' Scientific Thinking and Understanding of Physics Concepts

8.2.1 University Type

As presented in Chapter 3, the pre-service physics teachers involved in this study came from four Indonesian universities, comprising two public universities and two private universities. This study has demonstrated that the type of university attended by pre-service physics teachers had a direct positive effect on their scientific thinking (comprising epistemological beliefs, argumentation, and scientific reasoning) and on their understanding of physics concepts, either directly or indirectly. Based on the t-test results provided in Chapter 5 and the Structural Equation Modelling (SEM) analysis in Chapter 6, the type of university attended by the participants positively and directly influenced their epistemological beliefs, argumentation, and scientific reasoning. The SEM model also showed that the type of university attended by the participants indirectly influenced their understanding of physics concepts through the aspects of scientific thinking, including epistemological beliefs, argumentation, and scientific reasoning. As indicated in the SEM model, the participants' epistemological beliefs and argumentation had an indirect effect on their conceptual understanding of physics, mediated by their skills in scientific reasoning. Meanwhile, scientific reasoning was found to strongly and directly influence their understanding of physics concepts. The findings suggest that pre-service physics teachers from public universities performed significantly better than those from private universities with respect to epistemological beliefs, argumentation, and scientific reasoning, which, could lead to them being better at understanding physics concepts.

As described in Chapter 1, undergraduate students enrolled in public universities in Indonesia are those who have passed a very strict selection process with certain requirements. Based on *The Regulation of the Minister of Research, Technology, and Higher Education of the Republic of Indonesia Number 45 Year 2015*, admission to undergraduate programs is carried out through the *National Selection of State Universities (SNMPTN)*, the *Selection of Joint Entrance State University (SBMPTN)*, as well as an independent selection process (managed by each university). Students selected to study at a public university, especially universities or institutions that offer teacher preparation or teacher education programs, are those who generally have better academic knowledge and skills. Meanwhile, those who fail to pass one of these pathways (tests) may apply to a private university. This implies that the character and the quality of knowledge and skills of undergraduate students from public universities are different from those of undergraduate students from private universities. Such circumstances are consistent with the findings of this study showing that pre-service physics teachers from public universities performed better than pre-service physics teachers from private universities in all aspects of scientific thinking and the conceptual understanding of physics.

As demonstrated in the literature review, there has been no previous comprehensive research conducted on the effect of the type of university attended by pre-service physics teachers on their epistemological beliefs, argumentation, and scientific reasoning, as well as their conceptual understanding of physics, especially in the Indonesian context. The implications of this finding suggest that educational practitioners and policymakers in higher education should pay more attention to the quality of knowledge and skills of pre-service physics teachers, particularly those from private universities, because they will be future physics teachers in Indonesia.

8.2.2 Gender

Students' **epistemological beliefs** are regarded as having essential implications for their learning and academic performance (Adams et al., 2006; Chen et al., 2019; Lising & Elby, 2005; Vecaldo, 2017). More specifically, pre-service teachers' beliefs about knowledge and knowing are considered to play an important role in their teaching practices in their future work as teachers, because these beliefs can influence their students' learning (Pamuk et al., 2017). Understanding pre-service teachers' epistemological beliefs is important, particularly for teacher preparation programs, because these institutions are responsible for preparing future teachers with strong content knowledge, pedagogical competencies, and teaching skills (Langcay et al., 2019). It is worth stating that epistemological beliefs are one of the more important variables related to student learning. With regard to this variable, this section outlines the extent to which gender differences contributed to the pre-service physics teachers' epistemological beliefs involved in this study. Conducting research on the role of gender differences in students' epistemological beliefs contributes to the literature, as the previous research has revealed mixed results and conclusions. In other words, the existing literature has reported differing results about the effects of gender differences on students' epistemological beliefs (Langcay et al., 2019).

In the present study, the t-test results presented in Chapter 5 and the SEM model in Chapter 6 demonstrated that gender differences did not affect the epistemological beliefs of the participants. In other words, the findings demonstrated that there were no significant differences between male and female pre-service physics teachers involved in this study with regard to their epistemological beliefs; basically, they held similar beliefs towards knowledge and knowing. This finding is similar to the results reported by previous researchers who found no significant differences between male and female students with regard to epistemological beliefs (Chen et al., 2019; Efilti & Çoklar, 2016; Schommer-Aikins & Easter, 2006; Tanriverdi, 2012; Tumkaya, 2012; Yalcin & Yalcin, 2017). The quantitative findings of the present study also imply that gender was not a factor that influenced pre-service physics teachers' beliefs towards knowledge and knowing.

On the other hand, some other studies have indicated that students' epistemological beliefs differ significantly according to gender (Adams et al., 2006; Aslan, 2017; Kanadlı & Akay, 2019; Langcay et al., 2019; Muin, Abedalaziz, Hussin, Mohamed, & Md Saad, 2012; Muis & Gierus, 2014; Terzi et al., 2012). Mostly, these researchers have reported that female students' epistemological beliefs were more sophisticated and developed than those of male students. This may be because female students believe more strongly than male students in some points, such as "…the learning could take place over time, depending on the effort, and that learning ability is something that could be improved" (Kanadlı & Akay, 2019, p. 404). Other researchers found significant differences between male and female students with both holding more sophisticated epistemological beliefs in certain dimensions (Aslan, 2017; Langcay et al., 2019; Muis & Gierus, 2014). For instance, Langcay et al. (2019) found that the epistemological beliefs of female pre-service teachers were more sophisticated than male pre-service teachers in terms of the structure of knowledge. Meanwhile, epistemological beliefs of male pre-service teachers were found to be more sophisticated than female pre-service teachers in terms of the ability to learn. This suggests that students can have high levels of epistemological beliefs along particular dimensions, while having low levels of epistemological beliefs in other dimensions.

Undoubtedly, the effect of gender on epistemological beliefs is one of the most investigated research topics. However, the mixed results reported above indicate that there are disagreements between researchers regarding how epistemological beliefs differ by gender. The results of the present study suggest that gender was not a factor that had an influence on shaping pre-service physics teachers' epistemological beliefs. Certainly, the absence of the effect of gender differences in the epistemological beliefs of pre-service physics teachers in this study may be due to the various situations they faced during the period of their studies in higher education. The findings would seem to indicate that male and female pre-service physics teachers in this study were receiving similar learning opportunities and undertaking similar learning activities during the learning process in similar learning environments which, in turn, affected their epistemological beliefs in similar ways. Descriptions related to teaching and learning practices, including the opportunity to learn and other learning factors, are outlined in more detail in the next section (i.e., point 8.4).

Likewise, the t-test results presented in Chapter 5 and the SEM model in Chapter 6 indicated that gender did not have a significant impact on the **argumentation** skills of the participants involved in this study. This finding is consistent with previous studies reporting that there were no significant differences between male and female students in terms of their argumentation ability (Chen et al., 2019; Sampson & Clark, 2009; Widodo et al., 2016). In contrast to this, other studies have identified a gender gap in the development of students' argumentation skills (Hong et al., 2013; Salminen & Marttunen, 2018). Such researchers have reported that male students tended to generate higher quality arguments than female students. Meanwhile, female students tended to experience more difficulties and to lack confidence in constructing arguments that required higher levels of critical thinking skills during their learning process (Hong et al., 2013).

The findings of this study suggest that both male and female pre-service physics teachers have similar abilities in argumentation in terms of making scientific argumentation and challenging argumentation. In other words, pre-service physics teachers' skills in argumentation in this study were independent of demographic factors such as gender. This implies that both female and male participants might be receiving the same opportunities to engage in, and to practice their argumentation skills during the learning activities undertaken as part of their studies in higher education, regardless of whether or not these argumentation skills were presented and taught by their instructors in the classroom. In the research literature, few studies have focused on the role of gender differences with regard to argumentation skills among pre-service teachers, particularly in Indonesia. These results contribute to the literature by providing information about the effects of gender on the argumentation skills of pre-service physics teachers in the Indonesian context.

Furthermore, the t-test results presented in Chapter 5 and the SEM model in Chapter 6 indicated that there were significant differences between male and female pre-service physics teachers in terms of **scientific reasoning**. Male pre-service physics teachers were found to have higher levels of scientific reasoning skills than females. This finding is in line with that reported by Coletta et al. (2012), and Nieminen et al. (2013). These researchers revealed that male students tended to outperform female students on the ability to reason scientifically. On the other hand, Alshamali and Daher (2016) reported that the performance of female students was better than that of male students in scientific reasoning. However, many other studies have found no significant differences between male and female students in scientific reasoning (Novia et al., 2018; Piraksa et al., 2014; Talib et al., 2018; Thuneberg, Hautamäki, & Hotulainen, 2015). Apparently, research on whether gender plays an important role in affecting students' scientific reasoning is still limited (Talib et al., 2018), and researchers have found inconsistent results.

Further to this, the results of the SEM analysis in this study demonstrated that gender has a direct and positive effect on pre-service physics teachers’ **understanding of physics concepts,** as well as an indirect effect that was mediated by their scientific reasoning skills. The findings show that male pre-service physics teachers tended to have higher abilities compared to female pre-service physics teachers in understanding physics concepts. These results are consistent with previous studies indicating that male students outperformed female students on tests of mechanics as well as electricity and magnetism concepts (Coletta et al., 2012; Henderson et al., 2017; Karim et al., 2018; Kost-Smith, Pollock, & Finkelstein, 2010). Furthermore, Kost-Smith, Pollock, and Finkelstein (2009) noted that the gender gap was greatest in the conceptual understanding of physics among students with high levels of scientific reasoning skills. On the other hand, Hairan et al. (2019) indicated that there was no statistically significant difference between male and female undergraduate students with respect to the conceptual understanding of physics, especially in the understanding of Newtonian mechanics concepts. Other previous research has also found that gender did not significantly affect students' understanding of physics concepts, particularly in the understanding of electromagnetic induction concepts (Adolphus & Omeodu, 2016). The varied findings of the prior research with regard to gender differences may be triggered by variations in different populations, year level studied, and other factors.

The results of this study have found that there was a gender difference, particularly in pre-service physics teachers' scientific reasoning and conceptual understanding of physics. This gender gap observed in male and female pre-service physics teachers might be caused by various factors such as field of major study, and stereotype threat. For instance, instructional approaches or teaching strategies and standard instruments implemented in the classroom are factors that may contribute to gender differences in measuring students' physics achievement (Henderson et al., 2017). In natural science subjects, male students have been found to have higher levels of academic performance in physics, female students performed better in biology, and there were no differences between female and male students in chemistry (Yamtinah, Masykuri, Ashadi, & Shidiq, 2017). Male students might be more confident in participating and being independent, in conducting experiments or investigations related to physical phenomena than female students. If scientific investigations are often carried out by students either in or outside of the laboratory, male students are more likely to have better scientific reasoning skills than female students, because male students tend to be better in observing phenomena, controlling variables, and drawing conclusions (Yamtinah et al., 2017), where these abilities are needed to carry out scientific investigations. Certainly, gender differences cannot be easily explained by a single factor. Many studies have investigated the factors that contribute to gender differences such as prior knowledge, cognitive ability, standardised tests, self-efficacy, mathematics and science anxiety, and gender stereotype or stereotype threat (Henderson et al., 2017; Karim et al., 2018; Mallow et al., 2010; Miller & Halpern, 2014; Shapiro & Williams, 2012; Wang & Degol, 2017).

According to Makarova, Aeschlimann, and Herzog (2019), "gender stereotypes are part of a broader belief system that includes attitudes toward female and male family roles, female and male occupations, and gender-associated perceptions of the self." Li, Zhang, and Wang (2017) further pointed out that stereotypes contribute to gender differences, especially in students' mathematics and physics achievements. Female students have been found to be more anxious in mathematics and science than male students (Henderson et al., 2017), whereas students' scientific reasoning has been shown to be significantly correlated with their mathematical abilities (Nieminen et al., 2013). The stereotypical view held by female students is that science and mathematics is a male-dominated field (Makarova et al., 2019). This view may negatively affect female confidence for success in the fields of science (Amelink, 2009). More specifically, stereotypes are a potential factor contributing to the poor achievement of female students in the subject of physics (Marchand & Taasoobshirazi, 2013).

As noted by Henderson et al. (2017), the under-representation of female students in physics is significant and well-documented. This can be explained, in part, by the stereotypes present in society relating to careers and professional trends for male and female students. For instance, "engineer, architect, doctor, and police" are considered more suitable for males, while "elementary and secondary school teacher, nurse, housekeeper, and typist" are considered more suitable for females (Chen et al., 2019, pp. 637-638). In the qualitative part of the present study, participant P16_B expressed the same point of view, namely: "… I am a woman, being a teacher will allow me to spare my time for taking care of my family." Apparently, gender differences can also be attributed to the role of the female in the family, where females are seen to have the main responsibility for doing household chores. As also noted by Moss-Racusin, Dovidio, Brescoll, Graham, & Handelsman (2012), gender differences in science are caused by females taking a disproportionate amount of time caring for their children and families.

On the other hand, males may struggle to meet societal and family expectations to succeed in their careers, because they have the responsibility of fulfilling the needs of their family. Consequently, they might be more serious about learning science or physics because mastering these fields means they are more likely to have a better career in the future. Therefore, they would be more motivated to study science or physics, and this could result in them having a deep understanding of scientific phenomena and being able to view the world of physics through a range of different scientific viewpoints. It could also mean that they are more easily able to apply science in their daily lives. Undoubtedly, many factors have been investigated to describe gender inequality, especially in the field of physics. However, there is no single factor that can fully account for this gender gap. In fact, this complex phenomenon cannot be observed, measured, and explained simply or easily in order to draw valid conclusions (Madsen, McKagan, & Sayre, 2013).

In general, the findings of the prior research have indicated that gender differences play an important role in students' cognitive abilities and academic achievement (Yu, Shek, & Zhu, 2018). In contrast to this, some of the previous research has found that gender does not play a significant role in student learning achievement (Kost et al., 2009; Roohr, Liu, & Liu, 2017). The results of this study indicate that gender differences play an important role, particularly in pre-service physics teachers' scientific reasoning and conceptual understanding of physics. Nevertheless, gender did not affect their epistemological beliefs and argumentation skills. To the best of the researcher's knowledge, there has been no comprehensive research on the effect of gender on pre-service physics teachers' epistemological beliefs, argumentation, and scientific reasoning, as well as their conceptual understanding of physics, especially in the Indonesian context. Clearly, this demonstrates the need for a great deal of work to be done in this area in order both to explain and reduce the gender differences found in aspects of scientific thinking and the conceptual understanding of physics of Indonesian pre-service physics teachers.

8.2.3 Year Level

The results of this study have confirmed the relationship between the year level of pre-service physics teachers and aspects of their scientific thinking and conceptual understanding of physics. It could be expected that students at a higher year level of university study would have a higher level of knowledge and skills than those at lower year levels. However, the results of the one-way analysis of variance (ANOVA) shown in Chapter 5 and the SEM analysis in Chapter 6 indicated that there were no statistically significant differences in pre-service physics teachers' epistemological beliefs, argumentation, and scientific reasoning with regard to their year level of study. Meanwhile, year level was found to have a direct and positive effect on pre-service physics teachers' understanding of physics concepts. This implies that participants at the higher year levels are more likely to perform better in understanding physics concepts compared with participants at lower year levels. The explanation of each of these variables is outlined below.

In terms of **epistemological beliefs**, the findings indicate that there was no statistically significant difference in pre-service physics teachers' epistemological beliefs measured by the *Epistemological Beliefs Assessment for Physical Science* (EBAPS) survey with regard to the year level. This finding is consistent with previous research showing that year level did not have any significant effect on students' epistemological beliefs, especially at the tertiary level (Bates, Galloway, Loptson, & Slaughter, 2011; Brauer & Wilde, 2014; Ding & Zhang, 2016; Gire, Jones, & Price, 2009; Yalcin & Yalcin, 2017; Yenice, 2015). In contrast, other research (mentioned below) demonstrated that there was a statistically significant difference for year level with regard to the epistemological beliefs of students in higher education. For instance, Perry revealed that university students in their first year of study held the belief that knowledge is certain and simple, as well as handed down by an authority, but those in their final year held the belief that knowledge is complex and tentative (Schommer, Crouse, & Rhodes, 1992). Other research showed that the epistemological beliefs of university students are unstable; but, changing and developing over time (Aslan, 2017; Belet & Guven, 2011; Marzooghi, Fouladchang, & Shemshiri, 2008; Schommer-Aikins & Duell, 2013). These researchers found a positive trend across year level in which first-year students had more naive epistemological beliefs than higher year-level students. In other words, as students moved from the lower year levels to the higher year levels, their epistemological beliefs seemed to develop and become more mature and sophisticated. This could be attributed to students gaining more knowledge and experience over time during their course of study.

The findings of this study suggest that pre-service physics teachers' epistemological beliefs were relatively unchanged or developed little over the course of their undergraduate education or four years of study (see Figure 5. 10 for the mean of the epistemological beliefs score). Several possibilities can be offered to explain this situation. The possibility is that pre-service teachers starting undergraduate studies in physics education hold quite highly developed epistemological beliefs. The level of their epistemological beliefs could have reached a plateau and remained stable and did not grow until their fourth year of tertiary-level study. An alternative possibility is that teaching and learning practices in the university classroom did not significantly develop pre-service physics teachers' epistemological beliefs during the four years of their study, or over the course of their undergraduate program. As pointed out by Khine (2010), students' epistemological beliefs may change over time because they are influenced by their education. This study implies that the education received by pre-service physics teachers during the period of their study at university did not significantly influence their epistemological beliefs.

Some researchers have noted that the level of students' epistemological beliefs is strongly influenced by the teaching methods and strategies adopted in the teaching and learning process (Brauer & Wilde, 2014; Ekinci, 2017; Ogan-Bekiroglu & Sengul-Turgut, 2008; Winberg et al., 2019). For instance, teaching methods based on constructivist approaches are likely to promote students' epistemological beliefs to a more advanced level. In the constructivist teaching approach, students have opportunities to practice their skills to solve real-life problems, think critically, conduct investigations, and collaborate in group work. In other words, this method encourages students not to be passive in acquiring knowledge, but requires their active participation in the process of constructing knowledge. Therefore, a learning environment that offers a constructivist approach could support students to hold sophisticated epistemological beliefs (Ekinci, 2017). On the other hand, traditional teaching practices basically concentrate on transferring knowledge from teachers to students. In this case, students have fewer opportunities to construct their knowledge or to be actively engaged in their learning. Hence, traditional teaching methods do not seem to enhance students' epistemological beliefs. Clearly, constructivist teaching methods are closely related to more sophisticated epistemological beliefs, while traditional teaching methods are more closely related to naive epistemological beliefs (Ekinci, 2017; Kirmizigul & Bektas, 2019). It is thus reasonable to state that teaching and learning practices that tend to transfer information or knowledge to students may not contribute to enhancing the level of their epistemological beliefs.

Further to this, the ANOVA results presented in Chapter 5 and the SEM model in Chapter 6 indicate that there was no significant difference between participants' year level and their **argumentation** abilities. Previous research has reported that high school students' argumentation showed no significant difference over their grade levels (Kaya et al., 2012; Widodo et al., 2016). In contrast, Osborne et al. (2016, p. 838) found that the higher the grade level of middle school students, the higher their ability to master argumentation, such as "identifying and making claims, selecting evidence to support a claim, and providing reasoning that links claim and evidence." In addition, Berland and McNeill (2010) demonstrated that students at higher-grade levels seemed to be more experienced in constructing an argument using more complex evidence that requires a deeper understanding of the subject.

As mentioned previously, in this study, pre-service physics teachers' argumentation abilities did not develop significantly across the four year levels of their study in higher education (see Figure 5. 10 for the mean of the argumentation score). There might be several factors that influence the mastery of argumentation ability among Indonesian pre-service physics teachers. The findings suggest that classroom teaching and learning processes might provide little support for practising argumentation skills or engaging students through complex scientific practices that enable them to construct arguments or justify claims. As pointed out by Widodo et al. (2016), the teaching strategies implemented by Indonesian teachers in the classroom tend to focus on simply transferring content knowledge to students, while rarely encouraging students to provide evidence that supports their claims. In addition, the qualitative findings of this study also indicate that due to the small amount of time available for covering a lot of physics material, physics teaching and learning in the classroom tended to implement teacher-centred methods where students tended to only listen and take notes. For instance, a male participant, (P18_D) revealed that the lecture method gave him little opportunity to ask questions or express opinions and that, as a consequence, he tended to be a passive listener. This suggests that instructors who dominate the learning process in class inhibit students from getting the chance to practice and develop their argumentation skills.

The absence of argumentation practice in class might be because the teachers were more focused on delivering course material to prepare students for the examinations. Therefore, the students tended to memorise the physics materials, which left them with few opportunities to engage in argumentative activities or to practice their argumentation skills, such as constructing an argument, expressing their opinions using strong evidence, or defending an argument. An alternative possibility is that the educators or teachers might avoid opportunities to teach argumentation practices in their classes because they have limited pedagogical skills in teaching or organising classroom activities that can support the practice of their students' argumentation (Kaya et al., 2012; Nurlatifah et al., 2018; Wang & Buck, 2016). Consequently, argumentation practices rarely take place in the classroom (Aydeniz, Pabuccu, Cetin, & Kaya, 2012; Heng et al., 2014; Osborne et al., 2016). This could result in students lacking experience in constructing arguments and expressing or defending their opinions.

This suggests that teacher education programs should place more emphasis on the construction and development of pedagogical knowledge of pre- and in-service teachers, especially those pertaining to the enhancement of their skills in argumentation (Acar et al., 2015). Murphy et al. (2018) noted that a teaching approach based on class discourse could be implemented by teachers to develop students' knowledge of argumentation and their ability to generate arguments. Certainly, argumentation skills will not be present in the class unless the teacher has an advanced understanding of argumentation skills, is able to organise an active argumentation environment, and has the ability to both teach and train these skills to their students.

In terms of **scientific reasoning**, the findings of this study indicate that pre-service physics teachers' scientific reasoning, as measured by the *Lawson's Classroom Test of Scientific Reasoning* (LCTSR), did not develop significantly over the four years of their undergraduate program (see Figure 5. 10 for the mean of the scientific reasoning score). This is consistent with previous research reporting that the year level of university students did not have any significant effect on improving their scientific reasoning skills (Ding, 2018; Ding, Wei, & Liu, 2016). On the other hand, other research has shown a statistically significant difference for year level with respect to students' scientific reasoning at the tertiary level (Ding, Wei, & Mollohan, 2016). These researchers demonstrated that the progression trends of students' scientific reasoning skills showed little improvement across year levels. In other words, the existing literature shows that students' skills in reasoning scientifically do not progress satisfactorily during the period of their study at university.

The results of the present study suggest that there was no relationship between the year level of pre-service physics teachers and their scientific reasoning skills. Indeed, there was no improvement detected in the participants' scientific reasoning. Certainly, this finding cannot be explained by a single factor. This might be due to the fact that the teaching and learning process in higher education makes little contribution to the development of pre-service physics teachers' scientific reasoning. The instructors might focus more on students' content knowledge rather than enhancing their skills in reasoning scientifically. As demonstrated in the qualitative results in this study, instructors tended to put more emphasis on the importance of content knowledge, physics formula derivations, and solving physics problems. Therefore, pedagogical strategies were predominantly more traditional. In traditional classroom practice, students are required to go deeper into the content knowledge; consequently, their scientific reasoning skills are poorly developed and show little improvement (Ding, 2018). In the same vein, Bao, Fang, et al. (2009) indicated that lecture methods emphasising traditional problem-solving did little to promote the development of students' scientific reasoning. Consequently, the learning strategies adopted by students were more focused on memorisation or rote learning which might offer fewer opportunities for them to engage in the practice of using scientific reasoning skills. Such learning strategies are inadequate for students to practice and improve their scientific reasoning skills. Practicing scientific reasoning skills should require more inquiry or investigation as part of the learning process. The qualitative results in this study also demonstrated that laboratory activities were not fully utilised to support teaching and learning practices. Previous research has found that physics laboratory activities or experimentation strategies that engage students in scientific inquiry promote students' scientific reasoning skills (Koenig, Wood, Bortner, & Bao, 2019; Zeineddin & Abd-El-Khalick, 2010). These researchers pointed out that laboratory activities allow students to generate and test a hypothesis, create an experimental design, and engage in inquiry that allows them to practice their skills in scientific reasoning. As also pointed out by Ibrahim et al. (2015), scientific reasoning skills are crucial when students have to reconcile prior theories or existing knowledge by evaluating experimental results.

To sum up, the findings imply that teacher education institutions involved in this study did little to enhance the aspects of scientific thinking of pre-service physics teachers, regardless of how many years they spent studying in higher education. This may be because the teaching and learning approaches adopted in class placed little emphasis on developing undergraduate students' epistemological beliefs, or on their ability to construct arguments and use skills to reason scientifically. More specifically, in traditional education settings, instructors are likely to be oriented towards developing students' ability to derive physics formulae and solve physics problems using formulae and mathematical calculations. Therefore, students also tend to focus more on recalling content knowledge, memorising physics formulae, and deriving physics formulae rather than trying to understand physics theories and concepts, or developing their scientific thinking skills. These findings reveal that instructional design that focuses simply on content knowledge and a collection of physics formulae may be insufficient to promote the scientific thinking skills of Indonesian prospective physics teachers. Further, learning to pass examinations rather than learning for understanding may explain why pre-service physics teachers' scientific thinking did not improve across the four years of their studies in higher education. The results of this study should raise the alarm for educators in Indonesian teacher education institutions. There is a need to consider appropriate and effective pedagogical strategies and teaching materials, as well as the creation of supportive learning environments in order to develop the skills of scientific thinking of learners, especially pre-service physics teachers in Indonesia.

In relation to the variable of **physics conceptual understanding**, the t-test results in Chapter 5 and the SEM model in Chapter 6 show that pre-service physics teachers in higher year levels performed better than those in the lower year levels. However, participants’ physics conceptual understanding, as measured by the FCI (*Force Concept Inventory*) and the BEMA (*Brief Electricity and Magnetism Assessment*) instruments, showed unsatisfactory improvement across the entire undergraduate program (see Figure 5. 10 for the mean of the physics conceptual understanding score). Simply put, the teaching and learning activities received by the student-teachers of physics at the universities participating in this study may not be contributing to improving understanding of physics concepts over time. Similar results have been found in previous studies which revealed that although misconceptions about the topic of mechanics of student-teachers of physics decreased over their years of education at the tertiary level, the results were still unsatisfactory (Bayraktar, 2009).

As mentioned previously, current physics teaching and learning practices seemed to have a positive effect on improving the ability of pre-service physics teachers to understand physics concepts, although this improvement was still relatively small. It might be reasonable to say that the understanding of physics concepts has not been sufficiently emphasised and promoted in the pre-service physics teachers’ study programs involved in this research. The results of the qualitative study also indicated that physics questions in examinations were mostly related to traditional problem-solving that used physics formulae or the derivation of physics formulae, rather than testing for students’ understanding of physics concepts. Undoubtedly, students tended to memorise physics formulae and undertake practice in solving numerical questions from textbooks in order to pass the exams without trying to understand the underlying physics concepts. It seems reasonable to say that longer periods of study by pre-service physics teachers at university would not necessarily guarantee that they would gain a deeper understanding of physics concepts. As asserted by Hairan et al. (2019), the traditional physics teaching approach that places an emphasis on mathematical equations and physics problem-solving does not significantly promote students' understanding of physics concepts.

In the previous literature, there has been no comprehensive research conducted on the role that education level or year level has on pre-service physics teachers' understanding of physics concepts or their skills in thinking scientifically, in the Indonesian context. Thus, the results of this study provide important insights for educators in teacher education programs to fully support the development of pre-service physics teachers' scientific thinking and physics conceptual understanding across the year levels. These are the teachers of tomorrow who will have the responsibility of constructing and developing effective teaching and learning practices for their students. Other factors that might contribute to the improvement of students' scientific thinking and understanding of physics concepts are described in the following sections of this chapter.

8.3 The Relationship between Aspects of Scientific Thinking and Physics Conceptual Understanding

This section describes the relationship between the various aspects of scientific thinking (consisting of epistemological beliefs, argumentation, and scientific reasoning) and physics conceptual understanding.

Scientific Reasoning and Physics Conceptual Understanding

The results of the SEM analysis presented in Chapter 6 indicated a strong and direct effect of pre-service physics teachers’ scientific reasoning on their physics conceptual understanding in relation to the mechanics and the electricity and magnetism topics. This finding implies that participants who have good scientific reasoning skills tend to be more proficient in understanding physics concepts and hold fewer misconceptions. The findings of this study are consistent with those of the existing research, which found that students’ scientific reasoning contributes to their understanding of physics concepts (Coletta & Phillips, 2015; Ding, 2014c; Nieminen et al., 2012; Pyper, 2012).

The participants interviewed in this study acknowledged that in order to understand a natural phenomenon, it is necessary to have good scientific reasoning skills, so that the results of their thinking related to the concept underlying the natural phenomenon can be accepted and understood in logical and reasonable ways. They indicated that if they were more skilled in scientific reasoning, it would be easier for them to understand physics concepts. For instance, a male participant, P19_A, pointed out that to understand the natural phenomena or physics concepts, it is important for learners to have good skills in scientific reasoning, instead of simply memorising concepts and formulae. The findings of this study suggest that in order to make sense of natural physical phenomena and understand physics concepts or theories, a systematic way of thinking is needed to generate correct conclusions about the phenomena or physics concepts being studied.

Several previous studies have highlighted that scientific reasoning consists of inductive or deductive thinking that represents the scientific ways of thinking needed to construct scientific knowledge, generate and test hypotheses, solve scientific problems using hypothetical–deductive reasoning, conduct experiments and evaluate observations, and draw valid conclusions based on the evidence (Effendy, Hartono, & Ian, 2018; Ibrahim et al., 2015; Lawson et al., 2000; Psycharis, 2013). All of these abilities are essential to the process of scientific inquiry when conducting scientific investigations in order to make sense of the natural phenomena being studied. The process of scientific reasoning is also closely related to the coordination of theory and evidence (Kuhn, 2010). According to Ibrahim et al. (2015, p. 97), the ‘self-generated responses’ or ‘spontaneous predictions’ resulting from students' thinking about a natural phenomenon is identified as a theory. Meanwhile, “evidence is defined as the justifying information or data that students provide to expand on their underlying theory.” In cases where coordination between existing theories and evidence or experimental results are inadequate, there is a need to conduct new observations or investigations in order to draw new inferences that can lead to the development of conceptual understandings and the elimination of misconceptions (Ibrahim et al., 2015; Kuhn, 2010). Therefore, students who have advanced scientific reasoning skills are more likely to have better conceptual understanding and hold fewer misconceptions regarding the observed physical phenomenon or theory, than those with lower scientific reasoning skills.

Argumentation and Physics Conceptual Understanding

As presented in Chapter 6, the SEM model shows that the argumentation skills of the participants were found to have an indirect effect on their physics conceptual understanding, which is mediated by a single construct, that is scientific reasoning. Meanwhile, the participants' skills in argumentation had a direct and positive effect on their scientific reasoning skills. This result implies that the participants who are more highly skilled in argumentation are also more likely to be more skilled in scientific reasoning. This is in line with the previous literature showing a positive relationship between students' ability in argumentation and their scientific reasoning skills (Acar et al., 2015; Heng et al., 2014; Lancaster & Cooper, 2015; Psycharis, 2013), revealing that promoting students' skills in argumentation enhances their skills in reasoning scientifically. As mentioned previously, students with high-level scientific reasoning skills are more likely to have a deeper understanding of physics concepts.

The qualitative findings of this study also demonstrate that the participants recognised the relationship between argumentation and physics conceptual understanding. More specifically, a female participant, P15_B, stated that skills in argumentation positively affected her understanding of physics concepts. Other participants revealed that in order to have good argumentation skills, it was necessary to have prior insight or basic knowledge, because they needed to know the topic or content knowledge being studied or discussed. In the same vein, some researchers have pointed out that students' prior knowledge should be taken into account when engaging them in the argumentation process (Çelik & Kılıç, 2014; Eskin & Ogan-Bekiroglu, 2013; Von Aufschnaiter, Erduran, Osborne, & Simon, 2008). In addition, the interview data in this study indicated that physics learning activities that engage students in expressing their thoughts or defending their opinions, and even in refuting the opposing opinions of others, encourages students to think scientifically in order to make sense of the world. The thought processes that occur, in turn, provide a better understanding of the physics concepts being discussed, because students who engage in argumentation need to construct arguments that are accompanied by logical and scientific explanations. The participants further acknowledged that in order to engage in argumentation, having only prior knowledge or experience was not enough. They also need to have the courage and confidence to convey their opinions or express disagreements with certain ideas that differ from their own.

Previous studies have investigated the effect of argumentation skills on students' conceptual understanding of scientific concepts (Aydeniz & Dogan, 2016; Çelik & Kılıç, 2014; Kaya et al., 2012; McNeill, 2011; Nurlatifah et al., 2018; Osborne et al., 2016; Sadler & Zeidler, 2005a; Sampson & Blanchard, 2012; Venville & Dawson, 2010). The research pointed out that students' engagement in argumentative practices provides opportunities for them to justify scientific claims about specific natural phenomena using appropriate evidence and assists them to construct scientific conceptions. Such practices could broaden students' understanding and strengthen their content knowledge. In addition, the practice of argumentation allows students to get close to the subject matter; hence, "it would be reasonable to expect enhanced student understanding" (Von

Aufschnaiter et al., 2008, p. 127). The research cited above implies that students' argumentation skills contribute to developing their conceptual understanding of scientific knowledge.

As noted previously, the SEM model in this study demonstrated a more complex connection between argumentation and the conceptual understanding of the participants. In other words, the relationship between participants' skills in argumentation and their understanding of physics concepts is not straightforward. Their skills in argumentation could increase students' understanding of physics concepts through the improvement of their scientific reasoning skills. The SEM model showed that the argumentation skills of the participants (in terms of making scientific argumentation and challenging argumentation) have an indirect effect on their understanding of physics concepts, especially in the topics of mechanics as well as electricity and magnetism. This effect was mediated by their skills in scientific reasoning in relation to conservation of weight and volume, proportional reasoning, control of variables, probability and correlational reasoning, and hypothetical-deductive reasoning.

The existing literature has pointed out that argumentation represents the process of thinking used to connect evidence to a claim in order to construct scientific explanations about certain phenomena (Kuhn, Kenyon, & Reiser, 2006; Osborne et al., 2016; Sampson & Blanchard, 2012). To generate complex arguments, students need to develop a sense of how to analyse the appropriate evidence critically to support their claims. Kuhn et al. (2006) further noted that in the process of argumentation, learners take an active role in conveying and defending their understanding of scientific phenomena to evaluate knowledge claims using appropriate evidence, constructing knowledge, using reason, and drawing conclusions in order to make sense of the phenomena being studied. Indeed, argumentation involves reasoning processes, and as noted by Buber and Coban (2017), argumentation is closely related to scientific reasoning. Hence, the results of this study can be interpreted to mean that students who have difficulty in evaluating the association between evidence and claims are likely to have difficulty with their ability to coordinate theories and evidence. Consequently, students who have low argumentation abilities tend to have low scientific reasoning skills. In addition, Heng et al. (2014) asserted that argumentation skills are the main elements of scientific reasoning skills and play an important role in helping students to understand scientific concepts. This suggests that the process of generating arguments is regarded as a set of skills that may enhance students' scientific reasoning skills which, in turn, help develop their conceptual understanding. As indicated in the previous section, this study indicates that students' scientific reasoning skills strongly influence their understanding of physics concepts.

Teaching and learning science in the classroom should engage students in practicing their skills in argumentation, which might involve expressing and defending opinions as well as revising ideas if needed, or refining their understandings based on the content knowledge being studied. In so doing, students have the opportunity to use their relevant existing knowledge, organise their thinking, and to justify their claims supported by scientific evidence which, in turn, helps them to enhance their argumentation skills and deepens their conceptual understanding of scientific knowledge. Undoubtedly, the practice of argumentation skills plays an important role in students' construction of scientific knowledge, which allows them to link new information to their prior knowledge, leading to the elimination of misconceptions (Çelik & Kılıç, 2014; Cross et al., 2008).

Epistemological Beliefs and Physics Conceptual Understanding

In terms of the relationship between epistemological beliefs and conceptual understanding, the qualitative findings of this study indicate that the majority of participants interviewed acknowledged that their epistemological beliefs could affect their ability to understand physics concepts. The participants also highlighted that their epistemological beliefs can lead them to be either passive or active learners. Passive learners are those who tend to believe in the information obtained from their instructors as being an undeniable truth, so they absorb such information without exploring it deeply. These beliefs reflect a low level of epistemological beliefs or naïve epistemological beliefs (Kirmizigul & Bektas, 2019). Meanwhile, active learners are those who tend to be encouraged to think further about the truth of knowledge or information received from their instructors by seeking more information through various sources to gain a deeper understanding of the knowledge or phenomena being studied. These beliefs reflect a high level of epistemological beliefs or sophisticated epistemological beliefs (Ekinci, 2017). A male participant, (P6_C), further revealed that the higher the level of his epistemological beliefs, the better he was in the understanding of physics concepts. Clearly, the results of this study indicate that pre-service physics teachers' epistemological beliefs contributed to the enhancement of their ability to understand physics concepts. This finding is consistent with previous studies such as those reported by Ding (2014c), Stathopoulou and Vosniadou (2007), and Franco et al. (2012), who found that students with high levels of epistemological beliefs showed a better conceptual understanding of physics compared to students with low levels of epistemological beliefs. According to Ding (2014c), students who viewed physics knowledge as a coherent set of ideas tended to learn physics by making sense of the phenomena and deepening their conceptual understanding more than those who viewed physics as disconnected pieces of information, facts, or concepts.

In the quantitative study, the SEM model in Chapter 6 demonstrated multiple relationships between epistemological beliefs and understanding of physics concepts. The epistemological beliefs of the participants comprising the structure of scientific knowledge, the nature of knowing and learning, real-life applicability, evolving knowledge, and the source of the ability to learn, were found to have an indirect effect on their physics conceptual understanding, which is mediated by two constructs, namely argumentation and scientific reasoning. In the same vein, Stathopoulou and Vosniadou (2007) recognised that the relationship between epistemological beliefs and physics conceptual understanding is not straightforward. Hofer and Pintrich (1997) also argued that students' epistemological beliefs indirectly affect their academic achievement.

The results obtained from the SEM analysis showed that the epistemological beliefs of the participants directly and positively influenced their ability in argumentation. Simply put, participants with more advanced levels of epistemological beliefs seemed to have better abilities in argumentation. The qualitative results in this study also support this quantitative finding revealing a relationship between epistemological beliefs and argumentation ability. For instance, a female participant, P15_B, argued that belief in the truth of physics concepts contributed to her ability to defend arguments or to refute the opinions of others that contradicted her understanding about certain scientific knowledge. These findings are consistent with previous studies indicating that most students tend to rely on their epistemological beliefs or their beliefs about how knowledge is obtained when they generate arguments or express their ideas (Ku et al., 2014; Mason & Scirica, 2006; Noroozi, 2018; Nussbaum, Sinatra, & Poliquin, 2008; Weinstock & Cronin, 2003). This suggests that students' ways of thinking about the knowledge and knowing process make a significant contribution to the quality of their skills in constructing more complex and integrated arguments, counter-arguments, or rebuttals when facing a conflict of opinion or argument with others. Hence, students who have more advanced levels of epistemological beliefs are more likely to have better abilities to convey and defend their opinions or ideas. In contrast, students who have less developed epistemological beliefs tend to be reluctant to engage in argumentation activities in their classes (Nussbaum et al., 2008).

In addition, the SEM model in Chapter 6 showed that the participants' epistemological beliefs had a direct and positive effect on their skills in scientific reasoning. This suggests that participants with higher levels of epistemological beliefs tend to have better scientific reasoning skills than participants with less developed epistemological beliefs. This finding is consistent with the previous research which found that more advanced epistemological beliefs are correlated with higher levels of scientific reasoning of learners (Hotulainen & Telivuo, 2014; Zeineddin & Abd-El-Khalick, 2010). These researchers noted that students with low levels of epistemological beliefs were more likely to experience difficulties in coordinating theories with evidence, which reflected their inadequate skills in reasoning scientifically. As mentioned previously, the participants' argumentation and scientific reasoning skills contributed to their ability to understand physics concepts. This can be interpreted to mean that high levels of argumentation and scientific reasoning skills are strongly associated with a high level of students' epistemological beliefs which, in turn, contributes to their ability to understand physics concepts. In other words, students' epistemological beliefs seem to increase their ability in understanding physics concepts through the improvement of argumentation and/or scientific reasoning skills.

As indicated in the literature, students with low levels of epistemological beliefs are known as naïve or novice believers, while those who have higher levels of epistemological beliefs are referred to as sophisticated believers (Kirmizigul & Bektas, 2019; Ku et al., 2014). Students with naive epistemological beliefs generally believe that physics knowledge is absolute and consists of discrete facts and formulae, that the ability to acquire knowledge is endowed at birth, and that knowledge is handed down by authority or teachers. In addition, they tend to learn physics by rote memorisation. On the other hand, students with more sophisticated epistemological beliefs are those who believe that physics knowledge is a coherent system of ideas, representing a unified whole, highly-interconnected; is knowledge that is reasoned out through objective and subjective means; that this knowledge is tentative and evolving; that learning is a gradual process; and that the ability to learn is acquired through experience. They tend to learn physics by deepening their understanding and sense-making about certain phenomena (Ding & Zhang, 2016; Elby, 1999; Hammer & Elby, 2003; Schommer, 1994b). Hence, it is reasonable to say that students with sophisticated epistemological beliefs may have more advanced thoughts and sophisticated ways of thinking about knowledge that has the potential to promote their ability to express opinions or ideas, generate more complex arguments, and enable them to conduct scientific inquiry which, in turn, could act to enhance their understanding of scientific knowledge, such as physics concepts. With regard to teaching and learning practice, the existing literature indicates that naive epistemological beliefs are associated with traditional teaching or teacher-centred approaches where teachers have a dominant role in transferring content knowledge to students in the classroom. Meanwhile, sophisticated epistemological beliefs are related to constructivist teaching or student-centred methods where students are encouraged to construct knowledge, to be actively involved in learning, and to solve problems using scientific methods that lead to a deeper level of understanding (Ekinci, 2017; Khine, 2010; Kirmizigul & Bektas, 2019).

Epistemological beliefs should not be ignored in the classroom because they have positive effects on students' learning and in developing students' scientific knowledge and skills. Hence, pre- and in-service teachers should have sophisticated epistemological beliefs because these beliefs could potentially influence them in determining teaching methods, classroom management techniques, and evaluation strategies adopted in their teaching and learning practices in the classroom (Aslan, 2017; Ekinci, 2017; Kirmizigul & Bektas, 2019). In turn, pre- and in-service teachers who hold sophisticated epistemological beliefs might help their students to develop more advanced levels of epistemological beliefs, improve their ability to generate arguments, and use scientific reasoning skills to gain a better understanding of scientific knowledge.

To the best of the researcher's knowledge, the relationship between multiple variables consisting of epistemological beliefs, argumentation, scientific reasoning skills, and physics conceptual understanding has been investigated separately in the existing literature, where the measurement has relied mostly on the analysis of single variate or bivariate relationships. The present study investigated these multiple variables in an integrated way in order to acquire a more comprehensive understanding. Hence, this study contributes to the research literature by demonstrating how students' argumentation and scientific reasoning skills can be predicted through their epistemological beliefs. This study also indicates that the relationship between students' epistemological beliefs and their understanding of physics concepts is not straightforward but is mediated by their skills in argumentation and scientific reasoning. The following section outlines the participants' perceptions of the teaching and learning factors that might influence their scientific thinking and physics conceptual understanding.

8.4 Pre-Service Physics Teachers' Perceptions of Physics Teaching and Learning Factors that Influence Their Scientific Thinking and Physics Conceptual Understanding

Teachers play a vital role in effectively enhancing students' academic performance. Therefore, to become qualified teachers, prospective teachers must have adequate academic knowledge regarding the subjects they will be teaching in the future, as well as the necessary thinking skills required during their undergraduate education. Certainly, prospective teachers face various difficulties in acquiring adequate scientific knowledge and the necessary thinking skills. This section outlines pre-service physics teachers' perceptions of their classroom experience regarding the opportunities and barriers to promoting their skills in thinking scientifically and developing their physics conceptual understanding. Participants' responses in face-to-face interviews indicated that there were some aspects of physics teaching and learning practices that could have an impact on their scientific thinking and understanding of physics concepts, which were not investigated in the present quantitative study. The factors that have been taken into account include the facilities and learning resources, teaching methods, learning media and physics materials, learning activities in the classroom, homework or assignments, approaches to learning and learning assessment, as well as internal personal barriers experienced by participants and the external factors that might influence their learning. In so doing, this research provides more information related to the potential factors that might influence students' improvement of their scientific thinking and understanding of physics concepts.

Facilities or Learning Resources

In order to achieve academic goals, a university needs to provide learning resources for students that facilitate their learning and enhance their academic achievement. These learning facilities might include library facilities such as textbooks, learning materials, hand-outs, and technology such as the Internet network. Laboratory facilities that are equipped with scientific materials and experimental equipment are also important, especially for students in the field of science. Using the laboratory facilities provided by the university, students should have the opportunity to carry out inquiry activities through experiments which, in turn, can enhance their scientific thinking skills and help them acquire deep understanding of scientific concepts. In addition, the role of the instructor is important in encouraging students to optimally utilise the various learning resources that have been provided on campus.

In the qualitative study, a male participant, P11_A revealed that the university had provided adequate facilities or learning resources such as books in the library, an Internet connection, and laboratory equipment that could support students' learning and help them in understanding physics concepts. Undoubtedly, the availability of the Internet as a source of student learning is very important for acquiring knowledge and information. Textbooks are another student learning resource that students can use to gain knowledge and information about the subject matter being studied. Students can use textbooks and the Internet to support their learning, requiring that universities should facilitate good Internet connections and provide a variety of textbooks needed by students. Laboratory equipment is no less important to help students better understand the physics world. Consistent with the results of the interviews carried out in this study, some researchers have emphasised the important role played by a number of learning resources and facilities such as laboratory equipment, textbooks, libraries, and the classroom climate for promoting students' learning performance and academic achievement (Adeyemo, 2012; Ayaz et al., 2017; Daluba, 2012; Galarpe, 2017; Husnaini & Chen, 2019; Mbunde, 2017; Owoeye & Olatunde Yara, 2011; Sobremisana, 2017). For instance, Galarpe (2017) pointed out that the ready availability of laboratory equipment is beneficial for students to help them visualise phenomena by conducting scientific investigations to promote scientific thinking and foster scientific inquiry skills, as well as to help them better understand physics theories. In addition, the availability of adequate library facilities that are a source of knowledge and information plays an important role in supporting student learning and fostering student interest in reading (Ayaz et al., 2017). Students who are accustomed to using library facilities are likely to have better academic achievement compared to those who use them less, because they are likely to be more motivated to focus on learning, or the reading of various learning resources to broaden their horizons and better understand the concepts being studied.

Meanwhile, participants in the present study also indicated that the facilities or learning resources available to them were very limited and underutilised. Indeed, these barriers could hamper their efforts to go deeper into the physics materials and detract from their ability to conduct experiments independently. In other words, the poor quality or lack of learning resources or facilities, including not having adequate and current books or laboratory equipment, or having a slow Internet connection, are considered to be barriers to student learning, and this could contribute to their poor performance in physics. Furthermore, having non-conducive learning environments, such as having high temperatures in the classroom, and large class sizes or over-crowded classrooms, can affect students' motivation and enthusiasm to focus on their learning. Indeed, having a large number of students in one class may distract students' focus on learning and influence their academic outcomes. In such situations, instructors can find it difficult to provide personal attention to all their students making it difficult to implement instructional strategies appropriately and effectively. Ultimately, instructors are likely to implement traditional teaching methods when delivering physics learning material in large classes, rather than employing other teaching practices that place more emphasis on active student learning that benefits the learning process. As pointed out by Nepal & Maharjan (2015), comfortable classroom temperatures are essential for effective learning and teaching processes to occur for both students and teachers. Lack of adequate facilities such as heating or air conditioning in the classroom and having over-crowded classrooms, affects students' ability to concentrate in lessons which, in turn, can contribute to the low quality of their learning. As stated by Adeyemo (2012), over-crowded classrooms have a negative effect on students' academic performance. This could be due to noise that not only distracts students' interest and focus in learning, but also affects the teacher's effectiveness in presenting subject matter to students in class. Certainly, inadequate facilities and learning resources have a negative effect on students' interest in learning which, in turn, affects their academic performance (Daluba, 2012).

The results of this study imply that inadequate learning resources or lack of adequate teaching facilities and non-conducive learning environments experienced by participants presented barriers to their learning process. In turn, this may affect their motivation to explore various sources and conduct scientific investigations in order to foster their thinking skills and conceptual understanding of scientific knowledge. Furthermore, the lack of support from instructors to encourage students to utilise the various learning resources provided by their universities could also contribute to students' lack of motivation for learning which, in turn, could affect their academic performance. Hence, improving facilities and learning resources on campus may offer more opportunities to improve student learning outcomes.

In the case of this study, the researcher does not draw the conclusion that one type of university has better facilities than the other type. Participants' perceptions about the availability of facilities or learning resources provided by their universities seemed to vary. For instance, even though participants P11_A and P3_B were from the same university type, a public university, they had different perspectives on the availability of laboratory equipment at their university. It seems that the participants' responses in the interviews relied on the extent of their understanding about the availability of the facilities and learning resources that had been provided by their university, as well as their experiences in using the facilities. It also depended on their motivation to use the facilities at their university, as revealed by participants P3_B and P8_C. Therefore, further exploration is needed to obtain a comprehensive picture of the availability of facilities and learning resources at both public and private universities.

Teaching Methods

The participants interviewed in the present qualitative study revealed that the instructors tended to implement traditional lecturing or teacher-centred methods when presenting physics learning material in class. Other teaching practices that were sometimes implemented by lecturers were inquiry-based teaching methods, demonstrations, project-based learning, presentation methods, and group discussions.

By and large, the participants recognised that traditional teaching methods that simply transfer content knowledge and information through one-way communication or direct guidance were useful to help them acquire a good deal of knowledge or learning material, including the collection of formulae and the derivation of physics formulae. In the same vein, Dervić, Glamočić, Azra, and Mešić (2018) reported that researchers have asserted that teacher-centred methods were more effective than student-centred methods. Kirschner, Sweller, and Clark (2006, cited in Dervić et al., 2018, p. 289) also asserted that "… guided instruction is generally more effective compared to minimal guidance approaches." They argued that teacher-centred methods provided an opportunity for students to ask questions and receive answers immediately from the instructor concerning particular content knowledge they did not understand. Another study also demonstrated that implementing a teaching method such as interactive simulation-based learning required teacher guidance due to the fact that students were less familiar with simulations (Dervić et al., 2018). In this regard, implementing traditional teaching methods seemed more appropriate than the student-centred approach when the instructors need to transfer large amounts of knowledge and information to students, in addition to the use of computer-based instruction. Further to this, it is understandable that for physics subjects that consist of theories, principles, and laws, as well as a collection of formulae, it can be difficult and may require a lot of time for students to understand these ideas, particularly if they need to derive a variety of physics formulae without direct guidance from their teachers.

Other participants mentioned that several teaching methods, such as inquiry-based approaches and hands-on methods or laboratory work, encouraged them to be more actively engaged in the learning process, practiced solving problems, and prompted them to use their thinking skills and develop their curiosity about the scientific phenomena being studied. As stated in the constructivist learning theory, students play an active role in learning. In a constructivist environment, students are encouraged to be actively engaged in practicing complex thinking, such as problem solving and scientific thinking skills (Bada & Olusegun, 2015), to make sense of the world through investigations (Veletsianos, 2016).

In addition, through class presentations and discussions, they were more encouraged to practice speaking in class to convey or defend their ideas or opinions. The participants' perceptions indicate that implementing these teaching practices in the classroom offered many opportunities for them to engage more actively in developing higher-level thinking skills and to have rich learning experiences. Such teaching and learning practices also allowed them to engage in self-paced learning to gain a better understanding of the physics materials. However, these teaching practices seem to be implemented rarely in the classroom compared to traditional teaching methods.

Furthermore, the participants acknowledged that there was a great deal of physics material that needed to be covered in a small amount of time. This might be why the instructors tended to implement traditional teaching methods that do not require a long time both in the teaching preparation and the transferring material in the classroom as compared to other non-conventional teaching approaches. This implies that traditional lecturing makes it easy for the instructors to convey a large amount of learning material and information to students, and requires only a small amount of time, compared to other teaching methods that better engage students in taking an active role in their learning process. In the traditional classroom, the instructors tend to convey knowledge and information by writing material on the blackboard or using PowerPoint slides. This method allows students to listen passively to the instructors' explanations, to simply absorb information, and to take notes, which provided few opportunities for the students to practice their thinking skills. Therefore, the participants demonstrated that traditional teaching methods did not help them to enhance their thinking skills or to promote their understanding of physics concepts.

To sum up, the findings of this part of the study suggest that traditional teaching methods are useful for helping participants acquire a large amount of learning material related to the collection of formulae and the derivation of physics formulae. However, the teacher-centred approach is considered less effective in developing participants' thinking skills and their physics conceptual understanding. Previous research supports this finding indicating that traditional lecture methods did not significantly promote students' thinking skills, problem-solving, or their conceptual understanding (Adeyemo & Babajide, 2014; Adolphus & Omeodu, 2016; Bigozzi et al., 2018; Hairan et al., 2019; Sobremisana, 2017; Usmeldi, 2016). The lack of facilities such as laboratory equipment and teaching tools could explain why instructors preferred to implement lecturing methods and simple demonstrations in class. This approach could be contributing to the low skills of students, especially in the science field (Putri & Rusdiana, 2017). Sobremisana (2017) further pointed out that students' poor performance and misconceptions in physics can be caused by instruction that is based entirely on memorisation and limited interactions with experimental tools in the laboratory. The lecturing method seems to be unable to bring most students to a better understanding of the physical world. Students may find it difficult to understand certain physics concepts and to enhance their problem-solving abilities. Therefore, teaching methods that place more emphasis on memorisation of theories and facts appear to do little to help students develop their thinking skills and overcome their misconceptions in physics.

Ideally, the teaching methods applied in class should encourage students to learn in a way that promotes the long-term retention of physics concepts. Therefore, it is important for instructors to ensure that the teaching methods they use in class not only transmit knowledge, but also help students to gain meaningful learning experiences. The methods used should help students to overcome misconceptions and increase their understanding of physic concepts, as well as equip them with the skills needed to support their learning. The results of this study contribute to the understanding of current teaching practices in teacher education programs in Indonesia, and their effect on pre-service teachers' scientific thinking skills and understanding of physics concepts. However, to confirm these findings, further research is needed to obtain a broader picture of the teaching methods implemented in the classroom, by involving the instructors in the research.

Learning Media and Physics Material

Learning media as tools for transferring learning materials in the classroom are also taken into account as an aspect that might have an impact on scientific thinking skills and the understanding of physics concepts of the participants in this study. The findings from the interviews with the participants indicated that their instructors used various types of learning media to support their teaching practices. In classroom practice, instructors tend to use learning media such as chalk and a blackboard, and PowerPoint presentations. However, the participants interviewed indicated that these learning media did not encourage them to pay attention to the physics material presented by the instructors in class and did little to practice their thinking skills. In addition, participant P17_C revealed that such learning media provided little stimulation to students to actively ask questions and express opinions or ideas, leading to passive learning, so they were easily bored and unmotivated.

Other participants expressed the view that the use of demonstration aids, animations, or learning videos was more appealing and provided them with more learning opportunities to actively engage in the use of scientific thinking skills in order to gain a better understanding of physics phenomena. The participants' perceptions also suggested that if these types of learning media were adopted (i.e., animation and learning videos), it would positively affect their motivation to observe and deeply explore scientific phenomena. Participants' perceptions in this study are consistent with the results of the previous research showing that instruction assisted by learning media, such as learning videos and animations, can improve students' learning outcomes (Eguabor & Adeleke, 2017; Park, 2019). These researchers pointed out that multimedia consisting of collections of texts and visual information such as pictures and diagrams offered benefits for students to better understand physics concepts, which they often considered to be abstract and difficult to understand. Applying computer-based learning media that integrates audio-visual media may assist students to improve their ability to learn independently. Supporting this view, Jian and Wu (2015) highlighted that the use of pictures and diagrams can help students process information and get a better understanding of the subject matter being studied, rather than the use of texts alone.

According to Sobremisana (2017), instructors need to be innovative in helping students to understand the learning material and choosing appropriate teaching and learning aids that motivate and stimulate students' interest so that they can gain physics knowledge and skills effectively. These findings indicate that the instructors were using integrated learning media such as PowerPoint slides, animations, learning videos, and demonstration tools when conveying learning material in class in order to help students make connections between theory and the real physics world. The use of animations and learning videos could also provide opportunities for students to explore the learning material more independently. However, these types of learning media were rarely used by instructors as aids for presenting physics learning material in class, so the participants had only limited opportunity to practice their thinking skills. This suggests that learning media should be considered as another learning variable that may affect students' skills in thinking scientifically and in mastering physics concepts.

The results of the data analysis indicated that scientific thinking skills are needed to better understand the **physics material** that covers all aspects of nature and the relationships between the concepts of a specific phenomenon. This helps students to make sense of the natural phenomena around them instead of simply memorising physics materials.

The participants in this study recognised that studying physics allows them to connect with, and to be open-minded about the real world. However, they mentioned that the physics materials delivered in the classroom were more centred around the derivation of physics formulae and ways to solve the physics problems in the textbooks, rather than physics concepts or physics applications in the real world. Hence, the participants found it difficult to relate the physics materials covered in class to physics phenomena that occur in daily life, and this causes the learning process to be less effective and meaningful than it could be.

The results of this study are consistent with the findings reported by Kim and Pak (2002) that students were unable to overcome difficulties in their understanding of physics concepts, even though they had solved a large number of physics problems presented in their traditional physics textbooks. The findings of this study imply that if the content of the physics materials delivered by instructors in class is mostly related to the collection of physics formulae or derivations of physics formulae, this does not significantly promote students' thinking skills or their understanding of physics concepts. In addition, lots of practice in solving a large number of traditional physics problems where students tend to simply ‘plug-and-chug’ numbers into physics equations, does not necessarily overcome students' misconceptions in physics, and might have only limited effects on their conceptual understanding.

Learning Activities in the Classroom

Teachers are authorised to organise all activities and make a vital contribution to promoting effective teaching and learning practices for all students in the classroom in order to provide meaningful learning experiences for their students. In terms of learning activities, the findings indicated that various types of learning activities took place in the classroom. These learning activities included demonstrations and class presentations, small group discussions, simple experiments, as well as independent research projects, which provided opportunities for students to explore in-depth physics materials, to share ideas, and to practice their thinking skills. As presented in Chapter 7, the following participants stated:

In addition, a male participant, P1_A, indicated that in order to create effective teaching and learning practices, the active role of students to be engaged in learning activities, and instructors in providing feedback or supervising students must be present in class. However, the participants asserted that such learning activities were rarely carried out in class. The learning activities that took place in the classroom were dominated by transferring content knowledge by the instructors, and the students tended to receive the lectures passively, take notes, and solve example problems. Having a large amount of learning content to cover and time constraints were given as the reasons why the instructors tended to dominate the classroom and adopt the practice of simply transferring knowledge through one-way communication. This type of learning activity does not seem to provide many opportunities for students to be actively engaged in the learning process.

The previous research has shown a number of barriers experienced by teachers to implementing learning activities that encourage the active participation of students in the classroom. For instance, Kaya et al. (2012) pointed out that teachers face difficulties in managing and facilitating student learning activities such as class discussions or working in small groups, due to the fact that they must also consider the amount of learning material that needs to be covered by students and the associated time constraints. Hence, instructors or teachers tend to stand at the front of the class transmitting knowledge and information, while students spend most of their time taking notes and listening to the teachers' explanations as passive learners. Indeed, such teaching and learning practices provide limited opportunities for students to practice their thinking skills (Heng et al., 2014). In other words, learning activities that are more focused on the receiving and absorbing of content knowledge are likely to encourage students to spend more time memorising knowledge and physics formulae and practicing the solving of traditional physics problems in textbooks, instead of practicing their thinking skills to understand physics concepts.

On the other hand, learning activities that engage and increase student participation in the learning process tend to motivate and enable them to think more about the concepts. A number of previous researchers have stated that learning activities involving student participation in learning such as demonstrations, and working in small groups or group discussions are more attractive to students for learning, and that such activities lead to deeper conceptual understanding (Adolphus & Omeodu, 2016; Freeman et al., 2014). Learning activities that employ group discussions allow students to construct knowledge, convey their understandings, evaluate other understandings, and help them to refine their own understandings, if needed (Berland & McNeill, 2010). Thus, such learning activities permit students to understand concepts that are more deeply related to the material being discussed and have the potential to enhance their skills in argumentation and reasoning. In other words, classroom learning activities that engage students in taking an active role in the learning process may contribute to their ability to master physics concepts as well as to think scientifically about the physics materials.

The findings of this study suggest that learning activities that are dominated by transferring content knowledge by the instructors, in which students tend to receive lectures passively, take notes, and solve example problems, provide limited opportunities for them to be actively engaged in the learning process and to practice their thinking skills in order to enhance their understanding of physics concepts. Conversely, learning activities that emphasise the active role of students in the learning process are more likely to provide opportunities for them to explore in-depth physics material; however, this can inhibit the instructors from covering all the required learning material by the students. This implies that learning activities are one aspect that might affect participants' scientific thinking skills and their understanding of physics concepts in this study.

Homework

Homework or assignments given by instructors might be intended to provide opportunities for students to further explore the learning material in order to acquire a deeper understanding of the content knowledge. The results of this study indicate that instructors provided homework or assignments to students that needed to be completed either individually or in groups. This seems to be one of the strategies implemented by instructors with the aim of providing opportunities for students to overcome their difficulties in understanding the physics material being studied in class. Through doing homework, students were expected to acquire a better understanding of the physics material. Participants acknowledged that having appropriate homework could encourage them to better understand physics material and could foster their development as independent learners outside of school hours.

The perceptions of the participants in this study demonstrated that homework was centred around traditional physics problems. In addition, getting a large amount of homework without valuable feedback from the instructors did not necessarily help them to improve their understanding of physics concepts due to the fact that they were not sure whether the homework they had completed was correct or not. This view was given by participants P3_B and P7_B. There is no denying that checking students' homework and providing critical feedback on it or other assignments involves a lot of time for the instructors, so they are likely to provide little feedback on it. The lack of instructor feedback might, in turn, cause the students to complete their homework by copying answers from other students and to fail to take the homework seriously, leading them to submit their answers without understanding the material. Under these circumstances, it seems reasonable to say that homework offers few academic benefits for supporting student learning.

The existing literature suggests that homework plays an important role in students’ learning and affects their academic achievement (Buijs & Admiraal, 2013; Grodner & Rupp, 2013; Gu & Kristoffersson, 2015; Suárez et al., 2016). Researchers have pointed out that homework is beneficial for students, giving them the opportunity to consolidate their understanding of the material that has been covered during class, and involves them in thinking deeply about the content knowledge being covered. In addition, the research has found that homework has a positive effect on student learning when the teachers provide feedback. Feedback on homework is important because it allows students to review the material they have learned in class, improve their understanding, and strengthen their knowledge, as well as providing opportunities to practice learning independently. This autonomy is important because, as noted by Eveline et al. (2019), student independence in learning has an impact on their academic achievement.

To sum up, the findings of the present study highlight several issues raised by the participants in the context of homework completion or assignments. For instance, instructors gave students a large amount of homework, so the students found it difficult to manage their time to complete their homework and review the physics material that had been delivered in class. In addition, the types of questions or problems presented in their homework or assignments (e.g., solving physics problems) failed to encourage them to practice their skills in thinking scientifically and gain a deep understanding of physics concepts. It had little impact on improving their motivation to learn independently. The participants also argued that the lack of valuable feedback from instructors negatively affected their motivation and seriousness in completing their homework or assignments. This implies that homework should be considered as one aspect that can influence participants' scientific thinking skills and their understanding of physics concepts in this study.

Approaches to Learning and Learning Assessment

The results of the analysis indicate that the participants mentioned various learning strategies implemented by them to understand the physics material. As presented in Chapter 7, their **learning approaches** included memorising formulae, working on problem-solving questions using formulae and mathematical calculations, attempting to understand physics concepts by using pictures or diagrams, conducting experiments, discussing physics material in a study group, exploring learning videos, and actively engaging in physics study clubs.

The participants revealed that the learning strategy they relied on most was memorising physics material and solving traditional physics problems to prepare for examinations. Learning strategies that focus on memorising facts or theories and solving physics problems by simply plugging and chugging numbers into formulae could be referred to as surface learning approaches. As indicated in the previous literature, students who adopt a surface learning approach tend to memorise content knowledge and physics equations in their learning, have a fear of failure when in examinations, spend less time exploring the material in-depth, as well as making less effort to strengthen or refine their prior knowledge (Baeten, Kyndt, Struyven, & Dochy, 2010; Momsen et al., 2013). In turn, they are more likely to experience difficulties in connecting and applying physics materials to their real life.

Other participants acknowledged that learning independently, e.g., by conducting experiments, exploring learning videos and various sources from the Internet, attending physics study clubs, and engaging in group discussions provided them with opportunities to understand physics materials deeply and to practice thinking skills, compared with learning approaches that emphasised memorisation of the materials. They also revealed that discussing physics material in a group provided more chances for them to share knowledge and information as well as to enhance their ability to generate arguments or convey their ideas. Such learning strategies adopted by the participants enabled them to explore and go deeply into physics material, which enhanced their understanding of the physics concepts being studied. This type of learning approach can be referred to as a deep approach to learning. As summarised by Baeten et al. (2010) and Momsen et al. (2013), students who adopt a deep learning approach are more likely to be interested in exploring the material to acquire knowledge and develop their conceptual understanding as they use strategies to link the facts or ideas with the available evidence. It also helps them develop their ability to think critically. The deep learning approach is relevant to what is suggested in the constructivist learning theory, which states that students actively construct their understanding to acquire advanced knowledge (Bada & Olusegun, 2015). Thus, students could go beyond formulaic solutions, and become active learners who are able to build a meaningful connection between new knowledge and prior knowledge and experience.

The findings of this study imply that approaches to learning adopted by students may be either surface or deep learning approaches. Participants who tended to implement a surface learning approach did little to make a positive contribution to enhance their understanding of physics concepts and their skills in thinking scientifically, compared with those who implemented a deep learning approach. This suggests that approaches to learning adopted by students are related to their academic performance. In addition, Ogbeba, Odoh, and Adeke (2014) noted that students' approaches to learning affected their academic achievement.

Further to this, **assessment of learning** is another crucial consideration when investigating teaching and learning practices that take place in the classroom. Ideally, the tests developed by the instructors should be aligned with the knowledge and skills that are expected to be achieved by the students. In this study, the perceptions of the participants demonstrated that the types of test questions given in their mid-term and final exams were more about problem-solving questions that use physics formulae and mathematics calculations, instead of measuring the students' understanding of physics concepts or thinking skills. Working on questions or tests that emphasise the use of formulae and calculations, appears to encourage students to spend a lot of time memorising a collection of equations and solving traditional physics problems from textbooks in order to simply pass the test and achieve high scores, rather than furthering their understanding of physics concepts. The participants also mentioned that students obtaining high scores in the subject of physics did not necessarily have high ability in understanding physics concepts. The participants' perceptions further indicated that the reason for instructors tending to provide tests that mostly centred around traditional physics problems was that such questions were likely easier to create than developing items that measure students' understanding of physics concepts. This suggests that the types of test questions developed by the instructors influences the learning strategies chosen by the students.

The existing research literature supports this finding, reporting that the types of questions or tests used has a significant effect on students' learning (Kibble, 2017; Momsen et al., 2013). According to Baeten et al. (2010, p. 252), "… being successful in terms of assessment does not always require a deep approach to learning." This implies that the types of test questions used in the assessment of students' learning given by the instructors are likely to be related to the approach to learning employed by their students. For instance, the types of test questions that are more centred on problem-solving that use physics formulae and mathematics calculations might not require a deep learning approach, compared to types of tests that measure students' skills in higher-level scientific thinking and their understanding of physics concepts. Previous studies have also highlighted that students tend to use a surface learning approach when memorising facts and formulae, and plugging in numbers where the assessment measures basic comprehension and knowledge, as well as the ability to answer straight-forward problem-solving questions (Ciara, 2009; Momsen et al., 2013). Consequently, for the purpose of passing examinations, students might not need to employ higher-level thinking skills and in-depth understanding of physics concepts.

In this study, physics tests developed by instructors and distributed in class did not fully motivate the participants to enrich their understanding in terms of the physics knowledge and thinking skills they would need as future physics teachers. Therefore, instructors should consider the knowledge and skills needed by prospective physics teachers when creating test instruments to assess learning outcomes. A good test instrument might be one that encourages students to adopt a deep learning approach instead of a surface learning approach. Certainly, to confirm these findings, further research is needed to obtain a broader picture of the types of test questions developed by instructors in assessing students' learning outcomes, by involving the instructors in the research. To sum up, teaching and learning practices implemented in the classroom, as well as the type of learning assessment used by instructors, seems to play an important role in the decision's students make about adopting learning strategies or learning approaches that are appropriate to supporting their learning.

Internal Personal Barriers and External Factors

Students' skills in thinking scientifically and their ability to master physics concepts are certainly influenced by many factors. The findings of this study reveal various internal and external barriers that are experienced by participants during the period of their studies at university that may have an impact on their academic performance. These barriers relate to students' misconceptions of physics, lack of prior knowledge and motivation, parental expectations, poor time management, and students being assessed by a standard evaluation system (i.e., National Examination), as mandated by the Indonesian government.

Based on the participants' perceptions, these barriers affect their interest in, and enthusiasm for, studying physics, implementing learning strategies, and setting learning goals which, in turn, hinders their academic achievement. For instance, students who were less motivated to learn physics revealed that they tended to memorise physics materials when preparing for tests without trying to understand physics concepts and acquiring meaningful learning. One reason given for the participants' lack of motivation to achieve meaningful learning in physics was that they had little interest in studying in the Physics Education Department. They tended to force themselves to do this because they wanted to meet the expectations of their parents due to considerations such as being able to get a better job in the future. For example, a female participant, P16_B, revealed that "Studying in the Physics Education major is not my passion. My parents hoped that I become a teacher because there is a significantly high opportunity to be a teacher, especially in rural areas like my place of birth …". In addition, the findings of this study show that students who have held misconceptions about certain physics concepts since their time in secondary school were more likely to continue to hold these misconceptions, despite studying at university. Hence, students were experiencing difficulties in integrating new knowledge or information with their prior knowledge, which hampered their efforts to develop their conceptual understanding (Treagust & Chandrasegaran, 2007; Tuder & Urban-Woldron, 2015). In terms of time management, a male participant, (P19_A), expressed the view that he had difficulties in managing his time effectively both when studying at university and when being actively involved in various activities outside the classroom, such as being involved in the organization and giving private tutoring to secondary school students. Poor time management led him to be a passive learner in the classroom.

In addition, the National Examination system adopted by the Indonesian government as a standard evaluation system is one of the factors that can also affect the type of learning approach adopted by students. The National Examination is a standard evaluation system to assess students' learning outcomes, particularly at the primary and secondary school levels (Kemendikbud, 2013). As revealed by a male participant, P14_A, he tended to memorise physics formulae to face the national examinations, which involve a lot of mathematical calculations and require speed to complete them. Therefore, it is not surprising that most students tended to simply memorise physics formulae and practice solving physics problems. The existing literature indicates that the national examinations also affect teachers' performance in implementing teaching methods in the classroom. According to Astuti and Retnawati (2017, p. 59), "teaching and learning methods used by teachers weren't considering students understanding." It is understandable that although students were studying at the university level, they were maintaining the learning strategies they had adopted when studying in secondary school. This could mean that they rely more on memorising subject matter instead of understanding concepts. The existing literature supports the findings of this study which reveals that misconceptions, lack of motivation and interest, parental expectations or involvement, and national exams affected students' learning (Astuti & Retnawati, 2017; Emerson et al., 2012; Guido, 2018; Guo et al., 2019; Liu, Bridgeman, & Adler, 2012; Sobremisana, 2017; Widiyatmoko & Shimizu, 2018).

To the best of the researcher's knowledge, there has been no previous comprehensive research conducted that explores the various barriers faced by Indonesian pre-service physics teachers in enhancing scientific thinking and their understanding of physics concepts, as found in this study. This finding suggests that enhancing students' academic performance, which includes scientific thinking skills and understanding of physics concepts, is not an easy task for instructors. The findings of this study contribute to the world of education in general, and for physics teachers in particular. There is also a message for Indonesian teachers and pre-service teachers. They need to pay more attention to the various learning barriers that are experienced by their students because these barriers are likely to have negative effects either directly or indirectly on their students' academic achievement. This is especially important for pre-service physics teachers because they are the physics teachers of tomorrow who have the responsibility of constructing and developing effective physics teaching and learning practices for their students.

To sum up, the results of the present study contribute to minimising the gaps identified in the prior research by investigating various factors that may affect pre-service physics teachers' scientific thinking skills and understanding of physics concepts, as well as the relationships between the various factors. The results also provide information about how important it is to have sophisticated epistemological beliefs, and high levels of argumentation ability and scientific reasoning skills for students. In turn, these aspects of scientific thinking skills are expected to support students' learning and deep understanding of physics concepts. Furthermore, the findings have revealed some of the barriers faced by Indonesian pre-service physics teachers as well as highlighting a number of opportunities to support their learning process in order to improve academic achievement. Hence, educators working in physics teacher education programs need to be aware of, and carefully address, the various problems experienced by the pre-service physics teachers identified in this study. Efforts are needed to minimise such barriers and to consider effective teaching and learning physics practices combined with an appropriate evaluation system that promotes students' scientific thinking skills (consisting of epistemological beliefs, argumentation, and scientific reasoning) and their understanding of physics concepts.

8.5 Summary

This chapter has presented the findings from both the quantitative and qualitative parts of the study in association with the existing literature regarding the various factors that may affect the scientific thinking skills and physics conceptual understanding of Indonesian pre-service physics teachers. To begin with, demographic factors, comprising university type, year level, and gender, were found to affect the participants' scientific thinking and physics conceptual understanding. The results of the study demonstrate that the male participants were more likely to have higher level scientific reasoning skills and abilities to master physics concepts than the female participants. In addition, participants from public universities were more likely to perform at a higher level in scientific thinking and physics conceptual understanding than participants from private universities. In terms of the year level, it was found that participants at higher year levels of study were more likely to achieve higher scores in their understanding of physics concepts. This finding implies that participants' demographic factors play an important role in developing their cognitive factors, such as skills in thinking scientifically and physics conceptual understanding. In addition, based on the SEM model and the qualitative findings, the aspects of scientific thinking, consisting of epistemological beliefs, argumentation, and scientific reasoning, were found to significantly influence conceptual understanding of physics of the participants.

This study suggests that these aspects of scientific thinking must be practiced and developed by pre-service physics teachers during the teaching and learning processes in the classroom in order to enhance their understanding of physics concepts. In addition, based on the participants' perceptions, other factors such as facilities and learning resources, the teaching methods, learning media and the physics materials, the learning activities employed in class, homework, approaches to learning, and the type of learning assessment employed have been identified as having an impact on pre-service physics teachers' scientific thinking and physics conceptual understanding. Internal and external factors such as misconceptions about physics, lack of prior knowledge and motivation, parental involvement and expectations, students having poor time management, as well as the standard evaluation system (i.e., National Examination) adopted by the Indonesian government were also found to influence the participants' learning. The results of the study indicate that the teaching and learning practices that take place in the classroom do not provide adequate support for pre-service physics teachers to develop their scientific thinking and to promote their understanding of physics concepts.

In fact, there has been no research that integrates and correlates the multiple variables involved in this study. Therefore, these findings address a major gap in the prior research by providing a comprehensive picture of the relationship between the demographic factors of the pre-service physics teachers, aspects of scientific thinking, and understanding of physics concepts, as well as aspects of physics teaching and learning practices. This study contributes to the area of research, especially in the field of physics education at the higher education level, by providing insight into how, and elucidating ways in which, university lecturers can enhance pre-service physics teachers' understanding of physics knowledge and skills in thinking. Both of these components are crucial for successful physics teaching among future physics teachers.

CHAPTER 9
CONCLUSIONS AND RECOMMENDATIONS

9.1 Overview

This concluding chapter highlights the key contributions made in the present study. It begins with the research aims, the design of the study, and summarises the important findings that answer the research questions investigated in this study. This is followed by the conclusions drawn from the study and an explanation of the implications. Finally, some limitations of the study are identified and recommendations for potential future research are presented.

9.2 Aims and Objectives of the Study

The present study sought to investigate several aspects of scientific thinking and the conceptual understanding of physics among pre-service physics teachers in the province of Central Java, Indonesia. Scientific thinking was described as comprising epistemological beliefs, argumentation, and scientific reasoning. Specifically, this study aimed to achieve the following objectives. These are to:

1. Understand the extent to which pre-service physics teachers' demographic factors such as gender, year level, and university type, influence their scientific thinking and conceptual understanding of physics.
2. Develop a model to examine the relationships between the identified demographic factors, scientific thinking, and conceptual understanding of physics among Indonesian pre-service physics teachers.
3. Empirically validate the proposed model developed in this study using a structural equation modelling (SEM) approach.
4. Identify other factors that may contribute to pre-service physics teachers' scientific thinking and conceptual understanding of physics by asking for their perceptions about the existing teaching and learning practices they had experienced during their studies in higher education.

9.3 The Design of the Study

As presented in Chapter 1, this study was generally concerned with the pre-service physics teachers' scientific thinking (consisting of epistemological beliefs, argumentation, and scientific reasoning) and their understanding of physics concepts. The study of the extent to which students' epistemological beliefs, argumentation, and scientific reasoning are rarely investigated in the Indonesian context. On the one hand, these aspects of scientific thinking are believed to contribute to the development of students' understanding of physics concepts (Coletta & Phillips, 2015; Ding, 2014c; Ding, Wei, & Mollohan, 2016; Nurlatifah et al., 2018). In addition, the extent to which the aspects of teaching and learning practices affect the improvement of students' scientific thinking skills and understanding of physics concepts have not been explored in any significant depth, especially at the tertiary level in Indonesia. Therefore, this study seeks to fill the gap in research on physics education, particularly in the teacher education institutions in Indonesia. The rationale for the implementation of this study is that there has been no comprehensive empirical research on the aspects of scientific thinking and physics conceptual understanding of prospective physics teachers, as well as the relationships between these variables or constructs and their demographic factors in the context of Indonesian higher education. Likewise, the extent to which factors related to teaching and learning practices in the university classroom contribute to the improvement of pre-service physics teachers' scientific thinking and physics conceptual understanding has been the subject of little research in Indonesia. This lack of research has constrained understanding about the extent to which each of these factors may contribute to the improvement of pre-service physics teachers' skills in scientific thinking and understanding of physics concepts.

Having the knowledge and skills particularly to think scientifically and to understand physics concepts is crucial for students to meet the demands of the 21st century, where there is a focus on higher-level thinking skills and conceptual understanding (Collins, 2014). Therefore, it is important for pre-service teachers to have well-developed scientific thinking skills and deep conceptual understanding regarding the subject of physics that they will be teaching in the future, because they are responsible for developing and organising effective teaching and learning practices that will help to enhance the physics knowledge and thinking skills of their students.

In this study, pre-service physics teachers' scientific thinking and their understanding of physics concepts was investigated by employing a cross-sectional mixed-methods design. The relationships between multiple variables, i.e., epistemological beliefs, argumentation, scientific reasoning, physics conceptual understanding, and several demographic factors comprising gender, year level, and university type were examined by implementing multivariate procedures in a quantitative study. The several paper-based questionnaires used to obtain responses from the study participants were the *Epistemological Beliefs Assessment for Physical Science* (EBAPS), developed by Elby, Frederiksen, Schwarz, and White at the University of California, Berkeley (The Idea Behind EBAPS, 2002); the *Argumentation Test*, developed by Sampson and Clark (2006); the *Lawson's Classroom Test of Scientific Reasoning* (LCTSR), developed by Lawson (2000a); the *Force Concept Inventory* (FCI) test, developed by Hestenes, Wells, and Swackhamer (1992); and the *Brief Electricity and Magnetism Assessment* (BEMA), developed by Chabay and Sherwood (1997). These surveys were completed by 706 Indonesian pre-service physics teachers from Year 1 to Year 4, coming from two public and two private universities. The validity and reliability of the instruments were established through the multidimensional Rasch model analysis employing ACER ConQuest 4 and Confirmatory Factor Analysis (CFA) using the IBM SPSS Amos 25, as presented in Chapter 5. Prior to analysing the quantitative data, raw scores were transformed into measures using the Weighted Likelihood Estimation (WLE) technique through Rasch analysis to achieve uniformity for a more valid interpretation of the results. The analyses of the descriptive statistics, the t-test, and the one-way Analysis of Variance (ANOVA) were carried out using IBM SPSS 25. In this study, the proposed model had four latent variables or constructs consisting of **Epistemological Beliefs**, **Argumentation**, **Scientific Reasoning**, and **Physics Conceptual Understanding** (**PCU**). The relationship between these multiple variables or constructs was analysed using the Structural Equation Modelling (SEM) approach by employing IBM SPSS Amos 25. The findings indicated that all goodness of fit statistical values showed a good fit between the model and the observed data. In other words, the final SEM model in this study fitted the data well, as presented in Chapter 6.

Equally important was the perception of pre-service physics teachers regarding the factors in teaching and learning practices that might have contributed to shaping their skills in scientific thinking and their ability to understand physics concepts. The participants' perceptions with regards to the relationship between aspects of scientific thinking and physics conceptual understanding were also explored (see Chapter 7) in the qualitative part of the study, in which interview questions were developed and face-to-face semi-structured interviews were conducted with 25 pre-service physics teachers. The NVivo 11 software was used as the data analysis tool to store, organise, and code the interview data. In addition, a thematic analysis was carried out to combine related codes to develop common themes and identify the connections between significant themes (Creswell, 2012).

In this thesis, the quantitative part of the study was emphasised more than the qualitative, and the qualitative research was carried out for the purpose of complementing or enriching the findings of the quantitative investigation. As recommended by Creswell (2014a), the qualitative part of the study was carried out to support and expand the quantitative findings. To achieve the goals of this investigation, the quantitative and qualitative data analyses were combined in order to obtain a richer picture of pre-service physics teachers' scientific thinking and conceptual understanding of physics.

9.4 Summary of the Findings

This section presents a summary of the main findings generated from the analysis of the quantitative and qualitative data in response to the research questions formulated in Chapter 1. Research question 1 (RQ1), "*Are there any differences in demographic factors with regards to pre-service physics teachers' scientific thinking and conceptual understanding of physics?*" was answered based on the t-test analysis reported in Chapter 5 and the SEM model analysis presented in Chapter 6. Specifically, this question was more centred on differences in the demographic factors consisting of pre-service physics teachers' gender, year level, and university type. In terms of the influence of gender on pre-service physics teachers' scientific thinking and physics conceptual understanding, the findings demonstrated that there were no significant differences between male and female pre-service physics teachers with respect to their epistemological beliefs and argumentation. This result suggests that gender was not a factor that has an influence on shaping pre-service physics teachers' epistemological beliefs and argumentation. The absence of a gender effect on these variables could be due to the teaching and learning practices in the classroom having little influence on promoting epistemological beliefs and practicing argumentation skills among participants. Another alternative possibility is that the epistemological beliefs held by male and female pre-service teachers were similar when they started their studies in higher education, but these beliefs shifted at about the same rate as a consequence of being in similar learning environments. However, the findings from the t-test analysis and the SEM model indicated that there were significant differences between male and female pre-service physics teachers with regard to scientific reasoning and physics conceptual understanding. Male pre-service physics teachers were found to perform better than females on both these variables. The gender gap found in this part of the study was likely influenced by various factors such as field of major study and stereotype threat. For instance, in natural science subjects, male students have been found to have better performance in physics, while female students performed better in biology, while no differences between female and male students in chemistry have been found (Yamtinah et al., 2017). In addition, stereotypes have been found to contribute to gender differences, particularly in mathematics and physics subjects (Li et al., 2017). Female students have been found to be more anxious in mathematics and science than male students which, in turn, contributes to poor performance in these fields (Henderson et al., 2017; Nieminen et al., 2013). Another possibility is that male students are more confident to participate, and be independent, in conducting experiments or inquiries into physics phenomena than female students. This in turn leads male students to having better scientific reasoning skills than female students if they carry out investigations in the laboratory or outside the laboratory more often. As noted in the prior literature, male students tend to be better in observing phenomena, controlling variables, and drawing conclusions (Yamtinah et al., 2017). However, the results of this study suggest that further research is needed to establish in detail the role that gender plays in developing epistemological beliefs, argumentation, scientific reasoning, and conceptual understanding of physics.

Another demographic factor investigated in this study was the year level of pre-service physics teachers with respect to their scientific thinking and conceptual understanding of physics. The ANOVA analysis reported in Chapter 5 and the SEM model analysis presented in Chapter 6 indicated that there were no statistically significant differences by year of study for pre-service physics teachers' epistemological beliefs, argumentation, and scientific reasoning. Meanwhile, year level was found to have a direct and positive effect on pre-service physics teachers' understanding of physics concepts. However, the improvement was small in their understanding of physics concepts over time. The findings imply that pre-service teachers' skills in thinking scientifically were unchanged or developed little through each year of their undergraduate education. It is possible that the teaching and learning practices taking place in their university classroom did not significantly enhance the scientific thinking of these pre-service physics teachers. As indicated in the qualitative results, the participants' perceptions showed that instructors tended to implement traditional teaching methods that simply transferred content knowledge and presented scientific facts to them through direct instruction. This traditional lecture method is considered to have little impact on enhancing the level of students 'epistemological beliefs (Zhang, Ding, & Mazur, 2017), and is seen to give little opportunity for students to practice their abilities in argumentation (Wang & Buck, 2016). It has also been found to be less effective in enhancing students' scientific reasoning (Ding, 2018). In other words, this direct instruction teaching method encourages students to passively acquire knowledge and memorise it, rather than actively participating in constructing knowledge and practicing their thinking skills. Likewise, traditional direct instruction classroom practices tend to encourage pre-service teachers to solve traditional physics problems rather than promoting their understanding of physics concepts. If this approach is used commonly, students are more likely to only memorise physics formulae and practice solving physics problems found in their textbooks. As a result, improvements in the conceptual understanding of physics would not be realised.

In addition, the t-test analysis reported in Chapter 5 and the SEM model analysis presented in Chapter 6 showed that pre-service physics teachers from public universities performed better in all aspects of scientific thinking and the conceptual understanding of physics in comparison to pre-service physics teachers from private universities. Pre-service physics teachers studying in public universities have passed through a strict selection process to be accepted to study at the university. Meanwhile, those who failed in such a strict selection process were more likely to enrol in a private university. Generally, pre-service physics teachers studying in public universities could be expected to have higher level academic knowledge and skills. Consequently, the findings of the study appear reasonable in demonstrating that pre-service physics teachers from public universities scored higher on all surveys used in this investigation compared to pre-service physics teachers from private universities.

Research question 2 (RQ2) was: "*What are the relationships between pre-service physics teachers' scientific thinking (i.e., epistemological beliefs, argumentation, scientific reasoning), conceptual understanding of physics, and their demographic factors?*" This question was answered using SEM analysis, as presented in Chapter 6. Meanwhile, research question 3 (RQ3), "*What are pre-service physics teachers' perceptions of the relationships between scientific thinking (i.e., epistemological beliefs, argumentation, and scientific reasoning) and their conceptual understanding of physics?*" was answered based on the qualitative analysis presented in Chapter 7. More specifically, these questions focused on the relationships between three aspects of scientific thinking (i.e., epistemological beliefs, argumentation, and scientific reasoning) and the understanding of physics concepts of pre-service physics teachers. The results indicate that the epistemological beliefs, argumentation, and scientific reasoning of these pre-service physics teachers influenced their ability to understand physics concepts. These findings imply that those who held sophisticated epistemological beliefs were more skilled in generating arguments, and had good scientific reasoning skills which, in turn, could lead to them being better at understanding physics concepts or to holding fewer misconceptions. Another key finding indicated in the SEM model was the complex relationships between aspects of scientific thinking and physics conceptual understanding. The model shows that participants' epistemological beliefs and argumentation have an indirect effect on their conceptual understanding of physics. Pre-service physics teachers' epistemological beliefs influenced their understanding of physics concepts mediated by their argumentation and/or scientific reasoning. In addition, the effect of argumentation on the participants' conceptual understanding of physics was mediated by their scientific reasoning skills. In other words, those participants who had sophisticated levels of epistemological beliefs were more likely to be more skilled in generating arguments or opinions and in defending their ideas. They were also more likely to be better skilled in reasoning scientifically which, in turn, could assist them to develop a deeper understanding of physics concepts.

Finally, the response to research question 4 (RQ4), "*What are pre-service physics teachers' perceptions of the opportunities and barriers in enhancing their scientific thinking and conceptual understanding of physics?*" was answered based on the qualitative analysis presented in Chapter 7. The results of this qualitative study indicated that there were several factors in physics teaching and learning that had an impact on shaping pre-service physics teachers' skills in scientific thinking and their ability to understand physics concepts. These factors related to the facilities or learning resources provided, the teaching methods employed, the learning media and physics materials used, the kind of learning activities undertaken in the classroom, the type of homework or assignments provided, the learning approaches adopted, and the learning assessment implemented. In addition, these pre-service physics teachers outlined various factors relating to internal personal barriers as well as external ones. The internal and external barriers mentioned by the participants consisted of misconceptions, lack of prior knowledge and motivation, parental expectations, poor time management, and the adoption of a standard evaluation system (i.e., National Examination). Among the factors considered most influential in the development of pre-service teachers' scientific thinking and conceptual understanding of physics were the teaching methods implemented and the learning media used by the instructors, as well as the type of learning activities that were undertaken in class. The learning assessments employed likely determined the type of learning approach adopted by pre-service physics teachers and this, in turn, may also have influenced whether or not they were afforded the opportunity to enhance their scientific thinking and their ability to understand physics concepts.

9.5 Conclusion

The present study investigated aspects of scientific thinking and physics conceptual understanding of pre-service physics teachers. This investigation has outlined an explanation of how the demographic characteristics of pre-service physics teachers influence their aspects of scientific thinking and physics conceptual understanding, and how these aspects of scientific thinking contribute to their physics conceptual understanding. The main findings of the study provide empirical evidence demonstrating that aspects of scientific thinking comprising epistemological beliefs, argumentation, and scientific reasoning are potential factors that can influence pre-service physics teachers' understanding of physics concepts, either directly or indirectly. Epistemological beliefs and argumentation were found to affect pre-service physics teachers' understanding of physics concepts indirectly, mediated by scientific reasoning. Meanwhile, scientific reasoning was found to directly influence pre-service physics teachers' understanding of physics concepts.

The SEM model developed in this study provides strong evidence of the effect of pre-service physics teachers' scientific thinking on enhancing their understanding of physics concepts. In addition, the SEM model indicated that students' epistemological beliefs are the first variable that must be addressed and developed when learning physics. Students with high levels of epistemological beliefs are more likely to be able to construct knowledge and be active during the learning process. Having sophisticated epistemological beliefs helps students acquire new knowledge through their learning environment and this allows them to change their views and opinions about the phenomenon being studied. In turn, this aids them in constructing knowledge and generating arguments. Furthermore, in the process of argumentation, students take an active role in conveying and defending their understanding related to a scientific phenomenon, constructing knowledge through the process of reasoning, and drawing conclusions to make sense of the phenomenon being studied (Kuhn et al., 2006). This suggests that argumentation involves reasoning processes. Hence, students with high skills in argumentation tend to be more skilled in reasoning scientifically.

This finding implies that in addition to helping students develop their levels of epistemological beliefs, instructors should facilitate their practicing of skills in argumentation. Learning activities that encourage students to convey ideas, generate arguments, and defend opinions, help them to develop their reasoning skills. Pre-service physics teachers' scientific reasoning might be enhanced by instructors facilitating and encouraging them to conduct scientific investigations to observe natural phenomena which could help them gain a better understanding of the theories or facts being studied. As also demonstrated in the SEM model, the scientific reasoning of pre-service physics teachers was found to strongly influence their ability to understand physics concepts. These results suggest that pre-service physics teachers with high levels of epistemological beliefs, argumentation, and scientific reasoning would be more likely to have a deep understanding of physics concepts. Therefore, practicing and developing epistemological beliefs, argumentation, and scientific reasoning should not be ignored by instructors, because these variables play an important role in developing students' understanding of physics concepts.

However, the results of this study indicate that the improvement of pre-service physics teachers in relation to their epistemological beliefs, argumentation, scientific reasoning, and physics conceptual understanding was still far from satisfactory. This finding raises questions related to the learning process experienced by the participants during their studies in tertiary institutions, which is, the extent to which aspects of scientific thinking and understanding of physics concepts have been presented and practiced in their university classes. Hence, this study also described the perceptions of pre-service physics teachers in relation to the practice of teaching and learning physics in class that might contribute to facilitating and enhancing their scientific thinking and understanding of physics concepts.

The results of the qualitative analysis carried out through interviews with pre-service physics teachers indicate that there are a number of important factors related to teaching and learning practices in the classroom that influenced participants' thinking skills and their understanding of physics concepts. These important factors concerned the teaching methods implemented in the classroom, and the type of examination questions developed by university instructors. The findings suggest that as long as instructors mostly implement traditional teaching practices and use test questions that are centred around the derivation of physics formula and the solving of traditional physics problems, pre-service physics teachers are likely to find it difficult to practice and enhance their skills of scientific thinking and their ability to master physics concepts. This is because the teaching approach and exam questions developed by the instructors would seem to encourage pre-service physics teachers to memorise facts and physics formulae, as well as ‘plug-and-chug’ numbers into mathematical formulae to solve a large number of traditional physics problems in order to pass their examinations. A learning approach that focuses on memorising facts and solving traditional physics problems reflects a surface learning approach (Momsen et al., 2013). Furthermore, pre-service physics teachers’ perceptions demonstrate that teaching and learning practices in their university classes have not fully developed their beliefs about knowledge and knowing. Likewise, their perceptions showed that the ability to generate arguments, share ideas, or defend opinions was rarely practiced in their university classrooms. In addition, learning activities carried out in class had little impact on improving their scientific reasoning skills.

It is concerning that these pre-service physics teachers would be likely to implement similar teaching strategies in their secondary classrooms in the future, because they might not have sufficient pedagogical knowledge, skills, or experience to be able to know how to teach the skills of scientific thinking that could enhance the physics conceptual understanding of their students. This implies that university instructors need to be careful to construct and develop proper teaching methods to be implemented in the classroom as well as to develop appropriate types of test questions in order to measure the knowledge and skills that need to be achieved by students. These instructors also need to assist students to practice their thinking skills. In so doing, this could make a positive contribution to improving skills in scientific thinking and promoting understanding of physics concepts by their students. The empirical evidence revealed in this study adds valuable information to the existing literature and theory in the field of physics education and suggests various implications that need to be considered by policymakers and practitioners in order to find solutions to enhance students' higher-level thinking that is needed and encouraged in the 21st century.

9.6 Implications of the Study

The investigation of pre-service physics teachers’ scientific thinking and conceptual understanding of physics has implications related to theoretical, methodological, and practical concerns.

Theoretical Implications

Improving educational outcomes for students in physics education has been a perennial problem for decades with mixed results (Bayraktar, 2009; Ding, 2014a; Hairan et al., 2019). This is no less true for students in Indonesia as it is for those in Western nations. For instance, this study provides empirical evidence that demonstrates that epistemological beliefs, argumentation, and scientific reasoning are potential factors that affect pre-service physics teachers' conceptual understanding of physics. However, the findings of this study indicate that there was no significant improvement in the aspects of scientific thinking (consisting of epistemological beliefs, argumentation, and scientific reasoning) and only little improvement of pre-service physics teachers' conceptual understanding of physics across the year levels during their time spent studying at university. This should be of particular concern to university instructors and stakeholders in the field of physics education, especially in the Indonesian context. Previous research indicated that little work has been undertaken focusing on variables such as epistemological beliefs, argumentation, and scientific reasoning, as well as the relationship between these variables with the understanding of physics concepts. Furthermore, researchers analysing the relationships between these variables have relied mostly on the analysis of single variate or bivariate relationships that could be misleading when trying to deeply understand the case. For instance, previous research found that the relationship between epistemological beliefs and the conceptual understanding of physics is not straightforward (Stathopoulou & Vosniadou, 2007). Likewise, Hofer and Pintrich (1997) pointed out that students' epistemological beliefs can indirectly affect their academic achievement. Thus, information about the variables that mediate the relationship between these two variables is required. Analysing the relationship between two variables most likely would not provide a complete picture, because it would not take into consideration other variables that may affect such variables. Therefore, to analyse the relationship between multiple variables, an advanced statistical technique is needed. In this study, a multivariate statistical analysis technique was employed to test a proposed model resulting from a review of the existing literature, as presented in Chapter 2. The intention was to gain a more comprehensive understanding of how aspects of scientific thinking affect pre-service physics teachers' understanding of physics concepts.

The findings of this study provide several contributions to knowledge, particularly in the area of research into physics education. Firstly, this study presents empirical evidence about a number of factors that affect scientific thinking and the conceptual understanding of physics. In turn, this provides answers to why there has been little improvement in pre-service physics teachers' skills in thinking scientifically and understanding physics concepts. The use of traditional teaching methods and learning assessments that centre on memorising physics formula derivations and solving traditional physics problems could be considered as key factors that contribute to the unsatisfactory improvement of pre-service physics teachers' scientific thinking and physics conceptual understanding across the four years of their undergraduate university program. More specifically, the findings of this study indicate that epistemological beliefs, argumentation, and scientific reasoning were rarely presented in their university classes, even though these variables are considered to play an important role in enhancing students' understanding of physics concepts.

Furthermore, this study provides a comprehensive understanding of how the demographic characteristics of pre-service physics teachers and scientific thinking (consisting of epistemological beliefs, argumentation, and scientific reasoning) influence the conceptual understanding of physics. To the best of the researcher’s knowledge, there has been no other research that has integrated these multiple variables into a single model. The SEM model analysis indicated that pre-service physics teachers’ beliefs about knowledge and knowing and their skills in argumentation indirectly influence their understanding of physics concepts mediated by their scientific reasoning skills. Among all the factors involved in the SEM model analysis, scientific reasoning skills were shown to be the factor that has the strongest impact on pre-service physics teachers’ understanding of physics concepts. This finding implies that the most powerful single predictor of understanding of physics concepts was scientific reasoning, suggesting that scientific reasoning skills play a crucial role in enhancing the understanding of physics concepts. Nevertheless, other variables (i.e., epistemological beliefs and argumentation) must also be taken into account because they also contribute to shaping students' scientific reasoning skills.

Overall, the findings of this study provide an understanding of the various factors that have a positive impact on improving pre-service physics teachers' scientific thinking and understanding of physics concepts and the complex relationships between several factors. This study also holds the potential to not only improve the understanding of how to better teach physics at the secondary and tertiary levels, but also to provide an evidence base upon which future policy recommendations can be made, particularly within the Indonesian context.

Methodological Implications

As mentioned previously, this study aims to investigate aspects of the scientific thinking and physics conceptual understanding of pre-service physics teachers. There are several methodological implications arising from this research in relation to mixed-methods research design, the Rasch model approach, and the Structural Equation Modelling (SEM) technique.

In the present study, the cross-sectional research design was employed in both the quantitative and the qualitative parts of the study. This mixed-methods approach was used to elicit more information in order to acquire a deeper interpretation and understanding of the analysis of the results. A sequential explanatory design was employed in which the quantitative data collection (i.e., surveys) was carried out in the first stage followed by collection of the qualitative data (i.e., individual interviews) in the second stage. In this study, the quantitative data was prioritised over the qualitative data. Nevertheless, the use of the qualitative data to support and augment the interpretation of the quantitative data was no less important for gaining rich information and a deep understanding of the research topic under investigation.

Several procedures and techniques were used to process the data, which were determined based on the objectives to be achieved from this research. As this study employed surveys to collect the quantitative data, appropriate techniques were needed to establish the reliability and validity of the research instruments before conducting further data analysis. The ACER ConQuest 4 statistical software package was used to examine the validity of the research instruments and to obtain the Weighted Likelihood Estimates (WLE) scores through a Rasch model analysis. All raw scores obtained from the participants' responses to items in the surveys were transformed into an interval scale in the existing data sets by employing the Rasch model approach. In the case of this study, the WLE technique was used to transform raw scores into measures in order to reduce estimation bias. Commonly, schools or teachers report students' abilities based on raw scores obtained from responses to questionnaires or tests (Khairani & Razak, 2015). An estimate of a person's ability to survey items is expressed as a total raw score by summing up the correct or incorrect responses (Bond & Fox, 2015), which are then treated as measures. High total raw scores are associated with students' ability to correctly answer a large number of survey questions; conversely, low total raw scores represent a large number of incorrect answers to responses given by students to questionnaires or test questions. Using raw scores can pose several shortcomings and lead to bias and can raise concerns about the conclusions being drawn (see Chapter 4). This becomes more problematic when advanced statistical models with multiple variables, such as Structural Equation Modelling (SEM), are applied (Bond & Fox, 2015). Using the Rasch model analysis, it is possible for researchers to counter the challenges of scoring survey responses and test items in order to obtain a reliable numerical score as well as to increase the validity of the interpretation. As indicated in Chapter 3, the questionnaires used in this study have frequently been analysed using raw scores by researchers. In other words, these researchers did not employ the Rasch model analysis to process their data. The use of the Rasch model in this study is promising, where the Rasch model has not been widely used in educational research, especially in the Indonesian context. Thus, one methodological implication of this study is that it demonstrates the effectiveness of using Rasch model analysis in research, specifically in the field of physics education.

In this study, the validation of instruments was achieved by establishing construct validity. The construct validity of the instruments was confirmed by employing the Rasch Model analysis and Confirmatory Factor Analysis (CFA) approaches. Rasch analysis was used to verify if the items fitted well with the Rasch model, while CFA was employed to examine the underlying structure of a scale or construct. Rasch model and CFA analyses are considered to be validation techniques that are complementary to each other (Ben, 2010). Conducting the CFA measurement technique enables researchers to test the conceptual theories that explain how observed variables represent scales or constructs. In addition, in order to investigate the relationships between multiple variables in this study, Structural Equation Modelling (SEM) techniques were applied using the IBM SPSS Amos 25. Compared to simple and multiple regression analysis, SEM analysis offers more advantages that can test complex relationships between several constructs or scales of a proposed model in light of previous studies and a literature review. Thus, the use of the SEM model is promising for examining the relationship between multiple variables in order to understand complex research phenomena.

The use of mixed-methods design could provide benefits, especially in educational research. However, the use of mixed-methods design, the Rasch model, and SEM analyses seem not to have been truly widespread and implemented in physics education research, especially in the Indonesian context. Educational phenomena are quite complex which generally involves a large number of variables being investigated. As a result, appropriate research designs and statistical analytical techniques to analyse research data are needed to obtain proper interpretations and conclusions. Commonly, students' academic performance is expressed as a total raw score which poses some ambiguity to measurement leading to bias in the analysis process, which could raise concerns about the conclusions being drawn (Alagumalai & Curtis, 2005). The use of raw scores becomes more problematic when advanced statistical models that involve a large number of variables such as Structural Equation Modelling (SEM) are applied (Bond & Fox, 2015). Hence, this study has also demonstrated the use of the Weighted Likelihood Estimates (WLE) technique to transform raw scores into measures to be carried out before proceeding to further quantitative analysis, to achieve a more meaningful and valid interpretation of the research findings. This suggests that this study contributes to knowledge about methodological approaches that could improve the quality of the data analysis in efforts to achieve research objectives or to address research questions.

Policy and Practice Implications

The findings of the quantitative and qualitative analyses in this study have provided empirical evidence related to a number of factors that are considered to contribute to the improvement of pre-service physics teachers' scientific thinking and conceptual understanding of physics, as well as showing how these factors are related to one another. In addition, the results of this study provide an overview of how, and to what extent, physics teaching and learning practices facilitate pre-service physics teachers to practice and enhance their physics knowledge and skills. These will be needed by pre-service physics teachers when they are teachers of physics in secondary schools. The findings of this study offer a number of suggestions for policymakers and practitioners to improve the quality of pre-service physics teachers’ scientific thinking and understanding of physics concepts.

A serious concern arising from the findings of this study relates to the unsatisfactory improvement of pre-service physics teachers' scientific thinking and their ability to master physics concepts. This study supports the need for further research related to physics teaching and learning practices taking place in Indonesian classrooms. The results of this study should encourage universities to promote students' independent learning through the provision of facilities and learning resources that support students to practice their thinking skills and enhance their knowledge. Pre-service physics teachers' perceptions related to the use of traditional teaching methods assisted by teaching media, such as a blackboard and PowerPoint slides, as well as physics learning material centred on mathematics formulae and the derivation of physics formulae, indicated that such approaches lead them to become passive learners in class. In such cases, they would only listen and take notes from explanations given by their university instructors, which did not assist them with investigative physics activities or help them to practice their scientific thinking skills or develop their ability to understand physics concepts. This finding implies that university instructors need to learn how to combine traditional teaching strategies with other active teaching methods in order to provide opportunities for students to practice their thinking skills and improve their understanding of physics concepts. This might be achieved through the implementation of constructivist teaching approaches that emphasise the active involvement of students in their learning. Ekinci (2017) highlighted that creating a student-centred learning environment is important, so that students have the opportunity to construct knowledge and be actively involved in the learning process which would help them to improve their learning performance. It is clear that developing curricula, syllabi, and lesson plans that aim to improve pre-service physics teachers' scientific thinking and conceptual understanding of physics, as well as knowing how to implement them in the classroom, is not an easy task. However, the results of this study serve to provide a base upon which university instructors can develop curricula, syllabi, and teaching plans that enhance aspects of scientific thinking and the understanding of physics concepts of pre-service physics teachers.

Likewise, through the interviews, the pre-service physics teachers demonstrated that the questions given in the mid-term and final examinations were more focused on problem-solving that used physics formulae and mathematical calculations, and lead to the adoption of surface approaches to learning. Consequently, they tended to memorise physics materials in order to pass their exams, rather than fostering the use of deep learning approaches that would provide more meaningful learning. Developing test questions that measure students' scientific thinking (consisting of epistemological beliefs, argumentation, and scientific reasoning), as well as their understanding of physics concepts is certainly not an easy task for instructors. It would not only take more time to develop these test items, but it would also require a deep understanding of how these questions can measure all of these variables. This suggests the importance of providing opportunities for university instructors and prospective physics teachers to attend professional teacher development sessions designed to improve their ability to develop curricula, syllabi, and teaching plans, and training them to construct and develop exam questions that emphasise the improvement of thinking skills and conceptual understanding of physics. In turn, these approaches can be implemented in their classrooms as an attempt to improve their students’ skills in higher-level thinking and understanding of physics concepts. In so doing, students could be assisted to understand and apply physics concepts in their daily life which would help them understand the importance and usefulness of studying physics.

In addition, the findings indicated that private universities, when compared to public universities, made little impact on improving the skills of pre-service physics teachers in all aspects of scientific thinking and conceptual understanding of physics, which affords important policy implications for private universities. This shows the need to assist instructors in private universities to develop their pedagogical knowledge and skills for teaching scientific thinking and physics conceptual understanding. This is important because pre-service physics teachers are the physics teachers of tomorrow who will be responsible for constructing and developing effective physics teaching and learning strategies for their students. They need to be able to stimulate the development of their students' scientific thinking and understanding of physics concepts.

This research provides a new theoretical understanding of how the demographic characteristics of pre-service physics teachers and aspects of scientific thinking influence their conceptual understanding of physics. The findings of this study have also provided a broader picture of current teaching and learning practices in physics in the Indonesian context at the tertiary level. These findings make a useful contribution to the world of physics education by providing valuable insights that have the potential to improve the quality of Indonesian physics teachers in Indonesia.

9.7 Limitations of the Study and Recommendations for Future Research

Limitations

Although the present study makes its contribution to knowledge in the field of physics education research in terms of scientific thinking and the conceptual understanding of physics, especially for Indonesian pre-service physics teachers, there is no denying that this study has a number of limitations. For instance, the data collected in this study included only four universities, all from the Yogyakarta region of Indonesia. Pre-service physics teachers involved in this study can hardly be considered as representatives of other universities throughout Indonesia and even beyond it. As a result, any generalisations are limited to the participants involved in this study and may not apply to other contexts.

This study is the first to investigate the relationship between aspects of scientific thinking, understanding of physics concepts, and other relevant factors carried out among prospective physics teachers in the Indonesian context at the tertiary level. However, although the factors involved in this study are considered important, and are related to scientific thinking skills and understanding of physics concepts, these findings cannot be considered conclusive due to the fact that there are likely to be influences from other factors that have been excluded in this study. In other words, this study does not indicate that the factors that have been examined are the only ones that influence participants' scientific thinking skills and understanding of physics concepts. There are many other factors that might play an important role in shaping pre-service physics teachers' scientific thinking skills and conceptual understanding of physics. Thus, there is a need to broaden the scope of research to cover other demographic factors and skills, as well as other relevant factors involved in physics teaching and learning, particularly in teacher education institutions.

Another limitation is related to the research design employed in this study. Ideally, a longitudinal study should be carried out to obtain the best picture of the influence of variables over a specified period of time. This should be done to strengthen the findings from the participants being investigated. However, due to time constraints and resource limitations, the present study conducted a cross-sectional research design, in which multiple variables were examined in different population groups compared at the same time. As stated by Cohen et al. (2011), using a cross-sectional research design enables researchers to collect data from a large sample that is cheaper to administer and can be carried out quickly.

Further to this, the SEM model developed in this study tested the relationship between multiple variables only in the forward direction, with the construct variables in this structural model ordered from left to right. In other words, this study did not accommodate the possibility of reciprocal effects operating between two factors in the developed model. Thus, further research is needed to examine the possibility of reciprocal relationships (bi-directional modelling) arising between a number of factors investigated in the study. Employing appropriate statistical analyses enables researchers to not only analyse the complex relationships between multiple variables, but also to estimate the reciprocal effects between latent variables in the path models (Razak, Keeves, & Darmawan, 2014). In addition, the model analysis techniques used in this study were limited to analysing data obtained from pre-service physics teachers (the student-level). In other words, pre-service physics teachers are the only source of data or information in this study, and no confirmation was made of their instructors or the university they attended. Thus, there is a need to expand this research by applying multi-level analysis techniques that not only analyse student-level data, but also data at the instructor and university-levels (a three-level hierarchy).

Recommendations

The following recommendations are based on the conclusions of the investigation:

- Larger sample sizes are needed from universities located in different provinces and regions of Indonesia. Random sampling techniques adopted should be considered carefully so that the results can be representative of the target population. If this was done, the data collected would be useful for the purpose of comparing universities throughout Indonesia and could provide a broad picture of the physics teaching and learning practices taking place in university classrooms. The research could also be extended beyond Indonesia to other countries in order to confirm and compare the findings of the current study.
- Considering the findings of this study that showed pre-service physics teachers' scientific thinking (i.e., epistemological beliefs, argumentation, and scientific reasoning) did not demonstrate significant differences and their physics conceptual understanding showed little improvement across year levels, it would be highly desirable to investigate other factors in future research that might relate to these aspects of scientific thinking and physics conceptual understanding. In addition, several factors found in the present qualitative study (e.g., approaches to learning, physics misconceptions, and motivation) could be included in the SEM model analysis to examine the correlations that arise between the other factors involved in the quantitative study employed in this research. The reasons for pre-service teachers choosing physics as their teaching area could also be investigated. Thus, a new conceptual framework could be generated for future research.
- Further research could be focused on a longitudinal study that enables researchers to make observations in different population groups over different periods of time (e.g., over the four years of studies in higher education). Thus, changes in all aspects of pre-service physics teachers' scientific thinking and physics conceptual understanding would be able to be observed more accurately.
- The SEM model analysis presented in Chapter 6 only demonstrates the effects of aspects of scientific thinking, consisting of epistemological beliefs, argumentation, and scientific reasoning, on pre-service physics teachers' conceptual understanding of physics. This model

did not analyse the possibility of a reciprocal relationship between the variables involved in this study, thus indicating that this opportunity could be taken up in further research. In the interviews, pre-service physics teachers stated that in order to acquire a better understanding of physics concepts, it was necessary to have high skills in scientific reasoning. Likewise, other pre-service physics teachers believed that having a deep understanding of physics concepts enables them to increase their epistemological beliefs and generate better arguments. Therefore, a similar study is needed that analyses the reciprocal effects between the variables investigated in this study.

- Finally, there is a need to extend the scope of the study by including other significant variables at the instructor-level (e.g., teaching methods implemented, learning media used, and learning assessment adopted) and the university-level (e.g., facilities or learning resources provided, and curriculum implemented in public or private universities) in order to confirm the findings of this study and to acquire an in-depth understanding of the various potential factors that might contribute to pre-service physics teachers' scientific thinking and conceptual understanding of physics. The use of multi-level analysis could be considered as an appropriate analysis technique to explore the effects between variables examined at different levels or hierarchically (Darmawan & Keeves, 2009).

By outlining the limitations of this study along with the recommendations described above, it is hoped that future research will be able to address all the limitations of this study and implement the recommendations offered in order to obtain a broader understanding of scientific thinking and conceptual understanding of physics. Hence, the results of this study could be used as a starting point for conducting future research on the development of thinking skills and conceptual understanding in a similar research topic, particularly in the field of physics education. Such research offers hope in being able to effectively improve the quality of physics education in Indonesia now and into the future.

REFERENCES

Acar, Ö., Patton, B. R., & White, A. L. (2015). Prospective secondary science teachers' argumentation skills and the interaction of these skills with their conceptual knowledge. *Australian Journal of Teacher Education, 40*(9), 132-156.

Adams, R., Wilson, M., & Wang, W.-c. (1997). The multidimensional random coefficients multinomial logit model. In *Applied Psychological Measurement* (Vol. 21, pp. 1-23): Sage.

Adams, R., & Wu, M. (2003). *PISA 2000 technical report.* Paris: OECD Publishing.

Adams, R., & Wu, M. (2010). Multidimensional models. In Australian Council for Educational Research (ACER).

Adams, W. C. (2015). Conducting semi-structured interviews. In *Handbook of Practical Program Evaluation* (Fourth ed., pp. 492-505): Jossey Bass.

Adams, W. K., Perkins, K. K., Dubson, M., Finkelstein, N. D., & Wieman, C. E. (2005). *The design and validation of the Colorado Learning Attitudes about Science Survey.* Paper presented at the AIP Conference Proceedings 790, 45. https://doi.org/10.1063/1.2084697.

Adams, W. K., Perkins, K. K., Podolefsky, N. S., Dubson, M., Finkelstein, N. D., & Wieman, C. E. (2006). New instrument for measuring student beliefs about physics and learning physics: The Colorado Learning Attitudes about Science Survey. *Physical Review Special Topics-Physics Education Research, 2*(1), 010101, 1-14. doi:10.1103/PhysRevSTPER.2.010101.

Adeyemo, S. A. (2012). The relationship among school environment, student approaches to learning and their academic achievement in senior secondary school physics. *International Journal of Educational Research and Technology, 3*(1), 21-26.

Adeyemo, S. A., & Babajide, V. F. (2014). Effects of mastery learning approach on students' achievement in physics. *International Journal of Scientific & Engineering Research, 5*(2), 910-920.

Adhabi, E., & Anozie, C. B. (2017). Literature review for the type of interview in qualitative research. *International Journal of Education, 9*(3), 86-97. doi:10.5296/ije.v9i3.11483

Adolphus, T., & Omeodu, D. (2016). Effects of gender and collaborative learning approach on students' conceptual understanding of electromagnetic induction. *Journal of Curriculum and Teaching, 5*(1), 78-86. doi:10.5430/jct.v5n1p78

Alagumalai, S., & Curtis, D. D. (2005). Classical test theory. In *Applied Rasch Measurement: A book of Exemplars* (pp. 1-14): Springer.

Alagumalai, S., Curtis, D. D., & Hungi, N. (2005). Our experiences and conclusion. In *Applied Rasch Measurement: A Book of Exemplars* (pp. 343-346): Springer.

Alao, S., & Guthrie, J. T. (1999). Predicting conceptual understanding with cognitive and motivational variables. *The Journal of Educational Research, 92*(4), 243-254. doi:10.1080/00220679909597602

Aldous, C. R. (2014). Attending to feeling: It may matter more than you think. *Creative Education, 5*(10), 780-796.

Aldous, C. R. (2017). Modelling the creative process and cycles of feedback. *Creative Education, 8*(12), 1860-1877. doi:10.4236/ce.2017.812127

Allison, P. D. (2003). Missing data techniques for structural equation modeling. *Journal of Abnormal Psychology, 112*(4), 545–557. doi:10.1037/0021-843X.112.4.545

Alshamali, M. A., & Daher, W. M. (2016). Scientific reasoning and its relationship with problem solving: The case of upper primary science teachers. *International Journal of Science and Mathematics Education, 14*(6), 1003-1019. doi:10.1007/s10763-015-9646-1

Alshenqeeti, H. (2014). Interviewing as a data collection method: A critical review. *English Linguistics Research, 3*(1), 39-45. doi:10.5430/elr.v3n1p39
Alt, D., & Boniel-Nissim, M. (2018). Links between adolescents' deep and surface learning approaches, problematic internet use, and Fear of Missing Out (FoMO). *Internet Interventions, 13*, 30-39.
Alwan, A. A. (2011). Misconception of heat and temperature among physics students. *Procedia-Social and Behavioral Sciences, 12*, 600-614. doi:10.1016/j.sbspro.2011.02.074
Amelink, C. (2009). Literature overview: Gender differences in science achievement. *SWE-AWE CASEE Overviews*, 1-22. Retrieved from http://www.AWEonline.org.
Andayani, Y., Hadisaputra, S., & Hasnawati, H. (2018). *Analysis of the level of conceptual understanding.* Paper presented at the Journal of Physics: Conference Series 1095. doi :10.1088/1742-6596/1095/1/012045.
Anderman, E. M., Sinatra, G. M., & Gray, D. L. (2012). The challenges of teaching and learning about science in the twenty-first century: Exploring the abilities and constraints of adolescent learners. *Studies in Science Education, 48*(1), 89-117. doi:10.1080/03057267.2012.655038.
Arbuckle, J. L. (2009). Amos 18 user's guide. *Chicago*: Amos Development Corporation.
Arbuckle, J. L. (2017). IBM SPSS Amos 25 user's guide. USA: Amos Development Corporation.
Arends, R.I. (2012). *Learning to teach*. New York: McGraw-Hill Companies, Inc.
Arghode, V., Brieger, E. W., & McLean, G. N. (2017). Adult learning theories: Implications for online instruction. *European Journal of Training and Development, 41*(7), 593-609.
Arista, F. S., & Kuswanto, H. (2018). Virtual physics laboratory application based on the android smartphone to improve learning independence and conceptual understanding. *International Journal of Instruction, 11*(1), 1-16.
Arnadóttir, G., & Fisher, A. (2008). Rasch analysis of the ADL scale of the A-ONE. *American Journal of Occupational Therapy, 62*(1), 51-60.
Aslan, C. (2017). Examining epistemological beliefs of teacher candidates according to various variables. *Eurasian Journal of Educational Research, 67*, 37-50. doi:10.14689/ejer.2017.67.3.
Astuti, F. S., & Retnawati, H. (2017). The effect of national examination's policy on readiness, motivation, school test score, and national examination score. *The Online of New Horizon in Education, 7*(3), 58-66.
Ayaz, M., Ali, N., Khan, A. B., Ullah, R., & Ullah, M. (2017). Impact of school library on students' academic achievement at secondary school level in southern districts of Khyber Pakhtunkhwa. *International Journal of Academic Research in Business and Social Sciences, 7*(5), 95-103. doi:10.6007/IJARBSS/v7-i5/2880.
Aydeniz, M., & Dogan, A. (2016). Exploring the impact of argumentation on pre-service science teachers' conceptual understanding of chemical equilibrium. *Chem. Educ. Res. Pract., 17*, 111-119. doi:10.1039/c5rp00170f.
Aydeniz, M., Pabuccu, A., Cetin, P. S., & Kaya, E. (2012). Argumentation and students' conceptual understanding of properties and behaviors of gases. *International Journal of Science and Mathematics Education, 10*, 1303-1324.
Bada, S. O., & Olusegun, S. (2015). Constructivism learning theory: A paradigm for teaching and learning. *IOSR Journal of Research & Method in Education, 5*(6), 66-70. doi:10.9790/7388-05616670.
Baeten, M., Kyndt, E., Struyven, K., & Dochy, F. (2010). Using student-centred learning environments to stimulate deep approaches to learning: Factors encouraging or discouraging their effectiveness. *Educational Research Review, 5*, 243-260. doi:10.1016/j.edurev.2010.06.001.

Baker, F. B. (2001). *The basics of item response theory* (Second ed.). College Park, MD: ERIC.
Bao, L., Cai, T., Koenig, K., Fang, K., Han, J., Wang, J., . . . Luo, Y. (2009). Learning and scientific reasoning. *Science, 323*, 586-587. doi:10.1126/science.1167740.
Bao, L., Fang, K., Cai, T., Wang, J., Yang, L., Cui, L., . . . Luo, Y. (2009). Learning of content knowledge and development of scientific reasoning ability: A cross culture comparison. *American Journal of Physics, 77*(12), 1118-1123.
Bas, G., Senturk, C., & Cigerci, F. M. (2017). Homework and academic achievement: A meta-analytic review of research. *Issues in Educational Research, 27*(1), 31-50.
Bates, S. P., Galloway, R. K., Loptson, C., & Slaughter, K. A. (2011). How attitudes and beliefs about physics change from high school to faculty. *Physical Review Special Topics-Physics Education Research, 7*(2), 020114, 1-8. doi:10.1103/PhysRevSTPER.7.020114.
Bathgate, M., Crowell, A., Schunn, C., Cannady, M., & Dorph, R. (2015). The learning benefits of being willing and able to engage in scientific argumentation. *International Journal of Science Education, 37*(10), 1590-1612. doi:10.1080/09500693.2015.1045958.
Bayraktar, S. (2009). Misconceptions of Turkish pre-service teachers about force and motion. *International Journal of Science and Mathematics Education, 7*, 273-291.
Bazeley, P., & Jackson, K. (2013). *Qualitative data analysis with NVivo* (Second ed.). London: Sage Publications Limited.
Belet, S. D., & Guven, M. (2011). Meta-cognitive strategy usage and epistemological beliefs of primary school teacher trainees. *Educational Sciences: Theory and Practice, 11*(1), 51-57.
Ben, F. (2010). *Students' uptake of physics: A study of South Australian and Filipino physics students. (Unpublished Doctoral Thesis).* The University of Adelaide, Adelaide, South Australia.
Benford, R., & Lawson, A. E. (2001). Relationships between effective inquiry use and the development of scientific reasoning skills in college biology labs. ERIC.
Berben, L., Sereika, S. M., & Engberg, S. (2012). Effect size estimation: Methods and examples. *International Journal of Nursing Studies, 49*(8), 1039-1047. doi:10.1016/j.ijnurstu.2012.01.015.
Berland, L. K., & McNeill, K. L. (2010). A learning progression for scientific argumentation: Understanding student work and designing supportive instructional contexts. *Science Education, 94*(5), 765-793. doi:10.1002/sce.20402.
Bezuidenhout, S. (2011). *Toward assessing scientific thinking: A qualitative analysis of student reasoning among psychology undergraduates.* University of Cape Town, South Africa.
Biddulph, F., & Carr, K. (1999). Learning theories and curriculum. *Teachers and Curriculum, 3*(1), 31-35.
Biggs. (2003). *Teaching for quality learning at university* (Second ed.). Maidenhead: The Society for Research into Higher Education and Open University Press.
Biggs, J., Kember, D., & Leung, D. Y. (2001). The revised two-factor study process questionnaire: R-SPQ-2F. *British Journal of Educational Psychology, 71*, 133-149.
Biggs, J. B., & Moore, P. J. (1993). *The process of learning (3rd ed.)*. New York: Prentice Hall.
Bigozzi, L., Tarchi, C., Fiorentini, C., Falsini, P., & Stefanelli, F. (2018). The influence of teaching approach on students' conceptual learning in physics. *Frontiers in Psychology, 9*, 1-14. doi:10.3389/fpsyg.2018.02474.
Birbili, M. (2000). Translating from one language to another. Social Research(31), 1-7.
Available from: http://www.soc.surrey.ac.uk/sru/SRU31.html.
Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences (3rd ed.)*. New York: Routledge.
Boone, W. J., Staver, J. R., & Yale, M. S. (2013). *Rasch analysis in the human sciences.* USA: Springer.

Brauer, H., & Wilde, M. (2014). Pre-service science teachers' development of epistemological beliefes. *Reflecting Education, 9*(2), 44-62.
Breiner, J. M., Harkness, S. S., Johnson, C. C., & Koehler, C. M. (2012). What is STEM? A discussion about conceptions of STEM in education and partnerships. *School Science and Mathematics, 112*(1), 3-11.
Brislin, R. W. (1986). The wording and translation of research instruments. In W. J. Lonner & J. W. Berry (Eds.), In Field Methods in Cross-cultural Research (pp. 137–164). Beverly Hills, CA: Sage Publications.
Brownlee, J., Purdie, N., & Boulton-Lewis, G. (2001). Changing epistemological beliefs in pre-service teacher education students. *Teaching in Higher Education, 6*(2), 247-268. doi:10.1080/13562510120045221.
Buber, A., & Coban, G. U. (2017). The effects of learning activities based on argumentation on conceptual understanding of 7th graders about "force and motion" unit and establishing thinking friendly classroom environment. *European Journal of Educational Research, 6*(3), 367-384. doi:10.12973/eu-jer.6.3.367
Buijs, M., & Admiraal, W. (2013). Homework assignments to enhance student engagement in secondary education. *European Journal of Psychology of Education, 28*, 767-779. doi:10.1007/s10212-012-0139-0.
Bybee, R. W., & Fuchs, B. (2006). Preparing the 21st century workforce: A new reform in science and technology education. *Journal of Research in Science Teaching, 43*(4), 349-352. doi:10.1002/tea.20147.
Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (Second ed.). New York and London: Routledge.
Byrne, B. M. (2016). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (Third ed.). New York: Routledge.
Cahyaningrum, R., & Hidayat, A. (2018). *Contrasting-cases problems: Learning material to improve students' conceptual understanding on magnetism.* Paper presented at the Journal of Physics: Conference Series 1097. doi :10.1088/1742-6596/1097/1/012028.
Cameron, R. (2011). Mixed methods research: The five Ps framework. *The Electronic Journal of Business Research Methods, 9*(2), 96-108.
Canales, A., & Maldonado, L. (2018). Teacher quality and student achievement in Chile: Linking teachers' contribution and observable characteristics. *International Journal of Educational Development, 60*, 33-50.
Cavallo, A. M. L., Rozman, M., Blickenstaff, J., & Walker, N. (2003). Learning, reasoning, motivation, and epistemological beliefs. *Journal of College Science Teaching, 33*(3), 18-23.
Çelik, A. Y., & Kılıç, Z. (2014). The impact of argumentation on high school chemistry students' conceptual understanding, attitude towards chemistry and argumentativeness. *Eurasian Journal of Physics and Chemistry Education, 6*(1), 58-75.
Cetin, P. S., Erduran, S., & Kaya, E. (2010). Understanding the nature of chemistry and argumentation: The case of pre-service chemistry teachers. *Kırşehir Üniversitesi Eğitim Fakültesi Dergisi (KEFAD), 11*(4), 41-59.
Chabay, R., & Sherwood, B. (1997). Qualitative understanding and retention. *AAPT Announcer, 27*(2), 96.
Chang, M. C., Al-Samarrai, S., Ragatz, A. B., Shaeffer, S., De Ree, J., & Stevenson, R. (2013). *Teacher reform in Indonesia: The role of politics and evidence in policy making*: World Bank Publications.

Cheema, J. R. (2014). Some general guidelines for choosing missing data handling methods in educational research. *Journal of Modern Applied Statistical Methods, 13*(2), 53-75.
Chen, H.-T., Wang, H.-H., Lu, Y.-Y., & Hong, Z.-R. (2019). Bridging the gender gap of children's engagement in learning science and argumentation through a modified argument-driven inquiry. *International Journal of Science and Mathematics Education, 17*(4), 635-655. doi:10.1007/s10763-018-9896-9.
Chen, H.-T., Wang, H.-H., Lu, Y.-Y., Lin, H.-s., & Hong, Z.-R. (2016). Using a modified argument-driven inquiry to promote elementary school students' engagement in learning science and argumentation. *International Journal of Science Education, 38*(2), 170-191. doi:10.1080/09500693.2015.1134849.
Chen, L., Xu, S., Xiao, H., & Zhou, S. (2019). Variations in students' epistemological beliefs towards physics learning across majors, genders, and university tiers. *Physical Review Physics Education Research, 15*(1), 010106, 1-11. doi:10.1103/PhysRevPhysEducRes.15.010106.
Chen, P., & Schmidtke, C. (2017). Humanistic elements in the educational practice at a United States sub-baccalaureate technical college. *International Journal for Research in Vocational Education and Training (IJRVET), 4*(2), 117-145.
Chen, S., Chang, W. H., Lai, C. H., & Tsai, C. Y. (2014). A comparison of students' approaches to inquiry, conceptual learning, and attitudes in simulation-based and microcomputer-based laboratories. *Science Education, 98*(5), 905-935. doi:10.1002/sce.21126.
Ciara, O. (2009). Enhancing student learning through assessment. In *A Toolkit Approach* (pp. 1-24).
Çınar, D., & Bayraktar, Ş. (2014). Evaluation of the effects of argumentation based science teaching on 5th grade students' conceptual understanding of the subjects related to 'Matter and Change'. *International Journal of Education in Mathematics, Science and Technology, 2*(1), 49-77.
Clark, L., Birkhead, A. S., Fernandez, C., & Egger, M. J. (2017). A transcription and translation protocol for sensitive cross-cultural team research. *Qualitative Health Research, 27*(12), 1751-1764.
Clark, V. P., & Creswell, J. W. (2011). Designing and conducting mixed methods research. Thousand Oaks, CA: Sage.
Cohen, J. (1988). Statistical power analysis for the social sciences (Second ed.). New York: Academic Press.
Cohen, J. (1992). A power primer. Psychological Bulletin, 112(1), 155–159.
Cohen, L., Manion, L., & Morrison, K. (2011). *Surveys, longitudinal, cross-sectional and trend studies*. London: Routledge.
Coletta, V. P., & Phillips, J. A. (2005). Interpreting FCI scores: Normalized gain, preinstruction scores, and scientific reasoning ability. *American Journal of Physics, 73*(12), 1172-1182. doi:http://dx.doi.org/10.1119/1.2117109.
Coletta, V. P., & Phillips, J. A. (2015). *Thinking in physics.* Paper presented at the Physics Education Research Conference Proceedings.
Coletta, V. P., Phillips, J. A., & Steinert, J. (2012). *FCI normalized gain, scientific reasoning ability, thinking in physics, and gender effects.* Paper presented at the AIP Conference Proceedings 1413, 23. doi: 10.1063/1.3679984.
Collins, R. (2014). Skills for the 21st Century: Teaching higher-order thinking. *Curriculum & Leadership Journal, 12*(14).
Corpuz, E. G. (2006). Promoting understanding of physics concepts. Available at https://www.math.ksu.edu/math791/midterms03/edgarcomment.pdf.
Cramer, D. (2003). *Advanced quantitative data analysis*. UK: McGraw-Hill Education.

Creswell, J. W. (2012). *Educational research. Planning, conducting, and evaluating quantitative and qualitative research* (Fourth ed.). Boston: Pearson Education.
Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed methods approaches* (Third ed.). USA: Sage Publications.
Creswell, J. W. (2014a). *A concise introduction to mixed methods research*. USA: Sage Publications.
Creswell, J. W. (2014b). *Research design: Qualitative, quantitative, and mixed methods approaches.* (Fourth ed.). Thousand Oaks, California: SAGE Publications.
Creswell, J. W., & Clark, V. L. P. (2007). Designing and conducting mixed methods research. Thousand Oaks, CA: Sage.
Crocker, L., & Algina, J. (2006). *Introduction to classical and modern test theory.* ERIC.
Cross, D., Taasoobshirazi, G., Hendricks, S., & Hickey, D. T. (2008). Argumentation: A strategy for improving achievement and revealing scientific identities. *International Journal of Science Education, 30*(6), 837-861. doi:10.1080/09500690701411567.
Daluba, N. E. (2012). Evaluation of resource availability for teaching science in secondary schools: Implications for Vision 20: 2020. *Journal of Emerging Trends in Educational Research and Policy Studies, 3*(3), 363-367.
Darling-Hammond, L. (2006). Constructing 21st-century teacher education. *Journal of Teacher Education, 57*(3), 300-314. doi:10.1177/0022487105285962.
Darmawan, I. G. N. (2002). NORM software review: Handling missing values with multiple imputation methods. *Evaluation Journal of Australasia, 2*(1), 51-57.
Darmawan, I. G. N. (2003). *Implementation of information technology in local government in Bali, Indonesia*: Flinders University.
Darmawan, I. G. N., & Keeves, J. P. (2006). Suppressor variables and multilevel mixture modelling. *International Education Journal, 7*(2), 160-173.
Darmawan, I. G. N., & Keeves, J. P. (2009). *Using multilevel analysis. In C. R. Aldous, I. G. N. Darmawan & J. P. Keeves (Eds.), Change over time in learning numeracy and literacy in rural and remote schools* (Vol. 2). Adelaide, South Australia: Shannon Research Press.
Demirci, N. (2005). A study about students' misconceptions in force and motion concepts by incorporating a web-assisted physics program. *The Turkish Online Journal of Educational Technology, 4*(3), 40-48.
Dervic, D., Glamocic, D. S., Gazibegovic-Busuladzic, A., & Mesic, V. (2018). Teaching physics with simulations: Teacher-centered versus student-centered approaches. *Journal of Baltic Science Education, 17*(2), 288-299.
Dewey, I., & Dykstra, J. (2008). Physics classroom engagement: Constructing understanding in real time. *Latin-American Journal of Physics Education, 2*(1), 1-5.
DiCicco-Bloom, B., & Crabtree, B. F. (2006). The qualitative research interview. *Medical Education, 40*(4), 314-321. doi:10.1111/j.1365-2929.2006.02418.x.
Dilas, D. B., Mackie, C., Huang, Y., & Trines, S. (2019). Education in Indonesia. Retrieved from https://wenr.wes.org/2019/03/education-in-indonesia-2.
Dilber, R., Karaman, I., & Duzgun, B. (2009). High school students' understanding of projectile motion concepts. Educational Research and Evaluation, 15(3), 203-222.
Ding, L. (2014a). Long live traditional textbook problems!?—Constraints on faculty use of research-based problems in introductory courses. *International Journal of Science and Mathematics Education, 12*(1), 123-144.
Ding, L. (2014b). Seeking missing pieces in science concept assessments: Reevaluating the Brief Electricity and Magnetism Assessment through Rasch analysis. *Physical Review Special*

Topics-Physics Education Research, 10(1), 010105, 1-15. doi: 10.1103/PhysRevSTPER.10.010105.
Ding, L. (2014c). Verification of causal influences of reasoning skills and epistemology on physics conceptual learning. *Physical Review Special Topics-Physics Education Research, 10*(2), 1-5. doi:10.1103/PhysRevSTPER.10.023101.
Ding, L. (2018). Progression trend of scientific reasoning from elementary school to university: A large-scale cross-grade survey among Chinese students. *International Journal of Science and Mathematics Education, 16*(8), 1479-1498. doi:10.1007/s10763-017-9844-0.
Ding, L., Chabay, R., Sherwood, B., & Beichner, R. (2006). Evaluating an electricity and magnetism assessment tool: Brief electricity and magnetism assessment. *Physical Review Special Topics - Physics Education Research, 2*(1), 1-7. doi:10.1103/PhysRevSTPER.2.010105.
Ding, L., & Mollohan, K. N. (2015). How college-level introductory instruction can impact student epistemological beliefs. *Journal of College Science Teaching, 44*(4), 19-27.
Ding, L., Wei, X., & Liu, X. (2016). Variations in university students' scientific reasoning skills across majors, years, and types of institutions. *Research in Science Education, 46*, 613–632. doi:10.1007/s11165-015-9473-y.
Ding, L., Wei, X., & Mollohan, K. (2016). Does higher education improve student scientific reasoning skills? *International Journal of Science and Mathematics Education, 14*, 619–634. doi:10.1007/s10763-014-9597-y.
Ding, L., & Zhang, P. (2016). Making of epistemologically sophisticated physics teachers: A cross-sequential study of epistemological progression from preservice to in-service teachers. *Physical Review Physics Education Research, 12*(2), 1-16. doi:10.1103/PhysRevPhysEducRes.12.020137.
Docktor, J. L., & Mestre, J. P. (2014). Synthesis of discipline-based education research in physics. *Physical Review Special Topics-Physics Education Research, 10*(2), 1-58. doi:10.1103/PhysRevSTPER.10.020119.
Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education, 84*(3), 287-312.
Dunbar, K., & Fugelsang, J. (2005). Scientific thinking and reasoning. In *The Cambridge Handbook of Thinking and Reasoning* (pp. 705-725). USA: Cambridge University Press.
Efe, R., & Aslan-Efe, H. (2018). Science student teachers' approaches to studying. *International e-Journal of Educational Studies (IEJES), 2*(3), 53-63. doi:10.31458/iejes.376848.
Effendy, S., Hartono, Y., & Ian, M. (2018). *The ability of scientific reasoning and mastery of physics concept of state senior high school students in Palembang city.* Paper presented at the International Conference on Science and Education and Technology. Atlantis Press.
Efilti, E., & Çoklar, A. N. (2016). The analysis of the relationship between epistemological beliefs and TPACK education competence among preservice teachers. *Journal of Human Sciences, 13*(2), 2960-2971. doi:10.14687/jhs.v13i2.3593.
Eggen, P., & Kauchak, D. (2015). Educational psychology: Windows on classrooms (Tenth ed.; Global ed.). Pearson Higher Ed.
Eguabor, B., & Adeleke, A. (2017). Investigating differential learning outcomes of students in physics using animation and textual information teaching strategies in Ondo State secondary school. *European Journal of Sustainable Development Research, 1*(2), 1-7.
Ekinci, N. (2017). Examining the relationships between epistemological beliefs and teaching and learning conceptions of lower-secondary education teachers. *İnönü University Journal of the Faculty of Education, 18*(1), 344-358. doi:10.17679/inuefd.307065.

Elby, A. (1999). Another reason that physics students learn by rote. *American Journal of Physics, 67*(7), S52-S57.
Elby, A. (2001). Helping physics students learn how to learn. *American Journal of Physics, 69*(7), S54-S64. doi:10.1119/1.1377283.
Emerson, L., Fear, J., Fox, S., & Sanders, E. (2012). Parental engagement in learning and schooling: Lessons from research. In *A report by the Australian Research Alliance for Children and Youth (ARACY) for the Family–School and Community Partnerships Bureau*. Canberra.
Engelhardt, P. V., & Beichner, R. J. (2004). Students' understanding of direct current resistive electrical circuits. *American Journal of Physics, 72*(1), 98-115.
Engelmann, K., Neuhaus, B. J., & Fischer, F. (2016). Fostering scientific reasoning in education–meta-analytic evidence from intervention studies. *Educational Research and Evaluation*, 1-17. doi:10.1080/13803611.2016.1240089.
Engeström, Y. (1994). *Training for change: New approach to instruction and learning in working life*. International Labour Office Geneva.
Erduran, S. (2007). Methodological foundations in the study of argumentation in science classrooms. In *Argumentation in Science Education* (pp. 47-69). Springer.
Erduran, S., & Jiménez-Aleixandre, M. P. (2008). Argumentation in science education. In *Perspectives from Classroom-Based Research*. UK: Springer.
Erduran, S., Simon, S., & Osborne, J. (2004). TAPping into argumentation: Developments in the application of Toulmin's argument pattern for studying science discourse. *Science Education, 88*(6), 915-933.
Ertmer, P. A., & Newby, T. J. (2013). Behaviorism, cognitivism, constructivism: Comparing critical features from an instructional design perspective. *Performance Improvement Quarterly, 26*(2), 43-71. doi:10.1002/piq.21143.
Eskin, H., & Ogan-Bekiroglu, F. (2013). Argumentation as a strategy for conceptual learning of dynamics. *Research in Science Education, 43*(5), 1939-1956. doi:10.1007/s11165-012-9339-5.
Eveline, E., Wilujeng, I., & Kuswanto, H. (2019). *The effect of scaffolding approach assisted by PhET simulation on students' conceptual understanding and students' learning independence in physics.* Paper presented at the Journal of Physics: Conference Series 1233. doi:10.1088/1742-6596/1233/1/012036.
Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction*. Research Findings and Recommendation (pp. 1-112). ERIC.
Faisal, & Martin, S. N. (2019). Science education in Indonesia: Past, present, and future. *Asia-Pacific Science Education, 5*(4), 1-29. https://doi.org/10.1186/s41029-019-0032-0.
Faour, M. A., & Ayoubi, Z. (2018). The effect of using virtual laboratory on grade 10 students' conceptual understanding and their attitudes towards physics. *Journal of Education in Science, Environment and Health (JESEH), 4*(1), 54-68. doi:10.21891/jeseh.387482.
Faure, E. (1972). Learning to be. Paris: UNESCO. Findsen, B.
Fenanlampir, A., Batlolona, J. R., & Imelda, I. (2019). The struggle of Indonesian students in the context of TIMSS and PISA has not ended. *International Journal of Civil Engineering and Technology, 10*(2), 393-406.
Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (Fourth ed.). CA: Sage Publications.
Firdaus, T., Erwin, E., & Rosmiati, R. (2019). *Learning media free fall motion to reduce misconceptions and improve students' understanding of the concept.* Paper presented at the Journal of Physics: Conference Series 1157. doi:10.1088/1742-6596/1157/3/032072.

Firman, H. (2016). Diagnosing weaknesses of Indonesian students' learning. In *What Can PISA 2012 Data Tell Us?* (pp. 63-80). The Netherlands: Springer.

Fischer, G. H., & Molenaar, I. W. (1995). Rasch models: Foundations, recent developments, and applications. New York, NY: Springer-Verlag.

Fraenkel, J., & Wallen, N. (2003). Observation and interviewing. How to design and evaluate research in education, 5, 455-463.

Franco, G. M., Muis, K. R., Kendeou, P., Ranellucci, J., Sampasivam, L., & Wang, X. (2012). Examining the influences of epistemic beliefs and knowledge representations on cognitive processing and conceptual change when learning physics. *Learning and Instruction, 22*(1), 62-77. doi:10.1016/j.learninstruc.2011.06.003.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences, 111*(23), 8410-8415.

Frey, B. B., Ellis, J. D., Bulgren, J. A., Hare, J. C., & Ault, M. (2015). Development of a Test of Scientific Argumentation. *Electronic Journal of Science Education, 19*(4), 1-18.

Gable, G. G. (1994). Integrating case study and survey research methods: An example in information systems. *European Journal of Information Systems, 3*(2), 112-126.

Gaigher, E. (2014). Questions about answers: Probing teachers' awareness and planned remediation of learners' misconceptions about electric circuits. *African Journal of Research in Mathematics, Science and Technology Education, 18*(2), 176-187. doi:10.1080/10288457.2014.925268.

Gaigher, E., Rogan, J. M., & Braun, M. W. H. (2007). Exploring the development of conceptual understanding through structured problem-solving in physics. *International Journal of Science Education, 29*(9), 1089-1110. doi:10.1080/09500690600930972.

Galarpe, V. R. K. R. (2017). Students' perception of physics laboratory equipment in the university of science and technology of Southern Philippines. *Science International (Lahore), 29*(4), 777-781.

George, D., & Mallery, M. (2003). Using SPSS for Windows step by step: A simple guide and reference. 11.0 update (4th ed.).

Gire, E., Jones, B., & Price, E. (2009). Characterizing the epistemological development of physics majors. *Physical Review Special Topics-Physics Education Research, 5*(1), 010103, 1-6. doi:10.1103/PhysRevSTPER.5.010103.

Glaser, B., & Strauss, A. (1967). The discovery of grounded theory: Strategies for qualitative research. Chicago: Aldine, USA.

Graham, J. W. (2012). Multiple imputation with norm 2.03. In *Missing Data* (pp. 73-94). USA: Springer.

Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis, 11*(3), 255-274.

Greenhill, V. (2010). 21st Century knowledge and skills in educator preparation. In *Partnership for 21st Century Skills.* USA: Pearson.

Grodner, A., & Rupp, N. G. (2013). The role of homework in student learning outcomes: Evidence from a field experiment. *The Journal of Economic Education, 44*(2), 93-109. doi:10.1080/00220485.2013.770334.

Gu, L., & Kristoffersson, M. (2015). Swedish lower secondary school teachers' perceptions and experiences regarding homework. *Universal Journal of Educational Research, 3*(4), 296-305. doi:10.13189/ujer.2015.030407.

Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods, 18*(1), 59-82. doi:10.1177/1525822X05279903

Guey, C.-c., Cheng, Y.-y., & Shibata, S. (2010). A triarchal instruction model: Integration of principles from Behaviorism, Cognitivism, and Humanism. *Procedia-Social and Behavioral Sciences, 9*, 105-118. doi:10.1016/j.sbspro.2010.12.122.

Guido, R. M. D. (2018). Attitude and motivation towards learning physics. *International Journal of Engineering Research & Technology (IJERT), 2*(11), 2087-2094.

Gunawan, G., Nisrina, N., Suranti, N., Herayanti, L., & Rahmatiah, R. (2018). *Virtual laboratory to improve students' conceptual understanding in physics learning.* Paper presented at the Journal of Physics: Conference Series 1108. doi:10.1088/1742-6596/1108/1/012049.

Guney, A., & Al, S. (2012). Effective learning environments in relation to different learning theories. *Procedia-Social and Behavioral Sciences, 46*, 2334-2338. doi:10.1016/j.sbspro.2012.05.480.

Gunstone, R., Mulhall, P., & McKittrick, B. (2009). Physics teachers' perceptions of the difficulty of teaching electricity. *Research in Science Education, 39*(4), 515-538. doi:10.1007/s11165-008-9092-y.

Guo, Y. M., Klein, B. D., & Ro, Y. K. (2019). On the effects of student interest, self-efficacy, and perceptions of the instructor on flow, satisfaction, and learning outcomes. *Studies in Higher Education*, 1-18. doi:10.1080/03075079.2019.1593348.

Gürel, C., & Süzük, E. (2017). Pre-service physics teachers' argumentation in a model rocketry physics experience. *Educational Sciences: Theory & Practice, 17*(1), 5-26. doi:10.12738/estp.2017.1.0042

Gurria, A. (2016). *Pisa 2015 results in focus*. OECD.

Hailaya, W., Alagumalai, S., & Ben, F. (2014). Examining the utility of assessment literacy inventory and its portability to education systems in the Asia Pacific region. *Australian Journal of Education, 58*(3), 297-317.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (Seventh ed.). New Jersey: Pearson Education, Inc.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate data analysis*. UK: Pearson New International Edition.

Hairan, A. M., Abdullah, N., & Husin, A. H. (2019). Conceptual understanding of Newtonian mechanics among Afghan students. *European Journal of Physics Education, 10*(1), 1309-7202.

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics, 66*(1), 64-74.

Hake, R. R. (2007). Six lessons from the physics education reform effort. *Latin American Journal of Physics Education, 1*(1), 24-31.

Halloun, I. (1997). *Views about science and physics achievement: The VASS story.* Paper presented at the International Conference on Undergraduate Physics Education.

Halloun, I. A., & Hestenes, D. (1985). The initial knowledge state of college physics students. *American Journal of Physics, 53*(11), 1043-1055.

Hammer, D., & Elby, A. (2003). Tapping epistemological resources for learning physics. *The Journal of the Learning Sciences, 12*(1), 53-90. doi:10.1207/ S15327809JLS1201_3.

Han, J. (2013). *Scientific reasoning: Research, development, and assessment.* (*Unpublished Doctoral Thesis*). The Ohio State University, United States.

Hanson, W. E., Creswell, J. W., Clark, V. L. P., Petska, K. S., & Creswell, J. D. (2005). Mixed methods research designs in counseling psychology. *Journal of Counseling Psychology, 52*(2), 224–235. doi:10.1037/0022-0167.52.2.224

Harwell, M. R., & Gatti, G. G. (2001). Rescaling ordinal data to interval data in educational research. *Review of Educational Research, 71*(1), 105-131.

Hasene, E. Y., & Şekercioğlu, A. G. Ç. (2018). An analysis of teacher candidates' epistemological beliefs: A qualitative study. *Pegem Journal of Education and Instruction, 8*(1), 173-210. doi:10.14527/pegegog.2018.008.

Hays, R. D., Morales, L. S., & Reise, S. P. (2000). Item response theory and health outcomes measurement in the 21st century. *Medical Care, 38*(9 Suppl), II28–II42.

Hendayana, S., Asep, S., & Imansyah, H. (2010). Indonesia's issues and challenges on quality improvement of mathematics and science education. *Journal of International Cooperation in Education*, 4 (2), 41-51.

Henderson, C. (2002). Common concerns about the force concept inventory. *The Physics Teacher, 40*(9), 542-547.

Henderson, R., Stewart, G., Stewart, J., Michaluk, L., & Traxler, A. (2017). Exploring the gender gap in the conceptual survey of electricity and magnetism. *Physical Review Physics Education Research, 13*(2), 1-17. doi:10.1103/PhysRevPhysEducRes.13.020114.

Heng, L. L., Surif, J., & Seng, C. H. (2014). Individual versus group argumentation: Student's performance in a Malaysian context. *International Education Studies, 7*(7), 109-124. doi:10.5539/ies.v7n7p109.

Hestenes, D., & Wells, M. (1992). A Mechanics Baseline Test. In *The Physics Teacher* (Vol. 30, pp. 159-166).

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force Concept Inventory. In *The Physics Teacher* (Vol. 30, pp. 141-158).

Hofer, B. K. (2000). Dimensionality and disciplinary differences in personal epistemology. *Contemporary Educational Psychology, 25*(4), 378-405. doi:10.1006/ceps.1999.1026.

Hofer, B. K. (2001). Personal epistemology research: Implications for learning and teaching. *Journal of Educational Psychology Review, 13*(4), 353-383.

Hofer, B. K. (2006). Beliefs about knowledge and knowing: Integrating domain specificity and domain generality: A response to Muis, Bendixen, and Haerle (2006). *Educational Psychology Review, 18*(1), 67-76. doi:10.1007/s10648-006-9000-9.

Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research, 67*(1), 88-140.

Hofer, B. K., & Pintrich, P. R. (2012). *Personal epistemology: The psychology of beliefs about knowledge and knowing*. Routledge.

Holyoak, K. J., & Morrison, R. G. (2005). *Thinking and reasoning: A reader's guide.* In K.J. Holyoak & R.G. Morrison (Eds.), Oxford Handbook of Thinking and Reasoning. New York: Oxford University Press.

Hong, Z.-R., Lin, H.-s., Wang, H.-H., Chen, H.-T., & Yang, K.-K. (2013). Promoting and scaffolding elementary school students' attitudes toward science and argumentation through a science and society intervention. *International Journal of Science Education, 35*(10), 1625-1648. doi:10.1080/09500693.2012.734935.

Hotulainen, R., & Telivuo, J. (2014). Epistemological beliefs and scientific reasoning in Finnish academic upper secondary education. In *Kasvatus & Aika* (Vol. 9, pp. 92–106).

Hsu, P. S., Van Dyke, M., Chen, Y., & Smith, T. (2016). A cross-cultural study of the effect of a graph-oriented computer-assisted project-based learning environment on middle school students' science knowledge and argumentation skills. *Journal of Computer Assisted Learning, 32*(1), 51-76. doi:10.1111/jcal.12118.

Hunt, S. M., & Bhopal, R. (2004). Self report in clinical and epidemiological studies with non-English speakers: The challenge of language and culture. *Journal of Epidemiology & Community Health, 58*(7), 618-622. doi:10.1136/jech.2003.010074.

Husamah, & Pantiwati, Y. (2014). Cooperative learning STAD-PjBL: Motivation, thinking skills, and learning outcomes in biology students. *International Journal of Education Learning & Development, 2*(1), 77-94.

Husnaini, S. J., & Chen, S. (2019). Effects of guided inquiry virtual and physical laboratories on conceptual understanding, inquiry performance, scientific inquiry self-efficacy, and enjoyment. *Physical Review Physics Education Research, 15*(1), 1-16. doi:10.1103/PhysRevPhysEducRes.15.010119.

Hyytinen, H., Holma, K., Toom, A., Shavelson, R. J., & Lindblom-Ylänne, S. (2014). The complex relationship between students' critical thinking and epistemological beliefs in the context of problem solving. *Frontline Learning Research, 6*, 1-25. doi:10.14786/flr.v2i4.124.

Ibrahim, B., Ding, L., Mollohan, K. N., & Stammen, A. (2015). Scientific reasoning: Theory evidence coordination in physics-based and non-physics-based tasks. *African Journal of Research in Mathematics, Science and Technology Education, 20*(2), 93-105. doi:10.1080/10288457.2015.1108570.

Inhelder, B., & Piaget, J. (1958). The growth of logical thinking from childhood to adolescence. New York, NY: Basic Books.

Irwanto, Eli, R., & Prodjosantoso. (2019). Analyzing the relationships between pre-service chemistry teachers' science process skills and critical thinking skills. *Journal of Turkish Science Education, 16*(3), 299-313.

Jackson, T. R., Draugalis, J., Slack, M. K., Zachry, W. M., & D'Agostino, J. (2002). Validation of authentic performance assessment: A process suited for Rasch modeling. *American Journal of Pharmaceutical Education, 66*(3), 233-243.

Jalal, F., Samani, M., Chang, M. C., Stevenson, R., Ragatz, A. B., & Negara, S. D. (2009). *Teacher certification in Indonesia: A strategy for teacher quality improvement*. Departemen Pendidikan Nasional Republik Indonesia.

Jauhiainen, J. (2013). Effects of an in-service training program on physics teachers' pedagogical content knowledge: The role of experiments and interacting bodies in teaching Newtonian mechanics. Research Report 345. Helsinki.

Jeong, A., & Davidson-Shivers, G. V. (2006). The effects of gender interaction patterns on student participation in computer-supported collaborative argumentation. *Educational Technology Research and Developmental Review*, 54(6), 543–568.

Jeong, J. S., González-Gómez, D., Conde-Núñez, M. C., & Gallego-Picó, A. (2019). Examination of students' engagement with R-SPQ-2F of learning approach in flipped sustainable science course. *Journal of Baltic Science Education, 18*(6), 880-891. doi:10.33225/jbse/19.18.880.

Jian, Y.-C., & Wu, C.-J. (2015). Using eye tracking to investigate semantic and spatial representations of scientific diagrams during text-diagram integration. *Journal of Science Education and Technology, 24*(1), 43-55. doi:10.1007/s10956-014-9519-3.

Jiang, T., Wang, S., Wang, J., & Ma, Y. (2018). Effect of different instructional methods on students' conceptual change regarding electrical resistance as viewed from a synthesized theoretical framework. *Eurasia Journal of Mathematics, Science and Technology Education, 14*(7), 2771-2786. doi:10.29333/ejmste/90592.

Jingna, D. (2012). Application of humanism theory in the teaching approach. *Higher Education of Social Science, 3*(1), 32-36. doi:10.3968/j.hess.1927024020120301.1593.

Johnson, B., & Turner, L. A. (2003). Data collection strategies in mixed methods research. In *Handbook of Mixed Methods in Social and Behavioral Research. In Tashakkori, A.M. and Teddlie, C.B., Eds.* (pp. 297-319). Thousand Oaks: SAGE Publications.

Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research, 1*(2), 112-133.

Jufri, A., Setiadi, S., & Sripatmi, D. (2016). Scientific reasoning ability of prospective student teacher in the excellence program of mathematics and science teacher education in university of Mataram. *Jurnal Pendidikan IPA Indonesia, 5*(1), 69-74.

Kanadlı, S., & Akay, C. (2019). An investigation of the Schommers' epistemological beliefs model in terms of gender and academic success: A meta-analysis *education and science, 44*(198), 389-411. doi:10.15390/EB.2019.7992.

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*(1), 1-73.

Karaman, P., Demirci, I., & Özdemir, A. (2019). Modeling the relationship between motivation, learning approach, and academic achievement of middle school students in Turkey. *International Journal of Progressive Education, 15*(4), 187-199. doi:10.29329/ijpe.2019.203.14.

Karim, N. I., Maries, A., & Singh, C. (2018). Do evidence-based active-engagement courses reduce the gender gap in introductory physics? *European Journal of Physics, 39*(2), 1-31. doi:10.1088/1361-6404/aa9689.

Kaushik, V., & Walsh, C. A. (2019). Pragmatism as a research paradigm and its implications for social work research. *Social Sciences, 8*(255), 1-17. doi:10.3390/socsci8090255.

Kay, D., & Kibble, J. (2016). Learning theories 101: Application to everyday teaching and scholarship. *Advances in Physiology Education, 40*(1), 17-25. doi:10.1152/advan.00132.2015.

Kaya, E., Cetin, P. S., & Erduran, S. (2014). Adaptation of two argumentation tests into Turkish. *Elementary Education Online, 13*(3), 1014-1032.

Kaya, E., Erduran, S., & Cetin, P. S. (2012). Discourse, argumentation, and science lessons: Match or mismatch in high school students' perceptions and understanding. *Mevlana International Journal of Education (MIJE), 2*(3), 1-32.

Kaymak, E., & Ogan-Bekiroglu, F. (2013). How students epistemological beliefs in the domain of physics and their conceptual change are related? *European Journal of Physics Education, 4*(1), 10-25.

Keesee, G. S. (2012). Learning theories. *Learning theory and instutional design/technology*. Retrieved from http://teachinglearningresources.pbworks.com/w/page/19919565/Learning%20Theories

Keeves, J. P., & Watanabe, R. (2013). *The international handbook of educational research in the Asia-Pacific region* (Vol. 11): Springer Science & Business Media.

Kemendikbud. (2012). Dokumen kurikulum 2013. Jakarta: Kemendikbud.

Kemendikbud. (2013). Peraturan menteri pendidikan dan kebudayaan Republik Indonesia nomor 66 tahun 2013 tentang standar penilaian pendidikan. Jakarta: Kemendikbud.

Khairani, A. Z. b., & Abd Razak, N. (2015). Modeling a multiple choice mathematics test with the Rasch model. *Indian Journal of Science and Technology, 8*(12), 1-6. doi:10.17485/ijst/2015/v8i12/70650.

Khan, G.-A. N., Ishrat, N., & Khan, A. Q. (2017). Using item analysis on essay types questions given in summative examination of medical college students: Facility value, discrimination index. *International Journal of Research in Medical Sciences, 3*(1), 178-182. doi:10.5455/2320-6012.ijrms20150131.

Khine, M. S. (2010). *Knowing, knowledge and beliefs*. Australia: Springer.

Khine, M. S. (2013). *Application of structural equation modeling in educational research and practice*. The Netherlands: Sense Publishers.

Khoirina, M., Cari, C., & Sukarmin. (2018). Identify students' scientific reasoning ability at senior high school. *Journal of Physics: Conf. Series 1097*. doi:10.1088/1742-6596/1097/1/012024.

Kibble, J. D. (2017). Best practices in summative assessment. *Adv Physiol Educ, 41*, 110–119. Retrieved from doi:10.1152/advan.00116.2016.

Kim, E., & Pak, S.-J. (2002). Students do not overcome conceptual difficulties after solving 1000 traditional problems. *American Journal of Physics, 70*(7), 759-765. doi:10.1119/1.1484151.

Kimberlin, C. L., & Winterstein, A. G. (2008). Validity and reliability of measurement instruments used in research. *American Journal of Health-System Pharmacy, 65*(23), 2276-2284.

Kiong, S. S., & Sulaiman, S. B. (2010). Study of Epistemological Beliefs, Attitudes towards Learning and Conceptual Understanding of Newtonian Force Concept among Physics Education Undergraduates. In *Eprint, Universiti Teknologi Malaysia http://eprints. utm. my/14946/1/Study_of_Epistemological_Beliefs. pdf*.

Kirmizigul, A. S., & Bektas, O. (2019). Investigation of pre-service science teachers' epistemological beliefs. *Cypriot Journal of Educational Sciences, 14*(1), 146-157.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. Educational Psychologist, 41(2), 75-86.

Kizilgunes, B., Tekkaya, C., & Sungur, S. (2009). Modeling the relations among students' epistemological beliefs, motivation, learning approach, and achievement. *The Journal of Educational Research, 102*(4), 243-256.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (Third ed.). New York: The Guilford Press.

Kline, R. B. (2016). *Principle and practice of structural equation modeling, Fourth Edition*. New York: The Guilford Press.

Kline, T. (2005). *Psychological testing: A practical approach to design and evaluation*: Sage Publications.

Kock, N. (2014). Advanced mediating effects tests, multi-group analyses, and measurement model assessments in PLS-based SEM. *International Journal of e-Collaboration, 10*(1), 1-13.

Koenig, K., Wood, K. E., Bortner, L. J., & Bao, L. (2019). Modifying traditional labs to target scientific reasoning. *Journal of College Science Teaching, 48*(5), 28-35.

Koerber, S., Mayer, D., Osterhaus, C., Schwippert, K., & Sodian, B. (2015). The development of scientific thinking in elementary school: A comprehensive inventory. *Child Development, 86*(1), 327-336. doi:10.1111/cdev.12298

Kola, A. J. (2017). Investigating the conceptual understanding of physics through an Interactive Lecture-Engagement. *Cumhuriyet International Journal of Education-CIJE, 6*(1), 82-96.

Kortemeyer, G. (2007). Correlations between student discussion behavior, attitudes, and learning. *Physical Review Special Topics-Physics Education Research, 3*(1), 1-8. doi:10.1103/PhysRevSTPER.3.010101.

Kost-Smith, L. E., Pollock, S. J., & Finkelstein, N. D. (2010). Gender disparities in second-semester college physics: The incremental effects of a “smog of bias”. *Physical Review Special Topics-Physics Education Research, 6*(2), 020112, 1-17. doi:10.1103/PhysRevSTPER.6.020112.

Kost, L. E., Pollock, S. J., & Finkelstein, N. D. (2009). Characterizing the gender gap in introductory physics. *Physical Review Special Topics-Physics Education Research, 5*(1), 010101, 1-14. doi:10.1103/PhysRevSTPER.5.010101.

Krell, M., Redman, C., Mathesius, S., Krüger, D., & van Driel, J. (2018). Assessing pre-service science teachers’ scientific reasoning competencies. *Research in Science Education*, 1-25. doi:10.1007/s11165-018-9780-1.

Ku, K. Y., Lai, E. C., & Hau, K. (2014). Epistemological beliefs and the effect of authority on argument–counterargument integration: An experiment. *Thinking Skills and Creativity, 13*, 67-79. doi:10.1016/j.tsc.2014.03.004.

Kuhn, D. (2010). *What is Scientific Thinking and How Does it Develop?* (Second ed.). New York: Handbook of Childhood Cognitive Development (Blackwell).

Kuhn, D., & Dean Jr, D. (2004). Connecting scientific reasoning and causal inference. *Journal of cognition and Development, 5*(2), 261-288.

Kuhn, D., Iordanou, K., Pease, M., & Wirkala, C. (2008). Beyond control of variables: What needs to develop to achieve skilled scientific thinking? *Cognitive Development, 23*, 435-451. doi:10.1016/j.cogdev.2008.09.006.

Kuhn, D., & Udell, W. (2003). The development of argument skills. *Child Development, 74*(5), 1245-1260.

Kuhn, L., Kenyon, L. O., & Reiser, B. J. (2006). *Fostering scientific argumentation by creating a need for students to attend to each other's claims and evidence.* Paper presented at the Proceedings of the 7th International Conference on Learning Sciences.

Kwan, J. L., & Chan, W. (2011). Comparing standardized coefficients in structural equation modeling: A model reparameterization approach. *Behavior Research Methods, 43*(3), 730-745. doi:10.3758/s13428-011-0088-6.

Lacey, A., & Luff, D. (2001). *Qualitative data analysis*: The NIHR RDS for the East Midlands/Yorkshire & the Humber.

Lancaster, G., & Cooper, R. (2015). Investigation of a reflective pedagogy to encourage pre-service physics teachers to explore argumentation as an aid to conceptual understanding. *IL NUOVO CIMENTO, 38 C*(102), 1-9. doi:10.1393/ncc/i2015-15102-6.

Langcay, M., Gutierrez, J. P., Valencia, M.-M., & Tindowen, D. J. (2019). Epistemological beliefs of pre-service teachers. *Journal of Social Sciences and Humanities, 5*(2), 37-45.

Lasry, N., Rosenfield, S., Dedic, H., Dahan, A., & Reshef, O. (2011). The puzzling reliability of the Force Concept Inventory. *American Journal of Physics, 79*(9), 909-912. doi:10.1119/1.3602073.

Lawson. (1978). The development and validation of a classroom test of formal reasoning. *Journal of Research in Science Teaching, 15*(1), 11-24.

Lawson. (2000a). Classroom Test of Scientific Reasoning: Multiple choice version (Rev. ed.). Arizona: Arizona State University.

Lawson. (2000b). The generality of hypothetico-deductive reasoning: Making scientific thinking explicit. *The American Biology Teacher, 62*(7), 482-495.

Lawson. (2004). The nature and development of scientific reasoning: A synthetic view. *International Journal of Science and Mathematics Education, 2*(3), 307-338.

Lawson. (2005). What is the role of induction and deduction in reasoning and scientific inquiry? *Journal of Research in Science Teaching, 42*(6), 716-740. doi:10.1002/tea.20067.

Lawson, Banks, D. L., & Logvin, M. (2007). Self-efficacy, reasoning ability, and achievement in college biology. *Journal of Research in Science Teaching, 44*(5), 706-724. doi:10.1002/tea.20172.

Lawson, A. E., Clark, B., Cramer-Meldrum, E., Falconer, K. A., Sequist, J. M., & Kwon, Y. J. (2000). Development of scientific reasoning in college biology: Do two levels of general hypothesis-testing skills exist? *Journal of Research in Science Teaching, 37*(1), 81-101.

Lazonder, A. W., & Wiskerke-Drost, S. (2015). Advancing Scientific Reasoning in Upper Elementary Classrooms: Direct Instruction Versus Task Structuring. *Journal of Science Education and Technology, 24*(1), 69-77. doi:10.1007/s10956-014-9522-8.

Lee, C.-Q., & She, H.-C. (2010). Facilitating students' conceptual change and scientific reasoning involving the unit of combustion. *Research in Science Education, 40*(4), 479-504. doi:10.1007/s11165-009-9130-4.

Lee, M. C., & Sulaiman, F. (2018). The effectiveness of practical work in physics to improve students' academic performances. *PEOPLE: International Journal of Social Sciences, 3*(3), 1404-1419. doi:10.20319/pijss.2018.33.14041419.

Li, M., Zhang, Y., & Wang, Y. (2017). Gender gap in mathematics and physics in Chinese middle schools: A case study of a Beijing's district. In *The Urban Review* (Vol. 49, pp. 568-584): Springer.

Lising, L., & Elby, A. (2005). The impact of epistemology on learning: A case study from introductory physics. *American Journal of Physics, 73*(4), 372-382.

Little, R. J. A. & Rubin, D. B. (1987). *Statistical analysis with missing data.* New York: John Wiley.

Little, R. J. A. & Rubin, D. B. (1990), The analysis of social science data with missing values. In Fox J. & Long J. S. (eds), *Modern Methods of Data Analysis.* Newbury Park California: Sage Publications.

Liu, O. L., Bridgeman, B., & Adler, R. M. (2012). Measuring learning outcomes in higher education: Motivation matters. In *Educational Researcher* (Vol. 41, pp. 352-362).

Liu, O. L., Wilson, M., & Paek, I. (2008). A multidimensional Rasch analysis of gender differences in PISA mathematics. *Journal of Applied Measurement, 9*(1), 18-35.

Luo, X., Cappelleri, J. C., Cella, D., Li, J. Z., Charbonneau, C., Kim, S. T., . . . Motzer, R. J. (2009). Using the Rasch model to validate and enhance the interpretation of the functional assessment of cancer therapy–kidney symptom index—disease-related symptoms scale. In *Value in Health* (Vol. 12, pp. 580-586).

Madsen, A., McKagan, S. B., & Sayre, E. C. (2013). Gender gap on concept inventories in physics: What is consistent, what is inconsistent, and what factors influence the gap? *Physical Review Special Topics-Physics Education Research, 9*(2), 020121, 1-15. doi:10.1103/PhysRevSTPER.9.020121.

Madsen, A., McKagan, S. B., & Sayre, E. C. (2015). How physics instruction impacts students' beliefs about learning physics: A meta-analysis of 24 studies. *Physical Review Special Topics-Physics Education Research, 11*(1), 010115, 1-19. doi:10.1103/PhysRevSTPER.11.010115.

Makarova, E., Aeschlimann, B., & Herzog, W. (2019). *The gender gap in STEM fields: The impact of the gender stereotype of math and science on secondary students' career aspirations.* Frontiers in Education, 4, 1-11. doi: 10.3389/feduc.2019.00060.

Maley, C. R., & Bond, T. G. (2007). Measuring up for big school: a role for cognitive development. *Proceedings of AARE 2007 International Educational Research Conference: research impacts: proving or improving?* Fremantle, WA, Australia.

Mallow, J., Kastrup, H., Bryant, F. B., Hislop, N., Shefner, R., & Udo, M. (2010). Science anxiety, science attitudes, and gender: Interviews from a binational study. *Journal of Science Education and Technology, 19*(4), 356-369. doi:10.1007/s10956-010-9205-z.

Maloney, D. P., O'Kuma, T. L., Hieggelke, C. J., & Van Heuvelen, A. (2001). Surveying students' conceptual knowledge of electricity and magnetism. *American Journal of Physics, 69*(7), S12-S23. doi:10.1119/1.1371296.

Manz, E. (2015). Representing student argumentation as functionally emergent from scientific activity. *Review of Educational Research, XX*(X), 1 –38. doi:10.3102/0034654314558490.

Marchand, G. C., & Taasoobshirazi, G. (2013). Stereotype threat and women's performance in physics. *International Journal of Science Education, 35*(18), 3050-3061. doi:10.1080/09500693.2012.683461.

Martin, M. O., Mullis, I. V., Foy, P., & Arora, A. (2011). Creating and interpreting the TIMSS and PIRLS 2011 context questionnaire scales. In *Methods and Procedures in TIMSS and PIRLS* (pp. 1-11).

Martin, M. O., Mullis, I. V. S., Foy, P., & Stanco, G. M. (2012). TIMSS 2011 international results in science: TIMSS & PIRLS International Study Center.

Martin, M. O., Mullis, I. V., Foy, P., & Hopper, M. (2016). *TIMSS 2015 international results in science*. Lynch School of Education, Boston College: IEA TIMSS & PIRLS International Study Center.

Marton, F., & Säljö, R. (1976). On qualitative differences in learning: I—Outcome and process. British journal of educational psychology, 46(1), 4-11.

Marx, J., & Cummings, K. (2007). *What factors really influence shifts in students' attitudes and expectations in an introductory physics course?* Paper presented at the AIP Conference Proceedings 883. doi: 10.1063/1.2508701.

Marzooghi, R., Fouladchang, M., & Shemshiri, B. (2008). Gender and grade level differences in epistemological beliefs of Iranian undergraduate students. *Journal of Applied Sciences, 8*(24), 4698-4701.

Mason, L., & Scirica, F. (2006). Prediction of students' argumentation skills about controversial topics by epistemological understanding. *Learning and Instruction, 16*(5), 492-509. doi:10.1016/j.learninstruc.2006.09.007.

Masters, G. N. (2016). Partial credit model. In *Handbook of Item Response Theory, Volume One* (pp. 137-154). Chapman and Hall/CRC.

Mbunde, J. M. (2017). Head teachers' role in facilitating school facilities that influence pupils' performance in Kcpe. *European Scientific Journal, 13*(19), 146-155. doi:10.19044/esj.2017.v13n19p146.

McDermott, L. C. (1991). Millikan lecture 1990: What we teach and what is learned—closing the gap. *American Journal of Physics*, 59, 301–315.

McDermott, L. C., & Redish, E. F. (1999). Resource letter: PER-1: Physics education research. *American Journal of Physics, 67*(9), 755-767.

McGee, C., Cowie, B., & Cooper, B. (2010). Initial teacher eduction and the a New Zealand curriculum. *Waikato Journal of Education, 15*(1), 9-27.

McGrath, C., Palmgren, P. J., & Liljedahl, M. (2019). Twelve tips for conducting qualitative research interviews. *Medical Teacher, 41*(9), 1002-1006. doi:10.1080/0142159X.2018.1497149.

McNeill, K. L. (2011). Elementary students' views of explanation, argumentation, and evidence, and their abilities to construct arguments over the school year. *Journal of Research in Science Teaching, 48*(7), 793-823. doi:10.1002/tea.20430.

McNeill, K. L., & Knight, A. M. (2013). Teachers’ pedagogical content knowledge of scientific argumentation: The impact of professional development on K–12 teachers. *Science Education, 97*(6), 936-972. doi:10.1002/sce.21081.
Migiro, S., & Magangi, B. (2011). Mixed methods: A review of literature and the future of the new research paradigm. *African Journal of Business Management, 5*(10), 3757-3764. doi:10.5897/AJBM09.082.
Miller, D. I., & Halpern, D. F. (2014). The new science of cognitive sex differences. In *Trends in Cognitive Sciences* (Vol. 18, pp. 37-45).
MOEC. (2016). *Indonesia Educational Statistics in Brief 2015/2016*. Jakarta, Indonesia.
Momsen, J., Offerdahl, E., Kryjevskaia, M., Montplaisir, L., Anderson, E., & Grosz, N. (2013). Using assessments to investigate and compare the nature of learning in undergraduate science courses. In CBE—*Life Sciences Education* (Vol. 12, pp. 239-249).
MoNE. (2003). Act of the Republic of Indonesia number 20, year 2003 on national education system. The Minister of National Education Republic of Indonesia. Jakarta, Indonesia.
Morse, J. M. (1991). Approaches to qualitative-quantitative methodological triangulation. Nursing Research, 40(2), 120-123.
Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty’s subtle gender biases favor male students. In *Proceedings of the National Academy of Sciences* (Vol. 109, pp. 16474-16479).
Mueller, A. E., & Segal, D. L. (2014). Structured versus semistructured versus unstructured interviews. In *The Encyclopedia of Clinical Psychology* (pp. 1-7).
Mugot, D. C., & Sumbalan, E. B. (2019). The 21st Century learning skills and teaching practices of pre-service teachers: Implication to the new Philippine teacher education curriculum. *International Journal of Multidisciplinary Research and Publications (IJMRAP), 2*(1), 22-28.
Muin, W., Abedalaziz, N., Hussin, Z., Mohamed, M., & Md Saad, R. (2012). Epistemological beliefs of undergraduate students as function of gender and academic level. *OIDA International Journal of Sustainable Development, 5*(2), 109-118.
Muis, K. R., & Gierus, B. (2014). Beliefs about knowledge, knowing, and learning: Differences across knowledge types in physics. *The Journal of Experimental Education, 82*(3), 408-430. doi:10.1080/00220973.2013.813371.
Murphy, P. K., Greene, J. A., Allen, E., Baszczewski, S., Swearingen, A., Wei, L., & Butler, A. M. (2018). Fostering high school students’ conceptual understanding and argumentation performance in science through Quality Talk discussions. *Science Education, 102*(6), 1239-1264. doi:10.1002/sce.21471.
Murtonen, M. (2019). *Redefining scientific thinking for higher education: Higher-order thinking, evidence-based reasoning and research skills.* USA: Springer.
Nagowah, L., & Nagowah, S. (2009). A Reflection on the dominant learning theories: Behaviourism, cognitivism and constructivism. *International Journal of Learning, 16*(2), 1-13.
Nepal, B., & Maharjan, R. (2015). Effect of school's physical facilities on learning and outcomes of students in Nepal. *Journal for Studies in Management and Planning, 1*(6), 266-279.
Neuman, W. L. (2005). Social research methods: Quantitative and qualitative approaches. *International Journal of Medical Informatics, 81*(2), 88-97.
Newton, N. (2010). The use of semi-structured interviews in qualitative research: Strengths and weaknesses. *Exploring Qualitative Methods, 1*(1), 1-11.
Newton, P., Driver, R., & Osborne, J. (1999). The place of argumentation in the pedagogy of school science. *International Journal of Science Education, 21*(5), 553-576. doi:10.1080/095006999290570.

Nieminen, P., Savinainen, A., & Viiri, J. (2012). Relations between representational consistency, conceptual understanding of the force concept, and scientific reasoning. *Physical Review Special Topics-Physics Education Research, 8*(1), 1-10. doi:10.1103/PhysRevSTPER.8.010123.
Nieminen, P., Savinainen, A., & Viiri, J. (2013). Gender differences in learning of the concept of force, representational consistency, and scientific reasoning. *International Journal of Science and Mathematics Education, 11*(5), 1137-1156. doi:10.1007/s10763-012-9363-y.
Noroozi, O. (2018). Considering students' epistemic beliefs to facilitate their argumentative discourse and attitudinal change with a digital dialogue game. *Innovations in Education and Teaching International, 55*(3), 357-365. doi:10.1080/14703297.2016.1208112.
Noroozi, O., & Hatami, J. (2018). The effects of online peer feedback and epistemic beliefs on students' argumentation-based learning. *Innovations in Education and Teaching International*, 1-10. doi:10.1080/14703297.2018.1431143.
Novia, N., Syamsu, S., & Riandi, R. (2018). *Student's achievement in Lawson's classroom scientific reasoning (LCTSR): The effect of gender and age on scientific reasoning ability.* Paper presented at the International Conference on Mathematics and Science Education of Universitas Pendidikan Indonesia.
NRC. (1996). *National science education standards*. National Academy Press.
Nugroho, O., Permanasari, A., & Firman, H. (2019). The movement of STEM education in Indonesia: Science teachers' perspectives. *Jurnal Pendidikan IPA Indonesia, 8*(3), 417-425. doi:10.15294/jpii.v8i3.19252.
Nunkoo, R., & Ramkissoon, H. (2011). Structural equation modelling and regression analysis in tourism research. In Current Issues in Tourism (pp. 1-26). London, UK: Routledge. doi:10.1080/13683500.2011.641947
Nurlatifah, S., Tukiran, T., & Erman, E. (2018). *The development of learning material using learning cycle 7e with socio-scientific issues context in rate of reaction to improve student's argumentation skills in senior high school.* Paper presented at the Mathematics, Informatics, Science, and Education International Conference.
Nussbaum, E. M., Sinatra, G. M., & Poliquin, A. (2008). Role of epistemic beliefs and scientific argumentation in science learning. *International Journal of Science Education, 30*(15), 1977-1999. doi:10.1080/09500690701545919.
O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity, 41*(5), 673-690. doi:10.1007/s11135-006-9018-6.
Obura, C. A. (2019). The mediating role of learning strategy in the relationship between achievement goals and academic achievement. *International Journal of Learning and Development, 9*(3), 45-65. doi:10.5296/ijld.v9i3.15295.
OECD/Asian Development Bank (2015), Education in Indonesia: Rising to the Challenge, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264230750-en.
Ogan-Bekiroglu, F., & Sengul-Turgut, G. (2008). *Does constructivist teaching help students move their epistemological beliefs in physics through uppers levels?* Paper presented at the Conference of Asian Science Education.
Ogbeba, J., Odoh, C. O., & Adeke, S. (2014). Students' learning styles, gender and their effects on physics students' achievements in senior secondary schools. *Journal of Scientific Research and Studies, 1*(5), 73-80.
Öhrstedt, M., & Lindfors, P. (2019). First-semester students' capacity to predict academic achievement as related to approaches to learning. *Journal of Further and Higher Education, 43*(10), 1420-1432. doi:10.1080/0309877X.2018.1490950.

Opitz, A., Heene, M., & Fischer, F. (2017). Measuring scientific reasoning–a review of test instruments. *Educational Research and Evaluation, 23*(3-4), 78-101. doi:10.1080/13803611.2017.1338586.

Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching, 41*(10), 994-1020.

Osborne, J. F., Henderson, J. B., MacPherson, A., Szu, E., Wild, A., & Yao, S. Y. (2016). The development and validation of a learning progression for argumentation in science. *Journal of Research in Science Teaching, 53*(6), 821–846. doi:10.1002/tea.21316.

Owoeye, J. S., & Olatunde Yara, P. (2011). School facilities and academic achievement of secondary school agricultural science in Ekiti State, Nigeria. *Asian Social Science, 7*(7), 64-74. doi:10.5539/ass.v7n7p64.

Özmen, K., & Özdemir, Ö. F. (2019). Conceptualisation and development of the physics related personal epistemology questionnaire (PPEQ). *International Journal of Science Education, 41*(9), 1207-1227. doi:10.1080/09500693.2019.1597397.

Ozsevgec, L. C., & Azakli, T. K. (2018). An investigation on the vocational high school students' learning approaches in terms of various variables. *Universal Journal of Educational Research, 6*(1), 184-189. doi:10.13189/ujer.2018.060119.

Pallant, J. (2011). *SPSS survival manual* (4th ed.). NSW Australia: Allen & Unwin.

Pamuk, S., Sungur, S., & Oztekin, C. (2017). A multilevel analysis of students' science achievements in relation to their self-regulation, epistemological beliefs, learning environment perceptions, and teachers' personal characteristics. *International Journal of Science and Mathematics Education, 15*(8), 1423-1440. doi:10.1007/s10763-016-9761-7

Park, M. (2019). Effects of simulation-based formative assessments on students' conceptions in physics. *Eurasia Journal of Mathematics, Science and Technology Education, 15*(7), 1-18.

Perkins, K., Gratny, M., Adams, W., Finkelstein, N., & Wieman, C. (2006). *Towards characterizing the relationship between students' interest in and their beliefs about physics.* Paper presented at the 2005 Physics Education Research Conference.

Phanphech, P., Tanitteerapan, T., & Murphy, E. (2019). Explaining and enacting for conceptual understanding in secondary school physics. *Issues in Educational Research, 29*(1), 180-204.

Piaget, J. (1965). The stages of the intellectual development of the child. In *Educational Psychology in Context: Readings for Future Teachers* (pp. 98-106). Thousand Oaks, California: Sage Publications.

Pigott, T. D. (2001). A review of methods for missing data. *Educational Research and Evaluation, 7*(4), 353-383.

Piquero, A. R., MacIntosh, R., & Hickman, M. (2000). Does self-control affect survey response? Applying exploratory, confirmatory, and item response theory analysis to Grasmick et al.'s self-control scale. In *Criminology* (Vol. 38, pp. 897-930).

Piraksa, C., Srisawasdi, N., & Koul, R. (2014). Effect of gender on student's scientific reasoning ability: A case study in Thailand. *Procedia-Social and Behavioral Sciences, 116*, 486-491. doi:10.1016/j.sbspro.2014.01.245.

Pituch, K. A., & Stevens, J. P. (2016). *Applied multivariate statistics for the social sciences: Analyses with SAS and IBM's SPSS* (Sixth ed.). New York: Routledge.

Pollock, S. J. (2005). *No single cause: learning gains, student attitudes, and the impacts of multiple effective reforms.* Paper presented at the AIP Conference Proceedings 790.

Pritchard, A. (2017). *Ways of learning: Learning theories for the classroom* (Second ed.). London and New York: Routledge.

Psycharis, S. (2013). Examining the effect of the computational models on learning performance, scientific reasoning, epistemic beliefs and argumentation: An implication for the STEM agenda. *Computers and Education, 68*, 253-265. doi:10.1016/j.compedu.2013.05.015.

Punch, K. F. (2013). *Introduction to social research: Quantitative and qualitative approaches* (3rd Revised ed.). London, United Kingdom: Sage Publications Ltd.

Putra, A. (2019). *How student worksheet oriented of content complexity and cognitive processes can improve conceptual understanding and critical thinking skill of student in physics learning in high school.* Paper presented at the Journal of Physics: Conference Series 1185. doi:10.1088/1742-6596/1185/1/012045.

Putri, M. D., & Rusdiana, D. (2017). Identifying students' scientific argumentation skill at junior high school 1 Argamakmur, North Bengkulu. *IJAEDU- International E-Journal of Advances in Education, 3*(9), 566-572.

Pyper, B. A. (2012). *Changing scientific reasoning and conceptual understanding in college students.* Paper presented at the AIP Conference Proceedings 1413.

QSR International Pty Ltd (2012). *NVIVO qualitative software*.

QSR International Pty Ltd (2016). *NVivo 11 Pro for Windows.*

Raduta, C. (2005). General students' misconceptions related to electricity and magnetism. Retrieved from http://arxiv.org/ftp/physics/papers/0503/0503132.pdf.

Rahayu, D. P., & Widodo, A. (2019). *The profile of scientific argumentation skill student's using Toulmin argumentation pattern analysis in the solving energy problem on the students of class VII.* Paper presented at the International Conference on Educational and Information Technology.

Razak, N., Keeves, J. P., & Darmawan, I. G. N. (2014). *Developing a modelling approach and the examination of models. In I. G. N. Darmawan, & C. R. Aldous (Eds.), The Processes of Change on Learning Literacy and Numeracy in South Australian Primary Schools.* Adelaide: Shannon Research Press.

Redish, E. F., Saul, J. M., & Steinberg, R. N. (1998). Student expectations in introductory physics. *American Journal of Physics, 66*(3), 212-224.

Retnawati, H., Arlinwibowo, J., Wulandari, N. F., & Pradani, R. G. (2018). Teachers' difficulties and strategies in physics teaching and learning that applying mathematics. *Journal of Baltic Science Education, 17*(1), 120-135.

Retnawati, H., Djidu, H., Apino, E., & Anazifa, R. D. (2018). Teachers' knowledge about higher-order thinking skills and its learning strategy. In *Problems of Education in the 21st Century* (Vol. 76, pp. 1-16).

Richardson, R. M., & McCallum, W. G. (2003). The third R in literacy. In *Quantitative Literacy: Why Numeracy Matters for Schools and Colleges* (pp. 99-106).

Roadrangka, V., Yeany, R., & Padilla, M. (1982). GALT, Group Test of Logical Thinking. Athens, GA: University of Georgia.

Roohr, K. C., Liu, H., & Liu, O. L. (2017). Investigating student learning gains in college: A longitudinal study. *Studies in Higher Education, 42*(12), 2284-2300.

Rosdiana, R., Siahaan, P., & Rahman, T. (2019). *Mapping the reasoning skill of the students on pressure concept.* Paper presented at the Journal of Physics: Conference Series 1157. doi:10.1088/1742-6596/1157/2/022036.

Rosser, A. (2018). Beyond access: Making Indonesia's education system work. Australia: Lowy Institute, Melbourne University.

Rossman, G. B., & Wilson, B. L. (1985). Numbers and words: Combining quantitative and qualitative methods in a single large-scale evaluation study. *Evaluation Review*, 9(5), 627-643.

Rozgonjuk, D., Saal, K., & Täht, K. (2018). Problematic smartphone use, deep and surface approaches to learning, and social media use in lectures. *International Journal of Environmental Research and Public Health, 15*(92), 1-11. doi:10.3390/ijerph15010092.

Rubin, M., Scevak, J., Southgate, E., Macqueen, S., Williams, P., & Douglas, H. (2018). Older women, deeper learning, and greater satisfaction at university: Age and gender predict university students' learning approach and degree satisfaction. *Journal of Diversity in Higher Education, 11*, 82-96. doi:10.1037/dhe0000042.

Rush, B. R., Rankin, D. C., & White, B. J. (2016). The impact of item-writing flaws and item complexity on examination item difficulty and discrimination value. In *BMC Medical Education* (Vol. 16, pp. 1-10).

Rusmansyah, R., Yuanita, L., Ibrahim, M., Isnawati, I., & Prahani, B. K. (2019). Innovative chemistry learning model: Improving the critical thinking skill and self-efficacy of pre-service chemistry teachers. *Journal of Technology and Science Education, 9*(1), 59-76. doi:10.3926/jotse.555.

Sadler, T. D., & Zeidler, D. L. (2005a). Patterns of informal reasoning in the context of socioscientific decision making. *Journal of Research in Science Teaching, 42*(1), 112-138.

Sadler, T. D., & Zeidler, D. L. (2005b). The significance of content knowledge for informal reasoning regarding socioscientific issues: Applying genetics knowledge to genetic engineering issues. *Science education, 89*(1), 71-93. doi:10.1002/sce.20023.

Salminen, T., & Marttunen, M. (2018). Defending either a personal or an assigned standpoint. *Journal of Argumentation in Context, 7*(1), 72-100. doi:10.1075/jaic.17015.sal.

Sampson, V., & Blanchard, M. R. (2012). Science teachers and scientific argumentation: Trends in views and practice. *Journal of Research in Science Teaching, 49*(9), 1122-1148. doi:10.1002/tea.21037.

Sampson, V., & Clark, D. (2006). *The development and validation of the nature of science as argument questionnaire (NSAAQ).* Paper presented at the Annual Conference of the National Association for Research in Science Teaching, San Francisco, CA.

Sampson, V., & Clark, D. (2009). The impact of collaboration on the outcomes of scientific argumentation. In *Science Education* (Vol. 93, pp. 448-484): Wiley InterScience.

Sampson, V. D., & Clark, D. B. (2006). *Assessment of argument in science education: A critical review of the literature.* Paper presented at the Proceedings of the 7th International Conference on Learning Sciences.

Sandoval, W. A., & Millwood, K. A. (2007). What can argumentation tell us about epistemology? In *Argumentation in Science Education* (pp. 71-88). Springer.

Saputro, D., Sarwanto, S., Sukarmin, S., & Ratnasari, D. (2019). *Pre-services science teachers' conceptual understanding level on several electricity concepts.* Paper presented at the Journal of Physics: Conference Series 1157. doi:10.1088/1742-6596/1157/3/032018.

Savinainen, A., & Scott, P. (2002). The Force Concept Inventory: A tool for monitoring student learning. In *Physics Education* (Vol. 37, pp. 45-52).

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. In *Methods of Psychological Research Online* (Vol. 8, pp. 23-74).

Schleicher, A. (2019). PISA 2018: Insights and interpretations. OECD.

Schmaltz, R. M., Jansen, E., & Wenckowski, N. (2017). Redefining critical thinking: teaching students to think like scientists. *Frontiers in Psychology, 8*(459), 1-4. doi:10.3389/fpsyg.2017.00459.

Schommer-Aikins, M., & Duell, O. K. (2013). Domain specific and general epistemological beliefs. Their effects on mathematics. *Revista de Investigación Educativa, 31*(2), 317-330. doi:10.6018/rie.31.2.170911.

Schommer-Aikins, M., & Easter, M. (2006). Ways of knowing and epistemological beliefs: Combined effect on academic performance. In *Educational Psychology* (Vol. 26, pp. 411-423).

Schommer, M. (1990). Effects of beliefs about the nature of knowledge on comprehension. *Journal of Educational Psychology, 82*(3), 498-504.

Schommer, M. (1994a). *An emerging conceptualization of epistemological beliefs and their role in learning. In R. Garner & P. A. Alexander (Eds.), Beliefs about Text and Instruction with Text.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Schommer, M. (1994b). Synthesizing epistemological belief research: Tentative understandings and provocative confusions. In *Educational Psychology Review* (Vol. 6, pp. 293-319).

Schommer, M., Crouse, A., & Rhodes, N. (1992). Epistemological beliefs and mathematical text comprehension: Believing it is simple does not make it so. *Journal of Educational Psychology, 84*(4), 435-443.

Schoua-Glusberg, H. J. (1998). A questionnaires in translation. In *ZUMA-Nachrichten Spezial* (pp. 87-126).

Schraw, G. (2013). Conceptual integration and measurement of epistemological and ontological beliefs in educational research. *ISRN Education, 2013*, 1-19. doi:10.1155/2013/327680.

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research, 99*(6), 323-338. doi:10.3200/ JOER.99.6.323-338.

Schultze, U., & Avital, M. (2011). Designing interviews to generate rich data for information systems research. In *Information and Organization* (Vol. 21, pp. 1-16). Elsevier Ltd.

Schumacker, R. E. (2004). Rasch measurement: The dichotomous model. In E. V. Smith Jr. & R. M. Smith (Eds.). In *Introduction to Rasch Measurement* (pp. 226-257). Maple Grove: MN: JAM Press.

Schumacker, R. E., & Lomax, R. G. (2010). *A beginner’s guide to structural equation modeling* (Third ed.). New York: Routledge.

Schumacker, R. E., & Lomax, R. G. (2016). *A beginner's guide to structural equation modeling* (Fourth ed.). New York: Routledge.

Schunk, D. H. (1991). Learning theories: An educational perspective. New York: Macmillan.

Scott, P., Asoko, H., Leach, J., Abell, S., & Lederman, N. (2007). *Student conceptions and conceptual learning in science*. UK: Erlbaum Publishers

Sellin, N., & Keeves, J. P. (1997). Path analysis with latent variables. In Keeves, J. P. (Ed.). In *Educational Research, Methodology, and Measurement: An International Handbook* (pp. 633-640). Oxford: Pergamon Press.

Seyranian, V., Madva, A., Duong, N., Abramzon, N., Tibbetts, Y., & Harackiewicz, J. M. (2018). The longitudinal effects of STEM identity and gender on flourishing and achievement in college physics. *International Journal of STEM Education, 5*(40), 1-14.

Shapiro, J. R., & Williams, A. M. (2012). The role of stereotype threats in undermining girls’ and women’s performance and interest in STEM fields. In *Sex Roles* (Vol. 66, pp. 175-183). Springer.

Shieh, Y.-Y. (2003). *Imputation methods on general linear mixed models of longitudinal studies.* Paper presented at the Federal Committee on Statistical Methodology Research Conference, Arlington, VA.

Siegel, H. (1998). Education reason: Rationality, critical thinking, and education. New York: Routledge, Chapman & Hall.

Sobremisana, V. S. (2017). Use of physics innovative device for improving students' motivation and performance in learning selected concepts in physics. *Asia Pacific Journal of Multidisciplinary Research, 5*(4), 1-9.
Stathopoulou, C., & Vosniadou, S. (2007). Exploring the relationship between physics-related epistemological beliefs and physics understanding. *Contemporary Educational Psychology, 32*(3), 255-281. doi:10.1016/j.cedpsych.2005.12.002.
Suárez, N., Regueiro, B., Epstein, J. L., Piñeiro, I., Díaz, S. M., & Valle, A. (2016). Homework involvement and academic achievement of native and immigrant students. *Frontiers in Psychology, 7*(1517), 1-8. doi:10.3389/fpsyg.2016.01517.
Suciatmoko, P. M., Suparmi, A., & Sukarmin, S. (2018). *An analysis of students' conceptual understanding: How do students understand some electricity concepts?* Paper presented at the AIP Conference Proceedings 2014. https://doi.org/10.1063/1.5054558.
Suprapto, N. (2014). Role of physics' questions on the improvement of thinking skills: A case of Indonesian student. *International Journal of Education and Research, 2*(12), 71-82.
Taasoobshirazi, G., & Hickey, D. T. (2005). Promoting argumentative discourse: A design-based implementation and refinement of an astronomy multimedia curriculum, assessment model, and learning environment. *Astronomy Education Review, 4*(1), 53-70.
Taasoobshirazi, G., & Sinatra, G. M. (2011). A structural equation model of conceptual change in physics. *Journal of Research in Science Teaching, 48*(8), 901-918. doi:10.1002/tea.20434.
Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics (4th ed.)*. Boston: MA: Allyn and Bacon.
Taber, K. S. (2017). The use of cronbach's alpha when developing and reporting research instruments in science education. In *Research in Science Education* (pp. 1-24). Springer.
Talib, C. A., Rajan, S. T., Hakim, N. W. A., Malik, A. M. A., Siang, K. H., & Ali, M. (2018). *Gender difference as a factor in fostering scientific reasoning skill among students.* Paper presented at the 10th International Conference on Engineering Education (ICEED).
Tanriverdi, B. (2012). Pre-service teachers' epistemological beliefs and approaches to learning. *Procedia-Social and Behavioral Sciences, 46*, 2635-2642. doi:10.1016/j.sbspro.2012.05.538.
Tashakkori, A., & Teddlie, C. (2010). *Handbook of mixed methods in social & behavioral research*. Thousand Oaks: Sage Publications.
Taştan, S. B., Davoudi, S. M. M., Masalimova, A. R., Bersanov, A. S., Kurbanov, R. A., Boiarchuk, A. V., & Pavlushin, A. A. (2018). The impacts of teacher's efficacy and motivation on student's academic achievement in science education among secondary and high school students. *EURASIA Journal of Mathematics Science and Technology Education, 14*(6), 2353-2366.
Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences*. London: Sage.
Terzi, A. R., Çetin, G., & Eser, H. (2012). The relationship between undergraduate students' locus of control and epistemological beliefs. *Educational Research, 3*(1), 30-39.
Thien, L. M., Razak, N. A., Keeves, J., & Darmawan, I. G. N. (2016). *What can PISA 2012 data tell us? Performance and challenges in five participating Southeast Asian countries*. Netherlands: Sense Publishers.
Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications.* Washington, DC: American Psychological Association.
Thornton, R. K., & Sokoloff, D. R. (1998). Assessing student learning of Newton's laws: The force and motion conceptual evaluation and the evaluation of active learning laboratory and lecture curricula. *American Journal of Physics, 66*(4), 338-352.

Thuneberg, H., Hautamäki, J., & Hotulainen, R. (2015). Scientific reasoning, school ahievement and gender: A multilevel study of between and within school effects in Finland. *Scandinavian Journal of Educational Research, 59*(3), 337-356.
Tiruneh, D. T., De Cock, M., Weldeslassie, A. G., Elen, J., & Janssen, R. (2017). Measuring critical thinking in physics: Development and validation of a critical thinking test in electricity and magnetism. *International Journal of Science and Mathematics Education, 15*(4), 663-682. doi:10.1007/s10763-016-9723-0.
Tobin, K. G., & Capie, W. (1981). The development and validation of a group test of logical thinking. In *Educational and Psychological Measurement* (Vol. 41, pp. 413-423).
Tonjo, V. A., Wirjawan, J. V. D., & Untung, G. B. (2017). Application of direct instruction with laboratory activity to improve students' participation and learning achievement. *PEOPLE: International Journal of Social Sciences, 3*(2), 1276-1284.
Toulmin, S. E., & Rieke, R. D. J. (1984). *An introduction to reasoning* (Second ed.). New York: Macmillan.
Traxler, A., Henderson, R., Stewart, J., Stewart, G., Papak, A., & Lindell, R. (2018). Gender fairness within the Force Concept Inventory. *Physical Review Physics Education Research, 14*(1), 010103, 1-17. doi:10.1103/PhysRevPhysEducRes.14.010103.
Treagust, D. F. (1995). Diagnostic assessment of students' science knowledge. In *Learning Science in the Schools: Research Reforming Practice* (Vol. 1, pp. 327-436).
Treagust, D. F., & Chandrasegaran, A. (2007). Assessment of students' conceptual understandings in science: The Taiwan national science concept learning study in an international perspective. *International Journal of Science Education, 29*(4), 391-403.
Tuder, I., & Urban-Woldron, H. (2015). Conceptual understanding of Newton's second law – Looking behind the FCI. In *Learning Science: Conceptual Understanding* (pp. 109-115). University of Vienna: AECC Physics.
Tumkaya, S. (2012). The investigation of the epistemological beliefs of university students according to gender, grade, fields of study, academic success and their learning styles. *Educational Sciences: Theory and Practice, 12*(1), 88-95.
Usmeldi, U. (2016). The development of research-based physics learning model with scientific approach to develop students' scientific processing skill. *Jurnal Pendidikan IPA Indonesia, 5*(1), 134-139. doi:10.15294/jpii.v5i1.5802.
van Griethuijsen, R. A., van Eijck, M. W., Haste, H., den Brok, P. J., Skinner, N. C., Mansour, N., . . . BouJaoude, S. (2015). Global patterns in students' views of science and interest in science. *Research in Science Education, 45*(4), 581-603. doi:10.1007/s11165-014-9438-6.
Vecaldo, R. T. (2017). Epistemological beliefs, academic performance and teaching competence of pre-service teachers. *Asia Pacific Journal of Multidisciplinary Research, 5*(2), 114-124.
Veletsianos, G. (2016). *Emergence and innovation in digital learning: Foundations and applications*. Street, Edmonton: Athabasca University Press.
Venville, G. J., & Dawson, V. M. (2010). The impact of a classroom intervention on grade 10 students' argumentation skills, informal reasoning, and conceptual understanding of science. *Journal of Research in Science Teaching, 47*(8), 952-977. doi:10.1002/tea.20358.
Von Aufschnaiter, C., Erduran, S., Osborne, J., & Simon, S. (2008). Arguing to learn and learning to argue: Case studies of how students' argumentation relates to their scientific knowledge. *Journal of Research in Science Teaching, 45*(1), 101-131. doi:10.1002/tea.20213.
Von Korff, J., Archibeque, B., Gomez, K. A., Heckendorf, T., McKagan, S. B., Sayre, E. C., . . . Sorell, L. (2016). Secondary analysis of teaching methods in introductory physics: A 50 k-student study. *American Journal of Physics, 84*(12), 969-974.

Vosniadou, S., Ioannides, C., Dimitrakopoulou, A., & Papademetriou, E. (2001). Designing learning environments to promote conceptual change in science. *Learning and Instruction, 11*(4), 381-419.

Wang, J., & Buck, G. A. (2016). Understanding a high school physics teacher's pedagogical content knowledge of argumentation. *Journal of Science Teacher Education, 27*(5), 577-604. doi:10.1007/s10972-016-9476-1.

Wang, M.-T., & Degol, J. L. (2017). Gender gap in science, technology, engineering, and mathematics (STEM): Current knowledge, implications for practice, policy, and future directions. *Educational Psychology Review, 29*(1), 119-140. doi:10.1007/s10648-015-9355-x.

Wang, S., & Wang, T. (2001). Precision of Warm's weighted likelihood estimates for a polytomous model in computerized adaptive testing. In *Applied Psychological Measurement* (Vol. 25, pp. 317-331): Sage Publications.

Wang, W.-C., Yao, G., Tsai, Y.-J., Wang, J.-D., & Hsieh, C.-L. (2006). Validating, improving reliability, and estimating correlation of the four subscales in the WHOQOL-BREF using multidimensional Rasch analysis. In *Quality of Life Research* (Vol. 15, pp. 607-620).

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. Psychometrika, 54(3), 427-450.

Weinstock, M., & Cronin, M. A. (2003). The everyday production of knowledge: Individual differences in epistemological understanding and juror-reasoning skill. *Applied Cognitive Psychology, 17*(2), 161-181. doi:10.1002/acp.860.

Weller, S. C., Vickers, B., Bernard, H. R., Blackburn, A. M., Borgatti, S., Gravlee, C. C., & Johnson, J. C. (2018). Open-ended interview questions and saturation. *PloS one, 13*(6), e0198606, 1-18.

Wenning, C. J., & Vierya, R. (2015). *Teaching high school physics, Vol. I*. United States.

Weston, R., & Gore Jr, P. A. (2006). A brief guide to structural equation modeling. In *The Counseling Psychologist* (Vol. 34, pp. 719-751).

White, B. (2011). *Mapping your thesis: The comprehensive manual of theory and techniques for masters and doctoral research Victoria*. Australia: ACER Press.

Widiyatmoko, A., & Shimizu, K. (2018). Literature review of factors contributing to students' misconceptions in light and optical instruments. *International Journal of Environmental and Science Education, 13*(10), 853-863.

Widodo, A., Waldrip, B., & Herawati, D. (2016). Students argumentation in science lessons: A story of two research projects. *Jurnal Pendidikan IPA Indonesia, 5*(2), 199-208. doi:10.15294/jpii.v5i2.5949.

Wiggins, G. P., & McTighe, J. (2011). *The understanding by design guide to creating high-quality units*. Alexandria, VA: ASCD.

Wilson, M. (2005). Constructing measures: An item response theory approach. *Lawrence Erlbaum, Mahwah*.

Winberg, T. M., Hofverberg, A., & Lindfors, M. (2019). Relationships between epistemic beliefs and achievement goals: Developmental trends over grades 5–11. *European Journal of Psychology of Education, 34*(2), 295-315. doi:doi.org/10.1007/s10212-018-0391-z.

Woolfolk, A. (2005). *Educational psychology, 9th ed.* Boston: Allyn & Bacon.

Wright, B. D. (1994). Reasonable mean-square fit values. *Rasch Meas Trans, 8*, 370.

Wright, B. D., & Mok, M. (2000). Understanding Rasch measurement: Rasch models overview. *Journal of Applied Measurement, 1*(1), 83-106.

Wright, B. D., & Mok, M. M. (2004). An overview of the family of Rasch measurement models. In *Introduction to Rasch Measurement* (pp. 1-24).

Wu, M. L., & Adams, R. (2007). *Applying the Rasch model to psycho-social measurement: A practical approach*. Melbourne: Educational Measurement Solutions.

Wu, M. L., Adams, R., Wilson, M. R., & Haldane, S. A. (2007). ACER ConQuest: generalized item response modeling software (version 2). *Camberwell, Australia: ACER Press, an imprint of Australian Council for Educational Research Ltd*.

Yalcin, M., & Yalcin, F. A. (2017). The investigation of pre service science teachers' epistemological beliefs according to some variables. *Journal of Education and Training Studies, 5*(10), 207-217. doi:10.11114/jets.v5i10.2224.

Yamtinah, S., Masykuri, M., Ashadi, & Shidiq, A. S. (2017). *Gender differences in students' attitudes toward science: An analysis of students' science process skill using testlet instrument.* Paper presented at the AIP Conference Proceedings, 1868. https://doi.org/10.1063/1.4995102.

Yang, F.-Y., Huang, R.-T., & Tsai, I.-J. (2016). The effects of epistemic beliefs in science and gender difference on university students' science-text reading: An eye-tracking study. *International Journal of Science and Mathematics Education, 14*(3), 473-498.

Yang, F. Y. (2004). Exploring high school students' use of theory and evidence in an everyday context: The role of scientific thinking in environmental science decision-making. *International Journal of Science Education, 26*(11), 1345-1364.

Yenice, N. (2015). An analysis of science student teachers' epistemological beliefs and metacognitive perceptions about the nature of science. *Educational Sciences: Theory and Practice, 15*(6), 1623-1636. doi:10.12738/estp.2015.6.2613.

Yu, L., & Shek, D. T. (2014). Family functioning, positive youth development, and internet addiction in junior secondary school students: Structural equation modeling using AMOS. *International Journal on Disability and Human Development, 13*(2), 227-238. doi:10.1515/ijdhd-2014-0308.

Yu, L., Shek, D. T., & Zhu, X. (2018). General education learning outcomes and demographic correlates in university students in Hong Kong. In *Applied Research Quality Life* (pp. 1-18): Springer.

Yue, X. (2019). Exploring effective methods of teacher professional development in university for 21st century education. *International Journal of Innovation Education and Research, 7*(5), 248-257. doi:10.31686/ijier.Vol7.Iss5.1506.

Zeineddin, A., & Abd-El-Khalick, F. (2010). Scientific reasoning and epistemological commitments: Coordination of theory and evidence among college science students. *Journal of Research in Science Teaching, 47*(9), 1064-1093. doi:10.1002/tea.20368.

Zhang, J., & Lu, T. (2007). Refinement of a bias-correction procedure for the weighted likelihood estimator of ability. In *ETS Research Report Series* (Vol. 2007, pp. 1-25).

Zhang, P., Ding, L., & Mazur, E. (2017). Peer Instruction in introductory physics: A method to bring about positive changes in students' attitudes and beliefs. *Physical Review Physics Education Research, 13*(1), 010104, 1-9. doi:10.1103/PhysRevPhysEducRes.13.010104.

Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review, 27*(2), 172-223. doi:10.1016/j.dr.2006.12.001.

Zimmerman, C., & Klahr, D. (2018). *Development of scientific thinking* (Fourth ed.). In Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience. doi:10.1002/9781119170174.epcn407.

Zohar, A. (2007). Science teacher education and professional development in argumentation. In S. Erduran & M. P. Jime´nez-Aleixandre (Eds.), *Argumentation in Science Education* (pp. 245–268). Netherlands: Springer.

APPENDICES

Appendix 1. Epistemological Beliefs Assessment for Physical Science (EBAPS)

“Removed due to copyright restriction.” The EBAPS survey can be viewed online at https://www.physport.org/assessments/.

Appendix 2. Argumentation Test (Reproduced with permission)

Part I: Making a Scientific Argument

Introduction: Once a scientist develops an explanation for why something happens, he or she must support the claim with an argument. The argument is used by the scientist in order to convince others that their claim is indeed true. How do you think scientists create convincing arguments?

Directions: The first three questions are designed to determine what you think counts as a good scientific argument. In each question, you will be given a claim. Following the claim are 6 different arguments. Your job is to rank the arguments in order using the following scale:

1 = This is the **most** convincing argument
2 = This is the **2nd most** convincing argument
3 = This is the **3rd most** convincing argument
4 = This is the **4th most** convincing argument
5 = This is the **5th most** convincing argument
6 = This is the **least** convincing argument

For each question, you can only use each ranking once

Question #1. Your task is to rank these 6 different arguments in terms of how convincing you think they are. *Remember that you can only rank one claim as 1, one claim as 2, one claim as 3, and so on.*

Claim: Objects that are in the same room are the same temperature even though they feel different because…	**Your Ranking**
…when we measured the temperature of the table, it was 23.4^{O}C, the metal chair leg was 23.1^{O}C, and the computer keyboard was 23.6^{O}C. (evidence only)	
…good conductors feel different than poor conductors even though they are the same temperature. (warrant only)	
…objects that are in the same environment gain or lose heat energy until everything is the same temperature. Our data form the lab proves that point: the mouse pad and plastic desk were both 23^{O}C. (explanation and evidence)	
…objects will release and hold different amounts of heat energy depending on how good of an insulator or conductor it is. (contradictory)	
…our textbook says that all objects in the same room will eventually reach the same temperature. (appeal to authority)	
…we measured the temperature of the wooden table and the chair leg and they were both 23^{O}C even though the metal chair leg feels colder. If the metal chair leg was actually colder it would have been a lower temperature when we compared it to the temperature of the table. (data, explanation and rebuttal)	

Question #2. Your task is to rank these 6 different arguments in terms of how convincing you think they are. Remember that you can only rank one claim as 1, one claim as 2, one claim as 3, and so on.

A pendulum is a string with a weght attached to one end of it. Suppose someone makes the following claim about pendulums:

Claim: The length of the string determines how fast a pendulum swings back and forth regardless of the weight on the end of the string because…	**Your Ranking**
…the weight on the end of a long string has a longer distance to travel when compared to a weight on a short string. As a result, pendulums with shorter swings make more swings per second than pendulum with longer strings. (warrant only)	
…we measured the swing rate of two different pendulums, one had a 10 cm string and the other two had a 20 cm string, The 10 cm pendulum had swing rate of 2 swings per second and the 20 cm pendulum has a swing rate of 1 swing per second. If the string length did not matter these two pendulums would have had the same swing rate (explanation and evidence)	
…a pendulum with a 14 cm string had a swing rate of 1 swing per second and a pendulum with a 15 cm string had a swing rate of 1 swing per second. (contradictory)	
…a pendulum with a 10 cm string had a swing rate of 2 swings per second and a pendulum with a 15 cm string had a swing rate of 1 swing per second. (evidence only)	
…our textbook says that the weight on the end of the string has nothing to do with how fast a pendulum swings. (appeal to authority)	
…we tested the swing rate of three pendulums, one with a 10 gram weight and 10 cm string, one with a 10 gram weight and 20 cm string, and one with 20 gram weight and a 20 cm string. The two pendulums with the 20 cm string had the same swing rate (1 swing per second) and were slower the pendulum with the shorter string (2 swings per second). If the weight on the end of the string mattered these two pendulums would have had different swing rates but the they were the same (data, explanation and rebuttal)	

Question #3. Your task is to rank these 6 different arguments in terms of how convincing you think they are. *Remember that you can only rank one claim as 1, one claim as 2, one claim as 3, and so on.*

Claim: Scientists should be allowed to use animals for research because…	**Your Ranking**
…a computer or other non animal model can be used instead. (contradictory)	
…animals are susceptible to many of the same bacteria and viruses as people, such as anthrax, smallpox, and malaria. Even though animals differ from people in many ways, they also are very similar to people in many ways. An animal is chosen for research only if it shares characteristics with people that are relevant to the research. (data, explanation and rebuttal)	
…public opinion polls have consistently shown that a majority of people approve of the use of animals in biomedical research that does not cause pain to the animal and leads to new treatments and cures. (appeal to authority)	
…animal research was essential in developing many life-saving surgical procedures once thought impossible. For example the technique of sewing blood vessels together was developed through surgeries on dogs and cats by Alexis Carrel, for which he was awarded a Nobel Prize in 1912. (explanation and evidence)	
…infecting animals with certain microbes allows researchers to identify the germs that cause different types of diseases. Once discovered scientists can develop vaccines to test the effectiveness of these vaccines without harming any people in the process. (warrant only)	
…humans have 65 infectious diseases in common with dogs, 50 with cattle, 46 with sheep and goats, 42 with pigs, 35 with horses, and 26 with fowl. (evidence only)	

Part II. Challenging Arguments

Introduction: Once a scientist develops an explanation for why something happens, he or she must support the claim with an argument. Sometimes other scientists agree with the argument; sometimes they do not. When they disagree, they challenge the accuarcy of the claim. How do you think scientists challenge the claims of other scientists? The last three questions on this test are designed to determine what you think counts as a good challenge to a s*cientific* argument.

Directions: In each question you will be given a claim supported by an argument. Following the claim are 6 different challenges. Your job is to rank the arguments in order using the following scale (**For each question, you can only use each ranking once)**:

1 = This comment is the **strongest** challenge to this argument
2 = This comment is the **2nd strongest** challenge to this argument
3 = This comment is the **3rd strongest** challenge to this argument
4 = This comment is the **4th strongest** challenge to this argument
5 = This comment is the **5th strongest** challenge to this argument
6 = This comment is the **weakest** challenge to this argument

Question #4—Jason, Angela, Sarah, and Tim are in physics class together. Their teacher asked them to design an experiment to determine if all objects in the same room are the same temperature even though they feel different. After they designed and carried out an experiment to answer this question on their own, they met in a small group to discuss what they have found out. Suppose Jason suggests that:

> *"I think that all objects in the same room are always different temperatures because they feel different and when we measured the temperature of the table, it was 23.4^{0}C, the metal chair leg was 23.1^{0}C, and the computer keyboard was 23.6^{0}C."*

Angela disagrees with Jason. Your task is to rank these 6 different challenges in terms of how strong you think they are. *Remember that you can only rank one challenge as 1, one challenge as 2, one challenge as 3, and so on.*

Angela: I disagree…	**Your Ranking**
…because your evidence does not support your claim. All of the objects that you measured were within one degree of each other. That small of difference is just measurement error. (rebuttal against grounds no grounds)	
…I think that all objects in the same room are the same temperature even though they feel different (counter-claim only)	
…if those objects were really different temperatures their temperature would have been much different. For example, when I measured the temperature of my arm it was 37^{O}C while the temperature of the table was 23^{O}C that is a difference of 14 degrees. Everything else was right around 23^{O}C. (rebuttal against grounds with grounds)	
…I think all objects become the same temperature even though they feel different because objects that are good conductors feel colder than objects that are poor conductors because heat transfers through good conductors faster (rebuttal against thesis with grounds)	
…because I know you always rush through labs and never get the right answer (emotive)	
…I think all objects become the same temperature because the temperature of all those objects you measured were within 1 degree (rebuttal against thesis no grounds)	

Question #5—Tiffany, Steven, and Yelena are in the same science class. Their teacher asked them to design an experiment to determine what makes some objects float and some objects sink. After they designed and carried out an experiment to answer this question on their own, they met in a small group to discuss what they have found out. Suppose Steven suggests that:

> *"I think heavy objects sink and light objects float. I know this is true because when I put the plastic block, which was 10 grams, in the tub of water it floated while the metal block, which was 40 grams, sank."*

Tiffany disagrees with Steven.
Your task is to rank these 6 different challenges in terms of how strong you think they are. *Remember that you can only rank one challenge as 1, one challenge as 2, one challenge as 3, and so on*

Tiffany: I disagree…	**Your Ranking**
…because Yelena is always right and she disagrees with you (emotive)	
…because you did not test enough objects. How can you be sure that it is the weight of an object that makes it sink or float if you only tested two things? (rebuttal against grounds with no grounds)	
…the metal block sank because it is very dense not because it is heavy and the plastic block floated because it has density that is less than water not because it is light (rebuttal against thesis no grounds)	
…because light objects can sink too. A paper clip only weighs one gram and it sinks. According to you claim all light objects should float. How can a paper clip that is lighter than a piece of plastic sink while the heavier piece of plastic floats? (rebuttal against thesis with grounds)	
…The plastic block may have been lighter than the metal block but that is not why it floated. The metal block has a density of 2.5 g/cm3, which is more than water so it sinks. The plastic block has a volume 16 cm3 which means its density is .6 g/cm3 which is less than water so it floats. (rebuttal against grounds with grounds)	
…I think objects that have a density greater than water sink and objects that have a density less than water float (counter-claim only)	

Question #6— Elena, Shauna, and Sam are in a science class together. At the beginning of class, their teacher poses the following question: "Should scientists be able to use animals in medical research?" The teacher then asked Elena, Shauna, and Sam to discuss what they think about the issue in a small group. Suppose Shauna begins the conversation by saying:

> *"I think using animals in medical is a bad idea because people and animals suffer from different disease and the bodies of animals and humans are completely different. So how can scientists justify performing painful experiments on animals if they are so different?'*

Sam disagrees with Shauna. Your task is to rank these 6 different challenges in terms of how strong you thinkthey are. *Remember that you can only rank one challenge as 1, one challenge as 2, one challenge as 3, and so on.*

Sam: I disagree…	**Your Ranking**
…I think using animals in medical research is a good idea because even with all of the advances in modern science, it would be impossible to prove that a specific germ is responsible for a disease without the use of laboratory animals. (rebuttal against thesis with grounds)	
… I think using animals in medical research is very useful and a good idea (counter-claim only)	
…although animals are different from people in many ways, they still are susceptible to many of the same bacteria and viruses that we are including anthrax, smallpox, and malaria. (rebuttal against grounds no grounds)	
…because you don't know what you are talking about. You just care more about animals then you do about people. (emotive)	
…an animal is chosen for research only if it shares characteristics with people that are relevant to the research. For example; animals share many of the same organs as people so they can be used to develop new surgical techniques. Organ transplantation, open heart surgery, and many other common procedures were developed using animal models without risking human life in the process. (rebuttal against grounds with grounds)	
… I think using animals in medical research is a good idea not a bad idea because even though animal and human bodies are comepletely different like you say I would rather have scientists experimenting on animals instead of humans. (rebuttal against thesis no grounds)	

Appendix 3. Lawson's Classroom Test of Scientific Reasoning (LCTSR)

"Removed due to copyright restriction." The LCTSR survey can be viewed online at https://www.physport.org/assessments/.

Appendix 4. Force Concept Inventory (FCI)

"Removed due to copyright restriction." The FCI survey can be viewed online at https://www.physport.org/assessments/.

Appendix 5. Brief Electricity and Magnetism Assessment (BEMA)

“Removed due to copyright restriction.” The BEMA survey can be viewed online at

https://www.physport.org/assessments/.

Appendix 6. Interview Questions

Q1. Do you have any comments or opinions about your experiences in studying physics subject?

Q2. In your opinion, which is more difficult between the Mechanics concept and Electricity and Magnetism concept?

Q3. In your opinion, is there any relationship between epistemological beliefs and the understanding of Physics concepts?

Q4. In your opinion, is there any relationship between argumentation and physics conceptual understanding?

Q5. In your opinion, is there any relationship between scientific reasoning and physics conceptual understanding?

Q6. What drove you to major in Physics Education? Was there any intervention from your parents?

Q7. Is it important for a prospective Physics teacher or a Physics teacher to have scientific thinking skills?

Q8. In your opinion, which factors can foster or enhance scientific thinking skills?

Q9. Have the learning facilities and sources provided helped you with your scientific thinking and Physics conceptual understanding?

Q10. Is there a conducive learning environment in class?

Q11. Which teaching method is the most frequently used in the classroom?

Q12. Has the Physics instruction in the classroom fostered you to think scientifically and understand the physics concepts well?

Q13. What kind of instructional media is often used in delivering Physics materials in the classroom? And according to you, are those learning media adequate to foster your scientific thinking and physics conceptual understanding correctly?

Q14. Could you please explain the learning activities usually conducted in the classroom?

Q15. Is there any homework assignment that must be worked on and submitted?
And do students receive feedback from lecturers after completing the assignment given in the classroom?

Q16. In your opinion, are the testing instruments employed in mid-semester and final semester exams appropriate for measuring students' scientific thinking and Physics conceptual understanding?

Q17. Which learning approach is the most frequently used to study physics?

Q18. In your opinion, what should teachers do to help their students think scientifically and understand the physics concepts well?

Q19. With the physics learning system that you have experienced in class, will your scientific thinking skills and understanding of physics concepts continue to improve until the end of your studies at university?

Q20. Understanding the instructional system has been applied thus far, which strategy that you can use to improve your scientific thinking skills and Physics conceptual understanding?

Q21. Which learning strategy do you think is suitable to study physics materials?

Appendix 7. Ethics Approval

Human Research Ethics
Tue 4/11/2017 8:36 AM
Dear Lina,

The Chair of the Social and Behavioural Research Ethics Committee (SBREC) at Flinders University considered your response to conditional approval out of session and your project has now been granted final ethics approval. This means that you now have approval to commence your research. Your ethics final approval notice can be found below.

FINAL APPROVAL NOTICE

Project No.:	**7606**
Project Title:	An Investigation of Scientific Thinking and Conceptual Understanding in Physics of Indonesian Teacher Candidates
Principal Researcher:	Mrs. Lina Aviyanti
Email:	aviy0002@flinders.edu.au

Approval Date:	10 April 2017	Ethics Approval Expiry Date:	**23 May 2020**

The above proposed project has been **approved** on the basis of the information contained in the application, its attachments and the information subsequently provided with the addition of the following comment(s):

Additional information required following commencement of research:

1. Permissions
   Please ensure that copies of the correspondence granting permission to conduct the research.

Appendix 8. Letter of Introduction

Dr. Carol Aldous
School of Education
Faculty of Education, Humanities and Law
Room # 5.60, Education Building Level 5
Flinders University, Bedford Park SA 5042
GPO Box 2100
Adelaide SA 5001
Email: carol.aldous@flinders.edu.au
http://www.flinders.edu.au
CRICOS Provider No. 00114A

LETTER OF INTRODUCTION

Dear Student,

This letter is to introduce Ms. Lina Aviyanti who is a PhD student in the School of Education at Flinders University. She will produce her student card, which carries a photograph, as proof of identity. She is undertaking research leading to the production of a thesis or other publications on the subject of "An Investigation into Indonesian Pre-Service Physics Teachers' Scientific Thinking and Conceptual Understanding of Physics." This project is supported by the School of Education, Flinders University.

I would be most grateful if you would volunteer to assist in this project by completing all surveys, and also granting individual interview which cover certain aspect of this topic. The research instrument consists of 5 questionnaires namely scientific reasoning survey, argumentation survey, epistemological beliefs survey, conceptual understanding in Mechanics survey, and conceptual understanding in Electricity and Magnetism survey. It requires 2 meetings to complete all surveys for around 100 minutes for each meeting. The first day of survey distribution, you will be asked to complete 3 questionnaires namely Epistemological Beliefs survey (for around 20 minutes), Argumentation survey (for around 30 minutes) and conceptual understanding in Mechanics survey (for around 30 minutes). The second day of survey distribution, you will be asked to complete 2 questionnaires namely Scientific Reasoning survey (for around 30 minutes) and conceptual understanding in Electricity and Magnetism survey (for around 45 minutes). The participants in the study will consist of all pre-service physics teachers from year 1 to year 4. Meanwhile, only a sub-set will be invited to participate in the interview session. The individual interview will take approximately one and a half hour (45 minutes–break–45 minutes) only for one meeting.

Be assured that any information provided will be treated in the strictest confidence and none of you, as the participants will be individually identifiable in the resulting thesis, report or any other publications. You will be entirely free to discontinue your participation at any time or to decline to answer particular questions.

Since she intends to use a digital voice recorder in the interview session, she will seek your consent, on the attached form, to record the interviews, to use the recording or a transcription in preparing the thesis, report or other publications, and your name or identity is codified. The participants involved in the interview session will be given the opportunity to review and edit their interview transcripts. The principal researcher is the only person who will have access to the interview recordings and will be the only one to transcribe them. The principal researcher will keep all the completed copies of consent forms and interview recordings as well.

Any enquiries you may have concerning this project should be directed to me at the address given above or by telephone on +61 8 82013352 or e-mail carol.aldous@flinders.edu.au

Thank you for your attention and assistance.

Yours sincerely

Dr. Carol Aldous
Co-ordinator: Science Programs

This research project has been approved by the Flinders University Social and Behavioural Research Ethics Committee (Project number:7005). For more information regarding ethical approval of the project the Executive Officer of the Committee can be contacted by telephone on 8201 3116, by fax on 8201 2035 or by email human.researchethics@flinders.edu.au

Appendix 9. Information Sheet

Lina Aviyanti
School of Education
Faculty of Education, Humanities and Law
Room # 3.12A, Education Building Level 3
Flinders University, Bedford Park SA 5042
GPO Box 2100
Adelaide SA 5001
Email: aviy0002@flinders.edu.au
http://www.flinders.edu.au
CRICOS Provider No. 00114A

INFORMATION SHEET FOR STUDENT

Title: An Investigation into Indonesian Pre-Service Physics Teachers' Scientific Thinking and Conceptual Understanding of Physics.

Investigators:
Ms. Lina Aviyanti
School of Education
Flinders University

Supervisor(s):
Dr. Carol Aldous
School of Education
Flinders University

Description of the study:
This study is part of the project entitled 'An Investigation into Indonesian Pre-Service Physics Teachers' Scientific Thinking and Conceptual Understanding of Physics'. The project will investigate the aspects of scientific thinking that focus only on scientific reasoning, argumentation, and epistemological beliefs of pre-service physics teachers. The relationships between scientific thinking and conceptual understanding in physics of pre-service teachers also will be examined. This project is supported by Flinders University, School of Education, Adelaide South Australia.

Purpose of the study:
This project aims to investigate the differences of scientific reasoning, argumentation, and epistemological beliefs, as well as conceptual understanding in physics among pre-service teachers in Yogyakarta, Indonesia. Furthermore, this study will examine the relationships between scientific thinking (i.e. scientific reasoning, argumentation, and epistemological beliefs) and conceptual understanding in physics. A model of the possible associations among these learning variables will be proposed and tested by using structural equation modeling (SEM).

What will I be asked to do?

In the first phase of research, you are invited to answer all questions from the surveys. The research instrument consists of 5 questionnaires namely scientific reasoning survey, argumentation survey, epistemological beliefs survey, conceptual understanding in Mechanics survey, and conceptual understanding in Electricity and Magnetism survey. It requires 2 meetings to complete all surveys for around 100 minutes for each meeting. The first day of survey distribution, you will be asked to complete 3 questionnaires namely Epistemological Beliefs survey (for around 20 minutes), Argumentation survey (for around 30 minutes) and conceptual understanding in Mechanics survey (for around 30 minutes). The second day of survey distribution, you will be asked to complete 2 questionnaires namely Scientific Reasoning survey (for around 30 minutes) and conceptual understanding in Electricity and Magnetism survey (for around 45 minutes). The participants in the study will consist of all pre-service physics teachers from year 1 to year 4. Your participation in the study is entirely voluntary. Your identity will be kept confidential and anonymous in the thesis and any other publications.

In the second phase of research, the participants who gain the highest and the lowest average score in the first phase are invited to participate in an individual interview in Bahasa Indonesia. Please tick YES for the question number 2 of demography survey and provide your phone number if you are interested in participating in the interview session. Then, the principal research will contact you by phone to ask again your willingness to participate in the interview session. You will have the opportunity to choose a convenient time and place for the interview.

The interviewer (principal researcher) will ask some questions which will be taken from the surveys in the first phase of research. The duration of the individual interview will be about one and a half hour (45 minutes–break–45 minutes) only for participants who are selected. The interviewer will record the interviews, transcribe them anonymously and stored as a computer file, recordings will be destroyed once the results have been finalised. The participants involved in the interview session will be given the opportunity to review and edit their interview transcripts.

What benefit will I gain from being involved in this study?

You might not directly gain the benefit to being involved in this study, but your contribution may help the educators to improve the planning and development of future physics teaching and learning programs.

Will I be identifiable by being involved in this study?

Anonymity cannot be guaranteed, but the information provided will be kept with maximum confidentiality by using pseudonym or codified in the data analysis or report. Particularly in the qualitative study, once the interview transcribed, the typed-up file stored on a password-protected computer. The principal researcher is the only person who will have an access and the voice file will then be destroyed once the results have been finalised, so that the data will be kept confidential.

Are there any risks or discomforts if I am involved?

There may be risk or discomforts if you are involved in this study such as you might be burdened with respect to your time and travel expenses, you may not feel confident when answering the test questions of scientific reasoning, argumentation, epistemological beliefs and conceptual understanding in physics. The investigator anticipates few risks from your involvement in this study. If you have any concerns regarding anticipated or actual risks or discomforts, please raise them with the investigator.

How do I agree to participate?

Participation is entirely voluntary and you are free to withdraw from this research project at any time without penalties or consequences at all if you are interested in participating in the project. If you agree to participate, please answer all surveys in the particular time that we will decide later. And if you are selected and interested in participating in the interview session, please read and sign the consent forms and send it back to me later.

How will I receive feedback?

A project summary is just provided to participants on project completion and given to you by the investigator if you would like to see it.

Thank you for taking the time to read this information sheet and we hope that you will accept our invitation to be involved.

This research project has been approved by the Flinders University Social and Behavioural Research Ethics Committee (Project number: 7005). For more information regarding ethical approval of the project the Executive Officer of the Committee can be contacted by telephone on 8201 3116, by fax on 8201 2035 or by email human.researchethics@flinders.edu.au

Appendix 10. Consent Form

CONSENT FORM FOR PARTICIPATION IN RESEARCH
(by individual interview)

An Investigation into Indonesian Pre-Service Physics Teachers' Scientific Thinking and Conceptual Understanding of Physics

I ……...
being over the age of 18 years hereby consent to participate as requested in the Letter of Introduction and Information Sheet for the research project on "An Investigation into Indonesian Pre-Service Physics Teachers' Scientific Thinking and Conceptual Understanding of Physics."

1. I have read the information provided.
2. Details of procedures and any risks have been explained to my satisfaction.
3. I am aware that I should retain a copy of the Information Sheet and Consent Form for future reference.
4. I understand that:
   - I may not directly benefit from taking part in this research.
   - I am free to withdraw from the project at any time and am free to decline to answer particular questions.
   - While the information gained in this study will be published as explained, I will not be identified, and individual information will remain confidential.
   - Whether I participate or not, or withdraw after participating, will have no effect on my progress in my course of study, or results gained.

Participant's signature……………………………………Date…………………...
I certify that I have explained the study to the volunteer and consider that she/he understands what is involved and freely consents to participation.
Researcher's name…………………………….…………………….................
Researcher's signature…………………………………..Date……………………
I, the participant whose signature appears below, have read a transcript of my participation and agree to its use by the researcher as explained.

Participant's signature……………………………………Date…………………...

This research project has been approved by the Flinders University Social and Behavioural Research Ethics Committee (Project number: 7005). For more information regarding ethical approval of the project the Executive Officer of the Committee can be contacted by telephone on 8201 3116, by fax on 8201 2035 or by email human.researchethics@flinders.edu.au