Predicting Performance in the Sri Lankan
General Certificate of Education Advanced Level Examination
By Cognitive Abilities of the Test Takers

Student

W.V.D.S. Madhawa Warakagoda

Supervisor

Dr. Grace Skrzypiec

January 2019

A Dissertation

Presented to

The College of Education, Psychology and Social Work

Flinders University

Adelaide

Submitted in Partial Fulfilment of the Requirements
for the Degree of Master of Education (Educational Research, Evaluation and Assessment)

Table of Contents

List of Figures

List of Tables

Abstract

The relationship between cognitive abilities and academic achievement is extensively discussed in the literature. The expectation of the level of cognitive abilities of students being reflected in the current assessment system in Sri Lanka is a significant concern, especially with the assessments carried out to determine further opportunities for candidates, such as university entrance. The purpose of this study was to determine the predictability of the achievement of test takers of the Sri Lankan General Certificate of Education Advanced Level (GCE AL) Examination based on test takers' cognitive abilities. The study was carried out as a non-experimental correlational study with a cross-sectional design. A random sample of 2,623 candidates who took three core subjects along with the Common General Test (CGT) in the 2017 examination formed the sample of participants in this study. Their achievements at the examination and their general cognitive abilities were represented with average standardised scores for the core subjects and the scores for the CGT respectively.

The usability of the CGT in measuring its constructs was assessed in terms of construct validity and reliability. A four-factor model proposed to explain the factor structure of the CGT fitted well with the data and possessed a good level of construct reliability. A significant relationship between reasoning (RS) and problem solving (PS) was observed. The four-factor model with RS and PS as two separate constructs was not reported with satisfactory fit indices, while a three-factor model, which contained a latent factor of RS and PS together, was a satisfactory fit with the data. Despite this finding, the proportion of the total variance of achievement in GCE AL explained by general cognitive abilities was rather small. Thus, it was concluded that the high correlation between the constructs RS and PS was due to the ineffectiveness of the test items in highlighting constructs separately. Moreover, the use of cognitive ability to solely predict academic achievement was affirmed as insufficient.

Declaration

I certify that this thesis does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university. 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.

Name	:	W.V.D.S. Madhawa Warakagoda
Signature	:	
Date	:	28.01.2019

Acknowledgements

Firstly, I extend gratitude to my late parents who brought me up and always encouraged me to make my prospects with academic studies.

This study would have not been successful without the assistance of my supervisor, Dr. Grace Skrzypiec, who graciously educated me and guided me to enhance my knowledge.

I am thankful for the support of the Australian Government and the Australian Department of Foreign Affairs and Trade (DFAT) that offered me a scholarship for my further studies in Australia. Thus, I extend my appreciation to the officials in Australia and Sri Lanka who work with Australia Awards Scholarships.

My studies in Flinders University would not have been successful without the support I had from the International Student Service of the Flinders University. Therefore, I am very much grateful to Mr. Jose Paulino and his team for their great assistance.

The support I had from the Department of Examinations - Sri Lanka is remarkable. I specially extend my gratitude to Mr. Sanath Pujitha (Commissioner General of Examinations) and Ms. Gayathri Abeygunasekera (Commissioner of Examinations -Research & Development) for their great support.

Also, I am grateful for Mr. David Langdon's support in helping me to improve the quality of my dissertation.

Finally, I extend my sincere gratitude to my friends who supported my studies while I was in Australia.

Glossary and Abbreviations

CC	Comprehension and Communication
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
CGT	Common General Test
DOE SL	Department of Examinations – Sri Lanka
DWLS	Diagonally Weighted Least Squares
EFA	Exploratory Factor Analysis
GA	General Awareness
Gc	Crystallised Intelligence
GCE AL	General Certificate of Education Advanced Level
Gf	Fluid Intelligence
GK	General Knowledge
ML	Maximum Likelihood
NEC	National Education Commission
PS	Problem Solving
RMSEA	Root Mean Square Error of Approximation
RS	Reasoning
SEM	Structural Equation Modelling
SRMR	Standardized Root Mean Square Residual
TLI	Tucker-Lewis Index
UGC	University Grants Commission

Chapter 1: Introduction

1.1 Overview

The association between the cognitive abilities and the scholastic achievement of students has been well established and extensively discussed in the literature (Gustafsson & Balke, 1993; 2018; Rosén, Yang Hansen, & Wolff, 2017). Thus, the differences in cognitive abilities concerned in diverse dimensions of intelligence are widely used in explaining educational achievement in terms of social and individual variability (Kaufman, Reynolds, Liu, Kaufman, & McGrew, 2012; Mackintosh, 2011; Rosén et al., 2017). However, the effectiveness in considering cognitive abilities solely for predicting academic achievement is debated since some research has underpinned the view that the association of motivation with cognitive abilities is a critical factor in predicting academic achievement (Meece, Anderman, & Anderman, 2006; Pedaste et al., 2015; Yousefy, Ghassemi, & Firouznia, 2012).

However, it can be argued that the involvement of causal factors such as motivation in a study of cognitive abilities is conditional on the context. Sackett (2012) states that motivation is not a vital fact in determining differences of measurements in cognitive abilities when test administration is done under high-stake conditions. Sackett argues that due to the high-stake conditions test takers would respond with their maximum effort.

This study uses secondary data from the candidates of the Sri Lankan General Certificate of Education – Advanced Level (GCE AL) Examination to investigate the association between the cognitive abilities and test-takers' achievements at the examination. GCE AL Examination is the terminal assessment at the senior secondary level of education in Sri Lanka and it possesses high-stake conditions since the results are used in selections for the university entrances (University Grants Commission, 2017).

1.2 Background of the Study

The assessment of the predictability of the achievement in GCE AL Examination using the test scores of CGT is the primary focus of the current study. Thus, the background of the study is coherently bounded with the circumstances of the GCE AL Examination. Regarding the origination of the Sri Lankan GCE AL Examination which is conducted by the Department of Examinations – Sri Lanka (DOE SL), Abeygunasekera (2011) acknowledges its relationship with GCE Advanced Level Examinations held in Britain. The first administration of the Sri Lankan GCE AL Examination was done in 1951 (Abeygunasekera, 2011) and since then it is held annually as the terminal assessment of the senior secondary education in the country.

A situational analysis of the assessment and evaluation system of the country done by Sedere, Karunaratne, Karunanithy, and Jayasinghe-Mudalige (2016) asserts that GCE AL Examination has become a high-stake examination since the results are used in the selections for government universities. The GCE AL is held on a large scale; in recent years it has consisted of more than 200,000 candidates annually (Department of Examinations - Sri Lanka, 2018). According to the Department of Examinations – Sri Lanka (2015), the structure of the examination consists of

i. Three core subjects that can be chosen out of 60 subjects;

ii. A test module that measures cognitive abilities (Common General Test); and

iii. A test module that measures English knowledge (General English).

Passing all three core subjects and scoring above a cut-off mark for the CGT (a minimum achievement level) is mandatory for candidates to apply for university entrance (University Grants Commission, 2017, p. 9). General English (GE) is not a compulsory subject of the GCE AL Examination nor is its performance accounted for in the requirements for university entrance (Department of Examinations Sri Lanka, 2015, p. 1).

In issuance of the results of the examination, an average of the standardised scores (Z Score) of three core subjects is issued with the results of the examination. The University Grants Commission (UGC), which is the authoritative body for university entrance, prioritises the candidates according to this average of the standardised scores (University Grants Commission, 2017). The *subject stream*, which is a categorisation of candidates according to the combination of three core subjects (Ministry of Education, 2016), is considered by the UGC in selection of the candidates for undergraduate courses; yet, it is observed that universities offer some courses irrespective of the subject stream (i.e. irrespective of the subjects taken at the GCE AL Examination) (University Grants Commission, 2017). Due to the compulsoriness for university entrance, the CGT has become the test module that encompasses the maximum number of enrolments for the GCE AL Examination (Department of Examinations – Sri Lanka, 2015).

1.3 Statement of the Problem

The current educational measurement system in Sri Lankan schools has been criticised for its competitiveness and its emphasis on students’ ability to recall information rather than measuring higher order cognitive abilities. Such negative critiques are manifested at the societal level (BBC Sinhala News, 2018) as well as in academic discussions (Dundar, Béteille, & Riboud; Sedere et al., 2016). The competitiveness and narrow scope of educational measurements in schools have caused inattention of teachers on the true aims and purpose of education as determined by the national goals of education (National Education Commission, 2009). Also, it has made teachers and parents focus on rehearsing and preparing students for the examinations. Therefore, students are encouraged to memorise the subject content in the classroom, engage paid private tutoring, and practise for examinations. This behaviour is commonly manifested in summative examinations, such as GCE AL (Sedere et al., 2016). However, the focus on preparation for the final examination is rather contradictory with developing the students' higher order cognitive skills in critical thinking, comprehension, problem solving, self-management, and control of cognition (Collins, 2014).

A review of the literature reveals that most research done on the performance in the GCE AL Examination is not extensively discussed with the use of empirical evidence. This study, which reviews empirical data, is aimed at determining whether the performance of students in the core subjects of the GCE AL Examination is an accurate and satisfactory reflection of their cognitive abilities. In this regard, the measurements that were taken by a cognitive ability test module of the GCE AL Examination (Common General Test - CGT) are assessed with respect to the overall students' performance in the core subjects, which is represented with an average standardised score.

1.4 Theoretical Framework

The current study investigates whether the scores of CGT predict the achievement of the candidates at the GCE AL Examination. The theoretical framework of the current study is based on the following elements:

- Identification of CGT as a measure of cognitive ability;
- Assumption of predictivity of the achievement in the GCE AL Examination using scores of CGT due to the relationship between the academic achievement and cognitive ability; and
- Analysis of factor structure of the corresponding constructs.

The identification of CGT as a cognitive ability test is based on the analysis of the composition of its constructs. The corresponding constructs of CGT can be derived in terms of its objectives. The CGT is purported to measure the abilities in four domains: General Awareness (GA), Reasoning (RS), Problem Solving (PS), and Comprehension and

Communication (CC) (Department of Examinations – Sri Lanka, 2000). In the current study, these abilities are considered the latent constructs that will be measured by CGT. A description about the abilities measured by CGT, given by Perera (1999), supports this idea.

In the current study, the explanation of these constructs is primarily done in relation to Fluid Intelligence (Gf) and Crystallised Intelligence (Gc) (Cattell, 1963, 1987). The description of abilities (constructs) given by Perera (1999, pp. 3-4) supports the claim that GA measures General Knowledge (GK). It is evident that a positive relationship exists between GK and cognitive abilities (Batey, Furnham, & Safiullina, 2010; Chamorro-Premuzic, Furnham, & Ackerman, 2006; Furnham & Chamorro-Premuzic, 2006; Furnham, Swami, Arteche, & Chamorro-Premuzic, 2008). It can be asserted that the abilities RS and PS are more proportional to Gf due to the explanation that is given for Gf in the literature (Mackintosh, 2011). According to a number of authors, Gf is concerned with solving novel, complex problems using inductive and deductive reasoning, perceiving associations, identifying patterns in problems, and extrapolation using logic (Kyllonen & Kell, 2017; Mackintosh, 2011; Rosén et al., 2017). The construct, Comprehension and Communication (CC), can be considered a factor which is very proportional to Gc, since Gc is identified as the ability which is related to verbal and reading comprehension, lexical knowledge, aptitude and proficiency in foreign languages, listening and communication, spelling, grammar, and phonetic coding (Carroll, 1993; Mackintosh, 2011).

An assumption of predictivity of the achievement of the GCE AL Examination using scores of CGT can be made due to the existence of the relationship between academic achievement and cognitive ability. The relationship between academic achievement and cognitive ability is well established in the literature (Rosén et al., 2017). It can be further explained in relation to the Gf-Gc correlation that is emphasised in the *investment theory* (Cattell, 1987). Gf is invested in learning where the rate of learning in several tasks is subjected to Gf and where it is accompanied by motivation and opportunities to learn (Cattell, 1987). Since school achievement indicates the rate of learning and is proportional to Gc, it can be argued that scholastic achievement is proportional to Gf (Kyllonen & Kell, 2017; Rosén et al., 2017).

Analysis of the factor structure in the model proposed to assess whether the cognitive abilities of students, as measured by the CGT, are reflected in their scholastic achievement (as indicated by the GCE AL test scores) can be done by identifying the most significant associations among the variables within the model and identifying the most appropriate nested model that provides the best fit with data. Ullman (2006) acknowledges that Structural Equation Modelling can be utilised effectively in such circumstances.

1.5 Significance of the Research

With the findings of this research, the Department of Examinations - Sri Lanka (DOE SL), the institution that conducts the GCE AL Examination, will benefit from several inputs for their research and development tasks for the CGT test module. Since the factorial analyses undertaken within the multidimensional scale of CGT, and between CGT and other scales (subjects of the GCE AL), the current research will be important for them in two main aspects:

i. quality improvement and modifications for CGT,
ii. assessment of usability of CGT for inter-subject test score comparability in GCE AL Examination.

Pragmatic evidence for the quality and robustness of the multidimensional scale of CGT is important in terms of its quality improvement. There are suggestions made by the National Education Commission of Sri Lanka to restructure the CGT paper (Wijetunge & Rupasinghe, 2014). Such a task requires rigorous empirical evidence for the effectiveness of

CGT in measuring the cognitive abilities in terms of redesigning the test paper. Martin (1988) and Hubley and Zumbo (2013) acknowledge four main approaches that are considered in the development of scales and measures: rational–theoretical, factor analytic, empirical criterion keyed, and projective. The factor analytic approach that is taken to review the current multidimensional scale of the CGT in terms of validity and reliability would be beneficial to assert or redesign the factor structure and corresponding indicators (test items).

A search of the published literature and enquiries done by the author of this study supports the idea that consistent in-depth analyses of the performance for CGT have not been done by the DOE. Also, it was noted that itemised data were not gathered for CGT until 2017, though it is done usually for the subjects that have a higher number of enrolments for the purpose of item analysis (Department of Examinations - Sri Lanka, 2017b). One reason for this could be that less attention is given by the DOE SL to this test module compared to the core subjects of the examination. Therefore, the findings of this study would be beneficial for the DOE SL and would make a contribution in reviewing the existing test (CGT) in terms of a factorial analysis.

The second most important aspect of this research is the assessment of the relationship of the achievement of the candidates at the GCE AL Examination using the performance for the CGT. The findings of this study allow forming a basis for further discussions of using test scores of CGT for inter-subject test score comparability. Currently DOE does not provide a scaled score for the GCE AL Examination. Based on personal discussions, the researcher of this study noted that Inter-subject test score comparability is a major concern of the DOE in this regard.

Use of CGT in inter-subject score comparability can be further discussed with respect to the association between its test scores and scores of other subjects. There are several approaches for comparability and linking test scores between different subjects (He, Stockford, & Meadows, 2018; Newton, Baird, Goldstein, Patrick, & Tymms, 2007; Ofqual, 2015). Using CGT as a common test for test score comparability would be useful for DOE SL for making decisions regarding test score comparisons of different subjects. Murphy (2007) acknowledges that a common test can be used for comparisons of different examinations (tests) where all the test takers have taken the common test. In such a case, the common test results are used as the basis for the comparisons of the standards of different tests. It is done by plotting regression lines so that the relationship between common test scores and other test scores can be estimated. In this study, it was found that CGT has the highest number of enrolments in GCE AL Examination (240,693 candidates - those who had taken all three core subjects and had taken CGT too). Since CGT has the highest number of enrolments in the GCE AL Examination, its use as a common test is more logical. However, it is noteworthy that effectiveness of the common test method significantly depends upon a strong and consistent educational and statistical relationship between the common test and the tests being compared (Murphy, 2007, p. 301).

The findings of this study would be beneficial for the DOE in terms of determining the use of CGT as a common test for test score comparability. Depending upon the findings of the current study that describes the correlation between performance for the CGT and performance for the core subjects, a proper decision could be made. It could lead to the discussion and further studies that determine the acceptance or rejection of the recommendations for the use of CGT as common test or for using a combination of selected sub scales (measures of constructs) of the CGT as a common test.

In the context of Sri Lankan education, it is observed that there is a paucity of research and published literature on empirical studies of scholastic achievement. Consequently, studies done on causal factors of academic achievement using empirical data and information are still needed in the field of this research. Therefore, it is expected that the current research will fill a gap in this important area of study and open pathways for further academic discussion.

1.6 Research Aims and Research Questions

This research aims to evaluate the structural relationship between the abilities measured by the CGT and the achievement in core subjects of the GCE AL Examination. Identification of latent factors (measured by CGT) that are most influential for the achievement of the candidates is significant in this regard. Moreover, an understanding of the correlations between the factors is also expected as a coherent objective.

The robustness of the measure of CGT significantly affects the consistency of the prediction done by the test scores of CGT. Rosén et al. (2017) state that validity and reliability of measures have always been a corresponding theme of any research question in educational research. The authors further emphasize that creating valid and reliable measures of constructs that are used for comparisons between individuals and groups is a significant task for educational research and is a challenge in itself. Therefore, as a part of the central objective of this study, the CGT module will be examined for its quality and robustness in terms of validity and reliability of the measurement of the constructs.

Thus, the following research questions are examined in order to address the areas of concern:

1. Is CGT valid in terms of measuring its constructs?
2. Is CGT reliable in terms of measuring its constructs?
3. Do the scores of CGT predict the achievement of the candidates for core subjects in the GCE AL Examination?

In this study, these questions were addressed in a framework of structural equation modelling, and corresponding models to address the questions were based on a priori that was identified in relation to the literature. This is discussed in the following chapters.

Chapter 2: Literature Review

2.1 Introduction

There is an abundance of research and literature in the academic discourse on the subject of formal assessment of student academic achievement in secondary education (Adesoji & Olatunbosun, 2008; Candrasekaran, 2013; Khan, 2005; Zhang, 2007). One of the particular interests is the increasing focus on effective methods of teaching and assessment that improve cognitive processes and enable students to reason, to think abstractly and creatively, and to solve problems (Collins, 2014). Moreover, the association between the cognitive abilities and scholastic achievement of students is frequently discussed by authors (Rosén et al., 2017); yet there remains considerable debate about the most effective methods of enhancing educational outcomes and student performance (Kyllonen & Kell, 2017).

Thus, this chapter discusses the body of literature that focuses on the association of cognitive abilities and academic achievement. In addition, the context of education in Sri Lanka is of particular interest in this study. However, in the context of the Sri Lankan education system, there is a paucity of research and published literature on empirical studies done on scholastic achievement in relation to its causal factors, such as cognitive abilities. Therefore, the relevant research and studies are reviewed in relation to the association between the cognitive abilities and scholastic achievement along with review of the critical factors pertaining to the robustness of the tools that are used in this study to measure cognitive ability.

2.2 Common General Test (CGT)

One of the crucial issues in general education in Sri Lanka is the trend toward more examination centric education which is adversely affecting the academic achievement of students who sit for the GCE AL Examination (Sedere et al., 2016). Consequently, the introduction of the Common General Test (CGT) as a compulsory test module in GCE AL Examination was aimed at enhancing GCE AL students' self-development and broader perspective on global issues (Wijetunge & Rupasinghe, 2014). As a specific recommendation of a Presidential Task Force that was formed due to the educational reforms of 1997 (National Education Commission, 1998), this introduction was made to the existing GCE AL Examination as a means of encouraging students to gain more knowledge about the environment and current affairs, as well as to enhance their higher order cognitive abilities (National Education Commission, 1998; Wijetunge & Rupasinghe, 2014).

2.2.1 Type of the Test

The characterisation of CGT in terms of the type of the test has been controversial to some extent. For example, Perera (1999) emphasises that CGT is not a selection test, yet it can be considered that CGT is administered as an achievement test. In support of this view, Perera points out that, unlike aptitude tests that measure a specific set of skills, CGT is not designed specifically for a particular course (subject in GCE AL) and is common to every course of the candidates who expect to enter government universities. The author also points to the inclusion of the construct General Awareness (GA) to the CGT test, which is not found in aptitude tests (Perera, 1999, p. 6).

However, it can be argued that Perera's (1999) view is somewhat contradictory to the explanation given by Reynolds and Livingston (2012) for achievement and aptitude tests. In relation to the *Standards for Educational and Psychological Testing* (American Educational

Research Association, American Psychological Association, National Council on Measurement in Education, 2014), Reynolds and Livingston assert that achievement tests are designed to evaluate the knowledge or skills of individuals in a content domain in which the individuals received instruction; whereas, aptitude tests are concerned with a broader scope and are designed to assess the cognitive skills, abilities, and knowledge that individuals have acquired as the result of overall life experiences. This is further supported since achievement tests are associated with a particular program of instructions, whereas aptitude tests signify the cumulative outcome of life experiences (Reynolds & Livingston, 2012).

2.2.2 Constructs and Structure of the CGT

Despite the lack of agreement over the terms used to identify the type of the test, it is clear that the purpose of CGT is to test cognitive abilities. CGT measures the abilities in four domains: General Awareness, Reasoning, Problem Solving, and Comprehension and Communication (Department of Examinations - Sri Lanka, 2000). Thus, these abilities can be considered the latent constructs that are measured by CGT. Perera (1999) further describes the subcategories of the abilities and they are presented in Table 2.1. These constructs and abilities are further discussed in the next section on CGT as a cognitive ability test.

Table 2.1

The abilities measured by the CGT

Construct	**Abilities measured**
General Awareness	The awareness in .. • economic (National and International), • political, • social, • cultural, • environmental, • legal, • scientific, and • technological sectors
Reasoning	• Analytical Reasoning, • Methodical Reasoning, • Numerical Reasoning and, • Spatial Reasoning
Problem Solving	• Identifying a fact or facts required to solve a problem, • Restructuring an instance (case) by combining a collection of facts in a particular order, • Identifying analogous facts and situations that are useful to solve a problem
Comprehension and Communication	• Identifying the core idea of a paragraph, • Understanding the underlying concept of a statement, • Identifying the relationships among the ideas contained in a statement, • Identifying the consequences of an action or a statement, • Identifying the difference between facts and opinions, • Following instructions, • Encoding and decoding messages, • Understanding the data presented in charts, tables, graphs and images, • Expression of an idea accurately, • Organising thoughts and ideas in a meaningful way

Note. Adapted from *Samanya Podu Pareekshanaya Sandaha Igenum Athwelak* (pp. 3-4), by L. Perera, 1999

The CGT test consists of 60 objective-type multiple choice questions that should be answered within 2.5 hours, where each test item offers five possible answers from which the student can choose. Only one answer is correct and marks are recorded according to the student's performance. The number of questions is distributed equally among the constructs and the weight of each item for the total score is considered equally, as shown in Table 2.2.

Table 2.2

Distribution of the number items and marks allotted for the CGT by construct

Construct	**No. of test items**	**Marks allotted**
General Awareness	15	25
Reasoning	15	25
Problem Solving	15	25
Comprehension and Communication	15	25
Total	60	100

Note. Adapted from *Samanya Podu Pareekshanaya Sandaha Igenum Athwelak (A Learning Guide for the Common General Test)* (p. 6), by L. Perera, 1999

Objectivity is a major characteristic of CGT. Reynolds and Livingston (2012) acknowledge the greater reliability in using selected-response items in objective tests since the items can be scored using a fixed key where subjectivity in rating is minimised. However, such objective tests may reduce the effectiveness of the measurement in some cognitive areas, such as in assessing the abilities of students in communication and inductive reasoning.

2.2.3 CGT as a Cognitive Ability Test

The general cognitive ability, which is often referred to as 'general intelligence' is broadly discussed along with the performance of individuals for the academic aptitude tests. Plomin (1999) acknowledges that general cognitive ability predicts social outcomes in the areas of education. Coyle, Snyder, Pillow, and Kochunov (2011) acknowledge the effectiveness of linear relations among general intelligence (g), achievements for aptitude tests such as Scholastic Assessment Test (SAT), American College Testing (ACT), Preliminary SAT (PSAT) and college Grade Point Averages (GPAs). In the present research, CGT is considered in terms of a test module that measures the cognitive abilities of the candidates that have taken core subjects of the CGE AL examination. The identification of CGT as a cognitive ability test with respect to the literature is significant in this regard.

Regarding the CGT test, arguments that Perera (1999) has put forward can be understood from different aspects. It can be argued that the author’s first point (i.e. unlike aptitude tests that measure specific set of skills, CGT is not designed specifically for a particular subject/topic) implies that CGT is aimed at measuring the “general ability” of the candidates who are endeavouring to enter university study, irrespective of the subjects they take for the examination. Because the constructs measured by CGT, as shown in Table 2.1, are associated with cognitive abilities, it can be argued that CGT is aimed at measuring the general cognitive ability of the candidates. Regarding the author’s second argument (i.e. the inclusion of the construct GA to the test, that is not generally found in aptitude tests), it can be acknowledged that the measurement of GA is more likely to be related to the measurement of General Knowledge (GK). That is evident by examining the abilities measured in GA, as shown in Table 2.1, and the type of the corresponding test items, which are shown in Appendix D. Since the positive relationship between GK and cognitive abilities is reasonably discussed in the literature (Batey et al., 2010; Chamorro-Premuzic et al., 2006; Furnham &

Chamorro-Premuzic, 2006; Furnham et al., 2008) it can be argued that inclusion of GA to the CGT test instrument supports effective measurement of students' general cognitive ability.

The compositions of the constructs of CGT can be further discussed with respect to the two aspects of general intelligence, fluid intelligence (Gf) and crystallized intelligence (Gc), which were originally identified by Cattell (1963, 1987). Hershey, Austin, and Gutierrez (2015) acknowledge that fluid intelligence abilities include a set of fundamental cognitive processing abilities that are required to assimilate and integrate vital information about a problem or decision, whereas crystallized abilities are concerned with a comprehension of culturally-based values and knowledge about the world.

Reasoning and problem solving are abilities that are considered with Gf (Mackintosh, 2011). The abilities pertaining to Gf are concerned with solving novel, complex problems using inductive and deductive reasoning. Moreover, the abilities to reason and perceive associations, identifying patterns that underpin problems and the extrapolation using logic are widely concerned with Gf (Kyllonen & Kell, 2017; Rosén et al., 2017). Thus, it is clear that the indicators (test items) that come under the constructs Reasoning (RS) and Problem Solving (PS) in CGT are concerned with the domain of Gf.

The construct, Comprehension and Communication (CC), can be appraised with respect to Gc. Mackintosh (2011) emphasises that Gc characterises individual differences in language, information, and notions of a culture. It is the ability concerned with answering questions or solving problems in familiar domains where the knowledge and strategies acquired through experience, schooling, training, or acculturation are used (Kyllonen & Kell, 2017). Regarding the relevance of language related abilities to Gc, Carroll (1993) identifies a range of factors that pertain to Gc, such as verbal and reading comprehension, lexical knowledge, aptitude and proficiency in foreign languages, abilities in listening and communication abilities, spelling, grammar, and phonetic coding. Therefore, it is clear that indicators (test items) of CGT that pertain to the construct CC are associated with Gc.

Though the presence of the construct General Awareness (GA) in the multidimensional scale of CGT has made the characterisation of CGT (as an aptitude test) more controversial (Perera, 1999), it can be argued that such an inclusion has provided another dimension of measurement of cognitive abilities. It is clear that GA is related to General Knowledge (GK). The association between GK and cognitive abilities has been extensively discussed in the literature (Chamorro-Premuzic et al., 2006; Furnham & Chamorro-Premuzic, 2006; Furnham et al., 2008). For example, a study by Chamorro-Premuzic, Furnham, and Ackerman asserts that GK is more positively correlated with cognitive ability (r = .46) than with abstract reasoning (r = .37) (Chamorro-Premuzic et al., 2006). Moreover, Furnham, Weis, and Gee (n.d.) acknowledge that the tests of vocabulary and GK are often included in measurements of Gc. Similarly, Hershey et al. (2015) accede that crystallized abilities take part of an understanding of culturally-based values and knowledge about the world. It is noteworthy that the Weschler Adult Intelligence Test includes an information sub-scale, which is associated with a test of GK (Wechsler, 1981).

Thus, it can be argued that this information supports the idea that CGT is concerned with the measurement of cognitive abilities. Furthermore, in the context of examination administration in Sri Lanka, it can be stated that CGT is a measurement of the cognitive abilities of the GCE AL test takers.

2.3 Cognitive Abilities and Academic Achievement

The focus of the current study is to analyse the predictability of the abilities measured by the CGT for the performance of the candidates of the GCE AL Examination. This is associated with the relationship between cognitive abilities and academic achievement. Therefore, the ideas that are discussed in the literature are important in terms of making reliable assumptions in examining the research questions. More specifically, the underlying hypothesis pertaining to research Question 3, "Do the scores of CGT predict the achievement of the candidates for core subjects in the GCE AL Examination?" is underpinned by the idea that variances in cognitive abilities apprehended by several dimensions of intelligence can be used to describe changeability in educational achievement and to predict academic performance (Rosén et al., 2017).

Using the measurements of general cognitive ability that is represented by the test scores of CGT can be further discussed in a theoretical point of view in order to predict the scholastic achievement represented by the average of the standardised scores of the core subjects. Regarding one of the assumptions that is considered with psychological tests, Reynolds and Livingston (2012) point out that the importance of test performance is not confined to the scope of the assessment itself, since it reveals the standing of the test takers on the measured construct along with the relationship of this standing to other constructs. This idea is supported by Kyllonen and Kell (2017) who put forward an argument regarding a common factor that accounts for several tasks. In relation to this matter, the authors refer to the general fluid ability as a common construct. Based on the empirical observation, the authors show that there is a tendency in individuals who perform well in a variety of cognitive tests which requires learning, memory, or thought, to be more successful. Thus, it can be argued that the prediction of academic achievement according to cognitive abilities is rather logical.

Elaborating on this idea, Kyllonen and Kell (2017) explain the relationship between scholastic achievement and cognitive abilities through the lens of a Gf-Gc correlation that is upheld by the *investment theory* (Cattell, 1987). In the investment theory, it is proposed that Gf is invested in learning diverse tasks which are subjected to Gc when it is accompanied by motivation and opportunities to learn. Kyllonen and Kell argue that school achievement reflects the rate of learning and is related to Gc. Accordingly, they argue that scholastic achievement is proportional to Gf. Due this fact, it is expected that the two Gf components of the CGT would be more predictive of achievement in GCE AL than the two Gc components.

Rosén et al. (2017) elaborate further on the idea of investment theory in relation to academic achievement. They emphasise the causality of Gf in achievement differences because Gf represents the capacity for solving novel problems, making inferences, recognizing relations, transforming information. They further acknowledge that, through investment of Gf in learning experiences, these capacities are transformed into crystallized intelligence that is more concerned with depth and breadth of knowledge gained through learning. In relation to this idea, it is noteworthy that Thorsen, Gustafsson, and Cliffordson (2014) show that applying Gf for Gc is not confined to an initial state, but is rather manifested continuously by schools throughout the learning process.

In addition to the above ideas, Kaufman et al. (2012) have stated that, though general cognitive abilities in reading, math, and writing achievement are not isomorphic, they correlate significantly. In a similar study, Spinath, Spinath, Harlaar, and Plomin (2006) affirmed that a considerable portion of common variance in school achievement is explained by differences in general mental ability and motivation. Moreover, in a study conducted to explain variation in academic achievement with general cognitive ability, specific cognitive abilities and academic achievement, Rohde and Thompson (2007) concluded that processing speed and spatial ability were noteworthy in predicting scores for the mathematical portion of the tests considered while holding general cognitive ability constant.

2.4 Robustness of the Measurement Done With the CGT

In the current study, the quality and worthiness of the measurement done by the CGT is discussed in terms of validity and reliability. Though these characteristics have been discussed by other researchers in various aspects, in the current research, validity, and reliability is more concerned with the constructs of the test.

2.4.1 Validity of the test

In the body of literature on the subject of assessment, the validity of the tests has been defined in many ways. However, it is noted that many of these definitions imply the same underlying idea that test validity refers to the degree that the test tool actually measures what it purports to measure (Sireci & Sukin, 2013). In the *Standards for Educational and Psychological Testing*, validity is referred to the degree to that evidence and theory underpin the interpretations of scores for proposed uses of assessments (American Educational Research Association et al., 2014)

Researchers have discussed and assessed a number of different models for determining validity of tests. For example, Kane (2013) identifies several models of validation: Content-criterion model, construct model, unified construct model, and argument-based approach model. In the current research, validity is rather discussed in terms of the construct model. Regarding the concept of the *construct* , Cronbach and Meehl (1955) described the construct as a hypothesised attribute of people that is assumed to be reflected in test performance. Kane (2013) acknowledges the close intertwinement of validity theory and construct theory. The conceptualisation of construct validity was rather focused on the degree of representation of an individual's standing on a theoretical construct by the test scores of a test that is designed to measure the construct (Sireci & Sukin, 2013). In application, theoretical constructs are tacitly defined by their characterisation in a theory, and the theory along with the construct-based interpretation are evaluated together (Kane, 2013).

Regarding the measurement of the construct validity, several methods of measurements for psychological constructs are discussed in the literature. As a common factor, measures of constructs are validated by tests to see the relationship with measures of other constructs that are specified by the theory. Thus, each test of associations between measures indicates for the validity of both the measures and the theory which are considered in the test. Kane (2013), and Strauss and Smith (2009) state that construct validation is concerned with the simultaneous process of measurement and theory validation.

Confirmatory factor analysis is widely used in the studies that are involved in estimation of construct validity. In the use of CFA, a hypothesised model is applied to estimate a population covariance matrix which is compared with the observed covariance matrix (Schreiber, Nora, Stage, Barlow, & King, 2006). Furthermore, in conditions of hypotheses concerned with the structure of the measure, Goodwin (1999) emphasises the selection of CFA as the appropriate method. In the current study, CFA is used in validation of the constructs of the scale (CGT) and further described later in this chapter.

2.4.2 Reliability of the Test

Because most measurements in education and psychology are done using indirect measurement tools, reliability of those measurements is required to demonstrate that scores for the responses are consistent and reproducible. This phenomenon of reliability implies that scales measure the same latent trait or construct (McNeish, 2017). In *Standards for Educational and Psychological Testing* (American Educational Research Association et al., 2014), the term reliability is discussed in two main aspects: the correlation between scores on two comparable forms of the test given the presumption that administering one test form has no effect on performance in the second form; and consistency of scores over repetitions of a testing procedure. However, McNeish (2017) notes that reliability can be understood as the correlation between scores on two successive administrations of tests, assuming that the test takers do not recall their answers from the first administration (McNeish, 2017). These two interpretations share common understandings of the importance of correlation and consistency in testing to achieve reliability.

There are four general classes of reliability: *Test-Retest Reliability*, that is concerned with the consistency of measures from one time to another; *Alternate Form Reliability*, that assesses the consistency of the results of two test forms crafted similarly from the same domain of content; *Inter-Rater Reliability* that assesses the degree that different observers (examiners) provide consistent ratings for the same phenomenon; and *Internal Consistency* that is used to assess the consistency of test scores across items within a test (Geisinger, 2013; Reynolds & Livingston, 2012).

Internal consistency refers to the estimate of the reliability of a measure done by evaluating the within-scale consistency of the responses to the items of the measure, which indicates the degree of interrelatedness among the items (Cortina, 1993). It is noteworthy, that internal consistency coefficients are critiqued as indirect measures of reliability, pointing out that they are theoretical estimates derived from classical test theory (Henson, 2001). Out of the various types of reliability estimates, the most frequently used is the internal consistency estimate because they are readily computed from a single administration of a test (Henson, 2001). In educational and psychological research, Cronbach's α (Cronbach, 1951) is the most extensively used reliability index for estimating the internal consistency, which is reported with almost two-thirds (66%) of studies reporting reliability measures (McNeish, 2017, p. 1).

Due to the extensive usage of Cronbach's α in estimating reliability of the scales, it has been widely discussed in the literature (Cortina, 1993; Sijtsma, 2009; Streiner, 2003; Trizano-Hermosilla & Alvarado, 2016). For a fair use of Cronbach's α in estimation of the reliability of a scale, the satisfaction of the assumptions of uncorrelated error scores of any pair of items (uncorrelated error terms), and the same true score for all test items, or equal factor loadings of all items in a factorial model (tau equivalence) is significant. (Raykov, 1997, 1998; Trizano-Hermosilla & Alvarado, 2016). In case of violation of these assumptions, reliability value will be over or underestimated (Graham, 2006; Raykov, 1997, 1998). However, Teo and Fan (2013) argue that the satisfaction of these assumptions is not viable in practice.

Consequently, the use of Cronbach's α in multidimensional scales is limited. Brunner and SÜβ (2005) argue that this is due to the requirement for satisfaction of its assumptions: uncorrelated error terms and essentially equivalent scale items (indicators). Dunn, Baguley, and Brunsden (2014) show that lack of perfect inter-correlations between true scores of items (lack of unidimensionality) cause erroneous estimation for Cronbach's α. In contrast to unidimensionality, CGT test which is primarily concerned in this study possesses the constructs which have different loadings for the underlying construct (ability in CGT). This is further discussed in the next chapters, Research Method, and Results.

In such a circumstance of deviation from unidimensionality, the values for the reliability that is estimated by the Cronbach's α would be rather questionable. Therefore, this study prefers multidimensional reliability coefficients. Thus, the construct reliability, that assesses the degree to which the different scale indicators reflect an underlying factor (Brunner & SÜβ, 2005), was utilised in estimating the reliability of the CGT multidimensional scale. Different indicators of CGT (for example, the difference of the items pertain more to GA compared to items that reflect PS), will be assessed fairly in terms of the degree which they reflect the general cognitive ability by using the construct reliability.

In terms of the method of estimation (coefficients of the construct reliability) differences of the corresponding coefficients can be further discussed. Thus, based on the weight given for the indicators in calculating the scale score, two coefficients can be primarily distinguished. It was identified that equal weighting given for the scale indicators in estimating the reliability is a major characteristic of the coefficient *Omega* - Ω (McDonald, 1970). In contrast to the equal weighting, *Canonical factor regression method coefficient Omega* (Allen, 1973) was concerned with weight given for each indicator according to the corresponding pattern coefficient. Brunner and SÜβ (2005) observe that Bacon, Sauer, and Young (1995) refer to the same as *weighted Omega* - Ω*w*; and Hancock and Mueller (2001) call it *Coefficient H.* Due to the importance of weight given for the indicators (test items) in circumstances of having different types of test items and constructs, In the present study, this reliability coefficient is used in estimation of the construct reliability of the CGT. It is reported as the *Coefficient H,* which is given by

$$H = 1 \Big/ \left[1 + \left(1 \Big/ \sum_{i=1}^{k} l_i^2 / (1 - l_i^2)\right)\right]$$

Where *k* is the number of measured variables associated with a given latent factor, l_i is the ith indicator's standardised loading.

2.5 Methods Used in Assessment of Structural Relationships

This section will further explain and clarify the adequacy and the applicability of the approaches taken to evaluate the structural relationships and the assessment of the measures in the present study, taking into account the information gleaned from the literature on the subject. In this section, these approaches are reviewed in terms of their fundamental characteristics that are highlighted in literature.

In order for this study to assess whether the cognitive abilities of the GCE AL students reflect their scholastic achievement, use of Structural Equation Modelling (SEM) is identified as an effective method of analysis. SEM is a powerful framework for statistical analysis that combines path models with the analysis of latent factors (Hox & Bechger, 1998; Mueller, 1997; Ullman, 2006). The use of SEM with respect to the questions which are supposed to be answered can be further reviewed. Ullman (2006) identifies three main aspects of research questions that the SEM can be applied to:

i. Do the parameters of the model combine to estimate a population covariance matrix that is highly similar to the sample covariance matrix?;
ii. What are the significant associations among variables within the model?;
iii. Which nested model provides the best fit to the data? (p. 38). 25

However, in approaching those questions, it should be realised that use of SEM is not identified as a specific technique. Rather, it is a set of procedures that are concerned in several stages (Hoyle, 1995; Mueller, 1997; Mueller & Hancock, 2008). In most cases, the stages, which are (i) initial model conceptualization and specification, (ii) parameter identification and estimation, (iii) assessment of data-model fit, and (iv) potential model modification, are significantly identified (Mueller & Hancock, 2008).

SEM outlines a set of methods for data analysis which allows testing for theoretically derived and priori specified causal hypotheses (Hoyle, 1995; Mueller & Hancock, 2008). More specifically, regarding the foundations of the SEM, Mueller and Hancock (2008) identify *measured variable path analysis* and *confirmatory factor analysis* as significant methods of analysis. The authors further acknowledge that traditional data analytical techniques, such as the *analysis of variance, analysis of covariance, multiple linear regression, canonical correlation, exploratory and confirmatory factor analysis,* and *measured variable path analysis* can be regarded as special cases which are encompassed in the SEM framework. It is noteworthy that theoretical conceptualization of SEM is associated with the classical path analytic and factor analytic techniques, and therefore they have become significant components in SEM framework.

In the present research, the validity and the reliability of CFA are analysed using a factorial model. This is a type of SEM that is concerned with measurement models and deals with observed and latent variables (Brown, 2014; Brown & Moore, 2012). More specifically, it assesses the relationships between observed measures or and latent variables or factors based on a priori (Brown, 2014). Factor analytic approach is used more frequently in developing scales and measures (Hubley & Zumbo, 2013; Martin, 1988), while CFA is frequently used to examine the latent structure of the test in this regard (Brown, 2014). Moreover, CFA is carried out to verify the number of underlying dimensions (factors) of the test instrument and the patterns or relationships (factor loadings) between the items (indicators) and the factors (Brown, 2014; Brown & Moore, 2012).

A feature of CFA is that it is hypothesis-driven, which is a significant difference when compared to its counterpart, Exploratory Factor Analysis (EFA). Exploratory Factor Analysis has a data-driven approach, where the number of initial factors or the pattern of associations between the common factors and the indicators are usually not specified (Brown, 2014).

Conversely, pre-specification of the model is required in CFA (Mueller, 1999). In this case, based on past evidence and theory, the researcher should have a firm priori sense of the existence of the number of factors (constructs), and of the relationship between the indicators and the factors (Mueller, 1999). This prespecified solution is evaluated in terms of how well the corresponding factor structure regenerates the sample correlation (or covariance) matrix of the measured variables. Brown and Moore (2012) emphasise the significance of strong empirical or conceptual foundations, both in the specification and evaluation of the factor model.

The current research examines the validity of CGT in terms of its factorial structure. As Brown and Moore (2012) state, CFA can be considered a more crucial analytical tool that can be used for construct validation. Furthermore, Brown (2014) demonstrated that the results of CFA can provide evidence of the *convergent validity* which is indicated in the existence of strong interrelations among different indicators (test items) of theoretically similar or overlapping constructs. In addition, Brown & Moore (2012) have shown that CFA gives evidence of *discriminant validity* which is indicated by results showing that the indicators of theoretically dissimilar constructs are not highly intercorrelated (Brown & Moore, 2012). In CFA, the resulting estimates of convergent and discriminant validity are adjusted for measurement error and an error theory (Brown, 2014). Thus, CFA provides a stronger analytic framework for this study to account for measurement error.

Chapter 3: Research Method

3.1 Introduction

This chapter includes a description of the method, research design, target population and the sample, data collection, instrumentation, data analysis, and ethical considerations of the study. The study primarily focuses on two areas of the CGT, which are the quality of the measurement of the constructs of the CGT and the association of those constructs with the performance of the candidates in the GCE AL Examination. Data for the study were obtained from data archives of the DOE SL. Therefore, a quantitative method is used to assess the factorial relationships within the scale of CGT and among the factors pertaining to the model proposed to explain the predictability of achievement at the GCE AL using scores of CGT. According to Creswell (2012), the objective of quantitative research is to describe a research problem through an explanation of trends or a need for an explanation of the relationship among variables.

3.2 Research Design

The present research was carried out in a correlational design framework (Airasian & Gay, 2003; Creswell, 2012; Johnson, 2001) using archival (secondary) data, which is useful in a study where there is no intervention with the individuals involved in the research. The correlational design is concerned with collecting data to determine whether a relationship exists between two or more quantifiable variables, and to examine to what degree they are related (Airasian & Gay, 2003). In addition, the present research is rather a cross-sectional study (Levin, 2006) since the data were confined to a certain time of a test administration (GCE AL 2017).

3.3 Study Population and Sample

The population of this study consisted of 261,041 candidates who had taken at least one subject for the GCE AL Examination in year 2017. These candidates were between 17 and 18 years of age and were the national group of students that took the examination after completion of senior secondary education. Out of this population, a random sample of 2,678 individuals had been selected by the DOE SL to collect data of the responses for the items of the CGT. The corresponding data set was used in this study and the sample of this study consisted of 2,623 cases of the above-mentioned data set after data screening.

3.4 Data Collection

Since this study uses secondary data for analysis, the relevant data of the 2017 GCE AL Examination were captured from the archives of the Department of Examinations - Sri Lanka (DOE SL) with the approval of the head of the institution. Generally, DOE SL does not publish data of its examinations except for specific research purposes and, therefore, these test scores are not publicly available. Sri Lankan CGE AL Examination is held on a large scale and is undertaken with rigorous and confidential processes. The interventions for instrumentation such as determining the indicators of the constructs of the measure (designing test items for CGT); and determining the data collection design were not practical and were omitted to assure the minimal disturbance for the DOE SL. Therefore, the instrumentation was done at the level of capturing required data from the bulk data which DOE SL collects after test administrations.

Data collection was done from the population of GCE AL 2017 and from a random sample of candidates who took CGT for the GCE AL Examination. The sample consisted of 2,678 cases and they had been selected randomly by the DOE SL to collect data of the item responses. Such itemised data is collected from samples of candidates of selected subjects

(DOE SL, 2017a) for the research and development purposes. DOE SL does not collect itemised data for the whole population in each subject since that would consume more time and resources. From the above-mentioned sample for CGT, 2,623 cases were selected for this study after data screening. These data contained the test scores for each item of the CGT, and they were primarily used for the analysis of this study. Final test scores of the target population were used to determine the relative position of the candidates.

3.4.1 Instrumentation

Data collection formats were designed for the instrumentation process of collecting archival data. These formats contained the data of the test scores of each candidate in the population; and data of the item scores of CGT of a sample of candidates. As shown in Table 3.1, data of (i) serial number of the candidate (instead of the index number), (ii) subject, and (iii) score for the subject were collected from the population using Format 1 (Appendix A). The binary data of the CGT test items (defined as 1 = correct answer, 0 = incorrect answer), were derived from the Format 2 and 3 (Appendix A). For the benefit of explanation, data which were captured in this thesis using Format 1 is named as "Dataset P", while the binary data for test items along with the final score which were created with Format 2 and Format 3 are called "Dataset S". Table 3.1 and Table 3.2 show the schema of these data sets respectively.

Table 3.1

Dataset P - Test scores of the candidates in the target population

Serial No of the Candidate	**Subject**	**Final Mark**

Table 3.2

Dataset S – Itemised binary data and the final test score of the candidates for CGT in the selected sample

Serial No of the Candidate	**Item 1**	**Item 2**	**……**	**Item 60**	**Final Mark**

3.4.2 Sample Data

The sample contained 2,623 cases which were selected from a data set of 2,678 candidates who were selected using simple random sampling technique. These data were included in the Dataset S.

3.4.3 Data Screening

Data screening was critical to assure the usability of data (DeSimone, Harms, & DeSimone, 2015) in the study. In the present study, it was assumed that an average of the standardised scores for three core subjects indicates the scholastic achievement of candidates. Similarly, an assumption was made that the test scores of CGT were an indicator of their cognitive ability. Therefore, the selection of candidates who took all three core subjects along with the CGT was required to assure the representativeness of the data in relation to the assumptions made. This is analogous to removing missing data listwise. Thus, out of the total population that contained 261,041 cases, only 241,070 (candidates who took all three core subjects along with the CGT) were used for analyses, where it can be considered the target population of the study and included in the Dataset P.

The scores of the candidates for the CGT in Dataset P and Dataset S were compared since the data taken for the two data sets had been originally entered in two divisions of the DOE SL. Thus, 29 cases with different test scores for matching serial numbers and subjects were identified and removed listwise from the Dataset S. In comparison of these test scores, it was assumed that data in the Dataset P were correct since its original data had been populated by the data entry branch of the DOE SL which follows rather consistent data entry and verification procedures.

3.4.3.1 Outliers

Some data in the Dataset P that contained test scores for three core subjects were identified as multivariate outliers. It is a known that some candidates, such as gifted students, outperform in some subjects whereas their performance may be average or even weaker in other domains (Johnsen, 2004). However, since the main purpose of the current study is to determine whether cognitive abilities predict the performance of the candidates at the GCE AL Examination, it was decided to avoid the “extreme” cases in order to establish a general idea. These multivariate outliers were determined with the chi square distribution of the corresponding Mahalanobis distances (Ben-Gal, 2005). The cases that possessed the significance values less than 0.001 were omitted. Thus, the corresponding cases in both Datasets P and S were removed listwise. Out of 261,041 candidates who took at least one subject for GCE AL, 241,070 were identified as the candidates who took three core subjects along with the CGT. After the removal of outliers 240,693 cases (candidates) were observed in the dataset P. Accordingly, some cases in the sample (dataset S) were removed since they had no matching records of test scores in the dataset P. Thus, 2,623 cases were finally observed in the dataset S. Figure 3.1 illustrates the data screening and outlier removal procedure in the present study.

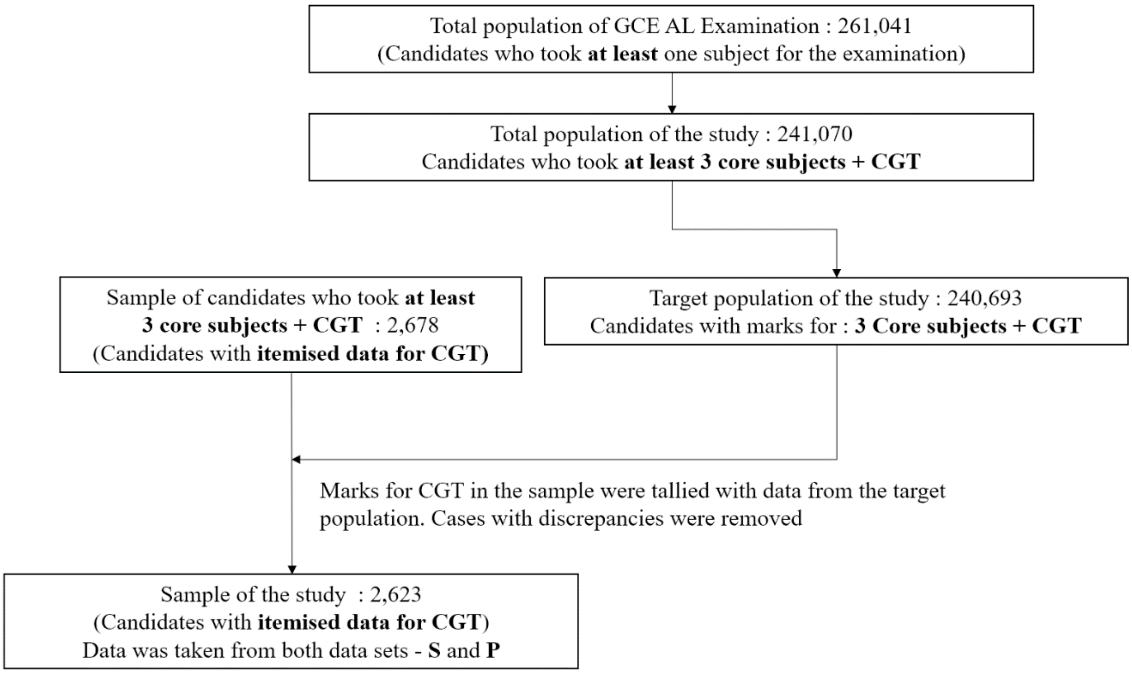


Figure 3.1 Data screening and removal of outliers

3.4.3.2 Missing Data

There were no missing data in the target population since only the candidates who took all three core subjects along with the CGT were concerned. No missing data were in the sample of the study (Dataset S).

3.5 Data Analysis

In the present study, data analysis was carried out in a framework of Structural Equation Modelling (SEM) that combines path models with the analysis of latent factors (Hox & Bechger, 1998; Mueller, 1997; Ullman, 2006). A Confirmatory Factor Analysis (CFA), which is a special case of SEM, was carried out in assessing the construct validity. In this study, “R” (R Core Team, 2018) statistical software was used for data analysis with the software package *lavaan - latent variable analysis* (Rosseel, 2012). The source data files were manipulated with PostgreSQL open source relational database management system (PostgreSQL, 2018).

3.5.1 Research Question 1: Measurement of Construct Validity of the CGT

Techniques of factor analysis can be used to provide evidence that a test is measuring the constructs it is supposed to measure (Sireci & Sukin, 2013). CFA requires an empirical or conceptual basis for both factor model and evaluation (Brown & Moore, 2012; Mueller, 1999). Since a priori for the factor structure of CGT could be established when examining the objectives of introduction of it to the GCE AL Examination (Department of Examinations - Sri Lanka, 2000; Perera, 1999), it was decided to carry out a CFA with an identified model (see Figure 3.2). One of the main advantages of using CFA in studies of construct validity is the possibility of comparison for alternative models that illustrate the relationships among constructs (Strauss & Smith, 2009).

In the present study, the Model 1, shown in Figure 3.2, that includes 15 indicators (items) in each construct, was proposed after the observation of the pattern of the questions (Appendix D) and their order in the test papers of CGT in several previous years. Each indicator contains dichotomous data where correct and incorrect answers are signified with “1” and “0” respectively.

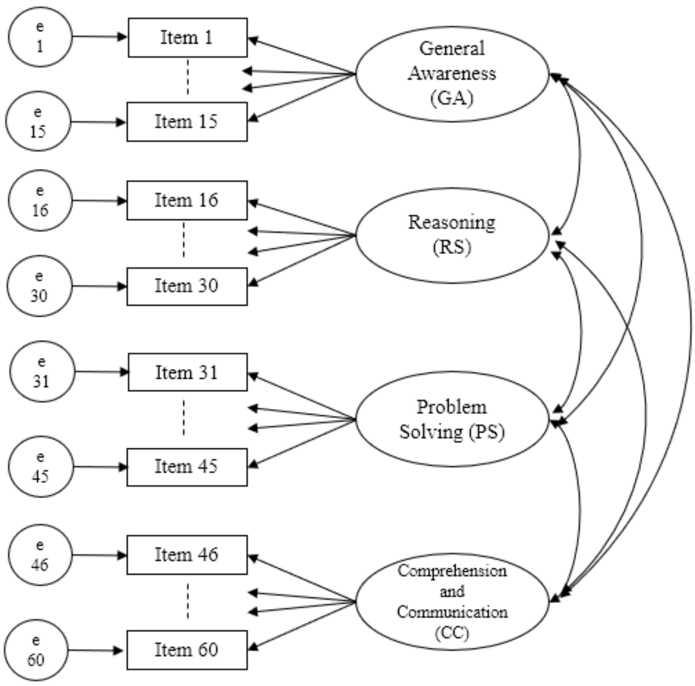


Figure 3.2 Model 1 - The ability measured by the CGT and corresponding four latent factors

In CFA, Maximum Likelihood (ML) is used widely as an estimator with the continuous observed variables (Li, 2016). ML is associated with the assumption of continuousness of the observed variables and multivariate normality of them for factor extraction. However, in contradiction of those assumptions it does not provide reliable estimates, especially for ordinal and categorical observed variables (Li, 2016; Myung, 2003). Therefore, use of Diagonally Weighted Least Squares (DWLS) is recommended for the cases with ordinal data (Flora & Curran, 2004). With the use of binary variables (indicators), DWLS ensures the estimation of a tetrachoric correlation matrix for factor extraction (Debelak & Tran, 2016). The software package, *lavaan*, is capable of switching in to DWLS to estimate the model parameters for the categorical and binary data (Rosseel, 2014).

In the test of CGT in 2017, some test items are grouped together, and some questions are based on common information (shared information). For example, the test items 32,33, and 34 are based on a graph provided in the examination. In the proposed model, it is assumed that such indicators (test items) are correlated with each other in addition to the assumption of correlation among the constructs. The indicators with negative factor loadings and the indicators which have, P(>|z|) ≥ 0.05 (p-values for testing the null hypothesis, such that the parameter equals zero in the population) were omitted from the initial model and the revised model was reanalysed. Thus, the revised model contained 58 indicators, which included 13 indicators for GA, and 15 each for other factors (constructs).

3.5.2 Research Question 2: Measurement of Reliability of the CGT

In the Chapter 2 : Literature Review there was a discussion of the reliability of the measure and the inconsistent results that Cronbach's α (Cronbach, 1951) provides in circumstances where the assumptions unidimensionility and Tau-equivalence are not satisfied. Therefore, use of construct reliability as an alternative approach was further discussed. This study uses Coefficient H (Hancock & Mueller, 2001) to estimate the construct reliability of the measure of CGT. Coefficient H is given by

$$H = {}^{1}\Big/\left[1 + \left({}^{1}\big/_{\sum_{i=1}^{k} l_i^2/(1 - l_i^2)}\right)\right]$$

Where k is the number of measured variables associated with a given latent factor, l_i is the i$^{\text{th}}$ indicator's standardised loading.

Research Question 3: The Predictability of the Performance in the GCE AL Using Test Scores of CGT

In the present study, the model shown in Figure 3.3 was proposed to assess the association between general cognitive ability (measured by the CGT) and achievement in the GCE AL Examination. Each construct of the CGT is indicated with a composite score of 15 items except the GA which is indicated with 13 items. The composite score was determined by summing raw scores corresponding to all items loading on a factor, which is given as a linear combination of rows scores and factor loadings ($Y = W_1X_1 + ... + W_pX_p$, W_i = factor loading for the ith item, X_i = row score of the ith item). Tabachnick and Fidell (2007) and DiStefano, Zhu, and Mindrila (2009) acknowledge that summed factor scores preserve the variation in the original data. The relative position of the candidates in each core subject was determined with T Scores, where T=[(x−μ)σ×10]+50. It was assumed that the relative position of the candidates in the target population is represented with the average of the standardised scores (T Scores) for core three subjects. Furthermore, it was assumed that their relative position in the target population indicates their achievement in the GCE AL Examination. .

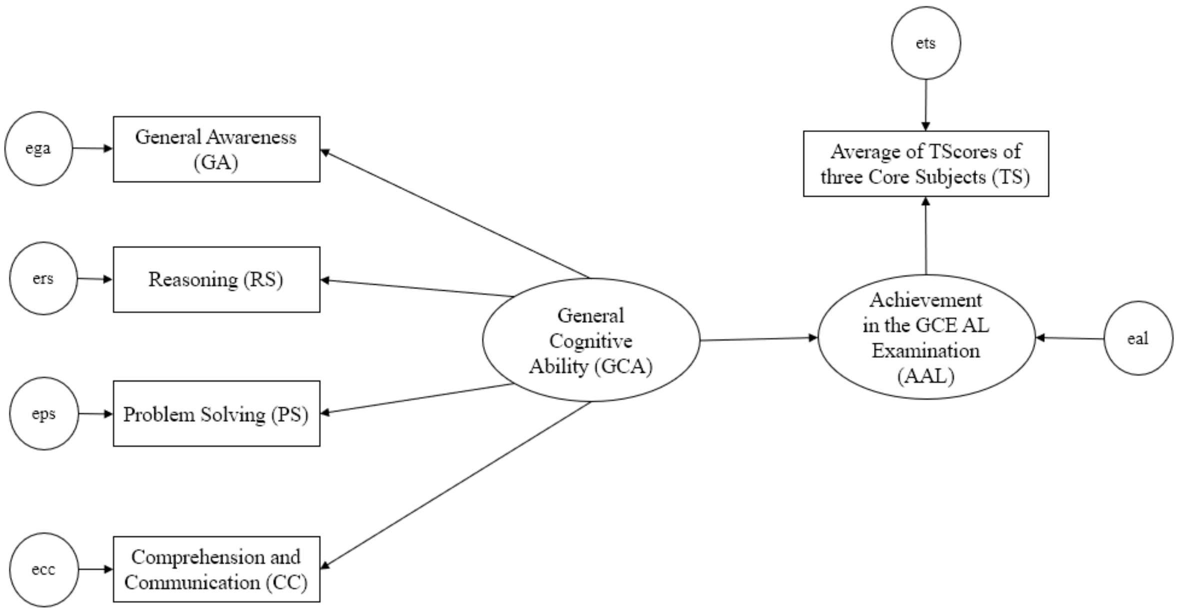


Figure 3.3 . Model 2A – Achievement in GCE AL predicted by four factors of the General Cognitive Ability

In the assessment of the Model 2A, a considerably higher correlation between PS and RS compared to the other constructs was observed. The Variance Inflation Factors (VIF) were computed to detect the multicollinearity (Alin, 2010), yet it was found each variable was reported with the VIF values were considerably small (GA = 1.163, RS = 1.926, PS = 2.065, CC = 1.537, TScore = 1.135). However, given the assumption that the abilities that items of the CGT pertain to RS and PS measure are originated with a common latent factor Model 2B was proposed. The new construct was named as "Reasoning and Problem Solving". Figure 3.4 illustrates the corresponding model.

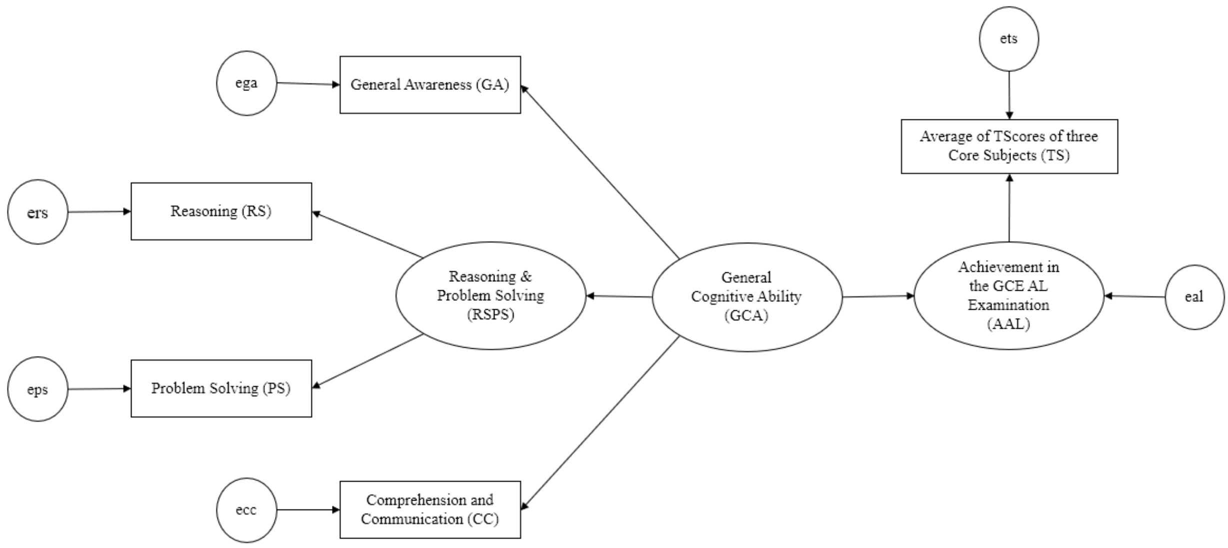


Figure 3.4 Model 2B – Achievement in GCE AL predicted by four factors of the General Cognitive Ability

3.5.3 Goodness of Fit of the Models

In order to assess the goodness of fit of the models proposed in research questions (Model 1, Model 2A and 2B), a Chi-Square test was performed to determine whether the population covariances were consistent with those predicted by the model. However, since the Chi-Square Goodness Fit Statistic is reported statistically significant for the models with higher number of cases (Cheung & Rensvold, 2002; Kenny, 2015), the fit indices, Root Mean Square Error of Approximation (RMSEA)(Steiger, 1980), Bentler's Comparative Fit Index

(CFI) (Bentler, 1990; Bentler & Bonett, 1980) , Tucker Lewis Index (TLI) (Tucker & Lewis, 1973), Standardized Root Mean Square Residual (SRMR) were also considered in order to evaluate the goodness of fit of the model.

3.6 Ethical Considerations

The current research involved the analysis of archival data from the Sri Lankan CGE AL Examination, which ensures confidential processes for all students undertaking the examination. Therefore, the researcher sought and obtained permission from the DOE SL to conduct the present study and to receive and use raw data of the GCE AL Examination 2017, which are relevant to the research. Currently, raw marks of the core subjects are not published by the DOE in the issuance of the results of the GCE AL Examination. Candidates receive results of the examination with gradings for each subject except for the CGT, which is provided with a final mark. Therefore, as a means of assuring the confidentiality and privacy of the candidates, the identities (index numbers) of the examinees considered in this research are not revealed during the research or in the written thesis. All information of individuals gathered relating to the study will be kept confidential.

A further aim of the study was to cause a minimum disturbance or inconvenience for the DOE SL and for other institutions involved in conducting the GCE AL Examination. Thus, the research design was adjusted accordingly by being aligned with a correlational framework rather than conducting an experiment or a survey that would have required involvement of personnel or candidates. DOE SL as contributor to the present study will benefit from the results of this study and the information can be used to enhance its test development, data analyses and policy decisions in future.

Chapter 4: Results

4.1 Introduction

In this chapter, the results of the analyses for the three research questions (Question 1. Is CGT valid in terms of measuring its constructs? Question 2. Is CGT reliable in measuring its constructs? Question 3. Do the scores of CGT predict the achievement of the candidates for core subjects of the GCE AL Examination?) are presented along with relevant findings. After initial application of CFA in answering the research Question 1, it was found that Item_13 has a negative factor loading. It was observed that a low negative value (-0.01) had been reported with the discrimination index of this item which had been calculated using the point biserial correlation. In Item_7, the p-value for testing the null hypothesis, such that the parameter equals zero in the population, P(>|z|), was greater than 0.05. The Model 1 was therefore refined by removing these items and they were not considered in creating composite scores for the indicators in the Models 2A and 2B. Dichotomous data were used for the indicators of the Model 1, whereas the indicators of the Models 2A and 2B were continuous scalar variables.

4.2 Research Question 1: Construct Validity of the CGT – 2017

The model proposed for the factor structure (Model 1) resulted with satisfactory fit indices (see Table 4.1). Although the χ^2 goodness-of-fit statistic was significant, χ^2 (1572) = 2387.781, p<.001, the measurement model was considered to be an acceptable approximation of relationships of the constructs due to the values of the other model fit indices. Value for the Root Mean Square Error of Approximation (RMSEA) was 0.014, 95% CI [0.013, 0.015]. The Standardized Root Mean Squared Residual (SRMR) value was 0.044. These were below the cut-off values of those indices, which were 0.6 and 0.8 respectively (Schreiber et al., 2006). The Comparative Fit Index (CFI) was above the benchmark value of 0.95 which is proposed for acceptable values (L.-t. Hu & Bentler, 1998; L. t. Hu & Bentler, 1999). The Tucker–Lewis Index (TLI) was 0.976 which is above its cut-off value of 0.95 for acceptance of the model (Schreiber et al., 2006). Thus, it was determined that the Model 1 adequately explains the factor structure of the CGT (composition of constructs) and the test scores of the CGT in 2017 test administration. This observation indicates the construct validity.

Table 4.1

Model fit indices of Model 1

Model	χ^2	*df*	RMSEA	CFI	TLI	SRMR
Model 1	2387.781*	1572	0.014	0.978	0.976	0.045

*$P < 0.01$

In addition to the association of the indicators with corresponding constructs, the correlation between the latent constructs was also measured. Table 4.2 shows the respective correlation matrix.

Table 4.2

*Correlation between each construct of CG*T

Measure	1	2	3	4
1. General Awareness (GA)	-			
2. Reasoning (RS)	0.467	-		
3. Problem Solving (PS)	0.485	**0.974**	-	
4. Comprehension and Communication (CC)	0.612	0.766	0.891	-

According to the correlations given in Table 4.2, the lowest is observed between GA and RS, whereas RS and PS correlate considerably with each other. It is noteworthy, that the correlation of GA with CC is relatively higher than its correlation with RS and PS, whereas CC correlates with RS and PS more than it does with CC. The construct CC of CGT comprises the abilities, such as identifying the relationships among the ideas contained in a statement; identifying the consequences of an action or a statement; identifying the difference between facts and opinions; encoding and decoding messages; and understanding the data presented by charts, tables, graphs, and images. Therefore, it can be argued that verbal reasoning and problem-solving abilities are required for the manifestation of the abilities of CC, especially the comprehension.

4.3 Research Question 2: Reliability of the Constructs of the CGT – 2017

The reliability of the constructs was examined with respect to the values of Coefficient *H* (Hancock & Mueller, 2001). Regarding the matter of how large Coefficient *H* should be for a given construct, Hancock and Mueller (2001) refer to the magnitudes of 0.7 or 0.8 which Nunnally and Bernstein (1994, p. 265) recommend within single measured variable context and the magnitude of 0.7 which Hair, Tatham, Anderson, and Black (1998, p. 612) recommend for the reliability coefficients. Thus, the entire multidimensional scale of CGT was satisfactory since the value for the reliability coefficient ($H = 0.938$) which it has for the entire scale was acceptable. Though the reliability of the subscales, RS ($H = 0.844$) and PS ($H = 0.814$), are not considered excellent, they can be reported as acceptable measures in terms of reliability. However, it is noteworthy that the subscale for GA is reported with a smaller value compared to the other subscales (constructs), yet it is reasonably consistent with the margins mentioned above. Relative lower value for the reliability for GA subscale is due to the comparatively lower factor loadings of the items to the construct (see Appendix B). The Coefficient *H* concerns standardised factor loadings of the indicators for computation of the reliability measure. Thus, it provides a lower value for the value of the coefficient in relation to subscale GA.

Table 4.3

Reliability of CGT and its constructs estimated by Coefficient H

Measure	*H*
All Constructs	0.938
General Awareness (GA)	0.695
Reasoning (RS)	0.844
Problem Solving (PS)	0.814
Comprehension and Communication (CC)	0.743

4.4 Research Question 3: Relationship between Cognitive Ability and Achievement

In answering the research Question 3, which focuses on assessing the effectiveness of the cognitive abilities in predicting the achievement in GCE AL Examination, the Model 2A was primarily considered. However, due the observation of high intercorrelation of the observed variables, RS and PS, Model 2B was proposed by introducing a latent construct such that its variance is explained by RS and PS (see Figure 3.4). This modification of the original model (Model 2A) significantly improved the fit of the model with data (see Table 4.4).

Unlike the Model 1 that contained dichotomous data in observed variables, Model 2 contained continuous observed variables since composite scores were used for the indicators. Thus, the parameter estimation was supposed to be done using the Maximum Likelihood (ML) method. Since the normality of the observed variables is required in use of ML for estimation (Li, 2016), normality was examined with Henze-Zirkler's multivariate normality test (Henze & Zirkler, 1990) and the results (HZ = 2.004, $p < 0.001$) indicated deviation from multivariate normality. The MVN software package (Korkmaz, Goksuluk, & Zararsiz, 2018) of R statistical software was used for computation and the corresponding results for multivariate normality, and the relevant statistics and histograms for univariate normality are reported in detail in Appendix C.

As a remedy for the non-normality of the data, Finney and DiStefano (2006) suggest using a robust ML estimator that corrects for non-normality-induced bias in the standard errors. Thus, the MLM estimator of lavaan software package (Rosseel, 2012) was used since it produces maximum likelihood estimation with robust standard errors and Satorra-Bentler scaled χ^2 (Satorra & Bentler, 2010) fit statistic (lavaan, n.d.). Therefore, in reporting fit of the Model 2A and Model 2B with data, Satorra-Bentler scaled χ^2 fit statistics are reported instead of the standard χ^2 fit statistics.

Table 4.4

Model fit indices of Model 2A, 2B

Model	χ^2	*df*	χ^2_{diff}	df_{diff}	RMSEA	CFI	TLI	SRMR
Model 2A	123.673*	5			0.095	0.961	0.923	0.041
Model 2B	14.178*	3	112.41**	2	0.038	0.997	0.998	0.013

*P < 0.01 , ** p< 0.001

Regarding Model 2A, the model acceptance was underpinned with the value SRMR, which was less than its benchmark value 0.08, and the value for CFI, which was greater than its benchmark value 0.95 (L. t. Hu & Bentler, 1999). TLI was close to its cut-off value of 0.95 (Schreiber et al., 2006). However, the model was not considered as fitting adequately with the data since Satorra-Bentler scaled χ^2 statistic was significant, χ2(5) = 123.673, p<.001, and RMSEA value was 0.095, 95% CI [0.081, 0.110]. For the acceptance of the Model 2A, RMSEA should be in a range which is less than .06 to .08 with confidence interval (Schreiber et al., 2006). In this model, the probability of the RMSEA for being equal to or less than 0.05 was 0.

In contrast to the above observations, Model 2B was reported with satisfactory fit indices. Despite the Satorra-Bentler scaled $\chi 2$ statistic being significant, χ^2 (3) = 14.178, $p<.001$, RMSEA was 0.038, 95% CI [0.020, 0.058], which satisfied the requirement of being less than 0.6., SRMR value of 0.013 was well below its cut-off value of 0.8 (Schreiber et al., 2006). The CFI and TLI values were satisfactory, being above their benchmark values of 0.95 (L.-t. Hu & Bentler, 1998; L. t. Hu & Bentler, 1999; Schreiber et al., 2006). Therefore, these indices indicated the acceptance of the model. Moreover, the statistically significant difference between the two models χ^2 (2) = 112.41, $p<.001$ indicates the better explanation of variance by Model 2B than by Model 2A.

Though Model 2B explained the variance of the factors better than Model 2A, and the fit indices of the Model 2B indicated its acceptance, it can be argued that fit of the model with data is still a subject of debate. For example, Kenny, Kaniskan, and McCoach (2015) argue that RMSEA provides inconsistent results with the models that possess small degrees of freedom and small sample sizes. The measure of degrees of freedom of the Model 2B takes a small value (2).

The values for the coefficient of determination for the latent factor “Achievement in the GCE AL Examination” in Model 2A and Model 2B are 0.116 and 0.155 respectively (see Appendix C). This indicate that corresponding predictor variable “General Cognitive Ability” does not explain more than 11.6% and 15.5% of the variance of “Achievement in the GCE AL Examination” respectively in each model. Therefore, it is observed that, “General Cognitive Ability” which is derived with test scores for each construct of CGT explicitly does not provide a consistent prediction of the achievement in GCE AL Examination.

However, it is evident that PS and RS factors in CGT are highly correlated with each other and converged into an underlying construct, which results in improvement of the overall explanation of the variance of the measure.

Chapter 5: Discussion

5.1 Introduction

The purpose of this study was to investigate the relationship between the cognitive abilities of the GCE AL students and their academic achievement. Additionally, the study examined the validity and the reliability of the CGT scale in terms of measuring its constructs. Thus, the test scores of a sample of 2,623 candidates were analysed in terms of answering the research questions:

1. Is CGT valid in terms of measuring its constructs?
2. Is CGT reliable in terms of measuring its constructs?
3. Do the scores of CGT predict the achievement of the candidates for core subjects in the GCE AL Examination?

The results of the present study suggest that CGT is robust in terms of measuring its constructs. However, the present study revealed that the abilities measured by the CGT do not effectively predict the performance for the core subjects of the GCE AL Examination. Put another way, this finding supports the view that sole use of cognitive abilities does not provide a consistent prediction of academic achievement.

5.2 Robustness of CGT in Terms of Measuring Its Constructs

The proportionality of the constructs of the measure (CGT) has been discussed extensively by a number of authors in the literature on the measurement of cognitive ability (Carroll, 1993; Kyllonen & Kell, 2017; Mackintosh, 2011; Rosén et al., 2017). The research questions 1 and 2 were rather focused on the robustness of the measure concerned in terms of the validity and the reliability.

A significant association between the RS and PS was observed in this study; for example, the higher correlation of RS and PS can be seen in Table 4.2. It can be argued that this is a manifestation of interrelatedness of reasoning and problem solving that has been widely discussed in the literature (Jonassen & Hernandez-Serrano, 2002; Mackintosh, 2011; Sternberg, 1980). This close association has been discussed by authors in several aspects. For example, Mackintosh (2011) has stated that reasoning and problem solving are abilities that pertain to common factors of Gf. Moreover, Mayer (1992, 2011) and Simon (1983) have argued that reasoning can be explained as a type of problem solving which is needed in deductive and inductive reasoning tasks.

Despite the close relationship between reasoning and problem solving, it was acknowledged that the objectives of introducing CGT highlights the identification of abilities reasoning and problem solving separately. Perera (1999) states, “Though thinking with reasoning is an ability that is needed in problem solving, a part of the CGT is dedicated to measure it (reasoning) separately” (p. 12). Thus, including the constructs RS and PS, specifications and sample items are given by Perera (1999) as a guidance to measure each construct of CGT.

The capability of the indicators (test items) pertaining to RS and PS in highlighting the corresponding construct is questionable. Since the researcher of this study was unable to obtain the item specifications and blueprint of the test, the arrangement of the test items was determined through the informal communication with the people who designed the CGT in 2017. Thus, it was identified that items 16-30 and items 31-45 pertain to RS and PS respectively. However, regarding the CGT in 2017, it is questionable whether the corresponding indicators (test items) are able to highlight the two constructs RS and PS separately. For example, though one can argue that item no 26 (see Appendix D) is purported to measure the numerical reasoning, another supposition would be that the item no 26 is aimed at problem solving since it asks for the solution of a mathematical problem.

Despite the effectiveness in the test items relating to RS and PS in highlighting the abilities reasoning and problem solving, both these constructs of the CGT can be identified as the factors which are proportional to Gf. The review of the body of literature on this subject affirms that Gf is one of the major causes of achievement variances of the individuals, where it signifies the individual's capacity, especially in solving novel problems, making inferences, identifying patterns and relationships, reasoning and the extrapolation using logic, and transforming information (Kyllonen & Kell, 2017; Mackintosh, 2011; Rosén et al., 2017). It has been stated in this study that the constructs, PS and RS, are more proportional to Gf due to the corresponding abilities which are supposed to be measured in these constructs (see Table 2.1). This suggestion has been discussed in the Chapter 2 in relation to the ideas of Kyllonen and Kell (2017) about a common factor which accounts for several tasks. The authors emphasise that Gf develops into a general factor since it drives knowledge and skill acquisition in a variety of domains, such as vocabulary acquisition, rule induction, and developing associations between performances in those domains.

Results pertaining to the construct GA leads to further discussion regarding its validity and reliability. Compared to other constructs of the scale (CGT), GA which reflected the ability in GK was weakly correlated with the ability in CGT. This is rather contradictory with the finding of the study of Chamorro-Premuzic et al. (2006) which showed a higher positive correlation of GK with cognitive ability ($r = 0.46$) than with abstract reasoning ($r = 0.37$). However, it was found that more than half of the items (8 items out 13) pertaining to GA are reported with the standardised loading less than 0.3 which indicates less correlation of test items to the factor (construct).

5.3 Predictability of the Academic Achievement by Cognitive Abilities

The main focus of this study was to assess the predictability of the performance of the candidates in the GCE AL Examination by their cognitive abilities which are represented with the scores of CGT. Thus, models 2A and 2B were proposed to explain the predictability. The main finding of the results was the impact of the cognitive abilities pertaining to RS and PS and their interrelationship in predicting the achievement in GCE AL. Nesting the original model by introducing a three-factor model such that the factors (constructs) RS and PS have a common underlying factor significantly improved the fit of the model with data. It can be argued that the combination of the current type of test items which are used to measure PS and RS would improve the interpretation of the test results in terms of the objectives of the measure (CGT).

Despite the finding that fit indices indicated satisfactory model fit in relation to Model 2B with the data, the unique variance pertaining to the achievement in the GCE AL is reported with a higher value (see Appendix C). Moreover, the corresponding coefficient of determination was 0.155. This implies that the general cognitive ability measured by the CGT does not explain more than 15.5% of the variance of the achievement in GCE AL Examination. Such a phenomenon indicates the involvement of factors other than the cognitive ability of the candidates in explaining the variance of the achievement in the GCE AL Examination. The potential causal factors in addition to the cognitive ability can be discussed in relation to the literature.

The causal factors which were not concerned in Model 2A and 2B, yet are affective for the achievement of the candidates in the GCE AL Examination, can be further discussed in relation to assessment of the cognitive ability. Mackintosh (2011) describes distinctions of cognitive processes with several aspects, such as education, social class, family, and environment. Consequently, Kyllonen and Kell (2017) refer to the studies of Zigler and

Trickett (1978) and Kvist and Gustafsson (2008), and acknowledge that intelligence tests measure three distinct components: formal cognitive processes, school learning, and motivation. More importantly, the authors emphasise their view that, in the case that school learning and motivation components are found to be unequal, the intelligence tests provide poor measures of cognitive processing abilities. Thus, it can be argued that unequal opportunities for the education of students in Sri Lanka (Liyanage, 2014) may cause inconsistent measurement of their (cognitive) abilities.

Though Kyllonen and Kell (2017) identify that inequalities in motivation cause inconsistencies in intelligence tests, Sackett (2012) argues that motivation does not play a major role in determining differences in measurements of cognitive abilities when respective tests are administered under high-stakes conditions, since the test takers' maximum effort is given in such circumstances. However, it can be argued that motivation plays a major role in academic achievement since it has been widely discussed and demonstrated in the literature (Duchesne & McMaugh, 2018; Lucidi, 2011; Meece et al., 2006; Yousefy et al., 2012). Moreover, focusing on the self-regulated learning, Zimmerman (1990) acknowledges that self-regulated learning is a major factor that is concerned in determining how learners promote their own academic achievement: cognitively, motivationally, and behaviourally.

Inequalities in school learning, the second major issue that Kyllonen and Kell (2017) identify as a causal factor of producing inconsistencies in intelligence tests, can also be discussed in relation to academic achievement. Rosén et al. (2017) acknowledge that educational attainment is significantly affected by the school organisational features and academic processes, such as curriculum, resources, and teacher competence. Similarly, Mackintosh (2011) acknowledge the different contents that teaching and learning process in schools focuses on, as a factor which causes deficiencies in correlation between cognitive abilities and academic performance. In addition to the content, the teacher behaviour in the teaching-learning process is found to be a factor which is strongly correlated with student achievement (Creemers & Kyriakides, 2013; Scheerens, Witziers, & Steen, 2013). Overall, the quality of the school is a major factor which affects growth in school achievement (Finn et al., 2014). Thus, it can be argued that school learning affects both the cognitive abilities of the students and their academic achievement.

5.4 Limitations

Several limitations were identified in the current study and it is believed that they cause inconsistencies for the generalisation of the results of the study. Inability to have control over the data collection is one of the major weaknesses in the current study. It is based on secondary data which have been collected from the archives of the DOE SL. The simple random sampling which had been followed to select the sample is somewhat questionable. A one-sample t-test was conducted to assess the representation of the target population by the sample, and it was noted that the test scores for the CGT of the sample (M = 52.217, SD = 12.978) were significantly different from the population, $t(2,622) = 5.560$, $p < 0.05$. A comparison with metadata of the candidates determined that candidates from some regions (administrative districts) of Sri Lanka have been omitted in the sample. This was a crucial inconsistency in the study, and it weakens the generalisability of the results to the target population.

The present research was carried out in a cross-sectional design. Confinement of the data collection to a particular year was another limitation in generalising the results of the study. Since the itemised data for the CGT was available only for 2017, the current study reflects results from only that year. A more comprehensive and dependable approach to the research questions would have been possible with the data from several years.

The use of measures with relatively less reliable subscales caused limitations in dependable decision making. Though the reliability of the subscales PS (H = 0.814) and RS (H = 0.844) were at a satisfactory level, the subscales GA (H = 0.695) and CC (H = 0.743) were not up to the level of making consistent decisions. Schumacker (2005) argues that a value for the reliability should be more than 0.8 for dependable decision making.

The determination of the validity and the reliability of the test only in the aspect of constructs is another confinement. It is known that validity is discussed in several aspects apart from the construct validity. For example, GA subscale is supposed to measure the general knowledge (GK) in economic, political, social, cultural, environmental, legal, scientific, and technological spheres at the local and international levels. Yet, it was observed that some domains, such as cultural and legal spheres, have not been concerned, which leads to doubts about the content validity.

5.5 Implications

As a contribution for the quality improvement of CGT, the test items pertaining to the construct GA can be revisited and a model test paper can be created so that they are more applicable to the abilities of Gc. In this regard, rather than questioning on isolated subjects (for example, “Who is the current president of United States of America?”) the concept of the test items can be bounded by the cultural, social, and lawful norms, knowledge, and beliefs more relevant to the Sri Lankan and global contexts (for example, questioning on a place or a city where unique features of it are given).

5.6 Future Study

Extensions of the current study could focus more on quality improvements of the CGT using empirical evidence. In this context, the assessment of the validity and reliability of the test in relation to several factors is more important. The current study was confined to the assessment of quality of the test and the association of its test scores with the performance for the core subjects of the GCE AL Examination. As an extension of the study, the assessment of robustness of the test in several aspects would be advantageous. The test bias concerned in terms of subgroups of the population of the CGT would be significant in research and development tasks of the measure. Such subgroups can be identified mainly in terms of subject stream, mode of instruction (language used for the test), and region. Thus, statistically significant differences among the groups which are defined with the above factors and factor loadings of the items for corresponding constructs would be useful for the test designers to distinguish the inconsistent test items.

Chapter 6: Conclusions

The relationship between the cognitive abilities and academic achievement is widely discussed and well established in the literature (Rosén et al., 2017). The narrow scope of assessment content in national examinations in Sri Lanka is a significant and concerning issue, whereas improved methods of assessment reflecting the higher order cognitive abilities are now being discussed widely (Dundar et al.; Sedere et al., 2016). The purpose of this study was to determine the predictability of the achievement of the test takers of the Sri Lankan GCE AL Examination by their cognitive abilities. In the literature, it is often stated that the association between motivation and cognitive abilities is crucial in predicting academic achievement (Meece et al., 2006; Pedaste et al., 2015; Yousefy et al., 2012). However, a contrary argument claims that motivation is not significant in high-stake test administrations, since test takers theoretically would respond with their maximum effort in such circumstances (Sackett, 2012).

This study used secondary data from the candidates of the Sri Lankan GCE AL Examination, which is a high-stake examination conducted on a large scale. The study was carried out as a non-experimental correlational study with a cross-sectional design. A random sample of 2,623 candidates selected from a population of 240,693 who took three core subjects along with the CGT in 2017 examination were considered in the study. Average standardised scores for the core subjects and the scores for the CGT represented the achievement of the test takers at the examination and their general cognitive ability respectively. Thus, the research questions:

1. Is CGT valid in terms of measuring its constructs?
2. Is CGT reliable in terms of measuring its constructs?

3. Do the scores of CGT predict the achievement of the candidates for core subjects of the GCE AL Examination?

were addressed in a structural equation modelling framework.

Based on the priori of the composition of the constructs of the CGT (Department of Examinations - Sri Lanka, 2000; Perera, 1999), a four-factor model was proposed to explain the factor structure. The proposed model was reported with satisfactory fit indices and possessed an acceptable level of construct reliability. The relationship between RS and PS was significant. In assessing the predictability of the achievement in the GCE AL Examination using cognitive ability, the four-factor model which was considered such that RS and PS are two separate constructs was not reported with satisfactory fit indices. Yet, a three-factor model which contained a latent factor of RS and PS, fitted with data satisfactorily. Thus, it was concluded that the high correlation between the constructs RS and PS were due to the ineffectiveness of the test items in highlighting constructs distinctly.

Even though the three factor model fits with data, the proportion of the total variance of achievement in the GCE AL explained by the general cognitive ability was noticeably small. Such a finding affirmed the inadequacy of using cognitive ability solely in predicting academic achievement. This finding underpins the suggestion that inequality in school learning and motivation provides inconsistent results of measurements of cognitive abilities (Kyllonen & Kell, 2017). Put in another way, this finding implies the significance of causal factors relating to social, environmental, and behavioural circumstances of students in predicting the academic achievement in addition to the cognitive ability.

References

Abeygunasekera, G. (2011). History of the examinations in Sri Lanka. In K. M. H. Bandara (Ed.), *Vajra Jayanthi Samaru Kalapaya (Souvenir for the 60th Anniversary of the Department of Examinations - Sri Lanka)*. Department of Examinations - Sri Lanka (National Evaluation and Testing Service).

Adesoji, F. A., & Olatunbosun, S. M. (2008). Student, Teacher And School Environment Factors As Determinants of Achievement In Senior Secondary School Chemistry in Oyo State, Nigeria. *Journal of international social research, 1*(2).

Airasian, P. W., & Gay, L. (2003). *Educational research: Competencies for analysis and application*: Prentice Hall.

Alin, A. (2010). Multicollinearity. *Wiley Interdisciplinary Reviews: Computational Statistics, 2*(3), 370-374. doi:doi:10.1002/wics.84

Allen, M. P. (1973). Construction of composite measures by the canonical-factor-regression method. *Sociological methodology, 5*, 51-78.

American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (2014). *The Standards for Educational and Psychological Testing*.

Bacon, D. R., Sauer, P. L., & Young, M. (1995). Composite Reliability in Structural Equations Modeling. *Educational and Psychological Measurement, 55*(3), 394-406. doi:10.1177/0013164495055003003

Batey, M., Furnham, A., & Safiullina, X. (2010). Intelligence, general knowledge and personality as predictors of creativity. *Learning and Individual Differences, 20*(5), 532-535. doi:https://doi.org/10.1016/j.lindif.2010.04.008

BBC Sinhala News (2018). Children's Education: "The children's intelligence is not measured by exams in Sri Lanka". Retrieved from https://translate.google.com/translate?sl=auto&tl=en&u=https%3A%2F%2Fwww.bbc.com%2Fsinhala%2Fsri-lanka-46602475Ben-Gal, I. (2005). Outlier Detection. In O. Maimon & L. Rokach (Eds.), *Data Mining and Knowledge Discovery Handbook* (pp. 131-146). Boston, MA: Springer US.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological bulletin, 107*(2), 238.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological bulletin, 88*(3), 588.

Brown, T. A. (2014). *Confirmatory factor analysis for applied research*: Guilford Publications.

Brown, T. A., & Moore, M. T. (2012). Confirmatory factor analysis. *Handbook of structural equation modeling*, 361-379.

Brunner, M., & SÜβ, H.-M. (2005). Analyzing the Reliability of Multidimensional Measures: An Example from Intelligence Research. *Educational and Psychological Measurement, 65*(2), 227-240. doi:10.1177/0013164404268669

Candrasekaran, S. (2013). Creativity and academic achievement of higher secondary school students in Tamilnadu. *International Journal of Humanities and Social Science Invention, 3*(8), 32-36.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*: Cambridge University Press.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54*(1), 1.

Cattell, R. B. (1987). *Intelligence: Its structure, growth and action* (Vol. 35): Elsevier.

Chamorro-Premuzic, T., Furnham, A., & Ackerman, P. L. (2006). Ability and personality correlates of general knowledge. *Personality and Individual Differences, 41*(3), 419-429. doi:https://doi.org/10.1016/j.paid.2005.11.036

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural equation modeling, 9*(2), 233-255. doi:10.1207/S15328007SEM0902_5

Collins, R. (2014). Skills for the 21st Century: teaching higher-order thinking.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of applied psychology, 78*(1), 98.

Coyle, T., Snyder, A., Pillow, D., & Kochunov, P. (2011). SAT predicts GPA better for high ability subjects: Implications for Spearman's Law of Diminishing Returns. Personality and Individual Differences, 50(4), 470-474.

Creemers, B., & Kyriakides, L. (2013). *Improving quality in education: Dynamic approaches to school improvement*: Routledge.

Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative*: Prentice Hall Upper Saddle River, NJ.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological bulletin, 52*(4), 281.

Debelak, R., & Tran, U. S. (2016). Comparing the Effects of Different Smoothing Algorithms on the Assessment of Dimensionality of Ordered Categorical Items with Parallel Analysis. *PloS one, 11*(2), e0148143.

Department of Examinations - Sri Lanka. (2000). *Samaanya Podu Parikshanaya Pilibanda Vigrahayak*. Department of Examinations - Sri Lanka.

Department of Examinations - Sri Lanka. (2015). *G.C.E. (A/L) Examination - Test papers and prototype test Items for the examinations conducted after 2015*. Department of Examinations - Sri Lanka.

Department of Examinations - Sri Lanka. (2017a). Evaluation Report 2016 - Sinhala Medium. Retrieved from https://www.doenets.lk/exam/EVALUATION_REPORT-2016_Sinhala.jsf

Department of Examinations - Sri Lanka. (2017b). Evaluation Reports 2016 - Sinhala Medium.

Department of Examinations - Sri Lanka. (2018). *G.C.E (A.L) Examination 2017 - Performance of Candidates*. Retrieved from https://www.doenets.lk/exam/docs/comm/G.C.E.(A.L.)%20Examination%202017-Perfomance%20of%20Candidates.pdf.

DeSimone, J. A., Harms, P. D., & DeSimone, A. J. (2015). Best practice recommendations for data screening. *Journal of Organizational Behavior, 36*(2), 171-181.

DiStefano, C., Zhu, M., & Mindrila, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research & Evaluation, 14*(20), 1-11.

Duchesne, S., & McMaugh, A. (2018). *Educational psychology for learning and teaching*: Cengage AU.

Dundar, H., Béteille, T., & Riboud, M. Monitoring Learning Outcomes: Student Assessment Systems. In *Student Learning in South Asia: Challenges, Opportunities, and Policy Priorities* (pp. 295-319).

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology, 105*(3), 399-412. doi:doi:10.1111/bjop.12046

Finn, A. S., Kraft, M. A., West, M. R., Leonard, J. A., Bish, C. E., Martin, R. E., . . . Gabrieli, J. D. (2014). Cognitive skills, student achievement tests, and schools. *Psychological Science*, *25*(3), 736-744.

Finney, S. J., & DiStefano, C. (2006). Non-normal and categorical data in structural equation modeling. *Structural equation modeling: A second course, 10*(6), 269-314.

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological methods, 9*(4), 466.

Furnham, A., & Chamorro-Premuzic, T. (2006). Personality, intelligence and general knowledge. *Learning and Individual Differences, 16*(1), 79-90. doi:https://doi.org/10.1016/j.lindif.2005.07.002

Furnham, A., Swami, V., Arteche, A., & Chamorro-Premuzic, T. (2008). Cognitive ability, learning approaches and personality correlates of general knowledge. *Educational Psychology, 28*(4), 427-437. doi:10.1080/01443410701727376

Furnham, A., Weis, L., & Gee, M. (n.d.). The Assessment of General Knowledge Online: A cross-cultural study using two platforms. Retrieved from http://discovery.ucl.ac.uk/1508881/1/Furnham_Knowledge%20of%20mental%20illnesses%20two%20studies%20using%20a%20new%20test_AAM.pdf

Geisinger, K. F. (2013). Reliability. In *APA handbook of testing and assessment in psychology* (pp. 21-42).

Goodwin, L. D. (1999). The Role of Factor Analysis in the Estimation of Construct Validity. *Measurement in Physical Education and Exercise Science, 3*(2), 85-100. doi:10.1207/s15327841mpee0302_2

Graham, J. M. (2006). Congeneric and (essentially) tau-equivalent estimates of score reliability: What they are and how to use them. *Educational and Psychological Measurement, 66*(6), 930-944.

Gustafsson, J.-E., & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research, 28*(4), 407-434.

Hair, J. F., Tatham, R. L., Anderson, R. E., & Black, W. (1998). *Multivariate Data Analysis*.

Hancock, G., & Mueller, R. O. (2001). *Rethinking construct reliability within latent variable systems*.

He, Q., Stockford, I., & Meadows, M. (2018). Inter-subject comparability of examination standards in GCSE and GCE in England. *Oxford Review of Education, 44*(4), 494-513. doi:10.1080/03054985.2018.1430562

Henze, N., & Zirkler, B. (1990). A class of invariant consistent tests for multivariate normality. *Communications in Statistics - Theory and Methods, 19*(10), 3595-3617. doi:10.1080/03610929008830400

Henson, R. K. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement and evaluation in counseling and development, 34*(3), 177.

Hershey, D. A., Austin, J. T., & Gutierrez, H. C. (2015). Chapter 16 - Financial Decision Making across the Adult Life Span: Dynamic Cognitive Capacities and Real-World Competence. In T. M. Hess, J. Strough, & C. E. Löckenhoff (Eds.), *Aging and Decision Making* (pp. 329-349). San Diego: Academic Press.

Horn, D., & Kiss, H. J. (2018). Which preferences associate with school performance?-Lessons from an exploratory study with university students. *PloS one, 13*(2), e0190163-e0190163. doi:10.1371/journal.pone.0190163

Hox, J. J., & Bechger, T. M. (1998). An introduction to structural equation modeling.

Hoyle, R. H. (1995). *Structural equation modeling: Concepts, issues, and applications*: Sage.

Hu, L.-t., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological methods, 3*(4), 424.

Hu, L. t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55.

Hubley, A. M., & Zumbo, B. D. (2013). Psychometric characteristics of assessment procedures: An overview. *APA handbook of testing and assessment in psychology, 1*, 3-19.

Johnsen, S. K. (2004). Definitions, models, and characteristics of gifted students. *Identifying gifted students: A practical guide*, 1-22.

Johnson, B. (2001). Toward a new classification of nonexperimental quantitative research. *Educational Researcher, 30*(2), 3-13. doi:10.3102/0013189X030002003

Jonassen, D. H., & Hernandez-Serrano, J. (2002). Case-based reasoning and instructional design: Using stories to support problem solving. *Educational Technology Research and Development, 50*(2), 65-77.

Kane, M. (2013). The argument-based approach to validation. *School Psychology Review, 42*(4).

Kaufman, S. B., Reynolds, M. R., Liu, X., Kaufman, A. S., & McGrew, K. S. (2012). Are cognitive g and academic achievement g one and the same g? An exploration on the Woodcock–Johnson and Kaufman tests. *Intelligence, 40*(2), 123-138. doi:10.1016/j.intell.2012.01.009

Kenny, D. A. (2015). Measuring Model Fit. Retrieved from http://www.davidakenny.net/cm/fit.htm

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The Performance of RMSEA in Models With Small Degrees of Freedom. *Sociological Methods & Research, 44*(3), 486-507. doi:10.1177/0049124114543236

Khan, Z. N. (2005). Scholastic Achievement of Higher Secondary Students in Science Stream. *Online Submission, 1*(2), 84-87.

Korkmaz, S., Goksuluk, D., & Zararsiz, G. (2018). MVN: An R Package for Assessing Multivariate Normality (Version 5.5). Retrieved from https://cran.r-project.org/web/packages/MVN/vignettes/MVN.pdf

Kvist, A. V., & Gustafsson, J.-E. (2008). The relation between fluid intelligence and the general factor as a function of cultural background: A test of Cattell's Investment theory. *Intelligence, 36*(5), 422-436.

Kyllonen, P., & Kell, H. (2017). What Is Fluid Intelligence? Can It Be Improved? In M. Rosén, H. K. Yang, & U. Wolff (Eds.), *Cognitive Abilities and Educational Outcomes*: Springer.

lavaan. (n.d.). Estimators. Retrieved from http://lavaan.ugent.be/tutorial/est.html

Levin, K. A. (2006). Study design III: Cross-sectional studies. *Evidence Based Dentistry, 7*, 24. doi:10.1038/sj.ebd.6400375

Li, C.-H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods, 48*(3), 936-949. doi:10.3758/s13428-015-0619-7

Liyanage, K. (2014). Education System of Sri Lanka : Strengths and Weaknesses.

Lucidi, F. (2011). Relationship Between Social Context, Self-Efficacy, Motivation, Academic Achievement, and Intention to Drop Out of High School: A Longitudinal Study AU - Alivernini, Fabio. *The Journal of educational research, 104*(4), 241-252. doi:10.1080/00220671003728062

Mackintosh, N. J. (2011). *IQ and human intelligence*: Oxford University Press.

Mardia, K. V. (1974). Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *Sankhyā: The Indian Journal of Statistics, Series B*, 115-128.

Martin, R. P. (1988). *Assessment of personality and behavior problems: Infancy through adolescence*: Guilford Press New York.

Mayer, R. E. (1992). *Thinking, problem solving, cognition*: WH Freeman/Times Books/Henry Holt & Co.

Mayer, R. E. (2011). Problem solving and reasoning. *Learning and cognition in education*, 112-117.

McDonald, R. P. (1970). The theoretical foundations of principal factor analysis, canonical factor analysis, and alpha factor analysis. *British Journal of Mathematical and Statistical Psychology*, *23*(1), 1-21.

McNeish, D. (2017). *Thanks Coefficient Alpha, We'll Take it From Here* (Vol. 23).

Meece, J. L., Anderman, E. M., & Anderman, L. H. (2006). Classroom goal structure, student motivation, and academic achievement. *Annu. Rev. Psychol.*, *57*, 487-503.

Ministry of Education (2016). *Subject Combinations for the GCE AL Examination and Subject combinations for the university entrance*.

Mueller, R. O. (1997). Structural equation modeling: Back to basics. *Structural Equation Modeling: A Multidisciplinary Journal, 4*(4), 353-369.

Mueller, R. O. (1999). *Basic principles of structural equation modeling: An introduction to LISREL and EQS*: Springer Science & Business Media.

Mueller, R. O., & Hancock, G. R. (2008). Best practices in structural equation modeling. *Best practices in quantitative methods, 488508*.

Murphy, R. (2007). Common Test Methods. In P. Newton, J.-A. Baird, H. Goldstein, H. Patrick, & P. Tymms (Eds.), *Techniques for monitoring the comparability of examination standards*.

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of mathematical Psychology*, *47*(1), 90-100.

National Education Commission, Sri Lanka. (1998). *Reforms in General Education 1997*. Retrieved from http://nec.gov.lk/post4-2/.

National Education Commission, Sri Lanka (2009). *New Education Act for General Education in Sri Lanka - Context, Issues and Proposals* Retrieved from http://nec.gov.lk/wp-content/uploads/2017/12/Towards-a-New-Education-Act.pdf.

Newton, P., Baird, J., Goldstein, H., Patrick, H., & Tymms, P. (2007). Techniques for monitoring the comparability of examination standards (London, Qualifications and Curriculum Authority). *Tognolini, J. & Andrich, D.(1996) Analysis of profiles of students applying for entrance to universities, Applied Measurement in Education, 9*(4), 323-353.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*.

Ofqual. (2015). Inter-Subject Comparability: A Review of the Technical Literature: ISC Working Paper No 2. In: Coventry: The Office of Qualifications and Examinations Regulation.

Pedaste, M., Must, O., Silm, G., Täht, K., Kori, K., Leijen, Ä., & Mägi, M.-L. (2015). *How do cognitive ability and study motivation predict the academic performance of IT students?* Paper presented at the ICERI conference.

Perera, L. (1999). *Samanya Podu Pareekshanaya Sandaha Igenum Athwelak (A learning guide for the Common General Test)*: Ministry of Education and Higher Education.

Plomin, R. (1999). Genetics and general cognitive ability. Nature, 402(6761supp), C25.

PostgreSQL. (2018). PostgreSQL: Open Source Relational Database (Version 10.2). Retrieved from https://www.postgresql.org/

Raykov, T. (1997). Scale reliability, Cronbach's coefficient alpha, and violations of essential tau-equivalence with fixed congeneric components. *Multivariate Behavioral Research, 32*(4), 329-353.
Raykov, T. (1998). Coefficient alpha and composite reliability with interrelated nonhomogeneous items. *Applied Psychological Measurement, 22*(4), 375-385.
Reynolds, C. R., & Livingston, R. B. (2012). *Mastering modern psychological testing: Theory and methods*: Pearson Education.
Rohde, T. E., & Thompson, L. A. (2007). Predicting academic achievement with cognitive ability. *Intelligence, 35*(1), 83-92. doi:10.1016/j.intell.2006.05.004
Rosén, M., Yang Hansen, K., & Wolff, U. (2017). *Cognitive Abilities and Educational Outcomes*: Springer.
Rosseel, Y. (2012). lavaan : An R Package for Structural Equation Modeling. *Journal of Statistical Software, 48*(2), 1-36.
Rosseel, Y. (2014). Structural Equation Modeling with categorical variables. Retrieved from https://personality-project.org/r/tutorials/summerschool.14/rosseel_sem_cat.pdf
Sackett, P. R. (2012). Cognitive Tests, Constructs, and Content Validity: A commentary on S chmidt (). *International Journal of Selection and Assessment, 20*(1), 24-27.
Satorra, A., & Bentler, P. M. (2010). Ensuring Positiveness of the Scaled Difference Chi-square Test Statistic. *Psychometrika, 75*(2), 243-248. doi:10.1007/s11336-009-9135-y
Scheerens, J., Witziers, B., & Steen, R. (2013). *A Meta-analysis of School Effectiveness Studies: Un metaanálisis de estudios de eficacia escolar*: Ministerio de Educación.
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.
Schumacker, R. E. (2005). Standards for interpreting reliability coefficients. *Retrieved February, 24*, 2008.
Sedere, U., Karunaratne, S., Karunanithy, M., & Jayasinghe-Mudalige, U. (2016). *Study on Evaluation & the assessment system in general education in Sri Lanka*. National Education Commision of Sri Lanka.
Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika, 74*(1), 107.
Simon, H. A. (1983). Search and reasoning in problem solving. *Artificial intelligence, 21*(1-2), 7-29.
Sireci, S. G., & Sukin, T. (2013). Test validity. *APA handbook of testing and assessment in psychology, 1*, 61-84.
Spinath, B., Spinath, F. M., Harlaar, N., & Plomin, R. (2006). Predicting school achievement from general cognitive ability, self-perceived ability, and intrinsic value. *Intelligence, 34*(4), 363-374. doi:10.1016/j.intell.2005.11.004
Steiger, J. H. (1980). *Statistically based tests for the number of common factors.* Paper presented at the the annual meeting of the Psychometric Society. Iowa City, IA. 1980.
Sternberg, R. J. (1980). *Reasoning, Problem Solving, and Intelligence*. Retrieved from
Strauss, M. E., & Smith, G. T. (2009). Construct Validity: Advances in Theory and Methodology. *Annual Review of Clinical Psychology, 5*(1), 1-25. doi:10.1146/annurev.clinpsy.032408.153639
Streiner, D. L. (2003). Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment, 80*(1), 99-103.
Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics*: Allyn & Bacon/Pearson Education.
Team, R. C. (2018). R: A Language and Environment for Statistical Computing. Retrieved from https://www.R-project.org/

Teo, T., & Fan, X. (2013). Coefficient alpha and beyond: Issues and alternatives for educational research. *The Asia-Pacific Education Researcher, 22*(2), 209-213.

Thorsen, C., Gustafsson, J. E., & Cliffordson, C. (2014). The influence of fluid and crystallized intelligence on the development of knowledge and skills. *British journal of educational psychology, 84*(4), 556-570.

Trizano-Hermosilla, I., & Alvarado, J. M. (2016). Best Alternatives to Cronbach's Alpha Reliability in Realistic Conditions: Congeneric and Asymmetrical Measurements. *Frontiers in psychology, 7*, 769-769. doi:10.3389/fpsyg.2016.00769

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38*(1), 1-10.

Ullman, J. B. (2006). Structural Equation Modeling: Reviewing the Basics and Moving Forward. *Journal of Personality Assessment, 87*(1), 35-50.

University Grants Commission (2017). *Admission to undergraduate courses of the universities in Sri Lanka - Academic Year 2017/2018*. University Grants Commission Retrieved from http://www.ugc.ac.lk/downloads/admissions/Handbook_2017_18/ENGLISH%20HANDBOOK%202017-2018.pdf.Wechsler, D. (1981). Manual for the adult intelligence scale-revised. *New York: Psychological Corporation*.

Wijetunge, S., & Rupasinghe, S. (2014). *The Senior Secondary School Curriculum (Grades 10-13)*. National Education Commission - Sri Lanka.

Yousefy, A., Ghassemi, G., & Firouznia, S. (2012). Motivation and academic achievement in medical students. *Journal of education and health promotion, 1*.

Zhang, L. F. (2007). Intellectual styles and academic achievement among senior secondary school students in rural China. *Educational Psychology, 27*(5), 675-692.

Zigler, E., & Trickett, P. K. (1978). IQ, social competence, and evaluation of early childhood intervention programs. *American Psychologist, 33*(9), 789.

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational psychologist, 25*(1), 3-17.

Appendix A

Instrumentation

Format 1 – Marks of the candidates by subject

Serial No	Subject	Marks

Format 2 – Answers of the candidates (sample) for the items of CGT

Serial No	Given answer		
	Item 1	……	Item 60

Format 3 - Correct answers for the items of CGT

Item No	Correct Answer	Item No	Correct Answer	Item No	Correct Answer
Item_1	2	Item_21	5	Item_41	3
Item_2	1	Item_22	3	Item_42	4
Item_3	4	Item_23	4	Item_43	3
Item_4	1	Item_24	4	Item_44	4
Item_5	5	Item_25	2	Item_45	5
Item_6	1	Item_26	1	Item_46	2
Item_7	3	Item_27	4	Item_47	5
Item_8	1	Item_28	3	Item_48	3
Item_9	2	Item_29	4	Item_49	1
Item_10	1	Item_30	4	Item_50	2
Item_11	5	Item_31	2	Item_51	3
Item_12	2	Item_32	3	Item_52	5
Item_13	3	Item_33	1	Item_53	1
Item_14	4	Item_34	3	Item_54	1
Item_15	5	Item_35	2	Item_55	4
Item_16	3	Item_36	3	Item_56	4
Item_17	4	Item_37	1	Item_57	5
Item_18	4	Item_38	1	Item_58	3
Item_19	2	Item_39	2	Item_59	4
Item_20	5	Item_40	2	Item_60	1

Appendix B

Fit Statistics of Model 1

Model Summary

lavaan 0.6-3 ended normally after 175 iterations

Optimization method	NLMINB	
Number of free parameters	139	
Number of observations	2623	
Estimator	DWLS	Robust
Model Fit Test Statistic	2387.781	2459.252
Degrees of freedom	1572	1572
P-value (Chi-square)	0.000	0.000
Scaling correction factor		1.171
Shift parameter		419.961

for simple second-order correction (Mplus variant)

Model test baseline model:

Minimum Function Test Statistic		38076.281	23757.412	
Degrees of freedom		1653	1653	
P-value		0.000	0.000	

User model versus baseline model:

Comparative Fit Index (CFI)		0.978	0.960	
Tucker-Lewis Index (TLI)		0.976	0.958	
Robust Comparative Fit Index (CFI)			NA	
Robust Tucker-Lewis Index (TLI)			NA	

Root Mean Square Error of Approximation:

RMSEA		0.014	0.015	
90 Percent Confidence Interval	0.013	0.015	0.014	0.016
P-value RMSEA <= 0.05		1.000	1.000	
Robust RMSEA			NA	
90 Percent Confidence Interval			NA	NA

Standardized Root Mean Square Residual:

SRMR	0.045	0.045

Parameter Estimates:

Information	Expected
Information saturated (h1) model	Unstructured
Standard Errors	Robust.sem

Latent Variables:	Estimate	Std.Err	z-value	P(>|z|)	Std.lv	Std.all
GA =~						
Item_1	1.000				0.089	0.089
Item_2	6.260	2.563	2.443	0.015	0.559	0.559
Item_3	1.728	0.811	2.130	0.033	0.154	0.154
Item_4	4.304	1.807	2.383	0.017	0.384	0.384
Item_5	5.926	2.446	2.423	0.015	0.529	0.529
Item_6	2.601	1.176	2.211	0.027	0.232	0.232
Item_8	3.312	1.419	2.333	0.020	0.296	0.296
Item_9	1.722	0.803	2.145	0.032	0.154	0.154
Item_10	3.283	1.397	2.350	0.019	0.293	0.293
Item_11	3.027	1.369	2.211	0.027	0.270	0.270
Item_12	3.744	1.560	2.400	0.016	0.334	0.334
Item_14	5.698	2.366	2.408	0.016	0.509	0.509
Item_15	6.013	2.529	2.377	0.017	0.537	0.537
RS =~						
Item_16	1.000				0.162	0.162
Item_17	3.281	0.555	5.915	0.000	0.532	0.532

Item_18	3.868	0.650	5.949	0.000	0.627	0.627
Item_19	2.872	0.494	5.809	0.000	0.465	0.465
Item_20	1.171	0.268	4.373	0.000	0.190	0.190
Item_21	3.413	0.582	5.865	0.000	0.553	0.553
Item_22	2.241	0.424	5.282	0.000	0.363	0.363
Item_23	3.862	0.655	5.900	0.000	0.626	0.626
Item_24	1.852	0.358	5.177	0.000	0.300	0.300
Item_25	3.647	0.617	5.912	0.000	0.591	0.591
Item_26	4.393	0.735	5.976	0.000	0.712	0.712
Item_27	3.068	0.517	5.930	0.000	0.497	0.497
Item_28	2.317	0.408	5.684	0.000	0.375	0.375
Item_29	2.739	0.471	5.814	0.000	0.444	0.444
Item_30	3.307	0.554	5.965	0.000	0.536	0.536
PS =~						
Item_31	1.000				0.459	0.459
Item_32	0.816	0.066	12.281	0.000	0.374	0.374
Item_33	1.459	0.080	18.244	0.000	0.669	0.669
Item_34	0.852	0.070	12.153	0.000	0.391	0.391
Item_35	1.070	0.070	15.344	0.000	0.491	0.491
Item_36	0.784	0.070	11.170	0.000	0.359	0.359
Item_37	1.284	0.076	16.928	0.000	0.589	0.589
Item_38	0.973	0.068	14.394	0.000	0.446	0.446
Item_39	1.292	0.075	17.128	0.000	0.592	0.592
Item_40	1.019	0.070	14.603	0.000	0.467	0.467
Item_41	0.386	0.065	5.910	0.000	0.177	0.177
Item_42	0.382	0.064	5.930	0.000	0.175	0.175
Item_43	1.116	0.073	15.242	0.000	0.512	0.512
Item_44	0.537	0.064	8.337	0.000	0.246	0.246
Item_45	0.462	0.060	7.646	0.000	0.212	0.212
CC =~						

Item_46	1.000				0.388	0.388
Item_47	1.246	0.111	11.177	0.000	0.484	0.484
Item_48	0.859	0.091	9.453	0.000	0.333	0.333
Item_49	0.940	0.091	10.277	0.000	0.365	0.365
Item_50	1.084	0.109	9.973	0.000	0.421	0.421
Item_51	0.359	0.080	4.499	0.000	0.139	0.139
Item_52	0.820	0.111	7.383	0.000	0.318	0.318
Item_53	0.959	0.147	6.520	0.000	0.372	0.372
Item_54	1.624	0.125	12.965	0.000	0.630	0.630
Item_55	0.859	0.093	9.234	0.000	0.333	0.333
Item_56	0.981	0.116	8.442	0.000	0.381	0.381
Item_57	0.771	0.112	6.896	0.000	0.299	0.299
Item_58	0.971	0.126	7.724	0.000	0.377	0.377
Item_59	0.443	0.085	5.196	0.000	0.172	0.172
Item_60	0.637	0.086	7.432	0.000	0.247	0.247

Covariances:

	Estimate	Std.Err	z-value	P(>|z|)	Std.lv	Std.all
GA ~~						
PS	0.020	0.008	2.403	0.016	0.485	0.485
CC	0.021	0.009	2.390	0.017	0.616	0.616
RS ~~						
PS	0.072	0.013	5.745	0.000	0.974	0.974
CC	0.048	0.009	5.412	0.000	0.771	0.771
PS ~~						
CC	0.160	0.014	11.064	0.000	0.897	0.897
.Item_22 ~~						
.Item_23	0.106	0.036	2.958	0.003	0.106	0.146
.Item_32 ~~						
.Item_33	0.249	0.027	9.196	0.000	0.249	0.361

.Item_34	0.039	0.033	1.196	0.232	0.039	0.046
.Item_33 ~~						
.Item_34	0.058	0.026	2.269	0.023	0.058	0.085
.Item_35 ~~						
.Item_36	0.042	0.028	1.494	0.135	0.042	0.051
.Item_37 ~~						
.Item_38	0.238	0.025	9.412	0.000	0.238	0.329
.Item_41 ~~						
.Item_42	0.500	0.026	19.291	0.000	0.500	0.516
.Item_43 ~~						
.Item_44	0.131	0.029	4.537	0.000	0.131	0.157
.Item_47 ~~						
.Item_48	0.064	0.032	2.040	0.041	0.064	0.078
.Item_49 ~~						
.Item_50	0.152	0.030	5.003	0.000	0.152	0.180
.Item_51	0.135	0.030	4.523	0.000	0.135	0.147
.Item_50 ~~						
.Item_51	0.066	0.031	2.100	0.036	0.066	0.073
.Item_52 ~~						
.Item_53	0.376	0.051	7.407	0.000	0.376	0.428
.Item_56 ~~						
.Item_57	0.231	0.051	4.563	0.000	0.231	0.262
.Item_58	0.301	0.049	6.169	0.000	0.301	0.352
.Item_57 ~~						
.Item_58	0.330	0.051	6.484	0.000	0.330	0.374
.Item_59 ~~						
.Item_60	0.124	0.032	3.857	0.000	0.124	0.130
GA ~~						
RS	0.007	0.003	2.259	0.024	0.467	0.467

Intercepts:

	Estimate	Std.Err	z-value	P(>|z|)	Std.lv	Std.all
.Item_1	0.000				0.000	0.000
.Item_2	0.000				0.000	0.000
.Item_3	0.000				0.000	0.000
.Item_4	0.000				0.000	0.000
.Item_5	0.000				0.000	0.000
.Item_6	0.000				0.000	0.000
.Item_8	0.000				0.000	0.000
.Item_9	0.000				0.000	0.000
.Item_10	0.000				0.000	0.000
.Item_11	0.000				0.000	0.000
.Item_12	0.000				0.000	0.000
.Item_14	0.000				0.000	0.000
.Item_15	0.000				0.000	0.000
.Item_16	0.000				0.000	0.000
.Item_17	0.000				0.000	0.000
.Item_18	0.000				0.000	0.000
.Item_19	0.000				0.000	0.000
.Item_20	0.000				0.000	0.000
.Item_21	0.000				0.000	0.000
.Item_22	0.000				0.000	0.000
.Item_23	0.000				0.000	0.000
.Item_24	0.000				0.000	0.000
.Item_25	0.000				0.000	0.000
.Item_26	0.000				0.000	0.000
.Item_27	0.000				0.000	0.000
.Item_28	0.000				0.000	0.000
.Item_29	0.000				0.000	0.000
.Item_30	0.000				0.000	0.000

.Item_31	0.000	0.000	0.000
.Item_32	0.000	0.000	0.000
.Item_33	0.000	0.000	0.000
.Item_34	0.000	0.000	0.000
.Item_35	0.000	0.000	0.000
.Item_36	0.000	0.000	0.000
.Item_37	0.000	0.000	0.000
.Item_38	0.000	0.000	0.000
.Item_39	0.000	0.000	0.000
.Item_40	0.000	0.000	0.000
.Item_41	0.000	0.000	0.000
.Item_42	0.000	0.000	0.000
.Item_43	0.000	0.000	0.000
.Item_44	0.000	0.000	0.000
.Item_45	0.000	0.000	0.000
.Item_46	0.000	0.000	0.000
.Item_47	0.000	0.000	0.000
.Item_48	0.000	0.000	0.000
.Item_49	0.000	0.000	0.000
.Item_50	0.000	0.000	0.000
.Item_51	0.000	0.000	0.000
.Item_52	0.000	0.000	0.000
.Item_53	0.000	0.000	0.000
.Item_54	0.000	0.000	0.000
.Item_55	0.000	0.000	0.000
.Item_56	0.000	0.000	0.000
.Item_57	0.000	0.000	0.000
.Item_58	0.000	0.000	0.000
.Item_59	0.000	0.000	0.000
.Item_60	0.000	0.000	0.000

GA	0.000				0.000	0.000
RS	0.000				0.000	0.000
PS	0.000				0.000	0.000
CC	0.000				0.000	0.000

Thresholds:

	Estimate	Std.Err	z-value	P(>|z|)	Std.lv	Std.all
Item_1|t1	-0.147	0.025	-5.991	0.000	-0.147	-0.147
Item_2|t1	-1.169	0.032	-36.946	0.000	-1.169	-1.169
Item_3|t1	0.266	0.025	10.744	0.000	0.266	0.266
Item_4|t1	-0.641	0.026	-24.285	0.000	-0.641	-0.641
Item_5|t1	0.326	0.025	13.076	0.000	0.326	0.326
Item_6|t1	1.002	0.030	33.951	0.000	1.002	1.002
Item_8|t1	0.045	0.024	1.855	0.064	0.045	0.045
Item_9|t1	0.582	0.026	22.354	0.000	0.582	0.582
Item_10|t1	0.091	0.025	3.728	0.000	0.091	0.091
Item_11|t1	0.942	0.029	32.639	0.000	0.942	0.942
Item_12|t1	-0.689	0.027	-25.786	0.000	-0.689	-0.689
Item_14|t1	-0.752	0.027	-27.677	0.000	-0.752	-0.752
Item_15|t1	-0.257	0.025	-10.355	0.000	-0.257	-0.257
Item_16|t1	-0.237	0.025	-9.577	0.000	-0.237	-0.237
Item_17|t1	-0.086	0.025	-3.494	0.000	-0.086	-0.086
Item_18|t1	-0.711	0.027	-26.457	0.000	-0.711	-0.711
Item_19|t1	-0.255	0.025	-10.277	0.000	-0.255	-0.255
Item_20|t1	0.649	0.026	24.548	0.000	0.649	0.649
Item_21|t1	0.743	0.027	27.420	0.000	0.743	0.743
Item_22|t1	-1.206	0.032	-37.468	0.000	-1.206	-1.206
Item_23|t1	0.270	0.025	10.900	0.000	0.270	0.270
Item_24|t1	0.666	0.027	25.074	0.000	0.666	0.666
Item_25|t1	-0.491	0.026	-19.181	0.000	-0.491	-0.491
Item_26|t1	0.445	0.025	17.526	0.000	0.445	0.445

Item_27|t1	0.042	0.024	1.698	0.089	0.042	0.042
Item_28|t1	-0.022	0.024	-0.918	0.359	-0.022	-0.022
Item_29|t1	-0.653	0.026	-24.661	0.000	-0.653	-0.653
Item_30|t1	0.221	0.025	8.954	0.000	0.221	0.221
Item_31|t1	0.029	0.024	1.191	0.234	0.029	0.029
Item_32|t1	-0.667	0.027	-25.112	0.000	-0.667	-0.667
Item_33|t1	-0.066	0.024	-2.713	0.007	-0.066	-0.066
Item_34|t1	0.382	0.025	15.208	0.000	0.382	0.382
Item_35|t1	-0.077	0.025	-3.143	0.002	-0.077	-0.077
Item_36|t1	0.451	0.025	17.757	0.000	0.451	0.451
Item_37|t1	-0.191	0.025	-7.746	0.000	-0.191	-0.191
Item_38|t1	-0.053	0.024	-2.167	0.030	-0.053	-0.053
Item_39|t1	-0.073	0.025	-2.987	0.003	-0.073	-0.073
Item_40|t1	-0.029	0.024	-1.191	0.234	-0.029	-0.029
Item_41|t1	0.472	0.025	18.527	0.000	0.472	0.472
Item_42|t1	0.449	0.025	17.680	0.000	0.449	0.449
Item_43|t1	-0.171	0.025	-6.966	0.000	-0.171	-0.171
Item_44|t1	0.465	0.025	18.258	0.000	0.465	0.465
Item_45|t1	-0.204	0.025	-8.291	0.000	-0.204	-0.204
Item_46|t1	0.042	0.024	1.698	0.089	0.042	0.042
Item_47|t1	0.345	0.025	13.774	0.000	0.345	0.345
Item_48|t1	-0.505	0.026	-19.680	0.000	-0.505	-0.505
Item_49|t1	-0.184	0.025	-7.473	0.000	-0.184	-0.184
Item_50|t1	0.550	0.026	21.249	0.000	0.550	0.550
Item_51|t1	0.287	0.025	11.561	0.000	0.287	0.287
Item_52|t1	-1.389	0.035	-39.327	0.000	-1.389	-1.389
Item_53|t1	-1.826	0.047	-38.899	0.000	-1.826	-1.826
Item_54|t1	-0.035	0.024	-1.425	0.154	-0.035	-0.035
Item_55|t1	-0.618	0.026	-23.530	0.000	-0.618	-0.618
Item_56|t1	-1.392	0.035	-39.343	0.000	-1.392	-1.392

Item_57|t1	-1.389	0.035	-39.327	0.000	-1.389	-1.389
Item_58|t1	-1.559	0.039	-39.936	0.000	-1.559	-1.559
Item_59|t1	0.533	0.026	20.676	0.000	0.533	0.533
Item_60|t1	0.483	0.026	18.912	0.000	0.483	0.483

Variances:

	Estimate	Std.Err	z-value	P(>|z|)	Std.lv	Std.all
.Item_1	0.992				0.992	0.992
.Item_2	0.687				0.687	0.687
.Item_3	0.976				0.976	0.976
.Item_4	0.852				0.852	0.852
.Item_5	0.720				0.720	0.720
.Item_6	0.946				0.946	0.946
.Item_8	0.913				0.913	0.913
.Item_9	0.976				0.976	0.976
.Item_10	0.914				0.914	0.914
.Item_11	0.927				0.927	0.927
.Item_12	0.888				0.888	0.888
.Item_14	0.741				0.741	0.741
.Item_15	0.712				0.712	0.712
.Item_16	0.974				0.974	0.974
.Item_17	0.717				0.717	0.717
.Item_18	0.607				0.607	0.607
.Item_19	0.784				0.784	0.784
.Item_20	0.964				0.964	0.964
.Item_21	0.694				0.694	0.694
.Item_22	0.868				0.868	0.868
.Item_23	0.608				0.608	0.608
.Item_24	0.910				0.910	0.910
.Item_25	0.651				0.651	0.651

.Item_26	0.493	0.493	0.493
.Item_27	0.753	0.753	0.753
.Item_28	0.859	0.859	0.859
.Item_29	0.803	0.803	0.803
.Item_30	0.713	0.713	0.713
.Item_31	0.790	0.790	0.790
.Item_32	0.860	0.860	0.860
.Item_33	0.552	0.552	0.552
.Item_34	0.847	0.847	0.847
.Item_35	0.759	0.759	0.759
.Item_36	0.871	0.871	0.871
.Item_37	0.653	0.653	0.653
.Item_38	0.801	0.801	0.801
.Item_39	0.649	0.649	0.649
.Item_40	0.782	0.782	0.782
.Item_41	0.969	0.969	0.969
.Item_42	0.969	0.969	0.969
.Item_43	0.738	0.738	0.738
.Item_44	0.939	0.939	0.939
.Item_45	0.955	0.955	0.955
.Item_46	0.849	0.849	0.849
.Item_47	0.766	0.766	0.766
.Item_48	0.889	0.889	0.889
.Item_49	0.867	0.867	0.867
.Item_50	0.823	0.823	0.823
.Item_51	0.981	0.981	0.981
.Item_52	0.899	0.899	0.899
.Item_53	0.861	0.861	0.861
.Item_54	0.603	0.603	0.603
.Item_55	0.889	0.889	0.889

.Item_56	0.855				0.855	0.855
.Item_57	0.910				0.910	0.910
.Item_58	0.858				0.858	0.858
.Item_59	0.970				0.970	0.970
.Item_60	0.939				0.939	0.939
GA	0.008	0.007	1.222	0.222	1.000	1.000
RS	0.026	0.009	3.029	0.002	1.000	1.000
PS	0.210	0.021	9.963	0.000	1.000	1.000
CC	0.151	0.021	7.219	0.000	1.000	1.000

Scales y*:

	Estimate	Std.Err	z-value	P(>|z|)	Std.lv	Std.all
Item_1	1.000				1.000	1.000
Item_2	1.000				1.000	1.000
Item_3	1.000				1.000	1.000
Item_4	1.000				1.000	1.000
Item_5	1.000				1.000	1.000
Item_6	1.000				1.000	1.000
Item_8	1.000				1.000	1.000
Item_9	1.000				1.000	1.000
Item_10	1.000				1.000	1.000
Item_11	1.000				1.000	1.000
Item_12	1.000				1.000	1.000
Item_14	1.000				1.000	1.000
Item_15	1.000				1.000	1.000
Item_16	1.000				1.000	1.000
Item_17	1.000				1.000	1.000
Item_18	1.000				1.000	1.000
Item_19	1.000				1.000	1.000
Item_20	1.000				1.000	1.000

Item_21	1.000	1.000	1.000
Item_22	1.000	1.000	1.000
Item_23	1.000	1.000	1.000
Item_24	1.000	1.000	1.000
Item_25	1.000	1.000	1.000
Item_26	1.000	1.000	1.000
Item_27	1.000	1.000	1.000
Item_28	1.000	1.000	1.000
Item_29	1.000	1.000	1.000
Item_30	1.000	1.000	1.000
Item_31	1.000	1.000	1.000
Item_32	1.000	1.000	1.000
Item_33	1.000	1.000	1.000
Item_34	1.000	1.000	1.000
Item_35	1.000	1.000	1.000
Item_36	1.000	1.000	1.000
Item_37	1.000	1.000	1.000
Item_38	1.000	1.000	1.000
Item_39	1.000	1.000	1.000
Item_40	1.000	1.000	1.000
Item_41	1.000	1.000	1.000
Item_42	1.000	1.000	1.000
Item_43	1.000	1.000	1.000
Item_44	1.000	1.000	1.000
Item_45	1.000	1.000	1.000
Item_46	1.000	1.000	1.000
Item_47	1.000	1.000	1.000
Item_48	1.000	1.000	1.000
Item_49	1.000	1.000	1.000
Item_50	1.000	1.000	1.000

Item_51	1.000	1.000	1.000
Item_52	1.000	1.000	1.000
Item_53	1.000	1.000	1.000
Item_54	1.000	1.000	1.000
Item_55	1.000	1.000	1.000
Item_56	1.000	1.000	1.000
Item_57	1.000	1.000	1.000
Item_58	1.000	1.000	1.000
Item_59	1.000	1.000	1.000
Item_60	1.000	1.000	1.000

Fit Measures

npar	139	rfi	0.934	rmsea.scaled	0.015
fmin	0.455	nfi	0.937	rmsea.ci.lower.scaled	0.014
chisq	2387.781	pnfi	0.891	rmsea.ci.upper.scaled	0.016
df	1572	ifi	0.978	rmsea.pvalue.scaled	1
pvalue	0	rni	0.978	rmr	0.044
chisq.scaled	2459.252	cfi.scaled	0.96	rmr_nomean	0.045
df.scaled	1572	tli.scaled	0.958	srmr	0.045
pvalue.scaled	0	cfi.robust	NA	srmr_bentler	0.044
chisq.scaling.factor	1.171	tli.robust	NA	srmr_bentler_nomean	0.045
baseline.chisq	38076.281	nnfi.scaled	0.958	crmr	0.045
baseline.df	1653	nnfi.robust	NA	crmr_nomean	0.046
baseline.pvalue	0	rfi.scaled	0.891	srmr_mplus	0.044
baseline.chisq.scaled	23757.412	nfi.scaled	0.896	srmr_mplus_nomean	0.045
baseline.df.scaled	1653	ifi.scaled	0.96	cn_05	1829.71
baseline.pvalue.scaled	0	rni.scaled	0.96	cn_01	1873.66

baseline.chisq.scaling.factor	1.648	rmsea	0.014	gfi	0.962
cfi	0.978	rmsea.ci.lower	0.013	agfi	0.959
tli	0.976	rmsea.ci.upper	0.015	pgfi	0.884
nnfi	0.976	rmsea.pvalue	1	mfi	0.856

Appendix C

Fit Statistics of Model 2A and 2B

Measured Variables (Indicators)

Test for Multivariate Normality with respect to Henze-Zirkler's test indicates deviation from multivariate normality (HZ = 2.004, $p < 0.001$)

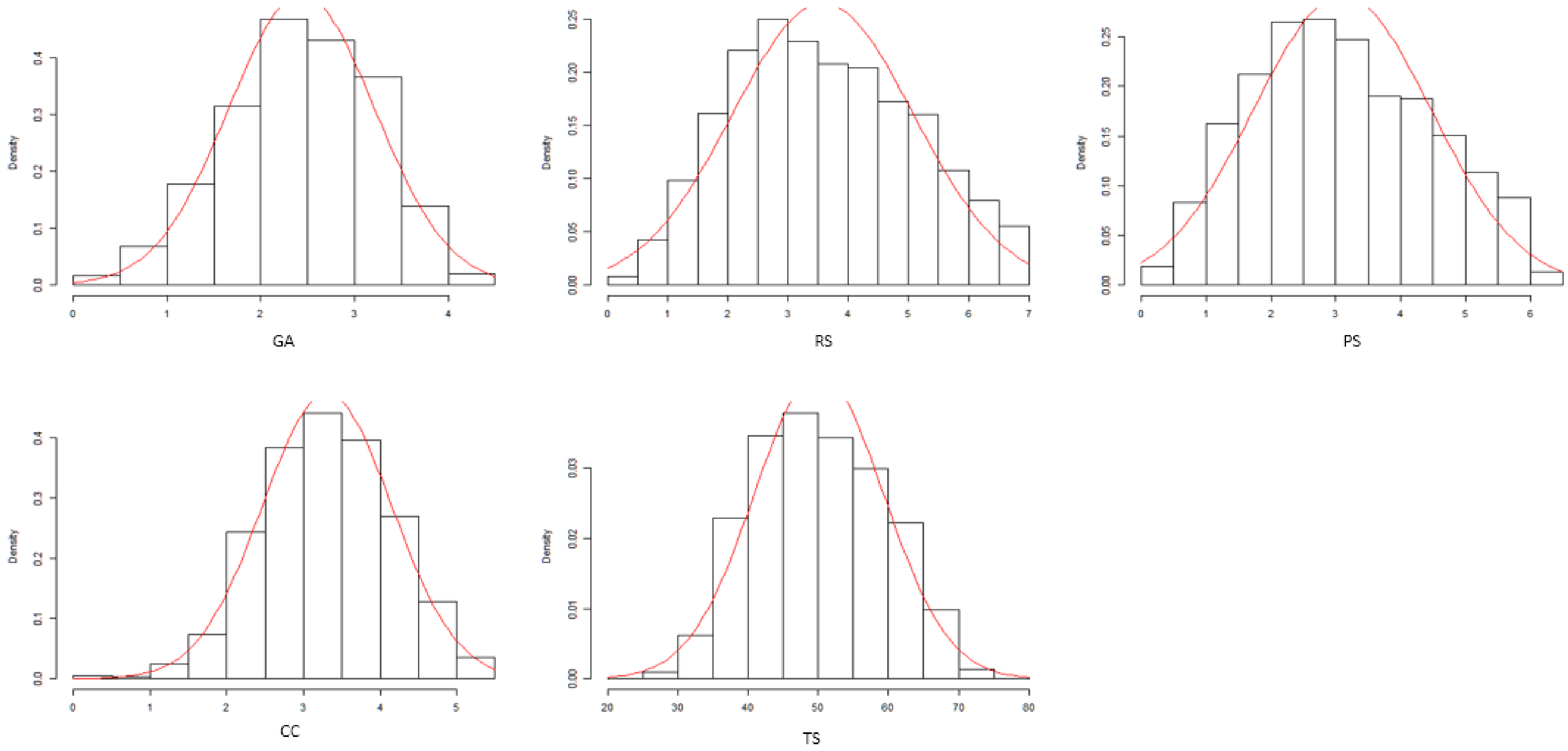
Univariate Normality with respect to Shapiro-Wilk test

Variable	Statistic	*p* Value
GA	0.991	<0.001
RS	0.983	<0.001
PS	0.982	<0.001
CC	0.995	<0.001
TScore	0.990	<0.001

Descriptive Statistics

	n	Mean	Std.Dev	Median	Min	Max	25th	75th	Skewness	Kurtosis
GA	2623	2.435	0.782	2.453	0.000	4.301	1.927	3.044	-0.263	-0.371
RS	2623	3.579	1.505	3.461	0.162	6.973	2.430	4.708	0.203	-0.751
PS	2623	3.101	1.370	2.982	0.000	6.159	2.046	4.127	0.194	-0.791
CC	2623	3.307	0.843	3.317	0.000	5.259	2.710	3.897	-0.131	-0.076
TScore	2623	50.103	9.203	49.676	24.987	79.023	42.924	57.298	0.088	-0.705

Histograms of the variables



Variance Inflation Factors

Variables	VIF
GA	1.163
RS	1.926
PS	2.065
CC	1.537
TS	1.135

Model 2A

Model Summary

Lavaan 0.6-3 ended normally after 135 iterations

Optimization method	NLMINB	
Number of free parameters	10	
Number of observations	2623	
Estimator	ML	Robust
Model Fit Test Statistic	122.670	123.673
Degrees of freedom	5	5
P-value (Chi-square)	0.000	0.000

Scaling correction factor for the Satorra-Bentler correction		0.992
Model test baseline model:		
Minimum Function Test Statistic	3249.088	3081.268
Degrees of freedom	10	10
P-value	0.000	0.000
ser model versus baseline model:		
Comparative Fit Index (CFI)	0.964	0.961
Tucker-Lewis Index (TLI)	0.927	0.923
Robust Comparative Fit Index (CFI)		0.964
Robust Tucker-Lewis Index (TLI)		0.927
Loglikelihood and Information Criteria:		
Loglikelihood user model (H0)	-23670.315	- 23670.315
Loglikelihood unrestricted model (H1)	-23608.980	- 23608.980
Number of free parameters	10	10
Akaike (AIC)	47360.630	47360.630
Bayesian (BIC)	47419.351	47419.351
Sample-size adjusted Bayesian (BIC)	47387.578	47387.578
Root Mean Square Error of Approximation:		

RMSEA		0.095	0.095	
90 Percent Confidence Interval	0.081	0.110	0.081	0.110
P-value RMSEA <= 0.05		0.000	0.000	
Robust RMSEA			0.095	
90 Percent Confidence Interval			0.081	0.110

Standardized Root Mean Square Residual:

SRMR	0.041	0.041

Parameter Estimates:

Information	Expected
Information saturated (h1) model	Structured
Standard Errors	Robust.sem

Latent Variables:

	Estimate	Std.Err	z-value	P(>|z|)	Std.lv	Std.all
GCA =~						
GA	1.000				0.294	0.376
RS	4.026	0.223	18.031	0.000	1.183	0.786
PS	3.912	0.219	17.856	0.000	1.150	0.839
CC	1.834	0.101	18.208	0.000	0.539	0.639
AAL =~						
TS	1.000				9.201	1.000

Regressions:

	Estimate	Std.Err	z-value	P(>|z|)	Std.lv	Std.all
AAL ~						
GCA	10.644	0.775	13.741	0.000	0.340	0.340

Variances:

	Estimate	Std.Err	z-value	P(>|z|)	Std.lv	Std.all
.GA	0.525	0.014	36.359	0.000	0.525	0.859
.RS	0.864	0.043	20.254	0.000	0.864	0.382
.PS	0.555	0.035	16.011	0.000	0.555	0.296
.CC	0.421	0.015	27.330	0.000	0.421	0.592
.TS	0.000				0.000	0.000
.GCA	0.086	0.009	9.206	0.000	1.000	1.000
.AAL	74.875	1.657	45.190	0.000	0.884	0.884

Fit Measures

npar	10	rni	0.964	rmsea.ci.lower.scaled	0.081
fmin	0.023	cfi.scaled	0.961	rmsea.ci.upper.scaled	0.11
chisq	122.67	tli.scaled	0.923	rmsea.pvalue.scaled	0
df	5	cfi.robust	0.964	rmsea.robust	0.095
pvalue	0	tli.robust	0.927	rmsea.ci.lower.robust	0.081
chisq.scaled	123.673	nnfi.scaled	0.923	rmsea.ci.upper.robust	0.11
df.scaled	5	nnfi.robust	0.927	rmsea.pvalue.robust	NA
pvalue.scaled	0	rfi.scaled	0.92	rmr	0.27
chisq.scaling.factor	0.992	nfi.scaled	0.96	rmr_nomean	0.27
baseline.chisq	3249.088	ifi.scaled	0.961	srmr	0.041

baseline.df	10	rni.scaled	0.961	srmr_bentler	0.041
baseline.pvalue	0	rni.robust	0.964	srmr_bentler_nomean	0.041
baseline.chisq.scaled	3081.268	logl	-23670.315	crmr	0.05
baseline.df.scaled	10	unrestricted.logl	-23608.98	crmr_nomean	0.05
baseline.pvalue.scaled	0	aic	47360.63	srmr_mplus	0.041
baseline.chisq.scaling.factor	1.054	bic	47419.351	srmr_mplus_nomean	0.041
cfi	0.964	ntotal	2623	cn_05	237.715
tli	0.927	bic2	47387.578	cn_01	323.582
nnfi	0.927	rmsea	0.095	gfi	0.98
rfi	0.924	rmsea.ci.lower	0.081	agfi	0.94
nfi	0.962	rmsea.ci.upper	0.11	pgfi	0.327
pnfi	0.481	rmsea.pvalue	0	mfi	0.978
ifi	0.964	rmsea.scaled	0.095	ecvi	0.054

Coefficients of Determination (R^2)

GA	RS	PS	CC	TS	AAL
0.141	0.618	0.704	0.408	1.000	0.116

Model 2B

Model Summary

lavaan 0.6-3 ended normally after 145 iterations

Optimization method	NLMINB	
Number of free parameters	12	
Number of observations	2623	
Estimator	ML	Robust
Model Fit Test Statistic	14.328	14.178
Degrees of freedom	3	3
P-value (Chi-square)	0.002	0.003
Scaling correction factor		1.011
for the Satorra-Bentler correction		

Model test baseline model:

Minimum Function Test Statistic	3249.088	3081.268
Degrees of freedom	10	10
P-value	0.000	0.000

User model versus baseline model:

Comparative Fit Index (CFI)	0.997	0.996

Tucker-Lewis Index (TLI)		0.988	0.988	
Robust Comparative Fit Index (CFI)			0.997	
Robust Tucker-Lewis Index (TLI)			0.988	
Loglikelihood and Information Criteria:				
Loglikelihood user model (H0)		-23616.144	-23616.144	
Loglikelihood unrestricted model (H1)		-23608.980	-23608.980	
Number of free parameters		12	12	
Akaike (AIC)		47256.288	47256.288	
Bayesian (BIC)		47326.753	47326.753	
Sample-size adjusted Bayesian (BIC)		47288.626	47288.626	
Root Mean Square Error of Approximation:				
RMSEA		0.038	0.038	
90 Percent Confidence Interval	0.02	0.059	0.02	0.058
P-value RMSEA <= 0.05		0.815	0.822	
Robust RMSEA			0.038	
90 Percent Confidence Interval			0.02	0.059
Standardized Root Mean Square Residual:				
SRMR		0.013	0.013	

Parameter Estimates:

Information	Expected
Information saturated (h1) model	Structured
Standard Errors	Robust.sem

Latent Variables:

	Estimate	Std.Err	z-value	P(>|z|)	Std.lv	Std.all
GCA =~						
GA	1.000				0.361	0.462
RSPA	2.527	0.194	13.053	0.000	0.772	0.772
CC	1.847	0.102	18.072	0.000	0.667	0.790
RSPA =~						
RS	1.000				1.182	0.786
PS	0.995	0.029	33.836	0.000	1.176	0.858
AAL =~						
TS	1.000				9.201	1.000

Regressions:

	Estimate	Std.Err	z-value	P(>|z|)	Std.lv	Std.all
AAL ~						
GCA	10.020	0.793	12.627	0.000	0.393	0.393

Covariances:

	Estimate	Std.Err	z-value	P(>|z|)	Std.lv	Std.all
.GA ~~						
.CC	-0.031	0.016	-1.893	0.058	-0.031	-0.087

Variances:

	Estimate	Std.Err	z-value	P(>|z|)	Std.lv	Std.all
.GA	0.481	0.017	27.842	0.000	0.481	0.787
.CC	0.267	0.028	9.377	0.000	0.267	0.375
.RS	0.866	0.044	19.666	0.000	0.866	0.383
.PS	0.494	0.037	13.275	0.000	0.494	0.263
.TS	0.000				0.000	0.000
GCA	0.130	0.015	8.654	0.000	1.000	1.000
RSPS	0.565	0.056	10.047	0.000	0.405	0.405
.AAL	71.576	1.703	42.028	0.000	0.845	0.845

Fit Measures

npar	12	rni	0.997	rmsea.ci.lower.scaled	0.02
fmin	0.003	cfi.scaled	0.996	rmsea.ci.upper.scaled	0.058
chisq	14.328	tli.scaled	0.988	rmsea.pvalue.scaled	0.822
df	3	cfi.robust	0.997	rmsea.robust	0.038
pvalue	0.002	tli.robust	0.988	rmsea.ci.lower.robust	0.02
chisq.scaled	14.178	nnfi.scaled	0.988	rmsea.ci.upper.robust	0.059
df.scaled	3	nnfi.robust	0.988	rmsea.pvalue.robust	NA
pvalue.scaled	0.003	rfi.scaled	0.985	rmr	0.09
chisq.scaling.factor	1.011	nfi.scaled	0.995	rmr_nomean	0.09
baseline.chisq	3249.088	ifi.scaled	0.996	srmr	0.013
baseline.df	10	rni.scaled	0.996	srmr_bentler	0.013
baseline.pvalue	0	rni.robust	0.997	srmr_bentler_nomean	0.013
baseline.chisq.scaled	3081.268	logl	-23616.144	crmr	0.016
baseline.df.scaled	10	unrestricted.logl	-23608.98	crmr_nomean	0.016
baseline.pvalue.scaled	0	aic	47256.288	srmr_mplus	0.013

baseline.chisq.scaling.factor	1.054	bic	47326.753	srmr_mplus_nomean	0.013
cfi	0.997	ntotal	2623	cn_05	1431.6
tli	0.988	bic2	47288.626	cn_01	2077.8
nnfi	0.988	rmsea	0.038	gfi	0.998
rfi	0.985	rmsea.ci.lower	0.02	agfi	0.989
nfi	0.996	rmsea.ci.upper	0.059	pgfi	0.2
pnfi	0.299	rmsea.pvalue	0.815	mfi	0.998
ifi	0.997	rmsea.scaled	0.038	ecvi	0.015

Coefficient of Determination (R^2)

GA	CC	RS	PS	TS	RSPS	AAL
0.213	0.625	0.617	0.737	1.000	0.595	0.155

Appendix D

Test Paper of the CGT in 2017

The test paper of the Common General Test - 2017 has been removed due to copyright restrictions