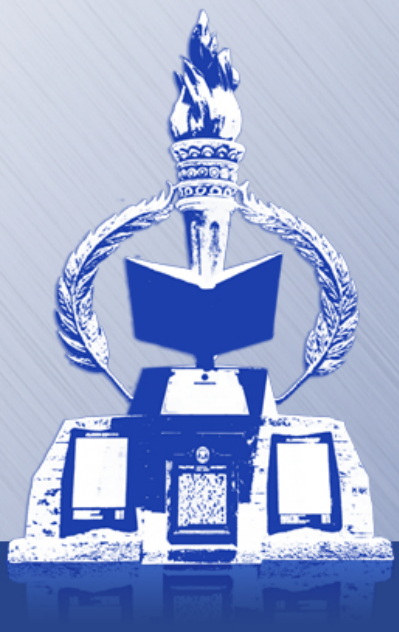
May 2018

Reducing Cognitive Load of Mathematics Test Items Effects on the Performance of High and Low Ability Students

Steve B. Anapi
Philippine Normal University



REDUCING COGNITIVE LOAD OF MATHEMATICS TEST ITEMS: EFFECTS ON THE PERFORMANCE OF HIGH AND LOW ABILITY STUDENTS

A Thesis
Presented to the
College of Graduate Studies
and Teacher Education Research
Philippine Normal University
Manila

In Partial Fulfillment of the Requirements
for the Degree Master of Arts in Education
with Specialization in Mathematics Education

by

STEVE B. ANAPI

May 2018

APPROVAL SHEET

This thesis entitled **"REDUCING COGNITIVE LOAD OF MATHEMATICS TEST ITEMS: EFFECTS ON THE PERFORMANCE OF HIGH AND LOW ABILITY STUDENTS"** prepared and submitted by **STEVE B. ANAPI** in partial fulfillment of the requirements for the degree **Master of Arts in Education with Specialization in *Mathematics Education***, is hereby recommended for approval and acceptance.

WILMER G. MARQUEZ, MA Ed.
Adviser

Approved in partial fulfillment of the requirement for the degree **Master of Arts in Education** by the Oral Examination Committee.

ADONIS P. DAVID, Ph.D.
Chairperson

MYLA E. DETECIO, MST
Member

ROBERT D. DIAZ, Ph.D.
Member

RENE R. BELECINA, Ph.D.
Member

Accepted in partial fulfillment of the requirements for the degree **Master of Arts in Education.**

_____
Date

ANTRIMAN V. ORLEANS, Ph.D.
Officer in Charge
College of Graduate Studies and
Teacher Education Research

ACKNOWLEDGMENTS

"*Sometimes our light goes out but is blown into a flame by another human being. Each of us owes deepest thanks to those who have rekindled this light.*"

- Albert Schweitzer

The researcher would like to extend his sincere and deepest gratitude to those significant people who have rekindled his spirit and who have significantly made this thesis possible. Gratitude goes to:

WILMER G. MARQUEZ, Mathematics professor of Philippine Normal University and the researcher's adviser, for his guidance, valuable insights, and wisdom;

DR. ADONIS P. DAVID, chair of the Oral Examination Committee for his sound judgment that directed the overall framework of this study;

DR. ROBERT DIAZ, PROF. MYLA DETECIO, and **DR. RENE BELECINA** research panel, for their valuable insights and suggestions in the study;

FAMILY members, who have never doubted the researcher's skill and potential, and have shown unconditional love, support, and understanding;

GEANHS Mathematics Department, for their encouragement to finish what the researcher has started;

FRIENDS, who never failed to give spiritual and moral support, for the joy you have given me in spite of the hard comings in completing this endeavor. They have been as always, thieves of time;

VIZALIE PORTO, a dear friend and companion, for the encouragement and support;

THE ALMIGHTY, most especially, for His never ending guidance and love for which the researcher had experienced continuous renewal of strength and spirit, and for the humble success of this undertaking for which it is rightfully returned in His Glory and Honor.

S.B.A

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

ABSTRACT

Name	: Steve B. Anapi
Title	: Reducing Cognitive Load of Mathematics Test Items: Effects on the Performance of High and Low Ability Students
Key Concepts	: Assessment, Cognitive Load Theory, Reduced Cognitive Load Test Forms, Mathematics Performance
Degree	: Master of Arts in Education
Specialization	: Mathematics Education
Adviser	: Wilmer G. Marquez, MA Ed.

This study aimed to determine the effect of reducing cognitive load of test items on the Mathematics performance. Particularly, it sought to determine how varying the test formats (i.e. traditional, modified item *stem* with essential information, and item stem with solution strategies as *alternatives/ options*) in a multiple-choice type test affects the performance of high and low ability high school students. Three Mathematics test instruments (traditional and two reduced cognitive load test forms) on operations on integers and polynomials were developed and validated. Psychometric properties of the test instruments were reported. The study employed a 2 x 3 between-subjects factorial design to determine the effect of reducing cognitive load of test items on the performance of 120 purposive samples of Grade 9 students. Two factors were

considered affecting Mathematics performance: ability level and test format. Results indicated that both factors significantly contributed to the performance of students. As to ability level, high- achieving students outperformed low- achieving students. As to the test format, post hoc analysis revealed that students who were randomly assigned to two reduced cognitive load test forms performed better than students who were assigned to traditional assessment. There was no significant difference on the performance of students on modified item stem condition and modified alternatives condition. No reported interaction effect on ability level and test format. In terms of preference, high- achieving students preferred test format with modified item stem. No outstanding test preference for low- achieving students. Implications for classroom assessment and future research work are discussed.

CHAPTER 1

THE PROBLEM AND LITERATURE REVIEW

Background of the Study

Mathematics is both an interesting and a very challenging subject. It has been considered as one of the most essential core subjects in school curriculum. Mathematics instruction and assessment play an influential and pivotal role in the Mathematics performance of students. Numerous studies conducted elsewhere have identified factors affecting Mathematics performance: individual (student), instructional (teacher), classroom management, and evaluation (Andaya, 2014). Similarly, Suan (2014) found that there was a significant relationship between student factors and academic performance in mathematics. These individual/student factors relate mainly on learner's affective and cognitive domain.

Individual factors together with self-directed learning, mathematics ability of students, attitude towards Mathematics, motivation of students, time spent in Mathematics lesson, student's interest, anxiety, reasoning ability, problem solving skill, and Mathematics phobia were related to Mathematics performance (Andaya, 2014; Nizoloman, 2013). Among the individual/ student factors that influence Mathematics performance, attitude towards Mathematics manifested significant influence to academic performance (Alpacion, Camaňan, Gregorio, Panlaan & Tudy, 2014; Suan, 2014). Aside from student factors, instructional factors influence academic performance in Mathematics. This means that teachers play a major role and consequently have

major effect on student's Mathematics performance (Andaya, 2014). She further goes on tackling important factors in mathematics instruction and mathematics performance which include curriculum, instructional strategies, method of teaching, math teacher's competency, school context, and facilities.

The problem of low performance in Mathematics is a major concern not only in the Philippine education but also across all countries. Considerably given much attention in today's educational setting are those school children receiving instruction but are not learning the basic numeracy skills, who are critically at-risk of not learning mathematics, are performing poorly than their typical peers (Kroesbergen & Van Luit, 2003), and are persistently having difficulty learning mathematics across the course of their school careers. Categorically, this group of children is considered low- achieving or at-risk in mathematics. Kroesbergen and Van Luit (2003) asserted that in almost every classroom, students at-risk of mathematics difficulty can be found. It is predicted that about 5% to almost 10% of school-age population have or experience mathematics difficulty (Fuchs & Fuchs, 2005; Fuchs, Compton, Fuchs, Paulsen, Bryant & Hamlett, 2005; Geary, 2004).

Studies show that students at-risk of mathematics difficulty suffer from specific cognitive deficits (i.e. executive function) like deficits in working memory, information retrieval (basic fact retrieval), inhibition, and attention regulation (Clements, Sarama & Germeroth, 2016; Compton, Fuchs, Fuchs, Lambert & Hamlett, 2012; Fuchs & Fuchs, 2005; Fuchs et al., 2010, 2014; Geary, 2004, 2010, 2013; Geary, Hoard, Byrd-Craven, Nugent & Numtee, 2007; Murphy, Mazzocco, Hanich & Early, 2007; Swanson &

Sachse-Lee, 2001; Swanson, 2016; Swanson, Jerman & Zheng, 2008). Aside from cognitive deficits, mathematical domain-specific deficits are being looked as identifying feature of students at-risk. Because of the underlying cognitive deficits experienced by this group, they suffer learning problems in the different areas/domains, thus showing symptomatic problems both on the procedural and conceptual mathematical knowledge (Lambert & Spinath, 2014). As such, there is a need to alleviate low-achieving students' Mathematics performance to an acceptable level of proficiency. But what other factors must be dealt with in consideration to the characteristics of this group of learners in particular and to increase Mathematics performance of all students in general?

As mentioned, evaluation factors affect Mathematics performance. This factor mainly concerns on how student's knowledge is being assessed and evaluated. It can be inferred that Mathematics assessment may be a contributory factor influencing student's performance in the subject area. With the intent of more valid process of assessing student mathematical knowledge, educators frequently write their own assessments. However, as Gillmor, Poggio and Embretson (2015) contend, these test items are often filled with problems that distract or confuse students. In their review of existing literature, it has shown that construct-irrelevant factors such as complexity of language and item format can hinder student performance on assessments. With that in mind, the study sought to address the influence of Mathematics assessment to Mathematics performance by looking into the cognitive demands of test items that interfere with students' test performance.

With Cognitive Load Theory's (CLT) emerging application in assessment or in testing context, the problem of Mathematics performance can be addressed by modifying features or aspects of test formats by looking into the cognitive loads demanded by test items. In general, the purpose of this study is to add to the existing and quite few emerging studies on the applicability of CLT to educational measurement with the specific intent of reducing cognitive load of Mathematics test items so students can use their available cognitive resources more efficiently for problem solving and in turn, improve their Mathematics performance.

To reduce cognitive loads in Mathematics test items, modifications to features of test were implemented for students to be able to demonstrate their knowledge of the target mathematics construct. Modifications in full sense are alterations made to a test or aspects of items (item format) that make a test more accessible for most learners while maintaining to assess grade-level content and skills at the same depth of knowledge as original items. Not only is the problem of increasing the academic performance in Mathematics being addressed but also the concept of access to assessment is hereby looked into. According to Elliott, Kettler, Beddow and Kurz (2011, p.319)

> "Access means more than participation in general education classes and end-of-year achievement tests. Real access involves the opportunity for all students to learn the intended curriculum and demonstrate what they have achieved as a result of their learning."

In line with the concept of access, the study aimed to explore the effect of reducing the cognitive load of test items on the Mathematics performance of high and low ability high school students. Although focused on the testing context, the study acknowledges the interrelation of assessment, instruction and learning. Assessment provides feedback as to how learning progresses and it informs the content and method of teaching. Assessment and instruction both inform each other of the learning outcome in terms of students' achievement or performance. Assessment helps teachers to put their efforts in a context where their learners have been before and where they are heading (Daro, Mosher & Corcoran, 2011). Hence, this study may provide productive information both to mathematics assessment and pedagogy on increasing mathematics performance of students.

Literature Review

This section examines the various findings of studies on low- achieving students in mathematics and the mathematics interventions for this group of learners in terms of instructional approaches/ designs, duration, domain, and grade level. Also, a discussion on studies and literature on mathematics education in the framework of Cognitive Load Theory is presented. Applications of Cognitive Load Theory in testing context are emphasized. Influence of Mathematics assessment to Mathematics performance through cognitive load framework and possibly addressing low performance of at-risk students in mathematics and performance of students in general is likewise discussed.

Low-achieving students in mathematics/ students at-risk of mathematics difficulty. This study, drawing from the Australian context (van Kraayenoord & Elkins, 2004, as cited in Hopkins & Egeberg, 2009) prefers to use the term *mathematics difficulty* rather than *mathematics disability*. Students at-risk of mathematics difficulty or low- achieving students in mathematics are generally recognized using professional judgments based on low achievement; whereas the term mathematics disability denotes a more severe type of learning difficulty. Kroesbergen and Van Luit (2003, p. 99) defined this group as those students 'who have more trouble with learning math than their peers, who perform at a lower level than their peers, or who need special instruction to perform at an adequate level'. Theoretically, students at-risk of mathematics difficulty are categorized when they exhibit inability 'to represent or process information in one or all of the mathematical domains (e.g. geometry) or in one or a set of individual competencies within each domain' (Geary, 2004, p. 4).

The ability to represent or process information as Geary claims, agrees with Crealock and Kronick (1993) on areas of functioning, which, as Morgan, Farkas, Hillemeier and Maczuga (2014) mentioned, lead to underlying cognitive impairments. This group of students suffers from several cognitive deficits like deficits in working memory, information retrieval, inhibition, and attention regulation. *Working memory* defined as cognitive processing resource of limited capacity involved in the presentation of information while simultaneously processing other information (Paas, Renkl & Sweller, 2003), is a key factor in problem solving (Swanson, 2016) and that it underlies mathematics difficulties (Geary, 2010, 2013; Swanson et al., 2008) because

at-risk students have problems in *maintaining* information in working memory (Murphy et al. 2007), have impairment in symbolic processing and visual-spatial reasoning (Brendefur et al. 2015), and have deficits in attention selection making irrelevant information to enter working memory thus occupying its capacity essential for able processing (Swanson, 2016).

Large body of literature shows that working memory plays a critical role and is an important variable both in intellectual ability and in academic performance (Musso, Kyndt, Cascallar & Dochy, 2012). The role of working memory is on the repetition of the information and processing that information systematically (Alloway & Copello, 2013). Hence, Swanson (2016) contended that deficit in working memory is a key feature for children who suffer from math difficulties since mathematical cognition involves complex mechanism such as identification of important quantities, encoding into an internal representation, mental comparisons, and calculations (Musso et al., 2012). As reviewed by Alloway and Copello (2013), there is a link between mathematics outcomes and working memory such that low working memory scores are correlated to poor performance on arithmetic word problems and poor computational skills.

Aside from deficit in working memory, low achieving students in mathematics are also characterized by their problems in information retrieval (Clements et al., 2016; Compton et al., 2012; Fuchs & Fuchs, 2005; Fuchs et al., 2010, 2014; Geary, 2004; Geary et al., 2007; Swanson & Sachse-Lee, 2001). The inability to use retrieval-based processes in solving simple arithmetic and simple word problems is the most consistent finding among children with mathematics difficulty and those children that have co-

morbid cases (both have math and reading difficulties) (Swanson, 2016). When they retrieve arithmetic facts from long term memory, they usually commit more errors (lack of inhibition) and show error and reaction time patterns (Geary, 2004). Thus, struggling students in mathematics have poor inhibition or irrelevant associations when doing mathematical task, have disrupted functions on working memory and problems in attention control (Swanson, 2016).

Underlying cognitive impairments of students at-risk of mathematics difficulty also present problem in learning different areas/ domains of mathematics. Most of literature and studies include number combination (NC) as signature deficits of at-risk students in the elementary age , in particular, the deficit in automatic retrieval of number combinations, of which, students at-risk continue to rely on counting to perform simple (single-digit) addition (Bryant, Bryant & Hammill, 2000; Geary & Hoard, 2005; Geary, 2004; Geary, Hanson & Hoard, 2000; Geary et al., 2007; Gersten, Jordan & Flojo, 2005; Hanich, Jordan, Kaplan & Dick, 2001; Hopkins & Egeberg, 2009; Jordan, Hanich, & Kaplan, 2003; Jordan, Kaplan & Hanich, 2002). This impairment of the acquisition of arithmetic skills is due to immature counting strategies/ difficulty with counting that contributed to the difficulties in computational fluency (Bryant, Bryant, Gersten, Scammacca & Chavez, 2008; Fuchs et al., 2010; Geary et al., 2007; Gersten, Chard, Jayanthi, Baker, Morphy & Flojo, 2008).

Deficits in many aspects of basic number sense (i.e. magnitude comparison) may be a good predictor of mathematics difficulty (Fuchs et al. 2005; Gersten et al. 2008; Jordan et al. 2003; Stock, Desoete & Roeyers, 2010). As what have been presented,

difficulties in the area of counting, number sense and number/ arithmetical combinations are identifying features of low- achieving students in mathematics; more so, they suffer great difficulty in solving word-problems (Fuchs et al., 2005). These students 'experienced considerable difficulty with problem representation or identifying relevant information, along with difficulties in reading, computation, and identifying operations' (Jitendra, DiPipi & Perron-Jones, 2002, p. 23). Chinese studies (Zhong et al., 2006; Zhu & Wang, 2014, as cited in Zhu, 2015) reached the same conclusions that elementary students at-risk of mathematics difficulty had limited ability to represent problems and had difficulty identifying effective strategies to use.

Literature and studies identify at-risk students based on cut-off scores on standardized mathematics test applying restrictive/ stringent criteria (10^{th} percentile; 1.5 SD below the mean) to lenient/moderate to high cut-offs (i.e. between 11^{th} and 25^{th}, 16^{th} , 25^{th} and 35^{th} percentile) (Dennis et al., 2016; Fuchs & Fuchs, 2002; Fuchs, Fuchs & Prentice, 2004; Fuchs et al., 2010; Geary & Hoard, 2005; Jitendra et al., 2002; Jordan et al., 2003; Lambert & Spinath, 2014; Murphy et al., 2007; Re, Pedron, Tressoldi & Lucangeli, 2014; Stock et al. 2010). The study conducted by Murphy et al. (2007) argued that different cut-offs (below 10^{th} percentile, between 11^{th} and 25^{th} percentile, and above 25^{th} percentile) show different profiles in mathematics and other related skills (e.g. reading, visual-spatial, and working memory skills), which means that students at-risk of mathematics difficulty vary as function of cut-off scores qualitatively.

Cut-off scores to identify at-risk students in mathematics are used mainly for selection of samples. But other literature suggest otherwise-- that consistent low

performance should also be looked into. Geary (2004) argued that a child with a lower score than was expected on a mathematics achievement test need not become as basis itself that the child has mathematics difficulty. Moreover, he argued that many of children who score low on achievement test subsequently have better scores as they progress. Theoretically, those students who are persistently struggling in mathematics across successive academic years warrant mathematics difficulty because of their underlying cognitive impairments (Geary, 2004, 2011; Mazzocco & Myers, 2003).

Studies on mathematics intervention for low- achieving students/ students at-risk of mathematics difficulty. On a meta-analysis of experimental researches on teaching students at-risk of mathematics difficulty conducted by Dennis et al. (2016), all 25 mathematics intervention studies had an overall positive impact on mathematics performance. Currently, most of the studies and literature focused on mathematics intervention of students at-risk or struggling have been mostly investigated during kindergarten and elementary school grades only (Bryant et al., 2008, 2011; Clarke et al., 2016; Fuchs & Fuchs, 2005; Fuchs et al., 2005, 2006, 2008, 2009, 2010; Hopkins & Egeberg, 2009; Jitendra et al., 2002; Jordan et al., 2003; Lambert & Spinath, 2014; Powell et al., 2009; Swanson, 2016; Zhu, 2015).

All of these studies conducted mathematics interventions that ranged from 8-week up to 18-week duration, 3 to 5 sessions per week, and 10 to 30 minutes per session. Most of the researches focused on number/ arithmetic combinations (addition/subtraction of whole numbers) for the purpose of building procedural fluency and conceptual knowledge. Studies relied greatly on the effects of instructional

approaches/ designs to the mathematics achievement/ performance of students at-risk, which means that most of them are experimental in nature particularly randomized control trial (RCT) trying out instructional approaches/ designs such as explicit systematic instruction, strategic counting instruction with deliberate practice, strategic instruction with guided and independent practice, and computer-assisted instruction.

Literature that focused on problem solving (Fuchs & Fuchs, 2005; Jitendra et al., 2002) used schema theory as guiding principle in designing instruction for low-achieving students in mathematics (e.g. schema-based instruction); while in the study of Zhu (2015), cognitive strategy instruction (CSI) was used to fourth grade at-risk students in mathematics. Interesting studies to note are those conducted by Jordan et al. (2003) and Zhu (2015). Aside from students with mathematics difficulty, they included co-morbid cases, such as students with mathematics and reading difficulties. Both studies concluded that the presence or absence of reading difficulty is a powerful predictor of the stability of a mathematics difficulty.

The research of Swanson (2016) is also an interesting study in this field. Not only did he used an instructional design in building accuracy on problem-solving of third graders with math difficulty but also looked into working memory capacity (WMC) as a variable. Findings of his study revealed that students with relatively higher WMC scored higher on posttest than children with lower WMC, and suggested that strategy instruction more likely benefited students at-risk with higher WMC than students at-risk with lower WMC. Similarly, Fuchs et al. (2014) investigated individual differences in WM as they examined how children's cognitive abilities (WM capacity) moderate the effects

of fraction intervention for at-risk 4th graders. Testing Geary et al.'s (2008) *fluency practice compensation hypothesis*: building automaticity with foundational skills compensating children's WM limitations revealed that students with very weak WM learned better with conceptual activities; but children with more adequate WM learned better with fluency activities. Treatment groups outperformed control group; thus, they have concluded that reducing cognitive load could improve efficacy of interventions that stress the measurement interpretation of fractions. Overall, findings and results of these studies differing in instructional approaches/ designs reported a significant impact on the mathematics performance of struggling students in mathematics.

Few studies have been conducted on high school students (Hopkins & Egeberg, 2009; Montague, Enders & Dietz, 2011; Montague, Krawec, Enders & Dietz, 2014). Randomized studies conducted by Montague and colleagues on grade seven students with mathematics difficulty using cognitive strategy instruction yielded positive effects; while Hopkins and Egeberg (2009) examined the effect of extended practice on reliance on retrieval for simple addition to at-risk Australian students ages 13-14 in mainstream secondary school setting and indicated that extended practice had positive effects on retrieval for simple addition.

All these studies reviewed pointed out that mathematics intervention should use an explicit systematic approach to instruction. Four meta-analysis studies on the mathematics intervention for students at-risk, struggling, or with mathematics learning disabilities (Baker, Gersten & Lee, 2002; Dennis et al., 2016; Gersten et al., 2008; Kroesbergen & Van Luit, 2003) concluded that approaches that used *explicit instruction*

had an overall positive significant effect on the mathematics achievement. **Explicit instruction** is defined as an 'unambiguous and direct teaching that includes both instructional design and delivery procedures' (Archer & Hughes, 2011). Clarke et al. (2016) specify the four essential features of explicit mathematics instruction, which are teacher modeling, deliberate practice, visual representation, and academic feedback. *Teacher modeling* entails overt explanations and demonstrations of specific mathematics content (Archer & Hughes, 2011). *Deliberate practice* is designed for the promotion of high rates of learning success and is also intended for eventual learner independence (Clarke et al., 2016). *Visual representation* is intended for the deepening of learner's conceptual understanding of the mathematical content being taught; while *academic feedback* is designed for immediate and specific error correction. **Systematic instruction**, in particular to mathematics teaching, focused on the logical order of introducing concepts that gradually builds learner proficiency on the specific mathematical concept being introduced (Gersten et al., 2009).

Gersten et al. (2008) synthesized experimental and quasi-experimental researches on instructions that enhanced the mathematics performances of students with learning disabilities and concluded that 'explicit instruction should play a key role in mathematics instruction for students with LD' (p.1) and should be a mainstay feature for teaching students with mathematics difficulty. This finding is also supported by the US National Mathematics Advisory Panel and recommends that struggling students receive explicit instruction (National Mathematics Advisory Panel, 2008).

Mathematics education in cognitive load theory perspective. The following reviewed literature and studies use the cognitive load perspective in their approaches in teaching and learning of different mathematical domains. In the study of Ngu, Yeung and Tobias (2014), they tested three approaches (unitary, pictorial and equation) in teaching percentage change problems to grade eight Australian students from a cognitive load standpoint. Their study found that the *equation approach* generated significantly more correct solutions, fewer non-attempted problems and errors compared to the unitary and pictorial approaches; thus, they have concluded that its superiority 'may have stemmed from the application of instructional design principle associated with cognitive load theory' (p. 703) specifically managing *intrinsic cognitive load* linked with the element interactivity.

Within the subject of *element interactivity*, Ngu and Phan (2016) studied the complexity of linear equations again from a cognitive load theory perspective. They reason that the amount of interacting elements determines the complexity, hence, the *intrinsic cognitive load* of linear equations. The level/number of operational and relational lines (e.g. series of steps and/or operations used) in solving linear equations allows its classification based on level of complexity. They found that there is difference in the degree of interacting elements between one-step, two-step, and multi-step equations; and for that reason, suggested that teaching and learning of linear equations can occur in a sequential manner in accordance with the degree of element interactivity, hence reducing working memory load.

The next two recent studies used *worked examples* (also referring to example-based learning) in their approach in teaching problem solving. Worked example comprises a problem statement followed by series of steps for solution and then the final solution (Renkl & Atkinson, 2003 as cited in Hancock-Niemic, Lin, Atkinson, Renkl & Wittwer, 2016; Renkl, Atkinson, Maier & Stanley, 2002). Paas et al. (2003) stated that the use of worked examples is one of the earliest and best known cognitive load reducing technique because it reduces *extraneous cognitive load* compared to conventional problems (referred as worked example effect), which causes learners to focus attention on the nature and structure of the problem and useful solution steps (van Merrienboer & Sweller, 2005). Hancock-Niemic et al. (2016) investigated the efficacy of using faded worked examples presented in matrices with problem structure variability with the aim of enhancing student's ability to recognize the underlying problem structure in learning probability. They used a 2×2 factorial design comparing the effects of matrix-format and linear-format faded worked examples along with equivalent problem structure versus contrast problem structure designating 113 undergraduate students under four experimental conditions in a computer-based environment. Their study found three significant interactions between instructional format and problem structure variability on accuracy of anticipation, near transfer and medium transfer, and thus reinforce the value of using matrices for example-based teaching when combined with contrast-structure problem to improve college students' learning in probability. They also reported *no main effects* found for the matrix-format lesson treatment.

On the other hand, Rao (2017) studied the role of worked examples in learning to solve number problems using commutative and associative properties, and designed and conducted an intervention to reduce the cognitive load with the aim of improving the performance of students in mathematics. An experimental study on 76 grade six Indian students revealed that students who studied worked examples committed fewer errors and performed better in the test phase. Interesting in Rao's study is the use of Short Self Report Instrument (SSRI). Participants in the study were subjectively asked to rate themselves using the SSRI of their mental effort during learning and test stage to measure their cognitive load experience. Results of the study revealed that those students who studied worked examples not only committed fewer errors but also experienced low cognitive load.

Jalani and Lai (2015) studied the effect of Example-Problem-Based Learning (EPBL), a combination of worked-examples and problem solving in line with the Cognitive Load Theory. A pretest-posttest design was implemented to explore the effect of EPBL on Malaysian students' knowledge acquisition, learning transfer, and mental effort/ efficiency. They have concluded that the EPBL as teaching method enhances students' knowledge acquisitions, learning efficiency and learning transfer, as well as reduces the mental efforts during the learning phase.

Cognitive load theory in testing context. Not only Cognitive Load Theory influences the instructional designs in teaching mathematics. Until recently, researchers are gaining interest in the theory's applicability in educational measurement. Three studies are reviewed concerning educational measurement in line with cognitive load

perspective. First is the study of Elliott and colleagues (2010) who investigated the effects of using modified items in reading and mathematics achievement tests to enhance accessibility intended for *students with persistent academic difficulties.* The study argues that by modification of items, students with academic difficulties can gain *access* (opportunity for test takers to demonstrate proficiency on the target construct of a test or a test item) to the test just like students without disabilities.

They modified items in line with the principles of universal design (UD), cognitive load theory (CLT), and research on item development through simplification of language and reduction in memory demands without alteration on the depth of knowledge the items assess. A sample of 755 grade eight students (with and without disabilities) from four (USA) states took original and modified forms of the reading and mathematics tests and found that there is an increase test accessibility, reduced cognitive load, and improved validity of test score inferences about the same concepts measured by the general education achievement test.

These findings were also reported by Kettler, Rodriguez, Bolt, Elliott, Beddow and Kurz (2011) in their article. They stated that CLT can be of guide in modifying regular assessments which are intended to help students who consistency fail to meet proficiency level. These articles reported studies conducted in four states in USA. A four-state collaboration called Consortium for Alternate Assessment Validity and Experimental Studies (CAAVES) reported findings on item modification strategies with the goal of constructing future alternate assessments intended for students with disabilities, who, as they have argued, are unable to reach proficiency because of their

disability. Further, they stated that acceptable modifications to standard test should be made with the intention in mind of providing universal access and reducing cognitive load ensuring that test scores allow valid inferences about the same constructs measured by the regular assessment.

In line with the hypothesis of reducing cognitive load in mathematical tasks, Gillmor, Poggio and Embretson (2015) modified fifteen multiple-choice mathematics assessment items with the hypothesis that reducing *extraneous cognitive load* of the items will improve their validity for measuring student knowledge, will improve students' mathematics performance, and will reduce students' mathematics anxiety. They contend that *extraneous cognitive load* contributes to construct-irrelevant variance that hinders more accurate assessment of the intended construct. Thus, such extraneous load should be curtailed or reduced by modifying the original wording or decreasing the amount of text and eliminating irrelevant details (e.g. pictures, story context, numerical values that are unnecessary, etc.). Findings of the study confirm their first hypothesis. No significant results were found in reducing students' mathematics anxiety.

Elliott, Kurz, Beddow and Frey (2009) supported the reduction of extraneous cognitive load in their paper presentation at the National Association of School Psychologists' Annual Convention at Boston, Massachusetts. They reasoned that the magnitude of cognitive load demands may influence a test taker's ability to demonstrate performance on assessments, and further argued that *extraneous cognitive load* stresses by an assessment item restricts with the test taker's capacity to respond and should be eliminated (reduced) from the assessment process. They have advanced the

application of cognitive load theory to the inclusion of students with disabilities in assessment through the concept of *accessibility*. They said that the 'increased test accessibility provide students greater access to the test construct by reducing construct-irrelevant variance. Greater accessibility thus permits more valid test score inferences' (p. 10). Moreover, they pointed out that 'unless an assessment task has the dual purpose of both instruction and assessment, the items on a test should demand only those cognitive resources intrinsic to the target constructs they are intended to measure' (p. 13).

Another interesting research that stimulated the present study is by Secolsky, Judd, Magaram, Levy, Kossar, and Reese (2016). They developed a prototype of quantitative literacy assessment instrument based on the students' misconceptions as revealed by think-aloud protocol and interviews of 22 college students. Not only were they able to identify students' mathematical conceptions/ misconceptions but were also able to use it in testing context. A 20-item test was developed incorporating four solution strategies (three incorrect/ineffective solution strategies and one correct solution strategy). The test instrument used was not a multiple-choice test. Students were asked to answer each item in open-ended fashion to evaluate each solution strategy for its correctness. The tryout of instrument was conducted to two groups of students: treatment group (tested with four solution strategies) and control group (tested with no solution strategies), which yielded favorable outcomes to the treatment group. The treatment group asked to evaluate alternative strategies while taking posttest had

significantly higher mean scores than the comparison group. They believed that this evidence is suggestive of the power of having students confront misconceptions.

Synthesis

As what have been presented, numerous causes underlie mathematics difficulty. Many studies targeting an increase in mathematical achievement and/ or performance of low- achieving/ at-risk /struggling students used evidence-based and empirically validated instructional designs. These studies have been successful in their end goals. Working within the Cognitive Load Theory perspective, the current study designed to test whether CLT could generate findings with direct application not only on instruction but also on mathematics assessment addressing the influence of Mathematics assessment on Mathematics performance. Modification of items through the reduction of cognitive loads demanded by Mathematics assessment informs both assessment and instruction about how students used what they have learned, and in turn improve their test performance. In addition to the existing studies on reduction of cognitive loads (i.e. extraneous cognitive load) in Mathematics assessment through modification, this study addresses the management of intrinsic cognitive load in testing context through presentation of solution strategies as alternatives. This adds to the volume of research on the application of CLT in test modification. Moreover, at the center of high quality instruction and testing practices, the study supports access of all students to valid assessments where they can demonstrate fully the depth of their knowledge without access barriers. As Elliott, Kettler, Beddow and Kurz (2011) pointed out, access is an

intervention between the individual test taker characteristics and features of the test itself. In sum, modified mathematics test forms may enhance test performance not only for low-achieving students but for all students varying in mathematics abilities.

Conceptual Framework

Three types of test instruments of different item formats were developed. Two of the assessment instruments were modification of the original assessment tool referred to as *reduced cognitive load mathematics test forms.* The objective of the study was to determine the effect of reducing the cognitive load or the working memory demand of test on the Mathematics performance of high school students. It particularly sought to determine how varying the item formats (i.e. traditional, modified item *stem* with essential information, and item stem with solution strategies as *alternatives/options*) in a multiple-choice type test affects the performance of high and low ability high school students.

The foundation of the current study stems from the research of Secolsky et al. (2016) citing Bjork's (1994) work on desirable difficulty, who then argues that learners stop to process information because of cognitive overload. Cognitive load has been largely expounded by Cognitive Load Theory (CLT), a learning theory and instructional design principles based on assumptions about human cognitive structural design (Sweller, 2004 as cited in Elliot, Kurz, Beddow & Frey, 2009; van Merrienboer & Ayres, 2005). Kirschner (2002) stated that CLT provides guiding principles in the presentation of information that optimizes intellectual performance. As a *learning theory*, CLT defines

learning as alteration of information stored in long-term memory (LTM) (Aditomo, 2009). It is concerned about how cognitive resources are used during the process of learning and problem solving (Chandler & Sweller, 1991). Thus, as an *instructional theory*, it focuses on how instructional materials and/ or presentations can be presented so as to require less cognitive effort (Ismail, Kuldas & Hamzah, 2013); and as a consequence, cognitive resources are directed towards activities germane/ important to learning (Chandler & Sweller, 1991).

In educational settings, students are expected to acquire and/ or construct knowledge from new information; hence, CLT argues that instruction must take into account working memory's (WM) severe processing limitations (Aditomo, 2009). Primarily, CLT is concerned on how information can be easily processed in working memory (van Merrienboer & Ayres, 2005). *Working memory* can be defined as the memory concern in all occurring conscious cognitive processes (Paas et al., 2003). It is characterized based on its limitations in processing capacity and duration (Chinnappan & Chandler, 2010). According to CLT, a limited working memory can only store about seven elements but operates on two to four bits of new/ novel information and can only hold it for about 20 seconds (Ismail, Kuldas and Hamzah, 2013; van Merrienboer & Ayres, 2005; van Merrienboer & Sweller, 2005). All new information will be lost unless rehearsed intentionally. Hence, CLT assumes that the limitation of working memory to process novel information might be a major factor influencing the effectiveness of instructional presentations (Kalyuga & Sweller, 2004).

According to van Merrienboer and Sweller (2005), working memory cannot process many elements for novel information; hence, its capacity is overloaded if more bits of new information are processed at the same time. There are three forms of cognitive loads: *intrinsic*, *extraneous*, and *germane* cognitive loads. Sweller (1994) describes *intrinsic cognitive load* as referring to the complexity of the core information. It represents the number and difficulties of the series of steps needed for completing the task as presented to the student without the nonessential elements (Secolsky et al., 2016). Intrinsic load is the amount of cognitive processing required to comprehend material and depends on the number of elements and their interactivity (referred as element interactivity) (Elliott et al., 2009). An *element* is any information that needs to be learned (Sweller, 1994). Basically, *intrinsic cognitive load* is the intrinsic nature of the learning tasks themselves (van Merrienboer & Ayres, 2005).

On the other hand, *extraneous cognitive load* or *ineffective cognitive load* (Paas et al., 2003) refers to the elements of the task that are unrelated to completing the task. It is anything that occupies the WM capacity brought about by ineffective format of instruction or instructional materials and is irrelevant to learning (Gillmor et al., 2015; Paas et al., 2003). It is simply the way in which the tasks are presented (van Merrienboer & Ayres, 2005). *Germane cognitive load,* on the other hand, is the load which directly leads to learning. It refers to the amount of cognitive resources that the learner willingly invested for schema construction and schema automation in LTM (Gillmor et al., 2015; van Merrienboer & Ayres, 2005). It enhances learning such that the increase in effort or motivation can increase the cognitive resources devoted to a

task (Paas et al., 2003). More so, it accounts for an increase in cognitive load due to content variability, which in turn, can promote the generalization of learning (Elliott et al., 2009).

Taking CLT's stand, instructional design basically needs to reduce extraneous load, manage intrinsic load, and promote germane load (Aditomo, 2009). CLT is not only concerned with instructional designs and optimization of cognitive loads during learning and problem solving; but until recently it has influenced educational testing and measurement. As reviewed by Gillmor et al. (2015), Kettler et al. (2009, 2011), they found that the modified items (reduced cognitive load forms) did improve performance of students with severe disabilities. In their study, Gillmor et al. (2015) examined the applicability of CLT for enhancing test score validity and fairness with the aim of reducing construct-irrelevant variance when writing mathematics test items. They have modified fifteen multiple-choice math items using research-based strategies to reduce extraneous cognitive load. They conducted an experimental research to test the effects of the reduced cognitive load items on 222 middle-school students' performance and anxiety.

They found that students who were randomly assigned the reduced cognitive load form performed better than the students assigned the form with typical assessment items, making them conclude that removing or reducing extraneous cognitive load of assessment items proved effective in increasing the accessibility of the measured construct, which consequently resulted in enhanced performance and potential test fairness. In addition, they have identified that (a) signaling important information, (b)

having aesthetic item organization, and (c) removing extraneous content, are three of the load-reducing item modifications particularly effective in reducing item difficulty.

Van Merrienboer and Sweller (2005) citing Paas and van Merrienboer (1994) stated that one assessment dimension of cognitive load is performance. They argued that a higher cognitive load often increases the number of errors and slows down performance. If instruction is aimed for schema construction and automation in LTM, then 'tests of working memory content may provide a de facto measure of levels of expertise…such tests should be designed to assess the extent to which working memory limits have been altered by information in long-term memory' (Kalyuga & Sweller, 2004, p. 559). It cannot be denied that cognitive load becomes an issue in mathematics education both in instruction and in testing context.

The severe processing limitations of the working memory and the high element interactivity which brings high informational complexity (Sweller, 1994), thus giving high intrinsic cognitive load are issues that cannot be avoided; and in this study is given much attention. Taking this into mathematical context, most of mathematical tasks consist of high intrinsic cognitive load on the basis of the degree of element interactivity since such tasks demand student's cognitive resources to draw upon numerous elements of information and integrate that information to solve a problem (Chinnappan & Chandler, 2010).

During mathematical tasks like problem solving and answering mathematics test especially if the test requires procedural knowledge or both procedural and conceptual knowledge, learners' attention is focused on mentally assimilating the numerous

sources of information (i.e. searching for operators) needed to solve a problem. This search process devotes a high demand of cognitive resources in the working memory and might interfere with learning and in the testing context, with performance as well (Chandler & Sweller, 1991). To cope with this cognitive demand, Kirschner (2002) stated that schema can reduce working memory load. Furthermore, he goes on arguing that "A schema can be anything that has been learnt and is treated as a single entity. If learning has occurred over a long period of time, a schema may incorporate a huge amount of information…schema construction aids the storage and organization of information in long-term memory and reduces working memory load" (p. 3).

As mentioned earlier, the study aimed to determine how varying the item formats (i.e. traditional, modified item *stem* with essential information, and item stem with solution strategies as *alternatives/ options*) in a multiple-choice type test affects the performance of low-achieving and high-achieving high school students. The test instruments were composed of 35 items, while each item stem has four alternatives/ options, three distracters, and one correct answer.

The first type of test is the *traditional format* (Test A). It is a typical multiple-choice type of assessment where most of the time, the stem is written in question form and alternatives offered are short responses. The second and third types of assessment that were developed are modifications of the first type. They are modified to reduce and/ or manage the extraneous and intrinsic cognitive loads respectively. As argued by Gillmor et al. (2015) acceptable modifications are made with the intent of providing *universal access* and reducing cognitive load to ensure that scores permit valid

inferences about the same constructs measured by the regular/ traditional assessment. The goal of the current study is to determine the effects of modifying items of Mathematics test by reducing working memory demands (load-reducing item modifications) with no alteration on the depth of knowledge the item assesses.

Table 1
Cognitive Load Reducing Strategies for Item Modification

Method/Technique	Description and Application to Testing
Translation	Reduce word count and simplify language. Text economy to reduce information load.
Weeding	Text economy. Strip content down to essentials. Eliminate extraneous visuals and text. All included visuals are necessary.
Visual Aid	Use diagram to represent spatial information.
Sequencing	Proper sequence is observed that includes asking questions first to give a direction to the item, and then followed by supporting information. Proper sequence is also observed in ordering answer options logically.
Aesthetics	Format item logically and aesthetically. Place text near corresponding figure.
Signaling	Focus attention with signal and cues to important visual and textual content. Use bold for vocabulary words and arrows for important elements of figures.

Table 1 shows the strategies/ techniques used in modifying items for managing the cognitive load in the developed assessment instruments (Test B and Test C). The study integrated the methods used in the study of Elliott et al. (2009) and Gillmor et al. (2015) for the purpose of item modifications.

The second type of assessment (Test B) developed measures the same constructs being tested by every item of the traditional format, but the item stems are modified. Instructions and/ or language are simplified without losing the essential construct being assessed of the item, and used other cognitive load reducing

techniques or methods mentioned in Table 1. In the second type, all item stems of the traditional format are text economized. Instructions are direct, concise, and clear. Modifications of item stem include visual aid such that figures and texts are integrated to lessen the extraneous information. The second type is focused on reducing the extraneous cognitive load demanded by the items. As reviewed by Gillmor et al. (2015), studies have shown that construct-irrelevant factors such as complexity of language can interfere with learner performance on assessments. In addition, Elliott et al. (2011) stated that for students with low reading ability, complex text in a math test is an access barrier hindering full demonstration of knowledge, skills, and/ or abilities in mathematics.

The current study in line with Elliott et al. (2009, 2010), and Gillmor et al. (2015) hypothesized that reducing the extraneous cognitive load of mathematics assessment item improves validity for assessing student knowledge and in turn improves student performance. Modification of the first type (traditional format) measures the same underlying construct being tested by the item while stripping the unrelated difference caused by extraneous cognitive load, thus resulting to a more valid measurement of student knowledge. Without being “loaded” with additional reading and complexity, student performances are likely to be higher on this item than the traditional assessment. This line of thinking is supported by Kettler et al. (2011) as cited in Gillmor et al. (2015) that reducing the length of the item stimulus is particularly effectual. *Modifications,* in full sense, are alterations made to a test or aspects of items (item format) that make a test more accessible for most learners while maintaining to assess

grade-level content and skills at the same depth of knowledge as unmodified items (Elliott et al., 2010).

The third type of assessment (Test C) developed measures the same constructs being tested by every item of the traditional format, where the item stems are the same with the traditional assessment but the alternatives/ options are modified into solution strategies. Every item offers four solution strategies, three incorrect/ ineffective solution strategies, and one correct solution strategy. Solution strategies offered as alternatives/options represent common conceptions on certain mathematical tasks. Each item was deconstructed with respect to their intrinsic difficulty (Secolsky et al., 2016) based on how high school students may conceive or misconceive mathematical tasks in operations on integers and polynomials.

The study hypothesized that presenting solution strategies reduces the intrinsic difficulty/ complexity of items on the assessment by way of presenting viable solution strategies as *alternatives/ options* in a multiple-choice test format. Intrinsic cognitive load is managed by integrating and connecting all necessary elements to solve the problem as reflected by each solution strategy. Moreover, Chandler and Sweller (1991) contend that physically integrating all disparate sources of information can save learners from performing unnecessary mental integrations that interfere with learning and in the position of the study, may enhance performance. Highly complex schema when dealt as *one* element significantly reduces working memory load and greatly extends its capacity (van Merrienboer & Ayres, 2005). Furthermore, these solution strategies offered as alternatives/ options are schemas themselves, hence activating

the long-term memory which according to Kirschner (2002), may help reduce the working memory load because the element interactivity is artificially reduced (van Merrienboer & Ayres, 2005). Solution strategies cue learners on the keywords of the series of processes necessary to attain the correct answer, thus activating the schema stored in long-term memory which will be passed on to working memory for conscious processing, thus extending the working memory capacity. This line of reasoning is supported by Ericsson and Kintsch (1995) as cited in van Merrienboer and Sweller (2005) that working memory has no known restrictions when dealing with previously learned information stored in long-term memory.

As mentioned earlier, working memory is limited in capacity when dealing with novel, unorganized information. As van Merrienboer and Sweller (2005) argued, as the number of elements that needs to be organized increases linearly, the number of possible combinations increases exponentially. But working memory capacity is greatly expanded when dealing with previously learned information. Memory capacity dramatically alters the manner of information processing as information increases in familiarity (Sweller, 2003, as cited in Kalyuga & Sweller, 2004). Alternatives/ options in Test C as solution strategies are presented as organized information such that all essential elements to solve a particular problem are combined singularly. A student answering a particular item does not necessarily have to search cognitively for elements stored in LTM and then cognitively organize them to be able to solve the problem; but because of offered solution strategies, she/he evaluates/ discriminates which schema (solution strategies) best solves the problem since these solution strategies are well

organized and can be processed simultaneously and automatically by the working memory. Understanding commences when all essential elements are simultaneously processed (Paas et al., 2003).

This type of assessment was developed deliberately with the researcher's intention to manage the intrinsic cognitive load (presenting solution strategies decreases intrinsic difficulty since students need not really have to solve the problem but to evaluate the correct solution strategy) and to scaffold test takers' retrieval of schema in long term memory. As Secolsky et al. (2016, p. 14) stated 'the ability to distinguish correct from incorrect solution strategies goes further into developing mathematics ability than merely solving the actual problems. It requires understanding what is wrong with a given solution strategy.'

Presenting them solution strategies in the form of alternatives/ options lessens the intrinsic item difficulty and working memory capacity. Based on the argument of the study, freeing up or "unloading" them certain cognitive resources in working memory will make these cognitive resources devoted more in evaluating the correct solution strategy, which when done, improves students' test performance. The use of means-ends analysis that generates extraneous expenses when solving traditional test in Mathematics can be avoided through presentation of solution strategies. Extraneous load rate is reduced by preventing students from irrelevant cognitive search processes so that they can concentrate all their working-memory capacity to study each solution strategy and activate schema in their long-term memory.

Table 2 shows the test format/ layout of the test instruments that were developed. Test instruments were compared in terms of item stem and alternatives/options. In terms of item stem, Test A and Test C have the same set of item stem. This means that the wordings and presentation of the questions were identical while Test A and Test B differ in item stem. Test B is a modification of Test A such that the item stem of Test B is modified to reduce the extraneous cognitive load. In terms of alternatives/ options, generally all test instruments have the same set of options; but Test C was modified into solution strategies to decrease the intrinsic difficulty of answering the test as it offers a ready step-by-step solution guide, thereby managing or artificially reducing the intrinsic cognitive load. Test A and Test B have identical set of alternatives/ options. For Test C, the order of the alternatives was changed for logical and aesthetic purposes (as recommended by one Mathematics expert).

Table 2
Format and Layout of Test Instruments

Parts of Test	Test Format		
	Test A	**Test B**	**Test C**
Item Stem	√	**Modified item stem** (item stem is text economized such that extraneous cognitive load is reduced; other extraneous cognitive load reducing technique such as use of diagrams and cues/ signals to focus attention to important visual and textual content).	√
Alternatives/ Options	√	√	**Modified Alternatives/ Options** (offered answers are modified into solution strategies such that intrinsic cognitive load is managed by reducing the intrinsic difficulty of answering the test).

√ (means that they have the same format)

The study agrees with Elliott et al. (2009, 2010) and Gillmor et al. (2015) that removing the added demand on the working memory will likely improve the validity of the measurement and increase students' mathematics performance. Beddow, Kurz and Frey (2011) stated that among the interaction domain in Accessibility Theory proposed by Beddow and colleagues, cognitive domain warrants the most attention in designing accessible test and test items because of the limitation of the working memory. Hence, Cognitive Load Theory warrants attention in test development. As Gillmor et al. (2015) concluded, CLT is suitable for providing guidance in the item writing development. It offers a constructive lens through which to evaluate and modify tests and their respective items to increase test takers' access to the target construct (Beddow, Kurz & Frey, 2011).

The study is also in support of access in educational testing whereby test takers have opportunity to demonstrate proficiency on the target construct of the test (Elliott, Beddow, Kurz & Kettler, 2011). Further, in response to the goals of modification stated in the *Handbook of Accessible Achievement Tests for All Students: Bridging the Gaps between Research, Practice, and Policy* by Elliott et al. (2011), the study noted the following when modifying test items: (a) preserve the target construct of the original test including the preservation of essential content, depth of knowledge, grade-levelness, and intended knowledge, skills and abilities intended to measure; and (b) isolate the target construct by reducing access barriers on the test score.

The present study argues that acceptable modification of items of an assessment could reduce the cognitive loads demanded by the items without changing the content

or constructs being tested, and doing this improves validity for assessing student knowledge, which consequently improves student performance. In line with the Cognitive Load Theory, the study contends that if extraneous cognitive load is reduced or intrinsic cognitive load is managed in the testing context, there is an increase in the test performance of students.

The diagram below shows the study's research paradigm.

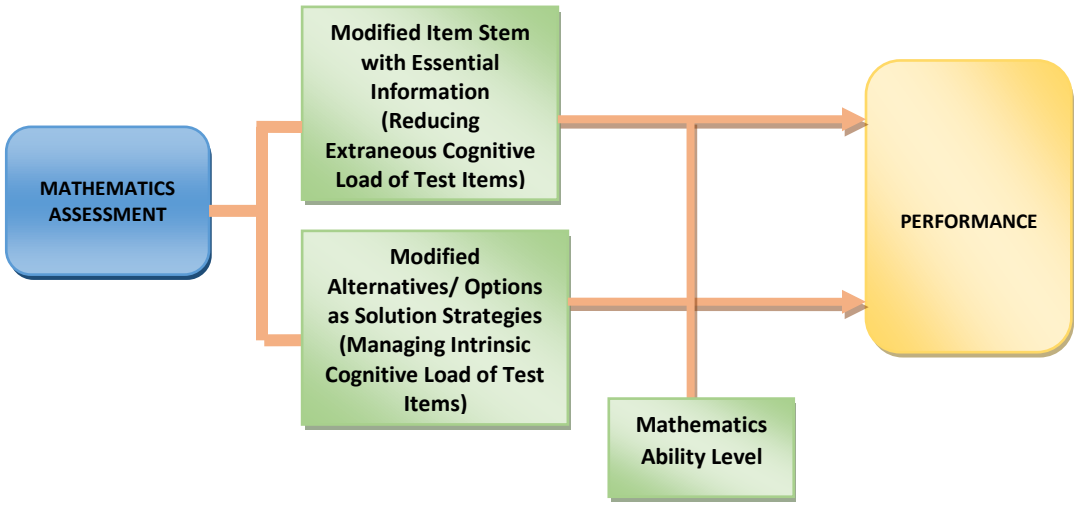


Figure 1. Conceptual framework

Figure 1 shows that Mathematics performance is influenced by test format and ability level. The present study argues that test format as a factor affects test performance in mathematics. Cognitive loads demanded by test items play an influential role in students' performance in assessment. As Cognitive Load Theory explains, since working memory is severely limited in processing capacity, extraneous and intrinsic cognitive loads in test items must be reduced and/or managed through

modifications of the test formats (i.e. modifying item stem or modifying alternatives) which are believed to improve students test performance.

Two modifications were made on the traditional assessment. First is modification of item stem with essential information. Item stems were stripped of extraneous cognitive loads through cognitive load reducing techniques previously mentioned in Table 1. The reduction of extraneous cognitive load of test is likely to address the cognitive demands of the test that are intrinsic to the target mathematics construct. Essential information needed to process cognitively is emphasized by stripping unrelated elements, thus focusing on the target construct of the test, with a purpose of improving test performance.

Second is modification of alternatives/ options. Alternatives were modified into solution strategies to scaffold test taker's retrieval of schema in LTM, thus reducing the intrinsic difficulty of the test and freeing up cognitive resources so that attention is devoted to the intrinsic nature of the targeted construct. This reduces the cognitive demands on test taker because all disparate sources of information where integrated and put into one, resulting to the lessening of mental effort on the part of test taker.

On the other hand, mathematics ability level also plays an important role influencing test performance. High-achieving or proficient students have higher levels of prior knowledge in the mathematics domain, while low-achieving students have lower level of prior knowledge in the mathematics domain. Blayney, Kalyuga and Sweller (2010) stated that experienced learners had a sufficient knowledge base to handle instructional materials without cognitive overload; moreover, this is also true for

proficient learners handling the cognitive demands of test materials. According to Kalyuga, Law and Lee (2013), organized knowledge structures in LTM are the most important variables of proficient performance in different areas. This is true for mathematics since mathematics concepts are interdependent. Students with high mathematics ability level have higher levels of knowledge in mathematics domain because of well-organized schematic structures in their LTM. Thus the amounts of information they can process in working memory are substantially higher than low-achieving students.

This means that high-achieving students have higher germane cognitive load than low-achieving students. The role of germane cognitive load in mathematics ability and in influencing test performance is important. Germane cognitive load accounts for the cognitive resources that are natural to learning. Specifically, it accounts for the constructed schema and its automatic retrieval. High ability students have high germane cognitive load indicated by their capacity to store well-constructed schema and they are likely to perform better than those at-risk/ struggling students who have less germane cognitive load indicated by their less capacity to store schema and their less capacity to automatically retrieve it when needed.

Statement of the Problem

The present study aimed to determine the effect of reducing cognitive load of test items on the Mathematics performance of high school students. Particularly, it sought to determine how varying the item formats (i.e. traditional, modified item *stem* with

essential information, and item stem with solution strategies as *alternatives/ options*) in a multiple-choice type test affects the performance of high and low ability high school students.

Specifically, the study sought to answer the following questions:

1. Do test formats and ability level influence Mathematics performance of high school students?
2. What are the students' perceptions on the different test formats in terms of test preference?

Definition of Terms

The following terms used in this study are operationally defined as follows:

Cognitive Load – this pertains to cognitive resources imposed or demanded by test items on working memory.

High- Achieving Students – these are students who perform well in the subject area Mathematics. They are students whose grade averaged 90 and above in the subject area Mathematics.

Low- Achieving Students – these are struggling students in Mathematics; are those groups of students at-risk for failure in the subject area Mathematics either by teacher nomination (teacher's professional judgment of the mathematics performance of the student) or through consistent low performances across their school careers. They are students whose grade averaged 79 or below in the subject area Mathematics.

Modified Alternatives/ Options Test Format (Test C) – it is a reduced cognitive load mathematics test format where alternatives/ options are modified into solution strategies with the intent of managing the intrinsic cognitive load through reduction of intrinsic difficulty of test items.

Modified Item Stem Test Format (Test B) – it is a reduced cognitive load mathematics test format where item stems are modified by reducing the extraneous cognitive load through text economization, use of visuals, cues and signals without losing the essential information and target construct measured.

Performance – this pertains to the Mathematics test scores of students obtained from the posttest.

Traditional Test Format (Test A) – it is a typical multiple-choice type of assessment where most of the time the stem is written in question form and alternatives offered are short responses.

CHAPTER 2

METHODS

Research Design

In pursuit of answers to the research questions, the study employed an experimental design, specifically a **2 x 3 between-subjects factorial design**. According to Johnson and Christensen (2004), experimental design is appropriate to show cause-effect relationships. To be able to achieve the identification of causality, subjects are exposed to different treatment/ experimental conditions (Charnessa, Gneezy, & Kuhn, 2012). Since in the real classroom setting and in any field of study, multiple factors or variables influence behavior as such, the present study addressing this need for an experimental design to investigate multiple factors or variables simultaneously used factorial design. **Factorial design** is a design in which the influences of two or more independent variables are being studied. This design allowed the researcher to study not only the effects of two independent variables separately (main effects) but also how they combine (interaction effects) to influence the dependent variable (Cavazzuti, 2013; Oehlert, 2010; Sekaran, 2003; Singh, 2006). The basic rationale of this design is to examine whether the independent variables interact and if they interact, what possible patterns of interaction they possess.

Two independent variables are addressed in the study. The first factor is the ability of high school students in Mathematics with two (2) levels: low-achieving and high-achieving; and the second factor is the test format of Mathematics assessment

with three (3) kinds: traditional (Test A); modified item stem with essential information (Test B); and item stem plus solution strategies as alternatives/options (Test C). Hence, a 2 x 3 factorial design was achieved. Combinations of the different levels of factors have lead to six treatment/ experimental conditions. In this study, the test format of Mathematics assessment was a true independent variable since it was manipulated. Manipulation was done by varying the test format through modification of item stem (Test B) and modification of alternatives/ options (Test C), while simultaneously reducing the different cognitive loads demanded by test items. The level of ability of students as a factor is a quasi-independent variable since it is a pre-existing participant characteristic. Thus, the study sought to investigate whether the variation of item formats (i.e. traditional, modified item stem, and modified alternatives) effects change in the Mathematics performance as a function of students' ability.

As to the assignment of participants to the six (6) treatment conditions, between-subjects design was employed. **Between-subjects design** is a design in which subjects/participants are exposed to only one treatment condition (Cavazzuti, 2013; Charnessa et al., 2012; Schepanski, Tubbs, & Grimlund, 1992). Since the focus of the study is to investigate the effects of varying test formats on the performance of high and low achieving high school students, one type of assessment is given once (posttest only) to each individual participant. Doing this helped the study to reduce the carry-over effects if participants are subjected to repeated measures (*within-subjects* design). In addition, because participants were tested only once, this lowered the likelihood of suffering boredom or accomplishing the test through practice and experience that might

skew the outcome of the study (Cavazzuti, 2013), consequently avoiding the experimental demand/ demand characteristics/ demand effects to the participants as a threat to construct validity (*see* Schepanski et al., 1992 for details).

Further, the researcher's decision to use posttest only was that pretest might be a threat to the internal validity of the study because exposure to a pretest might influence students' performance on a posttest. This concern is supported by Ross and Morrison (2003) arguing that the decision not to use a pretest may bias posttest results and since none of the participants are measured twice, testing specifically exposure to a pretest is not a possible threat to internal validity (Fraenkel & Wallen, 2008) of the study.

Additionally, to further strengthen the internal validity of the study, experimental conditions are run at once. In an intact class, all treatment conditions specifically the three levels of the second factor: Test A, Test B, and Test C were run at the same time; this meant that no differences in what takes place in the environment overtime can cause individuals to differ because they are all run at once (environmental threat) (Sekaran, 2003).

Sampling and Participants

All participants (subjects and non-subjects) that have been included in this study were high school students, particularly Grade Nine, ages 14-16 years old. They were chosen deliberately as respondents/ participants of the study because they have previous knowledge already in operations on integers and polynomials. Also, at this stage, they can demonstrate their knowledge on the mentioned mathematics construct

when interviewed. In getting samples of both high and low achieving students for the study's experimentation, a purposive sampling was used (*see* Cohen, Manion & Morrison, 2000, 2007 for definition). The researcher included participants based on their typicality, in this case, group of low-achieving and group of high-achieving students. The selection of student-participants to be included in the low-ability group is through teacher's professional judgment of low-achievement in Mathematics subject and through consistent low performances as indicated by their grade point/ general average (i.e. grades in the previous grading periods averaging 79 or below). For the selection of high-ability group, students who performed well in the subject area Mathematics and whose performance averaged 90 and above in the previous grading periods were included in the study. The grades from the first quarter up to the third quarter of the school year 2017-2018 were averaged and are used to get a purposive sample for the low and high-ability groups. Hence, the selection of samples is criterion-referenced.

The researcher, in coordination with the Schools Division Superintendent, the Principal, and the Mathematics Head Teacher communicated to the respective Grade Nine teachers the purpose of the study, specifically the selection of low and high-ability students in the subject area Mathematics. Through a letter, the researcher asked them a copy of their records on the performances of the selected intact classes for which the purposive samples of low and high-ability groups are chosen.

Included also in the selection process is mortality. The researcher excluded in the experimentation those participants with poor attendance in class as they might possibly drop or absent and eventually discontinue when actual data gathering/ testing

begins. Exclusion in this sense did not mean that they could not take the test or that the researcher prohibited them from taking the test; rather, it means that they were excluded in the list of subjects from which the researcher extracted the data. The decision to exclude them was based on the interview of the researcher to the Grade Nine teachers about each student's performance and attendance.

After they have been purposively selected, low-achieving and high-achieving students have been randomly assigned to six treatment conditions. Since the design of the experiment is a *between-subjects* design, a large number of participants were needed to ensure generalizable results and to make sure that the sample sizes could adequately detect significant differences among means (power calculations). For each treatment condition, a sample of 20 subjects was selected; hence, a total of 120 participants were selected for the experiment-- 60 low-achieving and 60 high-achieving student-subjects. In the randomization of low-ability and high-ability groups, a separate list of both groups was first secured. The lists were arranged alphabetically both for the two groups regardless of their average grades. To avoid assignment bias, a *systematic random assignment* was done where the first person was assigned to the first type of assessment (Test A), the second person to the second type (Test B), and the third person to the third type (Test C).

Random assignment of participants protects the study's results for confounding extraneous variables (Creswell, 2009; Fraenkel & Wallen, 2008; Johnson & Christensen, 2004; Ross & Morrison, 2003). By virtue of randomization for both low-ability and high-ability groups, it can be said that the three groupings made from high

and low achieving students are equivalent and any differences are due to chance since the variables other than the types/ test formats of assessment (i.e. traditional and two reduced cognitive load mathematics test formats) were equally distributed to each experimental condition (Cohen et al., 2007; Johnson & Christensen, 2004). Moreover, arranging the experiments in groups that are similar/ equivalent to one another reduces the sources of variability both for the groups of low-ability and high-ability students, consequently improving the precision of measurement.

For ethical issues and concerns, the researcher secured informed consents from their parents/guardian since the participants are minors (*see* Appendix A). All necessary permissions coming from the school administrators (*see* Appendix B, C, D and E) were also secured. The research locale was a sub-urban community in Region IV-A CALABARZON, a public national junior high school. The school is categorized as 'big school' with approximately 8, 000 student population.

Instruments

Three sets of Mathematics assessment instruments were used for the study. The scopes of the test instruments were on the topics of operations on integers and polynomials. The researcher purposively selected the content of the test instruments because in his seven years teaching the subject, he has always noticed the poor performance of high school students in more advanced mathematics concepts, attributed mainly to poor mastery of basic mathematical concepts such as knowledge in

operations on integers and polynomials. Each test was composed of 35 items in multiple-choice type format.

Table 3
Examples of Mathematics Test Items

Item No. 12

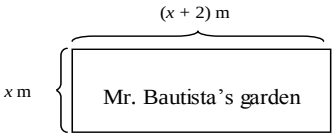
Test A

Mr. Bautista has a rectangular garden that is x meters wide and $(x + 2)$ meters long. What is the area of his garden?

A. $2x$ m^2
B. $(x^2 + 2)$ m^2
C. $(2x + 2)$ m^2
D. $(x^2 + 2x)$ m^2

Test B

What is the area of the **rectangular** garden?



A. $2x$ m^2
B. $(x^2 + 2)$ m^2
C. $(2x + 2)$ m^2
D. $(x^2 + 2x)$ m^2

Test C

Mr. Bautista has a rectangular garden that is x meters wide and $(x + 2)$ meters long. What is the area of his garden?

A. I.) Area of rectangle= length + width
II.) Add $(x + 2) + x \Rightarrow$ **$2x$ m^2**

B. I.) Area of rectangle= length + width
II.) Add $(x + 2) + x \Rightarrow$ **$(2x + 2)$ m^2**

C. I.) Area of rectangle = length × width
II.) Multiply $(x + 2)(x) \Rightarrow$ **$(x^2 + 2)$ m^2**

D. I.) Area of rectangle = length × width
II.) Multiply $(x + 2)(x) \Rightarrow$ **$(x^2 + 2x)$ m^2**

Item No. 25

Test A

If a pencil costs $(m + 5)$ pesos, how much does $(2m – 3)$ pencils costs?

A. $(3m + 2)$ pesos
B. $(2m^2 – 15)$ pesos
C. $(2m^2 – 7m – 15)$ pesos
D. $(2m^2 + 7m – 15)$ pesos

Test B

1 pencil = $(m + 5)$ pesos
$(2m – 3)$ pencils = ☐

What is the value of ☐ ?

A. $(3m + 2)$ pesos
B. $(2m^2 – 15)$ pesos
C. $(2m^2 – 7m – 15)$ pesos
D. $(2m^2 + 7m – 15)$ pesos

Test C

If a pencil costs $(m + 5)$ pesos, how much does $(2m – 3)$ pencils costs?

A. I.) Cost + No. of pencils = Total cost
II.) Add: $(m + 5) + (2m – 3)$
III.) Therefore, answer is
$(3m + 2)$ pesos

B. I.) Multiply the no. of pencils and cost; $(2m – 3)(m + 5)$
II.) $2m(m) = 2m^2$ and $(–3)(5) = –15$
III.) Add the terms. Hence,
$(2m^2 – 15)$ pesos

C. I.) Multiply the no. of pencils and cost; $(2m – 3)(m + 5)$
II.) Middle term = $10m + (–3m) = 7m$
III.) Add the terms. Hence,
$(2m^2 + 7m – 15)$ pesos

D. I.) Multiply the no. of pencils and cost; $(2m – 3)(m + 5)$
II.) Middle term = $10m + (–3m) = –7m$
III.) Add the terms. Hence,
$(2m^2 – 7m – 15)$ pesos

Each item had four alternatives/ options, three distracters, and one correct answer. All items were researcher-made. In the development of the instruments, a table

of specification was provided first (*see* Appendix I). No items were developed on the cognitive process dimension- Remembering. The researcher considered items that could truly reveal students procedural knowledge and both procedural and conceptual knowledge. Table 3 shows examples of the items in the three mathematics instruments.

Face and content validation. Before administering the three different types of Mathematics assessment, they were evaluated by two experts in the field of Mathematics and three experts in the field of Psychology (*see* Appendix G), who rated them based on three criteria: Content, Format and Layout, and Language Suitability; and whether the varying test formats reduced and/ or managed the different cognitive loads. Experts evaluated each test instrument using the evaluation form provided by the researcher (*see* Appendix H).

Each criterion has subcomponents for which the evaluators rated accordingly using the rating scale with their descriptive equivalent: 4- Strongly Agree (SA), 3- Agree (A), 2- Disagree (D) and 1- Strongly Disagree (SD). Ratings were tabulated according to subcomponents of each criterion. Descriptive statistics through means and standard deviations were used to quantitatively analyze each test instrument through the subcomponents and through the criteria in general with qualitative interpretation or remark for acceptability and suitability. Table 4 shows the interpretation of the mean rating for each subcomponent and criterion.

Table 4

Validation of Mathematics Assessment Instrument in Operations on Integers and Polynomials: Mean Ratings and their Interpretation

Mean Rating	Interpretation
3.26 - 4.00	Strongly Agree
2.51 - 3.25	Agree
1.76 - 2.50	Disagree
1.00 - 1.75	Strongly Disagree

Aside from the descriptive statistics used, inter-rater agreement was also reported. To measure the extent to which the five raters agreed when rating each test instrument, the researcher used percent of absolute agreement and intraclass correlation coefficient. Percent of absolute agreement is the percentage of cases when two raters agree (Gwet, 2014). It can be calculated by the number of times raters agree on a rating divided by the total number of ratings (Graham, Milanowski & Miller, 2012). Stated otherwise, percent of absolute agreement is equal to the number of agreements divided by the sum of number of agreements and number of disagreements multiplied by 100% (Araujo & Born, 1985). To compute the overall percent of absolute agreement among multiple raters in every criterion for each test instrument, the researcher referred to the calculations provided by Griffin (2011) and used *Microsoft Excel.* Table 5 shows the interpretation of percent of absolute agreement.

Table 5
Interpretation of Percent of Absolute Agreement

Range	Remark
0	No agreement
0.0 < P ≤ 20.0	Negligible agreement
20.0 < P ≤ 40.0	Low agreement
40.0 < P ≤ 60.0	Moderate agreement
60.0 < P ≤ 80.0	Marked substantial agreement
80.0 < P ≤ 90.0	High agreement
90.0 < P < 100.0	Very high agreement
100	Perfect agreement

Source: Ocampo et al. (2015)

Furthermore, the researcher calculated the intraclass correlation coefficient. Intraclass correlation coefficient (ICC) is used to know the context of reliability of the ratings of several judges (Shrout & Fleiss, 1979). ICC is an index that reflects both the degree of correlation and agreement between measurements (Koo & Li, 2016). The researcher used *SPSS* to calculate the ICC index. The ICC form used by the researcher is two-way mixed effects, absolute agreement, multiple raters/measurement (McGraw & Wong, 1996, as cited in Koo & Li, 2016). Table 6 shows how to interpret the ICC index computed for each criterion of each test instrument.

Table 6
Interpretation of Intraclass Correlation Coefficient

ICC index	Interpretation
ICC < 0. 50	Poor reliability
0.50 ≤ ICC ≤ 0.75	Moderate reliability
0.75 < ICC ≤ 0.90	Good reliability
ICC > 0.90	Excellent reliability

Source: Koo and Li (2016)

Test format A. The first developed item format (Test A) was the traditional type. Table 7 shows the evaluation of five experts on Test A. In terms of the Content criterion, it received a mean rating of 3.70 suggesting that evaluators strongly agreed on the content of Test A items as level appropriate and relevant to the objectives of the test. Format and Layout criterion received a mean rating of 3.65. Experts strongly agreed that proper format was observed including adherence to traditional way of writing a test. For language suitability, experts rated it 3.70 which implied that they strongly agreed that sentences used were grammatically acceptable, clear and simple.

The criteria used in evaluating Test A by evaluators are shown in Table 8. Mean, standard deviation, percent of absolute agreement, and intraclass correlation coefficient were indicated with corresponding remarks. Overall, evaluators strongly agreed that Test A followed appropriate content, format and layout, and language. Using percent of absolute agreement, they had moderate agreement across all three criteria. In sum, they moderately agreed that Test A followed the three criteria properly. In terms of intraclass correlation coefficient, evaluation on Content criterion had an excellent inter-rater reliability. Format and Layout criterion marked a good inter-rater reliability; and Language criterion marked a moderate inter-rater reliability. In totality using the ICC index, evaluators agreed rating Test A by assigning relatively same scores with excellent reliability index (ICC = 0.91, $p < 0.05$, Cronbach's Alpha = 0.92) (*see* Appendix Q for Overall ICC computation).

Table 7
Experts' evaluation on Test A

Criteria	*M*	*SD*	Remark
A. Content			
A.1 Conforms to DepEd K to 12 Mathematics learning competencies for Junior High School.	3.60	0.55	SA
A.2 The items are level- appropriate. Problems are presented within the level of understanding of learners.	3.80	0.45	SA
A.3 Relate assessment to the content under study and to student capacity.	3.80	0.45	SA
A.4 All problems are directly relevant to the objectives. Match assessment to intended learning objectives.	3.80	0.45	SA
A.5 Uses assessment that can truly reveal whether students understand the learning.	3.40	0.55	SA
A.6 Appropriate answers have been included.	3.80	0.45	SA
Grand Mean	**3.70**	**0.48**	**SA**
B. Format and Layout			
B.1 Questions are progressive in difficulty.	3.60	0.55	SA
B.2 Layout of pages is well organized making the whole test appears interesting and easy to answer.	3.60	0.55	SA
B.3 Appropriate in length for the allotted time.	3.40	0.55	SA
B.4 Followed the traditional way of writing a test.	4.00	0.00	SA
Grand Mean	**3.65**	**0.41**	**SA**
C. Language Suitability			
C.1 Questions and instructions are clear, easy to comprehend and grammatically acceptable.	3.60	0.55	SA
C.2 Sentences used in the problems are clear and simple.	3.80	0.45	SA
Grand Mean	**3.70**	**0.50**	**SA**

Table 8
Test A Evaluation Checklist

Criteria	Mean	*SD*	Remark	% Absolute Agreement	Remark	ICC Index [a]	Remark
A. Content	3.70	0.48	SA	53.33	Moderate Agreement	0.92	Excellent Reliability
B. Format and Layout	3.65	0.41	SA	55.00	Moderate Agreement	0.76	Good Reliability
C. Language Suitability	3.70	0.50	SA	50.00	Moderate Agreement	0.75	Moderate Reliability
Overall	**3.68[b]**	**0.46**	**SA**	**52.78[b]**	**Moderate Agreement**	**0.91**	**Excellent Reliability**

[a]*Two-way mixed effects, absolute agreement, multiple raters (average measures)*
[b]*Based on mean*

Overall, it can be gleaned from the mean rating, percent of absolute agreement, and ICC index from the experts' evaluation that Test A was acceptable, suitable, and face valid.

Test format B. The second item format (Test B) where each item stem in the traditional type was modified in terms of simplifying instruction without losing the essential information and the construct being tested. Test B had the same alternatives/options with Test A, but the item stems were text economized and/or simplified using cognitive load reducing technique.

Table 9 shows the evaluation on Test B. All five experts rated the test instrument according to appropriateness of content, format and layout, and language with “Strongly Agree” indicator having mean ratings for each criterion 3.67, 3.68, and 3.70 respectively. For Test B, the Format and Layout criterion had additional subcomponents compared to the same criterion on Test A. The additional subcomponents particularly B.4 to B.10 were descriptive conditions pertaining to extraneous cognitive load reducing techniques and/ or modifications made for Test B. Evaluators strongly agreed that the format and layout observed the cognitive load reducing technique.

The criteria used in evaluating Test B are shown in Table 10. Mean, standard deviation, percent of absolute agreement, and ICC were indicated with corresponding remarks. Raters strongly agreed that Test B followed appropriate content, format and layout, and language with an overall mean rating of 3.68.

Table 9
Experts' Evaluation on Test B

Criteria	*M*	*SD*	Remark
A. Content			
A.1 Conforms to DepEd K to 12 Mathematics learning competencies for Junior High School.	3.60	0.55	SA
A.2 The items are level- appropriate. Problems are presented within the level of understanding of learners.	3.80	0.45	SA
A.3 Relates assessment to the content under study and to student capacity.	3.80	0.45	SA
A.4 All problems are directly relevant to the objectives. Match assessment to intended learning objectives.	3.40	0.55	SA
A.5 Uses assessment that can truly reveal whether students understand the learning.	3.40	0.55	SA
A.6 Appropriate answers have been included.	4.00	0.00	SA
Grand Mean	**3.67**	**0.42**	**SA**
B. Format and Layout			
B.1 Questions are progressive in difficulty.	3.60	0.55	SA
B.2 Layout of pages is well organized making the whole test appears interesting and easy to answer.	3.80	0.45	SA
B.3 Appropriate in length for the allotted time.	3.60	0.55	SA
B.4 Item stems are stripped of construct irrelevant information.	3.60	0.55	SA
B.5 Simplification of item stems does not affect the construct being tested.	3.40	0.89	SA
B.6 Content of item stems is stripped down to essentials. Eliminate extraneous visuals and texts.	3.80	0.45	SA
B.7 Proper sequence is observed that includes asking questions first to give a direction to the item, and then followed by supporting information. Proper sequence is also observed in ordering answer options logically.	3.40	0.55	SA
B.8 Item stem instructions are precise, direct, clear and simple.	3.80	0.45	SA
B.9 Text economy is observed by reducing word count and simplifying language on item stems.	4.00	0.00	SA
B.10 The item format and layout are logically and aesthetically presented.	3.80	0.45	SA
Grand Mean	**3.68**	**0.49**	**SA**
C. Language Suitability			
C.1 Questions and instructions are clear, easy to comprehend, and grammatically acceptable.	3.60	0.55	SA
C.2 Sentences used in the problems are clear and simple.	3.80	0.45	SA
Grand Mean	**3.70**	**0.50**	**SA**

Table 10
Test B Evaluation Checklist

Criteria	Mean	*SD*	Remark	% Absolute Agreement	Remark	ICC Index[a]	Remark
A. Content	3.67	0.42	SA	56.67	Moderate Agreement	0.40	Poor Reliability
B. Format and Layout	3.68	0.49	SA	53.00	Moderate Agreement	0.53	Moderate Reliability
C. Language Suitability	3.70	0.50	SA	50.00	Moderate Agreement	0.75	Moderate Reliability
Overall	**3.68[b]**	**0.47**	**SA**	**53.22[b]**	**Moderate Agreement**	**0.76**	**Good Reliability**

[a]*Two-way mixed effects, absolute agreement, multiple raters (average measures)*
[b]*Based on mean*

Using percent of absolute agreement, raters had a moderate agreement across all three criteria. Using ICC, there was poor inter-rater reliability as indicated by raters' scores on Content criterion. The poor reliability is due to the score variability where one rater scored some subcomponents with 2 indicating a disagreement. The researcher failed to submit follow-up papers that would explain the content and scope of the test instrument. Raters' scores on Format and Layout and Language Suitability criterion are indicative of moderate inter-rater reliability. Across the three criteria, raters agreed rating Test B by assigning relatively same scores with good reliability index (ICC = 0.76, $p < 0.05$, Cronbach's Alpha = 0.76).

Based on the mean rating, percent of absolute agreement and ICC index from the experts' evaluation, it can be inferred from the overall mean rating that they strongly agreed on the appropriateness of test instrument across three criteria, although there was moderate agreement and good reliability index among the experts' scores as indicated by percent of absolute agreement and ICC; thus, Test B falls under "Acceptable" range. It can be said descriptively that Test B is acceptable, suitable, and face valid.

Test format C. The third item format (Test C) was a modified version of Test A such that the alternatives/ options offered were solution strategies, three incorrect solution strategies and one correct solution strategy. Test C had the same item stem with Test A but different in alternatives/ options. Alternatives of Test A which offered answers to the problem were modified such that they were deconstructed into series of steps necessary to attain that answer.

Table 11
Experts' Evaluation on Test C

Criteria	*M*	*SD*	Remark
A. Content			
A.1 Conforms to DepEd K to 12 Mathematics learning competencies for Junior High School.	3.80	0.45	SA
A.2 The items are level- appropriate. Problems are presented within the level of understanding of learners.	3.40	0.55	SA
A.3 Relates assessment to the content under study and to student capacity.	3.60	0.55	SA
A.4 All problems are directly relevant to the objectives. Match assessment to intended learning objectives.	3.80	0.45	SA
A.5 Uses assessment that can truly reveal whether students understand the learning.	3.40	0.55	SA
A.6 Appropriate answers have been included.	3.80	0.45	SA
Grand Mean	**3.63**	**0.50**	**SA**
B. Format and Layout			
B.1 Questions are progressive in difficulty.	3.60	0.55	SA
B.2 Layout of pages is well organized making the whole test appears interesting and easy to answer.	3.80	0.45	SA
B.3 Appropriate in length for the allotted time.	2.80	0.84	A
B.4 Proper sequence is observed that includes asking questions first to give a direction to the item, and then followed by supporting information. Proper sequence is also observed in ordering answer options logically.	3.40	0.55	SA
B.5 The item format and layout are logically and aesthetically presented.	3.40	0.89	SA
B.6 Alternatives/options cue learners to find the solution to the problem.	3.40	0.55	SA
B.7 Alternatives/options guide learners on the series of steps necessary to solve the problem.	3.40	0.55	SA
B.8 Alternatives/options activate previously stored knowledge.	3.40	0.55	SA
B.9 Alternatives/options motivate learners to find the answer to the problem.	3.40	0.55	SA
B.10 Solution strategies offered as alternatives/options invite learners to find correct solution to the problem.	3.20	0.45	A
Grand Mean	**3.38**	**0.59**	**SA**
C. Language Suitability			
C.1 Questions and instructions are clear, easy to comprehend and grammatically acceptable.	3.40	0.55	SA
C.2 Sentences used in the problems are clear and simple.	3.60	0.55	SA
C.3 Language used in alternatives/options is level appropriate for learners.	3.60	0.55	SA
Grand Mean	**3.53**	**0.55**	**SA**

Table 11 shows the experts' evaluation on Test C. In sum, evaluators strongly agreed on the appropriateness of the test instrument on all of the three criteria. For Test C, Format and Layout and Language Suitability criteria had additional subcomponents

compared to the same criteria on Test A. The additional subcomponents particularly B.6 to B.10 and C.3 were descriptive conditions pertaining to test modifications made for Test C. These subcomponents described the modifications made to manage the intrinsic cognitive load by reducing intrinsic difficulty of items through offering solution strategies. Evaluators strongly agreed that Test C observed the cognitive load reducing technique.

Table 12
Test C Evaluation Checklist

Criteria	Mean	*SD*	Remark	% Absolute Agreement	Remark	ICC Index[a]	Remark
A. Content	3.63	0.50	SA	50.00	Moderate Agreement	0.54	Moderate Reliability
B. Format and Layout	3.38	0.59	SA	41.00	Moderate Agreement	0.80	Good Reliability
C. Language Suitability	3.53	0.55	SA	40.00	Moderate Agreement	0.77	Good Reliability
Overall	**3.52[b]**	**0.55**	**SA**	**43.67[b]**	**Moderate Agreement**	**0.89**	**Good Reliability**

[a]*Two-way mixed effects, absolute agreement, multiple raters (average measures)*
[b]*Based on mean*

The criteria used in evaluating Test C are shown in Table 12. Raters strongly agreed that Test C followed appropriate content, format and layout, and language. Using percent of absolute agreement, raters had a moderate agreement across all three criteria. Using ICC, there was moderate inter-rater reliability as indicated by raters' scores on Content criterion; while raters' scores on Format and Layout and Language Suitability criterion were both indicative of good inter-rater reliability. Across the three criteria, raters agreed rating Test B by assigning relatively same scores with good reliability index (ICC = 0.89, $p < 0.05$, Cronbach's Alpha = 0.90). It can be inferred based on the overall mean rating, percent of absolute agreement, and ICC that Test C

fell within “Acceptable” range with moderate agreement and good reliable ratings from the evaluators, , descriptively interpreted as acceptable, suitable, and face valid.

Overall, the three sets of test instruments made were acceptable, suitable, and face valid based on experts’ rating. Table 13 shows suggestions of the experts that the researcher took constructively to make the face of each Mathematics test instrument more acceptable and suitable.

Table 13
Experts Suggestions and Comments

Test Format	Suggestion/ Comment
Test A	a. Items should be written in question form for consistency. b. Proper spacing should be observed. c. Options should be in ascending order and should also be arranged from shortest to longest. d. Do not italicize the exponents.
Test B	a. For items that have figures, make the measurement closer to the figure. b. Options should be in ascending order and should also be arranged from shortest to longest. c. Do not italicize the exponents.
Test C	a. Itemize the steps in each option of Test C. b. Alternatives/ options appear long and complicated. Appropriate for learners who are reflective. Impatient learners/ test takers may tend to be overwhelmed by long options and may resort to guessing.

The researcher revised the test instruments as suggested by the evaluators to make them more face valid. After the revision, preliminary assessments were conducted through two tryouts with large enough samples of Grade Nine students. All three test instruments were run at the same time during the first and second tryouts.

Psychometric properties of test instruments. In reporting the psychometric properties of each test instrument, quantitative item analysis was used. The *U-L Index Method* (Upper and Lower Group Index) proposed by Stocklein (1957) as cited in Oriondo and Dallo-Antonio (1984) was used to analyze the items. In selecting the

upper and lower group, the upper 27 percent and lower 27 percent rule was used (Kelley, 1939). The reported psychometric properties were difficulty indices, discrimination indices, reliability estimates, and plausibility of options. Difficulty and discrimination indices were calculated using *Microsoft* Excel. Table 14 shows the interpretation of the difficulty index of an item, while Table 15 reflects the interpretation of the discrimination index.

Table 14
Interpretation and Remark on Difficulty Index

Difficulty Index	Interpretation	Remark
0.00 - 0.10	Very Difficult Item	Reject
0.11 - 0.25	Difficult Item	Revise
0.26 - 0.75	Average Item	Retain
0.76 - 0.90	Easy Item	Revise
0.91 - 1.00	Very Easy Item	Reject

Source: Bermundo and Bermundo (2007)

Table 15
Interpretation and Remark on Discrimination Index

Discrimination Index	Interpretation	Remark
≥ 0.40	Very Good Item	Retain
0.30 - 0.39	Good Item	Retain
O.20 - 0.29	Reasonably Good Item	Retain
0.10 - 0.19	Marginal Item	Revise
≤ 0.09	Poor Item	Reject

Source: *Bermundo and Bermundo (2007)*

Table 16 was used as guide in deciding whether to accept, revise, or reject items. Furthermore, since the three tests were measuring the same construct per item, an item could not be rejected unless it was rejected across Tests A, B, and C.

Table 16
Action Table for Accepting, Revising and Rejecting Items based on the Quantitative Item Analysis

Difficulty Index	**Discrimination Index**		
	< 0.00 (Not Discriminating)	**0.00 - 0.19 (Moderately Discriminating)**	**≥ 0.20 (Discriminating)**
≥ 0.76 (Easy)	Reject	Need Revision	May Need Revision
0.26 - 0.75 (Moderately Difficult)	Need Revision	May Need Revision	Accept/ Retain
≤ 0.25 (Difficult)	Reject	Need Revision	May Need Revision

Option analysis was conducted for the plausibility of options as suggested by Padua and Santos (1997). Another psychometric property reported is the reliability estimate or reliability coefficient. Once a single administration of test was done, inter-item consistency was used to calculate the reliability coefficient using KR-20 (Kuder-Richardson formula 20). *Microsoft Excel* was used to calculate KR-20. Table 17 was used to interpret the KR-20 reliability coefficient.

Table 17
Interpretation of KR-20 reliability coefficient

Coefficient	**Interpretation**	**Remark**
≥ 0.90	Excellent	Standard Test
0.80 - 0.89	Very Good	Classroom Test
0.70 - 0.79	Good	Few Items to Improve
0.60 - 0.69	Moderate	Some Items to Improve
0.50 - 0.59	Poor	Revision of the Test
≤ 0.49	Very Poor	Questionable Reliability

Source: Bermundo and Bermundo (2007)

First tryout. On the first tryout, the researcher coordinated with two Mathematics teachers to conduct the tryout of the test instruments (*see* Appendix J). They were given clear instructions and guidelines on how to properly conduct the examination (*see* Appendix K). The researcher instructed them to divide each of their intact classes into six groups or six columns prior the conduct of the examination. The first group received Test A; second group received Test B; and third group received Test C. The fourth, fifth, and sixth groups received Test A, Test B, and Test C respectively. All in all, six sections/ intact classes had been selected for the first tryout. One hundred and two (102) respondents took Test A; one hundred and one (101) respondents took Test B; and one hundred and two (102) respondents took Test C with a total of 305 respondents on the first tryout.

The tryout took only one day to finish. Only the scores of 300 respondents were quantitatively analyzed for item analysis, one hundred per group (Test A, Test B and Test C). Answers of student-respondents were manually scored. To ensure that the manual scoring is reliable, the researcher used *Microsoft Excel* again to electronically score the students' answers and to quantitatively analyze the items. After the quantitative item analysis, test instruments were revised.

As a result, fifteen (15) items were revised for Test A; eleven (11) items for Test B; and fifteen (15) items for Test C (*see* Table 18). No item was rejected across the three tests; hence, test blueprint or table of specifications was not revised; instead, items were revised by either clarifying or improving the item stem or the options.

Table 18
Revisions based on the Results of First tryout

Item No.	Test A	Test B	Test C
4	Item Stem		Item Stem
5	Option D	Option D	Option B
7	Item Stem	Item Stem	Item Stem
10	Item Stem		Item Stem
14	Option B	Option B	Option C
15	Option C	Option C	Option C
20	Option A	Option A	Option A
21	Option D	Option D	Option B
23	Option D	Option D	Option D
26	Option A	Option A	Option C
27	Item Stem		Item Stem
28	Item Stem		Item Stem
32	Item Stem; Option D		Item Stem; Option D
33		Item Stem	
34	Option B	Option B	Option B
35	Option D	Item Stem; Option D	Option C

Table 19 shows the result of the quantitative item analysis on Test A for the first tryout (*see* Appendix L for all detailed analysis on the first tryout). As a result, eighteen (18) items were found to be retained; fourteen (14) items were to be revised; and three (3) items were to be rejected.

Table 20 shows the result of the quantitative item analysis on Test B for the first tryout. Twenty-one (21) items were to be retained; fourteen (14) items were to be revised; and no item was to be rejected.

Table 19
Test A First Tryout Result

Item No.	Difficulty Index	Interpretation/ Remark	Discrimination Index	Interpretation/ Remark	Decision
1	0.59	Average/ Retain	0.74	VG/ Retain	Retain
2	0.57	Average/ Retain	0.63	VG/Retain	Retain
3	0.19	Difficult/ Revise	0.07	PI/Reject	Revise
4	0.20	Difficult/ Revise	0.19	MI/ Revise	Revise
5	0.81	Easy/ Revise	0.30	RG/Retain	Revise
6	0.46	Average/ Retain	0.48	VG/Retain	Retain
7	0.15	Difficult/ Revise	0.15	MI/ Revise	Revise
8	0.69	Average/ Retain	0.33	G/Retain	Retain
9	0.28	Average/ Retain	0.04	PI/Reject	Revise
10	0.13	Difficult/ Revise	-0.04	PI/Reject	Reject
11	0.41	Average/ Retain	0.59	VG/Retain	Retain
12	0.48	Average/ Retain	0.67	VG/Retain	Retain
13	0.35	Average/ Retain	0.19	MI/ Revise	Revise
14	0.15	Difficult/ Revise	0.07	PI/Reject	Revise
15	0.19	Difficult/ Revise	-0.22	PI/Reject	Reject
16	0.46	Average/ Retain	0.41	VG/ Retain	Retain
17	0.54	Average/ Retain	0.70	VG/ Retain	Retain
18	0.28	Average/ Retain	0.56	VG/Retain	Retain
19	0.41	Average/ Retain	0.37	G/ Retain	Retain
20	0.24	Difficult/ Revise	0.41	VG/Retain	Revise
21	0.59	Average/ Retain	0.00	PI/Reject	Revise
22	0.59	Average/ Retain	0.52	VG/Retain	Retain
23	0.19	Difficult/ Revise	0.07	PI/Reject	Revise
24	0.30	Average/ Retain	0.52	VG/Retain	Retain
25	0.35	Average/ Retain	0.48	VG/Retain	Retain
26	0.35	Average/ Retain	-0.11	PI/Reject	Revise
27	0.17	Difficult/ Revise	0.04	PI/Reject	Revise
28	0.15	Difficult/ Revise	-0.07	PI/Reject	Reject
29	0.48	Average/ Retain	0.37	G/Retain	Retain
30	0.35	Average/ Retain	0.41	VG/Retain	Retain
31	0.48	Average/ Retain	0.59	VG/Retain	Retain
32	0.20	Difficult/ Revise	0.04	PI/Reject	Revise
33	0.31	Average/ Retain	0.33	G/Retain	Retain
34	0.13	Difficult/ Revise	0.11	MI/ Revise	Revise
35	0.35	Average/ Retain	0.48	VG/Retain	Retain

Table 20
Test B First Tryout Result

Item No.	Difficulty Index	Interpretation/ Remark	Discrimination Index	Interpretation/ Remark	Decision
1	0.54	Average/ Retain	0.48	VG/Retain	Retain
2	0.65	Average/ Retain	0.56	VG/Retain	Retain
3	0.59	Average/ Retain	0.52	VG/Retain	Retain
4	0.56	Average/ Retain	0.52	VG/Retain	Retain
5	0.96	Very Easy/ Reject	0.07	PI/Reject	Revise
6	0.61	Average/ Retain	0.56	VG/Retain	Retain
7	0.24	Difficult/ Revise	0.11	MI/ Revise	Revise
8	0.59	Average/ Retain	0.37	G/Retain	Retain
9	0.41	Average/ Retain	0.37	G/Retain	Retain
10	0.41	Average/ Retain	0.15	MI/ Revise	Revise
11	0.41	Average/ Retain	0.59	VG/Retain	Retain
12	0.26	Difficult/ Revise	0.37	G/Retain	Revise
13	0.50	Average/ Retain	0.56	VG/Retain	Retain
14	0.09	Very Difficult/ Reject	0.04	PI/Reject	Revise
15	0.22	Difficult/ Revise	0.15	MI/ Revise	Revise
16	0.52	Average/ Retain	0.37	G/Retain	Retain
17	0.54	Average/ Retain	0.70	VG/Retain	Retain
18	0.33	Average/ Retain	0.44	VG/Retain	Retain
19	0.37	Average/ Retain	0.52	VG/Retain	Retain
20	0.37	Average/ Retain	0.07	PI/Reject	Revise
21	0.72	Average/ Retain	0.11	MI/ Revise	Revise
22	0.50	Average/ Retain	0.63	VG/Retain	Retain
23	0.19	Difficult/ Revise	0.22	RG/Retain	Revise
24	0.43	Average/ Retain	0.70	VG/Retain	Retain
25	0.30	Average/ Retain	0.30	RG/Retain	Retain
26	0.13	Difficult/ Revise	0.26	RG/Retain	Revise
27	0.31	Average/ Retain	0.26	RG/Retain	Retain
28	0.33	Average/ Retain	0.30	RG/Retain	Retain
29	0.54	Average/ Retain	0.26	RG/Retain	Retain
30	0.28	Average/ Retain	0.48	VG/Retain	Retain
31	0.31	Average/ Retain	0.33	G/Retain	Retain
32	0.15	Difficult/ Revise	0.15	MI/ Revise	Revise
33	0.20	Difficult/ Revise	0.26	RG/Retain	Revise
34	0.13	Difficult/ Revise	0.11	MI/ Revise	Revise
35	0.24	Difficult/ Revise	0.19	MI/ Revise	Revise

Table 21 shows the result of the quantitative item analysis on Test C for the first tryout. Twenty-three (23) items are to be retained; ten (10) items were to be revised; and two (2) items were to be rejected.

Table 21
Test C First Tryout Result

Item No.	Difficulty Index	Interpretation/ Remark	Discrimination Index	Interpretation/ Remark	Decision
1	0.67	Average/ Retain	0.37	G/Retain	Retain
2	0.63	Average/ Retain	0.67	VG/Retain	Retain
3	0.28	Average/ Retain	0.26	RG/Retain	Retain
4	0.33	Average/ Retain	0.07	PI/Reject	Revise
5	0.74	Average/ Retain	0.37	G/Retain	Retain
6	0.56	Average/ Retain	0.59	VG/Accept	Retain
7	0.17	Difficult/ Revise	0.04	PI/Reject	Revise
8	0.59	Average/ Retain	0.52	VG/Retain	Retain
9	0.35	Average/ Retain	0.41	VG/Retain	Retain
10	0.30	Average/ Retain	0.37	G/Retain	Retain
11	0.46	Average/ Retain	0.70	VG/Retain	Retain
12	0.33	Average/ Retain	0.52	VG/Retain	Retain
13	0.30	Average/ Retain	0.30	RG/Retain	Retain
14	0.26	Difficult/ Revise	-0.07	PI/Reject	Reject
15	0.37	Average/ Retain	0.44	VG/Retain	Retain
16	0.30	Average/ Retain	0.22	RG/Retain	Retain
17	0.35	Average/ Retain	0.33	G/Retain	Retain
18	0.33	Average/ Retain	0.59	VG/Retain	Retain
19	0.39	Average/ Retain	0.56	VG/Retain	Retain
20	0.41	Average/ Retain	0.52	VG/Retain	Retain
21	0.65	Average/ Retain	0.63	VG/Retain	Retain
22	0.44	Average/ Retain	0.52	VG/Retain	Retain
23	0.22	Difficult/ Revise	0.15	MI/ Revise	Revise
24	0.59	Average/ Retain	0.22	RG/Retain	Retain
25	0.43	Average/ Retain	0.26	RG/Retain	Retain
26	0.07	Very Difficult/ Reject	-0.07	PI/Reject	Reject
27	0.31	Average/ Retain	0.04	PI/Reject	Revise
28	0.31	Average/ Retain	0.19	MI/ Revise	Revise
29	0.43	Average/ Retain	0.26	RG/Accept	Retain
30	0.24	Difficult/ Revise	0.04	PI/Reject	Revise
31	0.28	Average/ Retain	0.41	VG/Retain	Retain
32	0.24	Difficult/ Revise	0.19	MI/ Revise	Revise
33	0.30	Average/ Retain	0.15	MI/ Revise	Revise
34	0.19	Difficult/ Revise	0.15	MI/ Revise	Revise
35	0.22	Difficult/ Revise	0.37	G/Retain	Revise

Since the three tests are measuring the same content or construct per item, an item was accepted or retained if the decisions across the results on the item analysis of Test A, B ,and C were all “Retain”, consistently adhering to the decision rule that an item could not be rejected unless it is rejected across the three tests. Otherwise, any deviations from such rule called for item revision, strictly enforcing the decision rule.

Table 22

Summary of the Results of the First Tryout and Action Taken in Item Retention, Rejection, and Revision

Item No.	Test A Decision	Test B Decision	Test C Decision	Action
1	Retain	Retain	Retain	**Retained**
2	Retain	Retain	Retain	**Retained**
3	Revise	Retain	Retain	**Retained**
4	Revise	Retain	Revise	**Revised**
5	Revise	Revise	Retain	**Revised**
6	Retain	Retain	Retain	**Retained**
7	Revise	Revise	Revise	**Revised**
8	Retain	Retain	Retain	**Retained**
9	Revise	Retain	Retain	**Retained**
10	Reject	Revise	Retain	**Revised**
11	Retain	Retain	Retain	**Retained**
12	Retain	Revise	Retain	**Retained**
13	Revise	Retain	Retain	**Retained**
14	Revise	Revise	Reject	**Revised**
15	Reject	Revise	Retain	**Revised**
16	Retain	Retain	Retain	**Retained**
17	Retain	Retain	Retain	**Retained**
18	Retain	Retain	Retain	**Retained**
19	Retain	Retain	Retain	**Retained**
20	Revise	Revise	Retain	**Revised**
21	Revise	Revise	Retain	**Revised**
22	Retain	Retain	Retain	**Retained**
23	Revise	Revise	Revise	**Revised**
24	Retain	Retain	Retain	**Retained**
25	Retain	Retain	Retain	**Retained**
26	Revise	Revise	Reject	**Revised**
27	Revise	Retain	Revise	**Revised**
28	Reject	Retain	Revise	**Revised**
29	Retain	Retain	Retain	**Retained**
30	Retain	Retain	Revise	**Retained**
31	Retain	Retain	Retain	**Retained**
32	Revise	Revise	Revise	**Revised**
33	Retain	Revise	Revise	**Revised**
34	Revise	Revise	Revise	**Revised**
35	Retain	Revise	Revise	**Revised**

Table 22 summarizes the decisions and action taken in accepting/ retaining, rejecting and revising an item.

Overall, no item was rejected and sixteen (16) items were revised by either improving the item stem or revising the options. In revising the item stem, difficulty and discrimination indices were considered; and in revising the alternatives/ options, discrimination index and option analysis were considered (*see* Appendix M for detailed option analysis of first tryout).

Second tryout. For the second tryout, the researcher coordinated with three Mathematics teachers for field-testing, requiring a much larger number of respondents. Purposively, the second tryout was conducted to determine the improvement of the revised items in terms of difficulty and discrimination indices and to check if the retained items have been consistently good in terms of these two indices. As with the first tryout, the same instructions were given to the teachers. All in all, nine (9) sections/ intact classes were selected with 153 respondents taking Test A; 154 respondents taking Test B; and 151 respondents taking Test C with a total of 458 respondents. The tryout took only one day to finish. Only the scores of 450 respondents were gathered for quantitative item analysis, 150 each for the three tests. Answer sheets were scored manually and electronically.

Table 23
Revisions based on the Results of Second tryout

Item No.	Test A	Test B	Test C
5	Item Stem		Item Stem
9	Option A	Option A	Option A
10	Item Stem		Item Stem
28		Item Stem	

Table 23 shows the number and placement of the items revised as the result of the quantitative item analysis of the second tryout. Three (3) items were revised in Test A; two (2) items for Test B; and three (3) items for Test C. No item was rejected across the three tests; as such, test blueprint or table of specifications was not revised. Since none of the items across the three tests was rejected, reliability estimates for each test were calculated.

Results on the second tryout are summarized and presented on the following tables. Table 24 shows the result of the quantitative item analysis on Test A for the second tryout (*see* Appendix N for all detailed analysis on the second tryout). As a result, thirty-two (32) items were found to be retained; three (3) items were to be revised; and no item was to be rejected. In Test A, 33 items marked moderate difficulty; 1 item marked as easy; and 1 item marked as difficult. Additionally, in terms of the discriminating power of test, 32 items were marked discriminating; and 3 were marked moderately discriminating.

Table 25 shows the result of the item analysis on Test B for the second tryout. Thirty-three (33) items were to be retained; two (2) items were to be revised; and no item was found to be rejected. In Test B, 32 items were marked moderately difficult; 1 was marked easy; and 2 were marked as difficult. All items in Test B were discriminating.

Table 24
Test A Second Tryout Result

Item No.	Difficulty Index	Interpretation/ Remark	Discrimination Index	Interpretation/ Remark	Decision
1	0.66	Average/ Retain	0.34	G/Retain	Retain
2	0.56	Average/ Retain	0.49	VG/Retain	Retain
3	0.27	Average/ Retain	0.39	G/Retain	Retain
4	0.28	Average/ Retain	0.27	RG/Retain	Retain
5	0.84	Easy/ Revise	0.17	MI/ Revise	Revise
6	0.73	Average/ Retain	0.39	G/Retain	Retain
7	0.32	Average/ Retain	0.39	G/Retain	Retain
8	0.65	Average/ Retain	0.27	RG/Retain	Retain
9	0.50	Average/ Retain	0.17	MI/ Revise	Revise
10	0.28	Average/ Retain	0.17	MI/ Revise	Revise
11	0.49	Average/ Retain	0.59	VG/Retain	Retain
12	0.34	Average/ Retain	0.49	VG/Retain	Retain
13	0.35	Average/ Retain	0.41	VG/Retain	Retain
14	0.33	Average/ Retain	0.37	G/Retain	Retain
15	0.37	Average/ Retain	0.49	VG/Retain	Retain
16	0.62	Average/ Retain	0.22	RG/Retain	Retain
17	0.67	Average/ Retain	0.61	VG/Retain	Retain
18	0.52	Average/ Retain	0.80	VG/Retain	Retain
19	0.55	Average/ Retain	0.66	VG/Retain	Retain
20	0.38	Average/ Retain	0.22	RG/Retain	Retain
21	0.70	Average/ Retain	0.32	G/Retain	Retain
22	0.67	Average/ Retain	0.56	VG/Retain	Retain
23	0.24	Difficult/ Revise	0.39	G/Retain	Retain
24	0.48	Average/ Retain	0.76	VG/Retain	Retain
25	0.57	Average/ Retain	0.61	VG/Retain	Retain
26	0.40	Average/ Retain	0.41	VG/Retain	Retain
27	0.34	Average/ Retain	0.59	VG/Retain	Retain
28	0.28	Average/ Retain	0.32	G/Retain	Retain
29	0.71	Average/ Retain	0.44	VG/Retain	Retain
30	0.38	Average/ Retain	0.56	VG/Retain	Retain
31	0.48	Average/ Retain	0.56	VG/Retain	Retain
32	0.34	Average/ Retain	0.44	VG/Retain	Retain
33	0.33	Average/ Retain	0.56	VG/Retain	Retain
34	0.34	Average/ Retain	0.54	VG/Retain	Retain
35	0.46	Average/ Retain	0.49	VG/Retain	Retain

Table 25
Test B Second Tryout Result

Item No.	Difficulty Index	Interpretation/ Remark	Discrimination Index	Interpretation/ Remark	Decision
1	0.67	Average/ Retain	0.46	VG/Retain	Retain
2	0.59	Average/ Retain	0.29	RG/Retain	Retain
3	0.37	Average/ Retain	0.59	VG/Retain	Retain
4	0.43	Average/ Retain	0.41	VG/Retain	Retain
5	0.73	Average/ Retain	0.49	VG/Retain	Retain
6	0.52	Average/ Retain	0.66	VG/Retain	Retain
7	0.30	Average/ Retain	0.41	VG/Retain	Retain
8	0.54	Average/ Retain	0.63	VG/Retain	Retain
9	0.24	Difficult/ Revise	0.39	G/Retain	Revise
10	0.35	Average/ Retain	0.46	VG/Retain	Retain
11	0.49	Average/ Retain	0.63	VG/Retain	Retain
12	0.38	Average/ Retain	0.46	VG/Retain	Retain
13	0.49	Average/ Retain	0.68	VG/Retain	Retain
14	0.49	Average/ Retain	0.44	VG/Retain	Retain
15	0.49	Average/ Retain	0.29	RG/Retain	Retain
16	0.61	Average/ Retain	0.34	G/Retain	Retain
17	0.73	Average/ Retain	0.39	G/Retain	Retain
18	0.62	Average/ Retain	0.46	VG/Retain	Retain
19	0.67	Average/ Retain	0.41	VG/Retain	Retain
20	0.41	Average/ Retain	0.29	RG/Retain	Retain
21	0.85	Easy/ Revise	0.24	RG/Retain	Retain
22	0.61	Average/ Retain	0.63	VG/Retain	Retain
23	0.33	Average/ Retain	0.56	VG/Retain	Retain
24	0.57	Average/ Retain	0.66	VG/Retain	Retain
25	0.60	Average/ Retain	0.56	VG/Retain	Retain
26	0.48	Average/ Retain	0.51	VG/Retain	Retain
27	0.57	Average/ Retain	0.61	VG/Retain	Retain
28	0.26	Difficult/ Revise	0.37	G/Retain	Revise
29	0.73	Average/ Retain	0.39	G/Retain	Retain
30	0.55	Average/ Retain	0.76	VG/Retain	Retain
31	0.46	Average/ Retain	0.39	G/Retain	Retain
32	0.34	Average/ Retain	0.39	G/Retain	Retain
33	0.33	Average/ Retain	0.56	VG/Retain	Retain
34	0.49	Average/ Retain	0.78	VG/Retain	Retain
35	0.35	Average/ Retain	0.46	VG/Retain	Retain

For Test C second tryout results, thirty-three (33) items were found to be retained; two (2) items were to be revised; and no item was to be rejected (*see* Table

26). In terms of difficulty, 34 items were marked moderately difficult and 1 was marked easy. In addition, 34 items were marked as discriminating and 1 was marked moderately discriminating.

Table 26
Test C Second Tryout Result

Item No.	Difficulty Index	Interpretation/ Remark	Discrimination Index	Interpretation/ Remark	Decision
1	0.76	Average/ Retain	0.24	RG/Retain	Retain
2	0.57	Average/ Retain	0.41	VG/Retain	Retain
3	0.30	Average/ Retain	0.41	VG/Retain	Retain
4	0.34	Average/ Retain	0.34	G/Retain	Retain
5	0.82	Easy/ Revise	0.22	RG/Retain	Revise
6	0.65	Average/ Retain	0.51	VG/Retain	Retain
7	0.37	Average/ Retain	0.34	G/Retain	Retain
8	0.55	Average/ Retain	0.51	VG/Retain	Retain
9	0.38	Average/ Retain	0.27	RG/Retain	Retain
10	0.43	Average/ Retain	0.17	MI/ Revise	Revise
11	0.57	Average/ Retain	0.32	G/Retain	Retain
12	0.41	Average/ Retain	0.54	VG/Retain	Retain
13	0.48	Average/ Retain	0.46	VG/Retain	Retain
14	0.43	Average/ Retain	0.61	VG/Retain	Retain
15	0.37	Average/ Retain	0.39	G/Retain	Retain
16	0.40	Average/ Retain	0.46	VG/Retain	Retain
17	0.63	Average/ Retain	0.44	VG/Retain	Retain
18	0.59	Average/ Retain	0.54	VG/Retain	Retain
19	0.66	Average/ Retain	0.34	G/Retain	Retain
20	0.48	Average/ Retain	0.51	VG/Retain	Retain
21	0.57	Average/ Retain	0.66	VG/Retain	Retain
22	0.66	Average/ Retain	0.59	VG/Retain	Retain
23	0.37	Average/ Retain	0.44	VG/Retain	Retain
24	0.60	Average/ Retain	0.51	VG/Retain	Retain
25	0.43	Average/ Retain	0.46	VG/Retain	Retain
26	0.35	Average/ Retain	0.61	VG/Retain	Retain
27	0.55	Average/ Retain	0.61	VG/Retain	Retain
28	0.45	Average/ Retain	0.32	G/Retain	Retain
29	0.51	Average/ Retain	0.63	VG/Retain	Retain
30	0.45	Average/ Retain	0.51	VG/Retain	Retain
31	0.51	Average/ Retain	0.73	VG/Retain	Retain
32	0.38	Average/ Retain	0.37	G/Retain	Retain
33	0.29	Average/ Retain	0.39	G/Retain	Retain
34	0.37	Average/ Retain	0.44	VG/Retain	Retain
35	0.37	Average/ Retain	0.44	VG/Retain	Retain

In summary, four (4) items were revised and no item was rejected based on the second tryout results (*see* Table 27).

Table 27
Summary of the Results of the Second Tryout and Action Taken in Item Retention, Rejection, and Revision

Item No.	Test A Decision	Test B Decision	Test C Decision	Action
1	Retain	Retain	Retain	Retained
2	Retain	Retain	Retain	Retained
3	Retain	Retain	Retain	Retained
4	Retain	Retain	Retain	Retained
5	Revise	Retain	Revise	Revised
6	Retain	Retain	Retain	Retained
7	Retain	Retain	Retain	Retained
8	Retain	Retain	Retain	Retained
9	Revise	Revise	Retain	Revised
10	Revise	Retain	Revise	Revised
11	Retain	Retain	Retain	Retained
12	Retain	Retain	Retain	Retained
13	Retain	Retain	Retain	Retained
14	Retain	Retain	Retain	Retained
15	Retain	Retain	Retain	Retained
16	Retain	Retain	Retain	Retained
17	Retain	Retain	Retain	Retained
18	Retain	Retain	Retain	Retained
19	Retain	Retain	Retain	Retained
20	Retain	Retain	Retain	Retained
21	Retain	Retain	Retain	Retained
22	Retain	Retain	Retain	Retained
23	Retain	Retain	Retain	Retained
24	Retain	Retain	Retain	Retained
25	Retain	Retain	Retain	Retained
26	Retain	Retain	Retain	Retained
27	Retain	Retain	Retain	Retained
28	Retain	Revise	Retain	Revised
29	Retain	Retain	Retain	Retained
30	Retain	Retain	Retain	Retained
31	Retain	Retain	Retain	Retained
32	Retain	Retain	Retain	Retained
33	Retain	Retain	Retain	Retained
34	Retain	Retain	Retain	Retained
35	Retain	Retain	Retain	Retained

Table 23 shows the details of item revisions made as a result of the item analysis on the second tryout (*see* also Appendix O for details on option analysis). There was improvement on difficulty indices and discriminating indices of items on the second tryout compared to the first tryout. Since no item was rejected, the test blueprint or table of specifications was maintained. Consequently, since fewer items were revised, the researcher decided to get the reliability estimate/ coefficient of each test instrument.

In summary, all three test instruments exhibited acceptable psychometric properties in terms of difficulty and discrimination indices in the second tryout.

Table 28
Reliability Estimates on Each Test Instrument

	Statistic	*N*	*k*	*M*	*SD*	σ^2	*R*	Interpretation	Remark
Test Format									
Test A	KR-20	150	35	15.75	6.48	42.03	0.84	Very Good	Classroom Test
Test B	KR-20	150	35	17.07	6.83	46.67	0.86	Very Good	Classroom Test
Test C	KR-20	150	35	16.41	6.58	43.32	0.84	Very Good	Classroom Test

Note: N = No. of students; k = No. of items in the test; r = reliability coefficient

Table 28 shows the reliability estimates as reflected by inter-item reliability. As presented, all three tests marked a very good reliability index, which means that each test had a high inter-item reliability and that they could be used as a good classroom test.

Overall, the three test instruments developed where the two were modifications of the traditional format were evaluated as valid and reliable.

Data Gathering Procedure

Administration of test instruments: exploring the effectiveness of reducing the cognitive loads of mathematics test items. To investigate whether the variation of test formats effects change in the Mathematics performance as a function of students' ability, a 2 x 3 between-subjects factorial design was implemented. One hundred and twenty (120) purposive samples of Grade Nine high school students were randomly assigned to six (6) treatment conditions. A sample of twenty (20) participants for each treatment condition was selected and subjected to the experimentation. To ensure that each grouping made from the two ability groups (low-achieving and high-achieving students) were statistically equivalent, a systematic random assignment was used. Since the study was conducted and implemented in intact classes, problems that were encountered were on the assignment of participants, both low-achieving and high-achieving students (**subjects**), to each of the six experimental conditions (assigning each participant to the three types of assessment), and on the access to the test instruments of those students who were picked to be part of the experimentation but whose data were not subjected to data analysis (**non-subjects**).

To address the two problems, the researcher first secured permission from school administrators (i.e. the Schools Division Superintendent, Principal, and Mathematics Head Teacher) by writing them a letter requesting to conduct the study. An informed consent addressed to parents was also given to all students whose sections/ intact classes were selected for the conduct of the experimentation. After all ethical concerns were addressed, arrangements were made prior the actual data gathering.

The researcher met with the teachers and with the intact classes, and communicated the seating arrangement during the testing period. The seating arrangement was prepared so that the systematic random assignment of “subjects” were organized in a manner that the researcher had full knowledge whether the subject was receiving the right type of test assigned to him/ her. Communication of the seating arrangement during the testing period and reminders on the conduct of the test proper were done during their Mathematics classes approximately 30 minutes for each participating intact classes. This was done for duration of two days prior to testing/ experimentation.

Second potential ethical concern addressed was the problem of access of “non-subjects” to the test instruments. To avoid any questions and confusions on the part of non-subjects, they were given the privilege to take any type of Mathematics assessment depending on the seating arrangement prepared prior to the conduct of testing; but as already mentioned, their scores in the test were not extracted for data analysis.

Three hundred and forty-one (341) respondents (subjects and non-subjects) were given informed consents/ parental consents (*see* Appendix A). A day before the scheduled date of examination, the researcher collected the informed consents. Three hundred and thirty-three (333) respondents consented to be part of the experiment. All of the subjects (120) consented to be part of the experimentation. In the administration of the test instruments, each participating intact classes was given a day with the same scheduled date for the conduct of testing. The researcher and one Mathematics teacher conducted the administration of the tests or experimentation to the seven (7) sections/ intact classes (*see* Appendix J). Six out of seven were assigned to the researcher and

one out of seven was assigned to the teacher. The same examination procedures in the tryouts were observed for the experimentation. Strict implementation on the seating arrangement was observed. In answering the test instruments, participants were given a duration of 50 minutes to finish the examination. Before answering the test, a 5-minute reminder on the test instructions was given. Overall, the administration of test instruments for each participating intact class lasted for 55 minutes. The researcher communicated the date and schedules of participating intact classes to the Mathematics teachers based on each intact class's schedule specifically coinciding during their Mathematics class schedule only. This phase of the study covered a duration of one day only so as not to interrupt classes.

After the participants have completed answering the test instruments, follow-up interviews were conducted six or seven days after they took the examination. Twenty-four (24) out of the 120 subjects were interviewed: twelve males and twelve females. Eight subjects were interviewed per test (Test A, Test B, and Test C) with the following breakdown: 2 females and 2 males of high- achieving group (HA); and 2 females and 2 males of low- achieving group (LA). A semi-structured interview was conducted. All conversations were audio recorded and transcribed. Each subject was coded for anonymity and confidentiality. They were coded based on which ability group they belonged (ex. H for high- achieving group and L for low- achieving group); sequence number (ex. 1, 2, 3 or 4); and which test they took (ex. A for Test A, B for Test B, and C for Test C). The main objective of the interview was to know their perceptions on the different item formats in terms of their test preference.

Specifically, the questions asked during the interview were:

1. Which set of test did you take?
2. How do you remember that test?
3. How do you find that test in terms of difficulty?
4. Which of these tests do you prefer most or comfortable with to answer? Why?

Six (6) out of thirty-five (35) items were selected to represent each test. Each test was shown to the subjects. They were given ample time to recognize the similarities and differences of each test in terms of the item stem (question) and alternatives/ options (answers) before asking them question number 4.

Data Analysis

Before the actual administration of test instruments, a preliminary data analysis was conducted to know if the groupings made to both low-achieving and high-achieving groups achieved a statistical equivalence in terms of general average/ grade point average. Descriptive statistics such as arithmetic means and standard deviations quantified the general average for both levels of ability groups across the different test formats. One-way analysis of variance interpreted the results of the general average of both levels of ability groups.

To show that test format and mathematics ability influence Mathematics performance of high school students, descriptive statistics such as arithmetic mean and standard deviation were used to interpret the differences on test scores for all treatment

conditions. To analyze if there exists significant main effects and interaction effect on the Mathematics performance of high school students of each independent variable or factor (ability level of students and test format), a two-way analysis of variance (ANOVA II) was used. Two-way ANOVA is appropriate when there are two or more factors being studied. It is used to determine whether the effects of two factors on a dependent variable are significant or not (main effects), and in addition, interaction effect can be tested (Bluman, 2012; Moore, Mccabe & Craig, 2009). A post hoc analysis was conducted when main effect for test format yielded significant results. Tukey's HSD test (Honestly Significant Difference) was used to test all pair wise comparisons among means. It is an appropriate post hoc test to use when all the groups have the same sample size (Bluman, 2012). Power calculation was also done to detect a true difference, which means getting a significant *p-value* (Moore, Mccabe & Craig, 2009). All of the above mentioned statistical techniques/ treatments were calculated using *SPSS*. The researcher used an alpha level of 0.05 for all statistical tests.

To address the research question pertaining to students' test preference, a qualitative data analysis was conducted. The researcher used interview method to get the perception of student-subjects on which type of test or test format they prefer most to answer. Interviews were audio recorded and transcribed (*see* Appendix P for the transcription of the interviews). In qualitatively analyzing the data, frequency count was used based on the emerging themes.

Ethical Considerations

Everyday, an educator is faced with the problem of how his/her learners will be able to learn most. This has always been an enigma not only for a classroom teacher but also for school administrators and for public in general. As such, countless educational researches had been conducted all for “learning”. In pursuit of higher achievement of learning, educational researchers find means to understand the complex ways of how students construct knowledge, give meaning to it, store it and use it. Henceforth, to fully comprehend this complexity, they have been driven to the root of this mystery, the central focus of all educational efforts to achieve learning-- that is, the learners. Most of the educational researches involved students as their subjects. To fully understand how students learn, testing learning theories to students as subjects is the most reasonable way an educational researcher would do. Studies conducted in the field of social sciences used human subjects. More so, it cannot be denied that educational experiments are more liable and responsible for ethical grounds since it is experimenting with children, who for ethical and moral reasons, must not be subjected to situations that may harm them. As Singh (2006) stated, any researcher studying human subjects has certain responsibilities towards them. More so, a bigger responsibility is tasked to any educational researcher because her/his human sample subjects are students whose rights are not yet fully known to them. Indeed, it is a delicate subject that the researcher must deal with utmost sincerity and confidentiality.

The present study aimed to determine the effect of reducing the cognitive load of test items on the Mathematics performance of Grade Nine high school students.

Specifically, it sought to determine how varying the test formats (i.e. traditional, modified *item stem* with essential information, and item stem with modified *alternatives/options* as solution strategies) in a multiple-choice type test affects the performance of low-achieving and high-achieving high school students. As such, a large number of student-participants were selected in this research. As a general rule in any social science research, the researcher valued the rights of the participants selected in the research study.

Two ethical issues were addressed in the research. First is the selection of both low-achieving and high-achieving high school students. Both groups of learners were selected based on their typicality, in this case with regard to their Mathematics ability level. Low-achieving students were selected based on their grade point averages of 79 or below; while high-achieving students were selected with grade point averages of 90 and above. An *informed consent* (*see* Appendix A) (Cohen et al., 2007; Creswell, 2014; Crowl, 1993) was communicated through the participants' parents/guardians. Since the participants were minors (ages 14-16 years old), parent consent was secured. Parents/guardians as well as the student-participants were fully informed of the aims and the possible consequences of the study as stated explicitly in the informed consent document (Moore et al., 2009). Strict implementation of confidentiality and anonymity of each participating individual was observed and secured.

Before the actual data gathering, that is, the administration of the three test instruments, the researcher first secured permits from the school administrators to conduct the study in the researcher's selected school. Afterwards, the researcher

communicated the aims of the study to the Grade 9 Mathematics teachers and requested permission on the records of students, specifically on the performances in the previous grading periods to be able to identify the samples of low and high achieving students in the subject area Mathematics as potential subjects. Sixty (60) low-achieving and sixty (60) high-achieving students were involved in the study. Separate alphabetical lists of low-achieving and high-achieving student-subjects according to sex were made regardless of their general averages. A systematic random assignment was used to assign each low-achieving and high-achieving student to the three types of assessment. Hence, three groups for low-achieving students and three groups for high-achieving students were formed. These groupings comprised the six treatment/experimental conditions of the study.

Another ethical issue dealt with was the accessibility of non-subjects to the test instruments. It could be recalled that the study was conducted in intact classes where the subjects could not be pulled out of their classes, and the test schedule required them to take the tests in a complete different setting, deviating from their normal class schedules. This schedule meant that subjects taking the test were inside their classrooms, while others were not, a set-up deemed to create confusion on the part of non-subjects. The researcher therefore acknowledged the right and privilege of non-subjects to take any type of Mathematics assessment depending on the seating arrangement prepared prior to the conduct of testing; but their scores in the test were not gathered for data analysis. All participating students whether subject or non-subject were given informed consent/parent consent prior to the administration of the test

instruments. The high percentage of students who participated in the study was due to the researchers' communication to their Mathematics teachers of incentives in terms of additional class participation points.

The researcher informed participating students of the aims of the experimentation. Aside from the informed consent given to the students, the researcher, upon finalizing the approved schedules, personally communicated the aims, risks, and benefits of being in the study. They were also informed that the data gathered were confidential, anonymous and were kept secured and maintained as suggested by Moore et al. (2009). In cases when a few students have not consented to participate or refused to take the test, they were allowed by the researcher to do so. Their decision to decline participation was taken with utmost sincerity and respect (Sekaran, 2003; Singh, 2006). As a social researcher, it is the duty of the researcher to preserve the subjects' self-worth as human beings, protect their dignity, and guard any invasion to their privacy.

CHAPTER 3

RESULTS AND DISCUSSION

The research aimed to determine the effect of reducing the cognitive load of test items on the Mathematics performance of Grade Nine high school students. Specifically, it sought to determine how varying the test formats (i.e. traditional-Test A, modified *item stem* with essential information-Test B, and item stem with modified *alternatives/ options* as solution strategies-Test C) in a multiple-choice type test affected the performance of high and low ability high school students. Three Mathematics tests instruments were developed and used. Test B and Test C were modifications of Test A. They were modified by reducing the extraneous cognitive load using cognitive load reducing techniques previously mentioned (i.e. Test B) and artificially reducing/ managing the intrinsic cognitive load by reducing its intrinsic difficulty through presentation of solution strategies (i.e. Test C). These two set of tests were referred to as the *reduced cognitive load mathematics test forms*. The present study hypothesized that acceptable modification of items of an assessment, thereby reducing the different cognitive loads demanded by items without changing the content or constructs being tested, improve their validity for assessing student knowledge, and in turn improving student performance.

As mentioned, a 2 x 3 between-subjects factorial design was implemented in the study. The two factors were ability level (low and high achieving) and test format (Test A, Test B, and Test C), totaling six experimental/ treatment conditions. To achieve

equivalence, the three groupings made to both low- achieving and high-achieving students were matched in terms of general average, as indicated in Table 29 presenting the means and standard deviations of general averages of both levels of ability group across tests A, B, and C.

Table 29
Descriptive Statistics on General Averages of High and Low Achieving Groups across Different Test Formats

Ability Level	Test Format: Test A *N*	Test A Mean	Test A *SD*	Test B *N*	Test B Mean	Test B *SD*	Test C *N*	Test C Mean	Test C *SD*
High-Achieving Group	20	91.50	1.70	20	91.65	1.84	20	91.75	2.02
Low-Achieving Group	20	77.25	1.16	20	77.40	1.14	20	77.85	1.14

Table 30 presents the preliminary analysis on the data gathered on the two ability groups prior to the administration of test instruments.

Table 30
One-way Analysis of Variance on General Average of High and Low Achieving Groups across Different Test Formats

Ability Level	Statistic Test	*df*	Value	*p-value*	Remark
High- Achieving Group	ANOVA I	59	0.09	0.91	Not Significant
Low- Achieving Group	ANOVA I	59	1. 48	0.24	Not Significant

It shows that the three groupings made as to test format to high ability students were matched and are statistically equivalent in terms of general average $F(2, 57) = 0.09$, $p > 0.05$. The same finding was also interpreted for the low ability

students with general average indicating non-significant result, $F(2, 57) = 0.1.48$, $p > 0.05$.

Student Performance

Student Performance refers to the result obtained on the posttest composed of 35-item multiple choice type varying in test format, which was administered to six groups of subjects. The following results summarize the findings of the study to answer the research question: *Do test formats and ability level influence Mathematics performance of high school students?* Table 31 summarizes the differences between the test scores of the six experimental/ treatment conditions.

Table 31
Descriptive Statistics and ANOVA Results on Test Scores by Ability Level and Test Format

Ability Level	Test Format	*M*	*SD*	*N*
HA	A	24.15	5.71	20
	B	27.30	4.05	20
	C	26.65	4.28	20
Composite		26.03	4.86	60
LA	A	10.80	2.76	20
	B	15.45	3.75	20
	C	14.60	2.66	20
Composite		13.62	3.66	60
Total	A	17.48	8.08	40
	B	21.38	7.13	40
	C	20.63	7.04	40
Composite		19.83	7.56	120

Source	*SS*	*df*	*MS*	*F*	*p-value*	Observed Power[a]
Ability Level	4625.21	1	4625.21	288.40	0.00	1.00
Test Format	342.60	2	171.30	10.68	0.00	0.99
Interaction	13.27	2	6.63	0.41	0.66	0.12
Within	1828.25	114	16.04			
Total	53973.00	120				

Note: HA = High Achieving, LA = Low Achieving, A = Test A, B = Test B and C = Test C
[a] Computed using alpha = 0.05

Subjected to a two-way analysis of variance having two levels of Mathematics ability (High and Low Achieving) and three levels of test format (Test A, Test B, and Test C), results yielded a main effect for the students' ability level, $F(1, 114) = 288.40$, $p < 0.05$, such that the mean score was significantly higher for high- achieving students ($M = 26.03$, $SD = 4.86$) than for low- achieving students ($M = 13.62$, $SD = 3.66$) with *power* = 1.00. The main effect of test format yielded also a significant effect, $F(2, 114) = 10.68$, $p < 0.05$, rejecting the null hypothesis that there is no significant difference on the test scores of high school students on different test formats 99% of the time (*power* = .99). The interaction effect (ability level and test format) was non-significant, $F(2, 114) = 0.41$, $p > 0.05$.

A post hoc analysis was conducted using Tukey's HSD for the main effect of test format (*see* Table 32). Results indicated that the mean score of high school students were significantly higher in Test B condition ($M = 21.38$, $SD = 7.13$) than those in Test A condition ($M = 17.48$, SD = 8.08), $p < 0.05$. The mean score was significantly higher in Test C condition ($M = 20.63$, $SD = 7.04$) than that in Test A condition ($M = 17.48$, SD = 8.08), $p < 0.05$. However, pair wise comparison between the mean scores of Test B condition and Test C condition yielded a non-significant result, $p > 0.05$.

Table 32
Comparison of Mean Differences in Test Score by Test Format

Comparison	Estimated Mean Difference	Standard Error of Difference	*p-value*
Test A vs Test B	-3.90	0.90	0.00
Test A vs Test C	-3.15	0.90	0.00
Test B vs Test C	0.76	0.90	0.68

As hypothesized, students who were randomly assigned to the two reduced cognitive load forms performed better than those students who were assigned the form with typical assessment items. In terms of ability level as a factor, it significantly contributed to high test performance of students. High- achieving students performed better than low- achieving students. This result indicates that since high ability students have high germane cognitive load indicated by their proficiency in the targeted construct of the test items, they are likely to perform better than those students who have less germane cognitive load indicated by lack of schema previously constructed, hence resulting to deficiency in schema automation.

Further, regardless of test format (traditional or modified test formats), high-achieving students outperformed low-achieving students. Hence, mathematics ability level influenced test performance such that more proficient students with high germane cognitive load indicated by their capacity to store well-constructed schema and their capacity to automatically retrieve it with less mental effort are likely to perform better than those at-risk/ struggling students who have less germane cognitive load indicated by their less capacity to store schema and their less capacity to automatically retrieve it when needed. Further, high-achieving students are likely to perform better than low-achieving students because they have higher levels of well-organized knowledge structure in their long term memory (Blayney et al., 2010).

Test format as factor also significantly contributed to high test performance of high school students. Modifications made to the test format significantly contributed to high performance of both ability groups. As hypothesized, the reduction of cognitive

loads of test items is particularly effective in increasing the performance of students in assessment. Reduction of the extraneous cognitive load of assessment items (Test B) proved effective in increasing the accessibility of the measured construct, resulting to enhanced test performance. A higher proportion of students answering the items correctly resulted from the students in Test B condition having a better opportunity to exhibit their knowledge as compared to those students in Test A condition assessed with typical assessment items. This finding is supported by the same findings of Elliott et al. (2009), Elliott et al. (2010), and Gillmor et al. (2015) stripping test items of construct-irrelevant information and using cognitive load reducing technique previously mentioned proved effective in increasing the performance of test takers. This is due to the fact that the reduction of extraneous cognitive load of test items addressed the cognitive demands of the test that are intrinsic to the target constructs; hence, inferences made from test results are likely to represent the test taker's actual competence on the constructs.

Similarly, modification of alternatives/ options as solution strategies (Test C) significantly contributed to high test performance of students. Reducing the intrinsic difficulty of test items through offering solution strategies and through student's engagement of finding the correct solution strategy hypothetically proved effective in increasing the accessibility of students to the target mathematics construct that resulted in higher performance of students in Test C condition than Test A condition. This finding is supported by Secolsky et al. (2016) claiming that offering solution strategies lessens the intrinsic difficulty of the test items and improves the test scores of students.

Reducing the intrinsic difficulty of test items through artificially reducing the intrinsic cognitive load by offering solution strategies is particularly effective because solution strategies activated constructed schemas in long term memory, thus extending the working memory capacity limit. Although students need to read carefully each solution strategy, making extraneous cognitive load higher because of additional reading, the *intrinsic nature* of the targeted construct is being emphasized through the series of steps in the solution strategies, consequently lessening other cognitive demands on students while taking the test because all disparate sources of information were integrated and put into one, as such lessening mental effort on the part of test taker. Solution strategies as alternatives in multiple-choice test were made deliberately by the researcher with the intent of scaffolding the test takers' retrieval of schema stored in long term memory. This strategy is particularly effective to low-achieving students because according to the existing reviewed literature, they have deficits in working memory like maintaining information and lacking inhibition.

As indicated, there was no significant difference between the performances of high school students in Test B and Test C conditions. It can be inferred that although there is a difference in the mean scores of students between the two conditions with Test B condition having a higher mean score than Test C condition, it is not statistically significant. Thus, in both cases the modified test formats have higher mean scores. This means that both modifications are helpful in increasing the test performance of high school students.

Results indicated a no interaction effect. This means that all groups of students demonstrated improved performance on the mathematics assessments with modified items. Further, modified tests were generally easier than their original test, yet the tests were still aligned with the mathematics grade-level content and required the same depth of knowledge.

Student's Perception in terms of Test Preference

To answer the research question: *What are the students' perceptions on the different test formats in terms of test preference?*, out of 120 subjects who took the examinations and were assigned to different treatment conditions, twenty-four (24) were interviewed to get their perception on which test format they preferred the most, with 12 came from high- achieving group and another 12 from low- achieving group.

In the qualitative analysis of high- achieving students' perception on which test format they prefer most, three out of twelve (25%) chose Test A; 8 out of 12 (67%) chose Test B; and 1 out of 12 (8%) chose Test C. All of them were able to distinguish the similarities and differences of each test. High- achieving chose Test B over the two other tests because of the item stem instructions as being direct, simple, easy to understand, and with added figure in the item stem. The following were verbatim comments from the interviewed students on why they have chosen Test B:

"*Ahm sir…dito may figure na (referring to Test B) habang dito sa Test A wala pa kaso…parang gagawa ka pa ng figure para ma-solve mo yung …yung…no. 5*."

(Ahm sir…it has figures (referring to Test B) while here in Test A, none…you still need to draw the figure to be able to solve the…the…no. 5.)

"*Eh kasi po dito sa Test A po…yung…yung tanong po sa kanya kailangan mo po siyang…kailangan mo po siyang…magisip, kailangan mo pa siyang i-imagine. Sa Test B hindi na po kasi nakalagay na po dito yung…yung mismong object na…na dine-define sa Test A."*

(Eh here in Test A …the…the question needs to…you need to…analyze, you still have to imagine it. In Test B, it is given here…the needed object…that is being defined in Test A.)

"*Parang mas okay po na straight to the point na po yung …parang wala na pong paligoy-ligoy. Kung ano po yung dapat kunin…paanong yon na po yung kukuhanin."*

(Much better if it is straight to the point that…that it is not roundabout. What is being asked…it should be directly asked.)

"*I take po with Test B. Kasi po as a student, mas gusto ko po doon sa direct and mas simple lang po yung questions and hindi mahihirapan ang estudyante sa pagsagot lalo na sa Mathematics…kasi po yung iba dito more on words…mas okay pa po sa question dito na more on figures (referring to Test B)".*

(I take Test B. Because as a student, I prefer direct and questions are simpler and students are not having difficulty answering tests especially in Mathematics…because some here are more on words…question here is better because it has more on figures (referring to Test B).)

Further, high- achieving students did not prefer Test C because they thought that it was way too complicated because of long text on options. The following are verbatim comments from high- achieving students on Test C:

"Student (*S): Kasi po 'pag dito po sa Test A…mahirap po siya. Mayroon po siyang letters. Dito naman po sa Test C masyado po siyang masakit sa mata kasi po napakaraming given,…pareparehas naman po yung questions pero sa pagpipilian po masyado po siyang nakakalula.*

Researcher (R): In what sense nakakalula?

S: Parang… ang hirap.....tinatamad kasi kami magbasa."

S: Because here in Test A…it is difficult. It has letters. While here in Test C it is an eye soar because of too many given…questions are the same but the options are too complicated.

R: In what sense complicated?

S: Like…it's difficult…we don't like to read.

"Sa Test C po parang ang dami mo pang kailangan basahin tapos parang… mas nakakalito po siya. Tapos ano… kailangan mo po talagang ano… sundin yung method na ano… of yung process kasi halimbawa kung may technique or method para mas mapabilis yung sagot. Pero dito sa Test C parang kailangan mo pa isa-isahin yung mga ano… yung mga choices."

(Here in Test C like you have to read more, then like…it's more confusing. Then…you have to…follow the method of…of the process, what if example you have technique or method to be able to quickly answer it. But here in Test C like you need to go through each of…of the choices.)

For low- achieving students, all of the tests received the same percentage of preference. One-third of the interviewed students preferred Test A; one-third preferred Test B; and one-third preferred Test C. No outstanding preferences were indicated. Low- achieving students who chose Test B commented that the instructions were easy to understand and figures were helpful in solving the problem. The following are verbatim of students' preference for Test B:

"…madali lang po i-analyze po yung instruction."

(…instruction is easy to analyze.)

"Meron po kasi siyang figure na pwedeng makuha mo agad yung…parang mas madali po siyang…makakuha ng…solu…pwede po siyang, mas madali po siyang makuhanan ng ano…solution, kasi may figure po siya."
(Because it has figure that helps you to easily get the…like it's easier to…get the..solu…you can, easily get the…solution because there is figure.)

For both groups, their preference on Test B (reduction of extraneous cognitive load) supports the findings of Roach et al. (2010), as cited in Elliott, Beddow, Kurz and Kettler (2011) that based on cognitive interviews and survey data, students responded in favor of the item developed with a focus on accessibility (modified to reduce the construct-irrelevant information of the test).

Some of low- achieving students' preferred Test C based on how the options presented as solution strategies helped them to answer the problem. The following are some of the students' comments in preference of Test C:

"Kasi po ano…meron po siyang ano…yung paano po siya i-solution tapos yung bawat choices nya, ano pagpipilian na lang po kung paano mo siya iso-solve tapos magiisip ka na lang po kung ano yung tamang sagot doon."

(Because…it has…process of solution then every choice, you just choose from them based on how it is solved then you just analyze which of these is the correct solution.)

"Kasi sir meron na po siyang…kung paano i-solve, tsaka kung paano mo talaga made-determine yung…yung…sagot na kailangan ng tanong."

(Because it already has…way of solving, and it helps you on how you can really determine the…the…answer need by the question.)

"Ano na…may…computation na, nandito na lahat ng ano…sagot. Tapos hindi mo na kailangan mag-solve andito na sir."

(It has…has…already computation, all of the…solutions. Then, you don't need to really solve because it is already here sir.)

Some of them did not prefer Test C because they found it complicated. The long options presented made the test confusing on their part. Here is a comment from one of them for not preferring Test C over the other tests.

“Kasi po yung Test C na nasagutan ko ay…nakakalito kasi halos…halos parehas lang sila ng explanation magkakaiba lang sila ng sagot.”

(Because Test C that I have taken is…confusing because most…most of them have the same explanation but different in answer.)

In sum, high- achieving students preferred Test B than the other two tests. Reduction of extraneous cognitive load specifically simplification of language, making instruction direct and concise, and addition of figure are some of the key features why they preferred such kind of test format. These features were also the same for low-achieving students who preferred Test B over the other two tests. Some of the low-achieving students preferred Test C over the other two tests simply because the offered solution strategies as alternatives/ options cued them on the correct answer. No outstanding test preference for low- achieving students was reported.

CHAPTER 4

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

Summary of the Study

This study aimed to determine the effect of reducing the cognitive load of test items on the Mathematics performance of high school students. In particular, it sought to determine how varying the item formats (i.e. traditional, modified item *stem* with essential information, and item stem with solution strategies as *alternatives/ options*) in a multiple-choice type test affected the performance of high and low ability high school students. Three mathematics test instruments were developed and tested: traditional assessment and two reduced cognitive load mathematics test forms. Face and content validation of the experts and psychometric properties of the test instruments were reported. A 2 x 3 between-subjects factorial design was implemented. Two factors were considered affecting test performance of high school students in Mathematics: ability level and test format. All participants in the study were Grade 9 high school students. Samples of 120 subjects were purposively selected, randomly assigned, and subjected to six treatment conditions to determine the influence of reducing the cognitive loads of mathematics test instruments as function of students' ability level. In addition, the study also probed the students' perception in terms of their test preference.

Summary of Results

The following are the findings of the study:

1. Results indicated that mathematics ability level significantly contributed to high test performance of high school students with high- achieving students outperforming low- achieving students. Further, test format as a factor significantly contributed to high test performance of both ability groups of high school students. A post hoc analysis confirmed that students who were randomly assigned to two reduced cognitive load mathematics test forms (Test B and Test C) outperformed students who were randomly assigned in traditional assessment (Test A). No significant interaction effect for ability level and test format was noted.
2. Results of the interview confirmed that high- achieving students preferred Test B where extraneous cognitive load was reduced by text economization of item stem and through addition of figures. They wanted questions that asked direct to the point, simple and with clear instructions. No outstanding test preference for low-achieving students was observed.

Limitations of the Study

The following were the limitations of the present study:

1. In the selection of high and low ability groups, the study used criterion-referenced. The study falls short in categorizing high school students' mathematics ability level. Stringent categorization of mathematics ability

level should be made for them to qualify as samples for high or low achieving. The study failed to look into other factors such as results of diagnostic test and level of motivation as contributory factors in categorizing high and low ability students.

2. There was a shortcoming on the study's finding on students' perception in terms of test preference. An interview was conducted to get the students' perception on which test format they preferred the most. They were presented with the three test formats and were asked to identify the similarities and differences of each test based on item stem (question/instruction) and alternatives (answers) before being asked of their test preference. Students did not experience answering all the test formats to be able to give a bigger picture of their experience/ perceived difficulty of the tests and to give a more qualitative assessment of their test preference.

Conclusions

Considering the findings of this study, the following conclusions were drawn:

1. *Cognitive loads imposed by test items affect test performance.* Teachers and test developers must take into consideration when designing test the memory/cognitive resources demanded by test items on students' working memory, and on the performance. Cognitive Load Theory can be of guiding principle in the item writing process.

2. *Management of cognitive loads of test improves student performance.* Test items that are stripped of extraneous/ construct-irrelevant information improve student test performance. In addition, scaffolds on assessment like offering solution strategies as part of test improve student test performance. Modifications made to test format with the intent of reducing the working memory demands of items improve test performance of students and consequently, reflections of their understanding and depth of their knowledge in the targeted construct. Students of varying abilities can be assisted to improve their test performance through reduced cognitive load test forms. In Mathematics education context, remedial mathematics class programs should include in their remediation the findings of this study as an essential feature of the program especially in assessing the depth of knowledge of at-risk /struggling students.

Recommendations

The following recommendations are presented for future research work on this field and possible applications of the findings of this study.

1. Since the study used a *between-subjects* design with the intent of getting a large sample of subjects to gain a more analyzable data and generalizable result, for future research work, *repeated measures counter-balanced* design is suggested to determine the effect of reduced cognitive

load test forms. This design may enhance both the external and internal validity of the study.

2. Researcher made test instruments were developed for the purpose of the research. It is recommended that existing achievement and summative tests (i.e. test in the test bank with known reliability index) can be used for future research endeavor. These tests can be modified using existing cognitive load reducing techniques and/or item modifications for students to gain access to the intended construct. This informs test developers and as well as educators of the possible implications of reduced cognitive load test forms on the test performance of students and also to the psychometric properties of the test.

3. It is suggested that summative assessments may be modified to reduce the cognitive loads of test items to ensure access for all students most especially those students who are persistently performing low and in need of remedial class programs. Alternate assessments may be given as part or feature of remedial program.

4. The study included high- achieving and low- achieving students based on their performance in terms of grade. For future research, it is suggested that a stringent categorization may be used like results on National Achievement Test to classify students qualitatively. It is suggested also that normally achieving students be part of the study to get a bigger

picture of how reducing the cognitive load of test items affects performance of students.

REFERENCES

Aditomo, A. (2009). Cognitive Load Theory and Mathematics Learning: A Systematic Review. *Anima, Indonesian Psychological Journal , 24*(3), 207-217.

Alloway, T. P. & Copello, E. (2013). Working memory: The what, the why, and the how. *The Australian Educational and Developmental Psychologist, 30* (2), 105-118. doi:10.1017/edp.2013.13

Alpacion, N. J. D., Camanan, C. J., Gregorio, A. J., Panlaan, J. M. R. & Tudy, R. A. (2014). Attitude, self-efficacy and students' academic performance in mathematics. *IAMURE International Journal of Social Sciences , 12*, 21-34. doi: http://dx.doi.org/10.7718/ijss.v12i1.920

Andaya, O. J. (2014). Factors that affect mathematics achievement of students of Philippine Normal University-Isabela Campus. *Researchers World , 5* (4), 83-91.

Araujo, J.& Born, D. G. (1985). Calculating percentage agreement correctly but writing its formula incorrectly. *The Behavior Analyst, 8*(2), 207-208.

Archer, A. L. & Hughes, C. A. (2011). *Explicit instruction: Effective and efficient teaching.* New York: The Guilford Press.

Baker, S., Gersten, R. & Lee, D. (2002). A synthesis of empirical research on teaching mathematics to low-achieving students. *The Elementary School Journal, 103*, 51-73.

Beddow, P. A., Kurz, A. & Frey, J. R. (2011). Accessibility theory: Guiding the science and prcatice of test item design with the test-taker in mind. In S. N. Elliott, R. J. Kettler, P. A. Beddow, & A. Kurz (Eds.), *Handbook of accessible achievement test for all students: Bridging the gaps between research, practice, and policy* (pp. 163-182). NY, USA: Springer. doi:10.1007/978-1-4419-9356-4.

Bermundo, C. B. & Bermundo, A. B. (2007). *Micro analysis: Test checker and item analyzer with statistics workbook.*

Blayney, P., Kalyuga, S. & Sweller, J. (2010). Interactions between the isolated–interactive elements effect and levels of learner expertise: experimental evidence from an accountancy class. *Instr Sci , 38*, 277-287. doi: 10.1007/s11251-009-9105-x

Bluman, A. G. (2012). *Elementary statistics: A step by step approach, 8th edition.* NY, USA: The McGraw-Hill Companies, Inc.

Brendefur, J., Johnson, E. S., Thiede, K. W., Smith, E. V., Strother, S., Severson, H. H. & Beaulieu, J. (2015). Developing a comprehensive mathematical assessment tool to improve mathematics intervention for at-risk students. *International Journal for Research in Learning Disabilities, 2*(2), 65-90.

Bryant, D. P., Bryant, B. R., & Hammill, D. D. (2000). Characteristic behaviors of students with learning disabilities who have teacher-identifies math weaknesses. *Journal of Learning Disabilities, 33*, 168-199. doi:10.1177/002221940003300205

Bryant, D. P., Bryant, B. R., Gersten, R., Scammacca, N. & Chavez, M. M. (2008). Mathematics intervention for first- and second-grade students with mathematics difficulties: The effects of tier 2 intervention delivered as booster lessons. *Remedial and Special Education, 29*(1), 20-32.

Bryant, D. P., Bryant, B. R., Roberts, G., Vaughn, S., Pfannenstiel, K. H., Porterfield, J., & Gersten, R. (2011). Early numeracy intervention program for first-grade students with mathematics difficulties. *Exceptional Children, 78*(1), 7–23. doi:10.1177/0014

Cavazzuti, M. (2013). Design of experiments. In *Optimization methods: from theory to design, scientific and technological aspects in mechanics* (Vol. 18). Berlin Heidelberg, Germany: Springer-Verlag. doi:10.1007/978-3-642-31187-1_2

Chandler, P. & Sweller, J. (1991). Cognitive Load Theory and the Format of Instruction. *Cognition and Instruction, 8*(4), 293-332. Retrieved from http://ro.uow.edu.au/edupapers/128

Charnessa, G., Gneezy, U. & Kuhn, M. A. . (2012). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization, 81*, 1– 8.

Chinnappan, M. & Chandler, P. A. (2010). Managing cognitive load in the mathematics classroom. *Australian Mathematics Teacher, 66*(1), 5-11.

Clarke, B., Doabler, C., Smolkowski, K., Nelson, E., Fien, H., Baker, S., & Kusty, D. (2016). Testing the immediate and long-term efficacy of a tier 2 kindergarten mathematics intervention. *Journal of Research on Educational Effectiveness*, 1-63. doi:10.1080119345747.2015.1116034.

Clements, D. H., Sarama, J. & Germeroth, C. (2016). Learning executive function and early mathematics: Directions of causal relations. *Early Childhood Research Quarterly, 36*, 79-90. doi:10.1016/j.ecresq.2015.12.00.

Cohen, L., Manion, L. & Morrion, K. (2000). *Research methods in education, 5th ed.* London, United Kingdom: Routledge Falmer.

Cohen, L., Manion, L. & Morrison, K. (2007). *Research methods in education: 6th edition.* United Kingdom: Routledge.

Cohen, R. J. & Swerdlik, M. (2010). *Psychological testing and assessment: An introduction to tests and measurement, 7th edition.* USA: The McGraw−Hill Companies .

Compton, D. L., Fuchs, L. S., Fuchs, D., Lambert, W. & Hamlett, C. (2012). The cognitive and academic profiles of reading and mathematics learning disabilities. *Journal of Learning Disabilities, 45*, 79-95.

Crealock, C., & Kronik, D. (1993). *Children and young people with specific learning disabilities.* Paris: UNESCO.

Creswell, J. W. (2009). *Research design: qualitative, quantitative, and mixed methods approaches, 3rd edition.* United States of America: SAGE Publicatíons, Inc.

Creswell, J. W. (2014). *Research design: qualitative, quantitative, and mixed methods approaches, 4th edition.* United States of America: SAGE Publications, Inc.

Crowl, T. K. (1993). *Fundamentals of educational research .* Dubuque, Iowa, United States of America: Wm. C. Brown Communications, Inc.

Daro, P., Mosher, F. A. & Corcoran, T. B. (2011). *Learning trajectories in mathematics: A foundation for standards, curriculum, assessment, and instruction.* Consortium for Policy Research in Education. Retrieved from http://repository.upenn.edu/cpre_researchreports/60

Dennis, M. S., Sharp, E., Chovanes, J., Thomas, A., Burns, R. M., Custer, B., & Park, J. (2016). A meta-analysis of empirical research on teaching students with mathematics learning difficulties. *Learning Disabilities Research and Practice, 0*(0), 1-13. doi:10.1111/ldrp.12107.

Elliott, S. N., Beddow, P. A., Kurz, A., & Kettler, R. J. (2011). Creating access to instruction and tests of achievement: Challenges and solutions. In S. N. Elliott, R. J. Kettler, P. A. Beddow, & A. Kurz (Eds.), *Handbook of accessible achievement tests for all students: Bridging the gaps between research, practice, and policy* (pp. 1-18). NY, USA: Springer. doi:10.1007/978-1-4419-9356-4.

Elliott, S. N., Kettler, R. J., Beddow, P. A. & Kurz, A. (2011). Accessible test of student achievement: Access and innovations for excellence. In S. N. Elliott, R. J. Kettler, P. A. Beddow, & A. Kurz (Eds.), *Handbook of accessible achievement tests for*

all students: Bridging the gaps between research, practice, and policy (pp. 319-328). NY, USA: Springer. doi:10.1007/978-1-4419-9356-4.

Elliott , S. N., Kurz, A., Beddow , P., & Frey, J. (2009). Cognitive Load Theory: Instruction-based Research with Applications for Designing Tests. *Paper Presented at the National Association of School Psychologists' Annual Convention*, (pp. 1-22). Boston, MA.

Elliott, S. N., Kettler, R. J., Beddow, P. A., Kurz, A., Compton, E., McGrath, D., Bruen, C., Hinton, K., Palmer, P., Rodriguez, M. C., Bolt, D. & Roach, A. T. (2010). Effects of using modified items to test students with persistent academic difficulties. *Exceptional Children , 76*(4), 475-495.

Fraenkel, J. R. & Wallen, N. E. (2008). *How to design and evaluate research in education, 7th ed.* New York, New York, United States of America: McGraw Hill Companies, Inc.

Fuchs, D., Compton, D. L., Fuchs, L. S., Bryant, J., & Davis, G. N. (2008). Making ‘‘secondary intervention’’ work in a three-tier responsiveness-to-intervention model: findings from the first-grade longitudinal reading study of the National Research Center on Learning Disabilities. *Read Writ, 21*(4), 413-436. doi:10.1007/s11145-007-9083-9.

Fuchs, L. S. & Fuchs, D. (2005). Enhancing mathematical problem solving for students with disabilities. *The Journal of Special Education, 39*(1), 45-57.

Fuchs, L. S. & Fuchs, D. (2002). Mathematical problem-solving profiles of students with mathematics and disabilities with and without comorbid reading disabilities. *Journal of Learning Disabilities, 35*, 563–573.

Fuchs, L. S., & Vaughn, S. (2012). Responsiveness-to-intervention: A decade later. *Journal of Learning Disabilities, 45*, 195-203. doi:10.1177/0022219412442150

Fuchs, L. S., Compton, D. L., Fuchs, D., Paulsen, K., Bryant, J. D. & Hamlett, C. L. (2005). The prevention, identification, and cognitive determinants of math difficulty. *Journal Of Educational Psychology, 97*, 493-513.

Fuchs, L. S., Fuchs, D., & Prentice, K. (2004). Responsiveness to mathematical problem-solving instruction: Comparing students at-risk of mathematics disability with and without risk of reading disability. *Journal of Learning Disabilities, 37*(4), 293-306.

Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P., Capizzi, A. & Fletcher, J. M. (2006). The cognitive correlates of third-grade skill in arithmetic,

algorithmic, computation, and arithmetic word problems. *Journal of Educational Psychology, 98*, 29-45.

Fuchs, L. S., Powell, S. R., Seethaler, P. M., Cirino, P. T., Fletcher, J. M., Fuchs, D., & Hamlett, C. L. (2010). The effects of strategic counting instruction, with and without deli berate practice, on number combination skill among students with mathematics difficulties. *Learning and Individual Differences, 20*(2), 89-100. doi:10.1016/j.lindif.2009.09.003

Fuchs, L. S., Powell, S. R., Seethaler, P. M., et. al. (2009). Remediating number combination and word problem deficits among students with mathematics difficulties: A randomized control trial. *Journal of Educational Psychology, 101*, 561-576. doi:10.1037/a0014701

Fuchs, L. S., Schumacher, R. F., Sterba, S. K., Long, J., Namkung, J., Malone, A., et al. (2014). Does working memory moderate the effects of fraction intervention? An aptitude-treatment interaction. *Journal of Educational Psychology, 106*(2), 499-514. Retrieved from http://dx.doi.org/10.1037/a0034341

Fuchs, L. S., Seethaler, P. M., Powell, S. R., Fuchs, D., Hamlett, C. L., & Fletcher, J. M. (2008). Effects of preventative tutoring on the mathematical problem solving of third- grade students with math and reading difficulties. *Exceptional Children, 74*(2), 155–173.

Geary, D. C. & Hoard, M. K. (2005). Learning disabilities in arithmetic and mathematics: Theoretical and empirical perspectives. (J. I. Campbell, Ed.) *Handbook of mathematical cognition*, 253-268.

Geary, D. C. (2004). Mathematics and learning disabilities. *Journal of Learning Disabilities, 37*(1), 4-15. doi:10.1177/00222194040370010201.

Geary, D. C. (2010). Mathematical disabilities: Reflections on cognitive, neuropsychological, and genetic components. *Learning and Individual Differences, 20*(2), 130-133. Retrieved from http://dx.doi.org/10.1016/j.lindif.2009.10.008

Geary, D. C. (2011). Consequences, characteristics, and causes of mathematical learning disabilities and persistent low achievement in mathematics. *Journal of Developmental and Behavioral Pediatrics, 32*(3), 250-263. doi:10.1097/DBP.0b013e318209edef

Geary, D. C. (2013). Learning disabilities in mathematics: Recent advances. (H. L. Swanson, K. Harris, & S. Graham, Eds.) *Handbook of learning disabilities*, 239-256.

Geary, D. C., Hanson, C. O. & Hoard, M. K. (2000). Numerical and arithmetical cognition: A longitudinal study of process and concept deficits in learning disabled children. *Journal of Experimental Child Psychology, 77*, 236-263.

Geary, D. C.,Hoard, M. K., Byrd-Craven, J., Nugent, L., & Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Development, 78*, 1343–1359. doi:10.1111/j.1467-8624.2007.01069.x

Gersten, R., Beckmann, S., Clarke, B., Foegen, A., Marsh, L., Star, J. R. & Witzel, B. (2009). *Assisting students struggling with mathematics: Response to Intervention (RtI) for elementary and middle schools (NCEE 2009-4060).* Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sci ences, U.S. Department of Education. Retrieved from http://ies. Ed.gov/ncee/wwc/publications/practiceguides/

Gersten, R., Chard, D., Jayanthi, M., Baker, S., Morphy, P., & Flojo, J. (2008). *Mathematics instruction for students with learning disabilities or difficulty learning mathematics: A synthesis of the intervention research.* Portsmouth,NH: RMC Research Corporation, Center on Instruction.

Gersten, R., Jordan, N. C. & Flojo, J. R. (2005). Early identification and interventions for students with mathematics difficulties. *Journal of Learning Disabilities, 38*, 293-304.

Gillmor, S. C., Poggio, J, & Embretson, S. (2015). Effects of reducing the cognitive load of mathematics test items on student performance. *Numeracy, 8*(1), 1-18. doi: http://dx.doi.org/10.5038/1936-4660.8.1.4

Graham, M., Milanowski, A. & Miller, J. (2012). *Measuring and Promoting Inter-Rater Agreement of Teacher and Principal Performance Ratings.* Center for Educator Compensation Reform and U.S. Department of Education.

Griffin, B. W. (2011, Spring). *Inter-rater agreement for nominal/categorical ratings.* Retrieved February 16, 2018, from Bryan W. Griffin: http://www.bwgriffin.com/gsu/courses/edur9131/content/InterRater_nominal_codes.pdf

Gwet, K. L. (2014). Intraclass correlation: A measure of raters' agreement. In K. Gwet, *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters, 4th edition* (pp. 185-192). MD, USA: Advanced Analytics, LLC.

Hancock-Niemic, M. A., Lin, L., Atkinson, R. K., Renkl, A. & Wittwer, J. (2016). Example-based learning: exploring the use of matrices and problem variability. *Education Tech Research Dev, 64*, 115–136. doi:10.1007/s11423-015-9403-8

Hanich, L., Jordan, N. C., Kaplan, D. & Dick, J. (2001). Performance across different areas of mathematical cognition in children with learning difficulties. *Journal of Educational Psychology, 93*(3), 615-626.

Hopkins, S. & Egeberg, H. (2009). Retrieval of simple addition facts: Complexities involved in addressing a commonly identified mathematical learning difficulty. *Journal of Learning Disabilities, 42*(3), 215-229. doi:10.1177/0022219408331041.

Ismail, H. N, Kuldas, S. & Hamzah, A. (2013). Do students need more motivational resources or more cognitive resources for better learning? *Procedia - Social and Behavioral Sciences , 97*, 325 – 332. doi:doi: 10.1016/j.sbspro.2013.10.241

Jalani, N. H. & Lai C. S. (2015). The example-problem-based learning model: Applying cognitive load theory. *Procedia - Social and Behavioral Sciences, 195*, 872 – 880. doi:10.1016/j.sbspro.2015.06.366

Jitendra, A., DiPipi, C. M., & Perron-Jones, N. (2002). An exploratory study of schema-based word-problem-solving instruction for middle school students with learning disabilities: An emphasis on conceptual and procedural understanding. *The Journal of Special Education, 36*(1), 23-38.

Johnson, B. & Christensen, L. (2004). *Educational research: quantitative, qualitative, and mixed approaches, 2nd ed.* Boston, Massachusetts, United States of America: Pearson Education, Inc.

Jordan, N. C., Hanich, L. B. & Kaplan, D. (2003). A longitudinal study of mathematical competencies in children with specific mathematics difficulties versus children with comorbid mathematics and reading difficulties. *Child Development, 74*, 834-850.

Jordan, N. C., Kaplan, D. & Hanich, L. B. (2002). Achievement growth in children with learning difficulties in mathematics: Findings of a two-year longitudinal study. *Journal of Educational Psychology, 94*, 586-597.

Kalyuga , S. & Sweller, J. (2004). Measuring Knowledge to Optimize Cognitive Load Factor During Instruction. *Journal of Educational Psychology , 96*(3), 558–568. doi:10.1037/0022-0663.96.3.558

Kalyuga, S., Law, Y. K. & Lee, C. H. (2013). Expertise reversal effect in reading Chinese texts with added causal words. *Instr Sci , 41*, 481–497. doi: 10.1007/s11251-012-9239-0

Kelley, T. L. (1939). The selection of upper and lower groups for the validation of test items. *Journal of Educational Psychology, 30*, 17-24.

Kettler, R. J., Rodriguez, M. C., Bolt, D. M., Elliott, S. N., Beddow, P. A., & Kurz, A. (2011). Modified multiple choice items for alternate assessments: Reliability, difficulty, and differential boost. *Applied Measurement in Education, 24*(3), 210-234. doi:10.1080/08957347.2011.580620

Kirschner, P. A. (2002). Cognitive load theory: implication of cognitive load theory on the design of learning. *Learning and Instruction, 12*, 1-10.

Koo, T. K. & Li, M. Y. (2016). A Guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*, 155–163. doi:10.1016/j.jcm.2016.02.012

Kroesbergen, E. H. & Van Luit, J. E. (2003). Mathematics interventions for children with special educational needs: a meta-analysis. *Remedial & Special Education, 24*(2), 97-114. doi:10.1177/07419325030240020501

Kubiszyn, T. & Borich, G. (2007). *Educational testing and measurement: Classroom application and practice.* NJ, USA: John Wiley & Sons, Inc.

Lambert, K. & Spinath, B. (2014). Do we nee a special intervention program for children with mathematical learning disabilities or is private tutoring sufficient? . *Journal of Educational Research Online, 6*(1), 68-93.

Mazzocco, M. M. M., & Myers, G. F. (2003). Complexities in identifying and defining mathematics learning disability in the primary school age years. *Annals of Dyslexia, 53*, 218–253.

Montague, M., Enders, C., & Dietz, S. (2011). Effects of cognitive strategy instruction on math problem solving of middle school students with learning disabilities. *Learning Disability Quarterly, 34*, 262–272. doi:10.1177/0731948711421762

Montague, M., Krawec, J., Enders, C., & Dietz, S. (2014). The effects of cognitive strategy instruction on math problem solving of middle-school students of varying ability. *Journal of Educational Psychology, 106*, 469–481. doi:10.1037/a0035176

Morgan, P. L., Farkas, G., Hillemeier, M. M., & Maczuga, S. (2014). Who is at risk for persistent mathematics difficulties in the United States? *Journal of Learning Disabilities*, Advanced online publication. doi:10.1177/0022219414553849

Moore, D. S., Mccabe, G. P. & Craig, B. A . (2009). *Introduction to the practice of statistics, sixth edition.* NY, USA: W. H. Freeman and Company.

Murphy, M. M., Mazzocco, M. M. M., Hanich, L. B. & Early, M. C. (2007). Cognitive characteristics of children with mathematics learning disability (MLD) vary as a function of the cutoff criterion used to define MLD. *Journal of Learning Disabilities, 40*(5), 458-478.

Musso, M., Kyndt, E., Cascallar, E., & Dochy, F. (2012). Predicting mathematical performance: The effect of cognitive processes and self-regulation factors. *Hindawi Publishing Corporation Education Research International*, 1-13. doi:10.1155/2012/250719

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the national mathematics advisory panel.* Washington, DC: U.S. Department of Education.

Nizoloman, O. N. (2013). Relationship between mathematical ability and achievement in mathematics among female secondary students in Bayelsa State Nigeria. *Procedia-Socila and Behavioral Sciences , 106*, 2230-2240.

Ngu, B. H. & Phan, H. P. (2016). Unpacking the complexity of linear equations from a cognitive load theory perspective. *Educ Psychol Rev, 28*, 95–118. doi:10.1007/s10648-015-9298-2

Ngu, B. H., Yeung, A. S., & Tobias, S. (2014). Cognitive load in percentage change problems: unitary, pictorial, and equation approaches to instruction. *Instr Sci, 42*, 685–713. doi:10.1007/s11251-014-9309-6

Ocampo, C. A., De Mesa, D. M. B., Ole, A. F., Auditor, E., Morales, M. P. E., Sia, S. R. D., & Palomar, B. C. (2015). Development and evaluation of physics microlab (P6-ulab) kit. *The Normal Lights, 9*(1), 134-158.

Oehlert, G. W. (2010). *A first course in design and analysis of experiments.* Creative Commons. Retrieved from http://creativecommons.org/licenses/by-nc-nd/3.0/

Oriondo, L. L. & Dallo-Antonio, E. M. (1984). *Evaluating educational outcomes: Tests, measurement and evaluation.* Manila, Philippines: Rex Bookstore, Inc.

Paas, F., Renkl, A. & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologists, 38*(1), 1-4.

Padua, R. N. & Santos, R. G. (1997). *Educational evaluation and measurement: Theory, practice and application.* Manila, Philippines: KATHA Publishing Co., Inc.

Powell, S. R., Fuchs, L. S., Fuchs D., Cirino, P., & Fletcher, J. M. (2009). Effects of fact retrieval tutoring on third-grade students with math difficulties with and without reading difficulties. *Learning Disabilities Research and Practice, 24*, 1-11. Retrieved from [PubMed: 19448840]

Rao, V. (2017). Problem solving in mathematics: role of worked examples in reducing cognitive load and improving scholastic performance. *International Journal of Indian Psychology, 4* (2), 172-178 . doi:18.01.120/20170402

Re, A. M., Pedron, M., Tressoldi, P. E. & Lucangeli, D. (2014). Response to specific training for students with different levels of mathematical difficulties. *Exceptional Children, 80*(3), 337–352. doi:10.1177/0014402914522424

Renkl, A., Atkinson, R. K., Maier, U. H., & Staley, R. (2002). From example study to problem solving: Smooth transitions help learning. *Journal of Experimental Education, 70*(4), 293–315.

Ross, S. M. & Morrison, G. R. (2003). Experimental research methods. *The Handbook of Research for Educational Communications and Technology, 2nd ed.*, 1021-1043.

Schepanski, A., Tubbs, R. M., & Grimlund, R. A. (1992). Issues of concern regarding within- and between-subjects designs in behavioral accounting research. *Journal of Accounting Literature, 121*, 1-20. Retrieved from https://search.proquest.com/docview/216310543?accountid=173015

Secolsky, C., Judd, T. P., Magaram, E., Levy, S. H., Kossar, B. & Reese, G. (2016). Using think-aloud protocols to uncover misconceptions and improve developmental math instruction: An exploratory study. *Numeracy, 9*(1). doi:10.5038/1936-4660.9.1.6

Sekaran, U. (2003). *Research methods for business: A skill-building approach Fourth Edition.* New York, USA: John Wiley & Sons, Inc.

Shrout, P. E. & Fleiss, J. L. (1979). Intraclass correlations : Uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420-428.

Singh, Y. K. (2006). *Fundamentals of research methodology and statistics.* New Delhi, India: New Age International (P) Ltd., Publishers.

Stock, P., Desoete, A., & Roeyers, H. (2010). Detecting children with arithmetic disabilities from kindergarten: Evidence from a 3-year longitudinal study on the role of preparatory arithmetic abilities. *Journal of Learning Disabilities, 43*(3), 250-268. doi:10.1177/0022219409345011.

Suan, J. S. (2014). Factors affecting underachievement in mathematics. *Global Summit on Education GSE 2014* (pp. 13-20). Kuala Lumpur, Malaysia: Worlconferences.Net.

Swanson, H. L. & Sachse-Lee, C. (2001). Mathematical problem solving and working memory in children with learning disabilities: Both executive and phonological processes are important. *Journal of Experimental Child Psychology, 79*, 294-321.

Swanson, H. L. (2016). Word problem solving, working memory and serious math difficulties: Do cognitive strategies really make a difference? *Journal of Applied Research in Memory and Cognition*, 1-16. Retrieved from http://dx.doi.org/10.1016/j.jarmac.2016.04.012.

Swanson, H. L., Jerman, O. & Zheng, X. (2008). Growth in working memory and mathematical problem solving in children at risk and not at risk for serious math difficulties. *Journal of Educational Psychology, 100*(2), 343-379. Retrieved from http://dx.doi.org/10.1037/0022-0663.100.2.343

Sweller, J. (1994). Cognitive load theory, learning difficulty and instructional design. *Learning and Instruction, 4*, 295-312.

van Merrienboer, J. J. G. & Sweller, J. (2005). Cognitive Load Theory and Complex Learning: Recent Developments and Future Directions. *Educational Psychology Review, 17*(2), 147-178. doi:10.1007/s10648-005-3951-0

van Merriënboer, J. J. G., & Ayres, P. (2005). Research on cognitive load theory and its design implications for e-learning. *Educational Technology, Research and Development, 53* (3), 5-13.

Zhu, N. (2015). Cognitive strategy instruction for mathematical word problem-solving of students with mathematics disabilities in China. *International Journal of Disability, Development and Education, 62*(6), 608-627. doi:10.1080/1034912X.2015.1077935

APPENDICES

Appendix A

Informed Consent Document
(Parental Consent)

Reducing Cognitive Load of Mathematics Test Items: Effects on the Performance of High And Low Ability Students

Your daughter/son is invited to be part of a research study on the effect of reduced cognitive load mathematics test-items on the performance of high school students. She/he is selected as a possible participant because she/he is enrolled as Grade 9 student. I ask that you read this form and ask any questions you may have before deciding to let your child participate in the study.

This study is being conducted by: **Steve B. Anapi**, Teacher I (GEANHS)

Background:

The purpose of this study is to determine the effect of reducing the cognitive load of test items on the Mathematics performance of high school students. Particularly, it will seek to investigate how varying the item formats (i.e. traditional, modified item *stem* with essential information, and modified alternatives/options such that options are solution strategies) in a multiple-choice type test affects the performance of high school students.

The main goal of my study is to test this hypothesis. If students were given a reduced cognitive load form of test will there any improvement on their achievement and if so, how will it may help Mathematics teachers improve their writing of test instruments and as well how it will help in teaching Mathematics.

Procedures:

Agreeing to participate does not require your child to complete any work beyond normal class requirements. Participation in this study means that you provide permission to use the data we collect from your child in the assessments that we will provide. Responses of your child will be combined with those of other participants and reported in aggregate form. **Information about individual students will not be used in any published reports.**

Risks and Benefits of being in the Study:

There is no risk in participating in this project. Although your child may not receive direct benefit from her/his participation, others may ultimately benefit from the knowledge obtained in this study.

Compensation:

Your child will receive compensation for the work she/he completes, in the form of class participation points. Even if they choose not to participate in this study – meaning that they do not want their data to be used in this research project – they will receive the same compensation. Refusal to participate in this project will not affect their grades.

Confidentiality:

The records of this study will be kept confidential to the extent provided by national and local law. In any reports on this study, **I will not include any information that will make it possible to identify an individual student.**

Voluntary Nature of the Study:

Your child's participation in this project is voluntary. Even if you sign the informed consent document, your child may decide to leave the study at any time without penalty or loss of benefits to which they may otherwise be entitled.

Contacts and Questions:

The researcher conducting this study is Steve B. Anapi. If you have questions about this study, you may contact him at GEANHS, Mathematics Department, +639167992045 or at this email address: steve13anapi@gmail.com. A copy of this document will be kept together with the research records of this study. **The information contained in this document will be kept confidential.**

Statement of Consent (Please check the appropriate boxes here):

Consent: I have read and understood the above information.

[ ] I agree to the terms and conditions written above and consenting my child, _____ to participate in the study.
Name of the Student

[ ] I do **NOT** want my child to participate in the study.

_____ _____
Printed Name Signature

Appendix B

REQUEST LETTER TO THE SCHOOLS DIVISION SUPERINTENDENT

January 10, 2018

DR. EDITHA M. ATENDIDO
Assistant Schools Division Superintendent
OIC-Schools Division Superintendent

Madam:

Greetings!

I am currently enrolled in thesis writing as a final requirement for the degree Master of Arts in Education with Specialization in Mathematics Education at Philippine Normal University, Manila. My research titled "Reducing Cognitive Load of Mathematics Test Items: Effects on the Performance of High And Low Ability Students" aims to explore the effect of reducing the cognitive load of test items on the Mathematics performance of high school students. Particularly, it will seek to investigate how varying the item formats (i.e. traditional, modified item *stem* with essential information, and item stem with solution strategies as *alternatives/options*) in a multiple-choice type test affects the performance of low-achieving and high-achieving high school students.

In line with the objective of the study, I would like to seek your permission to allow me to conduct my study at General Emilio Aguinaldo National High School where I am currently teaching. Data gathering will take place somewhere between last week of January 2018 and first week of February 2018.

Attached herewith is the Clearance to Proceed signed by the Research Ethics Committee of my university should any query on ethical concerns as to the conduct of my study.

Your generous support is of great importance for the success of this humble undertaking.

Thank you very much.

Respectfully yours,

(SGD) STEVE B. ANAPI
Candidate, MAEd- Mathematics

Approved:

(SGD) EDITHA M. ATENDIDO
Assistant Schools Division Superintendent
OIC-Schools Division Superintendent

Appendix C

INDORSEMENT LETTER FROM THE SCHOOLS DIVISION

Republic of the Philippines
Department of Education
Region IV-A CALABARZON
Toclong I-C, City of Imus
SCHOOLS DIVISION OF IMUS CITY

1st Endorsement
January 12, 2018

Respectfully returned to **Steve B. Anapi, MAEd-Mathematics, Philippine Normal University, Manila,** with the information that this office interposes no objection to the attached request to distribute questionnaires to the teachers of **Gen. Emilio Aguinaldo National High School** in this Division relevant to the research study entitled "**Reducing Cognitive Load of Mathematics Test Items: Effects on the Performance of High And Low Ability Students**" provided the following conditions will be taken into consideration.

1. There should be **proper coordination with the School Head of Gen. Emilio Aguinaldo National High School, this Division;**

2. The activity **should be scheduled as approved by the School Head of Gen. Emilio Aguinaldo National High School; and**

3. The activity **should not disrupt classes nor interfere with the regular activities of the school.**

(SGD) EDITHA M. ATENDIDO
Assistant Schools Division Superintendent /
OIC/ Schools Division Superintendent

osds/lha2018

Appendix D

REQUEST LETTER TO THE SCHOOL HEAD

January 17, 2018

LERMA V. PEÑA
Principal IV
Gen. Emilio Aguinaldo National High School

Madam:

Greetings!

I am currently enrolled in thesis writing as a final requirement for the degree Master of Arts in Education with Specialization in Mathematics Education at Philippine Normal University, Manila. My research titled "Reducing Cognitive Load of Mathematics Test Items: Effects on the Performance of High And Low Ability Students" aims to explore the effect of reducing the cognitive load of test items on the Mathematics performance of high school students. Particularly, it will seek to investigate how varying the item formats (i.e. traditional, modified item *stem* with essential information, and item stem with solution strategies as *alternatives/options*) in a multiple-choice type test affects the performance of low-achieving and high-achieving high school students.

In line with the objective of the study, I would like to seek your permission to allow me to conduct my study at General Emilio Aguinaldo National High School where I am currently teaching. Participants of my study are Grade Nine high school students. Data gathering will take place somewhere between last week of January to February 2018.

Attached herewith are the Indorsement Letter from the Schools Division of Imus City signed by Dr. Editha M. Atendido and the Clearance to Proceed signed by the Research Ethics Committee of my university should any query on ethical concerns as to the conduct of my study.

Your generous support is of great importance for the success of this humble undertaking.

Thank you very much.

Respectfully yours,

(SGD) STEVE B. ANAPI
Candidate, MAEd- Mathematics

Approved:

(SGD) LERMA V. PEÑA
Principal IV

Appendix E

REQUEST LETTER TO THE HEAD TEACHER

February 12, 2018

ELIZABETH Z. VILLANUEVA
Head Teacher IV- Math Dept.

Madam:

Greetings!

I am currently enrolled in thesis writing as a final requirement for the degree Master of Arts in Education with Specialization in Mathematics Education at Philippine Normal University, Manila. My research titled "Reducing Cognitive Load of Mathematics Test Items: Effects on the Performance of High And Low Ability Students" aims to explore the effect of reducing the cognitive load of test items on the Mathematics performance of high school students. Particularly, it will seek to investigate how varying the item formats (i.e. traditional, modified item *stem* with essential information, and item stem with solution strategies as *alternatives/options*) in a multiple-choice type test affects the performance of low-achieving and high-achieving high school students.

In line with the objective of the study, I would like to seek your permission to allow me to conduct my study at General Emilio Aguinaldo National High School where I am currently teaching. Participants of my study are Grade Nine high school students. Data gathering will take place somewhere between last week of January to February 2018.

Attached herewith are the Indorsement Letter from the Schools Division of Imus City signed by Dr. Editha M. Atendido, the approved letter signed by our principal, Mrs. Lerma V. Peña and the Clearance to Proceed signed by the Research Ethics Committee of my university should any query on ethical concerns as to the conduct of my study.

Your generous support is of great importance for the success of this humble undertaking.

Thank you very much.

Respectfully yours,

(SGD) STEVE B. ANAPI
Candidate, MAEd- Mathematics

Approved:

(SGD) ELIZABETH Z. VILLANUEVA
HT IV- Mathematics Department

Appendix F

LETTER OF INVITATION TO EVALUATORS

January 15, 2018

LOVELY AIREIN M. DE GUZMAN
Faculty, CFLEX
Philippine Normal University- Manila

Madam:

Greetings!

I am currently enrolled in thesis writing as a final requirement for the degree Master of Arts in Education with Specialization in Mathematics Education at Philippine Normal University, Manila. My research titled "Reducing Cognitive Load of Mathematics Test Items: Effects on the Performance of High And Low Ability Students" aims to explore the effect of reducing the cognitive load of test items on the Mathematics performance of high school students.

In line with the objective of the study, it is with great pride and honor to invite you as one of my evaluators. Your expertise is of much help to my present study. I am humbly requesting you to evaluate the test instruments I am about to use in my research.

Attached herewith are the Indorsement Letter from the Schools Division of Imus City signed by Dr. Editha M. Atendido and the Clearance to Proceed signed by the Research Ethics Committee of my university should any query on ethical concerns as to the conduct of my study.

Your generous support is of great importance for the success of this humble undertaking.

Thank you very much.

Respectfully yours,

(SGD) STEVE B. ANAPI
Candidate, MAEd- Mathematics

Noted:

(SGD) WILMER G. MARQUEZ
Associate Professor
Faculty of Science, Technology & Mathematics
Philippine Normal University- Manila

January 15, 2018

CELINA P. SARMIENTO, Ph.D.
Faculty, CFLEX
Philippine Normal University- Manila

Madam:

Greetings!

I am currently enrolled in thesis writing as a final requirement for the degree Master of Arts in Education with Specialization in Mathematics Education at Philippine Normal University, Manila. My research titled “Reducing Cognitive Load of Mathematics Test Items: Effects on the Performance of High And Low Ability Students” aims to explore the effect of reducing the cognitive load of test items on the Mathematics performance of high school students.

In line with the objective of the study, it is with great pride and honor to invite you as one of my evaluators. Your expertise is of much help to my present study. I am humbly requesting you to evaluate the test instruments I am about to use in my research.

Attached herewith are the Indorsement Letter from the Schools Division of Imus City signed by Dr. Editha M. Atendido and the Clearance to Proceed signed by the Research Ethics Committee of my university should any query on ethical concerns as to the conduct of my study.

Your generous support is of great importance for the success of this humble undertaking.

Thank you very much.

Respectfully yours,

(SGD) STEVE B. ANAPI
Candidate, MAEd- Mathematics

Noted:

(SGD) WILMER G. MARQUEZ
Associate Professor
Faculty of Science, Technology & Mathematics
Philippine Normal University- Manila

January 17, 2018

MARIA ANGELA L. DIOPOL
Program Coordinator
Department of Social Sciences and Humanities
Cavite State University- Imus Campus

Madam:

Greetings!

I am currently enrolled in thesis writing as a final requirement for the degree Master of Arts in Education with Specialization in Mathematics Education at Philippine Normal University, Manila. My research titled “Reducing Cognitive Load of Mathematics Test Items: Effects on the Performance of High And Low Ability Students” aims to explore the effect of reducing the cognitive load of test items on the Mathematics performance of high school students.

In line with the objective of the study, it is with great pride and honor to invite you as one of my evaluators. Your expertise is of much help to my present study. I am humbly requesting you to evaluate the test instruments I am about to use in my research.

Attached herewith are the Indorsement Letter from the Schools Division of Imus City signed by Dr. Editha M. Atendido and the Clearance to Proceed signed by the Research Ethics Committee of my university should any query on ethical concerns as to the conduct of my study.

Your generous support is of great importance for the success of this humble undertaking.

Thank you very much.

Respectfully yours,

(SGD) STEVE B. ANAPI
Candidate, MAEd- Mathematics

Noted:

(SGD) WILMER G. MARQUEZ
Associate Professor
Faculty of Science, Technology & Mathematics
Philippine Normal University- Manila

January 23, 2018

TERESITA T. RUNGDUIN, Ph.D.
Head, REC Secretariat
Faculty
Philippine Normal University-Manila

Madam:

Greetings!

I am currently enrolled in thesis writing as a final requirement for the degree Master of Arts in Education with Specialization in Mathematics Education at Philippine Normal University, Manila. My research titled “Reducing Cognitive Load of Mathematics Test Items: Effects on the Performance of High And Low Ability Students” aims to explore the effect of reducing the cognitive load of test items on the Mathematics performance of high school students.

In line with the objective of the study, it is with great pride and honor to invite you as one of my evaluators. Your expertise is of much help to my present study. I am humbly requesting you to evaluate the test instruments I am about to use in my research.

Attached herewith are the Indorsement Letter from the Schools Division of Imus City signed by Dr. Editha M. Atendido and the Clearance to Proceed signed by the Research Ethics Committee of my university should any query on ethical concerns as to the conduct of my study.

Your generous support is of great importance for the success of this humble undertaking.

Thank you very much.

Respectfully yours,

(SGD) STEVE B. ANAPI
Candidate, MAEd- Mathematics

Noted:

(SGD) WILMER G. MARQUEZ
Associate Professor
Faculty of Science, Technology & Mathematics
Philippine Normal University- Manila

February 2, 2018

LIZA C. COSTA, Ph.D.
Campus Dean
Cavite State University- Imus Campus

Madam:

Greetings!

I am currently enrolled in thesis writing as a final requirement for the degree Master of Arts in Education with Specialization in Mathematics Education at Philippine Normal University, Manila. My research titled “Reducing Cognitive Load of Mathematics Test Items: Effects on the Performance of High And Low Ability Students” aims to explore the effect of reducing the cognitive load of test items on the Mathematics performance of high school students.

In line with the objective of the study, it is with great pride and honor to invite you as one of my evaluators. Your expertise is of much help to my present study. I am humbly requesting you to evaluate the test instruments I am about to use in my research.

Attached herewith are the Indorsement Letter from the Schools Division of Imus City signed by Dr. Editha M. Atendido and the Clearance to Proceed signed by the Research Ethics Committee of my university should any query on ethical concerns as to the conduct of my study.

Your generous support is of great importance for the success of this humble undertaking.

Thank you very much.

Respectfully yours,

(SGD) STEVE B. ANAPI
Candidate, MAEd- Mathematics

Noted:

(SGD) WILMER G. MARQUEZ
Associate Professor
Faculty of Science, Technology & Mathematics
Philippine Normal University- Manila

Appendix G

PROFILE OF EVALUATORS

Area of Specialization: Mathematics

Name	**Celina P. Sarmiento**
Designation	**Faculty- College of Flexible Learning and ePNU**
School/ University Affiliation	**Philippine Normal University- Manila**
Highest Educational Attainment	**Doctorate (Ph.D. in Science Education Major in Mathematics)**
University Graduated From	**De La Salle University- Manila Philippine Normal University- Manila**

Name	**Lovely Airein M. De Guzman**
Designation	**Faculty- College of Flexible Learning and ePNU**
School/ University Affiliation	**Philippine Normal University- Manila**
Highest Educational Attainment	**Masters (Master of Arts in Education with Specialization in Mathematics Education)**
University Graduated From	**Philippine Normal University- Manila**

Area of Specialization: Psychology

Name	**Liza C. Costa**
Designation	**Associate Professor V; Campus Dean**
School/ University Affiliation	**Cavite State University**
Highest Educational Attainment	**Doctorate (Ph.D. in Education Major in Educational Psychology)**
University Graduated From	**University of the Philippines- Diliman**

Name	**Teresita T. Rungduin**
Designation	**Associate Professor IV**
School/ University Affiliation	**Philippine Normal University- Manila**
Highest Educational Attainment	**Doctorate (Ph.D. in Psychology Major in Developmental Psychology and Minor in Clinical Psychology)**
University Graduated From	**University of the Philippines- Diliman**

Name	**Maria Angela L. Diopol**
Designation	**Program Coordinator; Instructor I**
School/ University Affiliation	**Cavite State University- Imus Campus**
Highest Educational Attainment	**Masters Master in Psychology**
University Graduated From	**Polytechnic University of the Philippines- Manila**

Appendix H

EVALUATION FORMS

VALIDATION OF MATHEMATICS ASSESSMENT INSTRUMENT IN OPERATIONS ON INTEGERS AND POLYNOMIALS (TEST A)
(Evaluation Form)

Rater's Name: _____ Position: _____
No. of Years in Service: _____ Subjects Handled: _____
School Affiliation: _____ _____
Highest Educational Attainment: _____

Directions: Please accomplish the questionnaire by indicating your answer with a check mark (/) under the appropriate column. Use the following rating scale with its verbal interpretation.

Numerical Rating	Descriptive Equivalent
4	Strongly Agree (SA)
3	Agree (A)
2	Disagree (D)
1	Strongly Disagree (SD)

Criteria	4	3	2	1
A. Content				
1. Conforms to DepEd K to 12 Mathematics learning competencies for Junior High School.				
2. The items are level- appropriate. Problems are presented within the level of understanding of learners.				
3. Relate assessment to the content under study and to student capacity.				
4. All problems are directly relevant to the objectives. Match assessment to intended learning objectives.				
5. Uses assessment that can truly reveal whether students understand the learning.				
6. Appropriate answers have been included.				
Average				
B. Format and Layout				
1. Questions are progressive in difficulty.				
2. Layout of pages is well organized making the whole test appears interesting and easy to answer.				

3. Appropriate in length for the allotted time.				
4. Followed the traditional way of writing a test.				
Average				
C. Language Suitability				
1. Questions and instructions are clear, easy to comprehend and grammatically acceptable.				
2. Sentences used in the problems are clear and simple.				
Average				

Comments/Suggestions:

_____
_____
_____
_____
_____
_____

_____
Signature over Printed Name

VALIDATION OF MATHEMATICS ASSESSMENT INSTRUMENT IN OPERATIONS ON INTEGERS AND POLYNOMIALS (TEST B)
(Evaluation Form)

Rater's Name: _____ Position: _____
No. of Years in Service: _____ Subjects Handled: _____
School Affiliation: _____ _____
Highest Educational Attainment: _____

Directions: Please accomplish the questionnaire by indicating your answer with a check mark (/) under the appropriate column. Use the following rating scale with its verbal interpretation.

Numerical Rating	Descriptive Equivalent
4	Strongly Agree (SA)
3	Agree (A)
2	Disagree (D)
1	Strongly Disagree (SD)

Criteria	4	3	2	1
A. Content				
1. Conforms to DepEd K to 12 Mathematics learning competencies for Junior High School.				
2. The items are level- appropriate. Problems are presented within the level of understanding of learners.				
3. Relate assessment to the content under study and to student capacity.				
4. All problems are directly relevant to the objectives. Match assessment to intended learning objectives.				
5. Uses assessment that can truly reveal whether students understand the learning.				
6. Appropriate answers have been included.				
Average				
B. Format and Layout				
1. Questions are progressive in difficulty.				
2. Layout of pages is well organized making the whole test appears interesting and easy to answer.				
3. Appropriate in length for the allotted time.				
4. Item stems are stripped of construct irrelevant information.				
5. Simplification of item stems does not affect the construct being tested.				
6. Content of item stems is strip down to essentials. Eliminate extraneous visuals and texts.				
7. Proper sequence is observed that includes asking questions first to give a direction to the item, and then followed by supporting				

information. Proper sequence is also observed in ordering answer options logically.				
8. Item stem instructions are precise, direct, clear and simple.				
9. Text economy is observed by reducing word count and simplifying language on item stems.				
10. The item format and layout are logically and aesthetically presented.				
Average				
C. Language Suitability				
1. Questions and instructions are clear, easy to comprehend and grammatically acceptable.				
2. Sentences used in the problems are clear and simple.				
Average				

Comments/Suggestions:

_____
_____
_____
_____
_____
_____
_____

_____
Signature over Printed Name

VALIDATION OF MATHEMATICS ASSESSMENT INSTRUMENT IN OPERATIONS ON INTEGERS AND POLYNOMIALS (TEST C)
(Evaluation Form)

Rater's Name: _____ Position: _____
No. of Years in Service: _____ Subjects Handled: _____
School Affiliation: _____ _____
Highest Educational Attainment: _____

Directions: Please accomplish the questionnaire by indicating your answer with a check mark (/) under the appropriate column. Use the following rating scale with its verbal interpretation.

Numerical Rating	Descriptive Equivalent
4	Strongly Agree (SA)
3	Agree (A)
2	Disagree (D)
1	Strongly Disagree (SD)

Criteria	4	3	2	1
A. Content				
1. Conforms to DepEd K to 12 Mathematics learning competencies for Junior High School.				
2. The items are level- appropriate. Problems are presented within the level of understanding of learners.				
3. Relate assessment to the content under study and to student capacity.				
4. All problems are directly relevant to the objectives. Match assessment to intended learning objectives.				
5. Uses assessment that can truly reveal whether students understand the learning.				
6. Appropriate answers have been included.				
Average				
B. Format and Layout				
1. Questions are progressive in difficulty.				
2. Layout of pages is well organized making the whole test appears interesting and easy to answer.				
3. Appropriate in length for the allotted time.				
4. Proper sequence is observed that includes asking questions first to give a direction to the item, and then followed by supporting information. Proper sequence is also observed in ordering answer options logically.				
5. The item format and layout are logically and aesthetically presented.				
6. Alternatives/options cue learners to find the solution to the				

	problem.				
7.	Alternatives/options guide learners on the series of steps necessary to solve the problem.				
8.	Alternatives/options activate previously stored knowledge.				
9.	Alternatives/options motivate learners to find the answer to the problem.				
10.	Solution strategies offered as alternatives/options invite learners to find correct solution to the problem.				
	Average				
C.	**Language Suitability**				
1.	Questions and instructions are clear, easy to comprehend and grammatically acceptable.				
2.	Sentences used in the problems are clear and simple.				
3.	Language used in alternatives/options is level appropriate for learners.				
	Average				

Comments/Suggestions:

_____
_____
_____
_____
_____
_____
_____
_____
_____
_____

_____
Signature over Printed Name

Appendix I

Mathematics Assessment Instrument in Operations on Integers and Polynomials

Table of Specifications

Learning Competencies	No. of days	Percentage	No. of Items	Cognitive Process Dimensions						Item Placement
				60%			30%		10%	
				Remem-bering	Understand-ing	Apply-ing	Analyz-ing	Evaluat-ing	Creat-ing	
Perform operations on the set of integers **(M7NS-Id-2)** (a) Addition of integers	2	8	3		3					1, 2, 6
(b) Subtraction of integers	2	8	3		3					3, 4, 7
(c)Multiplication of integers	1	4	1		1					5
(d) Division of integers	1	4	1		1					8
Adds and subtracts polynomials **(M7AL-IId-2)** (a) Addition of polynomials	2	8	3		1		1		1	9, 10, 16

(b)Subtraction of polynomials	2	8	3		1	1	1			14, 15, 20
Multiplies and divides polynomials **(M7AL-IId-e-2)** (a)Multiplication of polynomials	3	12	4		1	1	2			11, 12, 17, 26
(b) Division of polynomials	2	8	3		1		1	1		13, 27, 28
Uses models and algebraic methods to find the: **(M7AL-IIe-g-1)** *(a) product of two binomials*	2	8	3			1	1	1		18, 19, 25
(b) product of the sum and difference of two terms	2	8	3		2	1				21, 22, 29
(c) square of a binomial	2	8	3				2		1	24, 30, 31
(d) cube of a binomial	2	8	3			2	1			33, 34, 35
(e) product of a binomial and a trinomial	2	8	2			1			1	23, 32
TOTAL	**25**	**100**	**35**		**14**	**7**	**9**	**2**	**3**	

Appendix J

REQUEST LETTER TO COLLEAGUES

February 13, 2018

FACULTY OF MATHEMATICS
Gen. Emilio Aguinaldo National High School

Colleagues:

Greetings!

I am currently enrolled in thesis writing as a final requirement for the degree Master of Arts in Education with Specialization in Mathematics Education at Philippine Normal University, Manila. My research titled “Reducing Cognitive Load of Mathematics Test Items: Effects on the Performance of High And Low Ability Students” aims to explore the effect of reducing the cognitive load of test items on the Mathematics performance of high school students. Particularly, it will seek to investigate how varying the item formats (i.e. traditional, modified item *stem* with essential information, and item stem with solution strategies as *alternatives/options*) in a multiple-choice type test affects the performance of low-achieving and high-achieving high school students.

In line with the objective of the study, I would like to seek your permission to allow me to conduct my study to the following sections and with the following dates.

Phase	Date	Section	Teacher Concerned
First Tryout	Feb. 14, 2018	9-Amos, 9-Kings, 9-Nahum, 9-Songs of Solomon, 9-Habakuk and 9- Nehemiah	Ms. Katherine Anne Bagtas and Mr. Aaron Igat
Second Tryout	Feb. 21, 2018	9- Ruth, 9- Samuels, 9-Chronicles, 9- Esther, 9-Job, 9- Proverbs, 9-Isaiah, 9- Jeremiah and 9- Ezekiel	Mr. Mario Legaspi, Ms. Jhinnefer Rafael and Mr. Marvan Llanto
Administration of Test	Feb. 28, 2018	9- Genesis, 9- Exodus, 9-Leviticus, 9- Numbers, 9-	Ms. Marina Mandapat and Researcher

(Experimentation)		Deuteronomy, 9- Joshua and 9- Judges	
Interview	March 6 and 7, 2018	Selected Subjects	Ms. Marina Mandapat and Researcher

Attached herewith are the Indorsement Letter from the Schools Division of Imus City signed by Dr. Editha M. Atendido, the approved letter signed by our principal, Mrs. Lerma V. Peña, the approved letter signed by our Head Teacher, Mrs. Elizabeth Z. Villanueva and the Clearance to Proceed signed by the Research Ethics Committee of my university should any query on ethical concerns as to the conduct of my study.

I will at my very own expense explain to you the details on the conduct of the first tryout, second tryout and the administration of the test. Your cooperation and willingness is highly acknowledged.

Your generous support is of great importance for the success of this humble undertaking.

Thank you very much.

Respectfully yours,

(SGD) STEVE B. ANAPI
Candidate, MAEd- Mathematics

Appendix K

GUIDELINES IN TEST ADMINISTRATION

GENERAL INSTRUCTIONS:

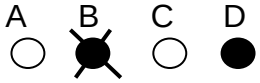
1. The time allotted to complete the test is 50 minutes.

2. Read each item carefully and choose the correct option. Shade the corresponding circle on your Answer Sheet.

3. **For Test A and Test B Instructions:**
   This is a test of your **ability to think out solutions** to the given problems. Choose the correct answer from the given choices and shade the corresponding circle on your **Answer Sheet**.

   For Test C Instructions:
   This is a test of your ability to **identify the correct solution strategy** that best answers the given problem. Choose the correct answer from the given choices and shade the corresponding circle on your **Answer Sheet**.

4. In changing your answer, put a cross mark on your first option and shade your final answer. You can only do this once per item.
   Example:



5. If you have completed the test before the allotted time, raise your hand to signal that you are done. The examiner will get your answer sheet and questionnaire; you may leave the room as soon as the examiner has checked it.

6. Do not write anything on the test questionnaire. Provide your own scratch paper. All scratch papers will be collected.

7. Refrain asking questions to your seatmates. If you want a clarification, raise your hand so the examiner may answer your questions.

Appendix L

QUANTITATIVE ITEM ANALYSIS ON THE FIRST TRY OUT

Item Analysis Based On The First Try Out

TEST A

Item No.	Key	F U 27%	F L 27%	P U 27%	P L 27%	Difficulty Index	Interpretation/ Remark	Discrimination Index	Interpretation/ Remark	Decision
1	**B**	26	6	0.96	0.22	0.59	Average/ Retain	0.74	VG/ Retain	Retain
2	**A**	24	7	0.89	0.26	0.57	Average/ Retain	0.63	VG/Retain	Retain
3	**A**	6	4	0.22	0.15	0.19	Difficult/ Revise	0.07	PI/Reject	Revise
4	**B**	8	3	0.30	0.11	0.20	Difficult/ Revise	0.19	MI/ Revise	Revise
5	**A**	26	18	0.96	0.67	0.81	Easy/ Revise	0.30	RG/Retain	Revise
6	**A**	19	6	0.70	0.22	0.46	Average/ Retain	0.48	VG/Retain	Retain
7	**D**	6	2	0.22	0.07	0.15	Difficult/ Revise	0.15	MI/ Revise	Revise
8	**C**	23	14	0.85	0.52	0.69	Average/ Retain	0.33	G/Retain	Retain
9	**B**	8	7	0.30	0.26	0.28	Average/ Retain	0.04	PI/Reject	Revise
10	**A**	3	4	0.11	0.15	0.13	Difficult/ Revise	-0.04	PI/Reject	Reject
11	**D**	19	3	0.70	0.11	0.41	Average/ Retain	0.59	VG/Retain	Retain
12	**D**	22	4	0.81	0.15	0.48	Average/ Retain	0.67	VG/Retain	Retain
13	**A**	12	7	0.44	0.26	0.35	Average/ Retain	0.19	MI/ Revise	Revise
14	**D**	5	3	0.19	0.11	0.15	Difficult/ Revise	0.07	PI/Reject	Revise
15	**B**	2	8	0.07	0.30	0.19	Difficult/ Revise	-0.22	PI/Reject	Reject
16	**C**	18	7	0.67	0.26	0.46	Average/ Retain	0.41	VG/ Retain	Retain
17	**D**	24	5	0.89	0.19	0.54	Average/ Retain	0.70	VG/ Retain	Retain
18	**C**	15	0	0.56	0.00	0.28	Average/ Retain	0.56	VG/Retain	Retain
19	**C**	16	6	0.59	0.22	0.41	Average/ Retain	0.37	G/ Retain	Retain
20	**B**	12	1	0.44	0.04	0.24	Difficult/ Revise	0.41	VG/Retain	Revise
21	**B**	16	16	0.59	0.59	0.59	Average/ Retain	0.00	PI/Reject	Revise
22	**D**	23	9	0.85	0.33	0.59	Average/ Retain	0.52	VG/Retain	Retain
23	**A**	6	4	0.22	0.15	0.19	Difficult/ Revise	0.07	PI/Reject	Revise

24	**D**	15	1	0.56	0.04	0.30	Average/ Retain	0.52	VG/Retain	Retain
25	**D**	16	3	0.59	0.11	0.35	Average/ Retain	0.48	VG/Retain	Retain
26	**C**	8	11	0.30	0.41	0.35	Average/ Retain	-0.11	PI/Reject	Revise
27	**B**	5	4	0.19	0.15	0.17	Difficult/ Revise	0.04	PI/Reject	Revise
28	**B**	3	5	0.11	0.19	0.15	Difficult/ Revise	-0.07	PI/Reject	Reject
29	**C**	18	8	0.67	0.30	0.48	Average/ Retain	0.37	G/Retain	Retain
30	**D**	15	4	0.56	0.15	0.35	Average/ Retain	0.41	VG/Retain	Retain
31	**C**	21	5	0.78	0.19	0.48	Average/ Retain	0.59	VG/Retain	Retain
32	**A**	6	5	0.22	0.19	0.20	Difficult/ Revise	0.04	PI/Reject	Revise
33	**D**	13	4	0.48	0.15	0.31	Average/ Retain	0.33	G/Retain	Retain
34	**D**	5	2	0.19	0.07	0.13	Difficult/ Revise	0.11	MI/ Revise	Revise
35	**C**	16	3	0.59	0.11	0.35	Average/ Retain	0.48	VG/Retain	Retain

Legend:

VG – Very Good Item
G – Good Item
RG - Reasonably Good Item
MI – Marginal Item
PI – Poor Item
F – Frequency
P - Proportion

Item Analysis Based On The First Try Out

TEST B

Item No.	Key	F U 27%	F L 27%	P U 27%	P L 27%	Difficulty Index	Interpretation/ Remark	Discrimination Index	Interpretation/ Remark	Decision
1	B	21	8	0.78	0.30	0.54	Average/ Retain	0.48	VG/Retain	Retain
2	A	25	10	0.93	0.37	0.65	Average/ Retain	0.56	VG/Retain	Retain
3	A	23	9	0.85	0.33	0.59	Average/ Retain	0.52	VG/Retain	Retain
4	B	22	8	0.81	0.30	0.56	Average/ Retain	0.52	VG/Retain	Retain
5	A	27	25	1.00	0.93	0.96	Very Easy/ Reject	0.07	PI/Reject	Revise
6	A	24	9	0.89	0.33	0.61	Average/ Retain	0.56	VG/Retain	Retain
7	D	8	5	0.30	0.19	0.24	Difficult/ Revise	0.11	MI/ Revise	Revise
8	C	21	11	0.78	0.41	0.59	Average/ Retain	0.37	G/Retain	Retain
9	B	16	6	0.59	0.22	0.41	Average/ Retain	0.37	G/Retain	Retain
10	A	13	9	0.48	0.33	0.41	Average/ Retain	0.15	MI/ Revise	Revise
11	D	19	3	0.70	0.11	0.41	Average/ Retain	0.59	VG/Retain	Retain
12	D	12	2	0.44	0.07	0.26	Difficult/ Revise	0.37	G/Retain	Revise
13	A	21	6	0.78	0.22	0.50	Average/ Retain	0.56	VG/Retain	Retain
14	D	3	2	0.11	0.07	0.09	Very Difficult/ Reject	0.04	PI/Reject	Revise
15	B	8	4	0.30	0.15	0.22	Difficult/ Revise	0.15	MI/ Revise	Revise
16	C	19	9	0.70	0.33	0.52	Average/ Retain	0.37	G/Retain	Retain
17	D	24	5	0.89	0.19	0.54	Average/ Retain	0.70	VG/Retain	Retain
18	C	15	3	0.56	0.11	0.33	Average/ Retain	0.44	VG/Retain	Retain
19	C	17	3	0.63	0.11	0.37	Average/ Retain	0.52	VG/Retain	Retain
20	B	11	9	0.41	0.33	0.37	Average/ Retain	0.07	PI/Reject	Revise
21	B	21	18	0.78	0.67	0.72	Average/ Retain	0.11	MI/ Revise	Revise
22	D	22	5	0.81	0.19	0.50	Average/ Retain	0.63	VG/Retain	Retain
23	A	8	2	0.30	0.07	0.19	Difficult/ Revise	0.22	RG/Retain	Revise
24	D	21	2	0.78	0.07	0.43	Average/ Retain	0.70	VG/Retain	Retain
25	D	12	4	0.44	0.15	0.30	Average/ Retain	0.30	RG/Retain	Retain
26	C	7	0	0.26	0.00	0.13	Difficult/ Revise	0.26	RG/Retain	Revise
27	B	12	5	0.44	0.19	0.31	Average/ Retain	0.26	RG/Retain	Retain
28	B	13	5	0.48	0.19	0.33	Average/ Retain	0.30	RG/Retain	Retain
29	C	18	11	0.67	0.41	0.54	Average/ Retain	0.26	RG/Retain	Retain
30	D	14	1	0.52	0.04	0.28	Average/ Retain	0.48	VG/Retain	Retain

31	C	13	4	0.48	0.15	0.31	Average/ Retain	0.33	G/Retain	Retain
32	A	6	2	0.22	0.07	0.15	Difficult/ Revise	0.15	MI/ Revise	Revise
33	D	9	2	0.33	0.07	0.20	Difficult/ Revise	0.26	RG/Retain	Revise
34	D	5	2	0.19	0.07	0.13	Difficult/ Revise	0.11	MI/ Revise	Revise
35	C	9	4	0.33	0.15	0.24	Difficult/ Revise	0.19	MI/ Revise	Revise

Item Analysis Based On The First Try Out

TEST C

Item No.	Key	F U 27%	F L 27%	P U 27%	P L 27%	Difficulty Index	Interpretation/ Remark	Discrimination Index	Interpretation/ Remark	Decision
1	B	23	13	0.85	0.48	0.67	Average/ Retain	0.37	G/Retain	Retain
2	B	26	8	0.96	0.30	0.63	Average/ Retain	0.67	VG/Retain	Retain
3	D	11	4	0.41	0.15	0.28	Average/ Retain	0.26	RG/Retain	Retain
4	C	10	8	0.37	0.30	0.33	Average/ Retain	0.07	PI/Reject	Revise
5	D	25	15	0.93	0.56	0.74	Average/ Retain	0.37	G/Retain	Retain
6	C	23	7	0.85	0.26	0.56	Average/ Retain	0.59	VG/Accept	Retain
7	A	5	4	0.19	0.15	0.17	Difficult/ Revise	0.04	PI/Reject	Revise
8	D	23	9	0.85	0.33	0.59	Average/ Retain	0.52	VG/Retain	Retain
9	B	15	4	0.56	0.15	0.35	Average/ Retain	0.41	VG/Retain	Retain
10	A	13	3	0.48	0.11	0.30	Average/ Retain	0.37	G/Retain	Retain
11	D	22	3	0.81	0.11	0.46	Average/ Retain	0.70	VG/Retain	Retain
12	D	16	2	0.59	0.07	0.33	Average/ Retain	0.52	VG/Retain	Retain
13	C	12	4	0.44	0.15	0.30	Average/ Retain	0.30	RG/Retain	Retain
14	D	6	8	0.22	0.30	0.26	Difficult/ Revise	-0.07	PI/Reject	Reject
15	A	16	4	0.59	0.15	0.37	Average/ Retain	0.44	VG/Retain	Retain
16	C	11	5	0.41	0.19	0.30	Average/ Retain	0.22	RG/Retain	Retain
17	D	14	5	0.52	0.19	0.35	Average/ Retain	0.33	G/Retain	Retain
18	D	17	1	0.63	0.04	0.33	Average/ Retain	0.59	VG/Retain	Retain
19	A	18	3	0.67	0.11	0.39	Average/ Retain	0.56	VG/Retain	Retain
20	C	18	4	0.67	0.15	0.41	Average/ Retain	0.52	VG/Retain	Retain
21	A	26	9	0.96	0.33	0.65	Average/ Retain	0.63	VG/Retain	Retain
22	A	19	5	0.70	0.19	0.44	Average/ Retain	0.52	VG/Retain	Retain
23	A	8	4	0.30	0.15	0.22	Difficult/ Revise	0.15	MI/ Revise	Revise
24	B	19	13	0.70	0.48	0.59	Average/ Retain	0.22	RG/Retain	Retain
25	C	15	8	0.56	0.30	0.43	Average/ Retain	0.26	RG/Retain	Retain
26	D	1	3	0.04	0.11	0.07	Very Difficult/ Reject	-0.07	PI/Reject	Reject
27	B	9	8	0.33	0.30	0.31	Average/ Retain	0.04	PI/Reject	Revise
28	B	11	6	0.41	0.22	0.31	Average/ Retain	0.19	MI/ Revise	Revise

29	A	15	8	0.56	0.30	0.43	Average/ Retain	0.26	RG/Accept	Retain
30	C	7	6	0.26	0.22	0.24	Difficult/ Revise	0.04	PI/Reject	Revise
31	D	13	2	0.48	0.07	0.28	Average/ Retain	0.41	VG/Retain	Retain
32	A	9	4	0.33	0.15	0.24	Difficult/ Revise	0.19	MI/ Revise	Revise
33	C	10	6	0.37	0.22	0.30	Average/ Retain	0.15	MI/ Revise	Revise
34	D	7	3	0.26	0.11	0.19	Difficult/ Revise	0.15	MI/ Revise	Revise
35	D	11	1	0.41	0.04	0.22	Difficult/ Revise	0.37	G/Retain	Revise

Appendix M

OPTION ANALYSIS ON THE FIRST TRYOUT

Plausibility Index of Options Based On the First Try Out

TEST A

Item No.	Key	A		B		C		D		No Answer		Option Plausibility Index (U - L)							
		U	L	U	L	U	L	U	L	U	L	**A**	Remark	**B**	Remark	**C**	Remark	**D**	Remark
1	**B**	1	17	26	6	0	3	0	1	0	0	-16	Y	20	Y	-3	Y	-1	Y
2	**A**	24	7	0	7	0	2	3	11	0	0	17	Y	-7	Y	-2	Y	-8	Y
3	**A**	6	4	1	9	17	9	3	5	0	0	2	Y	-8	Y	8	N	-2	Y
4	**B**	17	10	8	3	1	8	1	6	0	0	7	N	5	Y	-7	Y	-5	Y
5	**A**	26	18	1	5	0	2	0	2	0	0	8	Y	-4	Y	-2	Y	-2	Y
6	**A**	19	6	5	7	1	5	2	9	0	0	13	Y	-2	Y	-4	Y	-7	Y
7	**D**	12	5	3	8	6	12	6	2	0	0	7	N	-5	Y	-6	Y	4	Y
8	**C**	1	7	2	2	23	14	1	4	0	0	-6	Y	0	N	9	Y	-3	Y
9	**B**	4	3	8	7	11	9	4	8	0	0	1	N	1	N	2	Y	-4	Y
10	**A**	3	4	10	5	2	3	12	15	0	0	-1	Y*	5	N	-1	Y	-3	Y
11	**D**	6	13	1	1	1	10	19	3	0	0	-7	Y	0	N	-9	Y	16	Y
12	**D**	1	9	4	11	0	3	22	4	0	0	-8	Y	-7	Y	-3	Y	18	Y
13	**A**	12	7	9	11	1	3	5	6	0	0	5	Y	-2	Y	-2	Y	-1	Y
14	**D**	1	5	9	7	12	12	5	3	0	0	-4	Y	2	N	0	N	2	Y
15	**B**	8	8	2	8	14	6	3	5	0	0	0	N	-6	Y*	8	N	-2	Y
16	**C**	1	8	4	10	18	7	4	2	0	0	-7	Y	-6	Y	11	Y	2	N
17	**D**	0	7	1	8	2	7	24	5	0	0	-7	Y	-7	Y	-5	Y	19	Y
18	**C**	5	17	1	4	15	0	6	6	0	0	-12	Y	-3	Y	15	Y	0	N
19	**C**	1	13	1	5	16	6	9	3	0	0	-12	Y	-4	Y	10	Y	6	N

20	**B**	2	4	12	1	0	4	13	18	0	0	-2	Y	11	Y	-4	Y	-5	Y
21	**B**	0	4	16	16	3	6	8	1	0	0	-4	Y	0	Y*	-3	Y	7	N
22	**D**	0	5	1	4	3	9	23	9	0	0	-5	Y	-3	Y	-6	Y	14	Y
23	**A**	6	4	6	7	11	13	4	3	0	0	2	Y	-1	Y	-2	Y	1	N
24	**D**	4	12	4	10	4	4	15	1	0	0	-8	Y	-6	Y	0	N	14	Y
25	**D**	1	6	4	12	6	6	16	3	0	0	-5	Y	-8	Y	0	N	13	Y
26	**C**	7	3	6	8	8	11	6	5	0	0	4	N	-2	Y	-3	Y*	1	N
27	**B**	0	5	5	4	2	5	20	13	0	0	-5	Y	1	Y	-3	Y	7	N
28	**B**	0	3	3	5	4	1	20	18	0	0	-3	Y	-2	Y*	3	N	2	N
29	**C**	0	4	1	8	18	8	8	7	0	0	-4	Y	-7	Y	10	Y	1	N
30	**D**	6	9	5	9	1	5	15	4	0	0	-3	Y	-4	Y	-4	Y	11	Y
31	**C**	0	9	2	12	21	5	4	1	0	0	-9	Y	-10	Y	16	Y	3	N
32	**A**	6	5	2	4	6	15	13	3	0	0	1	Y	-2	Y	-9	Y	10	N
33	**D**	3	8	2	9	9	6	13	4	0	0	-5	Y	-7	Y	3	N	9	Y
34	**D**	1	11	12	10	9	4	5	2	0	0	-10	Y	2	N	5	N	3	Y
35	**C**	0	11	3	7	16	3	8	6	0	0	-11	Y	-4	Y	13	Y	2	N

Legend:

Y – Good Option

N – Not Good Option

*** -** Correct option but not plausible or not attractive enough to upper group

Plausibility Index of Options Based On the First Try Out

TEST B

Item No.	Key	A		B		C		D		No Answer		Option Plausibility Index (U - L)							
		U	L	U	L	U	L	U	L	U	L	**A**	Remark	**B**	Remark	**C**	Remark	**D**	Remark
1	**B**	4	6	21	8	1	9	1	4	0	0	-2	Y	13	Y	-8	Y	-3	Y
2	**A**	25	10	2	4	0	2	0	11	0	0	15	Y	-2	Y	-2	Y	-11	Y
3	**A**	23	9	2	9	2	5	0	4	0	0	14	Y	-7	Y	-3	Y	-4	Y
4	**B**	5	10	22	8	0	5	0	4	0	0	-5	Y	14	Y	-5	Y	-4	Y
5	**A**	27	25	0	1	0	1	0	0	0	0	2	Y	-1	Y	-1	Y	0	N
6	**A**	24	9	3	13	0	2	0	3	0	0	15	Y	-10	Y	-2	Y	-3	Y
7	**D**	12	9	1	11	6	2	8	5	0	0	3	N	-10	Y	4	N	3	Y
8	**C**	0	1	5	15	21	11	1	0	0	0	-1	Y	-10	Y	10	Y	1	N
9	**B**	2	3	16	6	7	14	2	4	0	0	-1	Y	10	Y	-7	Y	-2	Y
10	**A**	13	9	7	9	2	3	5	6	0	0	4	Y	-2	Y	-1	Y	-1	Y
11	**D**	6	16	1	6	1	2	19	3	0	0	-10	Y	-5	Y	-1	Y	16	Y
12	**D**	4	11	9	11	2	3	12	2	0	0	-7	Y	-2	Y	-1	Y	10	Y
13	**A**	21	6	6	14	0	2	0	5	0	0	15	Y	-8	Y	-2	Y	-5	Y
14	**D**	0	4	19	14	5	7	3	2	0	0	-4	Y	5	N	-2	Y	1	Y
15	**B**	8	9	8	4	9	6	2	8	0	0	-1	Y	4	Y	3	N	-6	Y
16	**C**	3	6	0	8	19	9	5	4	0	0	-3	Y	-8	Y	10	Y	1	N
17	**D**	0	6	0	8	3	8	24	5	0	0	-6	Y	-8	Y	-5	Y	19	Y
18	**C**	9	13	1	7	15	3	2	4	0	0	-4	Y	-6	Y	12	Y	-2	Y
19	**C**	8	10	0	5	17	3	2	9	0	0	-2	Y	-5	Y	14	Y	-7	Y
20	**B**	5	2	11	9	1	7	10	9	0	0	3	N	2	Y	-6	Y	1	N
21	**B**	1	3	21	18	3	1	2	5	0	0	-2	Y	3	Y	2	N	-3	Y
22	**D**	0	10	3	6	2	6	22	5	0	0	-10	Y	-3	Y	-4	Y	17	Y
23	**A**	8	2	3	11	10	12	6	2	0	0	6	Y	-8	Y	-2	Y	4	N
24	**D**	0	6	4	11	2	8	21	2	0	0	-6	Y	-7	Y	-6	Y	19	Y
25	**D**	3	4	8	13	4	6	12	4	0	0	-1	Y	-5	Y	-2	Y	8	Y

26	**C**	6	6	7	9	7	0	7	12	0	0	0	N	-2	Y	7	Y	-5	Y
27	**B**	0	3	12	5	6	9	9	10	0	0	-3	Y	7	Y	-3	Y	-1	Y
28	**B**	3	7	13	5	3	6	8	9	0	0	-4	Y	8	Y	-3	Y	-1	Y
29	**C**	1	5	3	8	18	11	5	3	0	0	-4	Y	-5	Y	7	Y	2	N
30	**D**	2	11	9	8	2	7	14	1	0	0	-9	Y	1	N	-5	Y	13	Y
31	**C**	5	9	8	12	13	4	1	2	0	0	-4	Y	-4	Y	9	Y	-1	Y
32	**A**	6	2	3	8	11	14	7	3	0	0	4	Y	-5	Y	-3	Y	4	N
33	**D**	7	8	5	9	6	8	9	2	0	0	-1	Y	-4	Y	-2	Y	7	Y
34	**D**	6	11	7	4	9	10	5	2	0	0	-5	Y	3	N	-1	Y	3	Y
35	**C**	1	8	9	14	9	4	8	1	0	0	-7	Y	-5	Y	5	Y	7	N

Plausibility Index of Options Based On the First Try Out

TEST C

Item No.	Key	A		B		C		D		No Answer		Option Plausibility Index (U - L)							
		U	L	U	L	U	L	U	L	U	L	**A**	Remark	**B**	Remark	**C**	Remark	**D**	Remark
1	**B**	2	5	23	13	1	3	1	6	0	0	-3	Y	10	Y	-2	Y	-5	Y
2	**B**	0	4	26	8	0	1	1	14	0	0	-4	Y	18	Y	-1	Y	-13	Y
3	**D**	6	10	5	5	5	8	11	4	0	0	-4	Y	0	N	-3	Y	7	Y
4	**C**	1	10	2	6	10	8	14	3	0	0	-9	Y	-4	Y	2	Y	11	N
5	**D**	0	3	1	5	1	4	25	15	0	0	-3	Y	-4	Y	-3	Y	10	Y
6	**C**	3	5	0	12	23	7	1	3	0	0	-2	Y	-12	Y	16	Y	-2	Y
7	**A**	5	4	4	12	2	5	16	6	0	0	1	Y	-8	Y	-3	Y	10	N
8	**D**	1	4	0	2	3	12	23	9	0	0	-3	Y	-2	Y	-9	Y	14	Y
9	**B**	4	5	15	4	6	13	2	5	0	0	-1	Y	11	Y	-7	Y	-3	Y
10	**A**	13	3	10	11	2	6	2	7	0	0	10	Y	-1	Y	-4	Y	-5	Y
11	**D**	5	12	0	2	0	10	22	3	0	0	-7	Y	-2	Y	-10	Y	19	Y
12	**D**	2	9	3	7	6	9	16	2	0	0	-7	Y	-4	Y	-3	Y	14	Y
13	**C**	2	11	6	9	12	4	7	3	0	0	-9	Y	-3	Y	8	Y	4	N
14	**D**	10	7	3	7	8	5	6	8	0	0	3	N	-4	Y	3	N	-2	Y*
15	**A**	16	4	4	9	3	12	4	2	0	0	12	Y	-5	Y	-9	Y	2	N
16	**C**	9	9	3	7	11	5	4	6	0	0	0	N	-4	Y	6	Y	-2	Y
17	**D**	3	7	3	8	7	7	14	5	0	0	-4	Y	-5	Y	0	N	9	Y
18	**D**	1	3	5	14	4	9	17	1	0	0	-2	Y	-9	Y	-5	Y	16	Y
19	**A**	18	3	1	6	3	11	5	7	0	0	15	Y	-5	Y	-8	Y	-2	Y
20	**C**	1	2	3	7	18	4	5	14	0	0	-1	Y	-4	Y	14	Y	-9	Y
21	**A**	26	9	1	9	0	7	0	2	0	0	17	Y	-8	Y	-7	Y	-2	Y
22	**A**	19	5	5	8	2	8	1	6	0	0	14	Y	-3	Y	-6	Y	-5	Y
23	**A**	8	4	12	12	0	7	7	4	0	0	4	Y	0	N	-7	Y	3	N
24	**B**	2	5	19	13	3	5	3	4	0	0	-3	Y	6	Y	-2	Y	-1	Y
25	**C**	5	6	5	8	15	8	2	5	0	0	-1	Y	-3	Y	7	Y	-3	Y
26	**D**	8	9	7	6	11	9	1	3	0	0	-1	Y	1	N	2	N	-2	Y*
27	**B**	7	4	9	8	11	11	0	4	0	0	3	N	1	Y	0	N	-4	Y
28	**B**	1	11	11	6	3	2	12	8	0	0	-10	Y	5	Y	1	N	4	N

29	**A**	15	8	8	8	3	8	1	3	0	0	7	Y	0	N	-5	Y	-2	Y
30	**C**	4	9	12	11	7	6	4	1	0	0	-5	Y	1	N	1	Y	3	N
31	**D**	2	5	3	9	9	11	13	2	0	0	-3	Y	-6	Y	-2	Y	11	Y
32	**A**	9	4	10	12	6	8	2	3	0	0	5	Y	-2	Y	-2	Y	-1	Y
33	**C**	5	8	5	7	10	6	7	6	0	0	-3	Y	-2	Y	4	Y	1	N
34	**D**	8	8	6	12	6	4	7	3	0	0	0	N	-6	Y	2	N	4	Y
35	**D**	7	8	5	10	4	8	11	1	0	0	-1	Y	-5	Y	-4	Y	10	Y

Appendix N

QUANTITATIVE ITEM ANALYSIS ON THE SECOND TRYOUT

Item Analysis Based On The Second Try Out

TEST A

Item No.	Key	F U 27%	F L 27%	P U 27%	P L 27%	Difficulty Index	Interpretation/ Remark	Discrimination Index	Interpretation/ Remark	Decision
1	**B**	34	20	0.83	0.49	0.66	Average/ Retain	0.34	G/Retain	Retain
2	**A**	33	13	0.80	0.32	0.56	Average/ Retain	0.49	VG/Retain	Retain
3	**A**	19	3	0.46	0.07	0.27	Average/ Retain	0.39	G/Retain	Retain
4	**B**	17	6	0.41	0.15	0.28	Average/ Retain	0.27	RG/Retain	Retain
5	**A**	38	31	0.93	0.76	0.84	Easy/ Revise	0.17	MI/ Revise	Revise
6	**A**	38	22	0.93	0.54	0.73	Average/ Retain	0.39	G/Retain	Retain
7	**D**	21	5	0.51	0.12	0.32	Average/ Retain	0.39	G/Retain	Retain
8	**C**	32	21	0.78	0.51	0.65	Average/ Retain	0.27	RG/Retain	Retain
9	**B**	24	17	0.59	0.41	0.50	Average/ Retain	0.17	MI/ Revise	Revise
10	**A**	15	8	0.37	0.20	0.28	Average/ Retain	0.17	MI/ Revise	Revise
11	**D**	32	8	0.78	0.20	0.49	Average/ Retain	0.59	VG/Retain	Retain
12	**D**	24	4	0.59	0.10	0.34	Average/ Retain	0.49	VG/Retain	Retain
13	**A**	23	6	0.56	0.15	0.35	Average/ Retain	0.41	VG/Retain	Retain
14	**D**	21	6	0.51	0.15	0.33	Average/ Retain	0.37	G/Retain	Retain
15	**B**	25	5	0.61	0.12	0.37	Average/ Retain	0.49	VG/Retain	Retain
16	**C**	30	21	0.73	0.51	0.62	Average/ Retain	0.22	RG/Retain	Retain
17	**D**	40	15	0.98	0.37	0.67	Average/ Retain	0.61	VG/Retain	Retain
18	**C**	38	5	0.93	0.12	0.52	Average/ Retain	0.80	VG/Retain	Retain
19	**C**	36	9	0.88	0.22	0.55	Average/ Retain	0.66	VG/Retain	Retain
20	**A**	20	11	0.49	0.27	0.38	Average/ Retain	0.22	RG/Retain	Retain
21	**B**	35	22	0.85	0.54	0.70	Average/ Retain	0.32	G/Retain	Retain
22	**D**	39	16	0.95	0.39	0.67	Average/ Retain	0.56	VG/Retain	Retain
23	**A**	18	2	0.44	0.05	0.24	Difficult/ Revise	0.39	G/Retain	Retain

24	**D**	35	4	0.85	0.10	0.48	Average/ Retain	0.76	VG/Retain	Retain
25	**D**	36	11	0.88	0.27	0.57	Average/ Retain	0.61	VG/Retain	Retain
26	**C**	25	8	0.61	0.20	0.40	Average/ Retain	0.41	VG/Retain	Retain
27	**B**	26	2	0.63	0.05	0.34	Average/ Retain	0.59	VG/Retain	Retain
28	**B**	18	5	0.44	0.12	0.28	Average/ Retain	0.32	G/Retain	Retain
29	**C**	38	20	0.93	0.49	0.71	Average/ Retain	0.44	VG/Retain	Retain
30	**D**	27	4	0.66	0.10	0.38	Average/ Retain	0.56	VG/Retain	Retain
31	**C**	31	8	0.76	0.20	0.48	Average/ Retain	0.56	VG/Retain	Retain
32	**A**	23	5	0.56	0.12	0.34	Average/ Retain	0.44	VG/Retain	Retain
33	**D**	25	2	0.61	0.05	0.33	Average/ Retain	0.56	VG/Retain	Retain
34	**D**	25	3	0.61	0.07	0.34	Average/ Retain	0.54	VG/Retain	Retain
35	**C**	29	9	0.71	0.22	0.46	Average/ Retain	0.49	VG/Retain	Retain

Legend:

VG – Very Good Item
G – Good Item
RG - Reasonably Good Item
MI – Marginal Item
PI – Poor Item
F – Frequency
P - Proportion

Item Analysis Based On The Second Try Out

TEST B

Item No.	Key	F U 27%	F L 27%	P U 27%	P L 27%	Difficulty Index	Interpretation/ Remark	Discrimination Index	Interpretation/ Remark	Decision
1	**B**	37	18	0.90	0.44	0.67	Average/ Retain	0.46	VG/Retain	Retain
2	**A**	30	18	0.73	0.44	0.59	Average/ Retain	0.29	RG/Retain	Retain
3	**A**	27	3	0.66	0.07	0.37	Average/ Retain	0.59	VG/Retain	Retain
4	**B**	26	9	0.63	0.22	0.43	Average/ Retain	0.41	VG/Retain	Retain
5	**A**	40	20	0.98	0.49	0.73	Average/ Retain	0.49	VG/Retain	Retain
6	**A**	35	8	0.85	0.20	0.52	Average/ Retain	0.66	VG/Retain	Retain
7	**D**	21	4	0.51	0.10	0.30	Average/ Retain	0.41	VG/Retain	Retain
8	**C**	35	9	0.85	0.22	0.54	Average/ Retain	0.63	VG/Retain	Retain
9	**B**	18	2	0.44	0.05	0.24	Difficult/ Revise	0.39	G/Retain	Revise
10	**A**	24	5	0.59	0.12	0.35	Average/ Retain	0.46	VG/Retain	Retain
11	**D**	33	7	0.80	0.17	0.49	Average/ Retain	0.63	VG/Retain	Retain
12	**D**	25	6	0.61	0.15	0.38	Average/ Retain	0.46	VG/Retain	Retain
13	**A**	34	6	0.83	0.15	0.49	Average/ Retain	0.68	VG/Retain	Retain
14	**D**	29	11	0.71	0.27	0.49	Average/ Retain	0.44	VG/Retain	Retain
15	**B**	26	14	0.63	0.34	0.49	Average/ Retain	0.29	RG/Retain	Retain
16	**C**	32	18	0.78	0.44	0.61	Average/ Retain	0.34	G/Retain	Retain
17	**D**	38	22	0.93	0.54	0.73	Average/ Retain	0.39	G/Retain	Retain
18	**C**	35	16	0.85	0.39	0.62	Average/ Retain	0.46	VG/Retain	Retain
19	**C**	36	19	0.88	0.46	0.67	Average/ Retain	0.41	VG/Retain	Retain
20	**A**	23	11	0.56	0.27	0.41	Average/ Retain	0.29	RG/Retain	Retain
21	**B**	40	30	0.98	0.73	0.85	Easy/ Revise	0.24	RG/Retain	Retain
22	**D**	38	12	0.93	0.29	0.61	Average/ Retain	0.63	VG/Retain	Retain
23	**A**	25	2	0.61	0.05	0.33	Average/ Retain	0.56	VG/Retain	Retain
24	**D**	37	10	0.90	0.24	0.57	Average/ Retain	0.66	VG/Retain	Retain
25	**D**	36	13	0.88	0.32	0.60	Average/ Retain	0.56	VG/Retain	Retain
26	**C**	30	9	0.73	0.22	0.48	Average/ Retain	0.51	VG/Retain	Retain
27	**B**	36	11	0.88	0.27	0.57	Average/ Retain	0.61	VG/Retain	Retain
28	**B**	18	3	0.44	0.07	0.26	Difficult/ Revise	0.37	G/Retain	Revise

29	**C**	38	22	0.93	0.54	0.73	Average/ Retain	0.39	G/Retain	Retain
30	**D**	38	7	0.93	0.17	0.55	Average/ Retain	0.76	VG/Retain	Retain
31	**C**	27	11	0.66	0.27	0.46	Average/ Retain	0.39	G/Retain	Retain
32	**A**	22	6	0.54	0.15	0.34	Average/ Retain	0.39	G/Retain	Retain
33	**D**	25	2	0.61	0.05	0.33	Average/ Retain	0.56	VG/Retain	Retain
34	**D**	36	4	0.88	0.10	0.49	Average/ Retain	0.78	VG/Retain	Retain
35	**C**	24	5	0.59	0.12	0.35	Average/ Retain	0.46	VG/Retain	Retain

Item Analysis Based On The Second Try Out

TEST C

Item No.	Key	F U27%	F L 27%	P U27%	P L 27%	Difficulty Index	Interpretation/ Remark	Discrimination Index	Interpretation/ Remark	Decision
1	**B**	36	26	0.88	0.63	0.76	Average/ Retain	0.24	RG/Retain	Retain
2	**B**	32	15	0.78	0.37	0.57	Average/ Retain	0.41	VG/Retain	Retain
3	**D**	21	4	0.51	0.10	0.30	Average/ Retain	0.41	VG/Retain	Retain
4	**C**	21	7	0.51	0.17	0.34	Average/ Retain	0.34	G/Retain	Retain
5	**D**	38	29	0.93	0.71	0.82	Easy/ Revise	0.22	RG/Retain	Revise
6	**C**	37	16	0.90	0.39	0.65	Average/ Retain	0.51	VG/Retain	Retain
7	**A**	22	8	0.54	0.20	0.37	Average/ Retain	0.34	G/Retain	Retain
8	**D**	33	12	0.80	0.29	0.55	Average/ Retain	0.51	VG/Retain	Retain
9	**B**	21	10	0.51	0.24	0.38	Average/ Retain	0.27	RG/Retain	Retain
10	**A**	21	14	0.51	0.34	0.43	Average/ Retain	0.17	MI/ Revise	Revise
11	**D**	30	17	0.73	0.41	0.57	Average/ Retain	0.32	G/Retain	Retain
12	**D**	28	6	0.68	0.15	0.41	Average/ Retain	0.54	VG/Retain	Retain
13	**C**	29	10	0.71	0.24	0.48	Average/ Retain	0.46	VG/Retain	Retain
14	**D**	30	5	0.73	0.12	0.43	Average/ Retain	0.61	VG/Retain	Retain
15	**A**	23	7	0.56	0.17	0.37	Average/ Retain	0.39	G/Retain	Retain
16	**C**	26	7	0.63	0.17	0.40	Average/ Retain	0.46	VG/Retain	Retain
17	**D**	35	17	0.85	0.41	0.63	Average/ Retain	0.44	VG/Retain	Retain
18	**D**	35	13	0.85	0.32	0.59	Average/ Retain	0.54	VG/Retain	Retain
19	**A**	34	20	0.83	0.49	0.66	Average/ Retain	0.34	G/Retain	Retain
20	**D**	30	9	0.73	0.22	0.48	Average/ Retain	0.51	VG/Retain	Retain
21	**A**	37	10	0.90	0.24	0.57	Average/ Retain	0.66	VG/Retain	Retain
22	**A**	39	15	0.95	0.37	0.66	Average/ Retain	0.59	VG/Retain	Retain
23	**A**	24	6	0.59	0.15	0.37	Average/ Retain	0.44	VG/Retain	Retain
24	**B**	35	14	0.85	0.34	0.60	Average/ Retain	0.51	VG/Retain	Retain
25	**C**	27	8	0.66	0.20	0.43	Average/ Retain	0.46	VG/Retain	Retain
26	**C**	27	2	0.66	0.05	0.35	Average/ Retain	0.61	VG/Retain	Retain
27	**B**	35	10	0.85	0.24	0.55	Average/ Retain	0.61	VG/Retain	Retain
28	**B**	25	12	0.61	0.29	0.45	Average/ Retain	0.32	G/Retain	Retain

29	**A**	34	8	0.83	0.20	0.51	Average/ Retain	0.63	VG/Retain	Retain
30	**C**	29	8	0.71	0.20	0.45	Average/ Retain	0.51	VG/Retain	Retain
31	**D**	36	6	0.88	0.15	0.51	Average/ Retain	0.73	VG/Retain	Retain
32	**A**	23	8	0.56	0.20	0.38	Average/ Retain	0.37	G/Retain	Retain
33	**D**	20	4	0.49	0.10	0.29	Average/ Retain	0.39	G/Retain	Retain
34	**D**	24	6	0.59	0.15	0.37	Average/ Retain	0.44	VG/Retain	Retain
35	**D**	24	6	0.59	0.15	0.37	Average/ Retain	0.44	VG/Retain	Retain

Appendix O

OPTION ANALYSIS ON THE SECOND TRYOUT

Plausibility Index of Options Based On the Second Try Out

TEST A

Item No.	Key	A		B		C		D		No Answer		Option Plausibility Index (U - L)							
		U	L	U	L	U	L	U	L	U	L	**A**	Remark	**B**	Remark	**C**	Remark	**D**	Remark
1	**B**	5	11	34	20	2	6	0	4	0	0	-6	Y	14	Y	-4	Y	-4	Y
2	**A**	33	13	1	2	1	3	6	23	0	0	20	Y	-1	Y	-2	Y	-17	Y
3	**A**	19	3	1	10	19	21	2	7	0	0	16	Y	-9	Y	-2	Y	-5	Y
4	**B**	14	8	17	6	8	24	2	3	0	0	6	N	11	Y	-16	Y	-1	Y
5	**A**	38	31	2	3	1	1	0	6	0	0	7	Y	-1	Y	0	N	-6	Y
6	**A**	38	22	1	12	0	3	2	4	0	0	16	Y	-11	Y	-3	Y	-2	Y
7	**D**	17	14	1	14	2	8	21	5	0	0	3	N	-13	Y	-6	Y	16	Y
8	**C**	4	8	4	7	32	21	1	5	0	0	-4	Y	-3	Y	11	Y	-4	Y
9	**B**	3	3	24	17	12	17	2	4	0	0	0	N	7	Y	-5	Y	-2	Y
10	**A**	15	8	5	8	8	4	13	21	0	0	7	Y	-3	Y	4	N	-8	Y
11	**D**	6	19	1	5	2	9	32	8	0	0	-13	Y	-4	Y	-7	Y	24	Y
12	**D**	4	20	11	13	2	4	24	4	0	0	-16	Y	-2	Y	-2	Y	20	Y
13	**A**	23	6	15	15	3	11	0	9	0	0	17	Y	0	N	-8	Y	-9	Y
14	**D**	0	3	10	12	10	20	21	6	0	0	-3	Y	-2	Y	-10	Y	15	Y
15	**B**	6	6	25	5	7	22	3	8	0	0	0	N	20	Y	-15	Y	-5	Y
16	**C**	5	4	2	15	30	21	4	1	0	0	1	N	-13	Y	9	Y	3	N
17	**D**	0	5	0	8	1	13	40	15	0	0	-5	Y	-8	Y	-12	Y	25	Y
18	**C**	1	22	0	7	38	5	2	7	0	0	-21	Y	-7	Y	33	Y	-5	Y
19	**C**	3	17	1	6	36	9	1	9	0	0	-14	Y	-5	Y	27	Y	-8	Y
20	**A**	20	11	6	7	1	2	14	21	0	0	9	Y	-1	Y	-1	Y	-7	Y
21	**B**	3	5	35	22	2	8	1	6	0	0	-2	Y	13	Y	-6	Y	-5	Y
22	**D**	0	2	1	7	1	16	39	16	0	0	-2	Y	-6	Y	-15	Y	23	Y

23	**A**	18	2	7	12	13	19	3	8	0	0	16	Y	-5	Y	-6	Y	-5	Y
24	**D**	2	17	2	11	2	9	35	4	0	0	-15	Y	-9	Y	-7	Y	31	Y
25	**D**	1	4	1	19	3	7	36	11	0	0	-3	Y	-18	Y	-4	Y	25	Y
26	**C**	1	7	6	6	25	8	9	20	0	0	-6	Y	0	N	17	Y	-11	Y
27	**B**	0	4	26	2	1	5	14	30	0	0	-4	Y	24	Y	-4	Y	-16	Y
28	**B**	1	5	18	5	7	10	15	21	0	0	-4	Y	13	Y	-3	Y	-6	Y
29	**C**	0	4	1	12	38	20	2	5	0	0	-4	Y	-11	Y	18	Y	-3	Y
30	**D**	6	14	8	11	0	12	27	4	0	0	-8	Y	-3	Y	-12	Y	23	Y
31	**C**	0	13	8	14	31	8	2	6	0	0	-13	Y	-6	Y	23	Y	-4	Y
32	**A**	23	5	1	12	9	14	8	10	0	0	18	Y	-11	Y	-5	Y	-2	Y
33	**D**	4	15	1	16	11	8	25	2	0	0	-11	Y	-15	Y	3	N	23	Y
34	**D**	4	14	6	16	6	8	25	3	0	0	-10	Y	-10	Y	-2	Y	22	Y
35	**C**	1	9	5	15	29	9	6	8	0	0	-8	Y	-10	Y	20	Y	-2	Y

Legend:

Y – Good Option

N – Not Good Option

Plausibility Index of Options Based On the Second Try Out

TEST B

Item No.	Key	A		B		C		D		No Answer		Option Plausibility Index (U - L)							
		U	L	U	L	U	L	U	L	U	L	**A**	Remark	**B**	Remark	**C**	Remark	**D**	Remark
1	**B**	3	17	37	18	0	6	1	0	0	0	-14	Y	19	Y	-6	Y	1	N
2	**A**	30	18	2	4	1	1	8	18	0	0	12	Y	-2	Y	0	N	-10	Y
3	**A**	27	3	2	11	8	18	4	9	0	0	24	Y	-9	Y	-10	Y	-5	Y
4	**B**	9	11	26	9	3	12	3	9	0	0	-2	Y	17	Y	-9	Y	-6	Y
5	**A**	40	20	1	6	0	7	0	8	0	0	20	Y	-5	Y	-7	Y	-8	Y
6	**A**	35	8	5	11	0	20	1	2	0	0	27	Y	-6	Y	-20	Y	-1	Y
7	**D**	12	20	4	12	4	5	21	4	0	0	-8	Y	-8	Y	-1	Y	17	Y
8	**C**	0	7	6	14	35	9	0	11	0	0	-7	Y	-8	Y	26	Y	-11	Y
9	**B**	4	8	18	2	14	22	5	9	0	0	-4	Y	16	Y	-8	Y	-4	Y
10	**A**	24	5	4	12	5	10	8	14	0	0	19	Y	-8	Y	-5	Y	-6	Y
11	**D**	2	25	1	2	5	7	33	7	0	0	-23	Y	-1	Y	-2	Y	26	Y
12	**D**	1	6	10	23	5	6	25	6	0	0	-5	Y	-13	Y	-1	Y	19	Y
13	**A**	34	6	7	21	0	4	0	10	0	0	28	Y	-14	Y	-4	Y	-10	Y
14	**D**	0	5	5	10	7	15	29	11	0	0	-5	Y	-5	Y	-8	Y	18	Y
15	**B**	8	11	26	14	3	8	4	8	0	0	-3	Y	12	Y	-5	Y	-4	Y
16	**C**	3	9	5	11	32	18	1	3	0	0	-6	Y	-6	Y	14	Y	-2	Y
17	**D**	1	6	0	4	2	9	38	22	0	0	-5	Y	-4	Y	-7	Y	16	Y
18	**C**	3	10	0	10	35	16	3	5	0	0	-7	Y	-10	Y	19	Y	-2	Y
19	**C**	1	7	1	7	36	19	3	8	0	0	-6	Y	-6	Y	17	Y	-5	Y
20	**A**	23	11	7	9	0	2	11	19	0	0	12	Y	-2	Y	-2	Y	-8	Y
21	**B**	0	3	40	30	1	3	0	5	0	0	-3	Y	10	Y	-2	Y	-5	Y
22	**D**	0	8	1	14	2	7	38	12	0	0	-8	Y	-13	Y	-5	Y	26	Y
23	**A**	25	2	5	8	10	28	1	3	0	0	23	Y	-3	Y	-18	Y	-2	Y
24	**D**	0	8	2	15	2	8	37	10	0	0	-8	Y	-13	Y	-6	Y	27	Y
25	**D**	0	8	2	13	3	7	36	13	0	0	-8	Y	-11	Y	-4	Y	23	Y
26	**C**	7	15	2	4	30	9	2	13	0	0	-8	Y	-2	Y	21	Y	-11	Y
27	**B**	0	10	36	11	2	5	3	15	0	0	-10	Y	25	Y	-3	Y	-12	Y
28	**B**	5	7	18	3	11	16	7	15	0	0	-2	Y	15	Y	-5	Y	-8	Y

29	**C**	1	2	2	9	38	22	0	8	0	0	-1	Y	-7	Y	16	Y	-8	Y
30	**D**	0	10	3	19	0	5	38	7	0	0	-10	Y	-16	Y	-5	Y	31	Y
31	**C**	4	14	6	16	27	11	4	0	0	0	-10	Y	-10	Y	16	Y	4	N
32	**A**	22	6	7	7	7	10	5	18	0	0	16	Y	0	N	-3	Y	-13	Y
33	**D**	5	7	5	19	6	13	25	2	0	0	-2	Y	-14	Y	-7	Y	23	Y
34	**D**	2	18	0	10	3	9	36	4	0	0	-16	Y	-10	Y	-6	Y	32	Y
35	**C**	4	5	3	15	24	5	10	16	0	0	-1	Y	-12	Y	19	Y	-6	Y

Plausibility Index of Options Based On the Second Try Out

TEST C

Item	Key	A		B		C		D		No Answer		Option Plausibility Index (U - L)							
No.		U	L	U	L	U	L	U	L	U	L	**A**	Remark	**B**	Remark	**C**	Remark	**D**	Remark
1	**B**	1	0	36	26	1	3	3	12	0	0	1	N	10	Y	-2	Y	-9	Y
2	**B**	1	6	32	15	1	2	7	18	0	0	-5	Y	17	Y	-1	Y	-11	Y
3	**D**	15	24	5	8	0	5	21	4	0	0	-9	Y	-3	Y	-5	Y	17	Y
4	**C**	7	14	3	16	21	7	10	4	0	0	-7	Y	-13	Y	14	Y	6	N
5	**D**	0	4	0	3	3	5	38	29	0	0	-4	Y	-3	Y	-2	Y	9	Y
6	**C**	1	4	3	17	37	16	0	4	0	0	-3	Y	-14	Y	21	Y	-4	Y
7	**A**	22	8	2	11	1	7	16	15	0	0	14	Y	-9	Y	-6	Y	1	N
8	**D**	0	4	1	9	7	16	33	12	0	0	-4	Y	-8	Y	-9	Y	21	Y
9	**B**	0	12	21	10	15	14	5	5	0	0	-12	Y	11	Y	1	N	0	N
10	**A**	21	14	11	12	3	8	6	7	0	0	7	Y	-1	Y	-5	Y	-1	Y
11	**D**	9	13	0	4	2	7	30	17	0	0	-4	Y	-4	Y	-5	Y	13	Y
12	**D**	2	15	3	9	8	11	28	6	0	0	-13	Y	-6	Y	-3	Y	22	Y
13	**C**	2	6	6	16	29	10	4	9	0	0	-4	Y	-10	Y	19	Y	-5	Y
14	**D**	2	7	1	14	8	15	30	5	0	0	-5	Y	-13	Y	-7	Y	25	Y
15	**A**	23	7	5	12	9	12	4	10	0	0	16	Y	-7	Y	-3	Y	-6	Y
16	**C**	3	14	5	13	26	7	7	7	0	0	-11	Y	-8	Y	19	Y	0	N
17	**D**	1	7	2	7	3	10	35	17	0	0	-6	Y	-5	Y	-7	Y	18	Y
18	**D**	2	6	2	15	2	7	35	13	0	0	-4	Y	-13	Y	-5	Y	22	Y
19	**A**	34	20	3	9	1	8	3	4	0	0	14	Y	-6	Y	-7	Y	-1	Y
20	**D**	5	13	0	7	6	12	30	9	0	0	-8	Y	-7	Y	-6	Y	21	Y
21	**A**	37	10	2	24	2	6	0	1	0	0	27	Y	-22	Y	-4	Y	-1	Y
22	**A**	39	15	2	10	0	9	0	7	0	0	24	Y	-8	Y	-9	Y	-7	Y
23	**A**	24	6	9	17	5	13	3	5	0	0	18	Y	-8	Y	-8	Y	-2	Y

24	**B**	2	14	35	14	3	11	1	2	0	0	-12	Y	21	Y	-8	Y	-1	Y
25	**C**	8	16	3	8	27	8	3	9	0	0	-8	Y	-5	Y	19	Y	-6	Y
26	**C**	5	17	4	15	5	7	27	2	0	0	-12	Y	-11	Y	-2	Y	25	Y
27	**B**	2	7	35	10	4	19	0	5	0	0	-5	Y	25	Y	-15	Y	-5	Y
28	**B**	3	7	25	12	8	10	5	12	0	0	-4	Y	13	Y	-2	Y	-7	Y
29	**A**	34	8	4	23	2	4	1	6	0	0	26	Y	-19	Y	-2	Y	-5	Y
30	**C**	1	17	10	11	29	8	1	5	0	0	-16	Y	-1	Y	21	Y	-4	Y
31	**D**	3	9	2	14	0	12	36	6	0	0	-6	Y	-12	Y	-12	Y	30	Y
32	**A**	23	8	4	10	8	16	6	7	0	0	15	Y	-6	Y	-8	Y	-1	Y
33	**D**	5	9	5	14	11	14	20	4	0	0	-4	Y	-9	Y	-3	Y	16	Y
34	**D**	3	13	7	12	7	10	24	6	0	0	-10	Y	-5	Y	-3	Y	18	Y
35	**D**	5	11	3	13	9	11	24	6	0	0	-6	Y	-10	Y	-2	Y	18	Y

Appendix P

INTERVIEW TRANSCRIPTIONS

Note: R – Researcher; S - Student

Day 1 Recording 1
Student Code: H1A

R: I just want to ask you, anong test yung kinuha mo?
S: Test A po.
R: Anong masasabi mo tungkol sa Test A?
S: Ahm…medyo madali lang po kasi yung iba napag-aralan na dati.
R: Ah..about difficulty nya? Mahirap ba siya, madali or tama lang?
S: Tama lang sir.
R: Tama lang (confimatory). Okay, so Test A yung sinagutan mo. Okay, so papakita ko dito sa iyo…diba may tatlong test akong pinatest, so ang napunta sa iyo ay Test A. So, do you recognize these, yung mga items na yan na sinagutan mo?
S: Di ko po naalala sir.
R: Di mo naalala, okay. So, ayan yung ating Test A. Kumuha lang ako ng six (6) items dyan, okay now, ipapakita ko sa iyo yung Test B. This is Test B, okay…and this is Test C okay. So, I'll be giving you time para marecognize yon but ahm…marecognize yung Test A, Test B and Test C. As you can see, number 1 sa Test A, number 1 sa Test B tsaka number 1 sa Test C, they are of the same construct yung tinetest, okay…sige, I'll be giving you time para makita si Test A, si Test B and si Test C.

(Student analyzing the similarities and differences of the tests)

S: Okay na sir.
R: Okay, ngayon, what do you observe sa Test A, sa Test B and Test C?
S: Magkakaparehas lang ng tanong sir.
R: Magkakaparehas lang ng tanong (confimatory). What about Test B… na napansin mo with Test A and Test B?
S: Uhm…yung dito sir, parang may ano…
R: Anong test yan?
S: Illustration.
R: Sa Test B may illustration. What else?
S: Yun lang sir.
R: Yon may illustration. Okay, ah…in terms of…if you're going to look at Test A and Test C, ano yung napansin mo sa kanya?
S: Yung Test C sir ano…komplikado, daming nakalagay na kung anu-ano.
R: Doon sa?
S: Test C.

R: Sa Test C (confirmatory). Ano yung komplikado sa kanya, yung question or yung options?
S: Yung options sir.
R: Options (confirmatory)
S: Daming nakalagay
R: Okay. Now, kung papipiliin with Test A, Test B and Test C, alin sa tingin mo ang mas pipiliin mo na test? Saan ka mas komportable?
S: Test A sir…Test A.
R: Test A, why?
S: Mas madali sir eh, babasahin mo na lang, tapos parang yun lang a-analyze mo lang yung ano…yung question.
R: How about Test B, in terms of difficulty?
S: Tama lang din sir.
R: Tama lang, okay, how about Test C?
S: Test C…parang mahirap sir eh kasi yun nga…parang magulo.
R: Magulo siya, okay. Ah…kung susundan mo yung mga options, sige, I want you to read the options.

(Student reading the options of Test C)

R: Okay, anong masasabi mo sa mga options?
S: Ahm…ano sir, yung parang…mas ano siya, tinuturuan ka dito.
R: Uhm…(confirmatory)
S: Kung ano yung gagawin sa Test C. Pero mas prefer ko pa rin yung Test A.
R: Okay no problem. Thank you very much.
(End of Record)

Day 1 Recording 2
Student Code: L1A

R: So, based doon sa tinest natin, anong test yung tinest mo last time?
S: Tungkol po sa, sir pwede Tagalog?
R: And…ah..yes, yes
S: Ano po, yung sums, about pos a sum, yung kukunin po yung sums tapos yung iba po given po yung,…given po yung sum, pero ang kukuning po yung …is yung part kung paano makukuha yung sum.
R: Ah…anong test yung napapunta sa iyo? Is it Test A, Test B or Test C?
S: Test A po.
R: Test A, so, how do you find Test A, in terms of difficulty?
S: Uhm…(Thinking)
R: Difficult ba siya? Tama Lang?
S: Medyo, o…ano lang po, tama lang po.
R: Tama lang (confirmatory). Okay sige, so ito yung mga items, six (6) items galing sa Test A, do you recognize them?

S: (Nods)
R: Narerecognize mo (Confirming). So ngayon, ipapakita ko sa iyo yung dalawa pang tests. Test B and Test C. Okay, ito si Test B (*giving it to the student*) and then ito si Test C, okay. Bibigyan kita ng time para marecognize yung mga test items, as you can see, yung item 1…item 1 here sa Test A, item 1 sa Test B tsaka item 1 sa Test C they are all asking the same questions.
S: Opo.
R: Okay, you go through 1 to 6, 1 to 6, and 1 to 6 of Test A, Test B and Test C so pare-parehas lang sila, okay.
S: Opo.
R: Sige, I want you to take a look at Test A, Test B and Test C then tatanungin kita later okay, so I'll be giving you time to take a look at Test A, Test B and Test C.

(Student analyzing the similarities and differences of the tests)

S: Okay na po.
R: Okay, so, if you're going to take a test, which type of test do you prefer? Test A, Test B or Test C?
S: Test B po.
R: Test B, why?
S: Kasi yung Test B po kasi, nakalagay na po dito sa…*(inaudible)…*ano…tanong po, yung…anong tawag dito, instruction po niya kung ano po yung gagawin. Kunwari, example po, multiply ($2m - 3n$) tapos ($2m + 3n$).
R: Okay, so anong ibig mong sabihin doon na… sa sinabi mo kanina?
S: Alin po?
R: Doon sa ah…pakiulit nga yung…yung sinabi mo?
S: Yung tungkol po dito sa Test B?
R: Sa Test B, yes.
S: Ano po…madali lang po i-analyze po yung instruction.
R: Okay. Madaling i-analyze yung instruction (confirmatory). Ah, how about Test A?
S: Sa Test A po kasi medyo nahihirapan po ko yung…yung… dito sa part po yung kukunin yung rectangular plot, yung ganoon ganoon po, yung area po ng rectangular plot po tapos po yung pencil cost po.
Pero po dito pos a Test B, pinaikli na lang po yung…yung question kaya madali na lang po i-analyze.
R: Okay, how about Test C? If you're going to, what…what item is that? Number?
S: Number 5 po.
R: Number 5, tingnan natin yung sa Test C.

(Student analyzing no. 5 of Test C)

S: Medyo, ano na din po…ok na din po.
R: Why?

S: Ano po, nandito na din po kung anong equation po yung gagamitin mo, kung divide po ganoon, multiply.
R: So, in term s of Test B and Test C, which do you prefer?
S: Test C.
R: Test C, why?
S: Sa Test C po, gaya ng Test B po, sa Test C po kasi, nakalagay na, nandito na po sa choices po kung anong equation na gagamitin para mas lalo pa pong mapadali…
R: Yung?
S: Mas lalo pa pong mapadali kung paano masasagutan yung question.
R: Okay, thank you very much.
S: Salamat po.
(End of Record)

Day 1 Recording 3
Student Code: L2A

R: If you're going to remember, anong test yung naibigay sa iyo?
S: Test B
R: Test B? Are you sure?
S: A…A
R: Ah…Test A, okay. Sa pagkakatanda mo, how do you find Test A…in terms of difficulty?
S: Ah…
R: Mahirap ba siya?
S: Medyo mahirap.
R: Medyo mahirap on your part, okay so ahm…I'm going to show you yung mga items sa Test A, so ang kinuha ko lang ay six (6) items, okay, so do you recognize those items?
S: Yes.
R: Yes, sa Test A. Ngayon, I'll be showing you yung Test B and Test C, okay, so, this is Test B and this Test C, okay *(giving the tests to the student)*. So, I'll be giving you time para marecognize yung Test A, Test B and Test C. Okay, now yung item number 1 sa Test A, item number 1 sa Test B and item number 1 sa Test C, they are asking the same questions.
S: Questions (confirmatory)
R: So, I want you to go through no's. 1 to 6 of Test A, Test B and Test C, then later on tatanungin kita, okay?

(Student analyzing the similarities and differences of the tests)

R: Okay, now, what can you say about Test A, Test B and Test C?
S: They…they have the same question but they derive in different way of…ah…sentence sir.

R: Sentence, okay. Now, based sa observation mo, if you're going to have a test, what kind of test or test format yung prefer mo? Saan ka mas magiging komportable? Test A? Test B? or Test C?
S: Test C
R: Test C, why? Bakit?
S: Because sir, they…the questions in Test C derives in a form of a sentence.
R: Uhm…
S: So that it can…it can ahm..
R: Sige pwedeng Tagalog.
S: Malinaw syang maintindihan ng kagaya kong estudyante, sir.
R: Why? In terms of? Paano sya naging malinaw?
S: Yung…ah…bali sir yung sagot na agad yung, what is ask na sir, yung…nakalagay na dito sir.
R: Uhm…In terms of instructions sa Test C, malinaw ba yung mga instructions nya?
S: Yes sir.
R: How about the options? What can you say about the options?
S: Ah sir, the equation sir…
R: Uhm…
S: Bali dinerive na rin po yung equation sa Test C.
R: So ibig sabihin?
S: Madali na lang po syang masagutan.
R: Why? Bakit mas madali syang masagutan?
S: Kasi sir napagaral…napagaralan na po namin yung mga ah…equation, bali sir, titignan na lang po namin kung tama yung mga equation dito sir.
R: Okay, ibig sabihin parang susundan mo yung mga steps?
S: Yes sir.
R: And then?
S: Masasagutan mo na sya ng maayos sir, makikita mo na yung tamang sagot.
R: Okay. Thank you very much. Okay, salamat.
(End of Record)

Day 1 Recoding 4
Student Code: L2C

R: If you're going to remember, what test did you take last time?
S: Ano sir, C.
R: Test C. Okay, how do you find Test C?
S: *(Thinking)*
R: In terms of difficulty, how do you find Test C?
S: Kung ano sir yung hinahanap?
R: Madali ba sya? Mahirap? Or tama lang?
S: Madali lang po eh…tama lang po.
R: Bakit naging tama lang siya?

S: Kasi meron po siyang ano sa mga choices…merong mga ano na doon equation na binibigay kung paano mo siya masasagutan ganoon.
R: Okay, sige. I have here Test A, Test B and then Test C. So ito si Test A, ito si Test B and then this is Test C *(giving each test one by one).* So, I want you to take a look and compare Test A, Test B and Test C. Test C, this is the one given to you last time right? Ok, so, si Test A, si Test B and Test C are of the same tests lang, meaning they are asking the same questions. For example, number 1 here in Test A, number 1 in Test B and number 1 here in Test C, they are asking preferably the same…the same questions. So I want you to recognize 1 to 6 of Test A, Test B and then Test C, then you can compare, then tatanungin kita.

(Student analyzing the similarities and differences of the tests)

R: Okay, now, ah…what do you observe about Test A, Test B and Test C?
S: Ano sir, ahm, yung Test A kasi parang malilito ka pa kung paano mo siya sasagutin.
R: Uhm…(confimatory)
S: Tsaka dito po nalilito din po ako ano po, pag may parenthesis.
R: Saang dito po?
S: Test B. Pag may parenthesis, nalilito din po ako.
R: Okay
S: Pero sa Test C, naa-analyze ko po na kung papaano ko po sya sasagutan.
R: Okay, now. Kung bibigyan ka ng pagkakataon mag-test, which kind of test do you prefer?
S: Yung Test C po.
R: Test C, why again?
S: Kasi po ano…meron po siyang ano…yung paano po siya i-solution tapos yung bawat choices nya, ano pagpipilian na lang po kung paano mo siya iso-solve tapos magiisip ka na lang po kung ano yung tamang sagot doon.
R: Okay, sige. Thank You!
(End of Record)

Day 1 Recording 5
Student Code: L3A

R: Based doon sa natatandaan mo na test na binigay sa iyo, anong test yung ahm…ibinigay sa iyo?
S: Test A
R: Test A. How do you find Test A?
S: Ahm…binigay sya sa akin ng teacher ko sa Math *(laughs)*
R: In terms of…I mean in terms of difficulty…difficulty in answering the test?
S: Konti lang…konti lang.
R: What do you mean 'konti lang'?
S: Kasi po yung iba ano…napagaralan na po namin noong nakaraang grade noong…noong…nakaraang grade po ako, Grade 7 or Grade 8.

R: So medyo madali na siya konti.
S: Opo.
R: Okay. I'll be giving you Test A, do you recognize these items? Ah…I just get six (6) items sa Test A. Di ba 35 items yon? Kinuha ko lang ay six items and then here is Test B and here is Test C *(giving the tests to the student)*. Okay so, I'll be giving you time para marecognize yung ah…compare, recognize yung ah…similarities ng Test A, Test B and Test C. But as you can see, number 1 in Test A, number 1 in Test B and number 1 in Test C, they are asking obviously almost the same questions.
S: Yes.
R: Okay so, I'll be giving you a time to recognize 1 to 6 of Test A, 1 to 6 of Test B and 1 to 6 of Test C.

(Student analyzing the similarities and differences of the tests)

S: Okay na sir.
R: Okay now ah…sa tingin mo, ano yung pagkakaiba ng mga tests?
S: Yung mga tanong po tsaka kung…kung ano yung kukunin mong sagot.
R: In terms of ah…doon sa tanong…ano yung ah…pagkakaiba po nila?
S: Yung tanong sir tsaka kung paano mo kukunin…kung papaano mo kukunin yung sagot gamit yung ibang *solve*.
R: Okay sige. If bibigyan ka ng pagkakataon, which among these tests you are comfortable or convenient na sagutan?
S: Test A po.
R: Test A, why Test A?
S: Kasi sir madali lang po siya masagutan konti. Tapos po yung iba *kering keri* naman po syang sagutin.
R: Okay, how about Test B?
S: Test B, masyado po siyang difficult sa tulad ko kasi sir…kasi ako sa totoo lang hindi naman po ako masyadong nakikinig sa Math.
R: Okay. Bakit mahirap yung Test B?
S: Kasi sir, sa ano pa lang…sa tanong, alam mo ng mahirap yun.
R: Okay. Take a look at number 1 in Test A and number 1 in Test B.

(Student analyzing both item no. 1 of Tests A and B)

S: Parang in-ano lang sir yung tanong.
R: Ano po yung pagkakaiba?
S: Yung ano…yung sa Test A po, what is the difference…kung ano po yung pagkakaiba nila, tapos po yung sa Test B, isu-subtract po yung -9 sa -15.
R: Okay, which of Test A and Test B yung ah… sa instructions pa lang…pagkakaiba ng instructions ay ah…mas madaling sundan?
S: Yung Test A sir.
R: Test A, okay. How about Test B…ah…Test C rather?
S: Parang magkaparehas po yung tanong sa Test A sa Test C po…

R: Okay
S: Pero po yung pagpipilian magkakaiba na po siya. Kung paano po i-solve.
R: Kung paano siya?
S: I-solve
R: Iso-solve, so ah…in terms of difficulty between Test A and Test C, which one do you prefer?
S: Test C
R: Oh…sabi mo kanina Test A ang mas madali?
S: People change *(laughs)*
R: Sige sige. Let's go back. Bakit naging Test C naman yung preference mo ngayon?
S: Kasi sir meron na po siyang…kung paano i-solve, tsaka kung paano mo talaga made-determine yung…yung…sagot na kailangan ng tanong.
R: Okay, sa tingin mo mas madali kung bibigyan ka ng mga test na ganito, mas madali syang ah…masasagutan?
S: I don't think so sir.
R: Bakit?
S: Kasi sir, parehas po naman kaya ko silang sagutan, parang…ganoon po.
R: Parehas kaya (confirmatory)
S: Kayang kaya naman.
R: Okay sige. Thank you very much.
S: Sige sir.
(End of Record)

Day 1 Recording 6
Student Code: L2B

R: Based doon sa nareremember mo, anong test yung naibigay sa iyo?
S: Yung ano po…Test B.
R: Test B. Okay, how do you find Test B in terms of difficulty?
S: Sa tagalog sir?
R: Sa Tagalog, papaano mo ah…mm…ma…nakikita yung Test B in terms sa difficulty niya?
S: Ano ba yung tagalong ng difficulty sir?
(Both laugh)
R: Ah, kung mahirap ba siya? Madaling sagutan…ganoon.
S: Actually po, madali naman siya kung nagrereview at kung may natutunan ka naman.
R: Okay, sige. So, ipapakita ko sa iyo si Test A…ito si Test A. Ito yung mga sinagutan mo sa Test B and this is Test C, okay (*giving the three tests to the student*). So actually yung tinest ninyo last time, Test A, Test B and Test C, they are all the same but ah,…namodify lang or nabago lang yung mga questions sa Test B and then sa Test C, may mga nabago lang sa kanila, but they are asking the same ah…questions, okay.

One to six ng Test A tsaka one to six ng Test B and then one to six ng Test C, as you can see, yung one (1) ng Test A, this is also asking the same question in Test B and

this is also asking the same question in Test C. So, bibigyan kita ng time para marecognize yung pagkakaiba ng Test A, Test B and then Test C then later on tatanungin kita.

(Student analyzing the similarities and differences of the tests)

R: Okay na diyan?...ah okay na. Okay now, if you are given a chance na magtetest, which among these tests, Test A, Test B and Test C yung mas prefer mo? Mas komportable kang sagutan?
S: Test B
R: Test B, why?
S: Because, kasi marami akong natutunan sa inyo sir, kasi yung mga natutunan ko sa inyo, sinautak ko, inalala ko at nireview ko.
R: Ah…in terms of instructions or yung directions sa Test B, bakit mas prefer mo siya kaysa sa Test A tsaka sa Test C?
S: Kasi po mas madali siyang sagutan.
R: Mas medaling sagutan. Bakit? Anong meron sa kanya na mas madali siyang sagutan?
S: Kasi po, i…i…ipa-plus lang po, ia-add lang sya.
R: So, ibig sabihin parang binibigay na yung?
S: Sagot.
R: Sagot? Or…mas…ah…mas ano sya?
S: Mas madali.
R: Mas madali yung tanong.
S: *(Nodding)*
R: Okay sige.
(End of Record)

Day 1 Recording 7
Student Code: L3B

R: Okay, based doon sa natatandaan mo, anong test yung ah…tinake mo last time?
S: Ano po…Test B.
R: Test B. Okay, how do you find Test B?
S: Ahm…
R: In terms of how difficult it is to answer?
S: Ahm…hindi naman siya masyado mahirap pero…lahat naman po siguro napagaralan kop o kasi po nagreview naman po ako.
R: Okay ngayon, ipapakita ko sa iyo si Test A and then ito si Test B. One to six lang yung kinuha ko. Six items lang yung kinuha ko sa Test A, six items sa Test B and then six items in Test C.

Okay now, based doon sa tinest ninyo, Test A, Test B and Test C, they are asking the same questions, okay. So, I just want you to recognize si Test A, si Test B tsaka si Test

C. You can compare, one to six of Test A, Test B and then Test C, then later on tatanungin kita.

(Student analyzing the similarities and differences of the tests)

S: Okay na po sir.
R: Okay na. Okay so, if given a chance, which test you are comfortable with?
S: Sa tingin ko po yung Test A.
R: Test A, why?
S: Kasi…medyo madali po siyang sagutin tsaka hindi po masyadong magulo.
R: Hindi masyadong?
S: Magulo.
R: Magulo, in terms of?
S: *(Thinking)*
R: Saan siya magulong part?
S: Yung mga…yung sa Test C po medyo po mahirap anuhin po…
R: Bakit mahirap siyang anuhin…sagutan?
S: Kasi po marami po siyang ano…marami po siyang mga choices na may pare-parehas po.
R: Pare-parehas na?
S: Sagot.
R: Sagot?
S: Ahm…tsaka ano po nakakalito.
R: Okay, ano ang meron sa Test A ah…or anong meron sa Test B na wala sa Test A?
S: Ahm…
R: And then, ano yung pagkakaparehas nila?
S: Yung…pagkakaparehas po siguro yung sagot nila.
R: Okay.
S: Tapos yung…
R: In terms of instruction, halimbawa, no. 1 and then no. 1 here sa Test A and then Test C?
S: Ahm…siguro po ano…uhm…magkaparehas ng sagot pero medyo nakakalito po.
R: Sino yung nakakalito?
S: Yung sa Test C po.
R: Test C, why?
S: Kasi po, marami po siyang ano…ahm…pare-parehas na sagot tsaka ano po…
R: Saan yung pare-parehas na sagot diyan?
S: Dito po.
R: Yung mga…?
S: Yung sign po.
R: Mga sign, mga symbols.
S: Opo.
R: Okay, but yung final answers, magkakaparehas ba?
S: Hindi po.

R: Okay, mas prefer mo si?
S: Test A
R: Test A, okay. Ano naman yung sa tingin mo, yung pagkakaiba ng Test A and then Test B?
S: Yung …mga tanong po.
R: Tanong. Okay, paano ba ang tanong sa Test A kumpara sa tanong sa Test B?
S: Magkaparehas lang po.
R: Pero?
S: Pero mas madali pa rin yung Test A.
R: Test A okay. Thank you.
(End of Record)

Day 1 Recording 8
Student Code: L4A

R: Based on your ah…what you remember, what test did you ah…took last time?
S: Yung Test A po. Sa pagkakatanda ko po, kung hindi po ako nagkakamali yung integers po.
R: Integers okay. How do you find Test A in terms of ah…difficulty? Is it difficult? Or it's just easy? Or average?
S: Average po. Kasi po matagal na po naming siya na-lesson noong Grade 7 pa po kaya medyo nahirapan din po ako.
R: Okay. So, what I'm going to do here, I'm going to show you Test A okay so, ito yung mga tinest mo okay. Anim lang doon yung kinuha ko. Six of the items there out of 35 I've picked it, so yan yung Test A. Do you recognize yung mga items diyan sa Test A?
S: Opo.
R: Medyo natatandaan pa, okay. Now, I'm going to show you Test B. This is Test B, okay and then this is also Test C okay so…as you can see Test A, Test B and Test C, all have the same number of items, six items. And then I want you to recognize Test A, Test B and then Test C and then you can compare items of Test A, Test B and Test C. You can compare item one of Test A, one of Test B, and one of Test C. Okay, so…bibigyan kita ng time para ma-recognize and ma-compare si Test A, Test Ba and Test C.

(Student analyzing the similarities and differences of the tests)

S: Yung sa Test A po tsaka sa Test C, magkapareho pong hinahanap yung difference tapos po yung Test B ah…su-subtract lang po pero pareho pa rin po yung given na nandoon.
R: Okay, so what else did you noticed sa Test A, sa Test B and then sa Test C?
S: Pare-pareho po yung given nila.
R: Uhm…how about the options?
S: *(Analyzing)*

R: Options of ah…, for example, item no. 1, options A, B, C and D, and then Test B no. 1, options A, B, C and D and then Test 1 of C, options A, B, C and D. What do you notice?
S: Pare-parehas.
R: Pare-parehas lang, okay. If given a chance, which test you are comfortable with na sagutan? Convenient sa iyo para sagutan?
S: A po.
R: Test A. Why Test A?
S: Kasi po ano…kasi po…na…na…parang kumportable na po ako dito, kasi…ito na po yung nasagutan ko noong nagtest po.
R: Okay ah…how about Test B, what can you say about Test B na hindi mo nakikita sa Test A?
S: Uhm…
R: In terms of instructions, which instructions is ah…mas madaling intindihin?
S: Dito sir?
R: Oo…you compare Test A and Test B, in terms of instructions.
S: Ahm…sir ulit.
R: Ah, okay…ganito ang tanong. Pagdating sa mga instructions ng Test A and Test B, sino yung mas ah…sa tingin mo ay ah…magiging kumportable ka kapag sinagutan mo siya?
S: Yung Test A po.
R: Test A, why?
S: Kasi po, kagaya po ng sinabi ko kanina, nasagutan ko na po siya noong nagtest.
R: Okay sige. Thank you.
(End of Record)

Day 1 Recording 9
Student Code: L4B

R: Okay so, based doon sa nare-remember mo, ano yung test na tinake mo last time?
S: Ano sir…*(inaudible)*
R: Last week.
S: Yung ano po ah…poly…pono…polo…ah…
R: Poly…
S: Polynomial po.
R: Okay ah…anong …ano yung na-assign na test sa iyo?
S: Yung…
R: Test A, Test B or Test C?
S: Test B po.
R: Test B. How do you find Test B, in terms of ah…in answering the test? Is it difficult?
S: Easy po.
R: Easy? Average?
S: Easy po.
R: Easy, why?

S: Kasi po ano ah…nag…ano po kasi talaga…nagreview po ako.
R. Okay. Nagreview ka talaga?
S: *(Laughs)*
R: Oh sige. Ngayon, ipapakita ko sa iyo yung Test B. Ito yung mga test na sinagutan mo. I've just picked six out of 35, do you recognize some of them?
S: Number 3 po.
R: Number 3, okay so, yan yung anim sa 35 …yung kinuha ko, okay. So, I'm going to show you Test A, ito si Test A *(giving to the student Test A)*. And then ito si Test C, okay *(giving to the student Test C)*.
Actually, Test A, Test B and Test C, they are all the same. They are just asking the same questions okay. So, kung makikita mo, one to six ang Test A, ang Test B one to six, and Test C one to six din, okay. So, yung test 1 ng A and then, number 1 ng A, number 1 ng Test B and number 1 ng C, nagma-match sila…lahat up to six. So I want you to recognize A, B and C and then compare…tingnan mo sila isa-isa, okay so, you can compare them, basahin mo yung mga options, you can look at them, then later, tatanungin kita ha?

(Student analyzing the similarities and differences of the tests)

R: Okay, if given a chance, which of these tests do you prefer? Or mas..saan ka mas kumportable or…ah…convenient na sagutan siya?
S: Sa B po.
R: B. Why B?
S: Kasi po ano…parang ano…ah, kumpleto po yung pagkaka-ano…kasi po may…may parang…ah…parang…may…
R: Figures?
S: Opo.
R: Diagrams. Okay, what else? Bukod sa may figures and diagrams sya…what else? In terms or instructions comparing to Test A?
S: Mas madali po siyang tingnan…*(laughs)*
R: Sino yung mas…oh…which test yung mas madaling intindihin ang instruction?
S: B po.
R: B, okay. How about Test C?
S: Uhm…madami po kasi…
R: Maraming?
S: Madami…ano…solution. Maraming…tinantanong ganoon.
R: Marami siyang…ah…anong tinanong? In terms of…?
S: Parang…ini-isa-isa niya pa po…na…doon po sa tinatanong. Ini-isa-isa niya yung…doon sa solution.
R: Yung mga steps or solutions?
S: *(Nodding)*
R: Okay. So, ang mas prefer mo is Test?
S: Test B.
(End of Record)

Day 1 Recording 10
Student Code: L3C

R: Okay so, based doon sa nare-remember mo, which test yung sinagutan mo?
S: Ang sinagutan ko pong test ay yung…Test C po.
R: Test C. How do you find Test C?
S: Sa akin po parang isa po siya sa mga ordinaryong …mga ordinaryong test para sa akin…ano po kasi…parang ganun pa rin yung itsura ng Math…, may solving, may solutions at kung anu-ano pa na…
R: In terms of…yes…yes…thank you. In terms of ah…difficulty? I-rate….can you ah…rate it if it is difficult? Or just average? Or just easy?
S: Uhm…average.
R: Average, okay. I'm going to show you Test C *(giving to the student Test C)*, anim na items yung kinuha out of 35, do you recognize these 6 items?
S: Yes.
R: Yes. Ganyan yung Test C natin diba? So, anim lang doon yung kinuha ko okay. I'm going to show you Test A, this is Test A *(giving to the student Test A)*, and then this is Test B, okay. Now, as you can see, pare-parehas may anim na items, actually, number 1 of Test A, number 1 of Test B and number 1 of Test C, they are match. Meaning, they are asking the same questions, okay.
So, I want to recognize each items of Test A you compare it with Test B, and then you compare it with Test C. And also the options, you take a look at the options of ah…A, B and C, and then later tatanungin kita, okay? Okay lang ba? So, I'll be giving you time to recognize.

(Student analyzing the similarities and differences of the tests)

R: So, if given ah…or ah…here Test A, Test B and Test C, what do you see?
S: Ang napansin ko lang po…halos wala po silang pinagkaiba.
R: Okay.
S: Or, wala talagang pinagkaiba. Ang napansin ko lang po na pagkakaiba nila eh yung…yung…pagkakadikit-dikit po nung ano po…nung mga…parang mas difficult tignan yung Test C kaysa sa Test A and Test B.
R: Uhm…
S: Kasi po parang mas marami po yung…mas marami pong pagpipilian…yung itsura ng pagpipilian kaysa sa dalawang test na ito.
R: Okay. Now, if given a chance, which among these tests you are comfortable or convenient sa iyo para sagutan? In terms of difficulty?
S: Para sa akin po yung Test B.
R: Test B. Why Test B?
S: Kasi po ano…mayroon po siyang pagkakatulad sa Test A at tsaka Test C…para po siyang pinaghalong Test A at tsaka Test C. May…yung itsura nya po…madali pero yung sagot para pong…wala naman pinagkaiba sa dalawang test na iyon.

R: Okay, in terms of instruction sa Test B, if you're going to compare Test A and Test B, in terms of instruction, which among these…Test A and Test B yung mas madali mong maiintidihan yung hinihingi ng test?
S: Yung Test A po.
R: Test?
S: Test A po.
R: Test A. Why Test A?
S: Kasi po yung Test A…para pong…interactive po yung ano nyo…yung pagtatanong nya po.
R: Uhm…
S: Tapos yung sa Test B po, ano…parang napakasimple po noong pinagagawa…, subtract…halimbawa. Tapos yung ano…yung Test A, mayroon po siyang parang…mayroon kang kausap na tao na…'Ano yung ganito?' Kaysa po sa …'Ganito po dapat.'
R: Okay so, ibig sabihin, mas prefer mong tanungin ka ng Test A or Test B?
S: Test A po.
R: Okay, balik tayo doon. Sa Test A, Test B and Test C, which among these three yung prefer mo?
S: Yung Test A po.
R: Test A. Okay. Thank you!
(End of Record)

Day 1 Recording 11
Student Code: L4C

R: If you're going to remember yung…ah…tinake mo na test last week…ah, anong test yung napapunta sa iyo?
S: C po.
R: Test C. How do you find Test C in terms of difficulty?
S: Uhm…ayos lang sir…hindi siya…
R: Ayos lang…not to difficult? Average? Mga ganoon.
S: Opo.
R: Okay. Ano yung natatandaan mo sa Test C?
S: Yung natatandaan ko lang talaga sir?
R: Yes.
S: Mayroon siyang ano…integers, tapos yung choices niya sir masyadong…every step nakalagay kaya…medyo…
R: Medyo?
S: Magulo po.
R: Magulo, okay. Pero…ah, nakikita mo yung series ng steps or solutions.
S: Opo, yung solution talaga .
R: Okay, sige. Ngayon, ipapakita ko sa iyo si Test A *(giving the copy of the test)*…si Test B and of course yung sinagutan mo, Test C, okay. I want you to recognize each items of Test A, Test B and Test C. I just get 6 of 35 items, six sa Test A, six sa Test B

and six sa Test C. Di ba 35 items yan? So, I want you to recognize ano yung pagkakaiba ng items. Actually, Test A, Test B and Test C,…they are all match, okay. One ng Test A, one ng Test B and one ng Test C, match yon. Meaning, they are just asking the same question. Okay, now I want you to recognize kung ano pa yung pagkakaiba ni Test A, ni Test B at tsaka ni Test C, okay, bibigyan kita ng time.

(Student analyzing the similarities and differences of the tests)

R: Based from what you've observed, ano yung ah…differences ng Test A sa Test B and sa Test C?
S: Yung A po yung nakalagay lang po yung pinaka-main na problem tapos yung mga sagot na po kaagad. Tas yung dito naman po sa Test B, mayroon naman po siyang figure tapos…yon…mayroon po siyang figure na nakalagay kung ano yung measure. Tapos yung dito naman po *(referring to Test C)*, masyadong…sobrang…ano po… detailed na nakalagay…hindi lang po detailed…parang…yung bawat every step po, tapos yung pati rin solution nakalagay sa choices, kaya mukhang mahaba.
R: Okay. But if you're going to ah…take an exam, which of these three tests do you prefer?
S: Test B or…Test B.
R: Test B, why Test B?
S: Meron po kasi siyang figure na pwedeng makuha mo agad yung…parang mas madali po siyang…makakuha ng…solu…pwede po siyang, mas madali po siyang makuhanan ng ano…solution, kasi may figure po siya.
R: Okay, sige. Thank you very much.
(End of Record)

Day 1 Recording 12
Student Code: H4C

R: Okay. Based on what you've remember, anong test yung tinake mo last week?
S: Test C po.
R: Test C. What do you remember about Test C?
S: Ahm…polynomials and…adding, subtracting, multiplying and dividing polynomials.
R: What else?
S: And integers.
R: Integers. Okay ah….doon sa…layout or format noon gating Test C, ano yung natatandaan mo sa kanya?
S: Sa choices niya po ahm…hindi siya direct na…sagot mismo. Ano siya, ahm…ini-isa-isa niya po bawat simplify nung ah…sagot, so parang ine-explain nya po yung sagot bago nya po…bago mo po malaman kung paano nangyari yun.
R: Okay. Now ah…ipapakita ko sa iyo yung Test C, six of the 35 items there kinuha ko. Do you recognize them?
S: Ahm…opo.

R: Okay, nare-recognize mo. Okay, ipapakita kita ko sa iyo si Test A…ito si Test A *(giving the copy of the test)* and then ito si Test B. Now, I want you to recognize each of the items of Test A, Test B and Test C. So, as you can see, anim ang items ng Test A, anim din sa Test B then six sa Test C. Ah…actually, matched yung…matched yung sa Test A number 1, number 1 ng Test B and number 1 ng Test C. Meaning, they are just all the same…they are just asking the same questions. So, I want you to compare each item. Two, two, and two *(pointing to item no. 2 of each tests)* of A, B and C and hangang number 6…and then compare mo sila in terms of instructions and also options…yung makikita mo, and then later tatanungin kita, okay.

(Student analyzing the similarities and differences of the tests)

R: Okay, sige. What do you observe about Test A, Test B and Test C?
S: Ahm…sa Test A po parang nililinaw nya yung questions nya po. Tapos sa Test B, diretso na po siya, kunwari, kagaya ng number 6, multiply ganyan ganyan…so yun po yung sa Test B. Sa Test C naman po, parang parehas lang din siya ng tanong sa Test A pero yung choices niya po kasi…talagang ine-explain nya po kung paano nangyari yung sagot .
R: Okay. Now, if given a chance, which among these tests you are comfortable with na sagutan?
S: Ahm…Test C po.
R: Why Test C?
S: Kasi po kapag sa Test C po kasi ine-explain nya po kasi kung paano nangyari yung sagot na yun eh…so doon pa lang po kun paano nya ginawa, masasagutan mo na po yung pinaka-tanong nya.
R: Okay, sige. Thank you!
(End of Record)

Day 1 Recording 13
Student Code: H1C

R: Based on what you remember…you can answer in English or in Tagalog ha. Based on what you remember, what test did you take last week?
S: About area uhm…quadratic equation?
R: May quadratic equation, may area? Okay. Which test did you ah…take? Test A, Test B or Test C?
S: Test C.
R: Test C. Ano yung natatandaan mo with regards to Test C?
S: Ahm…pinapakita yung mga solution.
R: Uhm…
S: Ang…parang magkapareho lang pero…dapat i-analyze.
R: I-analyze. Medyo dapat i-analyze yung Test C.
S: *(Nodding)*

R: In terms of difficulty, is it difficult, average, easy or what?
S: Average.
R: Average. Tama lang siya, okay. So, ipapakita ko sa iyo…ito si Test C. Anim lang yung kinuha ko out of 35, do you recognize them?
S: Yes.
R: Yes, okay. Now, ipapakita ko si Test B. Ito si Test B *(giving the copy of the test to the student)*, and then this is Test C...ah Test A rather. Okay, so as you can see, anim yung items sa Test A, anim din yung sa Test B and then anim din yung sa Test C. Actually, matched yung lahat ng items niya. Number 1 match sa one ng…number 1 ng Test A match sa number 1 ng Test B, match sa number 1 ng Test C okay. Meaning ah…Test A, Test B and Test C all ask the same questions okay, depende dun sa ah…numbering noong items. Now, I want you to recognize si Test A, si Test B tsaka si Test C and then recognize mo yung instructions…yung options niya and then you compare it to Test B and you compare it to Test C okay. Gets? Bibigyan kita ng time para mag-analyze. Go through every item. Tignan mo yung instructions…yung options.

S: In Test A, the instruction is in sentence.
R: Uhm…
S: And the answer in Test A and Test B is the same…the same in format.
R: Okay. The same yung format. How about sa Test C?
S: In Test C ahm…pinapakita po kung paano kinuha yung sagot.
R: Uhm…
S: Yung…yung question is in sentence? While in Test B is…*(thinking)*
R: You can describe in Tagalog.
S: *(Inaudible)*
R: Ha?
S: *(Inaudible)*
R: How about sa…If you're going to compare the instructions in Test A and the instructions in Test B. What do you observe?
S: *(Looking through the tests)*
R: Oh, for example in number 1. What is the difference of -9 and 15? *(Pointing to Test A)*. And here *(pointing to Test B)*, subtract: (-9) – 15. So, what is the difference? *(Referring to the test instructions)*
S: Sir…ahm…ikaw pa yung gagawa ng parang…solution *(referring to Test A)*, tapos dito *(referring to Test B)* naka-direct.
R: Okay. Direct yung?
S: Direct yung instructions.
R: Instructions sa…test?
S: B.
R: B, okay. If given a chance, which among these tests do you prefer?
S: Dito po?
R: Opo. Sa tatlong yan, alin yung mas convenient ka sagutan and then mas ah…prefer mong sagutan?
S: Ahm…Test B.

R: Test B, why Test B?
S: Ahm…babasahin ko na lang po kasi ito sir….tapos…
R: Tapos? Yes…
S: *(Thinking)*
R: Sige…pwedeng Tagalog. Babasahin mo na lang yung instruction and then?
S: Pag-nakuha ko na, titignan ko na lang yung sagot sa A, B, C and D.
R: Sa tingin mo, ano yung malaking pinagkaiba ni Test A compare to Test B? For example in number 4 *(pointing to Test A)* and in here number 4 *(pointing to Test B)*. Number 4 ng Test A and number 4 ng Test B.
S: Ah…parang dito sa Test A parang iko-collect mo pa yung mga given.
R: Okay.
S: Tapos…eh dito sa Test B ti…lalagay mo na lang yung missing …
R: Missing value?
S: Missing value.
R: Okay. How about number 5 here in Test B and number 5 in Test A?
S: Ahm sir…dito may figure na *(referring to Test B)* habang dito sa Test A wala pa kaso…parang gagawa ka pa ng figure para ma-solve mo yung …yung…no. 5.
R: Okay so, all in all, again I'm going to ask you, which of these tests do you prefer?
S: Test B.
R: Test B. Okay. Thank you!
(End of Record)

Day 1 Recording 14
Student Code: H2C

R: Okay, based on what you remember, anong test yung tinake mo last week?
S: Parang test na tungkol sa pinag-aralan namin noong Grade 7.
R: Grade 7. Okay ah…which format A, B or C?
S: C.
R: C. Ano yung natatandaan mo sa Test C?
S: Parang puro siya process.
R: Processes, okay. Ano pa?
S: Process then ano…parang ano…magkakaiba lang yung options sa ano…sa…magkakaiba lang siya ng…sa ginamit na pag ano po…yung ano sir, addition, multiplication doon lang siya magkaiba tapos…pero ano kahit na…pagtinignan mo siya yung kahit..yung mga sagot naman magkakaiba pa rin.
R: Okay, meaning pinapakita niya yung?
S: Yung process.
R: Yung process ng pag?
S: Pagso-solve.
R: Okay, so…thank you. Now, I want you to take a look at Test C. This is Test C *(giving the test to the student)*. I just get six items out of 35 items there, do you recognize them?
S: Yes,

R: Okay. Now, I want you to take a look at Test A *(giving the test to the student)*. And then, take a look at Test B *(giving the test to the student)*. So actually, Test A, Test B and Test C they are all the same. They are asking the same questions, okay. So as you can see, sa Test C kung ano yung tanong sa Test C, this is also being asked in Test A and then this is also being asked in Test B. But Test B and Test C are just modified okay. Medyo may binago lang sa Test B and sa Test C, okay. So, I want you to recognize each test, Test A, Test B and then Test C and then you compare each ah…each items. Tignan mo kung may pagkakaiba sa instructions and then also sa options, then later on tatanungin kita, okay? You can go through one by one sa every item.

(Student analyzing the similarities and differences of the tests)

R: Based on what you have observed from Test A, Test B and Test C okay so ah….what is your observation on the similarities and differences of Test A, Test B and Test C?
S: Yung similarities po…yung mga question po pare-parehas lang po. Tapos yung ano…bali magkakaiba naman yung…yung Test A kasi parang simple lang siya. Nandoon lang yung…yung sagot na yung mismong nadoon sa options. Yung sa Test B naman parang more on visual.
R: More on?
S: Visual.
R: Okay.
S: Tapos yung sa Test C naman puro siya process.
R: Processes.
S: Pero kung titignan po yung mga tanong...ay yung mga sagot…pare-parehas lang naman.
R: Okay, pare-parehas. Okay, now if given a chance, which of these tests do you prefer most?
S: Siguro po yung Test B po.
R: Test B. Why Test B?
S: Kasi po ano…yung mga kabataan ngayon mas visual learners po sila.
R: Okay.
S: So pag Test B po yung kinuha naming…yung ano yung…concept na pagkukuha ng…pagkukuha ng.. test, siguro po mas effective po yung result na makukuha namin.
R: Sa test?
S: Test B.
R: B, okay. If you're going to compare Test A and Test B, in terms of instructions, which do you prefer?
S: Yung Test B pa rin po.
R: Test B pa rin, okay. Why?
S: Eh kasi po dito sa Test A po…yung…yung tanong po sa kanya kailangan mo po siyang…kailangan mo po siyang…magisip, kailangan mo pa siyang i-imagine. Sa Test

B hindi na po kasi nakalagay na po dito yung…yung mismong object na…na dine-define sa Test A.
R: Okay. Thank you very much!
(End of Record)

Day 2 Recording 1
Student Code: L1C

R: Based on what you remember, anong test yung binigay sa iyo last week?
S: About integers and multi…multiplication of integers, dividing and subtracting
R: You can speak in English or in Tagalog.
S: Yung lang po yung alam ko sir eh.
R: Okay. What test has been given to you, Test A, Test B or Test C?
S: Test C.
R: Test C. How do you remember Test C?
S: Ahm…about dun siya…yung pag…how to…how to subtract and divide the integers then, ano pa ba yung iba doon?
R: Okay, based on what you remember, how do you find Test C in terms of how difficult it is? Is it too difficult? Is it just average… or easy?
S: It's just…average.
R: Average. Okay, I'm going to show you Test C, okay *(giving the copy of the test)*. I just get 6 out of 35 items there of Test C okay, do you recognize some of the items there?
S: *(Shaking head)*
R: Hindi na. So that is Test C. But you remember Test C being too long? The options are too long.
S: Yes sir.
R: Okay, so that is Test C. And this is Test A *(giving the copy to the student)*. Okay and this is Test B *(giving the copy to the student)*. Okay, now I want you to recognize Test A, Test B and Test C okay. As you can see, Test A has 6 items, Test B has 6 items, and then Test C has also 6 items okay. Actually, Test A, Test B and Test C they are all the same. They are all asking the same questions. For example, Test A number 1, Test B number 1 and then Test C number 1, they are asking the same question. Okay, so I want you to go over with Test A, Test B and then Test C…and then try to compare Test A, Test B and Test C.
S: *(Going through the tests)*
R: You can go over with the items…per item. Ah…tignan mo yung pagkakaparehas nila sa instructions and then sa alternatives. Bibigyan kita time then later on tatanungin kita, okay?

(Student analyzing the similarities and differences of the tests)

R: So…based on your observation about Test A, Test B and Test C ah…what are their similarities and differences?

S: Ang pag…ang…ang kanilang pagkakaparehas ay yung sagot nila pero yung question nila…ay parehas lang yung question…yung explanation lang doon sa…explanation doon sa choices pero yung pagkakaiba…
R: Which test is that?
S: Test A and B.
R: Test A and B. What do you mean “explanation sa sagot”?
S: Ano…parehas lang sila ng concept ng…sagot. Parehas lang sila ng concept ng …doon sa pagkaka-explain doon…
R: Okay.
S: Maliban doon sa Test C.
R: Ano naman yung pinagkaiba ng Test C?
S: Yung pagkakaiba ng Test C ay ano…meron…yung ano…meron siyang explanation hindi kagaya sa Test B and tsaka sa Test A.
R: Ano yung…saan part siya may explanation?
S: Doon sa choices.
R: Doon sa choices. Okay, if given a chance which of these tests ang comfortable ka na sagutan? And you think magkakaroon ka ng…
S: Test A.
R: Test A. Why Test A?
S: Kasi mas madali po siyang intindihin kaysa doon po sa Test C. Kasi po yung Test C na nasagutan ko ay…nakakalito kasi halos…halos parehas lang sila ng explanation magkakaiba lang sila ng sagot.
R: Okay, sige.
(End of Record)

Day 2 Recording 2
Student Code: L1B

R: Based on you remember, anong test yung tinake mo last week?
S: Tagalog…English?
R: You can speak in English or you can speak in Tagalog.
S: Tungkol sa…about doon sa…trigo…ano yun…trigonometric ba yun sir?
R: Is it trigonometric expressions? No. But the test you took last week?
S: Ahm…about sa ano sir…*(inaudible)*.
R: Okay, which test did you ah…take? Test A, Test B, Test C?
S: Test B.
R: How do you remember Test B?
S: About ano sir…about…PEMDAS sir?
R: PEMDAS, mga ganoon. Sige, how do you find Test B in terms of difficulty? Is it difficult, average, easy?
S: Average lang sir.
R: Average…tama lang, why?
S: Yung iba sir kasi yung iba may mga madali, yung iba mahihirap kailangan pang i-compute, yung iba i-mental na lang sir.

R: Okay, sige. Now, this is Test B *(giving the copy to the student)*. Ito yung kinuha mo last week, okay. Anim doon out of 35 yung kinuha ko, okay, do you remember or do you recognize some of this items?
S: Yes.
R: Yes, okay. Now, I'm going to show you Test A *(giving the copy to the student)*. This is Test A…and then this is Test C *(giving the copy to the student).* Okay, as you can see…may six items sa Test A, may six items sa Test B at may six items sa Test C, okay. Actually, Test A, Test B and Test C they are measuring the same thing or they are asking the same thing, okay. Now, I want you to recognize each item of Test A, each item of Test B, each item of Test C and then you compare. For example, one of Test A…one ng Test B, one ng Test C, i-compare mo sila, okay…hanggang to the sixth item, okay. Then, I'll be giving you time then later on I'll be asking you questions. Is it okay?

(Student analyzing the similarities and differences of the tests)

R: Okay, so based from what you observed, what do you think are the similarities and differences of Test A, Test B and Test C?
S: Yung pagkakaparehas nila sir is yung question nila tapos ano… dine-describe niya rin yung paano i-solve yung…item na kailangan sagutan. Tapos ang…yun lang sir.
R: Okay, so…if given a chance, which of these tests do you think na comfortable ka?
S: Test C sir.
R: Test C. Why Test C?
S: Ano na…may…computation na, nandito na lahat ng ano…sagot. Tapos hindi mo na kailangan mag-solve andito na sir.
R: Okay, so sa tingin mo madadalian kang i-solve yung problem sa Test C?
S: Yes sir.
R: Okay. Thank you very much!
(End of Record)

Day 2 Recording 3
Student Code: H2A

R: Based on what you remember last week, what test did you ah…take?
S: Ahm…test about mathematics
R: Mathematics… specifically ah…what set of test did you take?
S: Set A.
R: Set A, ok ah…how did you remember set A?
S: Ahm…meron po siyang mga addition and subtraction ng positive and negative numbers, tsaka ano po multiplication and division.
R: Ok, how did you find Test A?
S: Medyo madali po.
R: Medyo madali siya. Ok, now ah…ipapakita ko sayo…I have here Test A. Six of the 35 items kinuha ko sa kanya *(referring to Test A)*. Do you recognize these items?
S: Opo.

R: Yes. Ngayon, I'm going to show you Test B and then Test C. Ok, so as you can see Test B and Test C have also the same number of items, 6 items. So…actually Test A, Test B and Test C they are asking the same questions or the same construct. Okay, now I want you to recognize Test A, Test B and then Test C, then you can compare. For example, compare number 1 of Test A to number 1 of Test B and then number 1 of Test C and then take a look at the instructions and also the options so…take a look at the similarities and the differences and then later on I'll be asking you something okay?

(Student analyzing the similarities and differences of the tests)

R: Okay, so based from your observation, what can you say about the three tests?
S: Yung Test B po more on illustration po siya…mas pinapakita po kung ano po yung object na yun ganoon po tapos yung Test C po parang mas ini-isa-isa, mas pinapaliwanag yung…
R: Saan siya pinapaliwanag?
S: Kung ano po yung ginagawa doon or …
R: The processes?
S: Opo.
R: Okay.
S: Sa Test A po.
R: Saan siya pinapaliwanag sa mga…?
S: Ahm…pinapaliwanag po siya, ano po…parang mas madali po kasi siyang sagutan kasi mas andito na po yung parang keywords para masagutan mo yung tamang answer po.
R: Okay, how about Test A?
S: Ano po siya…parang simple lang po siya…pero hindi po siya ganoon ka-elaborate …parang kung ano yung andiyan yun na po yung sasagutan.
R: Ah okay, now if you're going to ah…take a test, which of these tests do you prefer most, where in ah…the difficulty…in terms of difficulty mas madadalian ka sagutan, you are comfortable with?
S: Ito pong Test A.
R: Test A, why?
S: Kasi po parang mas maliwanag na rin naman po siya, mas simple po siyang tingnan at madali rin po siyang ma-gets.
R: Okay. Salamat!
(End of Record)

Day 2 Recording 4
Student Code: H3A

R: Okay, so based on what you have remembered last week, which test or set of test did you take?
S: Set A po.
R: Set A, okay. How do you find set A in terms of ah…difficulty?

S: Medyo okay naman po. Mayroon pong mga natatandaan kung paano iso-solve pero mayroon nakaya naman po.
R: Ah, okay…so I'll be showing you six of the items in Test A so…I just ahm…get 6 out of 35 items there *(referring to Test A)*. Do you recognize these items?
S: Yes po.
R: Okay. Now, I'm going to show you Test B. This is Test B *(giving the copy to the student)* and then this is Test C, ok. So if you ah…if you're going to observe si Test B, six items. Si Test C six items then okay,…so…in ah…actually Test A, Test B and Test C, they are asking the same questions. Test B and Test C are just modified. Okay, now I want you to recognize the differences and the similarities between ah…for example, item number 1 of Test A, item number 1 of Test B and item number 1 of Test C…and then also take a look at the instructions and also the options. Okay, I'll be giving you time to recognize the similarities and differences sa instructions tsaka sa options and then later I'll be asking you, okay?

(Student analyzing the similarities and differences of the tests)

R: Based from what you've observed in Test A, Test B and Test C ah….what can you say about the similarities and differences of these tests?
S: So yung…parang magkakapare-parehas lang po sila ng hinihinging sagot sa…ano po yung dapat gawin. Parang nagkakaiba lang po sila kung paano po yung way ng pagtatanong po or paano po yung way ng pagkakasulat kung ano po yung tanong parang sa Test A po.
R: Okay.
S: Parang buong buo po siya, hindi katulad sa Test B straight to the point po na kung ano po yung dapat gawin yun na po yung nakalagay po…
R: Uhm…
S: So pagtiningnan naman po sa Test C, yung choices po kumpara sa Test A tsaka sa Test B parang naka-recognize po kung ano po yung way para makuha po yung final answer o kung paano mo po nakuha yung sagot mo.
R: Ah okay…now, if given a chance, if you're going to take an exam which of these tests would you like to or which of these tests you are comfortable or convenient with?
S:Yung Test B po.
R: Test B, why Test B?
S: Parang mas okay po na straight to the point na po yung …parang wala na pong paligoy-ligoy. Kung ano po yung dapat kunin…paanong yon na po yung kukuhanin.
R: Ah okay, do you think mas masasagutan mo na siya ng maayos if the instruction is direct?
S: Yes sir.
R: Yes. Okay, thank you!
(End of Record)

Day 2 Recording 5
Student Code: H4A

R: Okay, based on what you remember last week ah…which set of test did you take?
S: Set A
R: How did you find set A?
S: Sakto lang sir…okay naman.
R: Sakto. In terms of difficulty?
S: Ah…mga medium.
R: Mga medium or ah…average lang okay, how did you remember set A?
S: About ah…integers, squares and cubes.
R: Okay, integers, squares and cubes okay. I will ah…give you set A. This is set A, 6 out of 35 items kinuha ko sa kanya just to represent set A. Do you recognize these items?
S: *(Nodding)*
R: Okay, now I'm going to show you Test B. This is Test B *(giving the copy to the student)* and then this is Test c. Okay now ah…as you can see Test B and Test C have the same number of items, 6 items. So, basically Test A, Test B and Test C they are asking the same questions, okay. For example number 1 of Test A, number 1 of Test B and number 1 of Test C, they are asking the same questions or they are requiring you to find the same answer.
S: Answer *(in unison)*
R: Okay, I want you to just recognize Test A, recognize Test B, recognize Test C and then you can compare them one by one. One of A, one of B, one of C *(referring to item number 1 of the tests)*. Two of A, two of B, of C…and then identify the similarities and differences of the instructions and of course the options. Okay, so I'll be giving you time to recognize.

(Student analyzing the similarities and differences of the tests)

R: Okay, now based on what you've observed, what are the differences and similarities of these tests?
S: Example na po doon sa number 1. Yung keywords po pero syempre ano…subtraction kasi di ba yung difference po tapos ganoon din po sa Test C pati doon sa Test B so parang nagiiba lang po yung statement pero same answers nga po yung hinahanap so yung keywords lang po yung nagkakaiba pero yun po same answers po.
R: Okay. Now, if given a chance which among these tests you are comfortable with or convenient with to answer?
S: Test A sir.
R: Test A. Why Test A?
S: Because sir na-take ko na siya.
R: Na-take mo na siya…okay in terms of instructions, would you take Test B as a test?
S: Pwede rin sir kasi ano…ano siya parang less words.
R: Less words. Okay, thank you.

(End of Record)

Day 2 Recording 6
Student Code: H3C

R: Okay, so based on what you remember last week, which ah…test did you take?
S: Test C.
R: Test C. How do you remember Test C?
S: Ahm…mayroon siyang mga steps kung paano i-add yung mga integers o kaya i-subtract.
R: Okay…yung mga steps or processes okay ah…how did you find Test C in terms of difficulty? Is it difficult? Average? Easy?
S: Average po.
R: Average.
S: Nakakalito lang.
R: Okay. Nakakalito okay so ah…I'll be showing you Test C. Six out of 35 items kinuha ko sa kanya *(referring to Test C)*. Do you recognize these six items?
S: Yes.
R: Yes, okay. I'll be showing you Test A *(giving the copy to the student)*. And then this is Test B. Now ah…as you ah…observe yung Test A has 6 items, Test B has 6 items, and then also Test C has 6 items okay. So… actually Test A, Test B and Test C they are asking the same questions okay. So si Test A for example number 1 ng Test A, number 1 ng Test B and number 1 ng Test C, they are ah…requiring you to answer the same problem okay. So, I'll be giving you time to recognize Test A, Test B and then Test C. You can compare them…compare mo yung instructions, compare mo yung mga options and then later on I'll be asking you, okay?

(Student analyzing the similarities and differences of the tests)

R: Now, from what you have observed ah…what are the similarities and differences of these tests?
S: Parang pare-pareho lang sila ng sagot pero iba-iba yung ginamit na mga salita.
R: Okay, ah…what else?
S: Tsaka sa Test C…parang …parang pinahirapan.
R: Pinahirapan? In what sense pinahirapan siya?
S: Kasi maraming mga nakasulat?
R: Okay. Anu-ano yung mga nakasulat sa kanya? *(Referring to Test C)*
S: So…kunwari, ito yung A *(referring to option A of Test C)* -9 minus 15 tapos mayroon pang change sign ng 15.
R: Uhm…
S: Tapos ia-add mo po -9 plus -15.
R: Okay, so ah…if given a chance, which among these tests you are comfortable with to answer?
S: Siguro yung Test B.

R: Test B. Why Test B?
S: Kasi diretso na siya mayroon pang mga…
R: Ano yung diretso?
S: Ahm…sakto na yung pagkakasabi noong…
R: Sakto na yung pagkakasabi ng?
S: Equation noong
R: Equation *(confirmatory)*
S: Parang sinasabi na kaagad kung ano yung pinaka-operation na gagamitin.
R: Ah okay…so meaning the instruction is direct? Simple?
S: Yes.
R: Ah okay…sige.
(End of Record)

Day 2 Recording 7
Student Code: H1B
R: Okay, so based on what you remember last week, which set have you taken?
S:Set B po.
R: Set B. How did you find set B?
S: Ahm… I think it is more easy and I can... I can answer it enough ahm… knowledge because it's …it ask direct point.
R: Okay, now I'll be showing you Test B. Six of the items ah… in Test B kinuha ko sa kanya. Okay, do you recognize some of these items or all of these items?
S: Yes po.
R: Now I'm going to show you Test A. This is Test A *(giving the copy to the student).* It has six items. Test B has six items and then Test C has six items. I just get 6 out of 35 from all of these tests.
Okay, now ahh…in ah… actually Test A, Test B and Test C they are all asking the same questions. For example if you're going to look at number one of Test A, number one of Test B and then number one of Test C they are requiring you ah… of the same problem. Okay, so… I'll giving you time to ah…recognize Test A, Test B and Test C and also I'll be requiring you to compare the instructions on Test A, Test B and then Test C and also take a look on the options of Test A, B and C okay?

(Student analyzing the similarities and differences of the tests)

R: So what have you observed about these tests?
S: Sa una po yung Test B parang yung… question nya po ay kung ano lang mismo yung gagawin and direct. Tapos ganoon din yung choices…parang mas madali po siyang sagutan kasi may given pong mga…may mga given pong mga …
R: Figures?
S:Figures.
R: Okay.
S: Mas madali pong intindihin and,…Test A and Test C parang… parehas, nag kakaiba po sila sa choices.

R: Okay.
S: Madali pong sagutan yung sa Test A kasi kung ano lang po yung…ano… halimbawa may mga technique ako kung paano mas mapabilis na masasagutan… hindi na po pala choices na po agad. Sa Test C po parang ang dami mo pang kailangan basahin tapos parang… mas nakakalito po siya.
Tapos ano… kailangan mo po talagang ano… sundin yung method na ano… of yung process kasi halimbawa kung may technique or method para mas mapabilis yung sagot. Pero dito sa Test C parang kailangan mo pa isa-isahin yung mga ano… yung mga choices.
R: Ah… ok so, if given a chance which among these tests you are comfortable with to answer?
S: Isa lng po? Test B po.
R: Test B, why Test B?
S: Ahm… maliban po sa ano … sa direct po yung tanong ahm… mas madali po siyang intindihin at… yoon po mas madali ko siyang intindihin.
R: Okay, thank you.
(End of Record)

Day 2 Recording 8
Student Code: H2B

R: So based on what you remember on last week's test, which set have you taken?
S: Yung set B po.
R: Set B. Okay, how do you find set B in terms of difficulty?
S: Sakto lang po yung …mayroon pong items na medyo hindi ko po makuha…pero mayroon din pong nakukuha agad.
R: Okay so…on a scale, average lang siya. So, I'm going to show you Test B *(giving the copy to the student)*…so, 6 out of 35 items yung kinuha ko sa kanya *(referring to Test B)*. Okay, so do you recognize these six items on set B?
S: Yes po.
R: Okay, so…I'm going to show you Test A *(giving the copy to the student)*. Six out of 35 yung kinuha ko and then this is Test C *(giving the copy to the student)*. Six out of 35 kinuha ko din sa Test C, okay?
S: *(Nods)*
R: So, as you can see parehas may mga six items sila. Now, Test A, B, C actually are asking or requiring you to answer the same problems so…may ma modifications lang sa Test B and Test C. So I want you to recognize these tests individually …per item. Tingnan mo yung instruction for example, number 1 of Test C, you take a look at the instructions and also take a look at the options. You compare and identify what are the similarities in these tests. Identify also, what are the differences in these tests.

(Student analyzing the similarities and differences of the tests)

R: Okay na?
S: Okay na po.
R: So, what do you observe about the differences and similarities of these three tests?
S: Pare-parehas lang naman po sila. Ang pinagkaiba lang yung pagkaka-construct ng problem. Yung sa Test A po parang…pinupunto niya na yung hinahanap, tapos yung sa Test B naman po, parang mas madali po siyang intindihin. Tapos yung Test C naman po mahirap siyang intindihin parang ano …parang…
R: Options.
S: Option po parang mas pinahaba siya kasi po step by step.
R: Pinapakita yung steps or processes?
S: Opo.
R: Okay, if you're given a chance which among this tests you are comfortable or convenient na sagutan?
S: For me?
R: Just one.
S: For me, Test B.
R: Test B, why Test B?
S: Mas madali po siyang intindihin parang…pag tiningnan mo po siya maiso-solve mo po siya agad.
R: Okay, in what way madali siyang sagutan?
S: Parang ano lang po…ah…given na po yung gagawin kasi po sa Test A ia-analyze mo pa po? Tapos sa Test C parang…yung…ia-analyze pa po yung questions at ang options and sa Test B naman po mas madali po siya kasi…
R: Sa instructions, or in terms of instructions?
S: In terms of instruction po mas madali siyang intindihin.
R: Okay.
(End of Record)

Day 2 Recording 9
Student Code: H3B

R: So based on what you remember last week, anong test yung kinuha mo?
S: Test B po.
R: Test B, okay so…how do you remember Test B?
S: Ahm…okay naman po siya. Yung question naman po niya simple lang, makukuha mo agad yung answer na sa choices na po.
R: In terms of difficulty?
S: Madali lang.
R: So…on a scale, average lang?
S: Opo.
R: So, I'm going to show you Test B *(giving the copy to the student)*. Six out of 35 item yung kinuha ko sa Test B so…do you recognize these six items on Test B?
S: Opo.

R: Okay now, I'm going to show you Test A *(giving the copy to the student)* and this is also Test C *(giving the copy to the student)* okay? As you can see Test A, Test B and Test C, pare-parehas silang six items okay? So actually Test A, B, C, they are asking or requiring you to answer the same problem. So, I want you to recognize Test A, B, C…per item discriminate or differentiate. For example number 1 of Test A, number 1 of Test B, number 1 of Test C in terms of instructions and of course sa mga options. So bibigyan kita ng time para makita ang similarities and differences, okay?

(Student analyzing the similarities and differences of the tests)

S: Okay.
R: What do you observe?
S: Para po sa akin, in terms of questions po yung difference ng question is Test A and Test B, yung kapag naman po sa choices is yung Test C…yung sa Test A po kasi kumbaga yung pagkanagtatanong o yung tanong dito is…hindi siya direct kumbaga yung ahm…kunwari po yung sa one *(referring to no. 1 of Test A)* is 'What is the difference of -9 and 15?'. So yung sa Test B naman po is direct, subtract na agad po yung sinabi kaya alam na po agad noong bata na kung ano ang gagawin, subtract. Yung sa 'what is the difference po iisipin niya pa po ano ba yung difference di ba? Parang ganoon, tapos yung sa Test C naman yung…although yung question po sa Test C and Test A the same lang… yung choices po is…parang pinakita niya po kung paano kinuha doon po yung sagot niya po step by step para mas maintindihan nila.
R: Okay now, if you are given a chance which of these tests you are comfortable with or convenient with to answer?
S: I take po with Test B. Kasi po as a student, mas gusto ko po doon sa direct and mas simple lang po yung questions and hindi mahihirapan ang estudyante sa pagsagot lalo na sa Mathematics…kasi po yung iba dito more on words…mas okay pa po sa question dito na more on figures *(referring to Test B)*
R: So you go with Test B because of the figures and also yung instructions direct?
S: Opo.
R: Okay, sige. Thank you.
(End of Record)

Day 2 Recording 10
Student Code: H4B

R: So based on what you remember, anong set yung ah… tinake mo na test?
S: Set B po.
R: How do you find Test B?
S: Medyo may kadalian po.
R: Medyo may kadalian? So, on a scale…average lang siya okay?
S: *(Nodding)*

R: Now, I'm going to show you Test B… so this is Test B *(giving the copy to the student).* Six out of 35 items kinuha ko sa Test B okay...so do you recognize all of these items?
S: Opo.
R: Okay so now, I'm going to show you Test A…so this is Test A *(giving the copy to the student)* and then this is Test C. So… as you can see Test A, Test B and Test C pare-parehas six items okay?
S: Okay po.
R: Ah...which means to say match yung number 1 sa Test A at sa number 1 ng Test B and then number 1 ng Test C… meaning they are asking the same questions. So I want you to recognize all of these items and then pakitingnan yung differences sa instructions ng mga tests and also look at the options. Okay so, bibigyan kita ng panahon para ma-recognize ang similarities and differences sa instructions and sa options, okay?
S: Okay po.

(Student analyzing the similarities and differences of the tests)

R: Ok so...what do you observe about these three tests?
S: Para sa akin po mas madali yung Test B.
R: Why?
S: Kasi po 'pag dito po sa Test A…mahirap po siya. Mayroon po siyang letters. Dito naman po sa Test C masyado po siyang masakit sa mata kasi po napakaraming given,…pareparehas naman po yung questions pero sa pagpipilian po masyado po siyang nakakalula.
R: In what sense nakakalula?
S: Parang… ang hirap.....tinatamad kasi kami magbasa.
R: Ok.
S: Pero dito po sa Test B mas kaunti po yung question. May mga drawing pa po siya so… mas madali pa po sya maiintindihan ng mga estudyante...kaya para sakin po mas madali ang Test B.
R: Okay...so if given a chance, which among these three tests you are comfortable with or convenient with to answer?
S: Test B po.
R: Test B. Again, why Test B?
S: Kasi po yung Test B mayroong po siyang question at mayroon po siyang drawing kaya madali pong intindihin.
R: Okay.
S: Kumbaga… less scratch po siya.
R: Okay, salamat.
(End of Record)

Appendix Q

SPSS OUTPUT

A. Intraclass Correlation Coefficient

A.1. Test A

Criteria A (Content)

Case Processing Summary

		N	%
Cases	Valid	5	100.0
	Excluded [a]	0	.0
	Total	5	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.929	6

Intraclass Correlation Coefficient

	Intraclass Correlation [a]	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.649[b]	.308	.944	14.091	4.0	20	.000
Average Measures	.917[c]	.726	.990	14.091	4.0	20	.000

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. Type A intraclass correlation coefficients using an absolute agreement definition.

b. The estimator is the same, whether the interaction effect is present or not.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

Criteria B (Format and Layout)

Case Processing Summary

		N	%
Cases	Valid	5	100.0
	Excluded [a]	0	.0
	Total	5	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.812	4

Intraclass Correlation Coefficient

	Intraclass Correlation [a]	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.438[b]	.058	.891	5.308	4.0	12	.011
Average Measures	.757[c]	.173	.970	5.308	4.0	12	.011

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. Type A intraclass correlation coefficients using an absolute agreement definition.

b. The estimator is the same, whether the interaction effect is present or not.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

Criteria C (Language Suitability)

Case Processing Summary

		N	%
Cases	Valid	5	100.0
	Excluded [a]	0	.0
	Total	5	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.750	2

Intraclass Correlation Coefficient

	Intraclass Correlation [a]	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.600[b]	-.305	.948	4.000	4.0	4	.104
Average Measures	.750[c]	-.857	.973	4.000	4.0	4	.104

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. Type A intraclass correlation coefficients using an absolute agreement definition.

b. The estimator is the same, whether the interaction effect is present or not.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

TEST A OVERALL INTRACLASS CORRELATION COEFFICIENT

Case Processing Summary

		N	%
Cases	Valid	5	100.0
	Excluded [a]	0	.0
	Total	5	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.918	12

Intraclass Correlation Coefficient

	Intraclass Correlation [a]	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.467[b]	.192	.888	12.245	4.0	44	.000
Average Measures	.913[c]	.740	.990	12.245	4.0	44	.000

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. Type A intraclass correlation coefficients using an absolute agreement definition.

b. The estimator is the same, whether the interaction effect is present or not.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

A.2. Test B

Criteria A (Content)

Case Processing Summary

		N	%
Cases	Valid	5	100.0
	Excluded [a]	0	.0
	Total	5	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.420	6

Intraclass Correlation Coefficient

	Intraclass Correlation [a]	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.099[b]	-.080	.674	1.724	4.0	20	.184
Average Measures	.396[c]	-.787	.925	1.724	4.0	20	.184

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. Type A intraclass correlation coefficients using an absolute agreement definition.

b. The estimator is the same, whether the interaction effect is present or not.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

Criteria B (Format and Layout)

Case Processing Summary

		N	%
Cases	Valid	5	100.0
	Excluded [a]	0	.0
	Total	5	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.513	10

Intraclass Correlation Coefficient

	Intraclass Correlation [a]	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.100[b]	-.037	.632	2.053	4.0	36	.108
Average Measures	.526[c]	-.550	.945	2.053	4.0	36	.108

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. Type A intraclass correlation coefficients using an absolute agreement definition.

b. The estimator is the same, whether the interaction effect is present or not.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

Criteria C (Language Suitability)

Case Processing Summary

		N	%
Cases	Valid	5	100.0
	Excluded [a]	0	.0
	Total	5	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.750	2

Intraclass Correlation Coefficient

	Intraclass Correlation [a]	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.600[b]	-.305	.948	4.000	4.0	4	.104
Average Measures	.750[c]	-.857	.973	4.000	4.0	4	.104

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. Type A intraclass correlation coefficients using an absolute agreement definition.

b. The estimator is the same, whether the interaction effect is present or not.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

TEST B OVERALL INTRACLASS CORRELATION COEFFICIENT

Case Processing Summary

		N	%
Cases	Valid	5	100.0
	Excluded [a]	0	.0
	Total	5	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.758	18

Intraclass Correlation Coefficient

	Intraclass Correlation [a]	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.151[b]	.023	.656	4.135	4.0	68	.005
Average Measures	.763[c]	.294	.972	4.135	4.0	68	.005

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. Type A intraclass correlation coefficients using an absolute agreement definition.

b. The estimator is the same, whether the interaction effect is present or not.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

A.3. Test C

Criteria A (Content)

Case Processing Summary

		N	%
Cases	Valid	5	100.0
	Excluded [a]	0	.0
	Total	5	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.533	6

Intraclass Correlation Coefficient

	Intraclass Correlation [a]	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.162[b]	-.067	.745	2.143	4.0	20	.113
Average Measures	.537[c]	-.582	.946	2.143	4.0	20	.113

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. Type A intraclass correlation coefficients using an absolute agreement definition.

b. The estimator is the same, whether the interaction effect is present or not.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

Criteria B (Format and Layout)

Case Processing Summary

		N	%
Cases	Valid	5	100.0
	Excluded [a]	0	.0
	Total	5	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.811	10

Intraclass Correlation Coefficient

	Intraclass Correlation [a]	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.289[b]	.063	.804	5.292	4.0	36	.002
Average Measures	.803[c]	.403	.976	5.292	4.0	36	.002

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. Type A intraclass correlation coefficients using an absolute agreement definition.

b. The estimator is the same, whether the interaction effect is present or not.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

Criteria C (Language Suitability)

Case Processing Summary

		N	%
Cases	Valid	5	100.0
	Excluded[a]	0	.0
	Total	5	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.750	3

Intraclass Correlation Coefficient

	Intraclass Correlation[a]	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.529[b]	-.058	.928	4.000	4.0	8	.045
Average Measures	.771[c]	-.183	.975	4.000	4.0	8	.045

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. Type A intraclass correlation coefficients using an absolute agreement definition.

b. The estimator is the same, whether the interaction effect is present or not.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

TEST C OVERALL INTRACLASS CORRELATION COEFFICIENT

Case Processing Summary

		N	%
Cases	Valid	5	100.0
	Excluded[a]	0	.0
	Total	5	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.899	19

Intraclass Correlation Coefficient

	Intraclass Correlation[a]	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.305[b]	.105	.801	9.928	4.0	72	.000
Average Measures	.893[c]	.690	.987	9.928	4.0	72	.000

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. Type A intraclass correlation coefficients using an absolute agreement definition.

b. The estimator is the same, whether the interaction effect is present or not.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

B. One- Way Analysis of Variance (General Average)

B.1. High- Achieving Group

Descriptives

HAGrades

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	20	91.50	1.701	.380	90.70	92.30	90	96
2	20	91.65	1.843	.412	90.79	92.51	90	96
3	20	91.75	2.023	.452	90.80	92.70	90	98
Total	60	91.63	1.832	.236	91.16	92.11	90	98

ANOVA

HAGrades

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.633	2	.317	.091	.913
Within Groups	197.300	57	3.461		
Total	197.933	59			

B.2. Low- Achieving Group

Descriptives

LAGrades

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	20	77.25	1.164	.260	76.71	77.79	75	79
2	20	77.40	1.142	.255	76.87	77.93	76	79
3	20	77.85	1.137	.254	77.32	78.38	76	79
Total	60	77.50	1.157	.149	77.20	77.80	75	79

ANOVA

LAGrades

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3.900	2	1.950	1.480	.236
Within Groups	75.100	57	1.318		
Total	79.000	59			

C. Two- Way Analysis of Variance

Between-Subjects Factors

		N
AbilityLevel	HA	60
	LA	60
TestFormat	A	40
	B	40
	C	40

Descriptive Statistics

Dependent Variable: Score

AbilityLevel	TestFormat	Mean	Std. Deviation	N
HA	A	24.1500	5.71494	20
	B	27.3000	4.05359	20
	C	26.6500	4.28308	20
	Total	26.0333	4.85717	60
LA	A	10.8000	2.76444	20
	B	15.4500	3.74833	20
	C	14.6000	2.66359	20
	Total	13.6167	3.66426	60
Total	A	17.4750	8.08286	40
	B	21.3750	7.13132	40
	C	20.6250	7.04450	40
	Total	19.8250	7.56447	120

Tests of Between-Subjects Effects

Dependent Variable: Score

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power[a]
Corrected Model	4981.075[b]	5	996.215	62.119	.000	310.593	1.000
Intercept	47163.675	1	47163.675	2940.877	.000	2940.877	1.000
AbilityLevel	4625.208	1	4625.208	288.404	.000	288.404	1.000
TestFormat	342.600	2	171.300	10.681	.000	21.363	.988
AbilityLevel * TestFormat	13.267	2	6.633	.414	.662	.827	.116
Error	1828.250	114	16.037				
Total	53973.000	120					
Corrected Total	6809.325	119					

a. Computed using alpha = .05

b. R Squared = .732 (Adjusted R Squared = .720)

Post Hoc Analysis

Warnings

Post hoc tests are not performed for AbilityLevel because there are fewer than three groups.

2. AbilityLevel

Dependent Variable: Score

AbilityLevel	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
HA	26.033	.517	25.009	27.058
LA	13.617	.517	12.592	14.641

Multiple Comparisons

Dependent Variable: Score

Tukey HSD

(I) TestFormat	(J) TestFormat	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
A	B	-3.9000*	.89547	.000	-6.0265	-1.7735
	C	-3.1500*	.89547	.002	-5.2765	-1.0235
B	A	3.9000*	.89547	.000	1.7735	6.0265
	C	.7500	.89547	.681	-1.3765	2.8765
C	A	3.1500*	.89547	.002	1.0235	5.2765
	B	-.7500	.89547	.681	-2.8765	1.3765

Based on observed means.

*. The mean difference is significant at the .05 level.

APPENDIX R

CERTIFICATION OF LANGUAGE EDITING

Philippine Normal University

The National Center for Teacher Education

College of Graduate Studies and Teacher Education Research

(CGSTER)

24 April 2018

C E R T I F I C A T I O N

This certifies that I have edited the thesis titled "**Reducing Cognitive Load of Mathematics Test Items: Effects on the Performance of High And Low Ability Students"** by **Steve B. Anapi,** a candidate for the degree Master of Arts in Education with Specialization in Mathematics Education**.**

MERRY RUTH M. GUTIERREZ
CGSTER Faculty

APPENDIX S

Authenticity Result

Philippine Normal University
The National Center for Teacher Education
PUBLICATION OFFICE
Taft Avenue, Manila 1000, Philippines

PUBLICATION OFFICE

AUTHENTICITY RESULT

Date of Release:	April 19, 2018
Result Code:	2018-0414
Submission ID:	949553733
Title of Output:	Effect of Reduced Cognitive Load Mathematics Test-Items on the Performance of High School Students
Author (s):	Steve B. Anapi

Authenticity Report:

Similarity Report: 10%

- *Interpretation*: A similarity of 10% means that 10% of the entire paper is similar to sources in Turnitin's (authenticity software) online repository. The colored bibliographical references denote the sources of similarity to previously published work which maybe from internet sources (8%)publications (5%), and student papers (3%) for the remaining similarity percentage.

- *Highlighted text in the Manuscript:*

  - Red marks and highlights are set of texts directly or exactly copied from online sources.
  - Other color marks and highlights are set of text exactly copied from other sources (student papers and publications).

Marie Paz E. Morales, Ph.D.
Director

CURRUCULUM VITAE

Name : **STEVE B. ANAPI**

Birth date : **SEPTEMBER 13, 1990**

Civil Status : **SINGLE**

Address : **151 B8 L18 HUGO VIDAL ST. BPS, PASONG BUAYA 2, IMUS CITY, CAVITE**

Correspondence : **steve13anapi@gmail.com 09167992045**

Educational Background:

Level	School	Degree/Course	Year Graduated
Elementary	Pasong Buaya Elem. School	Primary Education	2003
Secondary	Gen. Emilio Aguinaldo Nat'l High School	High School	2007
Tertiary	Philippine Normal University- Manila	Bachelor of Secondary Education Major in Mathematics	2011
Graduate	Philippine Normal University- Manila	Master of Arts in Education with Specialization in Mathematics Education	2018

Eligibility: Licensure Examination for Teachers (Rating: 83.20%)

Work Experiences:

	Department/ Company	Position
Present	DepEd Imus City	Teacher I
Previous	La Trinidad Academy	Teacher