2019-12-01

A Validity Study of the Cognitively Guided Instruction Teacher Knowledge Assessment

Debra Smith Fuentes
Brigham Young University

A Validity Study of the Cognitively Guided Instruction

Teacher Knowledge Assessment

Debra Smith Fuentes

A dissertation submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Richard R Sudweeks, Chair
Steven R. Williams
Damon L. Bahr
Joseph A. Olsen

Educational Inquiry, Measurement, and Evaluation

Brigham Young University

ABSTRACT

A Validity Study of the Cognitively Guided Instruction
Teacher Knowledge Assessment

Debra Smith Fuentes
Educational Inquiry, Measurement, and Evaluation, BYU
Doctor of Philosophy

This study reports the development of an instrument intended to measure mathematics teachers' knowledge of Cognitively Guided Instruction (CGI). CGI is a mathematics professional development framework based on how students think about and solve problems and how that knowledge guides instruction for developing mathematical understanding.

The purpose of this study was to (a) analyze and revise the original CGI Teacher Knowledge Assessment (CGI TKA), (b) administer the revised CGI TKA, and (c) analyze the results from the revised CGI TKA. As part of the revision of the original CGI TKA, distractor analysis identified distractors that could be improved. Experts in CGI content were interviewed to identify ways in which the content of the CGI TKA could be improved, and some new items were created based on their feedback. Formatting changes were also made to administer the assessment electronically.

After the original CGI TKA was revised, the revised CGI TKA was administered to teachers who had been trained in CGI. Two hundred thirteen examinees completed the revised CGI TKA and the results were analyzed. Exploratory and confirmatory factor analyses showed 21 of the items loaded adequately onto one factor, considered to be overall knowledge of CGI. The Rasch model was used to estimate item difficulty and person abilities as well as to compare models using dichotomous and partial credit scoring. Advantages and disadvantages of using partial credit scoring as compared to dichotomous scoring are discussed. Except under special circumstances, the dichotomous scoring produced better fitting models and more reliable scores than the partial credit scoring. The reliability of the scores was estimated using Raykov's rho coefficient. Overall, the revised CGI TKA appears to validly and reliably measure teachers' CGI knowledge.

Keywords: cognitively guided instruction, mathematics education, teacher education, professional development, partial credit scoring, pedagogical content knowledge, teacher knowledge assessment

ACKNOWLEDGMENTS

This project would not have been possible without the tremendous support of my husband, Alejandro, and children, Rebeca, Jimena, Owen, and Sofia Fuentes. I would like to especially thank my parents, Marvin and Stephanie Smith, for producing the original Cognitively Guided Instruction Teacher Knowledge Assessment, and for their ongoing support during the revision process. I am also grateful for their influence and examples as exceptional mathematics educators. Special thanks to Richard Sudweeks, committee chair, for the many hours of counsel, patience, and expertise while analyzing, revising, and rewriting the many drafts created of this study and related documents. I appreciate the wisdom and insight of all my committee members, as well as their patience with me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

CHAPTER 1

Introduction

Cognitively Guided Instruction (CGI) in elementary school mathematics is a framework for professional development of mathematics instruction based on understanding how to develop students' mathematical thinking by learning how students think about and solve problems (Carpenter, Fennema, Franke, Levi, & Empson, 2000). The research on children's mathematical thinking upon which CGI is based shows that children are able to solve problems without direct instruction by drawing upon their informal knowledge of everyday situations (Carpenter, Fennema, Peterson, & Carey, 1988; Carpenter et al., 2000; Carpenter, Fennema, Franke, Levi, & Empson, 2015; Fennema et al., 1996; Peterson, Fennema, Carpenter, & Loef, 1989). At the core of this approach is the practice of asking questions and listening to children talk about their thinking, solution strategies, and understanding of mathematical ideas, and then using that problem-solving discourse as the foundation for mathematical learning in the classroom (Carpenter, Fennema, Peterson, Chiang, & Loef, 1989). In brief, CGI is a problem-solving approach to teaching mathematics for understanding, as emphasized in professional standards (National Council of Teachers of Mathematics [NCTM], 2000; National Governors' Association [NGA], 2010).

Many teachers across the United States have participated in professional development focused on CGI. In an attempt to determine the effects of professional development instruction on teachers, researchers (Carpenter et al., 1988; Carpenter et al., 1989; Enochs, Smith, & Huinker, 2000; Peterson et al., 1989; Philipp, 2007; Pianta, LaParo, & Hamre, 2008; Schoen Bray, Wolfe, Tazaz, & Nielsen, 2017) have used a variety of assessment measures, some of which are explained more fully in the literature review section of this document. For example, the Mathematics Teaching Efficacy Beliefs Instrument (MTEBI; Enochs, Smith, & Huinker, 2000) and Integrating Mathematics and Pedagogy (IMAP; Philipp, 2007) are intended to measure beliefs and efficacy of teachers' ability to teach elementary mathematics. The electronic Teacher Knowledge Assessment System (TKAS), which is part of the Learning Mathematics for Teaching (LMT) project, (Hill, Schilling, & Ball, 2004) has been used to assess teachers' knowledge of elementary mathematics content, in addition to the Knowledge for Teaching Early Elementary Mathematics (K-TEEM) assessment. In addition, researchers typically observe classroom instruction to determine changes based on CGI professional development using a tool such as the Classroom Assessment Scoring System (CLASS; Pianta, LaParo, & Hamre, 2008) observation tool or the Standards Based Learning Environment Classroom Observation Tool (SBLE COT; Tarr et al. 2008). In many cases, researchers correlate these assessments with student assessments, both standardized and researcher created (Carpenter et al., 1989) to see how student outcomes relate to teacher knowledge, beliefs, etc. to determine effects of these teacher attributes on student outcomes.

One might ask, why not directly test teachers' understanding of the principles and content of CGI? Such an assessment could inform the mathematics education field of the impact of CGI on mathematics instruction. In fact, teachers' CGI knowledge may correlate with student outcomes regardless of changes in practice or beliefs. It may be that increased knowledge of CGI principles affect students in ways that are not measured by beliefs scales or observable classroom practices.

For this reason, Smith and Smith (Unpublished, 2014) developed a test called the CGI Teacher Knowledge Assessment (CGI TKA) to measure the knowledge teachers gain through professional development around CGI (See Appendix A). Smith and Smith asserted that this assessment could be used to inform school and district administrators of teachers' potential capacity to effectively teach using the CGI framework in ways that none of the abovementioned assessments do. More distinctions between the CGI TKA and these other assessment instruments and procedures are described in the literature review section of this paper. Additionally, data from the CGI TKA could also help improve the professional development instruction offered to CGI teachers. The largest influence the CGI TKA may offer is that of providing evidence of the impact of CGI on mathematics instruction. However, questions remain regarding the extent to which scores obtained from the CGI TKA provide a reliable and valid representation of teachers' CGI knowledge. Since there are still many questions surrounding the potential uses of the CGI TKA, it is not included in its entirety in this document. The revised assessment produced by this study will not be made available until it is determined what some of those uses might be and if it is beneficial to allow public access to such a test.Statement of the Problem

Concisely write a logical 1-2 paragraph statement of the problem to be solved by your research. The problem should be demonstrated to be significant enough to warrant study (e.g., affecting a large number of individuals statewide, nationally, or internationally; limited or inconclusive research has been conducted on this topic with this population; research that has been conducted is outdated or not applicable; a need for replication of another research study; or a need for expanding another research study). Make sure you describe why it would be a problem if you didn't conduct this research to find answers to the presenting problem.

Purpose

The purpose of this study was threefold: first, to analyze the psychometric properties of scores obtained from the original version of the CGI TKA, including how the items functioned and how they were interpreted by CGI experts, teachers, and other mathematics educators; second, to produce a revised CGI TKA based on the identified weaknesses of the original version; and third, to estimate the reliability and validity estimates of teachers' knowledge obtained from the new revised instrument.

Research Questions

These purposes were accomplished by investigating the following research questions:

1. What revisions or modifications should be made to the original CGI TKA to make it a more valid and reliable measure of teachers' pedagogical content knowledge of CGI?
    a. Based on classical item analysis and distractor analysis, which items in the CGI TKA need to be revised or replaced?
    b. What evidence of content validity related to CGI is apparent based on subject-matter experts' judgment of the items in the original CGI TKA?
2. How validly and reliably do the scores obtained from the revised CGI TKA measure teachers' knowledge of CGI?
    a. To what extent does exploratory and confirmatory factor analysis provide evidence that supports the construct validity of the revised CGI TKA?
    b. To what extent does item response theory analysis provide evidence that supports proper functionality of the items and their distractors of the revised CGI TKA?
    c. What are the advantages or disadvantages of using partial credit scoring compared to dichotomous scoring of the items of the revised CGI TKA?
    d. How reliable are the scores obtained from administering the revised CGI TKA?

CHAPTER 2

Review of Literature

Measuring Teacher Knowledge

Most people would agree that teachers should understand the subject matter they teach. However, this knowledge is difficult to define and measure. Furthermore, we often assume that individuals who understand a subject will automatically know how to teach that subject to others who do not understand it. Nevertheless, this assumption is questionable. Common wisdom observes college and university faculties often include individuals who are experts in a subject such as mathematics or some other specialized subject-matter area but are unable to successfully plan and present effective instruction in that subject.

Shulman (1986) distinguished three different domains of knowledge that teachers need: (a) content knowledge, (b) pedagogical content knowledge, and (c) curricular knowledge. *Content knowledge* he defined as the amount and organization of knowledge in the teacher's mind, including knowledge of the subject and its organizing structures (e.g., Grossman, Wilson, & Shulman, 1989). This knowledge can be subject specific with regards to teaching content knowledge, but Shulman (1986) argued that teaching requires more than knowing a subject's facts and concepts. It requires understanding how students think about content and how that impacts instruction, known as *pedagogical content knowledge.*

Going beyond knowledge of subject matter, Shulman (1986) points out that teachers must possess this pedagogical content knowledge for effective teaching; namely the specific aspects of content that are most applicable to its teachability. He described this knowledge as the understanding of what makes content teachable. Pedagogical content knowledge as Shulman (1986) describes it includes,

> The most regularly taught topics in one's subject area, the most useful forms of representation of those ideas, the most powerful analogies, illustrations, examples, explanations, and demonstrations . . . [and] an understanding of what makes the learning of specific topics easy or difficult: the conceptions and preconceptions that students of different ages and backgrounds bring with them to the learning of those most frequently taught lessons (p. 9).

Shulman sums up the characteristics of pedagogical content knowledge by describing it as "the ways of representing and formulating the subject that make it comprehensible to others" (Shulman, 1986, p. 9).

Ball, Thames, and Phelps (2008) argue that pedagogical content knowledge is "essential to effective teaching" (p. 390), and "that it bridges content knowledge and the practice of teaching" (p. 389). As these scholars and others have argued, pedagogical content knowledge is of utmost importance in effective teacher education and teacher development. Ball, Thames, and Phelps (2008) further described this kind of pedagogical content knowledge as how educators use knowledge of how students learn combined with subject matter content knowledge to guide instruction. The pedagogical content knowledge associated with Cognitively Guided Instruction is the focus on student mathematical thinking as well as how instruction can support and further that thinking. Thus, it is the understanding of students' mathematical thinking that becomes the bridge between mathematics content and effective teaching.

However, the kind of specialized pedagogical content knowledge Shulman and Ball and their colleagues claim to be essential to teaching is not only difficult to define, but difficult to measure (Shulman, 1986; Ball et al., 2008). Ball and her colleagues have spent a great deal of their professional careers attempting to effectively measure pedagogical content knowledge of mathematics teachers (Ball & Forzani, 2010, 2011; Hill, Rowan, & Ball, 2005). As a result, projects like the Learning Mathematics for Teaching (LMT) Project and the Teacher Knowledge Assessment System (TKAS) assess general elementary mathematics content knowledge because measuring pedagogical content knowledge is illusive. Thus, we see that when confronted with the difficulty of measuring pedagogical content knowledge, researchers opt for measuring content knowledge or curricular knowledge.

Why Knowledge of Cognitively Guided Instruction?

Cognitively Guided Instruction has been heralded as one of the most significant contribution to mathematics education reform in the past thirty years (Empson & Jacobs, 2008). Since its inception, researchers of CGI have attempted to measure its effectiveness in a variety of ways. From their early work, the CGI researchers correlated teachers' abilities to anticipate and predict their students' strategies with children's problem-solving abilities (Carpenter et al., 1988). In fact, in 1988, Carpenter, Fennema, Peterson, and Carey measured teachers' knowledge of CGI by having teachers write story problems, compare problem difficulty across similar story problems, identify word problems according to the CGI framework, and demonstrate knowledge of solution strategies including direct modeling, counting, and number facts.

In a study of teachers who had not participated in the CGI professional development program, the CGI researchers found that teachers had a great deal of intuitive knowledge about children's mathematical thinking; however, because that knowledge was fragmented, it generally did not play an important role in most teachers' decision-making (Carpenter et al., 1988). This study indicated that teachers possessed informal knowledge of children's thinking that could be built upon during a CGI professional-development program. Teachers could identify differences between problem types, and they had some idea of many of the modeling and counting strategies that children often use and as described in the CGI literature (Carpenter et al., 1988). However, most teachers' understanding of problems and strategies was not well connected, and most did not appreciate the critical role that modeling and counting strategies play in children's thinking or understand that more than a few students are capable of using more sophisticated strategies. These early studies also showed that teachers' knowledge of their students' thinking related to student achievement. Students of teachers who knew more about their students' thinking had higher levels of achievement in problem solving than students of teachers who had less knowledge of their students' thinking. In a related study (Peterson et al., 1989), these same researchers found that classes of teachers whose beliefs about how students learn that were more consistent with principles of CGI tended to have higher levels of student achievement than classes of teachers whose beliefs were less consistent with principles of CGI.

In order to reach their conclusions, Peterson and colleagues (Peterson et al., 1989) utilized a teaching and learning beliefs survey and a beliefs interview to determine how teachers had changed their beliefs based on the CGI professional development. Teachers were also given assessments that attempted to assess relative problem difficulty, general knowledge of strategies, and knowledge of individual students' strategies. Students were given tests of number facts or computation, and problem solving (Peterson et al., 1989). To determine successfulness of the CGI professional development, they used this data from the teachers and compared it to student achievement data, as is so often done. However, there are so many factors that affect student achievement other thank professional development of CGI that it is difficult to make any statements of causality connected to the professional development of CGI.

Since that time, most of the research around CGI has not measured knowledge of CGI, but instead has focused on beliefs about teaching and learning and pedagogical practice without incorporating teachers' knowledge in the studies. Researchers have primarily measured teachers' beliefs about how students learn and observed classroom instruction, using student achievement to measure effectiveness (Peterson et al., 1989).

Beyond the scope of CGI, researchers measure teachers' knowledge of elementary mathematics with the *Teacher Knowledge Assessment System* (Hill, Schilling, & Ball, 2004). Other researchers have attempted to measure teachers' beliefs, and how those beliefs change over time as they participate in professional development, as was the case with the *Integrating Mathematics and Pedagogy* assessment (Philip, 2007). Still other researchers have observed classroom instruction to measure impact of professional development and incorporation of theory into practice, as with the *Classroom Assessment Scoring System* (Pianta et al., 2008).

Each of these studies assumes teachers possess the knowledge necessary to provide quality CGI instruction. However, Carpenter et al. (1988) discovered that many teachers overestimated the difficulty of some problem types. They claimed that the. teachers did not understand the structure of the problem and how it related to how children would likely solve the problem. Especially when it came to counting strategies, teachers had trouble predicting students' solution strategies (Carpenter et al., 1988). If the premise of CGI is that teachers will anticipate student thinking and orchestrate effective discourse around student strategies to develop conceptual mathematics understanding, findings such as these demonstrate a lack of preparation to do so. It is necessary to measure teachers' knowledge specific to CGI in ways that help inform professional development. This study suggests the need to measure the cognitive portion of Cognitively Guided Instruction as a way to measure pedagogical content knowledge.

Assessments Used to Measure Teacher Attributes

Several major instruments and procedures are currently used to determine teachers' content knowledge, beliefs about student thinking, efficacy of teaching practices, and classroom practice. This section will briefly describe each of these tests and some of the current research that employs each of them in the context of CGI. For a summary of the intended attributes each assessment attempts to measure, as well as the primary authors of each test and the year it was first utilized, see Table 1. Each of the attributes and related assessments is described more fully below, along with descriptions of some research that has attempted to measure such attributes.

Content knowledge. Teachers require deep and broad knowledge of mathematics to be effective in their teaching (Hill, 2010). Likewise, deep understanding of standards and effective practices are required to create Standards-Based Learning Environments (SBLEs; Tarr et al., 2008) that promote classroom discourse and foster conceptual understandings of mathematics. Multiple efforts have been attempted to define the exact mathematical knowledge needed for teaching, and several researchers (Ball & Forzani, 2010; Ball et al., 2008; Hill, 2010) have emphasized a specialized content knowledge (SCK) characterized as "mathematical knowledge needed to perform the recurrent tasks of teaching mathematics to students" (Ball et al., 2008, p. 399). The SCK for teaching mathematics extends Shulman's (1986) conceptualizations of subject matter knowledge (SMK) and pedagogical content knowledge (PCK) and includes teachers' abilities to: (a) analyze and interpret students' mathematical thinking and ideas, (b) use multiple representations of mathematical concepts, and (c) define terms in mathematically correct and accessible ways (Hill, 2010; Thames & Ball, 2010). CGI content knowledge would very likely be considered specialized content knowledge.

To measure teacher's mathematics content knowledge, the *Teacher Knowledge Assessment System* was created (Hill et al., 2004). Although this assessment does an excellent job of measuring teachers' elementary mathematics content knowledge, it has a few limitations. First, it focuses on assessing content knowledge rather than pedagogical knowledge. Second, scores obtained from this test are very reliable and are provided complete with Item Response Theory (IRT) analysis. However, the scores are only reported at a group level, such as a school or district, not at an individual teacher level. Since individual teacher data is not reported, any correlation between a teacher's knowledge and student achievement is impossible. Analysis can only be done for groups of examinees. The present study of the CGI TKA attempts to provide reliable results at the individual teacher level.

Table 1

Assessments Used to Measure Teacher Attributes

Assessment Name	Intended Characteristic of Measure	Primary Author(s)
Teacher Knowledge Assessment System/Learning Mathematics for Teaching (TKAS/LMT)	Elementary Mathematics Content Knowledge	Hill et al., 2004
Knowledge for Teaching Early Elementary Mathematics (K-TEEM)	Early Elementary Mathematics Content Knowledge	Schoen et al., 2017
Mathematics Teaching Efficacy Beliefs Instrument (MTEBI)	Teachers' Beliefs and Efficacy of Mathematics Instruction	Enochs et al., 2000
Integrating Mathematics and Pedagogy (IMAP)	Intensity of Mathematics Instruction Beliefs	Philipp, 2007
Standards Based Learning Environment Classroom Observation Tool (SBLE COT)	Classroom Instructional Practices	Tarr et al., 2008
Classroom Assessment Scoring System (CLASS)	Classroom Instruction Practices	Pianta et al., 2008

Like the LMT and TKAS, the *Knowledge for Teaching Early Elementary Mathematics* (K-TEEM; Schoen et al., 2017) instrument measures teachers' mathematical knowledge for teaching early elementary mathematics. Schoen and colleagues (2017) intend to use the K-TEEM to relate research among teachers' mathematical content knowledge for teaching, professional development, and student learning. However, this instrument, like others measuring content knowledge, measures knowledge of the mathematical matter and not how students understand or learn that content. A measure of specialized content knowledge is needed.

Beliefs and efficacy. Another salient factor influencing teacher effectiveness is teachers' beliefs about student thinking and efficacy towards individual teacher's practice. Over time, research has established a robust relationship between teachers' beliefs about student thinking by showing that those beliefs influence teacher thinking and behaviors, including instructional decision-making and use of curriculum materials (Buehl & Fives, 2009; Clark & Peterson, 1986; Philipp, 2007; Raymond, 1997; Romberg & Carpenter, 1986; Thompson, 1992; Wilson & Cooney, 2002). Pedagogical beliefs are considered the cognitive set of psychological understandings, premises, or propositions through which interpretations are made of the surrounding world (Philipp, 2007). Teachers have deep-rooted mathematical beliefs formed during their seminal years as students in K-12 classrooms (Lortie, 1975); they tend to resist changing these beliefs during teacher education (Bird, Anderson, Sullivan, & Swidler, 1992; Handal & Herrington, 2003; Philipp, 2007). It has been argued, "The lack of attention to substantive mathematics preparation, coupled with the questionable quality of appropriateness of the mathematics courses taken by . . . elementary teachers, provides little chance of changing teachers' beliefs" (Reys & Fennell, 2003, p. 278). However, teachers' beliefs about student thinking are influential in how and what they learn and should be targets for change during teacher education (Feiman-Nemser, 2001; Philipp, 2007; Richardson, 1996). Two belief constructs that have frequently been studied include pedagogical beliefs (i.e., beliefs about teaching and learning) and teaching efficacy beliefs (i.e., beliefs about capabilities to teach effectively and influence student learning).

Due to the importance of teachers' beliefs about student thinking to mathematics instruction and teacher change, several assessments have been developed to measure teachers' beliefs about student thinking and efficacy of their own teaching practices. The *Mathematics Teaching Efficacy Beliefs Instrument* (MTEBI) has been effective in several research studies (Enochs et al., 2000; Swars, 2005; Swars et al., 2007; Swars et al., 2009; Smith, Swars, Smith, Hart, & Haardörfer (2012) to demonstrate change in teachers' beliefs about student thinking and efficacy of teaching practices as a result of professional development in mathematics education. Likewise, the *Integrating Mathematics and Pedagogy* (IMAP) measures the intensity of teachers' beliefs about student thinking in mathematics education (Philipp, 2007).

Classroom teaching practices. The *Principles and Standards for School Mathematics* (PSSM; NCTM, 2000) and the *Common Core State Standards for Mathematics* (CCSS-M; NGA, 2010) recommend the intersection of mathematical content and process standards requiring a pedagogical approach different from the traditional direct instruction in computational skills still found in many U.S. classrooms. Research by Tarr et al. (2008) indicates that this change in pedagogy is more important for improving student achievement than the use of particular curriculum materials. Many of the suggested changes in teaching practices are grounded in social-constructivist methods of teaching, in which teachers engage students in authentic non-routine problem-solving tasks and discourse that are intended to develop students' understandings of concepts and mathematical practices in ways that foster their abilities to solve problems and to reason and communicate mathematically. Specifically, teachers are being asked to create standards-based learning environments where students are encouraged to explain their problem-solving strategies and reasoning and to make conjectures and other generalizations about mathematical ideas based on their specific problem-solving experiences and contextualized understanding. Student statements are used to build discussion or work toward a shared understanding for the class. Moreover, multiple perspectives are encouraged and valued, and enacted lessons foster the development of deep, well-connected conceptual understanding. According to the Tarr et al. (2008) study, the improved student achievement was linked to the extent of enactment of such a SBLE. This emphasis on the importance of improving pedagogy in mathematics education has continued in the exploration of a "common core for teaching practice" (Ball & Forzani, 2011, p. 19), including a set of high-leverage practices that underlie effective teaching. While it is important to measure classroom teaching practices, specifically within SBLEs, as mentioned in the *Teacher Knowledge* section, that practice is affected by several teacher characteristics. If one wants to determine teacher knowledge, it is better to attempt to assess that knowledge more directly (Hill, Rowan, & Ball, 2005).

As part of the Tarr et al. (2008) study, an observation tool was used to systematically observe classroom instruction—the *Standards Based Learning Environment Classroom Observation Tool* (SBLE COT). Similarly, researchers at the University of Virginia (Pianta et al., 2008) have developed an effective classroom observation tool for mathematics education—the *Classroom Assessment Scoring System* (CLASS). Both of these measures focus on instructional practices and are, therefore, time and labor intensive. While many studies are able to gather observational data, this is often one of the limitations of large-scale studies. An assessment of teacher knowledge may be more feasible for larger groups of teachers. Still, this study suggests the need to measure the cognitive part of Cognitively Guided Instruction, not just the Guided Instruction parts. In other words, this study suggests measuring teachers' understanding of student thinking that influence their practice without only directly observing their practice.

Each of the assessments described above and the attributes it is attempting to measure intends to measure variables related to CGI knowledge. However, none of them attempts to measure CGI knowledge directly, but rather the effects of CGI knowledge on teachers' mathematical content knowledge, beliefs about student thinking, efficacy of teachers' practice, and effectiveness of classroom practice. For this purpose of attempting to more directly measure knowledge specific to CGI, this study proposes to validate and refine a test of teacher CGI knowledge, the CGI TKA.

Rationale

The present study focuses on the specialized content knowledge necessary to teach CGI effectively, and how that knowledge can be reliably and validly assessed. Similar to the content knowledge Ball, Thames, and Phelps (2008) described, Carpenter and his colleagues (Carpenter et al., 1988, Peterson et al., 1989) have attempted to measure the specific knowledge needed by mathematics teachers regarding the CGI framework. This is a narrower type of specialized content knowledge needed than what is measured by the TKAS and other content assessments. However, Carpenter and his colleagues focused on other aspects of teaching in addition to knowledge, such as beliefs, efficacy, and practice. While all of these aspects are necessary for teaching, each is difficult to define and measure.

However, Smith and Smith (unpublished, 2014) created the CGI TKA in an attempt to measure teachers' CGI specialized content knowledge based on initial professional development in CGI. Thus far, the original form of this assessment appears to have preliminary psychometric promise (Myers, Swars, Smith, Smith, & Fuentes, in press, 2019). This study intends to further determine the need to revise the original assessment and then analyze this assessment's ability to reliably and validly measure teachers' CGI knowledge.

CHAPTER 3

Method

Participants

Preliminary sample participants using the original CGI TKA. Existing CGI TKA data gathered from 97 in-service teachers who received CGI training during in-school professional development or university settings in the Eastern United States is the basis for initial analysis of the scores. Participating teachers were from schools or university teacher preparation programs in the eastern United States. Approximately two-thirds were in-service teachers and one-third pre-service teachers. However, the authors of the assessment (Smith & Smith, unpublished, 2014) or their affiliates trained all the teachers assessed in CGI content and pedagogy. The teachers participated because their school leaders sought out CGI professional development, or they were part of a graduate elementary mathematics endorsement program where the assessment was given as part of the class. All the teachers and leaders in the professional development at the schools and continuing education classes were assessed shortly after completing the CGI professional development or coursework.

Revised CGI TKA sample participants. Smith and Smith, as well as the researcher of the present study, continue to train teachers in CGI. Those teachers and leaders in elementary and middle schools currently using CGI were asked to take the revised assessment. These teachers ranged in years of experience from 0 to 20+, taught preK-elementary or middle grade levels, and were located in all geographic regions of the United States of America.

Instruments

Original CGI TKA. The original CGI TKA was a 36-item selected-response test intended to assess six main constructs: (a) teachers' knowledge of CGI problem types, (b) teachers' knowledge of CGI student solution strategies, (c) teachers' understanding of the theoretical underpinnings of CGI as they pertain to pedagogy, (d) teachers' identification of evidence of base ten understanding, (e) the usefulness of student strategies in mathematics instruction, and (f) number sentences or equations that are associated with problems and strategies. The original test included 28 multiple choice and 8 true-false items. The table of specifications used to guide the development of items included in the revised CGI TKA is shown in Table 2, organized according to a revised version of Bloom's Taxonomy (Anderson & Krathwohl, 2001).

The cell entries in Table 2 list the number of each item based on the construct it was initially designed to measure, and the type of cognitive skill it is intended to assess. The distribution of the items of each construct across the modified taxonomy was analyzed as part of the revision process of the CGI TKA. For example, it may be appropriate to measure the usefulness of a mathematical solution strategy in a way that requires evaluation of that strategy, rather than simply identification or understanding. These types of considerations were made to support revisions of the original CGI TKA.

For the items intended to measure teachers' knowledge of CGI problem types, most of the item stems read, "Which of the following story problems is a good example of [problem type]." Then options a-e each included a different story problem, only one of which matches the correct problem type. The items written to measure teachers' knowledge of student solution strategies presented a scenario of a student solution and options a-e included names of possible solution strategies as presented by Carpenter and colleagues (Carpenter et al., 1999; Carpenter et al., 2015) and in the initial CGI professional development (Carpenter et al., 1989). Although understanding of the various problem types and their associated probable student solution strategies constitute a major portion of the initial CGI professional development, both knowledge of problem types and strategies seemed appropriate constructs to measure.

Table 2

Table of Specifications Based on the Original CGI TKA

Construct	Understand Concepts/Principles: Classify, Recognize	Apply: Interpret, Implement	Analyze: Differentiate, Organize, Compare, Examine	Evaluate: Appraise, Value, Select
CGI Problem Types (PT)	2, 3, 4, 5	6, 7		
CGI Solution Strategies (SS)		17, 18	8, 9, 10, 11, 12, 13, 14, 15, 16, 30	
CGI Principles (P)	1, 22, 23, 24, 25, 26, 27, 28, 29			
Evidence of Base Ten Understanding (BT)			16, 30	
Strategy Usefulness (SU)				31, 32, 33, 34, 35, 36
Equations or Number Sentences (NS)		6, 7, 19, 20, 21		

The original CGI TKA included eight true-false items regarding appropriate pedagogical behavior based on the theoretical foundation of CGI as taught in professional development. In a way, these items might be viewed as assessing teacher beliefs or practice, but here they are presented more in a way that attempts to determine teachers' understanding of the theoretical principles of CGI as presented in professional development. However, since there is only one distractor for these true or false items, the information provided by these items would likely be increased if these items were changed to a multiple-choice format with functioning distractors. This modification was made as part of revising the CGI TKA.

The final section of the original 36-item instrument presented base-ten strategies, appropriate number sentences, and useful base-ten concepts when solving CGI problems. These items incorporated elements of base-ten understanding but addressed different components of solution strategies and related equations within solutions. Base-ten evidence of understanding is often a difficult component of the initial CGI professional development for teachers to recognize in student solutions as well as conceptualizing how to support students in their base-ten development. It is also a significant portion of the two-day initial training, comprising nearly one fourth of the training. For these reasons, it makes sense that this construct was embedded in several items in the original CGI TKA, even though it was not heavily assessed directly. The manner in which it was assessed in the original CGI TKA focused primarily on student strategies as an embedded context for teachers to consider, attempting to mimic the classroom situation in which they will be required to recognize such evidence of base-ten understanding on the part of the students.

Revised CGI TKA. After considering the content, question format, and distractors of the original CGI TKA, a revised CGI TKA was created which consisted of 30 items. Much of the content was similar to the original assessment, but poorly functioning distractors were modified or replaced. Additional items were included that intended to assess areas considered by experts to have been omitted from the original assessment. Based on IRT analysis of the original assessment, the true-false items were removed and replaced with multiple-choice items intended to measure similar content. The table of specifications was modified based on the revisions of the CGI TKA after considering the results from the distractor analysis and expert judgements and is included in the results section of this document.

Procedure

The original CGI TKA was administered to examinees in a paper and pencil format and responses were entered into an Excel spreadsheet. They were then keyed for correct and incorrect answers, producing dichotomous data. The original paper version of the original assessment questions and answer key are included in Appendix A. The revised CGI TKA was administered via Qualtrics, an online assessment platform, which facilitated scoring and data analysis. Approval was obtained from the Brigham Young University Internal Review Board and the consent form and approval notice are included in Appendix B. The Qualtrics link was sent to school leaders who oversaw administration of the assessment, sometimes without a proctor. Two hundred thirteen participants completed the revised CGI TKA.

Analysis

Research Question 1. Research Question 1 was investigated in two phases. First, the researcher analyzed the original table of specifications for opportunities to improve the way the constructs were assessed. For example, the construct called CGI Principles was only measured at an understanding level according to the adapted Bloom's taxonomy. Since there is no opportunity for performance assessment in a multiple-choice format, the most in-depth form of measuring according to Bloom's taxonomy was deleted. Likewise, in an attempt to measure understanding beyond simple recall, the simplest form of measuring was also deleted, which left an adapted Bloom's taxonomy. Items were modified and created to include other thinking skills that would more broadly measure teachers' CGI knowledge. Likewise, only two items independently measured evidence of base ten understanding. It was determined that, rather than create new items to measure that construct, which seemed so related to student strategies that involve base ten understanding, this construct would be subsumed by student solution strategies.

In rewriting items or creating new items according to the table of specifications, the researcher conducted classical distractor analysis considering item difficulty, adjusted item-to-total correlations, and poorly functioning distractors. Items with poorly functioning distractors, meaning those that were not selected by at least 5% of the population, either had the stem modified to make the question clearer, or the nonfunctioning distractors were revised to make them more enticing to the examinees.

Second, by systematically collecting expert judgments from prominent CGI researchers, the researcher was able to analyze the extent to which the original CGI TKA assessed necessary and pertinent components of the CGI framework. Experts included original publishers of the CGI framework, as well as other local university CGI experts, and provided insights as to the ability of the CGI TKA to measure teachers' CGI knowledge. Expert interviews also offered feedback for revision of topics that ought to have been included in the CGI TKA that were possibly not included. See Appendix C for the questions used to gather systematic feedback about the intended constructs of the CGI TKA. Additional items were created in an attempt to satisfy expert suggestion for missing constructs and items within constructs.

Research Question 1 suggested the researcher would compile all of the analyses gathered to determine what revisions or extensions to the CGI TKA might improve the assessment. Based on an improved table of specifications, classical distractor analysis, and expert feedback, the researcher made informed revisions to the original version of the CGI TKA. The revised version of the assessment was formatted in Qualtrics and subsequently administered to a large group of teachers to conduct further analysis to answer Research Question 2.

Initial IRT analyses were performed using the WINSTEPS software to apply a Rasch model to the data. The point biserial correlation coefficient and the 2-PL discrimination parameters that are estimated post hoc so as not to distort the difficulty estimates were requested (Wright & Linacre, 1998). In IRTPRO (Cai, Thissen, & Du Toit, 2005), 2-PL analysis was used to estimate item difficulties and discrimination parameters. Based on the results, problematic items were deleted from the model to consider improvement. These suspect items were considered for deletion from the test or revision to improve items and response options during the test revision process. These deletions of items are discussed more in the results section of this document.

Research Question 2. Four different forms of analyses were conducted to investigate Research Question 2. First, exploratory and confirmatory factor analysis (CFA) were used to answer Research Question 2a. Since the items are each scored dichotomously, the response distribution for each item was not normally distributed. Since the maximum likelihood estimation procedure in Mplus software assumes that the input data are normally distributed, the "categorical" option in Mplus was used to override this default and invoke the Weighted Least Squares Mean and Variance (WLSMV) estimator instead. The use of the categorical option uses tetrachoric correlations instead of Pearson Product Moment correlations in the input correlation matrix for the exploratory factor analysis (EFA).

EFA was conducted first in order to obtain an estimate of how many factors should be retained and to identify any items that crossload on more than one factor or any items that do not clearly load on any factor. The decision about how many factors to retain was based on (a) the results of a parallel analysis, (b) the relative size of the eigenvalues, and (c) the interpretability of the resulting factors.

The initial CFA model was informed by both the intended six-factor model proposed by Smith and Smith (unpublished, 2014) and the EFA results. In addition, rival factor models, such as two, four, or six factors, were tested to determine existence of the intended constructs as compared to a single construct of CGI knowledge.

Second, to consider Research Question 2c, IRT was utilized to examine the functionality of each item, the information it provided, and to obtain person ability estimates. Since the sample size was smaller than originally anticipated, with 213 examinees, IRTPRO 2-PL estimates did not fully converge. Therefore, it was determined that Rasch scaling in WINSTEPS would be a better method of estimation, because of the small sample size.

Third, by definition, multiple-choice test items include two components: (a) a stem which presents a question to be answered or a problem to be solved, and (b) a set of two or more plausible answers to the question or problem presented. The examinee's task is to identify which of the alternative answer choices is the best or most correct answer to the question posed in the stem. Since the purpose of the test is to distinguish between knowledgeable and less knowledgeable examinees, the incorrect answer options are intentionally written so as to be plausible to examinees who either lack the necessary knowledge to answer the item correctly or who possess partial knowledge or a misconception.

The most commonly used method of scoring examinees' responses to multiple-choice items is to employ an answer key in which the correct answer to each item is coded "1" and all other options are coded "0." When this dichotomous scoring procedure is used, all incorrect options associated with an item are weighted as zero, even though some of them may represent a state of partial knowledge whereas other options are clearly incorrect. Some scholars (Andrich & Styles, 2011; Sideridis, Tsaousis, & Al Harbi, 2016) have argued that distractors which represent partial knowledge contain useful information and that examinees who choose such options should be awarded partial credit. Hence, they have advocated replacing dichotomous scoring with partial credit scoring where the best answer to an item is scored as "2," a partially correct option is scored as "1," and a clearly incorrect answer is scored as "0." The partial credit model (PCM) developed by Masters (1982) provides a way to implement this trichotomous scoring procedure using Rasch scaling.

The present study included an attempt to identify items with distractors representing partial knowledge that might contain useful information regarding teachers' CGI knowledge. Masters' PCM was used as an alternative way of scoring and analyzing such items. The results were then compared with the scores obtained using traditional dichotomous scoring.

Fourth, Raykov's rho reliability coefficient was estimated, which attends to question Research Question 2b. Rho was used because there were uncorrelated errors discovered while conducting factor analyses. Therefore, Cronbach's alpha was not an appropriate estimate of reliability.

CHAPTER 4

Results

Research Question 1 addresses the need for test revision, as suggested by the preliminary data analysis. This revision was informed by (a) expert feedback, (b) classical item analysis statistics, (c) distractor analysis, and (d) IRT. In addition, the revised test was formatted for delivery via a digital platform.

Research Question 1: Preliminary CGI TKA Data Analysis

To determine if the original CGI TKA was an assessment worth revising, preliminary analysis was conducted based on previously collected data. Reliability analysis in SPSS estimated an overall Cronbach's alpha reliability coefficient of .63, suggesting ample room for improvement. Other analyses also provided data to consider specific areas for revision, such as poorly functioning distractors, item difficulty, discrimination, and point biserial parameters for items that might need revision or removal from the test. After determining the need for revision and reformatting, the test was revised based on the table of specifications, expert judgment, distractor analysis, and IRT analysis of the preliminary data.

Research Question 1a: Preliminary distractor analysis. Due to the lack of unidimensionality suggested by the table of specifications, Classical Test Theory (CTT) was a more appropriate analysis method than IRT for distractor analysis. There are two types of classical discrimination indices. Both indices are reported in Table 3, which lists the following for each item: (a) the item number, (b) the difficulty index—proportion of respondents that answered correctly, (c) the adjusted item-to-total correlation coefficient—the correlation between an individual item and the total score without that item, (d) the D-index—a measure of the discrepancy between the percentage of the top third ability level of the respondents minus the percentage of the lower third of the respondents for each item (e) the number of distractors, not including the correct response to an item, (f) the non-functioning distractors—reporting less than 5% response rate or a response rate higher than that of the correct response, (g) the hyper-functioning distractors—distractors that were selected more frequently than the correct answer, and (h) the percent of omitted responses. These data were informative for improving distractors, as well as identifying problematic items, as part of the CGI TKA revision process.

From Table 3, it is apparent that many of the CGI TKA items would benefit from revision and rewriting of distractors to improve their functionality. CTT analysis also illuminates the lack of discriminating power of some items as justification for revising those items.

A few areas of need for revision became apparent from the classical item analysis statistics. One of the most obvious was the lack of difficulty of the true-false items, with six of the eight items ranging from .713 to .915 in difficulty. Except for Item 22, the first true-false item, the true-false items were very easy for examinees. The fact that the correct answer to all the original true-false items except the last one was true, may have made the items even easier to answer correctly once the pattern was discovered. Changing to a multiple-choice format accomplished more than simply avoiding this pattern. Likewise, several of the items in the original test were linked to a single common stem, i.e. Items 22-25 and 26-29. This was modified for the revised CGI TKA.

Poorly functioning distractors were defined as any distractor which attracted fewer than 5% of the examinees to select that response. All such options were revised to create distractors that would be more attractive to examinees who possessed incorrect or partial knowledge. Some item stems were also revised for clarity during the revision process. Thirteen of the 36 items

Table 3

Item Analysis and Distractor Analysis Statistics for the Original CGI TKA

Item	Difficulty Index	Adjusted Item-to-Total Correlation	*D*-Index	Number of Distractors	Non-functioning Distractors	Hyper-functioning Distractors	Percent Omitted
1	.628	.194	37.5	4	A		
2	.589	-.020	74.1	4			2.1
3	.526	.250	51.6	4	E		1.1
4	.684	.097	54.8	4	B		1.1
5	.344	.244	41.0	4	C		2.2
6	.695	.294	54.8	4	B, C, E		1.1
7	.663	-.009	51.6	4	C, D		1.1
8	.284	-.149	38.7	4			3.2
9	.511	.222	63.9	4	D		1.1
10	.366	.324	8.6	4	A, B		0.0
11	.468	.074	38.7	4	C		1.1
12	.613	.348	63.6	4	A, B		0.0
13	.430	.149	51.0	4	E		0.0
14	.543	.031	46.4	4	C, D		0.0
15	.581	.383	47.2	4			1.1
16	.198	.382	24.4	4		E	2.2
17	.380	.222	24.9	4			0.0
18	.446	.407	60.4	4			0.0
19	.272	.239	44.8	4	D	E	3.3
20	.255	.154	37.0	4	C	D	1.1
21	.242	.238	35.4	4	B	E	2.1
22	.284	.040	45.1	1		B	0.0
23	.747	.133	9.7	1			0.0
24	.839	.081	13.5	1			0.0
25	.817	-.033	7.1	1			0.0
26	.914	.167	20.7	1			0.0
27	.713	-.006	20.6	1			0.0
28	.915	.181	13.5	1			0.0
29	.419	.039	31.4	1		A	0.0
30	.543	.333	29.0	4	E		0.0
31	.554	.122	60.2	2			0.0
32	.570	.123	54.1	2			0.0
33	.366	.205	21.5	2			0.0
34	.484	.212	29.0	2			0.0
35	.372	.255	28.4	2		B	0.0
36	.479	.228	34.5	2			0.0

included options that were revised, which were Items 1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, and 14. Most of these items addressed CGI problem types and naming student strategies. New distractors were created based on insights of the original CGI TKA authors and the current researcher and common misconceptions they have observed from teachers.

The results from the preliminary IRT analyses were based on the potentially faulty assumption that the 36 items were essentially unidimensional, meaning that they all measured the same trait. As described previously, there were six potential constructs being evaluated in the CGI TKA. Factor analysis of the preliminary data suggested retaining four constructs, which resulted in the revised table of specifications shown in Table 4. If the new data had supported this revised table of specifications, IRT analysis would have needed to be conducted separately for each construct.

Table 4

Table of Specifications Based on the Revised CGI TKA

Construct	Understand Concepts/ Principles: Classify, Recognize	Apply: Interpret, Implement	Analyze: Differentiate, Organize, Compare, Examine	Evaluate: Appraise, Value, Select
CGI Problem Types (PT)	3, 4, 5, 6	10, 11, 12, 13		
CGI Solution Strategies (SS)	**16**, 17, **18**, 19, 20, 21, 22, 23, 24, 28, 29		14, 15	**26**, **30**
CGI Principles and Pedagogy (P)	1, **16**, **18**,	2, 8	7, 9, **27**	
Base Ten Understanding (BT)	**28**, **29**		25, **27**	**26**, **30**

Note. Items that were hypothesized to measure multiple constructs are in boldface font.

The variable map produced by the WINSTEPS Rasch model, as can be seen in Figure 1, indicates similar means of the item difficulties and the person ability estimates. This is observed as the "M" indicating the mean and highlighted in yellow on each side of the map, the person side and the item side, are lined up at almost the same level. Thus, the mean of person abilities is very similar to the mean of the item difficulties. Similarly, the standard deviations and tails of the examinees and the item difficulties are also very aligned, sometimes right across from one another, as highlighted in green. These findings suggest the test was not too difficult or too easy for the group. In fact, there were plenty of easy items, mostly the true-false, and there were only four individuals with abilities beyond the highest item difficulty as observed in Figure 1. The item statistics produced by the dichotomously scored Rasch model suggested good item fit with all items ranging between 0.5 and 1.5 in both infit and outfit. The point biserials of the true-false items (Items 22-29) and the "Never, Sometimes, Always" items (Items 31-36) were low, all below .25.

The original 36-item CGI TKA was revised based on the table of specifications, expert judgment, distractor analysis, and IRT analysis. It was also reformatted for digital delivery.

Research Question 1b: Revision of the original CGI TKA. Revisions to the original CGI TKA were made based on modifications to the table of specifications, feedback from CGI experts, and formatting and other testing modifications.

Revision based on the table of specifications. During the revision process, it was discovered that several items measured multiple constructs. Through reanalyzing the items, it was determined that the constructs were not as distinct as originally thought. The modified table of specifications included four constructs as shown in Table 4.

```
MEASURE Person - MAP - Item
           <more>|<rare>
    3            +
                 |
                 |
              X  |
                 |
                 |
             XX  |
                 |
                 |
              X  |
    2           T+T
              X  |
                 |  I0016
                 |
                 |
             XX  |  I0021
                 |  I0020
             XX  |
            XXX  |  I0008  I0019  I0022
                 |
    1     XXXXX S+S
                 |
          XXXXX  |  I0005  I0010  I0017  I0033  I0035
          XXXXX  |
          XXXXX  |
                 |
        XXXXXXX  |  I0013  I0018
           XXXX  |  I0011
              X  |  I0034  I0036
                M|  I0009
    0   XXXXXXX  +M I0003  I0014  I0030  I0031
       XXXXXXXX  |  I0032
           XXXX  |  I0015  I0029
                 |  I0002  I0012
              X  |
        XXXXXXX  |  I0001
                 |
      XXXXXXXXX  |
            XXX  |  I0004  I0006  I0007
                S|  I0027
   -1      XXXX  +S
         XXXXXX  |
                 |  I0023
             XX  |
                 |  I0025
                 |
                 |  I0024
                 |
                T|
                 |
   -2            +T
                 |
              X  |
                 |  I0026
                 |  I0028
                 |
                 |
                 |
              X  |
                 |
   -3            +
           <less>|<freq>
```

Figure 1. IRT Variable map of the original 36-item CGI TKA.

As can be seen in the revised table of specifications, there were still several items that were believed to measure multiple constructs. This was also taken into consideration when considering how many factors to retain during factor analysis. One of the constructs, “Base Ten Understanding,” was almost entirely measured by items that also measure something else. Hence, it was suspected that this construct would likely not be able to stand alone as a factor. In fact, “Base Ten Understanding” and “Number Sentences” were not distinct enough from “Student Strategies” and “Problem Types” to be retained as their own constructs.

Revision based on expert opinion. Based on interviews with three CGI experts using the expert judgement interview protocol shown in Appendix C, several modifications were made to the original test. One of the experts suggested assessing understanding of the relationship between problem type and difficulty, which resulted in the creation of two new items. Other areas that were considered missing from the original instrument but important to measure were (a) the advantages of invented algorithms for students, (b) strategies involving derivation or recall of fact knowledge, and (c) comparison of student solution strategies. These also resulted in additional items.

Some of the suggestions from the experts were considered but did not ultimately result in modifications to the assessment. For example, one of the experts noted that there were not contexts of multiplication or division for some of the problem type questions. Since this construct was already one of the constructs with the largest number of items, it was determined that more items, even if they were slightly varied in content, would not provide substantial new information and would only lengthen the test unduly.

Additionally, there were some suggestions made for additional content to be measured that the test authors determined were more related to practice or instructional pedagogy than to knowledge of the CGI content. Although these would be interesting constructs to consider, they are not part of this knowledge assessment and were not included in the revised assessment.

Formatting and other revisions. During the preliminary data analysis, it was clear that some of the formatting of the original CGI TKA did not allow for reliable item analysis. The testlet scenario where one stem was used for multiple responses was removed. This was done in some cases by simply repeating the stem for each item. For others, the content from two or more of the original items was combined. As for the true-false items, the stems were revised to be distractors for one item, meaning that the eight items were condensed into two items while assessing the same content. These revisions resulted in a reduction of ten items, which allowed for creation of other items that were pointed out as missing according to expert feedback while still shortening the test from 36 to 30 items.

When the test was imported into Qualtrics, special caution was taken to make each item clear and legible. The question stems were presented in large print while still easily displayed on one screen. Many teachers took the test on mobile devices, such as cellular phones, while the majority used computers to complete the test. Both formats were simple to access and easy to read, although some scrolling within an item was required. Additionally, when the test was digitized, the option to force examinees to answer each item was selected so that no items could be skipped or omitted by mistake, as had occurred in the original test. This mean there would be no missing data from completed examinations.

As part of the reformatting, the correct answers to the items were varied purposefully, with the same number of each of the five possible responses. In other words, there are six items to which the correct answer is A, six items to which the correct answer is B, etc. The pattern of answers also varies such that they sometimes repeat two of the same answer in a row, and other times they are spread out, to reduce guessing based on a predictable answer pattern.

Research Question 2: Statistical Analysis of Data from the Revised CGI TKA

Research Question 2 considered analysis of the revised CGI TKA after it was administered to a new group of examinees. Investigation of this question included a series of factor analyses, reliability estimation, IRT analyses, and distractor analyses. In order to satisfy the assumptions of these analyses, several analyses were conducted, including (a) analyzing unidimensionality, (b) testing for tau equivalence, and (c) measuring uncorrelated errors.

Research Question 2a: Exploratory and Confirmatory Factor Analysis using Mplus. Based on the Exploratory Factor Analysis (EFA), 21 items were retained from the 30-item revised test, as 22 items loaded on the one-factor EFA at the 0.3 threshold or higher. One of the items, Item 8, was problematic in Mplus because it was too difficult and had blank correlation table cells, meaning estimation was not possible (Muthén & Muthén, 2012). This resulted in the retention of the following 21 items: items 2, 3, 5, 6, 7, 9, 10, 11, 15, 16, 17, 18, 20, 21, 22, 23, 24, 26, 28, 29, and 30. The factor loadings of the items in Mplus EFA are listed in Table 5.

Although there was not a large enough sample to cross-validate the results of the EFA with a CFA using a different data set, it was decided that a CFA should be conducted on the same data to provide additional evidence as to the unidimensionality of the data. Based on a CFA of a one-factor model using Mplus, all 21 items loaded above a 0.32 threshold, which would be 10% of the variance, with excellent model fit (CFI 0.962 and TFI 0.958). This factor was interpreted as CGI specialized content knowledge. See Table 6 for the one-factor CFA model results.

Table 5

Geomin Factor Loadings for a One-Factor EFA Model of the 30-Item Revised CGI TKA

Item	Factor Loading
1	-.085
2	.300
3	.578
4	.164
5	.576
6	.380
7	.329
8	-.555
9	.386
10	.361
11	.333
12	.003
13	.199
14	.180
15	.730
16	.432
17	.340
18	.397
19	.064
20	.527
21	.589
22	.458
23	.381
24	.444
25	.214
26	.523
27	.014
28	.519
29	.598
30	.503

Table 6

Geomin Factor Loadings for a One-Factor CFA Model of the 21-Item Revised CGI TKA

Item	Factor Loading
2	.327
3	.727
5	.707
6	.439
7	.376
9	.418
10	.389
11	.345
15	.997
16	.449
17	.388
18	.447
20	.603
21	.759
22	.530
23	.340
24	.477
26	.656
28	.621
29	.720
30	.595

Research Question 2b: Distractor analysis. Distractor analysis was conducted using both classical test theory analysis and Item Response Theory analysis. Whereas CTT played an important role in revising the original assessment, the researcher desired to see how the revised distractors functioned. Table 7 shows the performance of the revised distractors. It is important to note that some of the distractors hyper-functioned, meaning they were selected more often than the correct answer while others poorly functioned, meaning they were selected by fewer than 5% of the examinees. Some items may have had poorly performing distractors due to the format of the question. For example, several of the items have responses such as “A, B, and C.” If an examinee were wanting to choose A and B and that was not an option, then “A, B, and C” would be an attractive distractor. However, this option cannot be an enticing distractor without the existence of options A and B. Therefore, some non-functioning distractors were retained in order to provide improved distractors that contain multiple options. Item 19, for example, reads as follows:

> Indicate the number sentence(s) that match the following story problem:
>
> *Matthew collects baseball cards. Each pack of cards he bought had 7 cards. He bought 5 packs of cards. How many baseball cards did Matthew buy?*
>
> a. $7 \times 5 = \square$
> b. $5 \times 7 = \square$
> c. a and b
> d. $7 + 7 + 7 + 7 + 7 = \square$
> e. a, b, and d

Several of the items with distractors that include multiple responses like this have hyper-functioning distractors, while others with a similar format have non-performing distractors. However, it would be difficult to modify these distractors and maintain the integrity of the item, as described. For other items, such as Item 30, the responses are somewhat related which means they cannot be changed without changing the nature of the item itself.

> You observed the following solution to this story problem: *How many eggs are in 59 dozen?* Mark whether the method used is NEVER, SOMETIMES, or ALWAYS useful in solving ANY whole-number multiplication problem?
> $60 \times 10 = 600$
> $60 \times 2 = 120$
> $600 + 120 = 720$
> $720 - 12 = 708$
>
> a. Always   b. Sometimes   C. Never

Changing the distractors would modify the meaning of the question, even if an option is not frequently selected. When considering only the 21 items retained based on factor analyses, the number of poorly functioning and hyper-functioning distractors was greatly reduced, as can be seen in Table 8.

Table 7

Distractor Analysis Results for the 30-Item Revised CGI TKA

Item	Classical Difficulty Index	Number of Distractors	Non-Functioning Distractors	Hyper-Functioning Distractors
1	85.9	4	A, B, D	
2	31.9	4	C	B
3	59.2	4		
4	32.4	4	A	E
5	68.5	4		
6	38.0	4		
7	57.3	4		
8	3.8	4		A
9	58.7	4	E	
10	85.0	4	B, C, E	
11	64.3	4		
12	13.6	4		A, E
13	40.4	4	B	C
14	39.0	4	D	C
15	55.4	4		
16	54.5	4	E	
17	35.7	4	A	
18	31.0	4		
19	36.6	4	A	
20	59.6	4		
21	72.3	4		
22	40.8	4		
23	58.7	4		
24	67.1	4		
25	29.6	4		E
26	53.5	2		
27	58.2	4		
28	37.1	4		
29	37.6	4		C
30	54.5	2		

Table 8

Distractor Analysis Results for the 21-Item Revised CGI TKA

Item	Classical Difficulty Index	Number of Distractors	Non-Functioning Distractors	Hyper-Functioning Distractors
2	31.9	4	C	B
3	59.2	4		
5	68.5	4		
6	38.0	4		
7	57.3	4		
9	58.7	4	E	
10	85.0	4	B, C, E	
11	64.3	4		
15	55.4	4		
16	54.5	4	E	
17	35.7	4	A	
18	31.0	4		
20	59.6	4		
21	72.3	4		
22	40.8	4		
23	58.7	4		
24	67.1	4		
26	53.5	2		
27	58.2	4		
28	37.1	4		
30	54.5	2		

Research Question 2c: Item Response Theory analysis using the Rasch Model. To determine whether the distractors included in the revised CGI TKA offer potential information beyond dichotomous scoring of correct and incorrect, a partial credit model (PCM) was compared to the dichotomous model. Using methodology outlined in both Andrich and Styles (2011) and Sideridis, Tsaousis, and Al Harbi (2016), as described in the method section of this document, each item was weighted using a 2-1-0 scoring method according to the following scoring key:

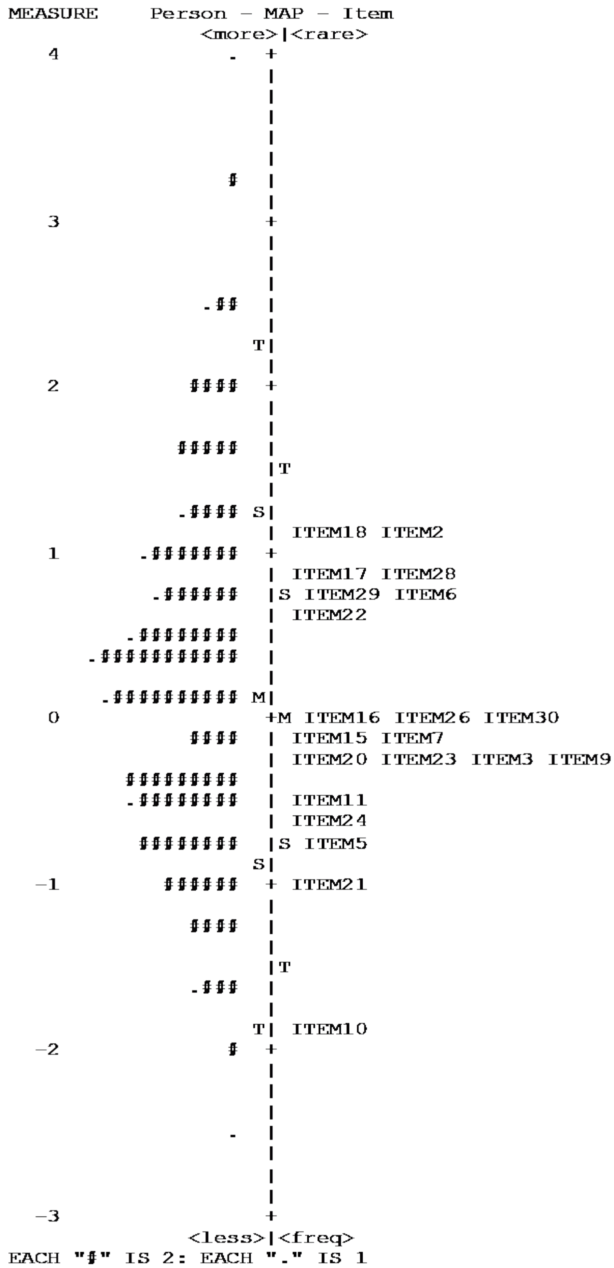
- 0 demonstrates no understanding or a misconception
- 1 is a distractor that demonstrates partial understanding of the question, and 0 demonstrates no understanding or a misconception
- 2 is the correct answer

Both the dichotomous and weighted PCM Rasch model data are compared in Figure 2 as well as Tables 9 and 10.

In both Andrich and Styles (2011) and Sideridis, Tsaousis, and Al Harbi (2016), the researchers used fit statistics to compare models with items they considered contained useful information, ultimately using PCM for three items. This study took a slightly different approach, as described in the method section of this document, by assigning partial credit scoring based on theoretical judgment of content before statistical item analysis. The researcher together with the authors of the CGI TKA examined the distractors in each item and identified potentially informative distractors, those that contain stochastic information. All but two items were considered to have distractors with information that demonstrated partial understanding. Therefore, the model reported in Table 9 and Figure 2 have scored all but two items assigning partial credit. Items 16 and 23 did not have distractors with useful information of partial understanding and were scored dichotomously, as shown in Table 10.

When a one-factor CFA of the 21-item CGI TKA was conducted using the partial credit model, three items did not load above the 0.32 threshold: Items 2, 6, and 22. Upon conducting item analysis using the partial credit model for the other 19 items, there were four items that did not function as expected; the person ability did not increase from the misconception to the partial understanding to the correct answer. Therefore, those four items, Items 6, 7, 20, and 22, were also scored dichotomously within the partial credit model. In Table 10, this distinction is

Dichotomous Items Variable Map



Partial Credit Model Variable Map

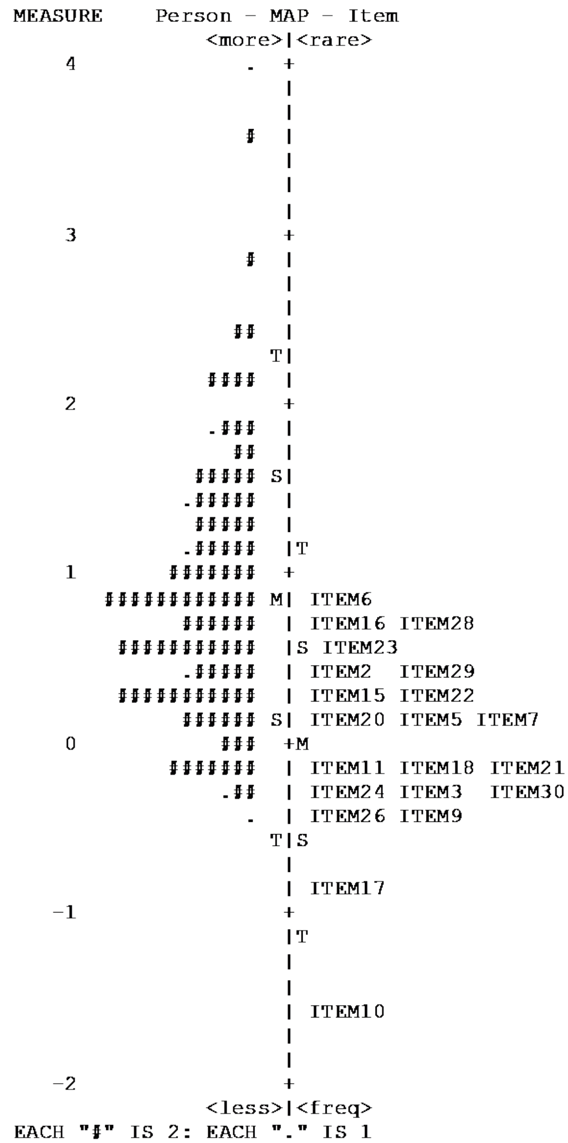


Figure 2. Contrasting variable maps for the dichotomous and PCM from the revised CGI TKA.

observed as Items 16 and 23 only have two scoring options of 0 and 2, while Items 6, 7, 20, and 22 have had the score of one changed to zero so that their responses include 0, 0, 2. This was done to distinguish between the two methods of determining if an item should be scored dichotomously or using partial credit scoring—theoretically and empirically. With the center of person ability and item difficulty aligned for the dichotomous and partial credit model variable maps, as seen in Figure 2, the spread of person ability is smaller for the partial credit model and there are fewer person ability estimates at the lower end of the scale. This makes sense as offering credit for partial understanding would likely make the test easier, raising person ability estimates.

However, in order to accurately compare the ability measures across the dichotomous and partial credit models beyond just an eyeball approach, the ability parameters needed to be transformed to the same scale. In Rasch modeling, the theta or person ability scale is arbitrarily defined, and is, therefore, incommensurable from one analysis to another as provided by the WINSTEPS software. The researcher transformed the parameter estimates from the scale of the dichotomous model to be on the same scale as the partial credit model according to de Ayala's (2009) formula for converting parameter estimates:

$$\xi^{*} = \zeta(\xi) + \kappa \tag{1}$$

where ξ = a theta parameter estimate on the initial metric that the user wants to transform, κ = the location constant, ζ = the scaling constant, and ξ^{*} = the new parameter estimate on the target metric. The scaling constant is calculated as follows:

$$\zeta = \frac{s_{\delta^{*}}}{s_{\delta}} = \left(\frac{\text{Standard deviation of the theta parameter on the target metric}}{\text{Standard deviation of the theta parameter on the initial metric}}\right) \tag{2}$$

The location constant is calculated as follows:

$$\kappa = \bar{\delta}^* - \zeta(\bar{\delta}) \tag{3}$$

where $\bar{\delta}^*$ = the mean of the theta parameters on the target PCM metric and $\bar{\delta}$ = the mean of the theta parameters on the initial dichotomous metric. In this case, the target metric was the PCM because the mean was higher, therefore the transformation calculations would be positive; the target metric is the destination metric onto which the initial metric is transformed. After the person ability parameters were transformed, they were compared as shown in Figure 3.

Table 9

Item Statistics of Dichotomous and Partial Credit Models from the Revised CGI TKA

	Difficulty Parameter		Standard Error		Outfit			
					Dichotomous Model		Partial Credit Model	
Item Number	Dichot-omous Model	Partial Credit Model	Dichot-omous Model	Partial Credit Model	MNSQ	ZSTD	MNSQ	ZSTD
2	1.11	.43	.16	.11	1.22	1.9	1.09	1.1
3	-.27	-.32	.15	.11	.87	-1.5	.82	-1.6
5	-.76	.13	.16	.09	.84	-1.4	.82	-1.0
6*	.78	.85	.16	.09	1.07	.7	1.23	2.4
7*	-.18	.17	.15	.09	1.09	1.0	1.13	1.1
9	-.25	-.45	.15	.12	1.19	2.0	1.08	.7
10	1.86	1.51	.20	.17	1.09	.4	.89	-.4
11	-.54	-.09	.16	.10	1.11	1.0	.99	.0
15	-.08	.22	.15	.09	.80	-2.5	.82	-1.7
16*	-.04	.69	.15	.08	1.04	.5	1.19	1.2
17	.90	-.90	.16	.14	1.15	1.4	1.01	.1
18	1.16	-.20	.16	.14	1.06	.6	1.01	.1
20*	-.29	.19	.15	.09	.92	-.8	1.10	.8
21	-.97	-.18	.17	.10	.82	-1.3	.82	-1.0
22*	.63	.26	.15	.11	1.01	.2	1.07	.8
23*	-.25	.58	.15	.08	1.09	1.0	1.26	1.4
24	-.68	-.33	.16	.11	1.10	.8	1.23	1.6
26	.01	-.46	.15	.12	.92	-.9	.85	-1.5
28	.83	.75	.16	.09	.97	-.2	.99	-.1
29	.80	.46	.16	.10	.87	-1.4	.86	-1.8
30	-.04	-.29	.15	.11	.98	-.2	.92	-.8
Mean	.00	.00	.16	.11	1.01	.1	1.01	.1
P.SD	.75	.56	.01	.02	.12	1.2	.15	1.2

*Scored dichotomously within the partial credit model.

Table 10

Rasch Dichotomous and PCM Person Ability Estimates from the 21-Item Revised CGI TKA

	Dichotomous Scoring				Partial Credit Scoring			
Item	Score	Person		Ability*	Score	Person		Ability*
Number	Value	Count	%	Mean	Value	Count	%	Mean
2	0	145	68	49.67	0	34	16	55.93
					1	111	52	57.33
	1	68	32	57.12	2	68	32	62.14
3	0	87	41	45.82	0	18	8	51.83
					1	69	32	55.46
	1	126	59	56.35	2	126	59	61.36
5	0	67	31	44.65	0	50	23	53.23
					1	17	8	55.23
	1	146	69	55.44	2	146	69	60.89
6	0	132	62	48.85	0	48	23	56.52
					0	84	39	55.87
	1	81	38	57.26	2	81	38	62.49
7	0	91	43	47.70	0	48	23	55.57
					0	43	20	55.34
	1	122	57	55.29	2	122	57	61.02
9	0	88	41	47.67	0	14	7	54.09
					1	74	35	56.21
	1	125	59	55.13	2	125	59	60.59
10	0	32	15	45.00	0	3	1	49.57
					1	29	14	54.45
	1	181	85	53.29	2	181	85	59.46
11	0	76	36	47.49	0	31	15	53.76
					1	45	21	56.70
	1	137	64	54.58	2	137	64	60.38
15	0	95	45	45.49	0	45	21	53.70
					1	50	23	54.68
	1	118	55	57.32	2	118	55	62.20

(Table Continues)

Table 10 Continued

	Dichotomous Scoring				Partial Credit Scoring			
Item Number	Score Value	Person Count	Person %	Ability Mean	Score Value	Person Count	Person %	Ability Mean
16	0	97	46	47.65	0	97	46	55.45
	1	116	54	55.72	2	116	54	61.31
17	0	137	64	49.21	0	3	1	51.29
					1	134	63	56.96
	1	76	36	57.16	2	76	36	61.90
18	0	147	69	49.49	0	10	5	54.25
					1	137	64	57.22
	1	66	31	57.74	2	66	31	62.25
20	0	86	40	46.27	0	39	18	55.03
					0	47	22	54.85
	1	127	60	55.96	2	127	60	61.14
21	0	59	28	44.12	0	31	15	52.48
					1	28	13	54.74
	1	154	72	55.08	2	154	72	60.59
22	0	126	59	48.33	0	93	44	56.77
					0	33	15	55.91
	1	87	41	57.43	2	87	41	62.27
23	0	88	41	47.88	0	88	41	55.61
	1	125	59	54.98	2	125	59	60.78
24	0	70	33	46.41	0	21	10	55.02
					1	49	23	55.28
	1	143	67	54.80	2	143	67	60.32
26	0	99	46	46.64	0	12	6	50.84
					1	87	41	55.77
	1	114	54	56.74	2	114	54	61.65

(Table Continues)

Table 10 Continued

	Dichotomous Scoring				Partial Credit Scoring			
Item Number	Score Value	Person Count	Ability %	Mean	Score Value	Person Count	Ability %	Mean
28	0	134	63	48.37	0	70	33	54.88
					1	64	30	57.41
	1	79	27	58.28	2	79	37	62.98
29	0	133	62	48.00	0	43	20	53.10
					1	90	42	57.30
	1	80	38	58.78	2	80	38	63.12
30	0	97	46	47.01	0	17	8	51.96
					1	80	38	56.36
	1	116	54	56.26	2	116	54	61.19

Note. Ability measures are incommensurable as units of the theta scale are arbitrarily defined.

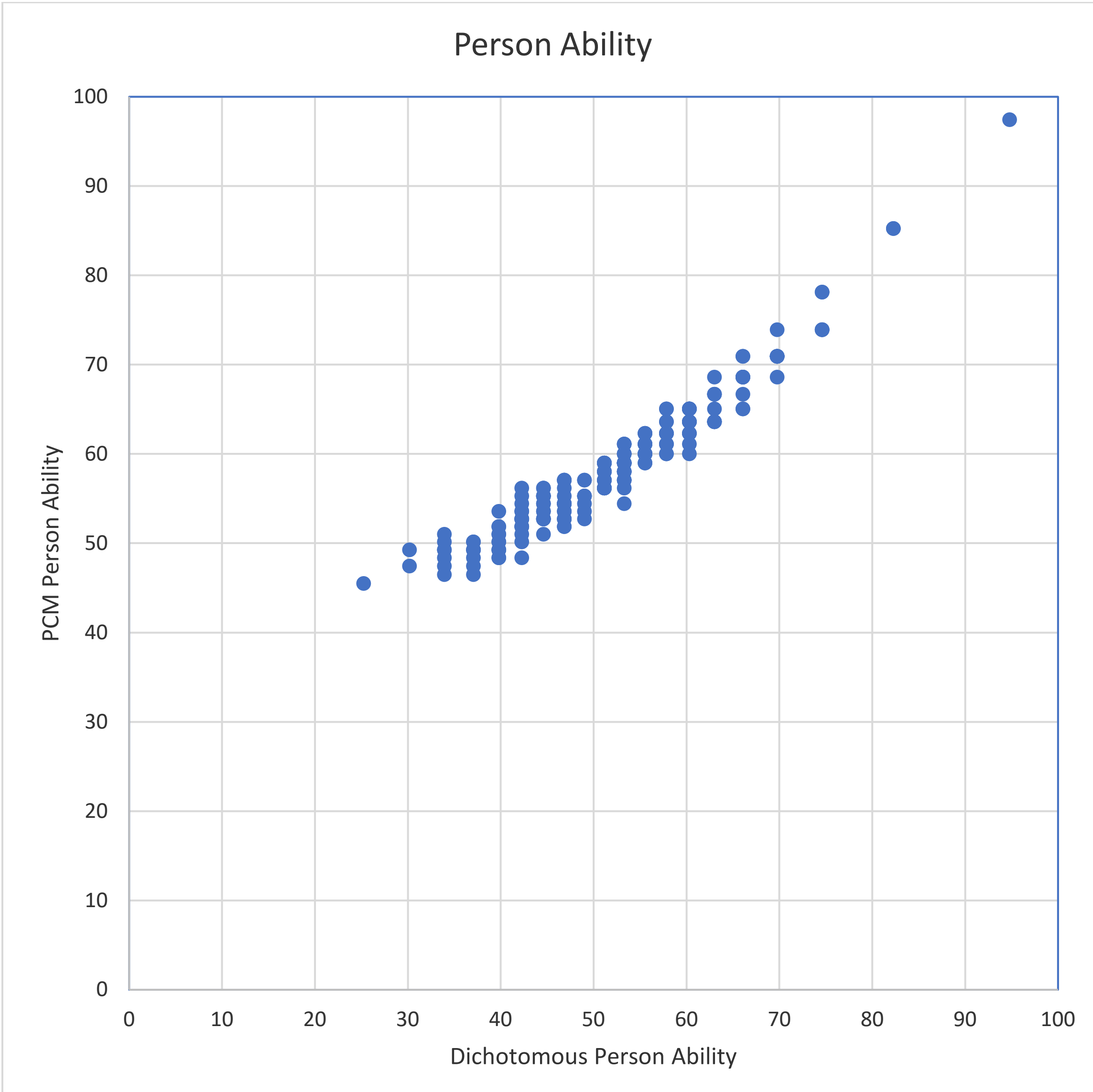


Figure 3. Comparison of person ability parameters of dichotomous and PCM.

Since Rasch analysis assigns the same person ability for any examinees with the same number correct, a column effect is evident in 20 different columns. This is because only one examinee answered all the items correctly. The lowest person ability score for the dichotomous model was 2 of a possible 21, and the lowest score for the PCM was 15 of 42. The mean person ability score was 52.0 for the dichotomous model and 58.6 for the PCM. After the transformation, the dichotomously scored model person ability mean was 59.5 on the same scale as the PCM.

As shown in the item information functions in Appendix D and in Figure 4, using the PCM, Items 16 and 23 provide more information than many of the other items but at a narrow range of person abilities. Other items, such as Items 17 and 18 are ideal candidates for partial credit modeling. As can be seen in the category probability curves shown in Figure 5 and Appendix D and the data count of Table 10, more than half the examinees selected distractors with partial information for these items. Table 9 simply reports these items as quite difficult using the dichotomous scoring. However, as can be seen in Figure 2 and Table 10, the difficulty of these items changes drastically when considering the partial understanding demonstrated by examinees that select distractors with useful information. These changes are even more evident when considering the bimodal output as observed in the item information functions in Figure 6 and in Appendix E for Items 17 and 18. These graphs visually indicate the broad range of information the partial credit model awards these items by giving value to distractors that contain useful information. These items in particular are excellent candidates for use of partial credit scoring.

Dichotomous Model Partial Credit Model

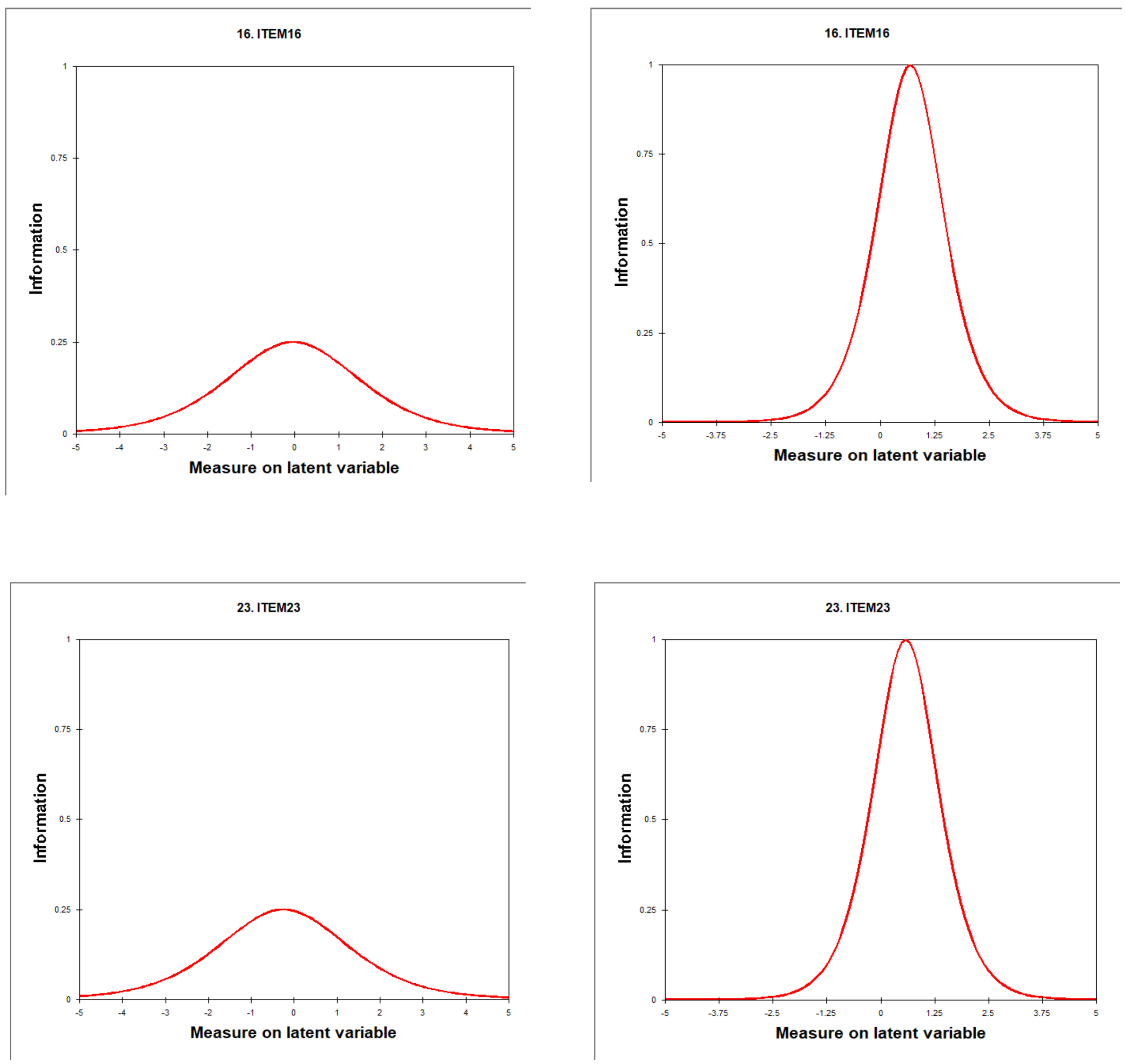


Figure 4. Item information functions for Items 16 and 23 from the dichotomous and PCM.

Other items, though scored using the PCM, do not offer as obvious advantages from one model to the other. For example, as can be seen in the characteristic curves of Figure 7 and Appendix D and the data count of Table 10, the distractors scored as “1” of Items 5 and 21 are not selected enough to produce results very different from the dichotomous model. Upon further investigation of the item characteristic curves from the Rasch model, Item 22 did not seem to have disordered thresholds as previously thought and was therefore scored using the partial credit model.

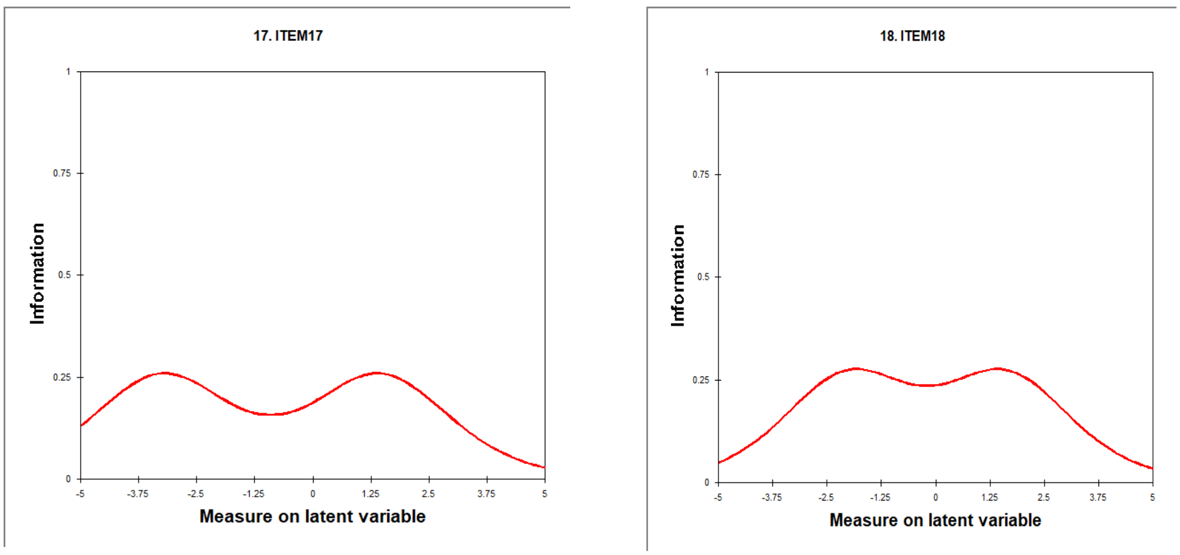


Figure 5. Item information functions for Items 17 and 18 from the PCM.

Dichotomous Model Partial Credit Model

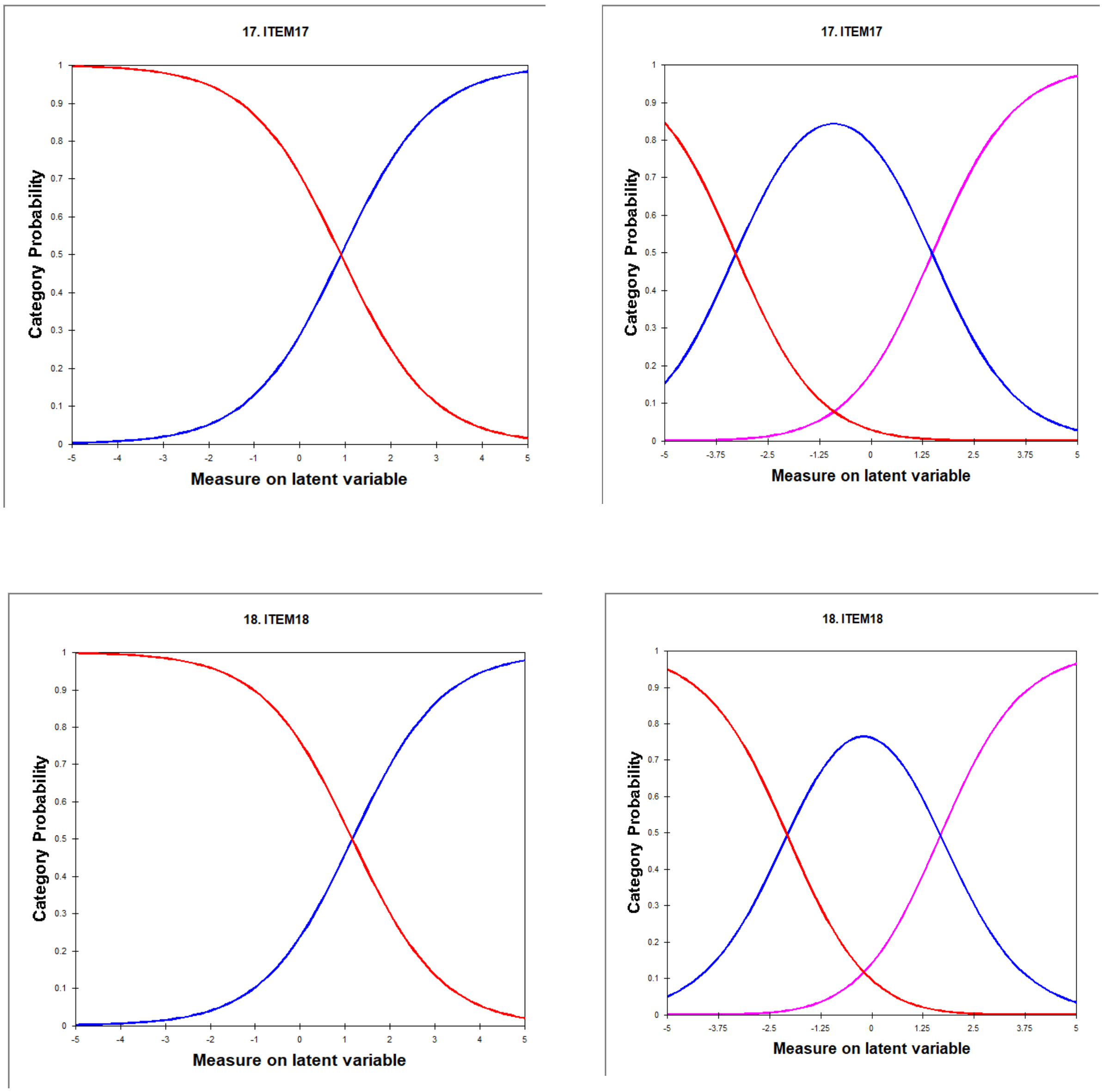


Figure 6. Item characteristic curves for Items 17 and 18 from the dichotomous and PCM.

Dichotomous Model

Partial Credit Model

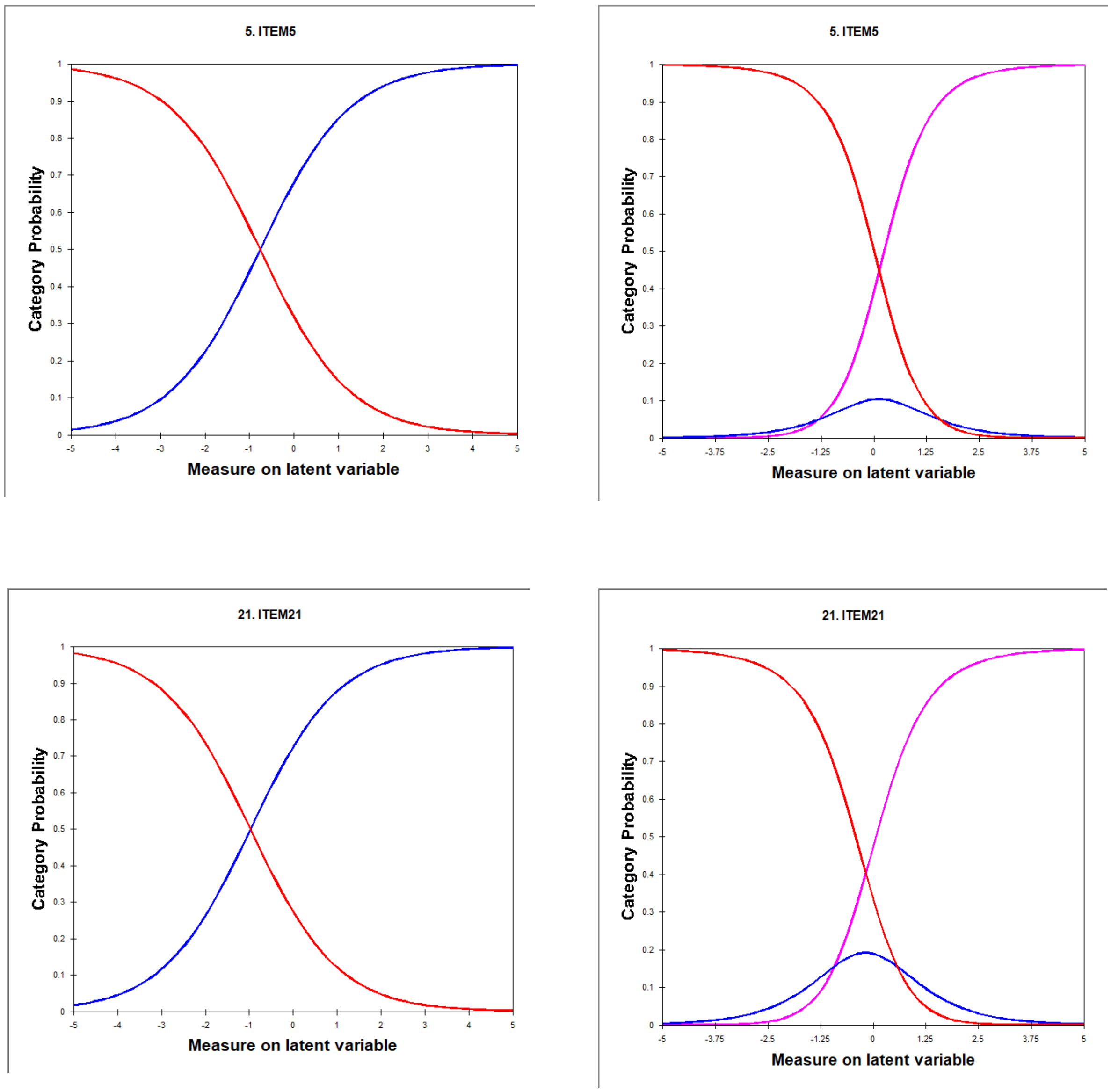


Figure 7. Item characteristic curves for Items 5 and 21 from the dichotomous and PCM.

As far as the overall test is concerned, the test information function has more information at a broader range for the dichotomous model, while the PCM provides additional information in the middle range of abilities, as shown in Appendix F. For most of the participants, there is very little difference in the person ability estimates for each examinee between the dichotomous or partial credit scoring. However, at the extremes of person ability estimates, meaning for those individuals who were estimated to have a very high or very low person ability estimate, the difference between the z-score for these individuals can be much larger, as shown in Appendix G. For the examinee with a perfect score, the difference in the z-scores of the person ability estimates between the two scoring models is 1.2, but this is because the person ability is not accurately estimated for someone who correctly answers all the items. However, there are still 8 examinees where the difference in the z-scores of their ability estimates is greater than 0.40.

To accurately compare the dichotomous and partial credit scoring models, information criteria were calculated from the log likelihood of each of the models as shown in Table 11. Information criteria are not statistics per se but are indices whose values can be compared. In general, the smaller the value of the information criteria, the better the model fits the data. Beretvas and Murphy (2013) suggest the use of different information criteria to correct for the lack of parsimony in fit statistics in various ways. Akaike's information criterion (AIC; Akaike, 1973) is commonly used and accounts for the number of parameters estimated in the model and does not rely upon the sample size:

$$AIC = -2\mathrm{LL} + 2p \tag{4}$$

where p represents the number of parameter estimates estimated in the model. The Bayesian information criterion (BIC; Schwarz, 1978) is designed to correct for $n = 213$, the sample size, and equals:

$$BIC = -2LL + \ln(n)\,p \tag{5}$$

Another consistent criterion is Bozdogan's consistent AIC (AICC; Bozdogan, 1987) and equals:

$$CAIC = -2LL + p[\ln(n) + 1] \tag{6}$$

Hurvich and Tsai (1989) recommend a modification to the AIC designed to correct the AIC's tendency to fit overparameterized models. The modification corrects for sample by including n and is known as the finite sample-corrected AIC (AICC) and equals:

$$AICC = -2\text{LL} + (2pn)/(n - p - 1) \tag{7}$$

Based on the calculation of each of these criterions, the dichotomous and partial credit models accounting for each partial credit item are shown in Table 11.

As shown in Table 11, the model with the lowest information criteria, indicating the best model fit, is the dichotomously scored model. Even considering removal of each of the 10 items identified as containing distractors with useful stochastic information, none of the models fit the data better than the dichotomously scored model.

Research Question 2d: Reliability. When conducting the CFA, the researcher observed disparate factor loadings, with the highest modification index in the Mplus output reported as 5.3. This indicated possible lack of tau equivalence and potentially correlated errors. Since the factor loadings ranged from .327 to .759, a test was conducted to determine if the factor loadings were essentially tau equivalent. The test resulted in statistical difference between the freely estimated and constrained models, meaning the factor loadings were not essentially tau equivalent. This nonequivalence was accounted for by estimating Raykov's (2001) rho for reliability.

Table 11

Information Criterion for Dichotomous and Partial Credit Models of the 21-Item Revised CGI TKA

Type of Model	Number of Parameters Estimated	Log Likelihood Statistic (-2LL)	Akaike IC (AIC)	Bayesian IC (BIC)	Consistent AIC (CAIC)	AIC Corrected (AICC)
21 Dichotomous Rasch Items	21	4979.19	5021.19	5091.78	5112.78	5026.03
10 PCM Items	31	5904.14	5966.14	6070.34	6101.34	5977.10
9 PCM Items: Item 2 Dichot	31	5749.20	5811.20	5915.40	5946.40	5822.16
9 PCM Items: Item 3 Dichot	31	5810.86	5872.86	5977.06	6008.06	5883.82
9 PCM Items: Item 9 Dichot	31	5832.38	5894.38	5998.58	6029.58	5905.34
9 PCM Items: Item 10 Dichot	31	5884.88	5946.88	6051.08	6082.08	5957.84
9 PCM Items: Item 17 Dichot	31	5884.30	5946.30	6050.50	6081.50	5957.26
9 PCM Items: Item 18 Dichot	31	5837.26	5899.26	6003.46	6034.46	5910.22
9 PCM Items: Item 22 Dichot	31	5745.44	5807.44	5911.64	5942.64	5818.40
9 PCM Items: Item 26 Dichot	31	5829.47	5891.47	5995.67	6026.67	5902.44
9 PCM Items: Item 29 Dichot	31	5741.63	5803.63	5907.83	5938.83	5814.59
9 PCM Items: Item 30 Dichot	31	5819.24	5881.24	5985.44	6016.44	5892.20

Likewise, the modification indices of the CFA were examined in order to identify any item pairs which showed evidence of correlated errors. The standardized and unstandardized factor loadings were the same since the scale had been set to 1.0. When determining which modification indices were within an acceptable range, the cutoff used was indices less than 3.84, meaning the expected parameter change should be less than .3 magnitude. A modification index of 3.84 means that probability of getting $p = .005$ with one degree of freedom, resulting in a statistically significant difference, i.e. anything above that cutoff would estimate statistically different results because of correlated errors. There were three pairs of correlated errors above the cutoff that were subsequently tested for significance, and the chi square differential test determined if the models were statistically different. The chi square test of model fit was 194.862 with 186 degrees of freedom and a p-value of 0.313, meaning the model was not statistically significant. In that model, each pair of correlated errors was also tested for statistical significance. It was determined that Item 29 correlated residuals with Items 10 and 16, and Item 30 with Item 26. The values of these correlations and their statistical significance are reported in Table 12.

Table 12

Highest Modification Indices and Related Parameters for Factor Loadings

Item Pair	Modification Index	Expected Parameter Change	Estimated Correlation	Two-tailed p Value
29 with 10	4.323	.378	.368	.016
29 with 16	4.576	.316	.315	.007
30 with 26	5.326	.337	.306	.006

Therefore, the assumptions of tau equivalence and uncorrelated errors were not satisfied and neither Cronbach's alpha nor McDonald's omega was an appropriate method for estimating reliability of the data. Instead, Raykov's rho was a more appropriate estimator of the reliability since these two assumptions were not satisfied. Raykov's rho was calculated using the following equation (Bacon, Sauer, & Young, 1995):

$$\rho = \frac{(\sum \lambda_i)^2}{k + (\sum \lambda_i)^2 - \sum \theta_{ii} + 2 \sum \theta_{ij}} \tag{8}$$

where k = the number of items, λ_i = the standardized factor loading of the i^{th} item, θ_{ii} = the residual or unique variance of the i^{th} item and θ_{ii} is essentially zero for each pair of items other than those in θ_{ij}, and θ_{ij} = the residual of each pair of items that is non-zero. These calculations produced a reliability estimate of .8370 for the scores of the 213 examinees for the dichotomous scoring model. Raykov's rho was subsequently calculated for the partial credit scoring of 10 items and was estimated to be .8264.

CHAPTER 5

Discussion

Revision of the Original CGI TKA Based on Results

Research Question 1 considered the need to revise the original CGI TKA to potentially improve the test. Sufficient evidence supported modifying the original CGI TKA in an attempt to improve it. This evidence was based on item response theory analysis, expert interviews, consideration of tables of specifications, and revision of the content of item stems and distractors. Since modifying the original assessment rendered it obsolete, most of this discussion focuses on the results of data obtained from the revised CGI TKA.

Factor Analyses of the Revised CGI TKA

Several iterations of Exploratory Factor Analysis were conducted in an attempt to determine how many factors were present in the data, and which items either did not load on any factors or crossloaded on more than one factor. This analysis was completed as a preliminary step to performing CFA. In each model where more than one factor was included, no more than 17 items loaded on the factors, even though model fit was good. When the one-factor analysis was conducted and 21 items loaded well enough, it was determined that some of the suspicions described in the preliminary results were accurate. It was hypothesized that the constructs were very related and actually measuring the same knowledge. For example, how could an examinee demonstrate understanding of base ten strategies without knowing about student solution strategies within CGI generally? Likewise, how could an examinee differentiate between types of problems without understanding the underlying principles of CGI, which includes how problem structures differ? Both conceptually and empirically, it was determined that the CGI

TKA measures the specialized content knowledge of CGI generally, and not discrete constructs within CGI knowledge.

To confirm the existence of only one factor, as well as to determine which items effectively measure that factor, a confirmatory factor analysis was conducted. As can be seen in Table 6, all 21 items identified by the EFA loaded satisfactorily on the CFA, confirming the existence of one factor. What this analysis means for the future of the CGI TKA is that only 21 items need to be administered, rather than all 30 items, while still effectively measuring teacher's knowledge of Cognitively Guided Instruction. Additionally, the existence of only one factor simplifies the reporting procedure for future test administrations. Given that the test appears to effectively measure teachers' CGI knowledge as a whole, only one total score should be reported for each examinee.

Distractor Analysis of the Revised CGI TKA

As is evident in Tables 7 and 8, another reason to retain the 21 items suggested by the factor analyses is that the distractors of the items function better. Items 9, 10, 16, and 17 still had some room for potential improvement of distractors, especially Item 10. Three of the current distractors are not attractive responses. The other three items, Items 9, 16, and 17, could likely have the one non-functioning distractor be revised on a future revision of the assessment, if considered necessary. Although Item 2 has a non-functioning distractor, it also has a hyper-functioning distractor. A hyper-functioning distractor is where the distractor is selected more often than the correct answer. As far as distractors goes, if a distractor functions so well that it is selected more frequently than the correct response, if functions extremely well as a distractor. This is why it is being referred to here as hyper-functioning, which is very desirable. When referring to Item 2, revision of option C may be possible, but doing so might affect how option B performs, and potentially could remove some of its hyper-functionality.

IRT Analysis of the Revised CGI TKA

In order to consider how well items and their distractors function in the revised CGI TKA, IRT was conducted using Winsteps. All the items behaved well according to the Rasch analysis, with the exception of Item 8. Although there did not appear to be anything very distinct theoretically about Item 8, in this group of 213 examinees, this item proved to be extremely difficult, with only 3.8% of examinees answering correctly. Item 8 was a new item, written in response to suggestions from CGI experts during the interviewing process. It was intended to assess underlying principles of CGI. It may be that this item was not worded clearly enough, or that the distractors were too similar to the correct answer. In the future, this item could be revised and included in future versions of the CGI TKA with additional analysis. However, it was determined to exclude this item from the analysis of the revised CGI TKA at this time. Therefore, the researcher would suggest eliminating this item and keeping only the 21 items that measure CGI knowledge well.

Advantages and Disadvantages of PCM

To determine the advantages and disadvantages resulting from using partial credit scoring as compared to dichotomous scoring, Rasch analyses were conducted with data from the revised CGI TKA. Since many of the distractors represented ideas that were partially correct, they may contain useful stochastic information when attempting to measure teachers' specialized content knowledge of CGI (Sideridis, Tsaousis & Al Harbi, 2016). On the other hand, many other items included distractors based on what the authors of the CGI TKA considered common misconceptions. It was, therefore, of interest to know if the distractors themselves contain useful stochastic information beyond the correct response. For this reason, the partial credit model was used to model items that included potentially informative distractors.

As can be seen in the variable maps of the dichotomous and partial credit models in Figure 2, there are advantages and disadvantages to each model, depending on the need for information for certain groups of examinees. The dichotomous model has a much wider spread of measuring person ability, meaning that the test information function is greater at the extremes of the abilities of the examinees, those that performed exceedingly well and those that performed at a low person ability level. The range of the information function is broader. However, the information function for the PCM is concentrated near the middle of the ability scale. Therefore, if one were interested in discriminating between persons as in a cutoff scenario, the partial credit model would provide more precise ability estimates to determine differences between examinees in this narrower band of scores. For example, if this test were used for determining whether a teacher should receive certification or an endorsement in CGI, the PCM data could provide evidence for both natural gaps in ability level, and more accuracy near the cutoff score. Likewise, the item and test information functions in Appendix D show that each item and therefore cumulatively the overall test, could potentially provide more information by including distractors that demonstrate partial understanding using the PCM, depending on the cutoff point.

When considering the scatter plot of the transformed person ability parameters in Figure 3, there are no outliers, which means there aren’t any data in which one model produces a drastically different result for an individual examinee than the other. However, after considering the difference in the z-scores for individual examinees, there is what is known as a *Matthew effect*—those at the extremes become more extreme. In this case, those with high person ability estimates using one scoring have a very high likelihood of having a high person ability estimate with the other scoring method. The only exceptions are at the very extremes. In those cases, there were some variations in the expected correlation, producing a larger than expected difference in z-scores. Therefore, the researcher suggests that the additional information provided by using the PCM diminishes the broad range of scores that is usually preferred by a measure. It is easier to determine when a teacher has CGI knowledge using the dichotomous model, as partial understanding of some distractors does not influence the overall score. For this reason, as well as ease of dichotomous scoring, it is suggested that the advantages of discrimination in the middle range of scores is not enough in most cases to warrant the increased labor of scoring using the partial credit model. The information provided by the dichotomously scored model is excellent at measuring a teacher's knowledge of Cognitively Guided Instruction.

When comparing the information criterions for the dichotomous and the partial credit scoring of the 10 items that are most likely to offer useful stochastic information, the dichotomous model fits better than any model incorporation partial credit scoring. This suggests yet again that the dichotomous model is the best fitting for scoring the CGI TKA.

Reliability of Scores from the Revised CGI TKA

When accounting for correlated errors and the lack of tau equivalence, using Raykov's rho, the reliability estimate for the dichotomous scores was .8370, placing the scores in an acceptable reliability range. The rho estimate for the partial credit scores was .8264, which is also acceptable, but below the estimate for the dichotomous scores.

Limitations

The researcher recognizes several limitations in this study. A few of the most obvious include the data collection method. Since not all the testing occasions were proctored, the researcher must assume the integrity of the results was similar in that teachers had the same restrictions as to material they could access during the test. The researcher must assume that the knowledge was based on the teachers' own understanding and was not the result of access to printed or other material during testing.

The small sample size was a limitation in not allowing 2-PL or 3-PL models to offer estimates of fully converged models, which could have also included guessing parameter estimation. This was primarily due to the busy nature of schoolteachers and leaders and the lack of proximity of the researcher to the participants.

Other limitations included the inability to conduct content validity analysis and teacher think-alouds as part of the revision process due to the lack of compensatory funding available to the researcher to warrant participation in such activities.

Recommendations

Suggestions for future research. For future distractor analysis of the revised CGI TKA, the researcher recommends utilizing an IRT nominal-response model that would provide a discrimination parameter for each of the distractors. Subsequent revisions of the new CGI TKA could be informed by such modeling.

Other suggestions for future research would include using demographic data to determine how the test functions for different subgroups of the population. Demographic data were gathered at the time of assessment to potentially perform Differential Item Functioning (DIF) testing and to conduct correlations between scores and professional characteristics of teachers. The demographic information gathered included the following: (a) who delivered CGI instruction to teachers, (b) years of teaching experience, (c) years of experience teaching CGI, (d) current position, and (e) grade level currently teaching. Personal demographics such as gender, race, age, etc., are not necessary to consider, and avoids personally identifiable information that may put teachers at risk. Although DIF analysis was not performed as part of this study, the data were collected and can be further analyzed in the future as the test is given to different demographic groups.

Future use of the instrument. The purpose of this study was to analyze the validity of the CGI TKA. The desire to do so comes with the assumption that this test would be useful to certain groups of educators. Depending upon the intended use of the assessment, one would want to score in a certain way, dichotomously or using partial credit scoring. This suggests that the CGI TKA could have a variety of uses. Up until now, the CGI TKA has been used to consider the effectiveness of professional development and need for further support of teachers and school leaders. However, this assessment could potentially be used for certification, advancement, or endorsement purposes if so desired. Because it is not fully determined how the CGI TKA will be used in its current revised form, the revised measure has not been included in its entirety here. In order to illustrate the content of the assessment, some sample items from the original CGI TKA have been included in Appendix A. However, in order to access the complete revised assessment, please contact Debra Fuentes at debra.fuentes@byu.net.

Conclusions

The 21 items of the revised CGI TKA constitute a valid assessment of teachers' specialized content knowledge of CGI. Not only do the items perform well with regards to difficulty, nearly all the distractors function properly, and in many cases can offer information of partial understanding if so desired. Although the use of partial credit scoring is justified for 10 items, the benefits do not outweigh the additional effort required. Dichotomous scoring of the 21 items is recommended for most testing situations. The reliability of the dichotomous scores obtained is good when estimated using Raykov's rho, and the shortened test means less examinee fatigue or intrusion. When considering the research questions of determining the validity of the revised CGI TKA, and the reliability of the scores obtained from the test, the researcher concludes that the current 21-item test is valid and psychometrically sound as a measure of CGI specialized content knowledge.

REFERENCES

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F Csaki (Eds.), *Second international symposium on information theory* (pp. 267-281).Oxford, England: Oxford University Press.

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York, NY: Longman.

Andrich, D. & Styles, I. (2011). Distractors with information in multiple-choice items: A rationale based on the Rasch model. *Journal of Applied Measurement, 12*(1), 67-95.

Bacon, D. R., Sauer, P. L., Young, M. (1995). Composite reliability in structural equations modeling. *Educational and Psychology Measurement, 55*(3), 394-406.

Ball, D. L., & Forzani, F. M. (2010). What does it take to make a teacher? *Phi Delta Kappa*, *92*(2), 8-12.

Ball, D. L., & Forzani, F. M. (2011). Building a common core for learning to teach: And connecting professional learning to practice. *American Educator*, *35*(2), 17-21.

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, *59*(5), 389-407.

Beretvas, N. S. & Murphy, D. L. (2013) An evaluation of information criteria use for correct cross-classified random effects model selection. *Journal of Experimental Education, 81*(4), 429-463.

Bird, T., Anderson, L. M., Sullivan, B. A., & Swidler, S. A. (1993). Pedagogical balancing acts: Attempts to influence prospective teachers' beliefs. *Teaching and Teacher Education*, *9*(3), 253-267.

Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika, 52*(3)*,* 345-370.

Buehl, M. M., & Fives, H. (2009). Exploring teachers' beliefs about teaching knowledge: Where does it come from? Does it change? *The Journal of Experimental Education*, *77*(4), 367-408.

Cai, L., Thissen, D., & Du Toit, S. (2005). *IRTPRO* (Version 2.1). Lincolnwood, IL: Scientific Software International.

Carpenter, T. P., Fennema, E., Franke, M. L., Levi, L., & Empson, S. B. (1999). *Children's mathematics: Cognitively Guided Instruction*. Portsmouth, NH: Heinemman.

Carpenter, T. P., Fennema, E., Franke, M. L., Levi, L., & Empson, S. B. (2000). *Cognitively Guided Instruction: A Research-based teacher professional development program for elementary school mathematics* (Research Report No. 003). Arlington, VA: National Science Foundation.

Carpenter, T. P., Fennema, E., Franke, M. L., Levi, L., & Empson, S. B. (2015). *Children's mathematics: Cognitively Guided Instruction* (2nd Ed.). Portsmouth, NH: Heinemann.

Carpenter, T. P., Fennema, E., Peterson, P. L., & Carey, D. A. (1988). Teachers' pedagogical content knowledge of students' problem solving in elementary arithmetic. *Journal for Research in Mathematics Education*, *19*(5), 385-401.

Carpenter, T. P., Fennema, E., Peterson, P. L., Chiang, C. P., & Loef, M. (1989). Using knowledge of children's mathematics thinking in classroom teaching: An experimental study. *American Educational Research Journal*, *26*(4), 499-531.

Clark, C. M., & Peterson, P. L. (1986). Teachers' thought processes. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 255-296). New York. NY: Macmillan.

de Ayala, R. J. (2009). *The theory and practice of item response theory*. New York, NY: Guilford.

Empson, S. B. & Jacobs, V. J. (2008). Learning to Listen to Children's Mathematics. In T. Wood (Series Ed.) & P. Sullivan (Vol. Ed.), *International handbook of mathematics teacher education: Knowledge and beliefs in mathematics teaching and teaching development* (vol.1, pp. 257-281). Rotterdam, the Netherlands: Sense Publishers.

Enochs, L. G., Smith, P. L., & Huinker, D. (2000). Establishing factorial validity of the mathematics teaching efficacy beliefs instrument. *School Science and Mathematics*, *100*(4), 194-202.

Feiman-Nemser, S. (2001). Helping novices learn to teach: Lessons from an exemplary support teacher. *Journal of Teacher Education*, *52*(1), 17-30.

Fennema, E., Carpenter, T. P., Franke, M. L., Levi, L., Jacobs, V. R., & Empson, S. B. (1996). A longitudinal study of learning to use children's thinking in mathematics instruction. *Journal for Research in Mathematics Education*, *27*(4), 403-434.

Grossman, P. L., & Wilson, S. M., & Shulman, LS (1989). Teachers of substance: Subject matter knowledge for teaching. In M. C. Reynolds (Ed.), *Knowledge base for the beginning teacher* (pp. 23-36). Oxford, UK: Pergamon Press.

Handal, B., & Herrington, A. (2003). Mathematics teachers' beliefs and curriculum reform. *Mathematics Education Research Journal*, *15*(1), 59-69.

Hill, H. C. (2010). The nature and predictors of elementary teachers' mathematical knowledge for teaching. *Journal for Research in Mathematics Education, 41*(5), 513-545.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, *42*(2), 371-406.

Hill, H., Schilling, S., & Ball, D. (2004). Developing measures of teachers' mathematics knowledge for teaching. *The Elementary School Journal, 105*(1), 11-30.

Hurvich, C. M., & Tsai, C. L. (1989). Regression and time series model selection in small samples. *Biometrika, 76*(2), 297-307.

Lortie, D. 1975. *Schoolteacher: A sociological study*. Chicago, IL: University of Chicago Press.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*(2)*,* 149-174.

Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide* (7th ed.), Los Angeles, CA: Author.

Myers, K. D., Swars Auslander, S., Smith, S. Z., Smith, M. E., Fuentes, D. S. (2019). Developing the pedagogical capabilities of elementary mathematics specialists during a K-5 mathematics endorsement program*. Journal of Teacher Education, 0*(0). Published online June, 2019.

National Council of Teachers of Mathematics (2000). *Principles and standards for school mathematics.* Reston, VA: National Council of Teachers of Mathematics.

National Governors' Association Center for Best Practices & Council of Chief State School Officers (2010). *Common core state standards for mathematics.* Washington, DC: Authors.

Peterson, P. L., Fennema, E., Carpenter, T. P., & Loef, M. (1989). Teachers' pedagogical content beliefs in mathematics. *Cognition and Instruction*, *6*(1), 1-40.

Philipp, R. A. (2007). Mathematics teachers' beliefs and affect. In FK Lester, Jr.(Ed.), *Second Handbook of research on mathematics teaching and learning* (pp. 257-315). Charlotte, NC: Information Age Publishing.

Pianta, R. C., LaParo, K. M., & Hamre, B. K. (2008). *The classroom assessment scoring system* (Pre-K Manual). Baltimore, MD: Paul H. Brookes Publishing Company.

Raykov, T. (2001). Bias of coefficient α for fixed congeneric measures with correlated errors. *Applied Psychological Measurement, 25*(1), 69-76.

Raymond, A. M. (1997). Inconsistency between a beginning elementary school teacher's mathematics beliefs and teaching practice. *Journal for Research in Mathematics Education*, *28*(5), 550-576.

Reys, B. J., & Fennell, F. (2003). Who should lead mathematics instruction at the elementary school level? A case for mathematics specialists. *Teaching Children Mathematics*, *9*(5), 277-282.

Richardson, V. (1996). The role of attitudes and beliefs in learning to teach. *Handbook of Research on Teacher Education*, (Vol. 2, pp. 102-119).

Romberg, T. A., & Carpenter, T. P. (1986). Research on teaching and learning mathematics: Two disciplines of scientific inquiry. *Handbook of Research on Teaching*, (Vol. 31, pp. 850-873).

Schoen, R. C., Bray, W., Wolfe, W., Tazaz, A., & Nielsen, L. (2017). Developing an assessment instrument to measure early elementary teachers' mathematical knowledge for teaching. *The Elementary School Journal, 118*(1), 55-81.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*(2), 461-464.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*(2), 4-14.

Sideridis, G., Tsaousis, I., & Al Harbi, K. (2016). Improving measures via examining the behavior of distractors in multiple-choice tests: Assessment and remediation. *Educational and Psychological Measurement*, *77*(1), 82-103.

Smith, M. E., Swars, S. L., Smith, S. Z., Hart, L. C., & Haardörfer, R. (2012). Effects of an additional mathematics content course on elementary teachers' mathematical beliefs and knowledge for teaching. *Action in Teacher Education*, *34*(4), 336-348.

Smith, S. Z., & Smith, M. E. (Unpublished, 2014). *Cognitively Guided Instruction: Teacher knowledge assessment.* Shared directly with researcher by Authors.

Swars, S. L. (2005). Examining perceptions of mathematics teaching effectiveness among elementary preservice teachers with differing levels of mathematics teacher efficacy. *Journal of Instructional Psychology*, *32*(2), 139-147.

Swars, S., Hart, L. C., Smith, S. Z., Smith, M. E., & Tolar, T. (2007). A longitudinal study of elementary pre-service teachers' mathematics beliefs and content knowledge. *School Science and Mathematics*, *107*(8), 325-335.

Swars, S. L., Smith, S. Z., Smith, M. E., & Hart, L. C. (2009). A longitudinal study of effects of a developmental teacher preparation program on elementary prospective teachers' mathematics beliefs. *Journal of Mathematics Teacher Education*, *12*(1), 47-66.

Tarr, J. E., Reys, R. E., Reys, B. J., Chavez, O., Shih, J., & Osterlind, S. J. (2008). The impact of middle-grades mathematics curricula and the classroom learning environment on student achievement. *Journal for Research in Mathematics Education, 39*(3), 247-280.

Thames, M. H., & Ball, D. L. (2010). What math knowledge does teaching require? *Teaching Children Mathematics*, *17*(4), 220-229.

Thompson, A. G. (1992). Teachers' beliefs and conceptions: A synthesis of the research. In D. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning* (pp. 127-146).

Wilson, M., Cooney T. J. (2002). Mathematics teacher change and development. In Leder G. C., Pehkonen E., Törner G. (Eds.), *Beliefs: A hidden variable in mathematics education?* (Vol. 31; pp. 127–147). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Wright, B. D., & Linacre, M. (1998). *Winsteps*. Chicago, IL: Mesa Press

APPENDIX A

Sample Items from the Original CGI TKA

Instructions: Mark the letter of the best answer for each numbered question.

1. Which of the following statements best explains the mathematical features of addition and subtraction story problems that most clearly influence problem difficulty and children's solutions?

a. Subtraction problems are more difficult than addition problems, so children must be able to recognize whether a problem is addition or subtraction to know how to solve it.
b. The number sizes or particular number combinations largely determine the problem difficulty and appropriate solution method.
c. A problem with a clearly described action and passage of time is easier to solve than a problem that describes a static relationship.
d. The structure of the story identifies the position of the unknown quantity, affects the problem difficulty, and influences the ways children will attempt to solve it.
e. Syntax or the specific wording of a problem is the most important factor in problem difficulty and selection of a solution method.

4. Which of the following story problems is an example of the Part-Part-Whole, Part Unknown problem type?

a. Kevin has 7 marbles. How many more marbles does he need to have 11 marbles?
b. Kevin had 11 marbles. He lost 4 of them. How many marbles does he have now?
c. Kevin had 11 marbles. He lost all of the red ones. Now he has 7 blue marbles. How many red marbles did Kevin lose?
d. Kevin has 7 blue marbles and 4 red marbles. How many marbles does Kevin have?
e. Kevin has 11 marbles. Some are red and 7 are blue. How many red marbles does Kevin have?

7. A child was asked to write a story problem for which $\square \times 12 = 139$ is the number sentence that matches the story. Which of the following story problems is the best response to this task?

a. The baker has 139 cupcakes. She wants to put them in 12 boxes, with the same number of cupcakes in each box. How many cupcakes are in each box?
b. The baker had some cupcakes in the display case. She baked 12 more cupcakes and put them in the display case. Now she has 139 cupcakes. How many cupcakes did she have at first?
c. The baker has 139 cupcakes. She wants to put them in 12 boxes, with the same number of cupcakes in each box. How many cupcakes will be left over?
d. The baker bakes 11 batches of cupcakes, with 12 cupcakes in each batch. How many cupcakes did she bake?
e. The baker baked 139 cupcakes. She wants to put them in boxes with 12 cupcakes in each box. How many boxes of cupcakes will she have?

Six children's solutions to the following story problem are provided: *Elisha has 37 dollars. How many more dollars does she need to have 53 dollars altogether?* For each of the children, give the specific name of the strategy used.

10. Nicole: Says, "37, 47," then raises 1 finger at a time while saying, "48, 49, 50, 51, 52, 53. That's 16."

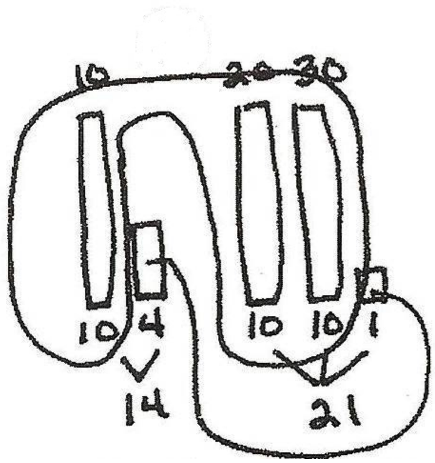
a. Direct Modeling Joining All
b. Direct Modeling Joining To
c. Counting On From First
d. Counting On To
e. Incrementing Invented Algorithm

17. The following story problem was given: *Last year I had 36 comic books in my collection. This year I collected 58 more comic books. How many comic books do I have now?* Which of the following solutions provide an example of the Incrementing Invented Algorithm?

a. 30 + 50 = 80, 80 + 8 = 88, 88 + 2 = 90, 90 + 4 = 94
b. 50 + 30 = 80, 8 + 6 = 14, 80 + 14 = 94
c. 58 is 2 less than 60, so 60 + 36 = 96, 96 – 2 = 94
d. 36 + 50 = 86, 86 + 4 = 90, 90 + 4 = 94
e. a and d

22-25. Children in Kindergarten should be asked to solve multiplication and division story problems because . . .	TRUE	FALSE
22. Those problem types are easier than many addition and subtraction problem types.	a.	b.
23. Multiplication and Measurement Division problems involving groups of 10 are the primary contexts for developing understanding of the base-ten number system.	a.	b.

30. Here is a child's solution strategy for the following story problem: *The little old lady who lived in a shoe had 14 sons and 21 daughters. How many children did she have?*



What is the name of the strategy and what is the evidence of base-ten understanding? (i.e., where on the Number Stories Observation Framework would you place this strategy?)

a. Direct Modeling Joining All with Base-Ten Evidence – More than one ten as a unit
b. Direct Modeling Joining All with Base-Ten Evidence – Direct Place Value Explanation
c. Direct Modeling Joining All with Base-Ten Evidence – Collects tens and ones
d. Direct Modeling Joining All with Base-Ten Evidence – Increments by tens
e. Counting On From First with Base-Ten Evidence – Increments by tens

APPENDIX B

Brigham Young University Internal Review Board Approval

and Study Consent Form

Memorandum
To: Debra Fuentes
Department: EIME
College: EDUC
From: Sandee Aina, MPA, IRB Administrator
Bob Ridge, PhD, IRB Chair
Date: January 25, 2018
IRB#: X17304
Title: *"A Validity Study of the Cognitively Guided Instruction Teacher Knowledge Assessment"*

Brigham Young University's IRB has approved the research study referenced in the subject heading as expedited level, category 7. The approval period is from **January 25, 2018 to January 24, 2019**. Please reference your assigned IRB identification number in any correspondence with the IRB. Continued approval is conditional upon your compliance with the following requirements:

1. CONTINGENCY: school district approval
2. A copy of the informed consent statement is attached. No other consent statement should be used. Each research subject must be provided with a copy or a way to access the consent statement.
3. Any modifications to the approved protocol must be submitted, reviewed, and approved by the IRB before modifications are incorporated in the study.
4. All recruiting tools must be submitted and approved by the IRB prior to use.
5. In addition, serious adverse events must be reported to the IRB immediately, with a written report by the PI within 24 hours of the PI's becoming aware of the event. Serious adverse events are (1) death of a research participant; or (2) serious injury to a research participant.
6. All other non-serious unanticipated problems should be reported to the IRB within 2 weeks of the first awareness of the problem by the PI. Prompt reporting is important, as unanticipated problems often require some modification of study procedures, protocols, and/or informed consent processes. Such modifications require the review and approval of the IRB.
7. A few months before the expiration date, you will receive a continuing review form. There will be two reminders. Please complete the form in a timely manner to ensure that there is no lapse in the study approval.

IRB Secretary
A 285 ASB
Brigham Young University
(801)422-3606

Brigham Young University - Consent to be a Research Subject

Introduction:
This research study is being conducted by Debra Fuentes (PhD Candidate) at Brigham Young University. You were invited to participate because you have previously been trained in Cognitively Guided Instruction (CGI). This study focuses on an assessment of teachers' knowledge of CGI. I am trying to improve the assessment based on the scores provided by you and other teachers. I know your time is valuable and I greatly appreciate your assistance.

Procedures:
For this study you will be asked to complete as assessment comprised of questions and prompts related to various aspects of CGI. The estimated time required to take the survey is approximately 30-40 minutes.

Risks/Discomforts:
You will be asked questions about CGI in an effort to assess your knowledge, which may cause some discomfort. However, the risks associated with participating in this study are likely no more than what you would encounter in everyday life.

Benefits:
There will be no direct benefits to you other than receiving your score from this test. It is hoped, however, that through your participation researchers may learn how to improve the assessment you took. Sharing your score with your school, district, or network leaders may be a way to justify further professional development for you.

Confidentiality:
You understand that if you participate in this study, the researchers will use a randomly generated alphanumeric identifier as a way to identify your responses for data analysis. Name and e-mail addresses will also be collected and used for reporting your test scores to you. The research data will be kept in a secure location/on password protected computers and only the research team will have access to the data. At the conclusion of the study, all identifying information will be removed and the data will be kept in the researcher's locked cabinet/office.

Compensation:
There is no monetary compensation for participation in this study.

Participation:
You understand that you do not have to participate in this research project. If you agree to participate, you can skip any question you are not comfortable with, and you can withdraw your participation without penalty.

Questions about the Research:
If you have questions regarding this study, you may contact Debra Fuentes at 801-422-3694, debra.fuentes@byu.net

Questions about Your Rights as Research Participants:
If you have questions regarding your rights as a research participant contact IRB Administrator at (801) 422-1461; A-285 ASB, Brigham Young University, Provo, UT 84602; irb@byu.edu.

CONSENT TO PARTICIPATE

I agree to participate in this study with the understanding that my participation in this study is voluntary. I may withdraw at any time without penalty or refuse to participate entirely without harming my relationship with the researchers or Brigham Young University. I have read and understand the above information.

We invite you to print a copy of this informed consent page for your records.

APPENDIX C

CGI Expert Judgment Interview Protocol

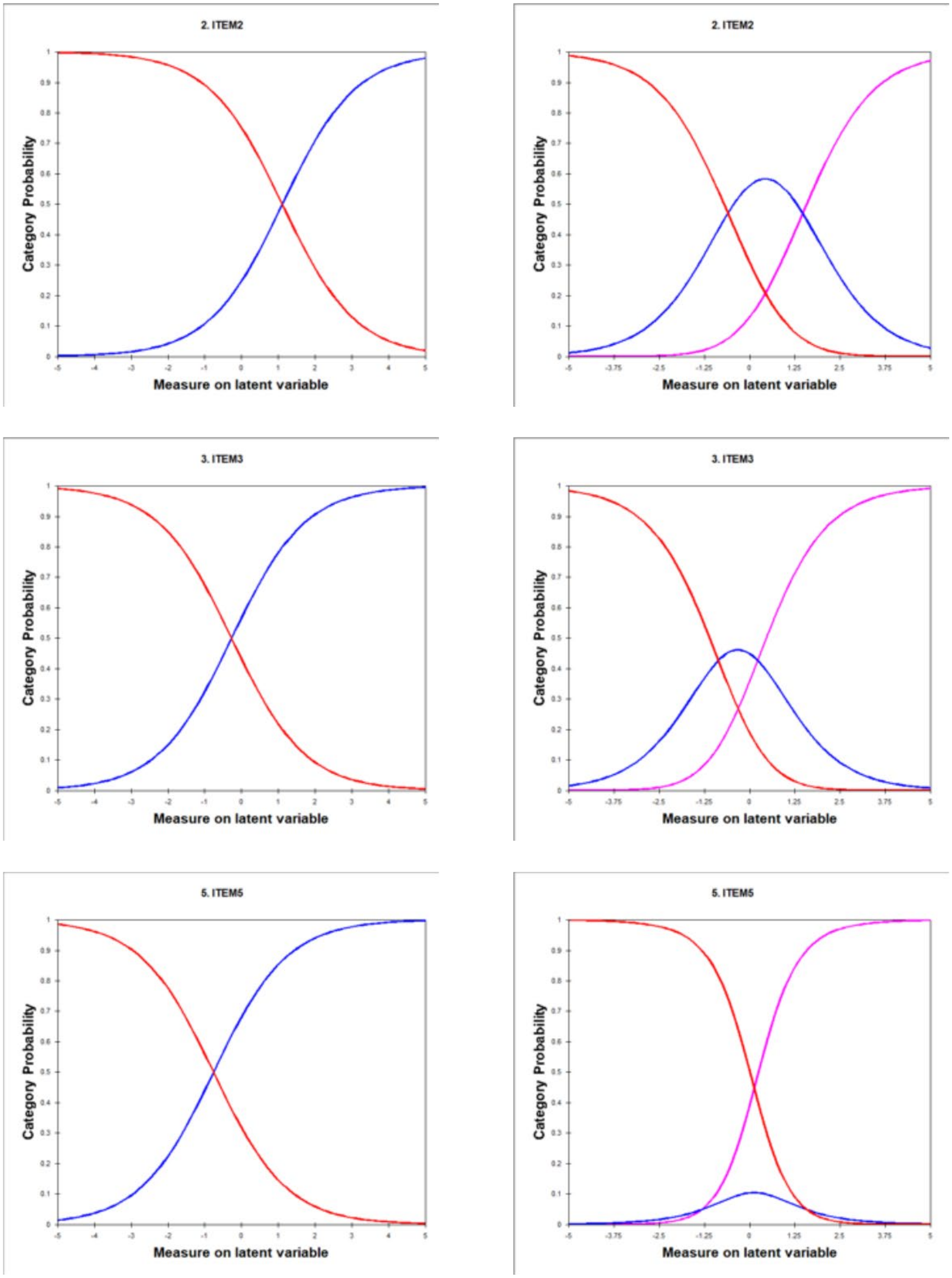
1. Are the constructs originally included in the CGI TKA appropriate?
   a. Of the original six constructs intended to be measured by the CGI TKA, identify the two most important and the two least important to measure.
   b. Which of the six constructs, if any, is not representative of CGI knowledge?
2. Are there other constructs pertinent to CGI that are not currently being measured by the CGI TKA?
   a. What aspects/facets of CGI knowledge, if any, are not included in the original list of constructs?
   b. How would you define any missing facets?
   c. Why is it important to include the missing facet in a test of CGI knowledge? Give reasons.

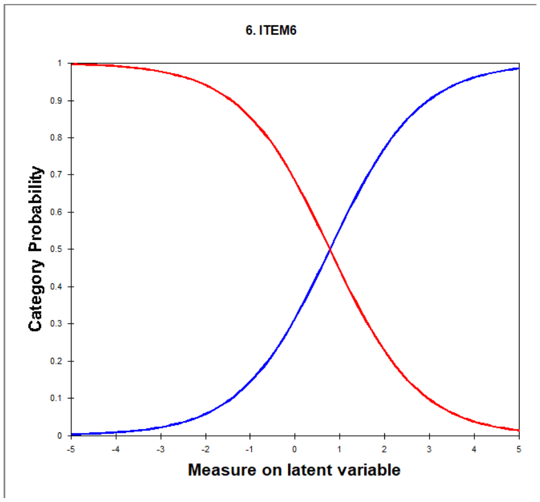
APPENDIX D

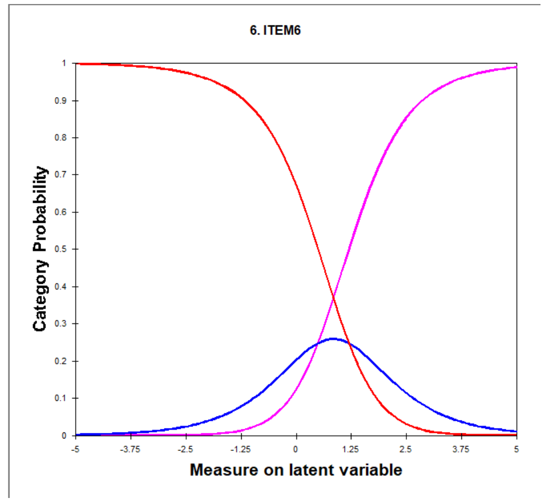
Category Probability Curves Comparing Dichotomous and

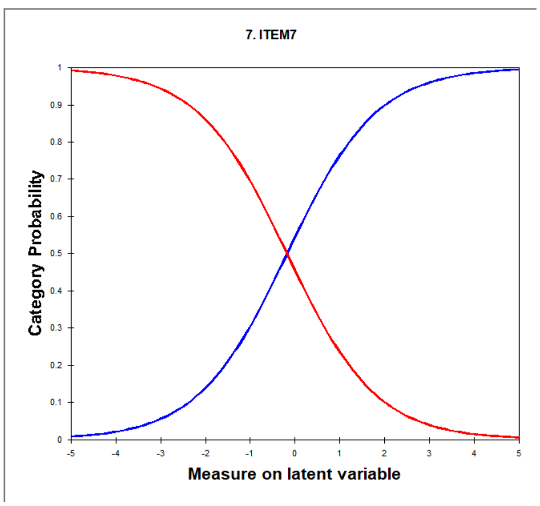
Partial Credit Scoring of the Revised CGI TKA

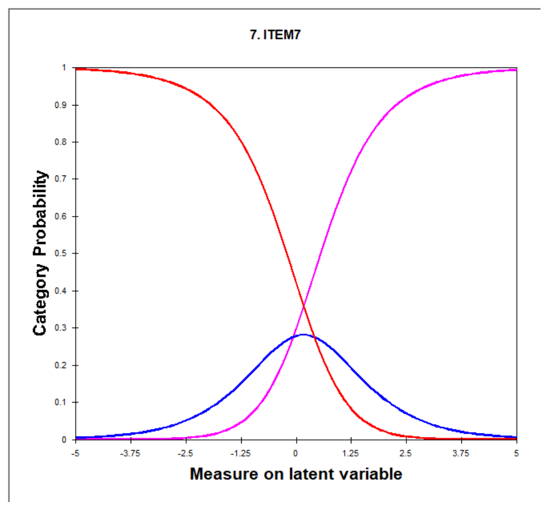
Dichotomous Model **Partial Credit Model**

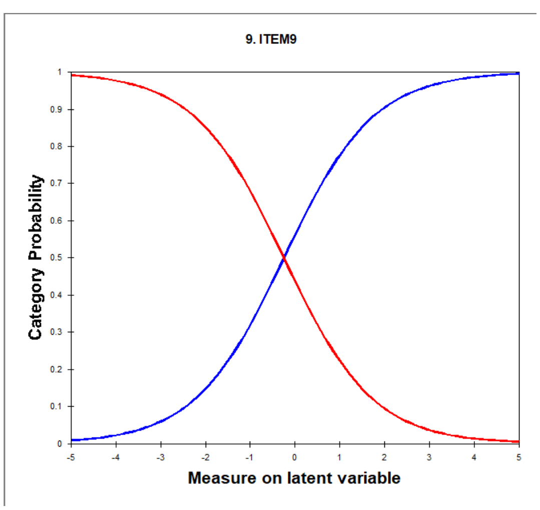


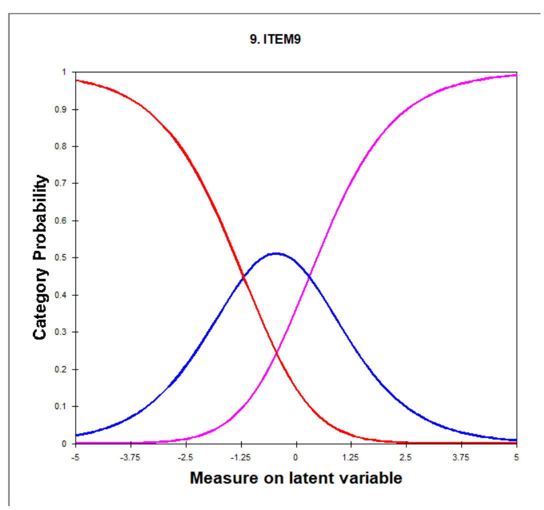

6. ITEM6
Category Probability
Measure on latent variable

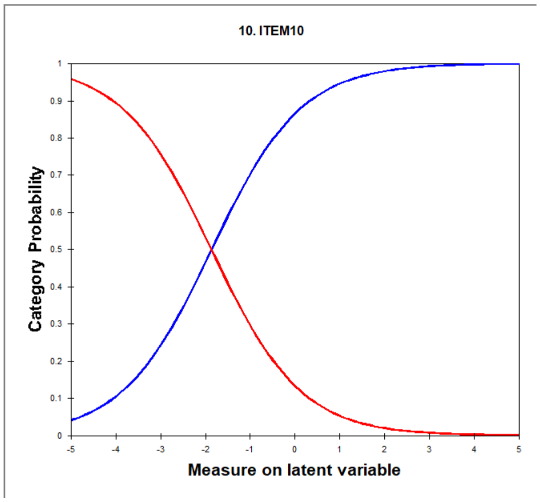

6. ITEM6
Category Probability
Measure on latent variable

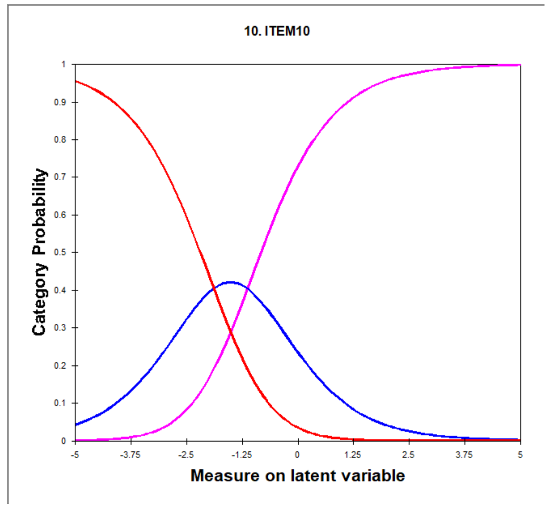

7. ITEM7
Category Probability
Measure on latent variable

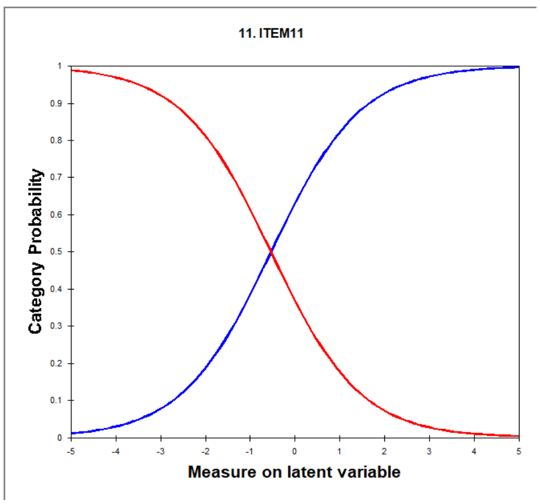

7. ITEM7
Category Probability
Measure on latent variable

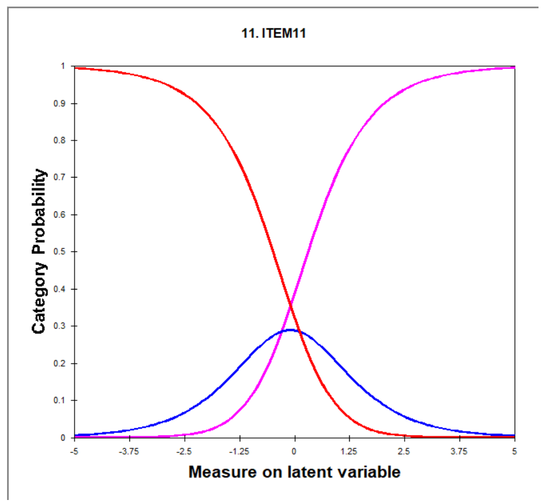

9. ITEM9
Category Probability
Measure on latent variable

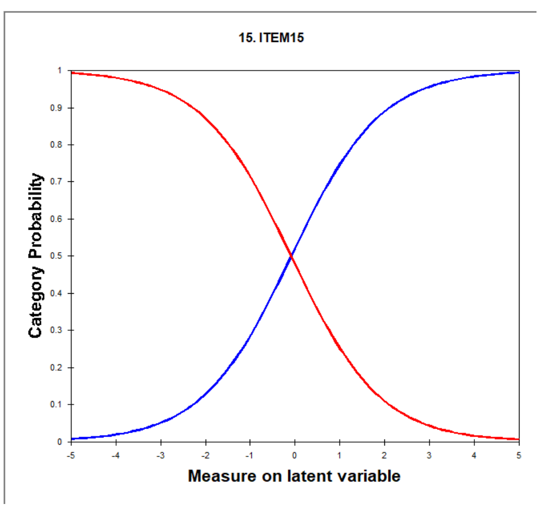

9. ITEM9
Category Probability
Measure on latent variable

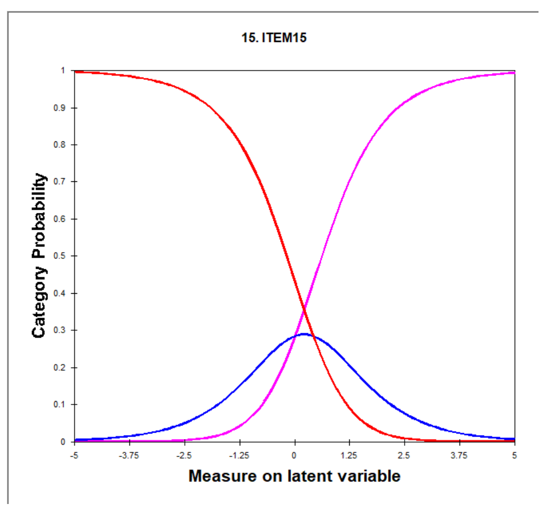

10. ITEM10
Category Probability
Measure on latent variable

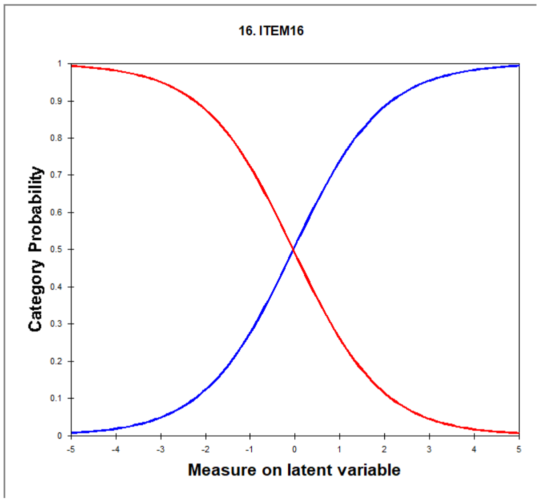

10. ITEM10
Category Probability
Measure on latent variable

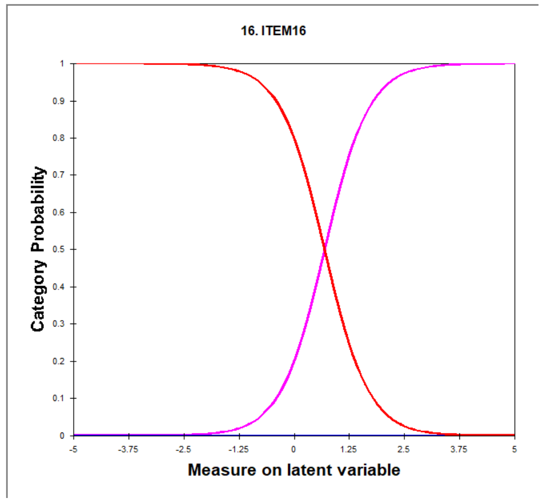

11. ITEM11
Category Probability
Measure on latent variable

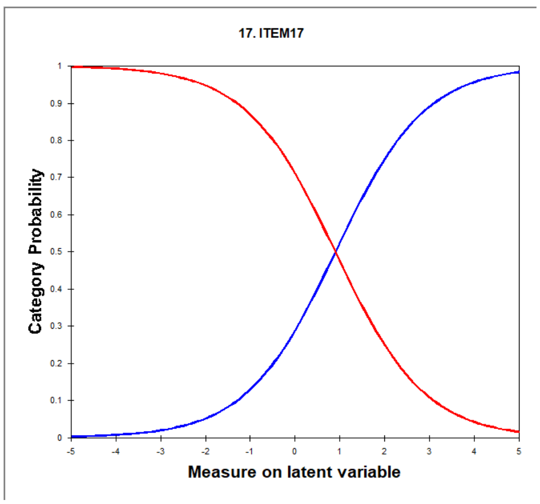

11. ITEM11
Category Probability
Measure on latent variable

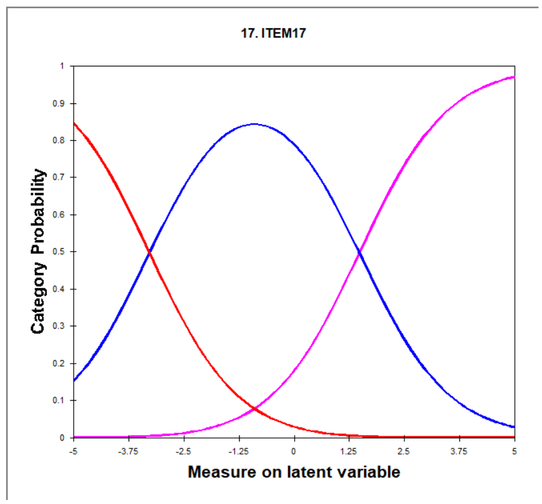

15. ITEM15
Category Probability
Measure on latent variable

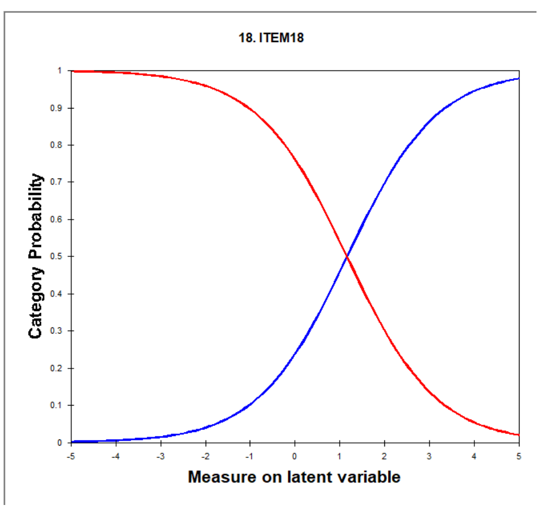

15. ITEM15
Category Probability
Measure on latent variable

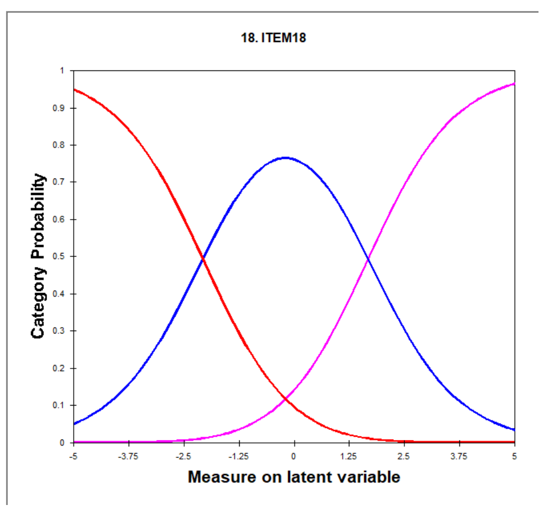

16. ITEM16
Category Probability
Measure on latent variable

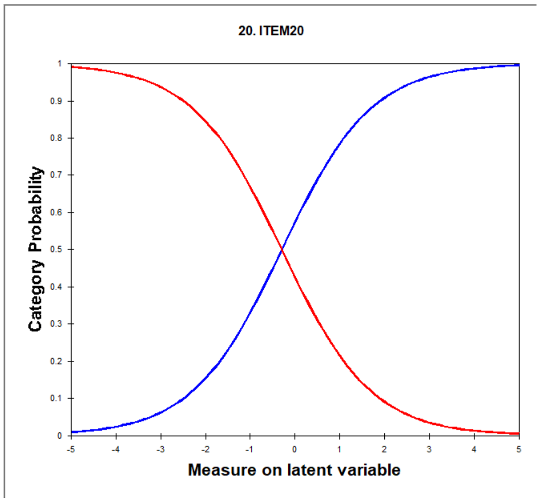

16. ITEM16
Category Probability
Measure on latent variable

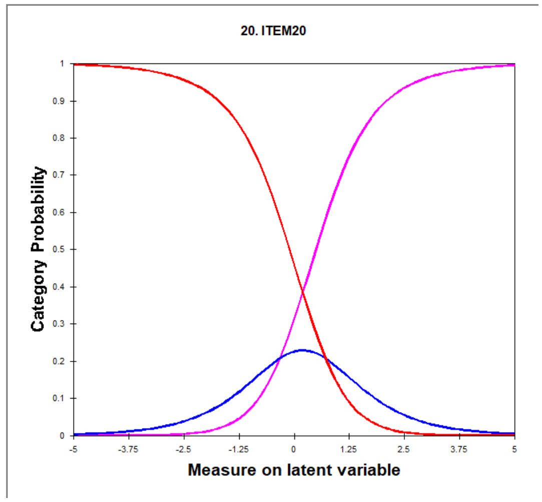

17. ITEM17
Category Probability
Measure on latent variable

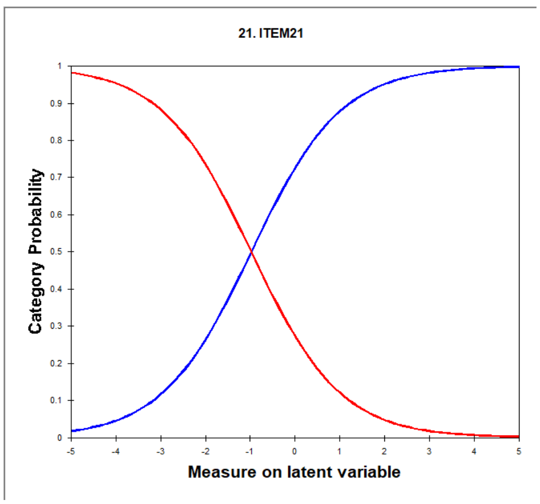

17. ITEM17
Category Probability
Measure on latent variable

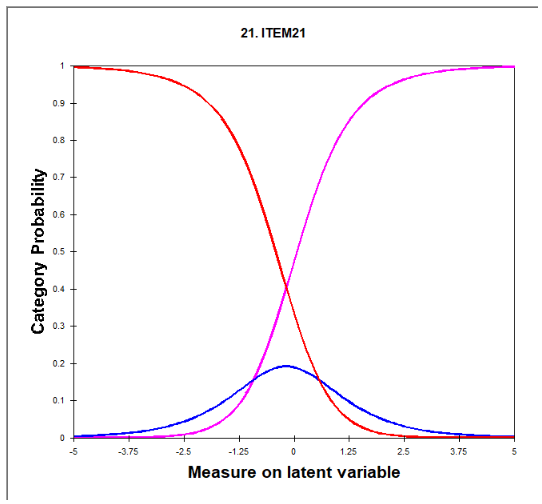

18. ITEM18
Category Probability
Measure on latent variable

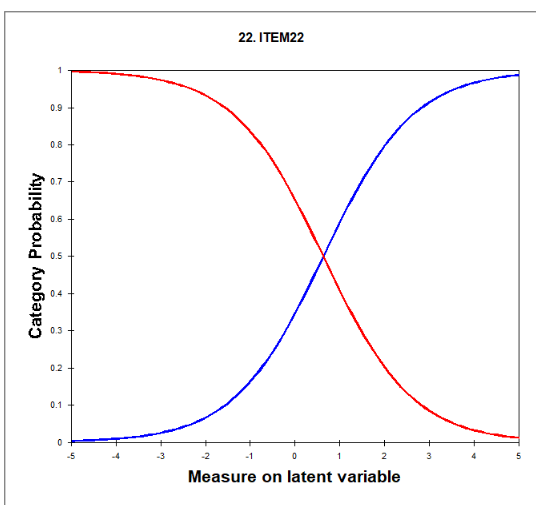

18. ITEM18
Category Probability
Measure on latent variable

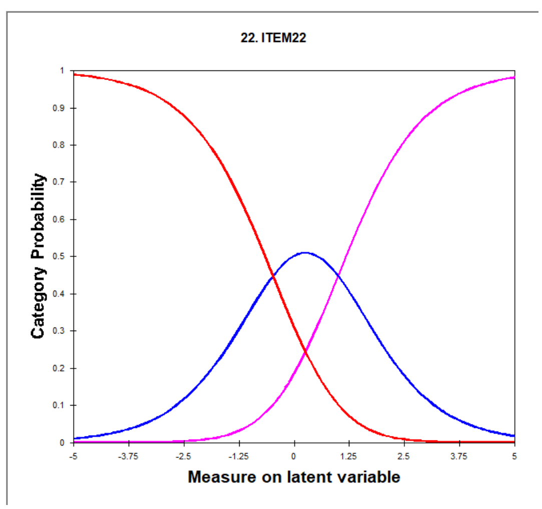

20. ITEM20
Category Probability
Measure on latent variable

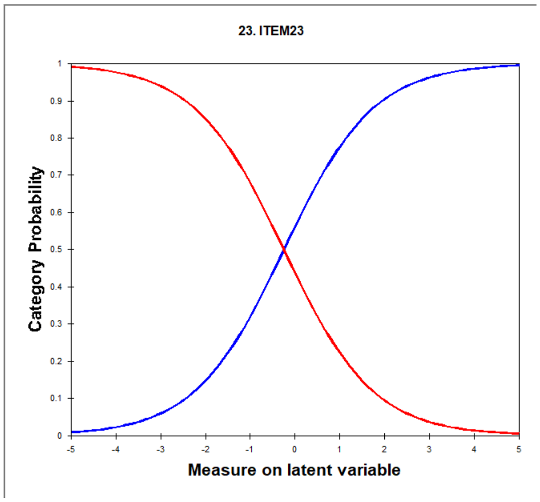

20. ITEM20
Category Probability
Measure on latent variable

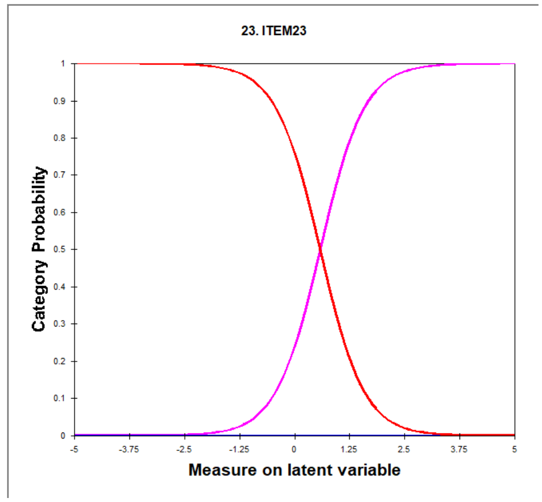

21. ITEM21
Category Probability
Measure on latent variable

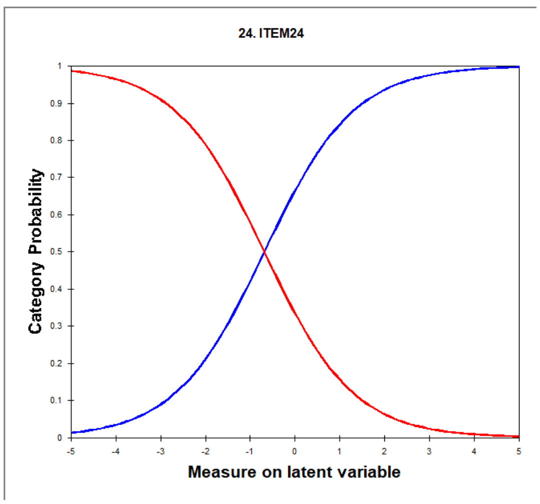

21. ITEM21
Category Probability
Measure on latent variable

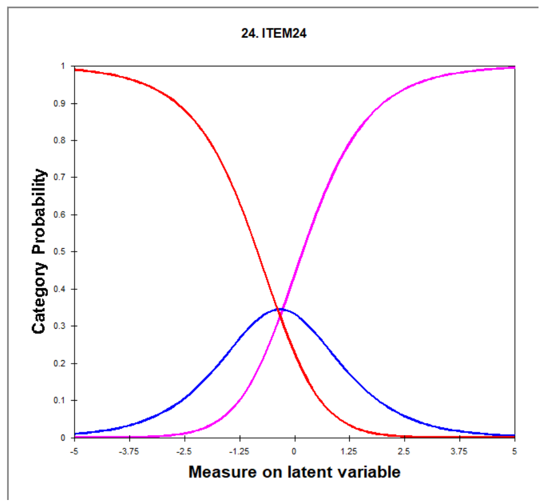

22. ITEM22
Category Probability
Measure on latent variable

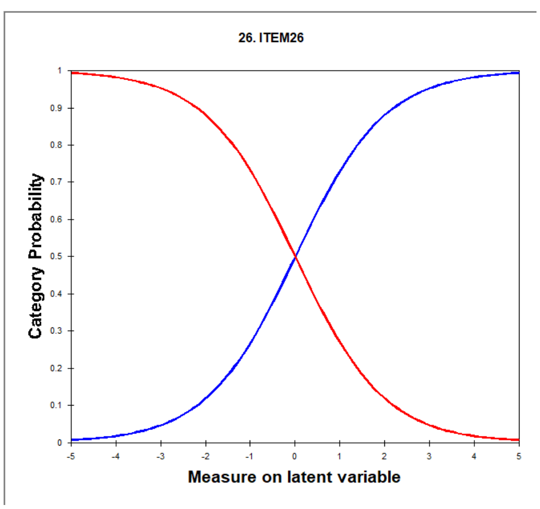

22. ITEM22
Category Probability
Measure on latent variable

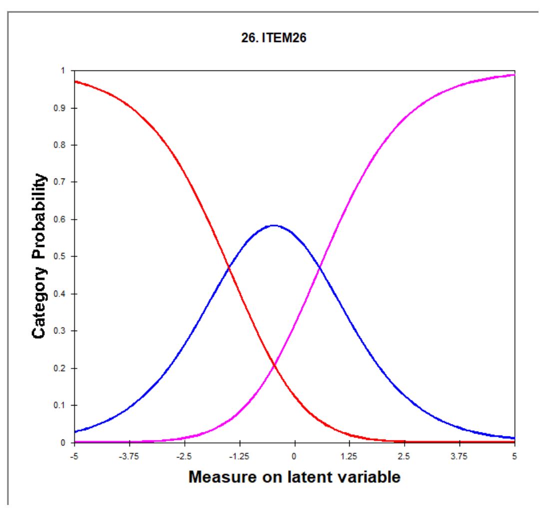

23. ITEM23
Category Probability
Measure on latent variable

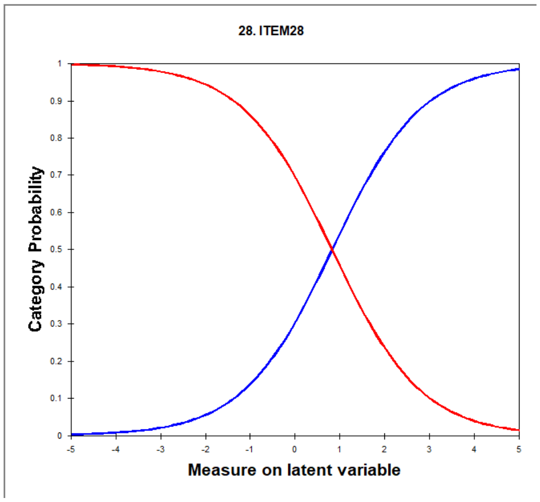

23. ITEM23
Category Probability
Measure on latent variable


24. ITEM24
Category Probability
Measure on latent variable

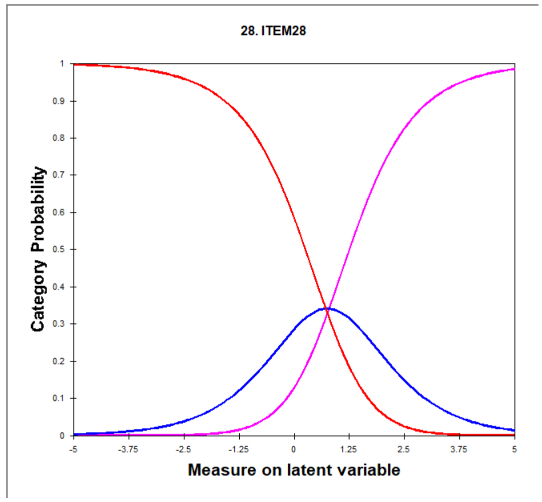

24. ITEM24
Category Probability
Measure on latent variable


26. ITEM26
Category Probability
Measure on latent variable

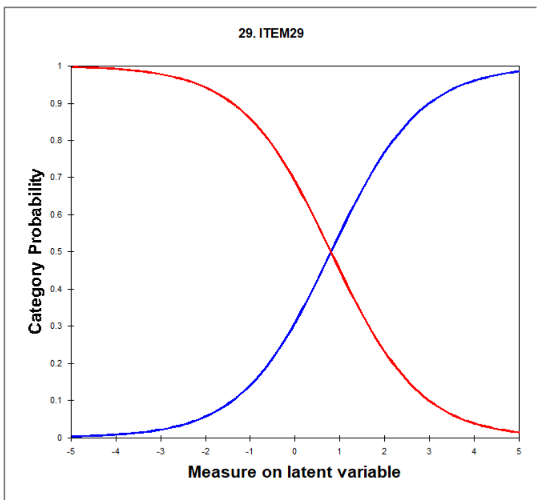

26. ITEM26
Category Probability
Measure on latent variable



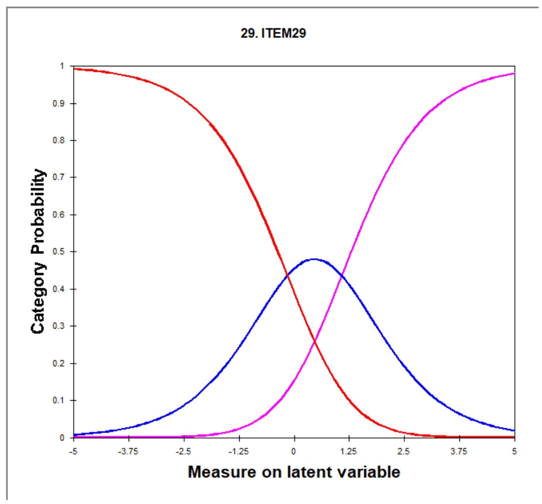
28. ITEM28
Category Probability
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
-5 -4 -3 -2 -1 0 1 2 3 4 5
Measure on latent variable



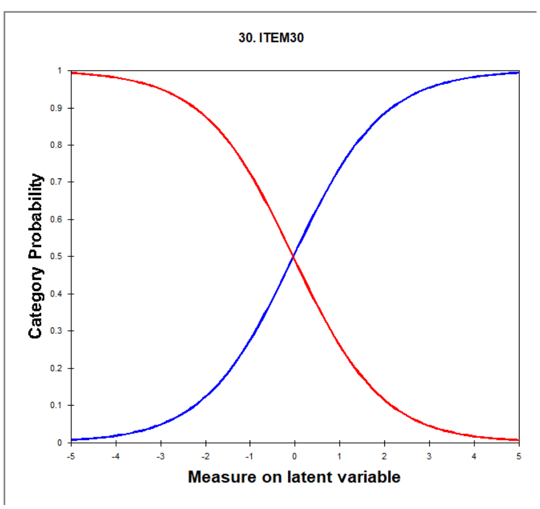
28. ITEM28
Category Probability
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
-5 -3.75 -2.5 -1.25 0 1.25 2.5 3.75 5
Measure on latent variable



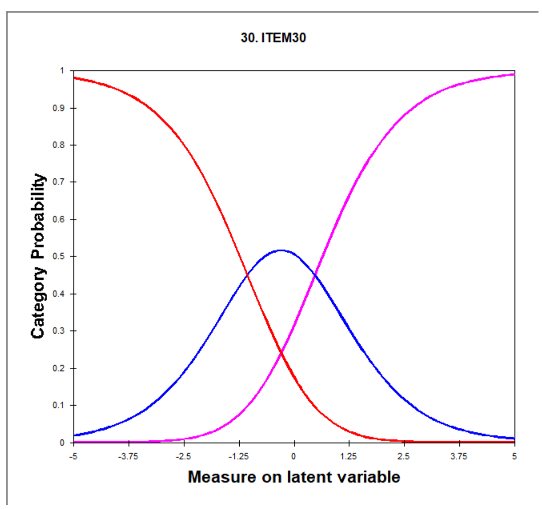
29. ITEM29
Category Probability
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
-5 -4 -3 -2 -1 0 1 2 3 4 5
Measure on latent variable



29. ITEM29
Category Probability
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
-5 -3.75 -2.5 -1.25 0 1.25 2.5 3.75 5
Measure on latent variable



30. ITEM30
Category Probability
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
-5 -4 -3 -2 -1 0 1 2 3 4 5
Measure on latent variable

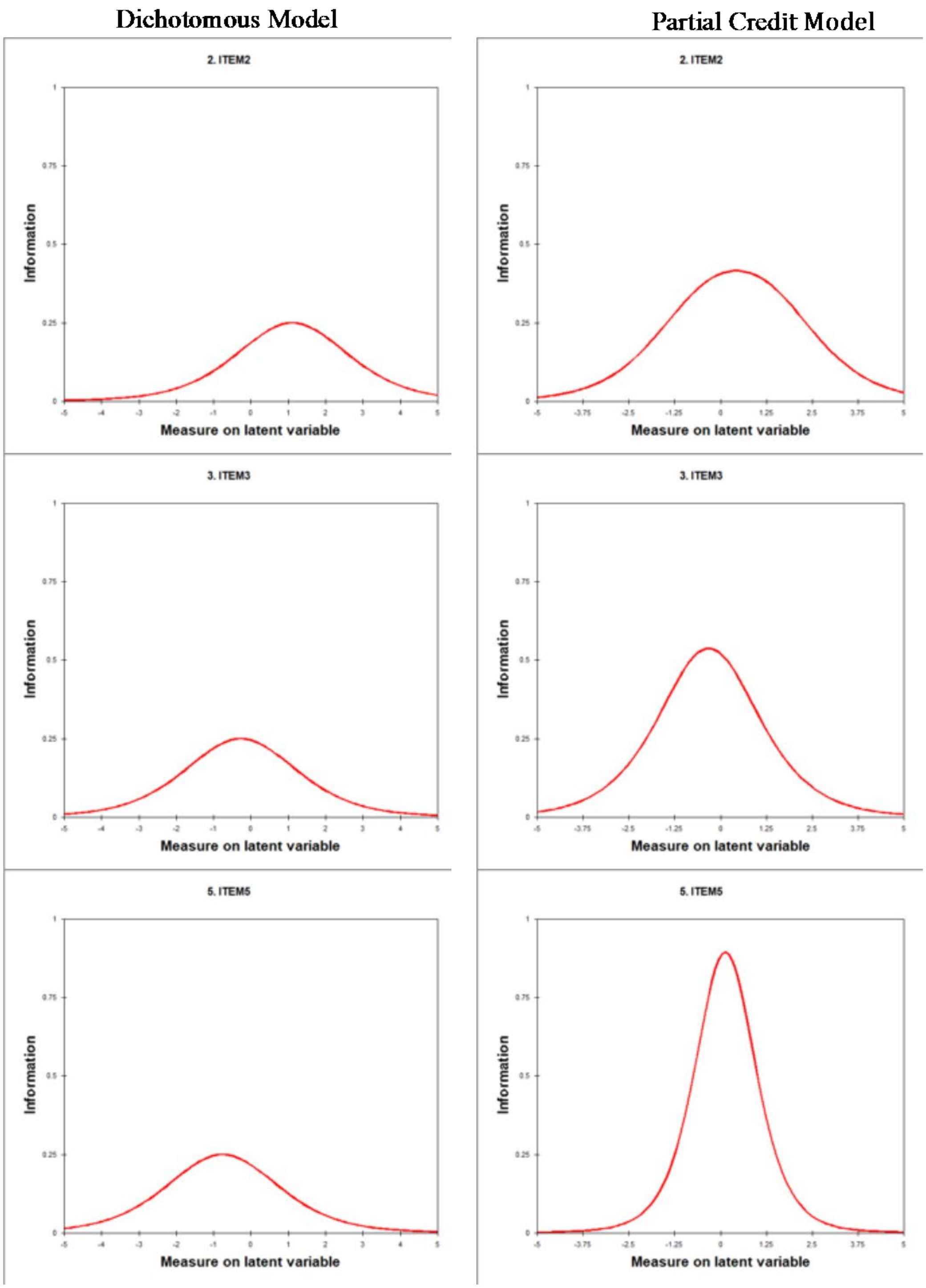


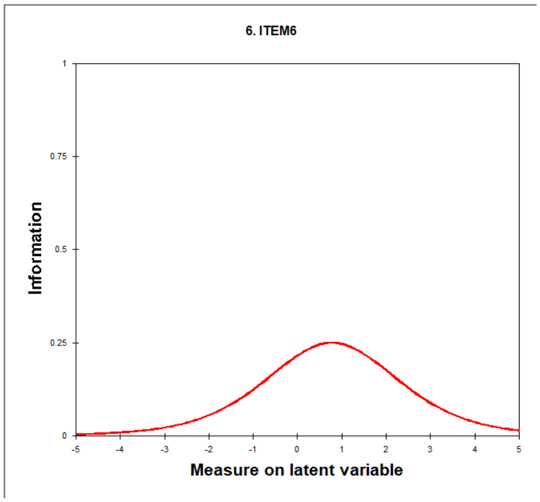
30. ITEM30
Category Probability
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
-5 -3.75 -2.5 -1.25 0 1.25 2.5 3.75 5
Measure on latent variable

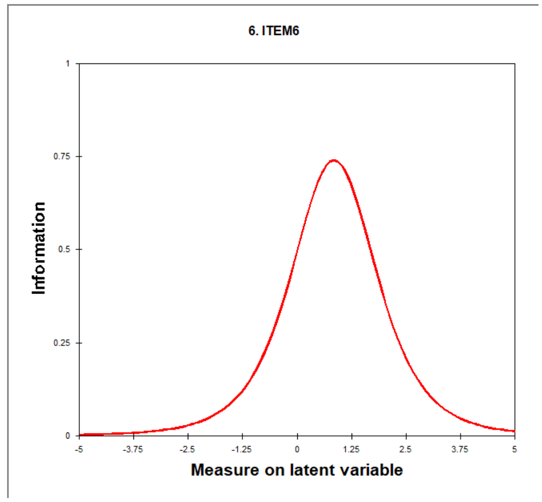
APPENDIX E

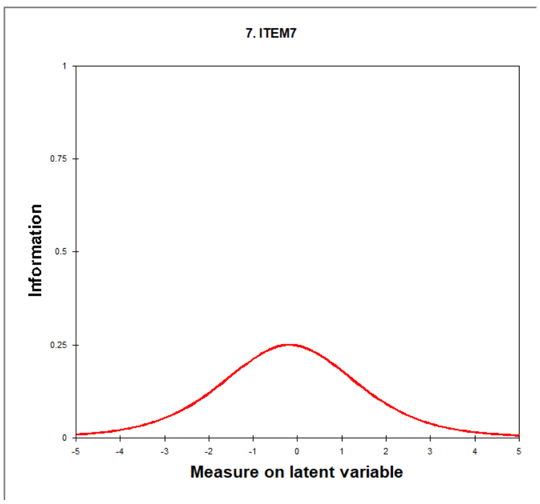
Item Information Functions Comparing Dichotomous and

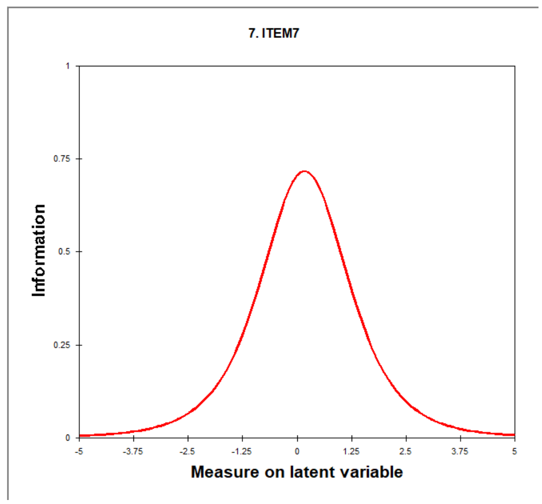
Partial Credit Scoring of the Revised CGI TKA

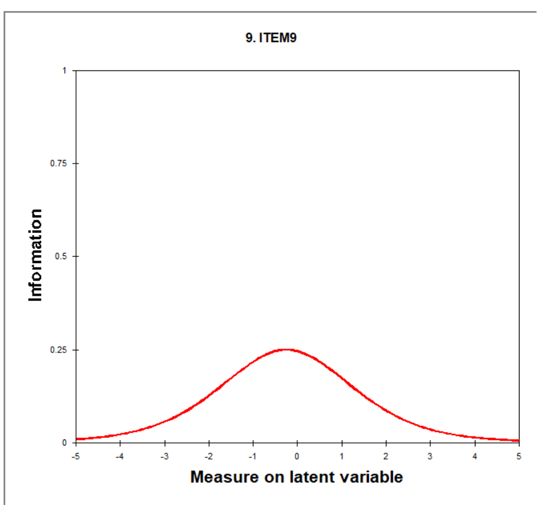

Dichotomous Model
Partial Credit Model
2. ITEM2
Information
Measure on latent variable
2. ITEM2
Information
Measure on latent variable
3. ITEM3
Information
Measure on latent variable
3. ITEM3
Information
Measure on latent variable
5. ITEM5
Information
Measure on latent variable
5. ITEM5
Information
Measure on latent variable

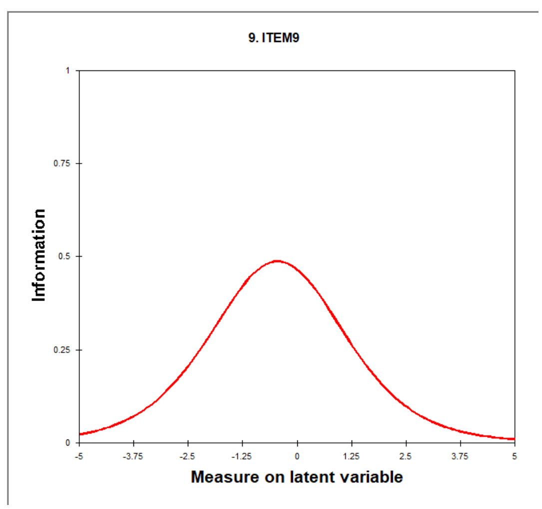

6. ITEM6
Information
1
0.75
0.5
0.25
0
-5 -4 -3 -2 -1 0 1 2 3 4 5
Measure on latent variable

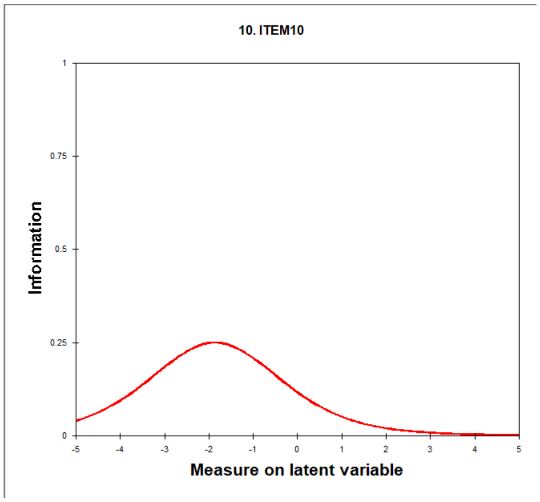

6. ITEM6
Information
1
0.75
0.5
0.25
0
-5 -3.75 -2.5 -1.25 0 1.25 2.5 3.75 5
Measure on latent variable

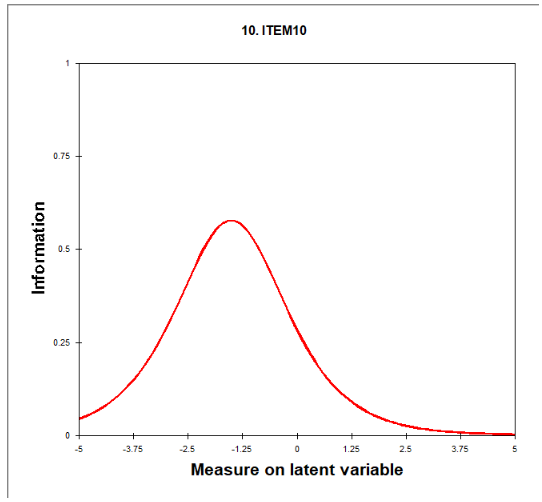

7. ITEM7
Information
1
0.75
0.5
0.25
0
-5 -4 -3 -2 -1 0 1 2 3 4 5
Measure on latent variable

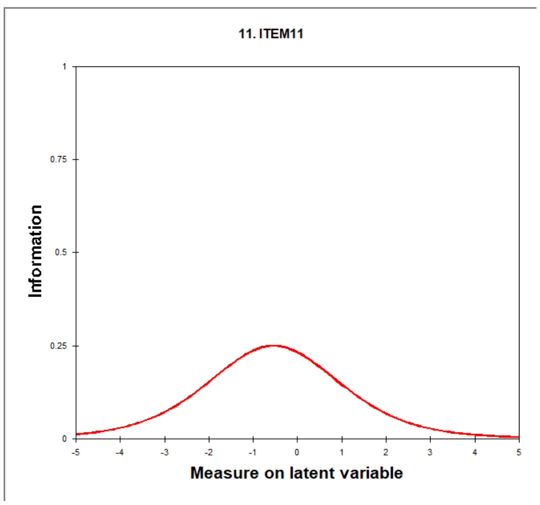

7. ITEM7
Information
1
0.75
0.5
0.25
0
-5 -3.75 -2.5 -1.25 0 1.25 2.5 3.75 5
Measure on latent variable

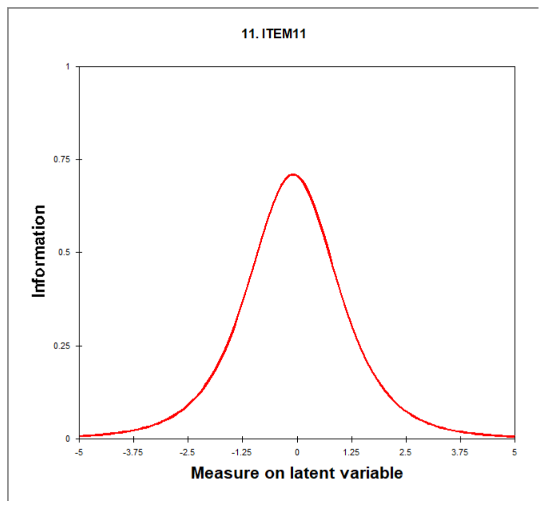

9. ITEM9
Information
1
0.75
0.5
0.25
0
-5 -4 -3 -2 -1 0 1 2 3 4 5
Measure on latent variable

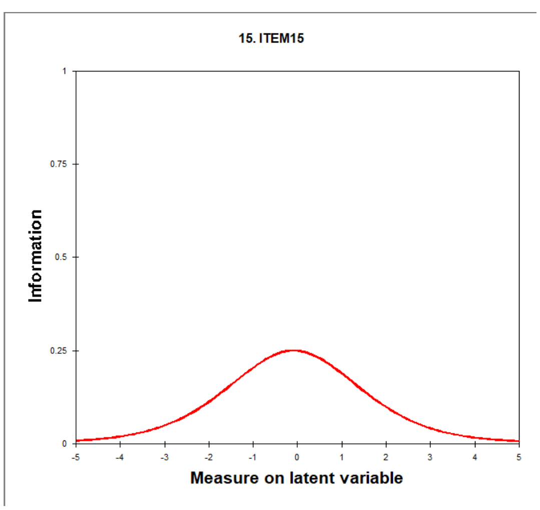

9. ITEM9
Information
1
0.75
0.5
0.25
0
-5 -3.75 -2.5 -1.25 0 1.25 2.5 3.75 5
Measure on latent variable

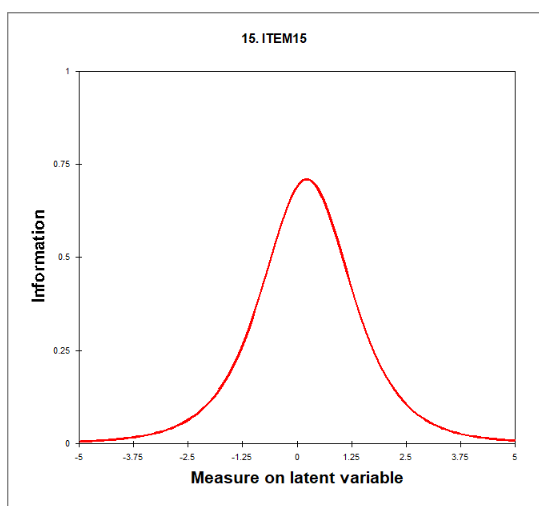

10. ITEM10
Information
1
0.75
0.5
0.25
0
-5 -4 -3 -2 -1 0 1 2 3 4 5
Measure on latent variable

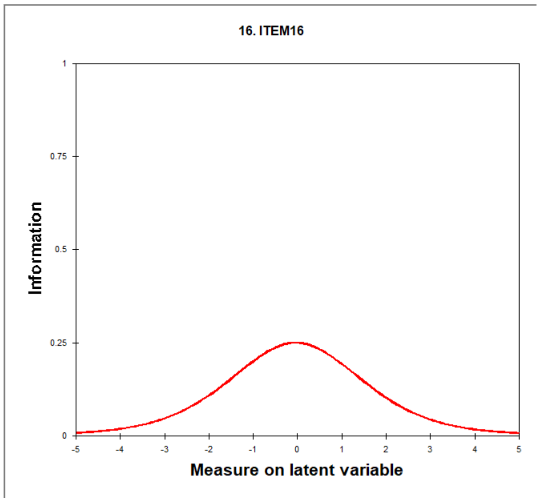

10. ITEM10
Information
1
0.75
0.5
0.25
0
-5 -3.75 -2.5 -1.25 0 1.25 2.5 3.75 5
Measure on latent variable

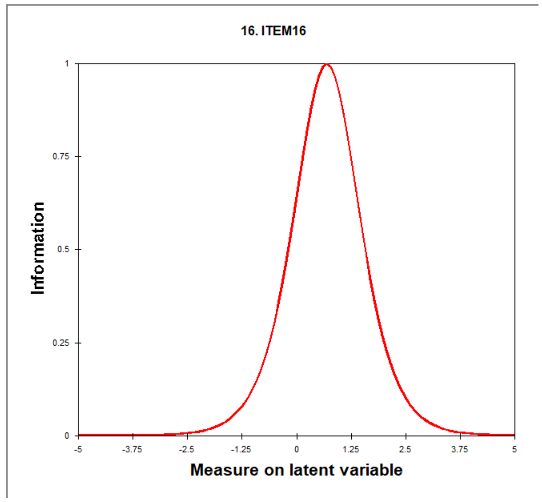

11. ITEM11
Information
1
0.75
0.5
0.25
0
-5 -4 -3 -2 -1 0 1 2 3 4 5
Measure on latent variable

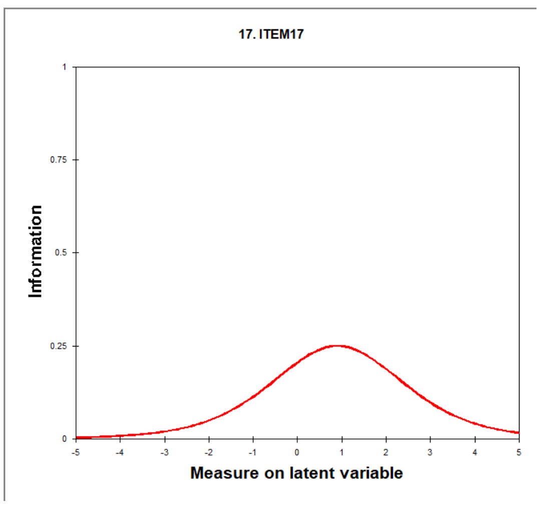

11. ITEM11
Information
1
0.75
0.5
0.25
0
-5 -3.75 -2.5 -1.25 0 1.25 2.5 3.75 5
Measure on latent variable

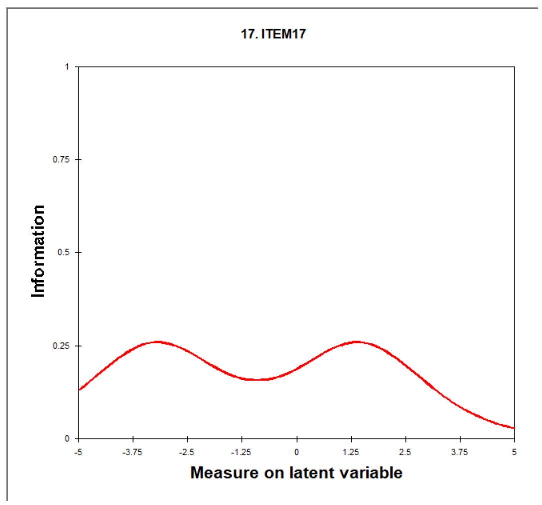

15. ITEM15
Information
1
0.75
0.5
0.25
0
-5 -4 -3 -2 -1 0 1 2 3 4 5
Measure on latent variable

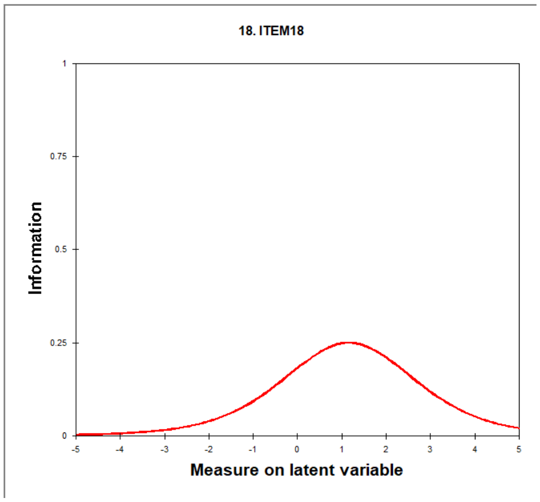

15. ITEM15
Information
1
0.75
0.5
0.25
0
-5 -3.75 -2.5 -1.25 0 1.25 2.5 3.75 5
Measure on latent variable

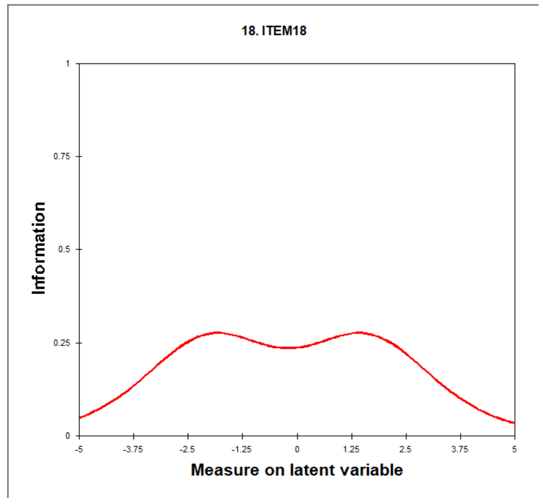

16. ITEM16
Information
Measure on latent variable

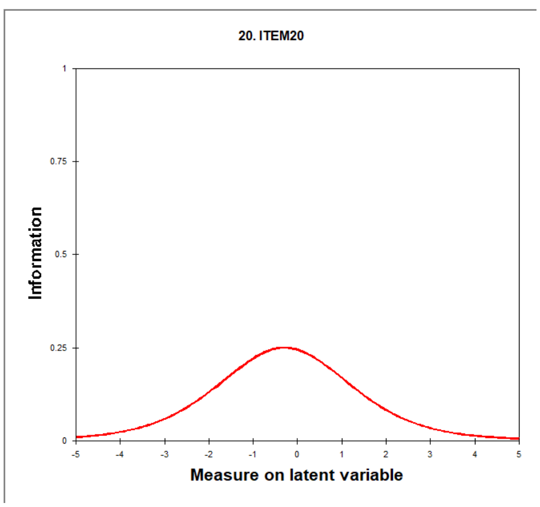

16. ITEM16
Information
Measure on latent variable

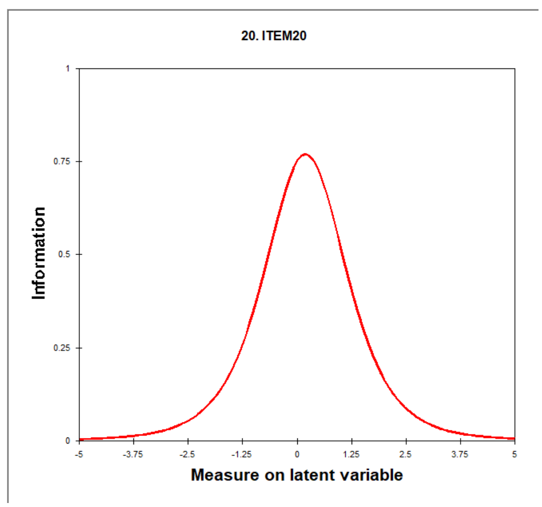

17. ITEM17
Information
Measure on latent variable

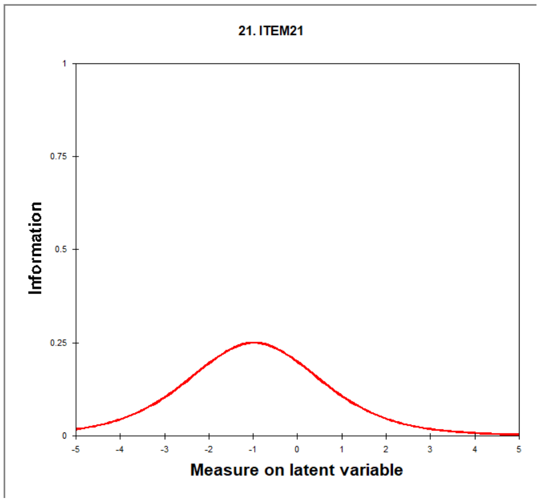

17. ITEM17
Information
Measure on latent variable

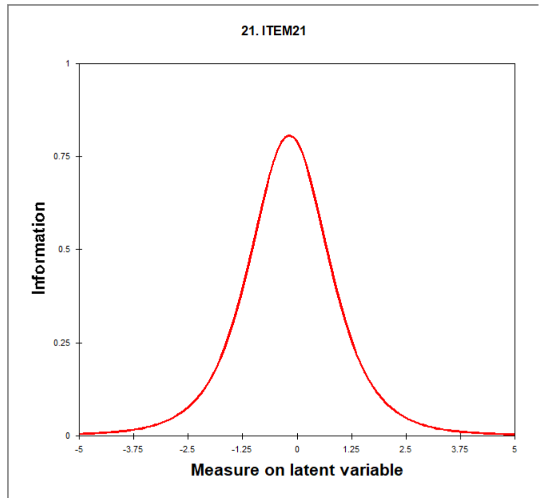

18. ITEM18
Information
Measure on latent variable

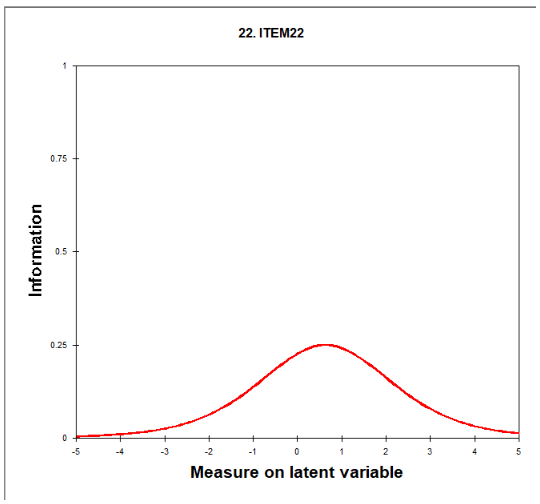

18. ITEM18
Information
Measure on latent variable

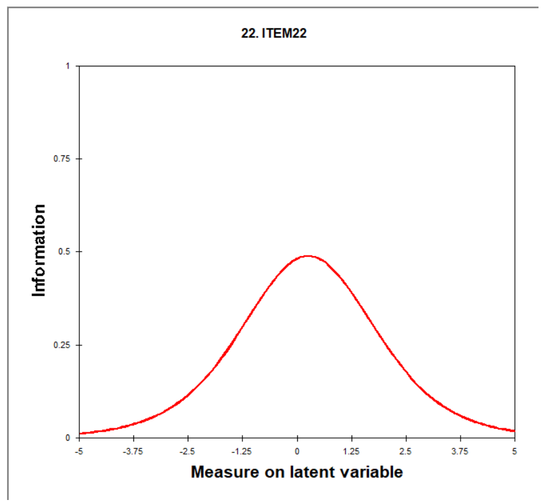

20. ITEM20
Information
Measure on latent variable

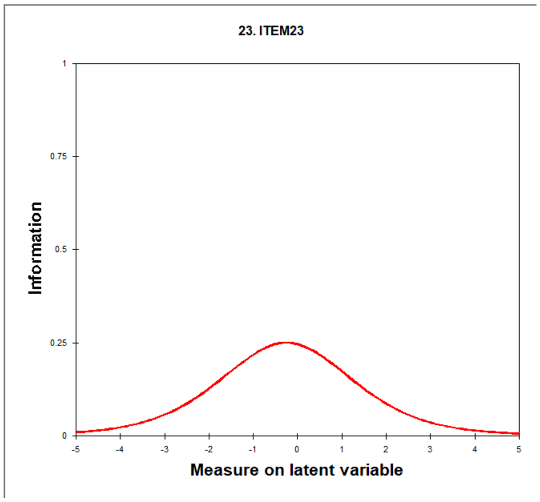

20. ITEM20
Information
Measure on latent variable

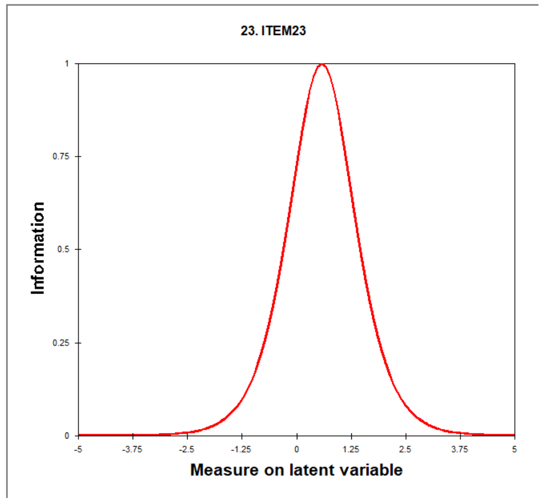

21. ITEM21
Information
Measure on latent variable

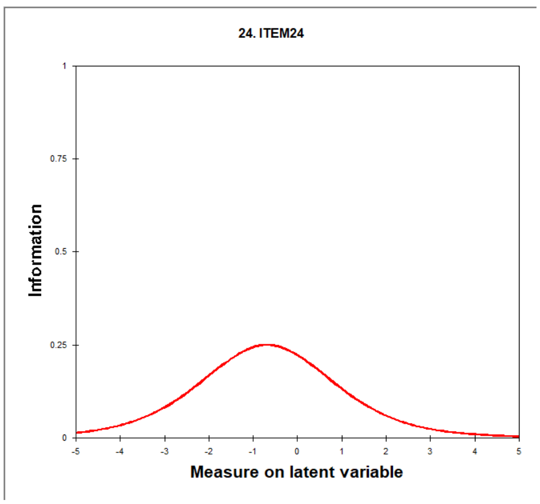

21. ITEM21
Information
Measure on latent variable

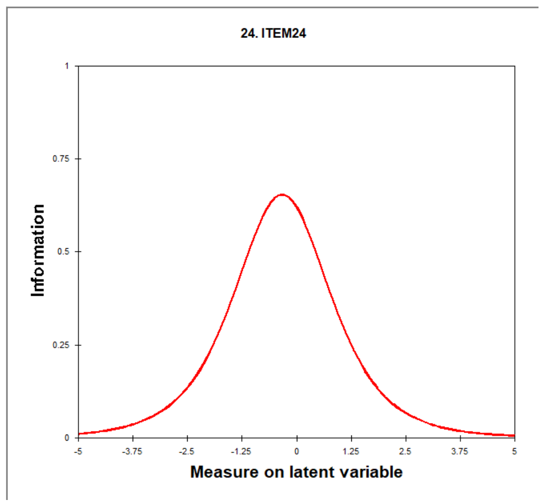

22. ITEM22
Information
Measure on latent variable

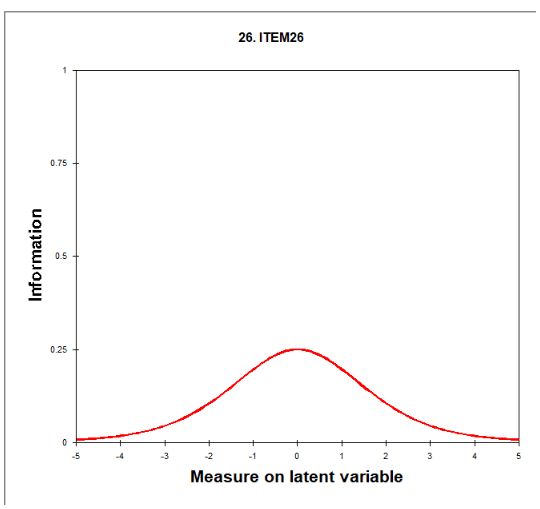

22. ITEM22
Information
Measure on latent variable

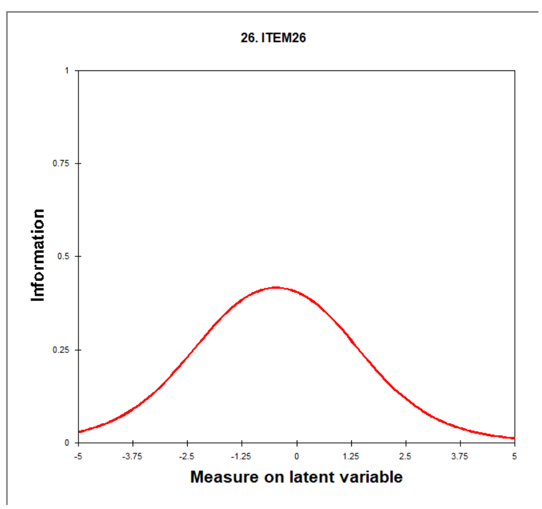

23. ITEM23
Information
1
0.75
0.5
0.25
0
-5 -4 -3 -2 -1 0 1 2 3 4 5
Measure on latent variable

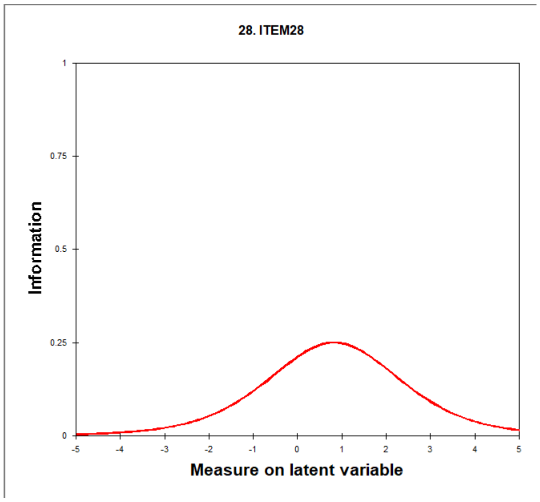

23. ITEM23
Information
1
0.75
0.5
0.25
0
-5 -3.75 -2.5 -1.25 0 1.25 2.5 3.75 5
Measure on latent variable

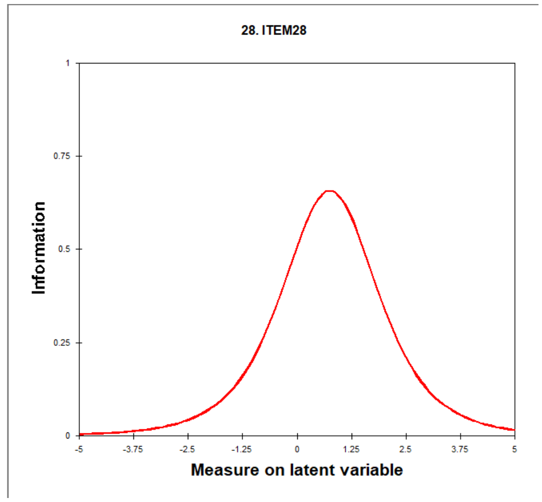

24. ITEM24
Information
1
0.75
0.5
0.25
0
-5 -4 -3 -2 -1 0 1 2 3 4 5
Measure on latent variable

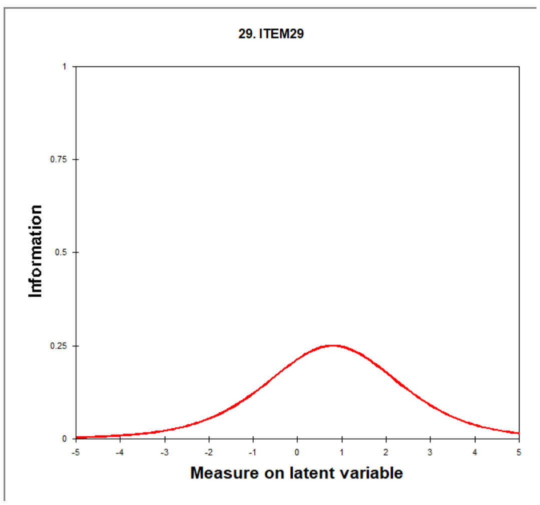

24. ITEM24
Information
1
0.75
0.5
0.25
0
-5 -3.75 -2.5 -1.25 0 1.25 2.5 3.75 5
Measure on latent variable

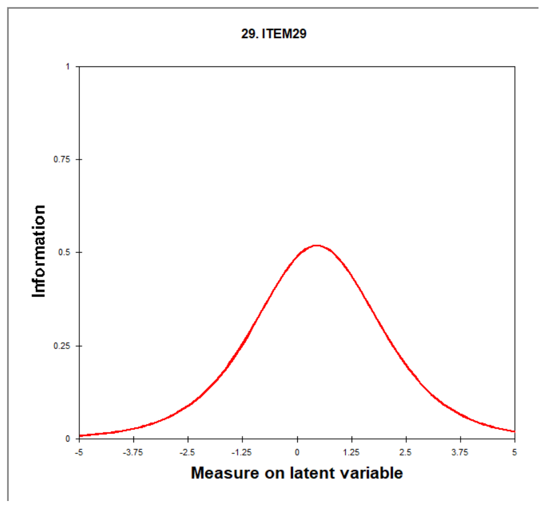

26. ITEM26
Information
1
0.75
0.5
0.25
0
-5 -4 -3 -2 -1 0 1 2 3 4 5
Measure on latent variable

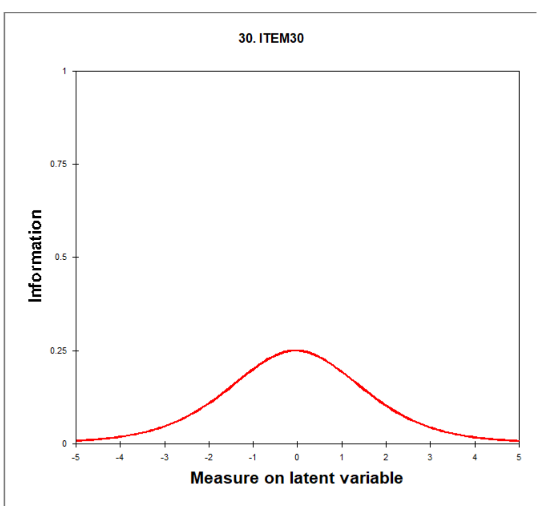

26. ITEM26
Information
1
0.75
0.5
0.25
0
-5 -3.75 -2.5 -1.25 0 1.25 2.5 3.75 5
Measure on latent variable

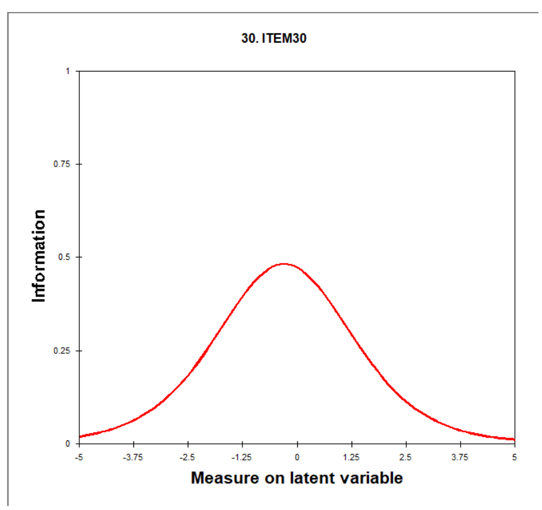

28. ITEM28
Information
1
0.75
0.5
0.25
0
-5 -4 -3 -2 -1 0 1 2 3 4 5
Measure on latent variable


28. ITEM28
Information
1
0.75
0.5
0.25
0
-5 -3.75 -2.5 -1.25 0 1.25 2.5 3.75 5
Measure on latent variable


29. ITEM29
Information
1
0.75
0.5
0.25
0
-5 -4 -3 -2 -1 0 1 2 3 4 5
Measure on latent variable


29. ITEM29
Information
1
0.75
0.5
0.25
0
-5 -3.75 -2.5 -1.25 0 1.25 2.5 3.75 5
Measure on latent variable


30. ITEM30
Information
1
0.75
0.5
0.25
0
-5 -4 -3 -2 -1 0 1 2 3 4 5
Measure on latent variable

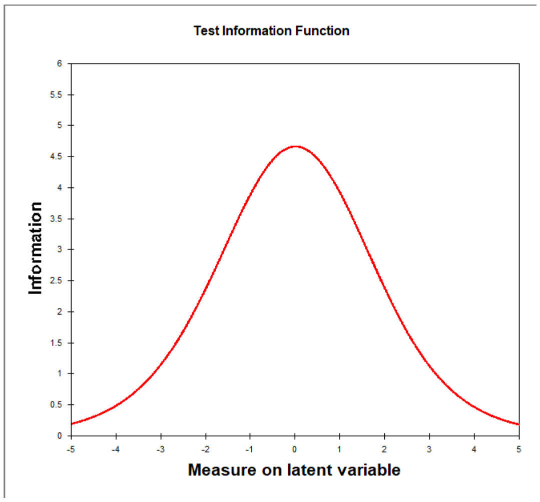

30. ITEM30
Information
1
0.75
0.5
0.25
0
-5 -3.75 -2.5 -1.25 0 1.25 2.5 3.75 5
Measure on latent variable

APPENDIX F

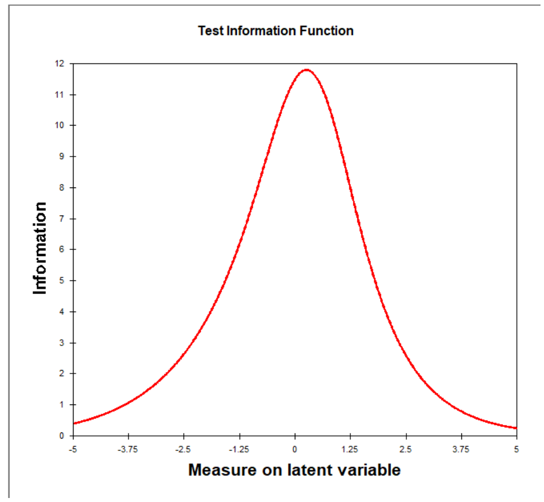
Test Information Functions Comparing Dichotomous and

Partial Credit Scoring of the Revised CGI TKA

Dichotomous Model



Partial Credit Model



APPENDIX G

Z-Score Difference Table of Person Ability Estimates for the Dichotomous and Partial Credit Scoring of the 21-Item CGI TKA with 10 Partial Credit Items

Dichotomous Person Ability	Dichotomous Z-Score	PCM Person Ability	PCM Z-Score	Z-Score Difference
51.80	-0.7122	49.00	-0.2865	-0.4257
54.70	-0.2887	53.31	0.1189	-0.4076
54.70	-0.2887	53.31	0.1189	-0.4076
68.42	1.7146	74.59	2.1201	-0.4055
59.41	0.3990	60.31	0.7772	-0.3781
61.30	0.6750	63.01	1.0311	-0.3561
61.30	0.6750	63.01	1.0311	-0.3561
61.30	0.6750	63.01	1.0311	-0.3561
51.05	-0.8217	46.83	-0.4905	-0.3311
52.53	-0.6056	49.00	-0.2865	-0.3191
63.56	1.0050	66.08	1.3198	-0.3148
63.56	1.0050	66.08	1.3198	-0.3148
63.56	1.0050	66.08	1.3198	-0.3148
63.56	1.0050	66.08	1.3198	-0.3148
55.43	-0.1821	53.31	0.1189	-0.3010
55.43	-0.1821	53.31	0.1189	-0.3010
44.96	-1.7109	37.05	-1.4103	-0.3006
56.94	0.0384	55.52	0.3267	-0.2883
58.55	0.2735	57.84	0.5449	-0.2714
48.70	-1.1648	42.28	-0.9184	-0.2464
48.70	-1.1648	42.28	-0.9184	-0.2464
60.32	0.5319	60.31	0.7772	-0.2453
60.32	0.5319	60.31	0.7772	-0.2453
60.32	0.5319	60.31	0.7772	-0.2453
60.32	0.5319	60.31	0.7772	-0.2453
60.32	0.5319	60.31	0.7772	-0.2453
66.49	1.4328	69.77	1.6668	-0.2340
66.49	1.4328	69.77	1.6668	-0.2340
66.49	1.4328	69.77	1.6668	-0.2340
66.49	1.4328	69.77	1.6668	-0.2340
66.49	1.4328	69.77	1.6668	-0.2340
50.29	-0.9326	44.61	-0.6993	-0.2333
50.29	-0.9326	44.61	-0.6993	-0.2333
51.80	-0.7122	46.83	-0.4905	-0.2216
53.25	-0.5004	49.00	-0.2865	-0.2140
54.70	-0.2887	51.15	-0.0843	-0.2044
54.70	-0.2887	51.15	-0.0843	-0.2044
54.70	-0.2887	51.15	-0.0843	-0.2044
62.37	0.8312	63.01	1.0311	-0.1999

(Table continues)

Dichotomous Person Ability	Dichotomous Z-Score	PCM Person Ability	PCM Z-Score	Z-Score Difference
62.37	0.8312	63.01	1.0311	-0.1999
56.18	-0.0726	53.31	0.1189	-0.1915
56.18	-0.0726	53.31	0.1189	-0.1915
56.18	-0.0726	53.31	0.1189	-0.1915
57.73	0.1537	55.52	0.3267	-0.1730
57.73	0.1537	55.52	0.3267	-0.1730
57.73	0.1537	55.52	0.3267	-0.1730
57.73	0.1537	55.52	0.3267	-0.1730
57.73	0.1537	55.52	0.3267	-0.1730
57.73	0.1537	55.52	0.3267	-0.1730
59.41	0.3990	57.84	0.5449	-0.1459
59.41	0.3990	57.84	0.5449	-0.1459
59.41	0.3990	57.84	0.5449	-0.1459
59.41	0.3990	57.84	0.5449	-0.1459
59.41	0.3990	57.84	0.5449	-0.1459
47.84	-1.2904	39.79	-1.1526	-0.1378
47.84	-1.2904	39.79	-1.1526	-0.1378
49.51	-1.0465	42.28	-0.9184	-0.1281
64.91	1.2021	66.08	1.3198	-0.1177
64.91	1.2021	66.08	1.3198	-0.1177
64.91	1.2021	66.08	1.3198	-0.1177
52.53	-0.6056	46.83	-0.4905	-0.1150
52.53	-0.6056	46.83	-0.4905	-0.1150
52.53	-0.6056	46.83	-0.4905	-0.1150
53.98	-0.3938	49.00	-0.2865	-0.1074
53.98	-0.3938	49.00	-0.2865	-0.1074
61.30	0.6750	60.31	0.7772	-0.1022
61.30	0.6750	60.31	0.7772	-0.1022
61.30	0.6750	60.31	0.7772	-0.1022
55.43	-0.1821	51.15	-0.0843	-0.0978
55.43	-0.1821	51.15	-0.0843	-0.0978
55.43	-0.1821	51.15	-0.0843	-0.0978
55.43	-0.1821	51.15	-0.0843	-0.0978
56.94	0.0384	53.31	0.1189	-0.0805
56.94	0.0384	53.31	0.1189	-0.0805
56.94	0.0384	53.31	0.1189	-0.0805
56.94	0.0384	53.31	0.1189	-0.0805
56.94	0.0384	53.31	0.1189	-0.0805
56.94	0.0384	53.31	0.1189	-0.0805
58.55	0.2735	55.52	0.3267	-0.0532

(Table Continues)

Dichotomous Person Ability	Dichotomous Z-Score	PCM Person Ability	PCM Z-Score	Z-Score Difference
58.55	0.2735	55.52	0.3267	-0.0532
58.55	0.2735	55.52	0.3267	-0.0532
58.55	0.2735	55.52	0.3267	-0.0532
70.94	2.0826	74.59	2.1201	-0.0375
70.94	2.0826	74.59	2.1201	-0.0375
70.94	2.0826	74.59	2.1201	-0.0375
63.56	1.0050	63.01	1.0311	-0.0261
63.56	1.0050	63.01	1.0311	-0.0261
50.29	-0.9326	42.28	-0.9184	-0.0142
50.29	-0.9326	42.28	-0.9184	-0.0142
60.32	0.5319	57.84	0.5449	-0.0130
60.32	0.5319	57.84	0.5449	-0.0130
60.32	0.5319	57.84	0.5449	-0.0130
60.32	0.5319	57.84	0.5449	-0.0130
51.80	-0.7122	44.61	-0.6993	-0.0128
51.80	-0.7122	44.61	-0.6993	-0.0128
51.80	-0.7122	44.61	-0.6993	-0.0128
51.80	-0.7122	44.61	-0.6993	-0.0128
51.80	-0.7122	44.61	-0.6993	-0.0128
48.70	-1.1648	39.79	-1.1526	-0.0122
48.70	-1.1648	39.79	-1.1526	-0.0122
46.95	-1.4203	37.05	-1.4103	-0.0100
46.95	-1.4203	37.05	-1.4103	-0.0100
46.95	-1.4203	37.05	-1.4103	-0.0100
46.95	-1.4203	37.05	-1.4103	-0.0100
53.25	-0.5004	46.83	-0.4905	-0.0099
53.25	-0.5004	46.83	-0.4905	-0.0099
53.25	-0.5004	46.83	-0.4905	-0.0099
53.25	-0.5004	46.83	-0.4905	-0.0099
53.25	-0.5004	46.83	-0.4905	-0.0099
53.25	-0.5004	46.83	-0.4905	-0.0099
53.25	-0.5004	46.83	-0.4905	-0.0099
53.25	-0.5004	46.83	-0.4905	-0.0099
53.25	-0.5004	46.83	-0.4905	-0.0099
53.25	-0.5004	46.83	-0.4905	-0.0099
44.96	-1.7109	33.93	-1.7037	-0.0072
44.96	-1.7109	33.93	-1.7037	-0.0072
54.70	-0.2887	49.00	-0.2865	-0.0022
56.18	-0.0726	51.15	-0.0843	0.0117
56.18	-0.0726	51.15	-0.0843	0.0117

(Table Continues)

Dichotomous Person Ability	Dichotomous Z-Score	PCM Person Ability	PCM Z-Score	Z-Score Difference
56.18	-0.0726	51.15	-0.0843	0.0117
56.18	-0.0726	51.15	-0.0843	0.0117
56.18	-0.0726	51.15	-0.0843	0.0117
56.18	-0.0726	51.15	-0.0843	0.0117
56.18	-0.0726	51.15	-0.0843	0.0117
56.18	-0.0726	51.15	-0.0843	0.0117
57.73	0.1537	53.31	0.1189	0.0349
57.73	0.1537	53.31	0.1189	0.0349
57.73	0.1537	53.31	0.1189	0.0349
57.73	0.1537	53.31	0.1189	0.0349
57.73	0.1537	53.31	0.1189	0.0349
57.73	0.1537	53.31	0.1189	0.0349
68.42	1.7146	69.77	1.6668	0.0478
68.42	1.7146	69.77	1.6668	0.0478
62.37	0.8312	60.31	0.7772	0.0541
62.37	0.8312	60.31	0.7772	0.0541
62.37	0.8312	60.31	0.7772	0.0541
62.37	0.8312	60.31	0.7772	0.0541
62.37	0.8312	60.31	0.7772	0.0541
62.37	0.8312	60.31	0.7772	0.0541
59.41	0.3990	55.52	0.3267	0.0723
59.41	0.3990	55.52	0.3267	0.0723
59.41	0.3990	55.52	0.3267	0.0723
59.41	0.3990	55.52	0.3267	0.0723
59.41	0.3990	55.52	0.3267	0.0723
52.53	-0.6056	44.61	-0.6993	0.0938
52.53	-0.6056	44.61	-0.6993	0.0938
52.53	-0.6056	44.61	-0.6993	0.0938
52.53	-0.6056	44.61	-0.6993	0.0938
52.53	-0.6056	44.61	-0.6993	0.0938
52.53	-0.6056	44.61	-0.6993	0.0938
51.05	-0.8217	42.28	-0.9184	0.0968
51.05	-0.8217	42.28	-0.9184	0.0968
51.05	-0.8217	42.28	-0.9184	0.0968
51.05	-0.8217	42.28	-0.9184	0.0968
51.05	-0.8217	42.28	-0.9184	0.0968
51.05	-0.8217	42.28	-0.9184	0.0968
51.05	-0.8217	42.28	-0.9184	0.0968
51.05	-0.8217	42.28	-0.9184	0.0968
55.43	-0.1821	49.00	-0.2865	0.1044

(Table Continues)

Dichotomous Person Ability	Dichotomous Z-Score	PCM Person Ability	PCM Z-Score	Z-Score Difference
55.43	-0.1821	49.00	-0.2865	0.1044
49.51	-1.0465	39.79	-1.1526	0.1061
49.51	-1.0465	39.79	-1.1526	0.1061
49.51	-1.0465	39.79	-1.1526	0.1061
49.51	-1.0465	39.79	-1.1526	0.1061
49.51	-1.0465	39.79	-1.1526	0.1061
49.51	-1.0465	39.79	-1.1526	0.1061
66.49	1.4328	66.08	1.3198	0.1130
66.49	1.4328	66.08	1.3198	0.1130
47.84	-1.2904	37.05	-1.4103	0.1199
47.84	-1.2904	37.05	-1.4103	0.1199
56.94	0.0384	51.15	-0.0843	0.1226
56.94	0.0384	51.15	-0.0843	0.1226
56.94	0.0384	51.15	-0.0843	0.1226
56.94	0.0384	51.15	-0.0843	0.1226
61.30	0.6750	57.84	0.5449	0.1301
61.30	0.6750	57.84	0.5449	0.1301
45.99	-1.5605	33.93	-1.7037	0.1432
58.55	0.2735	53.31	0.1189	0.1546
58.55	0.2735	53.31	0.1189	0.1546
58.55	0.2735	53.31	0.1189	0.1546
58.55	0.2735	53.31	0.1189	0.1546
64.91	1.2021	63.01	1.0311	0.1710
64.91	1.2021	63.01	1.0311	0.1710
53.25	-0.5004	44.61	-0.6993	0.1989
54.70	-0.2887	46.83	-0.4905	0.2018
54.70	-0.2887	46.83	-0.4905	0.2018
54.70	-0.2887	46.83	-0.4905	0.2018
60.32	0.5319	55.52	0.3267	0.2052
51.80	-0.7122	42.28	-0.9184	0.2063
51.80	-0.7122	42.28	-0.9184	0.2063
50.29	-0.9326	39.79	-1.1526	0.2200
50.29	-0.9326	39.79	-1.1526	0.2200
57.73	0.1537	51.15	-0.0843	0.2380
46.95	-1.4203	33.93	-1.7037	0.2834
46.95	-1.4203	33.93	-1.7037	0.2834
46.95	-1.4203	33.93	-1.7037	0.2834
62.37	0.8312	57.84	0.5449	0.2864
53.98	-0.3938	44.61	-0.6993	0.3055
53.98	-0.3938	44.61	-0.6993	0.3055

(Table Continues)

Dichotomous Person Ability	Dichotomous Z-Score	PCM Person Ability	PCM Z-Score	Z-Score Difference
53.98	-0.3938	44.61	-0.6993	0.3055
52.53	-0.6056	42.28	-0.9184	0.3129
44.96	-1.7109	30.18	-2.0564	0.3455
58.55	0.2735	51.15	-0.0843	0.3577
49.51	-1.0465	37.05	-1.4103	0.3638
68.42	1.7146	66.08	1.3198	0.3948
47.84	-1.2904	33.93	-1.7037	0.4133
70.94	2.0826	69.77	1.6668	0.4158
74.56	2.6112	74.59	2.1201	0.4911
45.99	-1.5605	30.18	-2.0564	0.4959
81.02	3.5544	82.27	2.8424	0.7121
81.02	3.5544	82.27	2.8424	0.7121
44.96	-1.7109	25.25	-2.5200	0.8091
92.75	5.2672	94.80	4.0208	1.2465