RICE UNIVERSITY

Reasoning, Fast and Slow:
Investigating Cognitive Abilities, Speed and Effects of Personality Traits

by

Amy Shaw

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Master of Arts

APPROVED, THESIS COMMITTEE:

Frederick L. Oswald, Committee Chair, Professor, Psychology

Margaret E. Beier, Associate Professor, Psychology

David M. Lane, Associate Professor, Psychology, Statistics, and Management

HOUSTON, TEXAS
APRIL 2015
ABSTRACT

Reasoning, Fast and Slow: Investigating Cognitive Abilities, Speed and Effects of Personality Traits

by

Amy Shaw

Previous research has suggested the existence of a general mental speed factor independent from general mental ability. However inconsistent empirical evidence suggests that the speed-ability relation might be more complicated than what was believed. Adopting the joint item response-response time modeling approach developed by van der Linden and colleagues (2006, 2007, 2009), the current study investigates the psychometric properties of the general speed factor and its relation to g in a reasoning task. Personality trait effects are examined as well to account for the speed and ability variances. In line with the earlier findings, results in the current study suggest that the reasoning speed and ability correlation is minimal and explanations for the mixed research findings are discussed within the context of individual differences and test situations.
ACKNOWLEDGMENT

I would like to thank my academic advisor and mentor, Dr. Fred Oswald, for his enlightening guidance, continuous support, and enormous encouragement in the completion of this thesis. I am very grateful for everything that he has taught me in the first half of my graduate school training. I would also like to thank my committee members, Dr. Margaret Beier and Dr. David Lane, for their extremely valuable input and feedback, which have improved this work greatly. Last but not least, I would like to express my deepest gratitude to my dear parents, Meilan Tong and Caiwen Xiao, for their unconditional love and endless support. To them I dedicate this thesis.
# TABLE OF CONTENTS

List of Tables ................................................................................................................iv
List of Figures ................................................................................................................v
Introduction .....................................................................................................................1
  Study ...............................................................................................................................18
    Model Specification .......................................................................................................18
    Method ..........................................................................................................................21
  Results .............................................................................................................................24
Discussion .........................................................................................................................31
References .........................................................................................................................35
 Appendix A .......................................................................................................................42
 Appendix B .......................................................................................................................45
 Appendix C .......................................................................................................................48
LIST OF TABLES

Table 1. Descriptive statistics for ability and personality measures ..................................25
Table 2. Observed sum score statistics for response model fit evaluation .............................26
Table 3. CIRT estimated parameter correlations .................................................................29
LIST OF FIGURES

Figure 1. The CIRT hierarchical framework ................................................................. 17
Figure 2. Bayesian residual plots for two example items ............................................... 27
Reasoning, Fast and Slow:

Investigating Cognitive Abilities, Speed and Effects of Personality Traits

General mental ability, or intelligence, is one of the most frequently discussed concepts across different subdisciplines of psychology. Although prominent researchers hold fundamentally different views about the definition of intelligence and the structure of cognitive ability models, the notion of g factor as the highest-order common factor of cognitive abilities has been well established since being discovered by Spearman’s factor analysis in 1904. Similar to the concept of g, a dominant general mental speed factor in power tests is implied in Carroll’s (1993) three-stratum theory; however there is much less theoretical discussion as well as empirical research on the structure of speed factors.

Although not explicitly, researchers tend to agree upon the existence of a general speed factor g-speed that can be extracted from single item or task response times through latent-variable modeling approaches. Nonetheless, it seems controversial to determine relations between the general speed factor and the well-known g factor. In Carroll’s landmark book on cognitive abilities, he summarized several inconsistent findings on the speed-ability relation from available studies at that time (e.g., DuBois, 1932; Horn, 1965; Kyllonen, 1985; Lord, 1956; Vernon, 1981) and concluded with a minimal or even orthogonal correlation between speed and level of intelligence in general. A handful of more recent studies on perceptual speed (considered as a component of the
broader concept of general speed factor; a closer examination of speed factors can be found in the following section) indicated that perceptual speed has a small to medium positive correlation with general mental ability and therefore may have predictive validity for task performance over and above intelligence (e.g., Ackerman, Beier, & Boyle, 2002; Mount, Oh, & Burns, 2008). In light of the above evidence, a common view (although may be not pervasive) of intelligence researchers on the speed-ability issue is that these two factors are positively correlated and the effect sizes tend to be small to moderate.

A spate of recent studies in the psychometric area, however, has suggested a correlation in the opposite direction (e.g., Goldhammer & Klein Entink, 2011; Klein Entink, Fox, & van der Linden, 2009; Klein Entink, Kuhn, Hornke, & Fox, 2009; Shaw, Oswald, Elizondo, & Wadlington, 2014). These studies followed a research paradigm where latent factors for item response and response time are simultaneously estimated via conjoint IRT modeling (CIRT; for more details see Fox, Klein Entink, & van der Linden, 2007; Klein Entink, Fox, & van der Linden, 2009; van der Linden, 2007). The CIRT modeling approaches originated from van der Linden’s (2006, 2007) work on incorporating response time into traditional IRT models so that collateral information can be extracted from response time for test design and diagnostic purposes. Previous studies have provided empirical support for the applicability of CIRT in computerized tests with a total test time limit (e.g., Goldhammer & Klein Entink, 2011; van der Linden, 2008) as well as in a relatively novel item-timed reasoning ability test (Shaw, Oswald, Elizondo, &
Wadlington, 2014). Although these researchers mainly focused on psychometric model development and application, their empirical results have converged on a general conclusion that the speed-ability correlation tends to be negative and moderate to substantial in magnitude. That is, high ability people tend to spend more time on the test, and are likely to be slower thinkers. The message conveyed by these findings directly challenged our previous impression of speed-ability relation and led to different theoretical implications.

From the individual differences lenses, a negative moderate-to-large speed-ability relation is somewhat counterintuitive, because if this conclusion is true, it implies that higher ability people tend to be slower on tasks and thus calls for caution when accuracy and speed are both preferred in practices. Early research on reflective and impulsive cognitive styles indicated individual differences on problem-solving approaches – some people tend to engage in slower but more accurate thinking (termed as more “reflective”) whereas the more impulsive counterparts tend to be faster and error-prone when working on tasks (Kagan, Rosman, Day, Albert, & Phillips, 1966). However no conclusion has been made upon the relationship between ability level and cognitive style. It is very possible and understandable, though, that the mixed research findings were due to the different test situations and measures used in the previous studies or statistical artifacts such as small sample sizes and range restriction, suggesting that the speed-ability relation can be even more complicated than what was believed. This inconsistency issue is
examined in more detail later, but it is worth noting in the introduction part that these two seemingly opposite viewpoints essentially address two different research questions. Intelligence psychologists are concerned with the role of processing speed (also considered a type of ability) in performance. Therefore, the relation between these two cognitive components would be ideally examined when processing speed and ability level are measured in separate cognitive tasks (e.g., Ackerman, Beier, & Boyle, 2002). In the psychometrics area however, researchers are interested in the additional information obtained by modeling response time (often recorded simultaneously during the ability task). Because response time in a test taking context is likely to depend on test taking strategy, pacing preference, motivation etc., the speed factor extracted from item response times in joint modeling reflects more than one’s cognitive processing speed.

In the current paper I attempt to review the prominent theories of cognitive abilities, speed factors, and effects of personality traits on response accuracy and speed, all of which inform the research questions and hypotheses that follow. The present study also further explores whether the CIRT model will appropriately fit data from a cognitive ability test where each item is timed (as opposed to the whole test, as is traditionally done). In addition, measures of the personality traits of Conscientiousness and Neuroticism are modeled into an extended CIRT model as person-level variables (Klein Entink, Fox, & van der Linden, 2009), so that more information about how personality influences cognitive ability test taking behavior and performance can be obtained.
Therefore, the goal of the current study is not to dispute the specific CIRT modeling techniques, but instead to further examine the applicability of this approach empirically as well as review the previous findings and clarify the terminological and methodological issues. IRT application is underrepresented in the integrative research on individual differences perhaps due to its complexity and model fitting issues. But it is my belief that it could contribute to our scientific understanding of individual differences and performance relationships by positioning the speed-ability relation topic within the broader context of psychological theories involving personality and test situation together with advanced psychometric methods. Note that although conceptually different, general reasoning ability (or mental speed during reasoning) is psychometrically interchangeable with the g factor (or general mental speed factor) in reasoning tests. The constructs discussed throughout this paper are considered general cognitive ability and speed assessed in a complex reasoning task.

**Structure of Reasoning Abilities and Speed**

General reasoning ability reflects an individual’s high-level abstract thinking capability and learning potential. The ability to reason is of crucial importance in many mental activities and considered as very similar to what the g factor (Spearman, 1923, 1927) and general fluid intelligence (Gf; Cattell, 1943, 1971) mean in practice. High reasoning ability individuals are better able to comprehend complex ideas, infer implicit relations, and generate rules based on available information. For decades, measures of
general mental ability have demonstrated some of the strongest validity coefficients for predicting job performance, and therefore such measures have been widely used in personnel selection and training practice.

In Horn and Noll (1997), general reasoning ability (Gf or g factor) is interpreted as inductive and deductive reasoning, which has a central place in the eduction of relations and correlates, that is, understanding the relations among stimuli and making meaning out of materials. Specifically, inductive reasoning refers to reasoning that proceeds from specific premises to a broader generalization, and deductive reasoning is about reasoning from multiple general premises to specific conclusions. Induction and deduction are considered two main methods of reasoning. Both inductive and deductive reasoning processes could involve verbal, quantitative and spatial contents (Horn & Cattell, 1967). The Cattell-Horn Gf-Gc model specifies quantitative reasoning and knowledge as a broad quantitative ability (Gq) separated from other factors such as reading/writing (Grw) and visual-spatial (Gv) abilities; however in Carroll’s three-stratum theory (1993), the g factor is defined as the Stratum III ability, broad visual perception ability is included in the Stratum II level, and quantitative reasoning and reading/writing reasoning are conceptualized as narrow abilities subsumed by Gf and Gc respectively. Despite their differences in some concepts, these two models share a lot of similarities in terms of the ability structure representation. The Cattell-Horn-Carroll (CHC) theory proposed by McGrew (1997) integrates both models, and has become the most influential
psychometric theory that guides cognitive ability test development and results interpretation.

The general reasoning ability measure (measure of g) used in the current study is Birkman Abilities Inventory (BAI; Birkman International, Inc., 2012), which measures an individual’s ability to identify and solve problems, plan and think abstractly, comprehend complex ideas, and learn quickly from experience. The BAI has each item timed up to 90 seconds and test takers can see a timer counting down on the screen, which might influence test takers’ perception of time passing and potentially affect test taking strategies/behaviors (addressed later in more detail). Like most reasoning tests, the BAI is an assessment that includes contents on quantitative, spatial, and verbal domains. Subtasks can have a mixture of content as well, such as a combination of quantitative and verbal information in arithmetic items. General reasoning ability is supposed to account for the common variance of subtask performance in different content areas. In the current study, the correctness of item responses is taken as an indication of an individual’s ability on the task; the assessment results and inferences can be drawn from either number of correct answers or person ability parameter estimates in an IRT model.

As pointed out by Jensen (1998), the g factor is ubiquitously reflected in all cognitive ability tests, especially where the eductive (“meaning-making”) ability is largely required (e.g., Raven’s Matrices reasoning ability tests). Individuals not only differ in their general reasoning ability but also general reasoning speed, which reflects
the cognitive processing fluency in reasoning tasks (i.e., reasoning problem solving speed). Due to social judgment and practical considerations, high efficiency in achieving a cognitive goal is generally endorsed and thought to be intelligent (without sacrificing much accuracy), although relying heavily on heuristic mental shortcuts to “think fast” in decision making can be problematic and error-prone as well.

In the Cattell-Horn-Carroll (CHC) theory, reasoning speed includes both cognitive processing speed and simple decision speed/reaction time components. The failure of clearly distinguishable speed factors might be due to the fact that many reasoning tasks are inevitably complex, involving multiple processes (such as induction and deduction) and content-related skills (such as language, number or spatial). In a reasoning task, several elements and complex processes are involved such as identifying and attending to stimuli, encoding or recognition, comprehension and comparison, retrieving associations, executing responses and monitoring performance (Carroll, 1993). Therefore, the concept of reasoning speed discussed in the current study involves more cognitive functions than perceptual speed assessed in elementary cognitive tasks. Perceptual speed determined with basic tasks is considered an underlying ability, but the “reasoning speed” construct (measured in power tests together with reasoning ability) reflects a combined effect of cognitive ability, item difficulty, non-cognitive factors and test situations. Individual differences in executing particular cognitive processes may cause remarkable performance variations with regard to both response accuracy and speed. Moreover, the
speed (or rate) of response and level (or accuracy) of performance aspects are often confounded because most power tests are given under a certain time limit. The extent to which speed can provide additional information about intelligence and different domains of cognitive ability remain questions to be further examined.

General reasoning speed affects the processing time needed across different mental activity stages and domains. However, it is useful and important to distinguish the true cognitive reasoning speed from the exhibited overall speed of response or rate of performance. The saying that “with other things being equal, the more quickly a test taker can produce the correct response, the greater is his/her ability” reflects a practical consideration of one’s ability combined with fastness of performing, but test takers may or may not choose to exhibit quickness of response due to other situation or personality factors. Researchers have been exploring response time analysis strategies for modeling the speed factor, which in actuality should be defined as speed of responding or rate of performing. As mentioned earlier, the concept of “speed” in a timed reasoning task reflects much more than perceptual or pure information processing speed. Test takers’ reasoning speed (as indicated by time taken to complete the task) at least involves multiple variables such as item difficulty, person ability, and test taking effort/motivation. Ideally, dimensions of speed and ability would be examined when individual ability and speed are measured separately through multiple tasks. For instance, pure speed tests may have very simple items administered with a strict time constraint, and test takers are
instructed to respond as quickly as possible. For low-difficulty questions, assuming that all test takers follow the instruction to answer as fast as possible, the measured response speed as indicated by the number of responses indicates the test taker’s information processing speed to a very large extent (because most examinees get all questions correct). However in practice, most cognitive ability tests are partially speeded (i.e., power tests administered with certain time limits; Lord & Novick, 1968), and the rate of test taking therefore involves more complex reasoning processes rather than the pure information processing speed in simple speed tests. Although implied by theories involving cognitive ability testing and speed-accuracy tradeoffs, the existence and psychometric properties of the general speed factor in tests with longer time limits (power tests) have not been well investigated empirically. Part of the reason for this might be that test scores in a time limited power test (or speeded test) are unknown functions of level of ability, reasoning speed, pacing style, test-taking motivation, amount of the time limit and difficulty of items. In the current framework, measures of response time spent on answering items are taken as indicants of the general reasoning factor, which relates to but differs from pure mental processing speed measured separately in a speed test.

*Research Question 1:* Does the general reasoning speed factor (or rate of test taking) exist as a unidimensional construct independent from general reasoning ability?
**Speed and Accuracy: Trade-off or Facilitation?**

In a timed ability test, test takers are often instructed to answer questions as fast and accurately as possible. The concept of test performance addresses the cognitive output in each time unit (Wilhelm & Schulze, 2002), that is, the amount of time spent on reading and answering an item might be indicative of the cognitive ability level as much as the response itself. Some basic cognitive processes may be too simple to reflect much variation in the population, but in a complex reasoning task where multiple processing stages and mental activities are involved (e.g., encoding, inference, application, monitoring), high processing speed helps improve test performance by releasing the information currently held in the direct access region more quickly. Wilhelm and Schulze (2002) argued that in a complex reasoning task where time constraints are applied, mental speed is likely to influence task performance. That is, the mental speed is viewed as an information processing ability component (Vernon, Nador, & Kantor, 1985); with a higher mental speed, the information stored in working memory can be processed more quickly, and thus it is less likely to reach one’s working memory capacity limit in a complex reasoning task. Therefore, complex reasoning tasks are often conceived as a compound of measuring both working memory capacity and mental speed ability (Jensen, 1998; Schnipke & Scrams, 1997, 2002; Wilhelm & Schulze, 2002). Working memory capacity as a source of individual differences in reasoning ability also affects information processing speed, so it offers a reason why speed and ability are correlated (Ackerman et
al., 2002). Based on these principles, it is reasonable to conclude that the reasoning speed factor may have a minimal correlation with the reasoning ability factor, yet positively contributes to test performance on timed tests.

Empirical studies have shown inconsistent evidence on the relation between reasoning ability and speed. Kyllonen (1985) examined the relation between perceptual speed and ability level – the results suggested a close-to-zero correlation between these two factors. CIRT modeling studies where specific domain contents were involved mostly revealed moderate negative correlations between reasoning abilities and their associated speed factors (e.g., negatively correlated quantitative reasoning ability and speed, Klein Entink, Fox & van der Linden, 2009; negatively correlated figural reasoning ability and speed, Klein Entink, Kuhn, Hornke, & Fox, 2009). The inconsistent results may be caused by different content measures, samples, or modeling approaches, and more probably, due to the confusion of terminology: the perceptual or information processing speed was indicative of cognitive speed itself in Kyllonen’s study, but the response speed in Klein Entink et al.’s studies should be regarded as a broader indicant of rate of performing that relates to one’s cognitive speed, and also in relation to other variables such as task difficulty, individual’s ability and experience, task performing preference, motivation, attention, test condition.

Klein Entink et al. (2009) concluded that the correlation is not necessarily positive, and also acknowledged that the dependency between ability and speed could vary from
different samples. They have found zero and positive correlations as well (Klein Entink, Fox & van der Linden, 2009). In Goldhammer and Klein Entink (2011), a no-time-limit pure power study was administered and a negative correlation was obtained, which was probably a reflection of test motivation issue. Thus, a caution should be sounded is that the specificities of studies such as the sample, measure, and modeling approach must be taken into consideration. The current study follows the CIRT approach and utilizes the same measure (i.e., BAI) as used in Shaw et al. (2014); the only difference is that the current study administered the tests in a real selection context whereas mixed sample was used in the 2014 study. This allows for qualifying research questions such as whether the results obtained from the selection context would show a similar or different pattern.

**Research Question 2: In the current framework where reasoning speed (or rate of performing reasoning tasks) and reasoning ability (or level of response accuracy) are simultaneously modeled in CIRT as person parameters, how does the speed factor relate to the ability factor?**

**Effects of Personality Traits on Reasoning Performance**

The measurement of individual differences in psychological constructs has a long tradition in the applied psychology field. A vast literature has been devoted to the exploring relationships between ability, personality, interests and other personal characteristics, in addition to their influence on the behaviors, thoughts and feelings of examinees. The current study investigates how Neuroticism and Conscientiousness are
involved in the test-taking process, and whether personality traits show effects on speed and ability.

Neuroticism (or lack of Emotional Stability) has been consistently demonstrated to be positively correlated with test anxiety (e.g., Ackerman & Heggestad, 1997; Deffenbacher, 1980; Liebert & Morris, 1967; Schmidt & Riniolo, 1999). An optimal level of anxiety for the task should be at an intermediate level – neither too high nor too low (Towle & Merrill, 1975; Yerkes & Dodson, 1908). According to research by Wine (1980) and Hill (1977, 1980, 1984), test anxiety adversely affects examinees’ test performance and self-efficacy, because anxiety leads to reduced or even negative motivation that prevents examinees from contributing all resources to focus on completing the task itself (Hill, 1977), and cognitive information-processing capabilities, such as memory-retrieval processes and problem-solving strategies, are also adversely interfered by these task-irrelevant concerns (Wine, 1980). Because neurotic test takers are likely to feel less confident about their answers and overcoming the interference caused by test anxiety may slow down overall processing speed, being high on this trait might mean greater impairment on both response speed and accuracy.

Conscientiousness may also come into play by influencing test-taking behaviors (such as effort or rapid guessing). Conscientious test takers tend to be more cautious and detail-oriented, so that they may spend more time reflecting, re-evaluating, affirming, validating, and modifying their answers. These efforts are likely to increase one’s test
performance and response time. A conscientious person evaluates conclusions and thus takes longer.

Whether these effects exist in item-timed tests remains an empirical question, although it is reasonable to hypothesize that having a countdown timer for each item might stimulate (or even increase) examinees’ anxiety and distraction throughout the test. According to the trait-activation-theory (TAT; Tett & Burnett, 2003), the test situation and characteristics provided cues for the behavioral expression of relevant traits (i.e., Conscientiousness and Neuroticism). Therefore, I speculate that Neuroticism is a critical personality trait as well as the detail-oriented aspects of Conscientiousness that may affect test taking behaviors. Qualifying these predictions by personality traits remains an exploratory question in a CIRT model, and incorporating personality as variables may in turn help explain certain test response strategies and behaviors. To my knowledge, this is the first empirical study that investigates how personality traits influence test taking speed and ability jointly.

Research Question 3: Neuroticism and Conscientiousness (modeled into CIRT as variables) help account for the variance of reasoning speed and ability and show differential effects. Specifically, Neuroticism correlates with rate of performing and level of accuracy negatively, whereas Conscientiousness negatively relates to rate but positively predicts accuracy.
CIRT Model

Historically, the speed factor is commonly ignored in traditional IRT models for the sake of simplicity. CIRT model makes use of traditional IRT models to incorporate response time explicitly, explains item and person characteristics in terms of latent variables (e.g., estimates of item difficulty, item time intensity, person ability, and person speed), and thus helps investigate the relationship between the speed component and accuracy component jointly. By incorporating person-level variables, CIRT helps explain whether and how a test taker’s response speed and accuracy can be influenced by non-ability factors such as personality traits.

The CIRT model is built within a Bayesian framework developed by van der Linden (2007; see Figure 1). The model uses a hierarchical framework with separate models for item responses and response time in the first level and a higher level model for the population distribution of ability and speed parameters in the current test (see van der Linden, 2007 for statistical demonstration details). Overall, the CIRT model allows for testing the latent effects of item parameters (e.g., item difficulty and item time intensity) and person parameters (person ability and speed) on the response accuracy and response time (RT) outcomes. Personality traits can be modeled as the person-level variables in an extended CIRT model within a multivariate multilevel regression structure (Klein Entink, Fox, & van der Linden, 2009).
In addition to classic IRT estimation parameters (e.g., item discrimination, item difficulty and person ability), CIRT provides another set of three parameters estimated from RT information: time discrimination, time intensity, and person speed parameters. As suggested by previous studies, the item and person parameter estimation accuracy (\textit{a posteriori error}) is expected to be improved in CIRT by obtaining additional information from response time and personality variables. According to the IRT models of task response, the probability that the item response is correct is a function of the test taker’s ability level and item difficulty; response time adds information that reflects the test taker’s rate of performing and item time demands; therefore parameter estimation can be improved by virtue of modeling response time spent on complex reasoning tasks. Whether modeling personality traits as variables improves estimation accuracy remains a research question to be explored in the current study, but I speculate that there are
interactive effects among person and item characteristics on test performance, so that modeling personality variables is likely to further improve parameter estimation in specific ways. For instance, high ability people are expected to make correct responses to more difficult items, whereas low ability people are expected to make correct responses only to items of low difficulty. When considering personality, low ability but conscientious people may engage more (and take more time in responding) in each decision-making and responding stage, however the less conscientious counterparts might be more prone to give up; or, neurotic but high ability people’s performance may be less likely affected by the timer presented on the screen compared to their low ability counterparts.

**CIRT Model Specification in the Current Study**

The CIRT model specification follows the formulas developed by van der Linden and colleagues (Klein Entink, Fox, & van der Linden, 2009; van der Linden, 2007); a detailed statistical demonstration can be found in van der Linden (2007).

**Level 1 measurement models.** For binary item response data, the two-parameter normal ogive (2PNO) model is used. The 2PNO model defines the probability of test taker $i$ answering item $k$ correctly as a function of the test taker’s ability $\theta_i$ as well as the item difficulty $b_k$ and item discrimination $a_k$, that is,

$$P(Y_{ik} = 1|\theta_i, a_k, b_k) = \phi(a_k \times \theta_i + b_k), \quad (1)$$

where $\phi(.)$ denotes the normal distribution function.
For item response time data, the two-parameter log-normal (2PLNO) model is used. The log-response time is denoted as $T_{ik}$, which is the function of the test taker’s speed $\zeta_i$ and the item’s time intensity $\lambda_k$ and time discrimination $\phi_k$, that is,

$$T_{ik} = -\phi_k \times \zeta_i + \lambda_k + \epsilon_{ik}, \tag{2}$$

Where the residual $\epsilon_{ik} \sim N(0, \tau^2_k)$. These two measurement models describe the Level 1 distributions of responses and response times. The above parameterization enables one to partition the variance in responses and response times into the person effects (i.e., ability and speed) and item effects (i.e., difficulty, discrimination and time intensity).

**Modeling item and person parameters at Level 2.** At the second level, a covariance structure $\Sigma_{person}$ is specified to allow for the dependencies between person speed and ability parameters at Level 1. The covariance parameter $\sigma_{\theta\zeta}$ reflects the possible correlation between person ability and speed within the population of test takers. The variance parameters $\sigma^2_{\theta}$ and $\sigma^2_{\zeta}$ indicate the individual differences in person ability and speed respectively. From a Bayesian perspective, this jointly bivariate normal distribution of speed ($\zeta$) and ability ($\theta$) can be considered as a common prior for both person parameters:

$$(\theta, \zeta) = \mu_{person} + e_{person}, \quad \mu_{person} = (\mu_\theta, \mu_\zeta), \quad e_{person} \sim N(0, \Sigma_{person}), \tag{3}$$

where $\Sigma_{person}$ specifies the covariance matrix given by:
\[\Sigma_{\text{person}} = \begin{bmatrix} \sigma_\theta^2 & \sigma_{\theta\zeta} \\ \sigma_{\theta\zeta} & \sigma_\zeta^2 \end{bmatrix}.\]

With regard to the items, a similar jointly multivariate normal distribution and covariance structure are specified for the item parameters in both response and response time models (i.e., item difficulty \(b_k\) and item discrimination \(a_k\); item time intensity \(\lambda_k\) and time discrimination \(\phi_k\)):

\[(b, \alpha, \lambda, \phi) = \mu_{item} + e_{item}, \quad \mu_{item} = (\mu_b, \mu_\alpha, \mu_\lambda, \mu_\phi), \quad e_{item} \sim N(0, \Sigma_{item}),\] (4)

where \(\Sigma_{item}\) specifies the covariance matrix given by:

\[\Sigma_{item} = \begin{bmatrix} \sigma_b^2 & \sigma_{ba} & \sigma_{b\lambda} & \sigma_{b\phi} \\ \sigma_{ba} & \sigma_\alpha^2 & \sigma_{a\lambda} & \sigma_{a\phi} \\ \sigma_{b\lambda} & \sigma_{a\lambda} & \sigma_\lambda^2 & \sigma_{\lambda\phi} \\ \sigma_{b\phi} & \sigma_{a\phi} & \sigma_{\lambda\phi} & \sigma_\phi^2 \end{bmatrix}.\]

**Incorporating person-level variables.** Personality traits (i.e., Neuroticism and Conscientiousness) are incorporated in the model as variables, denoted by \(N_l\) and \(C_l\), respectively. That is,

\[\begin{bmatrix} \theta_l \\ \zeta_l \end{bmatrix} = \begin{bmatrix} \gamma_{00} \\ \gamma_{10} \end{bmatrix} + \begin{bmatrix} N_l \times \gamma_{01} + C_l \times \gamma_{02} \\ N_l \times \gamma_{11} + C_l \times \gamma_{12} \end{bmatrix} + \begin{bmatrix} e_{0l} \\ e_{1l} \end{bmatrix}\] (5)

where \(e_{0l} \sim N(0, \Sigma_{person})\).

By regressing person parameters on person-level variables (i.e., \(N_l\) and \(C_l\)), the standardized regression coefficients are evaluated to compare whether the two personality traits predict ability and speed to the same extent or differently.
Method

Participants

The current study is based on $N = 300$ test-takers who responded to the Birkman Abilities Inventory (BAI) and The Birkman Method® (TBM). The former is a computerized measure of general reasoning ability in the quantitative, spatial, and verbal domains. The latter is a computerized assessment of self-reported personality, social perceptions, and vocational interests. Tests were administered between August 2012 and June 2014. Most subjects ($N = 277$) were from the United States, with a few from other countries (Canada, Britain, Austria, Singapore, Hong Kong, and Australia). In terms of the sample’s occupational and educational backgrounds, 81.3% of the participants finished university education and 63.3% are in Management, Business and Financial Operations occupations (with the rest of sample being spread out in various occupations such as computer science, sales, office and administrative support).

Measures

General reasoning ability scale. General reasoning ability is measured by Birkman Abilities Inventory (BAI), a computerized assessment that measures quantitative, spatial, and verbal abilities (Birkman International, Inc., 2012). The online test version contains 36 multiple-choice items (30 scored and 6 for field testing) with each item allocated a maximum of 90 seconds for answering, and although the maximum time to complete the test is 54 minutes, it typically takes 25 minutes to complete. A countdown timer is
presented on the screen when test takers respond to each item. The amount of time allocated to each item is intended to allow all people who have knowledge of the correct answer to be able to respond. Thus, the BAI was designed to measure accuracy without undue time pressure; in other words, the BAI is generally intended to be a power test. A summary of the scale development efforts and psychometric properties of BAI can be found in a technical report produced by the test vendor (Birkman International, Inc., 2012).

**Personality assessment.** Personality traits are measured by The Birkman Method® (TBM), a 298-item computerized questionnaire that assesses self-report personality, social perceptions, and vocational interests (Birkman International, Inc., 2013); with detailed psychometric properties available in its test manual (Birkman, Elizondo, & Wadlington, 2013). Five personality orientations are measured with 88 items and correspond with the Five-Factor Model of personality (FFM): Emotive Orientation (Neuroticism), Social Orientation (Extraversion), Process Orientation (Conscientiousness), Control Orientation (Agreeableness), and Change Orientation (Openness). The scores on Emotive Orientation and Process Orientation scales are used as the personality variables to be incorporated in CIRT. Specifically, Emotive Orientation (hereafter called Neuroticism) refers to an individual’s orientation to respond to life events emotionally. High emotive people tend to experience intense and less steady
emotions. Process Orientation (hereafter called Conscientiousness) represents the extent to which individuals are organized, goal striving, cautious and disciplined.

Data Analysis

**Estimation and software.** Data were analyzed in R, an open-source programming environment for various statistical analyses, and the CIRT package in R developed by Fox, Klein Entink, and van der Linden (2007) allows for the joint analysis of response accuracy and RT with person-level variables. A Bayesian Markov Chain Monte Carlo (MCMC) algorithm is utilized in the CIRT package to obtain parameter estimates by posterior simulation from the joint distribution of model parameters, given the observed data. The default non-informative priors are used in the CIRT package.

**CIRT model selection and model fit.** For model comparison and selection in CIRT, Bayes factors are computed (Kass & Raftery, 1995). Bayes factors serve a similar function in the context of Bayesian methods as likelihood ratios for selecting the best model. The Deviance Information Criterion (DIC; Spiegelhalter, Best, Carlin, & van der Linde, 2002) is used as the model fit statistic – it is the sum of a deviance measure and a penalty function for the model complexity based on the number of free parameters.
Results

Descriptive results

The descriptive statistics and correlations for the BAI and TBM scores are summarized in Table 1. Mean correctness for the current sample was 20.33 (out of 30) and mean total response time was 29.38 (out of 54 min). Both statistics indicate that the test had a reasonable time constraint and difficulty level. The Pearson correlation between the BAI total score and response time was close to zero ($r = -.07, ns$), suggesting that the estimated speed-ability relation may be minimal as well. CIRT analyses will help qualify this finding (because this estimated parameter correlation collapses across person ability/speed and item difficulty/time intensity in the current sample).
Table 1

*Descriptive Statistics for Ability and Personality Measures*

<table>
<thead>
<tr>
<th>Measure</th>
<th>Mean</th>
<th>SD</th>
<th>Correlation $r$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>1. BAI (total score)</td>
<td>20.33</td>
<td>4.93</td>
<td>1.00</td>
</tr>
<tr>
<td>2. BAI (total time [seconds])</td>
<td>1762.82</td>
<td>333.46</td>
<td>-.07</td>
</tr>
<tr>
<td>3. TBM Emotive</td>
<td>2.04</td>
<td>2.65</td>
<td>.07</td>
</tr>
<tr>
<td>4. TBM Process</td>
<td>11.59</td>
<td>2.23</td>
<td>-.23**</td>
</tr>
</tbody>
</table>

*Note.* $N = 300$. ** denotes that correlation is significant at the 0.01 level (2-tailed). BAI is short for Birkman Abilities Inventory; TBM is short for The Birkman Method.

**Model selection, estimation and fit**

First, for measurement models, the one-parameter logistic (1PL) and two-parameter logistic (2PL) IRT models were fitted to the data to explore the required number of parameters. The most restrictive 1PL model that includes only the difficulty and time intensity parameters shows a higher DIC (88195.93) compared to the two discrimination parameter model (88127.56), suggesting that 2PL would be the better performing model.

Bayesian posterior predictive checks were used to evaluate the fit of the response model. The check consists of comparing the replicated data from the posterior predictive distribution in the model to the observed data. The probability of the model-predicted test
statistic being greater than the observed test statistic is assessed (posterior predictive $p$
value). As shown in Table 2, the response model had a very good fit because there were
only two extreme $p$ values (lower than .025 or higher than .975) denoting unlikely
observations under the model.

Table 2

| Observed Sum Score Statistics for Response Model Fit Evaluation |
|---|---|---|---|
| Sum score | $p$ value | Sum score | $p$ value |
| 0 | .00 | 16 | .22 |
| 1 | .01 | 17 | .44 |
| 2 | .04 | 18 | .95 |
| 3 | .11 | 19 | .47 |
| 4 | .21 | 20 | .04 |
| 5 | .37 | 21 | .42 |
| 6 | .19 | 22 | .62 |
| 7 | .16 | 23 | .50 |
| 8 | .35 | 24 | .58 |
| 9 | .54 | 25 | .18 |
| 10 | .11 | 26 | .78 |
| 11 | .44 | 27 | .40 |
| 12 | .18 | 28 | .18 |
| 13 | .40 | 29 | .48 |
| 14 | .46 | 30 | .11 |
| 15 | .76 | |

Bayesian residual analysis was used to assess the fit of the response time model
(Klein Entink, Fox, & van der Linden, 2009). The observed response time of a test taker
in one item is evaluated under the respective posterior density of response times. The
probability values are plotted against the expected values under the U (0, 1) distribution.
for the 30 items (see Appendix C). By checking these QQ plots, most items achieved good fit, because the residuals followed the expected values so that the plots appear to be linear. For some items, the aberrancies were more salient, but there was no general deviation trend from the identity line suggesting under- or over- prediction, implying an overall acceptable fit across the entire range of items. Figure 2 shows two example items for illustration. For item 14 the RT model fits the observed RT data very well; for item 23 the RT model overpredicted faster responses and this deviation was noticeable.

Figure 2. Bayesian residual plots for two example items.
**Estimated parameters and correlations**

Table 3 presents the estimated parameters from CIRT. The estimated correlation between reasoning speed and ability, $\rho(\theta, \zeta) = .08$, indicating that there might be very low to zero correlation between speed and ability factors. Although the data variance could roughly be accounted for by a general speed factor, a different pattern might be shown if quantitative, spatial and verbal speed factors are separately examined.

Supplemental CIRT analyses of domain-specific factors were conducted, and the results suggested that after decomposition, the specific speed-ability relations tend to be positive ($\rho_{\text{quant}} = .23$, $\rho_{\text{spatial}} = .29$, and $\rho_{\text{verbal}} = .19$), meaning that a person who works faster than average tends to have an above-average ability level. The larger effect sizes of correlations revealed by examining the domain-specific factors also suggested the added value by looking at specific factors vs. the general factor alone. These results are consistent with findings in Kyllonen (1985) and Carroll’s theory (1993).
Table 3

*CIRT Estimated Parameter Correlations*

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Correlation ρ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>1. Item difficulty</td>
<td>1.00</td>
</tr>
<tr>
<td>2. Item time intensity</td>
<td>.88</td>
</tr>
<tr>
<td>3. Person ability</td>
<td>-</td>
</tr>
<tr>
<td>4. Person speed</td>
<td>-</td>
</tr>
</tbody>
</table>

Also, results gave insight into the relationship between item parameters. A strong positive relationship between item difficulty and item time intensity, $\rho (b, \lambda) = .88$. This is reasonable given that more difficult items require more time in general for answering.

In all, a general speed factor can be extracted from the CIRT model, and the amount of shared variance between ability and speed was found to be small, so the variance in speed cannot be accounted for by ability levels.

*Personality variables.* In the extended CIRT model, the effects of personality variables on person parameters were estimated through standardized regression coefficients. Ability was not predicted by either trait (Neuroticism on ability $\gamma_{01} = .00$, Conscientiousness on ability
\( \gamma_{11} = -.05 \), but speed was predicted by both traits (Neuroticism on speed \( \gamma_{02} = .35 \), Conscientiousness on speed \( \gamma_{12} = .19 \)). This result was unexpected and counterintuitive, because our theoretical frameworks indicated that Neuroticism and Conscientiousness should help explain some variance in test performance, and regarding the effects on speed, Conscientiousness and Neuroticism were both expected to be negatively correlated with speed. Comparing this result to the previous zero-order correlations (i.e., the negative correlation -.23 between Process and BAI scores, and the close-to-zero correlations between personality and time), the inconsistency could be reconciled by noticing the joint estimation of parameters in CIRT and the fact that the parameters were dependent on other factors (e.g., item difficulty) rather than simple zero-order correlations. As the descriptive results suggested a medium-size relation between Emotive and Process scores and a moderate correlation between Process and ability score, one might wonder whether there was a suppression effect in the CIRT multiple regression models (for details about suppression see Cohen, Cohen, Aiken, & West, 2013). Multiple regressions based on the raw scores were conducted and the results suggested no suppression effect: in the full model, \( \beta \) (Process) = -.23, \( p < .01 \), \( \beta \) (Emotive) = -.01, ns; in the reduced model where Emotive was removed as an IV, \( \beta \) (Process) did not change and R-square remained .05. However this inconsistency could be explained by noticing the joint estimation of parameters in CIRT and the fact that the parameters were dependent on other factors (e.g.,
item difficulty) rather than simple zero-order correlations. This also indicated a noticeable advantage of using IRT modeling approaches.

**Discussion**

Overall the current study supported the existence of a general speed factor which can be extracted from item response time through CIRT modeling (as suggested in Research Question 1). General speed factor measured within a reasoning test situation is distinct from pure cognitive processing speed (often measured through speed tests separately from ability level). Research Question 2 raised the issue of speed-ability relation in the current framework, where reasoning speed and ability were simultaneously modeled in CIRT. The finding suggested a close-to-zero relation and this was probably due to the strong situational effect in a selection context. Therefore, research findings of speed-ability relation in CIRT might differ depending on the test situation and thus should be interpreted with caution when generalizing these results to other contexts.

Research Question 3 explored the possible effects of personality traits (i.e., Neuroticism and Conscientiousness) on test taking behaviors. The findings suggested a different pattern from what was hypothesized, which will be discussed in the following section.

Despite the lack of consistent evidence about its structure, the idea of general cognitive speed is of interest to and often discussed by researchers and practitioners. It sounds a plausible idea that when other things are equal, test takers who can produce correct answers more quickly may be perceived more intelligent (see Thorndike et al.,
1926). This tenet provides a theoretical justification for imposing time limits in most power tests where the response accuracy rather than speed is measured. But what people can do and will do usually depends on a lot of other non-cognitive factors as well as situations. Claiming that being efficient in achieving cognitive goals equates to higher ability would be problematic, and it is also an argument about the nature of intelligence. Whether higher ability people are more rigid and less comfortable with ambiguity, so that they tend to think “more slowly” remains an empirical question, and it is beyond the scope of the current work to discuss all possible response patterns. Results from the current study suggest that in a high-stakes test condition where most people are concerned with their performance, they tend to put in more effort and the almost zero speed-ability relation reflects what early researchers believed, namely that speed may have minimal contribution to one’s intelligence at least under a high-stakes selection context. In a testing situation where test-taking attitude, effort and motivation might affect performance, it is likely that the best performers are not necessarily the most intelligent ones. So, perhaps it’s safer to speculate that high ability people have the option to perform well and excel in a selection context, but they may have different strategies and choose not to be performance-focused in a low-stakes context.

Supplemental analyses on the domain-specific speed factors indicated positive correlations between speed and ability. Although the results should be interpreted cautiously because of the aberrancies of some items in the current data, it is likely that
specific domain factors may reveal different patterns from general factors. In the
cognitive ability area, it might be promising to decompose broad facets, as the general
psychometric factor may contain cognitive modality components that exert different
effects that could have been overlooked if only the general factor is considered.

The personality effects obtained from the current study are not as expected. One
plausible alternative explanation might be that extreme Conscientiousness can cause extra
stress that harms one’s performance or leads to more subjective fatigue. But testing this
explanation would require further examination on the test length and fatigue levels in
reasoning tasks. Perhaps a more convincing and simple answer would be that neurotic
and conscientious test takers paid more attention to the timer presented on the screen,
which gave them a stronger sense of test time control and fast pacing demand. It would
be interesting to compare performance on different forms of the current measure (e.g., no
timer vs. timer on each item or even further manipulating the location and size of the
timer shown on the screen) to see whether personality played a role through this proposed
distraction avoidance and attention control mechanism. Because the BAI puts time limit
on each item rather than imposing a total time constraint, personality traits might be more
predictive in the traditional format tests where individuals have more flexibility in terms
of pacing themselves and adjusting test taking strategies. It is a pity that a lot of
demographic data were unavailable during the data collecting process, because in
addition to the test condition variance, discrepancy in educational level, age and even
occupation could lead to very different test performance and behaviors. Two potential research directions are suggested for future research agenda: 1) it might be interesting to explore the interactional effect of both personality traits to find out more details about how personality affects test taking behaviors, and 2) it would be useful to obtain facet-level personality traits (e.g., Achievement Striving and Dependability as facets of Conscientiousness) to illuminate the effects of personality in a more refined way.

Taken together, these results all indicated that response time can provide additional information such as test motivation and condition. Thus practitioners and researchers are encouraged to incorporate response time and also conduct future studies on test performance and integration of individual difference variables with the CIRT framework. The advantage of this model is that it statistically allows for simultaneous modeling of ability and speed as well as other person-level variables. Furthermore, criterion data would help investigate whether reasoning speed does show incremental predictive validity when reasoning tasks are used to predict future job performance.
References


Thorndike, E. L., Bregman, E. O., Cobb, M. V., Woodyard, E., & the staff of the Division of Psychology of the Institute of Educational Research of Teachers College,


Appendix A: Birkman Abilities Inventory Sample Items

<table>
<thead>
<tr>
<th>Instrument Name</th>
<th>BAI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item Number</td>
<td>IS0000</td>
</tr>
<tr>
<td>Item type</td>
<td>MC</td>
</tr>
<tr>
<td>Passage</td>
<td>NA</td>
</tr>
<tr>
<td>Description of item</td>
<td>Data identification</td>
</tr>
<tr>
<td>Key</td>
<td>C</td>
</tr>
<tr>
<td>Item Writer</td>
<td></td>
</tr>
<tr>
<td>Date</td>
<td></td>
</tr>
<tr>
<td>Current version of item (previous version saved to test file with date)</td>
<td></td>
</tr>
<tr>
<td>Response options</td>
<td></td>
</tr>
<tr>
<td>Comments</td>
<td></td>
</tr>
</tbody>
</table>

What makes figures 1 and 2 more like one another than like the third figure?

1.       2.       3.

a). The number of sides of each figure
b) The overall size of each figure
c) The shade within each figure
d) The length of the lines within each figure
### Appendix A: Birkman Abilities Inventory Sample Items (continued)

<table>
<thead>
<tr>
<th>Instrument Name</th>
<th>BAI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item Number</td>
<td>IV0000</td>
</tr>
<tr>
<td>Item type</td>
<td>MC</td>
</tr>
<tr>
<td>Passage</td>
<td>NA</td>
</tr>
<tr>
<td>Description of item</td>
<td>Verbal analogy with opposites</td>
</tr>
<tr>
<td>Key</td>
<td>B</td>
</tr>
<tr>
<td>Item Writer</td>
<td></td>
</tr>
<tr>
<td>Date</td>
<td></td>
</tr>
<tr>
<td>Current version of item (previous version saved to test file with date)</td>
<td>High is to Low as Big is to________</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Response options</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Short</td>
</tr>
<tr>
<td>b. Small</td>
</tr>
<tr>
<td>c. Stout</td>
</tr>
<tr>
<td>d. Great</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Instrument Name</th>
<th>BAI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item Number</td>
<td>SQ0000</td>
</tr>
<tr>
<td>Item type</td>
<td>MC</td>
</tr>
<tr>
<td>Passage</td>
<td>NA</td>
</tr>
<tr>
<td>Description of item</td>
<td>Complex quantitative analogy</td>
</tr>
<tr>
<td>Key</td>
<td>C</td>
</tr>
<tr>
<td>Item Writer</td>
<td></td>
</tr>
<tr>
<td>Date</td>
<td></td>
</tr>
<tr>
<td>Current version of item (previous version saved to test file with date)</td>
<td>♠ ♦ ♦ ♦ ♧ ♧ is to ♠ ♦ ♦ ♦ ♧ ♧ as 8 is to _____</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Response options</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. 88877</td>
</tr>
<tr>
<td>b. 88778</td>
</tr>
<tr>
<td>c. 88788</td>
</tr>
<tr>
<td>d. 87788</td>
</tr>
</tbody>
</table>

| Comments |             |
### Appendix A: Birkman Abilities Inventory Sample Items (continued)

<table>
<thead>
<tr>
<th>Instrument Name</th>
<th>BAI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item Number</td>
<td>SV0000</td>
</tr>
<tr>
<td>Item type</td>
<td>MC</td>
</tr>
<tr>
<td>Passage</td>
<td>NA</td>
</tr>
<tr>
<td>Description of item</td>
<td>Logic based</td>
</tr>
<tr>
<td>Key</td>
<td>A</td>
</tr>
<tr>
<td>Item Writer</td>
<td></td>
</tr>
<tr>
<td>Date</td>
<td></td>
</tr>
<tr>
<td>Current version of item (previous version saved to test file with date)</td>
<td>According to recent test results for your company, no account managers are creative thinkers. If this is true, what else must be true?</td>
</tr>
</tbody>
</table>
| Response options | a. Some creative thinkers are not account managers.  
|                  | b. All creative thinkers are account managers.  
|                  | c. Some account managers are creative thinkers.  
|                  | d. Some creative thinkers are account managers. |
| Comments         |       |
Appendix B: Descriptive statistics plots

1. Histograms of total score

![Histogram of total score](image)

- **Mean** = 20.33
- **Std. Dev.** = 4.932
- **N** = 300
Appendix B: Descriptive statistics plots (continued)

2. Histograms for total response time

![Histogram of total response time with mean 1762.82, std. dev. 333.455, and N = 300](image)
Appendix B: Descriptive statistics plots (continued)

3. Scatterplot of the relation between total score and response time
Appendix C: Bayesian residual plots for 30 items

Item 1

Item 2
Appendix C: Bayesian residual plots for 30 items (continued)

**Item 3**

![Graph for Item 3](image)

**Item 4**

![Graph for Item 4](image)
Appendix C: Bayesian residual plots for 30 items (continued)

**Item 5**

![Plot for Item 5](image)

**Item 6**

![Plot for Item 6](image)
Appendix C: Bayesian residual plots for 30 items (continued)

**Item 7**

**Item 8**
Appendix C: Bayesian residual plots for 30 items (continued)

**Item 9**

![Graph for Item 9]

**Item 10**

![Graph for Item 10]
Appendix C: Bayesian residual plots for 30 items (continued)

**Item 11**

![Graph for Item 11](image1)

**Item 12**

![Graph for Item 12](image2)
Appendix C: Bayesian residual plots for 30 items (continued)

**Item 13**

![Item 13 Graph]

**Item 14**

![Item 14 Graph]
Appendix C: Bayesian residual plots for 30 items (continued)

**Item 15**

![Graph for Item 15]

**Item 16**

![Graph for Item 16]
Appendix C: Bayesian residual plots for 30 items (continued)

Item 17

![Graph for Item 17]

Item 18

![Graph for Item 18]
Appendix C: Bayesian residual plots for 30 items (continued)

**Item 19**

![Graph for Item 19]

**Item 20**

![Graph for Item 20]
Appendix C: Bayesian residual plots for 30 items (continued)

**Item 21**

![Graph](image1)

**Item 22**

![Graph](image2)
Appendix C: Bayesian residual plots for 30 items (continued)

**Item 23**

![Graph showing cumulative proportion vs. Pr(t^*_k < t | k)]

**Item 24**

![Graph showing cumulative proportion vs. Pr(t^*_k < t | k)]
Appendix C: Bayesian residual plots for 30 items (continued)

**Item 25**

![Bayesian residual plot for Item 25](image)

**Item 26**

![Bayesian residual plot for Item 26](image)
Appendix C: Bayesian residual plots for 30 items (continued)

**Item 27**

![Graph for Item 27](image1)

**Item 28**

![Graph for Item 28](image2)
Appendix C: Bayesian residual plots for 30 items (continued)

**Item 29**

![Graph for Item 29](graph29)

**Item 30**

![Graph for Item 30](graph30)