Intelligence 56 (2016) 82–92

Contents lists available at ScienceDirect

Intelligence

Attempting to differentiate fast and slow intelligence: Using generalized item response trees to examine the role of speed on intelligence tests

Jack DiTrapani [a,*], Minjeong Jeon [a], Paul De Boeck [a], Ivailo Partchev [b]

[a] Department of Psychology, The Ohio State University, 1827 Neil Avenue, Columbus, OH 43221, United States
[b] Cito, Amsterdamseweg 13, 6814 CM Arnhem, Netherlands

ARTICLE INFO

Article history:
Received 6 October 2015
Received in revised form 14 February 2016
Accepted 29 February 2016
Available online xxxx

Keywords:
Intelligence
Item response theory
Response time
Item response tree models
Worst performance rule
Elementary cognitive tasks

ABSTRACT

Past research has indicated that a person's speed on cognitive tasks is correlated with his or her intelligence (Sheppard & Vernon, 2007). This has influenced the belief that faster respondents on intelligence tests may be more intelligent than those who are slower. Within this context, previous research has employed a one-parameter item response tree model to intelligence test data and concluded that there are two unique test-taking processes: one process for fast responses, and one for slow responses (Partchev & De Boeck, 2012). This study asks similar questions, but instead uses a two-parameter item response tree model. This model allows the researcher to calculate separate sets of item parameters for when an item is answered quickly versus when it is answered slowly. This item response tree model is fit to 503 respondents to a matrix intelligence test and 726 respondents to a verbal test. Results show that each item has separate parameters for fast and slow responses. Furthermore, for both matrix and verbal tests, the item discrimination parameters are consistently higher for fast responses, suggesting that fast responses to an item may contain more information about the ability of the respondent than slow responses.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

A common topic within intelligence research is the interplay between a person's intelligence and the speed at which one completes mental tasks. The possible relationship between speed and power has been investigated and discussed for decades (e.g., Beck, 1933; Brand, 1981; Eysenck, 1967; Kelley, 1927). Many early theorists reasoned that speed and power are a part of a single latent trait, such as Spearman's *g* (Spearman, 1927). At least semblances of this notion have lasted to today, as speed is often thought of as an indicator of intelligence.

Numerous studies have been conducted on the relationship between IQ scores and mental speed. Many of these investigations look at the speed of participants on cognitive tasks independent of an intelligence test, not the response times of actual intelligence test items. These tasks are often denoted as Elementary Cognitive Tasks, or ECTs. ECTs are not identical to intelligence test stimuli, but are similar and rudimentary enough that it is postulated that they measure basic levels of cognitive ability. Because performance in many of these tasks is measured by speed of completion (e.g. reaction time tests), it is common to compute correlations between intelligence test scores (calculated from a measure independent of the ECTs) and speed levels on ECTs, and therefore attempt to make a link between intelligence scores and cognitive processing speed.

When these relationships are examined, ECT speed and intelligence scores are often positively correlated. This implies that there is a negative correlation between the response time and intelligence (since a higher response time means a lower speed). There is an abundance of research that suggests that the faster the mental processing speed of an individual, the higher the intelligence score of this person (Neubauer & Bucik, 1996; Sheppard & Vernon, 2007; Vernon, 1983). A similarly positive correlation has been found between intelligence scores and inspection time speed. Inspection time is a slightly different concept from reaction time, and is defined as "the quickness of the brain to react to external stimuli prior to any conscious thought." (Kranzler & Jensen, 1989). The distinction between these two types of reaction times is subtle, but meta-analyses have shown that speeds on each of these tasks are both positively related to a subject's measure of intelligence (Kranzler & Jensen, 1989; Sheppard & Vernon, 2007).

One possible interpretation of this evidence is that more intelligent minds are simply "faster" than others, leading to faster response times in elementary cognitive tasks. This has helped fuel the idea that mental speed and mental power are not separate entities but are in fact intertwined as related traits. However, this does not necessarily imply that faster responses on the intelligence test itself leads to better scores.

One result that has been found with these ECTs is the Worst Performance Rule, or WPR (Coyle, 2003). This phenomenon suggests that the slowest (or "worst") responses on ECTs for a given person tend to

* Corresponding author.
E-mail address: ditrapani.4@osu.edu (J. DiTrapani).

http://dx.doi.org/10.1016/j.intell.2016.02.012
0160-2896/© 2016 Elsevier Inc. All rights reserved.

predict that person's intelligence score significantly better than that person's average or fast responses, across many different conditions and tasks (Baumeister & Kellas, 1968; Larson & Alderton, 1990). Extending this framework to intelligence testing, it may follow that, much like the WPR, a respondent's slowest responses to an intelligence test may tell us the most about the intelligence of the respondent. If this were found to be the case, it may be evidence of a connection between intelligence tests and performance on an ECT, suggesting that ECT speed and intelligence test response speed may represent similar aspects of mental speed. Like most ECT studies, WPR analyses are done at the person level, meaning an individual's slowest responses are used. This paper uses an unconventional approach to investigate if the WPR can be found at the item level as well.

ECTs that measure working memory (Baddeley, 1986) have also been used to study the relationship between cognitive speed and a construct of ability. Scores on these working memory tasks have consistently been found to be positively correlated with mental processing speed (Waiter et al., 2009) as well as with intelligence scores (Gray, Chabris, & Braver, 2003; Troche & Rammsayer, 2009), providing a link relating cognitive speed with working memory. However, these studies still do not focus on the response times directly from the intelligence tests themselves. ECT studies provide evidence that those who process information quickly tend to perform better on intelligence tests, but they give us little information regarding the relationship between a respondent's ability and his or her speed on intelligence test items. In this paper, we will investigate responses and response speed directly from the intelligence items, which has not commonly been investigated in ECT and WPR research.

There have been attempts to examine responses and response times not from ECTs but from the actual intelligence items (Davison, Semmes, Huang, & Close, 2011; Van der Linden, 2009). The majority of this research is done on participants under time-pressured conditions; however, there have still been some attempts to look at response times on tests that do not have a time limit. Correlations between the speed of an individual and his or her intelligence scores on a freely-timed test have ranged from negative (Klein Entink, Kuhn, Hornke, & Fox, 2009), to nonexistent (Davison et al., 2011), to highly variable from one test to the next (Van der Linden, 2009).

1.1. Fast and slow: different processes?

In a situation where there is no time pressure, one can postulate several reasons why a response may be quick or slow. One possibility parallels the ECT studies: those with a higher level of intelligence are able to arrive at the correct response at a faster rate. However, one could easily make the counterargument that impatient or unfocused individuals will tend to make rash, quicker decisions, resulting in faster answers but lower scores of intelligence, the opposite result to the first speculation.

Among this confusion, one theory suggests that fast and slow responses may not be a result of faster cognitive speed or levels of impatience, but of the usage of different processes by the respondents. Items may be approached and answered using different strategies, or processes, and these processes may vary in the time it takes to complete them.

Take for an example the math problem shown in Fig. 1. There are potentially two distinct methods of approaching this question. A geometry student with much experience in the subject would immediately and automatically recognize that the answer is 135° (180° − 45°). This approach uses "automatic" processing (Shiffrin & Schneider, 1977), where the respondent quickly and intuitively recognizes how to solve the problem just by initially glancing at the situation. An inexperienced geometry student, or a more thorough one, may instead sequentially work his or her way down to the angle in question. This process uses "controlled" processing, and will tend to be slower. However, it is difficult to say which type of approach is more likely to get the question correct. Those who use automatic processing often get to an answer at a quicker rate and go through fewer mental steps that may result in a mental error, but are more vulnerable to "tricky" items that try to take advantage of a respondent's initial, knee-jerk reaction to seeing the item. A more patient respondent who uses a controlled approach will not fall victim to "trap" items, but may be more likely to make a calculation error, since there are more opportunities to do so. Therefore, we hypothesize that response speed is relevant not because it may show a direct measure of a person's cognitive processing speed, but because it signifies that a different process was used, which may affect how difficult the item is. In essence, the speed of a response may be an indicator of whether a respondent used an automatic or a controlled approach to the item, possibly changing the properties of that item. These ideas are posed and explored by Goldhammer et al. (2014).

Unlike the above example, items from an intelligence test would not realistically have such stark differences in automatic and controlled strategies. However, the logic is that an item can have subtle differences in how it can be approached, leading to differences in the item's parameters dependent on what type of strategy (fast or slow) is adopted. If a fast response is observed, then one could theorize that a more automatic process was used, which may alter the properties of that item (for example, the geometry problem could be more difficult if the automatic process is adopted). Therefore, the strategy that may have been used by the respondent is explicitly included in the data as "fast" or "slow," and is treated as an observed, discrete, and dichotomous variable.

This formulation provides another question: if a given item can have different sets of parameters, can individuals have different levels of abilities on each type of process as well? It is reasonable to think that respondents may have a higher relative ability on faster processes than on the slower responses. This situation would be particularly relevant for intelligence testing, as it would be problematic if a test was attempting to measure one latent ability of respondents (general intelligence), when in fact two separate variables (fast/automatic and slow/controlled intelligence) were being measured.

1.2. Prior research

Partchev and De Boeck (2012) proposed a branching item response model to investigate if respondents have a "fast" and "slow" intelligence. When this model was applied to intelligence data, the model was able to differentiate separate latent abilities for fast and slow responses, meaning each respondent had two different measured latent traits: one for fast responses, and one for slow. Unsurprisingly, these propensities were highly correlated for inductive reasoning data ($r = 0.879$ for verbal analogies, $r = 0.869$ for Raven-like matrix comprehension), meaning that although there were two separate abilities being measured, they were quite similar within a given person. This would suggest that respondents who tended to succeed when answering quickly would be more likely to also succeed when answering at a slower rate.

It was also shown for matrix questions that there was a significant negative correlation between a person's "speed tendency" and inductive intelligence ($r = -0.422$), implying that those who tend to take more time on matrix tasks receive better scores. The speed-ability relationship found here is notably in the opposite direction compared to the aforementioned correlations found in meta-analyses comparing different reaction time task speeds and IQ scores (Kranzler & Jensen, 1989; Sheppard & Vernon, 2007). This finding, which will be explored further in this present study, provides evidence that the positive relationship between reaction time speeds and intelligence scores may only be present when looking at ECT response times, and may not be applicable to response times taken directly from intelligence test responses.

Looking at item parameters yielded similar findings, as each item was discovered to have different difficulties depending on how quickly it was answered. The branching model used in this study is often denoted as an item response tree, or "IRTree" model (e.g., De Boeck & Partchev,

2012; Jeon & De Boeck, in press), and a variation of this approach will be used and described in detail here.

1.3. Aims of this study

In this paper we further investigate the questions posed by Partchev and De Boeck (2012) utilizing an advanced version of the IRTree approach. The most immediate goal is to determine whether person and item parameters are different for fast responses versus slow responses. If the parameters at each speed are distinct, then there is evidence that response time plays a relevant role when an item is answered. A further interpretation could be that respondents approach items using different processes, meaning the speed of a response is highly relevant both to researchers and to test designers. If speed does impact a person's or an item's parameters, a further analysis can examine whether there are *systematic* relationships between the slow and fast parameters of one item/person. For example, we may be interested in discovering whether items are consistently more difficult and/or discriminating when answered quickly, instead of just knowing that they differ in some way.

As this investigation was being conducted, two more specific research questions surfaced. The first of these discusses the implications of treating response times from intelligence test items like they are from ECT reaction time tasks. ECT studies have often concluded that there is a positive link between cognitive power and cognitive speed. Will that relationship be evident when using IRTrees to intelligence item response data? Since we are using response times directly from intelligence test responses, it is unclear whether the positive relationship between speed and power found in ECT studies will be found here as well. This study will investigate whether this speed and intelligence link can be extended to the response times on items *internal* to the intelligence test, which is different from what standard studies using ECTs assess.

The other research question is concerned with the Worst Performance Rule (WPR). WPR studies are done at the within-person level, meaning the slowest response for each *individual* is often found to be highly related to the intelligence of that individual. However, the IRTree model in this study allows the researcher to estimate how relevant (or informative) a particular *item* is to intelligence at different speeds. This will allow us to examine whether slower responses for a given item are more strongly related to intelligence, possibly echoing the WPR conclusion but at the within-item level. We were curious to see if any effects are found using this item-level approach.

2. Methods

2.1. Dichotomization of fast versus slow responses

The data collected contain several different types of information. For every answer of every respondent, the response (either correct or incorrect) and response time (in seconds) is collected. To apply the proposed IRTree method, responses need to be categorically split into either a "fast" or a "slow" label. To do this, we calculate the median response time for each item. If a participant's specific response time to an item is faster than the overall median for that specific item, then that answer is labeled in the dataset as *fast*, and if it is slower than the median time, then it is categorized as *slow*. For instance, consider the data found in Table 1(a). If we wanted to label every response as either fast or slow, we would have to calculate the median response time for each of the three items in our dataset. Those would therefore be 23 s for item 1, 45.5 for item 2, and 71 for item 3. Each specific response time is then compared to the calculated median for that item. So, for subject 1's response to item 1, the answer would be labeled as "slow," since this response time (27 s) is slower than the calculated item 1 median of 23 s. Using this process will result in the labeled data found in Table 1(b). We call this method of dichotomization the *within-item median split*.

Table 1
An example dataset used in this analysis. These example data contain accuracy and response times of three items by three different respondents. An entry of "1" in the "Acc" column signifies a correct response, while an entry of "0" is incorrect. In Table 1(a), the "RT" column shows the amount of seconds needed before that response was entered. In Table 1(b), the response time has been dichotomized into "F" for a fast response and "S" for a slow one.

(a)

Subject	Item					
	1		2		3	
	Rt	Acc	Rt	Acc	Rt	Acc
1	27	1	60	0	79	0
2	19	1	42	1	38	1
3	11	1	49	1	75	0
4	59	1	30	0	67	0

(b)

Subject	Item					
	1		2		3	
	Sp	Acc	Sp	Acc	Sp	Acc
1	S	1	S	0	S	0
2	F	1	F	1	F	1
3	F	1	S	1	S	0
4	S	1	F	0	F	0

This approach has several useful properties. The method of splitting using the item median response time will lead to half of a given item's responses being labeled as fast and half labeled as slow (notice that exactly two responses are labeled as fast and slow for each item in Table 1(b)). However, this will not necessarily be the case for a given person; there are many respondents who will end up with primarily fast or slow responses. For example, note that subject 1 answered slowly for all three items in Table 1(b). We are then able to get an idea of a respondent's tendency to respond quickly or slowly relative to other participants, since all participants' response times are compared to the same median value per item. Therefore, a respondent who has more responses labeled as fast may have a faster tendency than a respondent who has fewer fast responses. This latent person tendency is explicitly modeled in the IRTree approach, which will be discussed in full detail later.

One can think of a different way of splitting item responses into a fast or slow label. For this alternative approach, the median speed for each *person* can be calculated. Each individual person's response times are then compared to that individual's median in order to label responses as fast or slow. We call this approach the *within-person median split*. For example, the median response time for subject 1 in Table 1(a) is 60 s. Using the within-person median split, subject 1's response to item 1 would be labeled as fast because it is faster than the individual median speed of subject 1. If the within-person approach is used, it no longer is possible to measure a respondent's speed relative to other participants, since every participant by definition will have the same amount of fast and slow labeled responses.

There are also technical reasons to avoid the within-person median split in this context. By defining exactly half of a respondent's answers as fast and half as slow, we are effectively setting the variance of the speed tendency across all participants to be zero, since all respondents have the same amount of fast and slow responses. In the two-parameter logistic framework that we are using (this will be described later), this brings computational issues when fitting the model. Specifically, this lack of variability will make it practically impossible for items to discriminate a person's speed

tendency, leading to nonconvergence or extreme estimates of the item discrimination parameters of this model.[1]

It should be noted that typical ECT and WPR analyses are similar in nature to the idea of a within-person median split, in that the notion of "speed" is defined at the *person* level. For example, in WPR analyses, slow or "worst" responses are defined to be the slowest response for each individual, despite the fact that that individual's slowest response could in theory be relatively fast compared to times from other respondents. In other words, in WPR (and ECT) analyses, speed is defined within-person, which is similar in logic to the within-person median split. By using a within-item median split with respect to speed, we take a slightly different approach.

2.2. The IRTree framework

After dividing responses, a proposed IRTree model is ready to be fit to the data, so that sets of parameters can be calculated for both fast and slow responses. This approach enables the researcher to see how an item and a person behave at different speeds. If, for example, it is found that respondents have different levels of ability when answering items at different speeds, it could imply that respondents have two measures of intelligence: one for fast responses (fast intelligence) and one for slow responses (slow intelligence). Similarly, the item difficulty measures could differ depending on the response speed, suggesting that items act differently as a result of response time.

IRTrees are extensions of item response theory, or IRT, which is a statistical item response modeling framework based on the idea that every respondent has a calculated latent trait that explains the person's underlying ability in the given test, and every item has certain fixed parameters that measure properties (such as easiness) of that given test item. In other words, in an IRT model, different items can have different characteristics. Observed responses are then modeled as the result of the interaction between item properties and the respondent's underlying ability. This logic is different from ECT analyses, which treat all items or response stimuli equivalently. When looking at responses and response speeds from intelligence test data, it may be naïve to use approaches that do not model the differences in the items' properties.

In an IRTree, observed responses are explained by a postulated tree that consists of nodes and branches. The choice, or "branch," coming from each node is explained by an IRT model, with sets of item and person parameters (Boeckenholt, 2012; De Boeck & Partchev, 2012; Jeon & De Boeck, in press). An intuitive conceptualization of this model is shown in Fig. 2.

Each circle on this image is a "node," and each line branching off of a node is a "branch," (hence, the name "IRTree"). For every given interaction between participant and item, there are two variables of interest: first, the speed of the response, categorized here as either fast or slow using the within-item median split described earlier. This is represented by the top node, which we will call the "Speed" node. If a participant responds quickly to the item, he or she will then follow the path of the left branch, leading to the "Fast" node. This node corresponds to the actual correctness of the response, given that the item was answered quickly. Similarly, the path traveling to the "Slow" node involves slower responses, and the two "leaves" coming out of this node correspond to a correct or incorrect response, given that the response was slow.

This format therefore contains four possible results for any participant answering any item: a given response can be fast and correct ("leaf" 1 in Fig. 2), fast and incorrect (2), slow and correct (3), or slow and incorrect (4). The practicality of this approach lies in the parameters that are estimated. Unlike a typical IRT model, this method calculates fixed parameter estimates for items and latent traits for persons, at every node. For example, item properties that are estimated for Node 1, or the "Speed" node, correspond with that item's tendency to be answered quickly versus slowly, regardless of whether the item is answered correctly. Similarly, the model calculates a Node 1 latent trait for each person, which measures the speed tendency of the respondent. Unlike Nodes 2 and 3 and most standard IRT models, the response to Node 1 corresponds to whether or not an item is answered quickly; a response of "1" in this case means that the item was responded at a fast rate (as opposed to Nodes 2 and 3, where a "1" response means a correct answer). Therefore, the parameters of Node 1 are concerned with the tendency of the item and person to have a quick response. As we will describe later, the probability that a response is fast can be modeled using these item and person parameters.

The parameters for Nodes 2 and 3 are a little easier to understand, since these nodes are concerned with whether the item is answered correctly. Node 2, or the "Fast" node, specifically models the probability that a response is correct, *given that it was answered quickly.* Therefore, the item parameters estimated at this node are concerned with the difficulty and relevance of that item only when it is answered quickly. Along the same lines, the person traits estimated for this node correspond to the ability of that specific respondent, for fast responses only. The same pattern follows for Node 3 parameters, except this node is for slow responses instead of fast.

This IRTree approach is similar in nature to other methods of jointly modeling response speed and response accuracy, such as Van der Linden (2009)'s hierarchical model and related models (e.g., Klein Entink et al., 2009; Molenaar, Tuerlinckx, & van der Maas, 2015). These models jointly estimate a probit or logit two-parameter item response model for accuracy and a similar two-parameter model for the log response time of respondents (with an identity link function). This

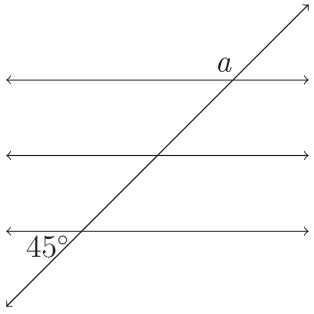


Fig. 1. What is the measure of angle *a*?

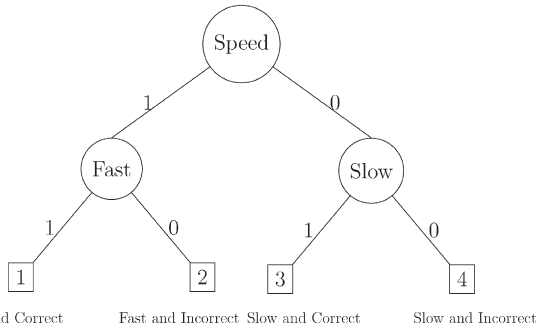


Fig. 2. A visual depiction of an IRTree model. Each node/circle corresponds to a latent ability of a respondent. The probability of each path/branch can be modeled using an item response model, where a response of "1" indicates a fast response if coming from the Speed node, or a correct response if coming from the Fast or Slow node. A "0" indicates a slow response if branching from the Speed node, or an incorrect response if branching from the Fast or Slow node.

[1] As expected, any attempts to run our model with within-person median split data faced issues. Due to the fact that the variance in the speed tendency was essentially zero, the item discrimination parameters (which are described later) carried no meaning, leading to models that either did not converge or had meaningless, extreme parameter estimates. Partchev and De Boeck (2012) were able to examine both the within-person and within-item median splits because the 1 parameter logistic model they used does not have the item discrimination issues that the 2 parameter model has. However, they found equivalent conclusions from both splitting techniques.

joint model estimates a covariance matrix for the ability and "effective speed" of participants, and in some versions also estimates a covariance matrix between item parameters (Klein Entink et al., 2009).

These models estimate person and item parameters, not only in terms of accuracy but also of speed tendency and the dependencies between accuracy and speed tendency. However, they assume that the same strategy is followed across persons and items. The IRTree approach diverges from the joint modeling literature because it looks at the role of speed *within* items and respondents as well as between items and respondents. Parameters for a given item can be estimated for both fast and slow responses to that item, leading to two sets of parameters for all items (parameter estimates for both fast responses and slow responses to that item). The same is done for all individuals, leading to potentially two sets of abilities for a respondent, one for when the person is answering quickly and one for slower responses. Allowing for within-item dimensionality (in the sense that for some respondents the item measures slow ability and for others fast ability) is unique to the IRTree approach.

One of the main specific goals of this study is to see if these Node 2 and Node 3 parameters are similar or systematically different within items or persons. These estimations allow us to investigate whether test-takers have higher latent abilities at certain speeds, or if the properties of items change as response time varies. This means that the ability required to have a 50% chance of getting a certain item correct may be different depending on whether the item is answered quickly or slowly. By calculating separate parameters for fast and slow answers, we are able to observe if there are systematic differences between the two types of responses. This makes it possible to, at the item level, investigate concepts such as the WPR, which states that the slowest responses contain more information about a respondent's intelligence than faster answers.

2.3. 1PL and 2PL IRTrees

The prior research conducted by Partchev and De Boeck (2012) used a one parameter logistic, or 1PL, IRTree to attempt to differentiate slow and fast abilities. This model only calculates one parameter per item, which is the easiness or "intercept" parameter (denoted as "β"). A relevant finding from the previous one-parameter investigation was that the variance estimate for fast latent abilities was noticeably higher than for slow abilities. This leads to the idea that if a two parameter logistic (2PL) IRTree model is fit, the item discrimination values may tend to be higher for fast responses.

For this study, we will attempt to push the IRTree framework further and apply 2PL IRT models within the IRTree context. These models will therefore calculate two properties of every item at every node: a difficulty parameter, which is also calculated in the 1PL model, and a discrimination parameter (denoted as "α"), which is best explained as the "amount of information" that the given item contains about the person trait being measured. This parameter is particularly relevant to our examinations of the WPR at the item level. The 2PL IRTree will allow us to compare a particular item's estimated discrimination parameters for both fast and slow responses, possibly allowing us to determine which type of response speed tends to contain more information about a respondent's intelligence.

This type of question cannot be answered with the 1PL IRTree models that have been used in the past. In the 1PL model, the discrimination parameter is automatically set to be constant across all items (often set to be 1, so that the variance of the random effect can be freely estimated[2]), meaning it is assumed that all items are equivalently related to the latent person trait in question. Because of this, it would be impossible to gather whether fast or slow responses are more related to ability. This model as a result has less parameters and is easier to run, yet does not pick up on subtle differences in the relationship between items and person latent variables. A result of interest for this study is how IRTree models will react to having two item parameters, and whether there will be differences between these models and the conclusions reached with the one parameter models used by Partchev and De Boeck (2012). Due to the evidence that variances for fast abilities are higher than slow abilities when using the 1PL IRTree model, it is hypothesized that the "fast" discrimination values will be higher than those of the same item's slow responses in the 2PL model.

A 2PL IRTree model is explained by the following function:

$$logit\left(P\left(Y_{isp}=1|\theta_{ps}\right)\right)=\alpha_{is}\theta_{ps}+\beta_{is}, \tag{1}$$

where $P(Y_{isp}=1|\theta_{ps})$ is the probability that person p ($p=1,2,\ldots P$) answers with a "1," or "goes left" at node s ($s=1,2,\ldots S$) for item i ($i=1,2,\ldots I$). For this model, the logit link function is used to model the binary response variable. This modeled probability is dependent on the values θ_{ps} and β_{is}, which correspond to the person propensity for person p to "go left" at node s, and the intercept, or "easiness" of item i at node s, respectively. Unlike a 1PL model, this equation also contains the α_{is} term, which is the discrimination parameter for item i at node s.

Note the presence of the s subscript, which does not exist in a standard item response model. This addition is what allows the IRTree model to estimate separate parameters for every item and person at every node. Therefore, at Node 1 (See Fig. 2), θ_{p1} would correspond to the tendency of person p to respond with a quick response (or respond with a "1" at Node 1), and so on. Like in the standard IRT convention, person abilities are treated as random effects, and are assumed to follow a multivariate normal distribution with a mean of zero and a covariance matrix where the variances are constrained to 1 for identification purposes.

A nice characteristic of the model is that the probabilities at Nodes 2 and 3 are independent of the Node 1 probability, conditional on the latent variables (θ_{ps}) for that node. Because of this conditional independence, we can easily calculate a probability for each of the four possible outcomes on the tree ("leaves" 1,2,3, and 4).

Here are the formulas for each of the four possible outcomes:

$$\begin{aligned}P(Fast\ \ and\ \ correct) &= P(Fast)\times P(Correct|Fast),\\ &= P\left(Y_{ip1}=1|\theta_{p1}\right)\times P\left(Y_{ip2}=1|\theta_{p2};Y_{ip1}=1\right),\end{aligned} \tag{2}$$

$$\begin{aligned}P(Fast\ \ and\ \ incorrect) &= P(Fast)\times P(Incorrect|Fast),\\ &= P\left(Y_{ip1}=1|\theta_{p1}\right)\times P\left(Y_{ip2}=0|\theta_{p2};Y_{ip1}=1\right),\end{aligned} \tag{3}$$

$$\begin{aligned}P(Slow\ \ and\ \ correct) &= P(Slow)\times P(Correct|Slow),\\ &= P\left(Y_{ip1}=0|\theta_{p1}\right)\times P\left(Y_{ip3}=1|\theta_{p3};Y_{ip1}=0\right),\end{aligned} \tag{4}$$

$$\begin{aligned}P(Slow\ \ and\ \ incorrect) &= P(Slow)\times P(Incorrect|Slow),\\ &= P\left(Y_{ip1}=0|\theta_{p1}\right)\times P\left(Y_{ip3}=0|\theta_{p3};Y_{ip1}=0\right).\end{aligned} \tag{5}$$

For example, the probability that a respondent answers an item quickly and correctly (denoted as "leaf 1" in Fig. 2) can be calculated by multiplying the estimated probability that the respondent answers quickly, with the probability that he or she answered correctly, given that the response was quick (and the latent ability of the respondent). Eq. (1) can be used to calculate both of these probabilities, where the value of s (the node) is different for each of the two formulas. Since node responses are conditionally independent, we can easily multiply these two probabilities together, like shown in Eq. (2). Another benefit of an IRTree model is that we can estimate correlations between the latent variables that are associated with each node. That way, we can

[2] This is done for model identification purposes. An alternative approach is to freely estimate the variance of the random effect, and fix one of the discrimination parameters to 1.

investigate whether the "speed tendency" of a respondent (θ_{p1} for Node 1) is correlated with performance (θ_{p2} and/or θ_{p3}), making it possible to examine if speed and power are related in some way.

2.4. Data

This current study will explore inductive reasoning intelligence using both verbal analogies and Raven-like matrix comprehension data. The data used here are a subset of two larger datasets, and come from the same "pool" as the data used by Partchev and De Boeck (2012). The matrix and verbal datasets are independent of one another, such that there is no respondent overlap between the two tests. However, the respondents of both data subsets are demographically similar, with ages ranging from 17 to 27 years (median = 20) and with the gender being primarily male. For both of these datasets, the tests were conducted with no tangible time pressure.

The matrix dataset consists of 35 Raven matrix items and a total of 503 respondents. This subset comes from a pool of 30,000 subjects, who were members of the German military, and a random sample of 503 subjects was selected (another subset of this data was used by Klein Entink et al. (2009)). The items are intended to measure the inductive reasoning ability of respondents with nonverbal tasks similar to Ravens matrices. The items used for this specific dataset are directly related to those constructed by Hornke and Habon (1986). More information on these items can be found in Hornke (2002). The test was constructed so that not every respondent answered every item; in fact, each subject only saw 11 items each. These items were answered with an average response time of 68.6 s, and 45.3% of all responses were correct. After the within-item median split was conducted, it was found that 21 respondents (4.2%) answered all items slowly, and 48 respondents (9.5%) answered all 11 items quickly.

The verbal dataset is similar in that it consists of 33 items, although each respondent only interacted with 12 items each.[3] The pool of verbal data contained around 12,000 participants, and the current subset contains a sample size of 726. These responses were answered at a quicker rate than matrix items, with an average response time of 17.7 s. The proportion of correct responses was also relatively higher than matrix responses, at 72.2% correct. After dichotomizing the speed of response, 24 respondents (3.3%) answered all 12 items slowly, and 30 respondents (4.1%) had all fast answers.[4]

2.5. Candidate models

The first model of interest is the model containing two item parameters at all three nodes, as well as three separate person latent abilities. If this model, deemed the "3P3I" model for 3 separate sets of person (3*P*) and item (3*I*) parameters, converges and has the most appropriate fit, it would follow a similar conclusion as 1PL investigations: item properties and person abilities both are differentiated as a result of a fast or slow response. However, we were also open to alternative models in which the fast and slow person abilities and/or item parameters are collapsed to be equal, while other parameters in the model still can be free to differ depending on response speed. Unlike the 3P3I model, these formulations can lead to models in which only two sets of item or person parameters are measured.

Following this logic, four possible model formulations can be analyzed and compared:

1. 3P3I: In this model, both items and persons are differentiated depending on the response time. This conclusion would follow that not only do items behave differently depending on how quickly they are responded, but people have different levels of ability at different speeds, meaning they truly have a "fast" and "slow" intelligence.
2. 2P2I: This model collapses both the fast and slow item and person parameters to be equal, so that there are two sets of person parameters (2*P*) and item parameters (2*I*). This is the most restricted model, and implies that response times do not differentiate either the item properties or the person abilities. This can be used as a baseline model to compare the other formulations with, as it is the theoretically most basic. If this model has the best model fit statistics and is endorsed, it would imply that response speed has little impact on the characteristics of the items or the intelligence of the respondent.
3. 2P3I: Here, items are differentiated by speed, but people are not. This would mean that items have different levels of difficulty and/or discrimination when they are answered quickly, suggesting that there may be two different processes underlying responses. In this formulation, respondents have the same level of ability, regardless of how quickly they answer an item. If this model is endorsed, we can go a step further and examine whether there are systematic relationships between the fast and slow parameters of an item by constraining the fast and slow parameters to differ by just a linear value that is constant across all items. Partchev and De Boeck (2012) conducted these analyses as well, and denoted these linear additional constraints between fast and slow parameters as "main effects." See Section 3.3 for details and the application of this approach.
4. 3P2I: In this model, item properties are identical regardless of response speed, but person intelligence can be split by fast and slow responses.

As a result of using an item-median split to dichotomize responses, the item intercept parameters for Node 1 (the "speed" node) are not useful, since the proportion of fast and slow responses for a given item is set at around 50%. Because of this, we constrained these item parameters to be equal to the logit-scale value that corresponded to the proportion of fast responses for that item.[5] For a portion of these items, this value is zero, since exactly 50% of the responses are defined as fast, and zero is the logit-scale value that corresponds with 50%. Some of these values were set to be slightly positive, as several items had more than 50% of responses defined as fast, because a fast response was any response that was faster or *equal to* the item median. As a result of this definition, many items had more than half fast responses, and the constrained intercept values reflected this. This approach simplified the estimation process, as the model no longer estimated parameters that were unnecessary. Partchev and De Boeck (2012) used a similar method to address this issue.[6]

Even though the intercept parameter is constrained at Node 1, it should be mentioned that the discrimination parameter at this node is still freely modeled. This parameter contains information that can be tricky to initially interpret. Discrimination parameters are by definition the relevance that the specific item in question has towards the latent variable being measured by that item. In this case, the Node 1 discrimination parameters measure the magnitude of the link between the response to this node (fast or slow), and the speed tendency of the respondent. In other words, an item with a very low discrimination

[3] The original dataset had 34 items, but one of the items was removed. Every response that fell into the fast category was correct for this item, making it unreasonable for item parameter estimation. This was the only item, for verbal and matrix data, that had this issue.

[4] Due to the exchangeability of subjects assumed in the proposed models, we can draw relevant inferences on all latent variables (that are associated with each node) even for the subjects who have only slow responses (no observed data for Node 2) or only fast responses (no observed data for Node 3) by "borrowing strength" from other subjects who have relevant information.

[5] This approach is similar to mean-centering the data (and not estimating the intercept) in factor analysis models. There were some concerns that fixing these parameters, however meaningless they seem, would alter the model as a whole and bring unwanted problems. However, the estimated parameters from the models containing constrained β_{i1} (Node 1) values were essentially equivalent to the unconstrained models, with negligible differences that did not change the interpretations of the model. Therefore, due to its better model fit, minor parameter changes, and computational ease, the models with the Node 1 intercept parameter fixed to a constant are used here.

[6] Partchev and De Boeck (2012) approached this problem by fixing all Node 1 intercept values to be equal to zero, which is the logit-scale equivalent to 50%.

estimate at Node 1 would not be effective at discriminating which respondents tend to be fast or slow on other items.

3. Results

3.1. 1PL analysis

All IRTree analyses were conducted with the R package **flirt** (Jeon, Rijmen, & Rabe-Hasketh, 2014), unless otherwise noted. 1PL analyses were first conducted to see if the results were similar in nature to the findings of Partchev and De Boeck (2012). For both matrix and verbal data, the 3P3I model had the best fit by AIC and BIC, echoing the findings of Partchev and De Boeck (2012) and showing that fast and slow intelligence can be differentiated at the item and person levels when using a 1PL approach. This implies that items behave differently when responded to at different speeds. This is potential evidence supporting the notion that typical WPR and ECT analyses should not be conducted with intelligence test responses, since these procedures treat all items as if there are no property differences between items. For example, some intelligence test items may have certain properties (e.g. more difficult concepts) that may lead to slower response times across all respondents. These slow responses would therefore reflect something other than the cognitive ability of the respondent. A WPR analysis, if done in the traditional manner, would possibly be problematic in these types of instances, since a person's slowest response is supposed to be an indicator of worst performance, not of a confounding property of an item. The typical approach to WPR analyses is not necessarily incorrect when dealing with reaction time ECT tasks, since it is reasonable to assume that those stimuli have similar properties. However, when extended to an intelligence test where items have different difficulties and speed tendencies, the analyses could be problematic.

Previous research showed that there was a high correlation between fast and slow latent abilities (0.869 and 0.879 for matrix and verbal, respectively). Our analysis found the same result, except the correlations were even stronger in magnitude (0.931 and 0.966 for matrix and verbal, respectively). Also like the past 1PL study, we found the variance of fast abilities to be higher than that of slow abilities. For matrix data, the past study had a slow ability variance of 1.02 and an estimated fast ability variance of 1.59, while we found corresponding variance estimates of 0.88 and 1.51. For verbal items, Partchev and De Boeck (2012) estimated the variance of slow abilities to be 1.22 and the fast variance to be 2.02, while we found similar values of 1.04 and 1.88. Although similar results are expected since the two studies stem from the same pools of data, it is encouraging to see that two separate datasets yield similar conclusions when a 1PL analysis is performed.

3.2. 2PL analyses: model comparisons

The four candidate 2PL models were then run (3P3I, 3P2I, 2P3I, and 2P2I), and the AIC and BIC of each model, as well as likelihood ratio tests, were used to assess which model had the most appropriate fit to the data. For both matrix and verbal data, the full 3P3I model had worse model fit statistics compared to the competitor models (see Tables 2 and 3). These results suggest that a 2PL IRTree model with three sets of person and item parameters is less appropriate for the data. This already diverges from the main conclusion from using 1PL models, in which the 3P3I model was endorsed.

The other three IRTree models were then run, and the model fit compared (results on Tables 2 and 3). The model that was endorsed was the 2P3I model, which collapses person parameters to be equal across response speeds, but allows item parameters to be different for fast and slow responses. This formulation (2P3I) was endorsed since it had the lowest AIC statistics for both matrix and verbal data. It should be mentioned that the 3P2I model, in which item parameters are collapsed to be equal for fast and slow responses, but fast and slow person abilities are still measured, had the lowest BIC for both matrix and verbal data.

Table 2
Model fit values for matrix data models. "p" denotes the number of parameters in the model. Deviance represents −2×log-likelihood of the model.

Model	p	Deviance	AIC	BIC
3P3I	184	12,976	13,344	14,121
2P3I	180	12,980	13,340	14,099
3P2I	114	13,124	13,353	13,834
2P2I	110	13,680	13,900	14,365

However, the likelihood ratio test shows that including fast and slow item parameters (like the 2P3I does) leads to a significantly lower deviance compared to the less restrictive 3P2I ($\chi^2 = 452$, $df = 62$, $p < 0.0001$). We conclude that the 2P3I model is the most appropriate model for the data.

The 2P3I model freely estimates unconstrained fast and slow sets of parameters for every item. The results suggest that having a separate set of item parameters for fast and slow responses does enhance the model (hence the lower AIC and deviance), but the amount of additional parameters may be inefficient. It may then be speculated that a simplified version of the 2P3I model would be most appropriate for the data. The candidate models would be 2P3I in nature, yet would restrict fast and slow item parameters in different ways such that they are a linear combination of one another. These constrained models will not only be more efficient in their parameter estimation, but they also will allow us to examine how the fast and slow parameters are structurally related to one another.

3.3. Additional analysis: structural relationships between item parameters

The next goal was to dig deeper and check for structural patterns within the fast and slow parameters. This was done by constraining the slow and fast item parameters from the 2P3I model to be different from one another only by a linear combination that is constant across all items. This approach is useful not only because it lessens the amount of parameters in the model and therefore improves parsimony, but also because it examines whether these slow and fast item parameters differ in systematic ways. A consistent difference between fast and slow item discrimination parameters for a given item may have implications for the WPR at the item level, as this parameter quantifies the amount of information the item has when measuring intelligence.

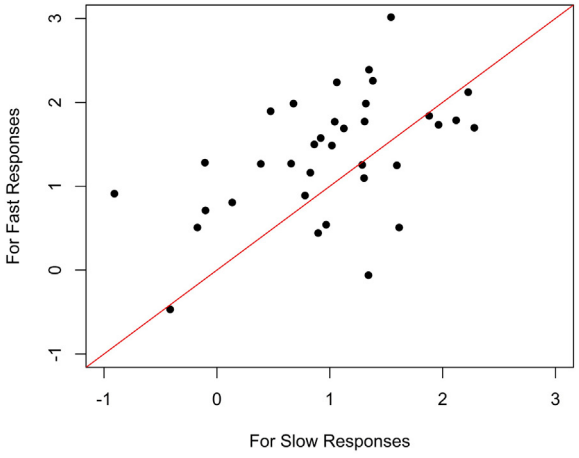
Before we ran any constrained models, we examined plots of these fast and slow parameters to see if there is any visual relationship between them. In Fig. 3 below, the item discrimination and intercept parameters are plotted for the matrix data.

It appears as though the fast discriminations are higher than the same item's slow response discriminations, echoing our hypothesis. The easiness parameters of an item also visually appear to be higher when a response is fast. Fig. 4 shows the same graphs for verbal data, with similar patterns. To investigate these observations quantitatively, constrained models were run for both the verbal and matrix datasets.

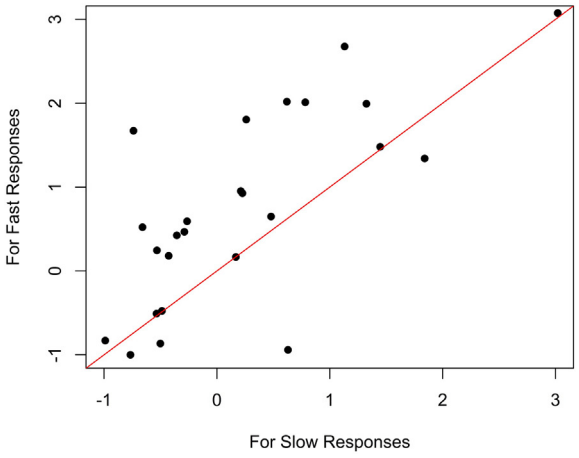
Table 3
Model fit values for verbal data models. "p" denotes the number of parameters in the model. Deviance represents −2×log-likelihood of the model.

Model	p	Deviance	AIC	BIC
3P3I	174	165,85[a]	16,933	17,731
2P3I	170	16,488	16,828	17,608
3P2I	108	16,724	16,940	17,435
2P2I	104	16,832	17,040	17,517

[a] Even though the 3P3I model has more model parameters than the 2P3I, the deviance for the 3P3I ends up being slightly higher than that of the 2P3I model. This occurs because the 3P3I model had some parameter estimates that were deemed extremely large (±15 on the logit scale). This resulted in a slightly altered deviance for this model, and provided further evidence that the 3P3I is probably unreasonable for the data.

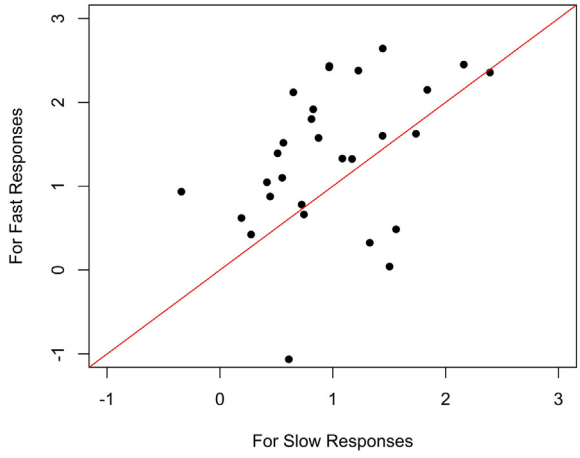


(a) Discrimination parameters for matrix item responses.

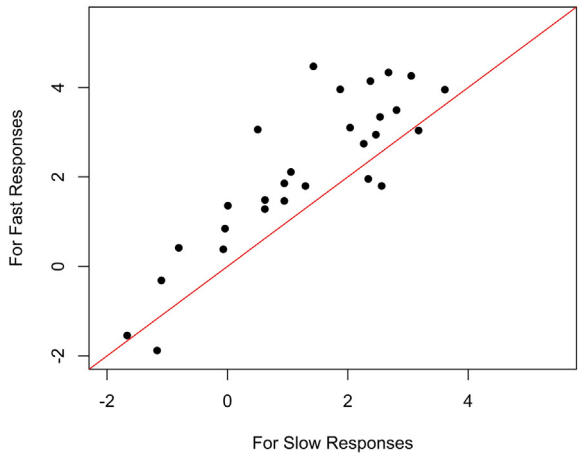


(b) Easiness parameters for matrix item responses.

Fig. 3. Item parameter estimates from the 2P3I model for matrix items. (a) Discrimination parameters for matrix item responses. (b) Easiness parameters for matrix item responses.



(a) Discrimination parameters for verbal item responses.



(b) Intercept parameters for verbal item responses.

Fig. 4. Item parameter estimates from the 2P3I model for verbal items. (a) Discrimination parameters for verbal item responses. (b) Intercept parameters for verbal item responses.

We then fit two restricted 2P3I models, constraining the slow and fast item parameters. The first model (which we will call "*M1*") imposes constraints on the intercept parameters of an item. Specifically, it sets an item's fast and slow intercept parameters to be different only by an additive constant, such that $\beta_i^{fast} = \beta_i^{slow} + c$, where c is the additive constant. Partchev and De Boeck (2012) called this additive constant the *main effect* of the difference between all fast and slow item intercept parameters. Only an additive relationship was explored for the intercept because an additive main effect was of the most interest. We wanted to see if there was a structural, consistent difference among an item's parameters across items of all levels of difficulties.

The second model (which will be denoted "M2") restricts an item's fast and slow discrimination parameters. A discrimination parameter is in practice a multiplicative parameter (recall Eq. 1) that is multiplied by the random effect θ_{ps}, which has a defined distribution. Therefore, discovering that an item's fast and slow discrimination parameters have structurally different values consequently means that the item responses follow different distributions, since the variances of the two types of responses are affected differently when the two distinct discrimination parameters are multiplied with the random person effect. This variance difference is much easier to conceptualize when the differences between the fast and slow discrimination parameters is multiplicative in nature. Therefore, the model in which fast and slow item parameters differ by only a multiplicative constant was run, such that $\alpha_i^{fast} = \alpha_i^{slow} \times w$, where the fast and slow discrimination parameters for item i only differ by the multiplicative constant w. This allows us to say that an item's fast discrimination parameters tend to be w times greater than that same item's slow discrimination values. Therefore, if $w > 1$, the fast discrimination parameter of items tend to be higher compared to a given item's discrimination parameter for slow responses. These models (M1 and M2) were run for both matrix and verbal data, and model fit criteria were compared.

3.3.1. Results: matrix data

All constrained models mentioned above were run and compared using MPlus[7] (Muthen & Muthen, 1998-2011). The AIC and BIC model fit criterion for M1, M2, the unconstrained 2P3I, and the 2P2I model were compared.

For the matrix data, it appears as though both of the constrained models (M1 and M2) have a more reasonable fit than the unrestricted 2P2I and 2P3I models. The AIC and BIC for M1 for this dataset (13,027.5 and 13,311.5, respectively) was lower than those of the 2P3I (13,327.0 and 13,511.2) and 2P2I (13,328.5 and 13,439.8) models. The resulting equation from this restricted model was $\beta_i^{fast} = \beta_i^{slow} - 0.561$ (SE = 0.086). The constant c for was estimated to be -0.561, which is found to be significantly lower than zero ($p < 0.001$). Since β_i is parameterized

[7] MPlus was used here since it can constrain model parameters by linear constants, and then estimate these constants and their standard errors, which is what was needed for these analyses.

here as the difficulty of the *i*-th item, this model suggests that when a given matrix item is answered quickly, it is more likely to be answered correctly. This finding could arguably parallel the theory that items can be answered using different "processes," and that some of these processes are easier than others. The pattern of β_is in this restricted model implies that for matrix items, there may often be a quicker, easier approach to correctly answering a question.

The restricted model constraining discrimination parameters by a multiplicative constant (M2) also seemed to have a more appropriate fit than the unrestricted 2P3I and 2P2I models. The AIC and BIC for M2 (13,297.9 and 13,446.5, respectively) was lower than the aforementioned fit criteria for the 2P3I and 2P2I model, suggesting more appropriate fit. The resulting equation for this model is $\alpha_i^{fast} = \alpha_i^{slow} \times 1.358$ (SE = 0.247). This equation suggests that, like Fig. 3 shows, the discrimination parameters for fast responses tend to be higher than the discrimination parameters for the same item's slow responses, since the value of *w* is greater than 1. This also echoes our original hypothesis that discrimination values would be higher for fast responses, since the variance for fast abilities was higher than for slow abilities when conducting 1PL models. This finding is contrary to the usual conclusion of the WPR, which claims that the slowest responses for a given individual contains more information about the respondent's intelligence.[8]

3.3.2. Results: verbal data

The same approach was taken for the verbal data to explore relationships between slow and fast parameters.

When looking at the model that constrains fast and slow intercept parameters (M1), it is unclear whether or not the additive restriction helps the fit of the model. The unconstrained 2P3I model has a lower AIC (16,812.2 for 2P3I vs. 16,824.2 for M1), but its BIC is worse than the restricted model (17,046.6 for 2P3I vs. 17,013.6 for M1). It is probably wise to be conservative and conclude that there is not enough clear evidence to claim that there is a systematic main effect between a verbal item's fast and slow difficulty parameters. So, unlike the findings with matrix data, there is no apparent structural relationship between the easiness of fast and slow responses for a given verbal item.

However, a conclusion consistent with the matrix data is found when examining the discrimination parameters. The restricted model (M2) has a better fit (AIC = 16,801.7, BIC = 16,991.0) than the 2P2I (16,899.1, 17,040.3) and 2P3I models (AIC and BIC mentioned above), resulting in the linear equation $\alpha_i^{fast} = \alpha_i^{slow} \times 1.500$ (SE = 0.333). Similar to the conclusion reached with the matrix dataset, this equation would imply that an item's fast responses tend to have higher discrimination parameter values than that same item's slow responses. This suggests that a faster response for a given item may tend to contain more information about a respondent's intelligence than a slower response. This is a noteworthy discovery because it further corresponds to the finding by Partchev and De Boeck (2012) that fast responses have higher variance. Since this result is once again contradictory to what would be expected if the WPR was apparent here, this is further evidence that intelligence test responses should not be treated like ECT or WPR tasks.

3.4. Correlations between person parameters

It is also possible to examine the correlations between person parameters of each node, using the constrained 2P3I models. This investigation illustrates the relationship between a given respondent's tendency to answer an item quickly (θ_{p1}) and that same person's latent ability (θ_{p23}, since fast and slow intelligence is collapsed into one trait in these models). This approach is similar to the correlations calculated in ECT analyses, except the IRTree correlations use the latent scores of individuals (for both speed tendency and ability) and not mean or sum scores. This leads to more pure measurements that are measurement error-free. Interestingly, this correlation is strongly negative for respondents on the matrix exam (−0.770 and −0.777 for the restricted intercept and restricted discrimination parameter models, respectively), yet almost nonexistent for verbal data (−0.009 and −0.023 for the unconstrained 2P3I and restricted discrimination models, respectively). This relationship shows that matrix respondents who tend to take more time to answer a question, thus perhaps showing more patience, are more likely to be successful on the given item. This pattern does not emerge for verbal data, implying that a respondent's propensity to be quick has little bearing on his or her ability. These results are starkly different from conclusions made when looking at the response times on ECT tasks. When investigating the response times on intelligence test items, the relationship between speed and power is no longer clearly positive (and may, at least for matrix items, be a negative relationship). Partchev and De Boeck (2012) and Klein Entink et al. (2009) (for matrix items) reported similar findings.

These correlations also create an interesting paradox: at the within-person level (at the item level), a matrix item tends to be easier when answered quickly, yet at the between-person level, respondents who have a higher ability tend to answer at a slow rate more frequently, therefore putting themselves in a more difficult situation.[9] This contradiction seems more coherent when thought through: those who have a higher ability may not answer quickly that often, but no matter how quickly they answer, their fast responses are more successful than their slow responses. Since a person's ability is constant for both fast and slow responses, this "advantage" is constant across both types of responses.

3.5. Worst performance rule investigations

The WPR states that the slowest, or "worst," ECT trial of a given respondent contains the most information, or is most strongly correlated with, the intelligence of that person. However, these analyses are dependent on giving a participant two separate tasks: an ECT, and an intelligence test that is somewhat independent from an ECT. Typically, these analyses are done at the person level: the slowest response for each person is found, and then the correlation between that response time and the same person's intelligence score is calculated. With the item parameters calculated from the IRTree models in this study, we can instead look into the WPR at the item level: do slower ("worse" under the definition of the WPR) responses to an item tell us more about the intelligence of respondents than faster responses? This question is inherently a bit different than a standard WPR analysis: instead of asking whether a person's slowest responses are more strongly related to his or her intelligence, we are asking whether an *item's* slower responses tell us more about the intelligence of respondents.

As previously mentioned, the IRTree models show that, for both matrix and verbal datasets, the discrimination parameters for an item tend to be higher for fast responses than for slow. Given that the item discrimination parameters indicate the amount of information about the latent variable of interest, this goes against what would be expected from the WPR, where slower responses tend to contain more information about a respondent's intelligence. This suggests that, at least when looking at item responses and response times directly from an intelligence examination, the WPR does not seem to exist at the within-item level.

As a follow-up, standard person-level WPR analyses were run in the typical fashion to see if the WPR is present when looking directly at intelligence test data. It was found that the WPR does not apply in this

[8] It is also possible to run models with both intercept and discrimination parameters constrained within the same model. We did run these models, but the detailed results are not reported here, because the conclusions were consistent with the current findings (there were a total of 25 combinations of constraints for each dataset, resulting in a total of 50 possible models to be compared).

[9] This situation is similar in nature to Simpson's paradox, where the within-group trend reverses or disappears when looking at the overall, between-person trend.

case either, at least not for the data used for this study. For matrix data, the correlation between the slowest response time (or "worst," according to the WPR) and that person's sum score on the test is actually positive ($r = 0.363$) and is smaller than the correlation between a respondent's *mean* response time and his or her sum score ($r = 0.435$). These results clash with the general WPR principles in two ways. The first is that the mean response time is actually more strongly related to performance than a participant's slowest time, which is by definition the opposite of what would be expected under the WPR. The other issue is that if the WPR was present with this data, the correlations above would be *negative*, since slower responses are considered "worse", meaning higher slowest response times would be associated with lower sum scores. This finding, and the general correlation provided earlier that slower respondents tend to have higher estimated abilities on matrix tests, suggests that on intelligence tests themselves, slow response times are not indicators of lower ability. They may in fact, for matrix items, indicate higher ability.

4. Discussion

4.1. Conclusions from applying the generalized IRTree model

Applying a generalized two-parameter IRTree model with item discrimination and difficulty parameters to verbal and quantitative intelligence data has yielded some intriguing findings. It appears as though fast and slow item parameters can be differentiated, suggesting that items (both matrix and verbal) behave differently as a result of response speed. For matrix and verbal items, we found that item discrimination parameters are consistently higher when a given item is responded to quickly. This supports Partchev and De Boeck's finding that the variance of fast abilities is higher than the variance of slow abilities. In the 1PL formulation, these high discrimination parameters show up as higher variance at the person level. By applying a 2PL IRTree model here, we were able to model these variances as higher item-level α_i^{fast} parameters.

A possible alternative hypothesis for why item discrimination parameters are higher for fast responses is variance compression. Larson & Alderton (1990) provided evidence that the correlation between reaction time speeds and general intelligence (g) was higher when the variance of the reaction times was higher. In other words, the slowest responses had the highest relationship with g, but also the highest variance. Therefore, if faster responses were found to have higher variability in this study, it could be a concerning explanation of why the discrimination parameters are higher for fast responses.

Fortunately, this does not appear to be the case. The response variability came out to be very similar between fast and slow responses. Across all items from the matrix dataset, the standard deviations of correct responses were 0.495 and 0.500 for fast and slow responses, respectively. For the verbal dataset, these standard deviations were 0.428 and 0.466 for fast and slow responses, respectively (since the responses are binary, these standard deviations are calculated using the proportion of correct responses out of total responses).

The standard deviation for the response speed was much higher for slow responses for both the matrix data (16.69 vs. 62.21 s) and the verbal data (3.81 vs. 18.52 s). After observing the variability for fast and slow responses, it becomes clear that the higher discrimination parameters for fast responses are not an artifact of variance compression.

We also can conclude that response speed impacts the difficulty of matrix items. For matrix items, an item is easier given that it is answered quickly, possibly suggesting that fast responses are an indicator of the recognition of an easier, faster process. This main effect was not found with verbal items, implying that there is no clear evidence that speed influences the easiness of these questions.

At the person level, the 2PL IRTree model did not differentiate fast and slow person abilities. Although this appears different from what was found by Partchev and De Boeck (2012), the fast and slow abilities they found were very highly correlated, leading us to similar conclusions. This conclusion implies that although item properties may change as a result of different response speeds, the person ability used to answer the item is constant regardless of the speed of response (or the process being used). The negative correlation found between response speed and estimated ability on matrix items is also consistent with Partchev and De Boeck (2012)'s 1PL IRTree analyses and the joint modeling procedure conducted by Klein Entink et al. (2009).

There is still much work to be done using IRTree models in substantive contexts. It is hoped that IRTree models start to be employed in other aspects of cognitive and educational research. They are a worthy complement of the speed-accuracy joint modeling frameworks described by Klein Entink et al. (2009) and Molenaar et al. (2015), due to their flexibility and their capacity to investigate measurement invariance at the person level. Measurement invariance is typically examined *between* different groups of people. For example, it is often a topic of interest to study whether an item behaves differently towards individuals of equal ability but from different groups. The models proposed by Klein Entink et al. (2009) and Molenaar et al. (2015) do well at estimating speed tendencies and dependencies of respondents and items, but do not examine how the same item or respondent behaves differently at different speeds. The IRTree approach enables researchers to look for *within-person* measurement invariance. Instead of checking whether different groups react differently to an item, we can now see whether a given item has different parameters when being answered under certain conditions (i.e. a fast or slow response time), resulting in a possible violation of measurement invariance at the person level.

There is also room for improvements in methodologically applying IRTrees to response time data, specifically with the strict dichotomization of speed. Using a median split of response times to define responses as either fast or slow certainly has its limitations, since it results in a given response that is barely faster than the median speed to be treated the same as the fastest response to that item in the dataset. However, the main methodological goal of this study was to extend the 1PL IRTree framework (which used a median split) by applying it to 2PL models. Since preliminary results using this 2PL method have been promising, more complex definitions of fast and slow can be applied in the future. Using a median split gives reasonable results, but a more complex manner of splitting responses should and will be explored. Work is being done on making this dichotomization a stochastic process, one in which the model automatically finds an appropriate cut-off that best fits the data, removing concerns over using a median split. This method will be based on a probabilistic model using latent classes to label a response as fast or slow, similar to the latent classification of ability groups using mixture models (e.g., Reynolds, Keith, & Beretvas, 2010).

4.2. Implications for ECT studies and the worst performance rule

Since this investigation found no positive link between speed and ability, we conclude that there is no evidence to claim that responding quickly on an intelligence test is a positive indicator of ability. ECT studies that have shown that cognitive speed and cognitive power are related use inherently different response time data than the times taken directly from an intelligence test. Our findings do not question those conclusions; they just provide evidence that speed on an intelligence test is not necessarily the same indicator of cognitive speed as speed on an ECT task.

Both traditional person-level WPR analyses (using a person's slowest item response) and item-level WPR analyses (using item discrimination parameters) detected no apparent WPR phenomenon for this dataset. It should be noted that this investigation is not critical of previous findings of the WPR. The WPR explores slow response times on ECTs where speed is explicitly an indicator of ability (such as a reaction time task). The fact that the WPR is not evident when solely looking at response times on intelligence tests is not evidence against the WPR as a whole, it merely implies that researchers should not equate the

response times on intelligence tests to those on ECTs. This suggests that response speed on ECTs and response speed on intelligence test items may reflect somewhat separate aspects of cognitive speed. "Speed = power" is a reasonable conclusion for many ECT and WPR analyses, but should not necessarily be applied to response times directly taken from an intelligence test, as this speed appears to be a different construct from the reaction times on an ECT.

4.3. Future directions

One important characteristic of this study is that it only explores data from a low-stakes test that was collected under conditions that had almost no time pressure. As a result, it would follow that the conclusions herein would not necessarily apply to tests with time or performance pressure, since variables such as order of item presentation and respondent ability would play much larger roles in the speed of response. An interesting follow-up study would be to extend what was done here to data in which respondents were under time constraints and were taking high-stakes examinations. It may also be possible to collect data where respondents self-report what process or approach they took while answering an item, which is information that would be very relevant to the interpretation of these models. Also, item-level WPR analyses, like the ones conducted here but with ECT data, could be an interesting future study.

This paper has extended the analysis of response speed and accuracy to a 2PL IRTree model. There are several reasons why attempting to use a 3 parameter logistic (3PL) IRTree model would be the next logical progression in investigating the relationship between fast and slow item responses. A 3PL IRT model contains a third item parameter, usually interpreted as the "guessing" parameter of an item, since it signifies the logistic probability of a correct response when a respondent makes no effort or has no knowledge relevant to the item. This IRTree extension would therefore have two guessing parameters for every item, one for fast or slow responses. Since many tasks (including the data presented in this study) are multiple choice items, it is likely that many responses are guesses, which the current 2PL model does not consider. True ability may be inaccurately estimated in a 2PL model when guessing is present in the data. It also stands to reason that response time is likely related to a participant's likelihood of guessing, which would be an interesting question that a 3PL IRTree could attempt to answer (Cao & Stokes, 2008). It would be intriguing to examine whether or not a systematic difference exists between the estimated guessing parameters of fast responses and slow responses. However, it should be noted that 3PL models need a very large sample size (larger even than the several hundred participants in our data) in order to effectively estimate parameters (e.g., Swaminathan, Hambleton, Sireci, Xing, & Rizavi, 2003).

This study brings several possible implications to the field of intelligence research. Response time is frequently explored to determine its effect on a respondent's probability of being correct. However, this IRTree approach claims that the properties of the items themselves shift as response time changes, leading to the same item almost behaving as two categorically different items dependent on response time. If ignored, this phenomenon could be problematic for certain tests, since the test would behave differently depending on the response time of the participant, possibly interfering with the measurement process. By modeling the effects of response speed on the properties of an item, we can estimate a more accurate person ability.

References

Baddeley, A. D. (1986). *Working memory*. New York, NY: Oxford Press.

Baumeister, A. A., & Kellas, G. (1968). Distributions of reaction times of retardates and normals. *American Journal of Mental Deficiency*, *72*, 715–718.

Beck, L. F. (1933). The role of speed in intelligence. *Psychological Bulletin*, *30*, 169–178.

Boeckenholt, U. (2012). Modeling multiple response processes in judgment and choice. *Psychological Methods*, *17*, 665–678.

Brand, C. (1981). General intelligence and mental speed: Their relationship and development. In M. P. Freeman, J. P. Das, & N. O'Conner (Eds.), *Intelligence and learning (14, 589–593)*. New York, NY: Springer.

Cao, J., & Stokes, L. (2008). Bayesian IRT guessing models for partial credit guessing behaviors. *Psychometrika*, *73*, 209–230.

Coyle, T. R. (2003). A review of the worst performance rule: Evidence, theory, and alternative hypotheses. *Intelligence*, *31*, 567–587.

Davison, M. L., Semmes, R. X., Huang, L. X., & Close, C. N. (2011). On the reliability and validity of a numerical reasoning speed dimension derived from response times collected in computerized testing. *Educational and Psychological Measurement*, *72*, 245–263.

De Boeck, P., & Partchev, I. (2012). IRTrees: Tree-based item response models of the GLMM family. *Journal of Statistical Software*, *48*, 1–28.

Eysenck, H. J. (1967). Intelligence assessment: A theoretical and experimental approach. *British Journal of Educational Psychology*, *37*, 81–98.

Goldhammer, F., Naumann, J., Stelter, A., Toth, K., Rolke, H., & Kleime, E. (2014). The time on task effect in reading and problem solving is moderated by task difficulty and skill: Insights from a computer-based large-scale assessment. *Journal of Educational Psychology*, *106*, 608–626.

Gray, J. R., Chabris, C. F., & Braver, T. S. (2003). Neural mechanisms of general fluid intelligence. *Nature Neuroscience*, *6*, 316–322.

Hornke, L. (2002). Item-generation models for higher order cognitive functioning. In S. Irvine, & P. Kyllonen (Eds.), *Item generation for test development (159–178)*. Mahwah, NJ: Erlbaum.

Hornke, L., & Habon, M. (1986). Rule-based item bank construction and evaluation within the linear logistic framework. *Applied Psychological Measurement*, *10*, 369–380.

Jeon, M., & De Boeck, P. (2016). A generalized item response tree model for psychological assessments. *Behavior Research Methods* in press.

Jeon, M., Rijmen, F., & Rabe-Hasketh, S. (2014). Flexible item response theory modeling with flirt. *Applied Psychological Methods*, *38*, 404–405.

Kelley, T. (1927). *Interpretation of educational measurements*. Yonkers, NY: World Book.

Klein Entink, R. H., Kuhn, J. T., Hornke, L. F., & Fox, J. P. (2009). Evaluating cognitive theory: A joint modeling approach using responses and response times. *Psychological Methods*, *14*, 54–76.

Kranzler, J. H., & Jensen, A. R. (1989). Inspection time and intelligence: A meta-analysis. *Intelligence*, *13*, 329–347.

Larson, G., & Alderton, D. (1990). Reaction time variability and intelligence: A "worst performance" analysis of individual differences. *Intelligence*, *14*, 309–325.

Molenaar, D., Tuerlinckx, F., & van der Maas, H. L. J. (2015). A generalized linear factor model approach to the hierarchical framework for responses and response times. *British Journal of Mathematical and Statistical Psychology*, *68*, 197–219.

Muthen, L. K., & Muthen, B. O. (1998-2011). *Mplus user's guide*. Los Angeles, CA: Muthen & Muthen.

Neubauer, A. C., & Bucik, V. (1996). The mental speed–IQ relationship: Unitary or modular? *Intelligence*, *26*, 23–48.

Partchev, I., & De Boeck, P. (2012). Can slow and fast intelligence be differentiated? *Intelligence*, *40*, 23–32.

Reynolds, M., Keith, T., & Beretvas, S. (2010). Use of factor mixture modelinf to capture spearman's law of diminishing returns. *Intelligence*, *38*, 231–241.

Sheppard, L. D., & Vernon, P. A. (2007). Intelligence and speed of information processing: A review of 50 years of research. *Personality and Individual Differences*, *44*, 247–259.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, *84*, 127–190.

Spearman, C. (1927). *The abilities of man*. New York, NY: Macmillan.

Swaminathan, H., Hambleton, R. K., Sireci, S. G., Xing, D., & Rizavi, S. M. (2003). Small sample estimation in dischotomous item response models: Effect of prior based on judgmental informationon the accuracy of item parameter estimates. *Applied Psychological Measurement*, *27*, 27–51.

Troche, S. J., & Rammsayer, T. H. (2009). The influence of temporal resolution power and working memory capacity on psychometric intelligence. *Intelligence*, *37*, 479–486.

Van der Linden, W. J. (2009). Conceptual issues in response-time modeling. *Journal of Educational Measurement*, *46*, 247–272.

Vernon, P. A. (1983). Speed of information processing and general intelligence. *Intelligence*, *7*, 53–70.

Waiter, G. D., Deary, I. J., Staff, R. T., Murray, A. D., Fox, H. C., Starr, J. M., & Whalley, L. J. (2009). Exploring possible neural mechanisms of intelligence differences using processing speed and working memory tasks: An fMRI study. *Intelligence*, *37*, 199–206.