Psychology Science, Volume 47, 2005 (2), p. 200 - 229

Decision speed in intelligence tasks: correctly an ability?

VANESSA DANTHIIR[1], OLIVER WILHELM & ANNEKATHRIN SCHACHT

Abstract

Relatively little is known regarding the broad factor of correct decision speed (CDS), which is represented in the theory of fluid and crystallized intelligence. The current study (*N* = 186) examined the possibility that distinct CDS factors may exist that are specific to the broad ability assessed by the tasks from which the correct response latencies are derived, in this instance fluid and crystallized intelligence (Gf and Gc) tasks. Additionally, the relationships between the correct response latencies and Gf, Gc, and processing speed (Gs) were investigated. Two distinct yet correlated factors of CDS were identified for Gf and Gc tasks, respectively. Both CDS factors were related to their ability factor counterparts, and CDS_{Gc} was lowly related to Gs. However, item difficulty moderated the relationships between CDS and the abilities. When item difficulty was considered relative to groups of participants differing in ability level, differences in the speed of responses were found amongst the ability groups. The pattern of differences in speed amongst the ability groups was similar across all levels of item difficulty. It is argued that this method of analysis is the most appropriate for assessing the relationship between ability level and CDS. The status of CDS as a broad ability construct is considered in light of these findings.

Key words: correct decision speed, response latencies, fluid and crystallized intelligence, computerized testing

[1] Vanessa Danhiir, Oliver Wilhelm, and Annekathrin Schacht, Institut für Psychologie, Humboldt-Universität zu Berlin, Germany; email: danthiiv@rz.hu-berlin.de
Correspondence concerning this article should be addressed to Vanessa Danthiir, Institut für Psychologie, Humboldt-Universität zu Berlin, Rudower Chaussee 18, 12489 Berlin, Germany.
The authors gratefully acknowledge the help of Eva Alshut, Marina Chernivsky, Franziska Heckel, Lea Joskowicz, Anna Pohlmeyer, Anne Ruhmland, and Ragna Runkel for their help in collecting the data. We are also grateful to Klaus Oberauer and Ralf Schulze for their helpful comments on a previous version of this manuscript.

Current attention pertaining to individual differences in mental speed focuses primarily upon the relationship between intelligence and reaction time (RT) in elementary cognitive tasks. The finding of a moderate negative relationship between these two constructs is now well established. The ability factor of processing speed (Gs) in the theory of fluid and crystallized intelligence (Gf-Gc theory) also represents the speed with which one responds to cognitive tasks that are so simple that almost everyone would get the answer correct if there were no time limit (see, e.g., Horn & Noll, 1997).

Although quickness of responding correctly in more complex cognitive tasks, such as typical intelligence test items, represents a popularly held notion that speed of solving a problem is indicative of one's intellectual power, this type of speed has arguably received little attention in latter day intelligence research. Despite this fact, a broad ability factor representing this particular type of speed seems firmly entrenched in descriptions of Gf-Gc theory.

(Correct) Decision speed in Gf-Gc theory

A capsule description of the broad ability factor representing speed in responding to complex intellectual tasks is found in most references concerning an account of Gf-Gc theory. The name and definition of such a factor varies somewhat amongst accounts, but it is most typically referred to as Correct Decision Speed (CDS) and described as representing speed in providing correct answers to problems of non-trivial, or moderate, or intermediate difficulty (see, e.g., Horn, 1987; Horn & Noll, 1997; Woodcock, 1998), or in "tasks that require one to think" (Horn & Hofer, 1992, p. 57; see also, Vanderwood, McGrew, Flanagan, & Keith, 2002). It is not clear in some descriptions of CDS that speed in intelligence tasks in general is being referred to, due to the use of the terms "moderate" or "intermediate" difficulty. However, this can be inferred from some of the earlier writings, mentioning this type of speed in more detail, in which variables were formed from either correct or correct and incorrect answers on typical marker tests of Gf and Gc, such as Matrices, Letter Series, Esoteric Analogies, and Vocabulary tasks (Horn, Donaldson, & Engstrom, 1981). Furthermore, Horn and Hofer (1992) state that the problems used when assessing CDS are "of the same kind used to measure Gf or Gc or other cognitive capabilities" (p. 62).

In other writings on Gf-Gc theory, a Decision Speed (DS; Horn, 1998; Horn & Donaldson, 1980) or Quickness in Deciding on Answers (QDA; Horn, 1988) factor is referred to, which does not differentiate between correctness of response. Horn (1988) states that QDA has been found in only three studies, which were reported by Horn, Donaldson, and Engstrom (1981), and summarises the relevant findings. The QDA factor is a second-order factor subsuming two first-order factors, one representing speed in giving correct answers (CDS) and the other indicating speed in giving incorrect answers, which arise from scoring a variety of intelligence tests for the speed of providing an answer (Horn, 1988). Finally, mention is made by Horn and Donaldson (1980) of another study which examined response latencies in intelligence tasks. In the analyses of this study, the speed taken in reaching decisions on three Gf and three Gc tasks was used to form composite variables representing correct decision speed in Gf tasks, correct decision speed in Gc tasks, and a conglomerate of the speed for incorrect answers in both types of tasks.

The main findings from Horn and colleagues' work investigating speed of response in intelligence tests are (see Horn, 1985; Horn, 1988; Horn & Donaldson, 1980; Horn & Hofer, 1992) (a) speed in obtaining correct and incorrect answers correlates positively (e.g., Horn, 1985); (b) both correct and incorrect speed measures and the higher-order factor of QDA correlate only very lowly with Gs and the difficulty with which one can succeed in intellectual problems. The typical correlation for speed in correctly solving answers with level of problem solved is .20, from several studies assessing several intellectual abilities (Horn, 1988); (c) Gs correlates higher with power measures of Gf than does CDS (Horn, 1985); and (d) CDS becomes longer with age yet is not implicated in the aging decline of Gf (Horn, 1988).

Horn, Donaldson, and Engstrom (1981) state that a CDS factor has been found under only a few conditions in a few studies and thus is not well replicated; similarly Horn and Noll (1994) classify this factor's status as "not replicated." Although there are only a limited number of studies that have investigated speed of responding in typical intelligence tasks in a manner similar to Horn and colleagues, few have come close to applying the stringent conditions under which the original studies were carried out. Horn and Hofer (1992) list several precautions that were used when investigating the relationship between CDS and accuracy measures in order to ensure that the two measures are not confounded, thereby producing an artifactually high correlation. These conditions were (a) all participants attempted all items from which speed and level measures were obtained; (b) the items used to measure level are different from those used to measure speed; (c) participants had the option of abandoning an item as opposed to guessing, and the measure of level was the average level of difficulty of correct solutions; and (d) the participant produces an answer rather than selecting it from several options. The last two precautions were supposed to reduce the effects of guessing (Horn & Hofer, 1992).

Theoretical considerations of response speed in intelligence tasks and its relationship to ability

Regarding the substantive nature of CDS, Horn (1988) points to a lack in our knowledge regarding the way findings pertaining to speed should be interpreted, and acknowledges that "the extent to which different forms of speediness represent capacity, stylistic features of personality, and test-taking strategies is not known with any certainty" (p. 668). Indeed, Horn (1987) presents evidence that personality or motivation-like variables, such as persistence and carefulness, influence performance on timed intelligence tests. Jensen's (1982) finding that total time taken on Raven's Matrices correlated negatively and moderately with extraversion, adds support to the notion that personality variables might play some role in speed of response in intelligence tasks.

The expectation, in line with the phenomenon of positive manifold, seems to be that if a relationship between accuracy and speed on intelligence tasks does exist, then more able people will be quicker than those less able. However, Sternberg (1984) emphasises the importance of metacomponents[2] in intelligence and suggests that ignoring metacomponents,

[2] Metacomponents are one type of component identified in Sternberg's componential subtheory of intelligence (Sternberg, 1984).

such as a metacomponent of *allocation of resources*, leads to erroneous conclusions regarding the nature of intelligence. For instance, contrary to the “smart is fast” assumption, Sternberg argues that there are tasks and task components in which being slower is correlated with higher intelligence. For example, more able participants spend more time on global (higher-order) planning in problem-solving (Sternberg, 1981) and more time overall on arithmetic and logic problems requiring “insight” as opposed to just knowledge (Sternberg & Davidson, 1982). Based on such evidence, Sternberg (1984) posits that in some situations, it is speed selection (or resource allocation), rather than speed per se, that differentiates more and less able people. If this is the case, then it is likely that the relationship between ability and speed on intelligence test items will not necessarily be as straightforward as a negative relationship, and may vary depending on what processes are required for task solution and possibly on how difficult the item is.

It seems many different factors may influence the speed at which one responds in intelligence tasks, hence it is unclear as to what correct decision speed in intellectual tasks represents. Of the variables that may be related to response speed, the level of ability has been most focused on. Although there is empirical evidence for many types of intelligence tasks suggesting that speed and accuracy are not highly correlated (Horn & Hofer, 1992), when examining latencies of only correctly solved items, a confound which potentially obscures any relationship which does exist has been noted by several researchers (Frearson, Eysenck, & Barrett, 1990; Jensen, 1982). This confound arises due to the fact that more able people will solve harder items, which require longer solution times[3], thereby increasing their average response time in comparison to less able people. To obviate this confound, Jensen (1982) advocates comparing only response speeds to items that all participants solve. The problem with this approach is that, if focusing on a normal sample ranging in ability level, even though the items are derived from intelligence tasks they are of limited generality because intelligence test items typically range in difficulty from easy to hard.

Another approach to obviate the aforementioned confound would be to examine the relationship between level of ability and speed on items grouped according to difficulty. This has been undertaken in some studies, to be mentioned shortly, with some evidence for item difficulty being a moderating factor of the relationship between item response speed and accuracy. However, considerable problems also exist with this method. One is that the number of higher ability people considered in the relationship will be increasingly greater, proportionate to the number of lower ability people, as the groups of items become harder. Similarly, the average correct response latency will always be based on a greater number of latencies for higher ability people, compared to the number for lower ability people (see also Lohman, 1989), and this discrepancy might be greater for the harder groups of items than the easier groups. Moreover, as the difficulty of the groups of items increases, the likelihood that the score will result from a lucky and not long considered guess also increases with decreasing ability. The end result of a combination of these factors is that an individual’s mean correct response latency score will be decreasingly stable and valid as their ability decreases, and the difficulty of the items increase. Furthermore, although an item’s difficulty is conventionally defined as the proportion of people in some standard group who succeed at the particular item, within a sample, the relative difficulty of an item will vary depending on the

[3] Jensen (1982) notes that, averaged over participants, the correlation between mean item solution time and item difficulty approaches unity.

particular individual's level of ability. Using this method, response speeds are being compared on items that are of differing difficulty for participants, depending on their ability level. Yet, without consideration of item difficulty as relative to ability level, it is impossible to fully determine any effect ability has on how quickly one responds to, for example, an "easy" item. Hence, a finding of a negative correlation between speed on intelligence items and ability can really only be interpreted as "more able people are faster than less able people on items that are easier for the more able people than they are for the less able people."

An approach that addresses all of the above-mentioned issues is to examine different groups of items that are of approximately equal difficulty across groups of participants, these participants being grouped according to level of ability. This approach affords a more rigorous examination of any relationship between speed of correct response and ability, in intelligence tasks, yet it has not been previously undertaken.

Empirical findings concerning response speed in intelligence tasks and its relationship to ability

The current study concerns the status of CDS, along with its relationship to Gf, Gc, and Gs. Relatively few studies exist that examine speed measures derived from intelligence tasks in regards to either factorial studies or their relationship to accuracy measures. Available studies have provided conflicting results, thus the picture regarding the relationship between speed and accuracy measures from intelligence tasks seems still far from clear, and the nature and number of different speed factors that arise from the measures is uncertain. Typically, both speed from correctly and incorrectly answered items have been examined in conjunction, therefore, in the following passages that briefly review some of the main findings, studies that use not only correct response speeds will be considered as well. Findings from factorial studies incorporating speed from intelligence measures will be reviewed first, and findings focusing upon the relationship between speed and accuracy measures from intelligence tasks will be addressed second.

Factorial findings using speed indices

A question that remains to be clarified is how many factors are required to account for the variability amongst decision speed indices derived from intelligence tasks. In earlier writings, Horn and colleagues proposed two composite variables for correct decision speed in Gf and Gc tasks (Horn & Donaldson, 1980), however only one Correct Decision Speed factor is present in contemporary work on the theory. The inference thus made is that CDS generalises across factors. This inference is similarly drawn regarding Carroll's Rate-of-Test-Taking (RTT; R9 in Carroll, 1993) factor, which in his three stratum model is considered a first stratum factor underlying Gs. RTT factors were reported in twelve studies reanalysed by Carroll, and represent the rate at which individuals perform tests of a cognitive nature but "do not appear to be associated with any particular type of test content" (Carroll, 1993, p. 475). The RTT factor seems to be similar to Horn's (1988) DS factor, with both factors not necessarily taking into account whether responses are correct. However, the RTT factor is loaded on by tests of low difficulty; indeed some of the tests are so easy that they are considered Perceptual Speed tests. Further, the notion of RTT loading on a higher-order Gs factor, as reported in Carroll, runs counter to Horn's finding of CDS and Gs being correlated at only a low level. Thus, the RTT factor, as defined in Carroll, could not be considered the same as DS or CDS. Carroll cites numerous other factors that are associated with speed of performance on ability tests, however the majority are related to relatively simple speeded tasks. The one exception relevant to the topic at hand is a Speed of Reasoning factor, for which limited evidence was found by Carroll.

Support for the possibility that more than one factor may exist that spans response speed in intelligence tasks is present in several studies. Kyllonen (1985) addressed the question of the dimensionality of response speeds on intelligence tasks, using a large sample. For a diverse battery of intelligence tasks, with response latencies recorded for the computerised portion, an exploratory factor analysis with oblique rotation indicated reasoning speed, verbal retrieval speed, and numerical-computation speed factors, alongside equivalent accuracy based factors, and Clerical Speed and Technical Knowledge factors. A second-order solution resulted in uncorrelated ($r = -.05$) general speed and general level factors.

Concordant evidence for correlated speed factors, based upon response latency measures from different types of intelligence tasks, is found in Roberts and Stankov (1999). In this study, separate speed factors were found for average time per item taken on visual and auditory intelligence tasks (Tv/a), and inductive reasoning tasks (Tir). The factor scores derived from Tv/a and Tir shared moderate negative (average $r = -.28$) and small positive ($r = .13$) relationships, respectively, with their accuracy based counterparts, and both correlated positively with time measures from Gs tasks ($r = .54$ and .22 respectively). Contrary to Horn and colleagues' findings, Roberts and Stankov reported that the response speeds defining the above-mentioned factors loaded on a second order factor interpreted as Psychometric Speed, together with the response speeds from the Gs tasks. Of note, Roberts and Stankov also reported that the determination and interpretation of these factors remained consistent when accuracy and speed of response were derived from two sets of experimentally independent variables for speed and accuracy, in line with Horn and Hofer's (1992) prescriptions.

The relationship between speed and accuracy at the task level

The studies that have investigated the relationship between accuracy and latency measures from intelligence tasks are conflicting, in that the correlations reported range from moderate and negative to small but positive. However, the majority of investigations have used two distinct types of tasks, that is, Gc and Gf type tasks. The evidence suggests that task type, in this broad sense of the term, may be a moderating factor of both the extent and direction of the relationship found.

For example, with regards to Gc tasks, MacLennan, Jackson, and Bellantino (1988) reported an overall correlation of -.40 between item response latencies from a computerised test of the verbal subscale of the Multidimensional Aptitude Battery (MAB) and accuracy measures from this test and a paper-and-pencil version. This relationship is similar to Rafaeli and Tractinsky's (1991) finding of an average correlation of -.60 between latencies and accuracies derived from a general knowledge task, averaged across two differently timed conditions. The relationship Stankov and Crawford (1997) found between latencies and accuracy from a vocabulary task was in the same direction of these findings, but smaller in magnitude (*r* = -.22).

More disparate results are found for the relationship between speed and accuracy in Gf tasks. For example, Rafaeli and Tractinsky (1991) reported that the average relationship between response speed and accuracy in a mathematical reasoning task, across two differently timed conditions, was negligible but negative. In contrast, Stankov and Crawford (1997) found a correlation of .49 between response latencies and accuracy in an untimed version of Raven's Progressive Matrices. Neubauer (1990) also found a negligible relationship between accuracy and average response latency, on Raven's Advanced Progressive Matrices (APM). However, when the latencies were grouped according to item difficulty, as conventionally defined, a different picture resulted: A moderate negative correlation was found between overall accuracy and latencies for the easy items (*r* = -.44), a near zero correlation for items of medium difficulty (*r* = -.10), and a small positive correlation for the hard items (*r* = .22). Furthermore, reaction time parameters from the Hick task exhibited a tendency of small positive correlations with the easy APM items. These findings suggest that the relationship between accuracy and response latencies on intelligence tasks may be moderated by the difficulty level of the items, and allude to a potential relationship between response latencies on easy Gf items and mental speed, as assessed by RT. Stankov and Cregan (1993) reported a contrasting moderate negative relationship between total time taken on a Letter Series task and accuracy from the APM, administered under stricter than usual time limits. However, Stankov and Cregan also found that the level of difficulty on the Letter Series task moderated this relationship, albeit to a much lesser extent than Neubauer (1990); the correlation decreased, in absolute terms, as the level of difficulty of the Letter Series task increased. This moderating effect of task difficulty on the relationship between response speed and Gf was not found in the studies of Stankov and Crawford (1993) and Stankov (1994).

The evidence reviewed supports the notion that the relationship between response latencies and accuracy on intelligence tests may be moderated by the type of task, in terms of what broad ability is assessed, that the measures are derived from. Overall, negative moderate sized correlations were reported for Gc tasks, whereas for Gf tasks the correlations ranged from negative to positive. The evidence that item difficulty is a moderating factor of the relationship between speed and accuracy is more equivocal and warrants further investigation.

The current study

Several studies exist that have found speed factors based on response latencies derived from intelligence tasks. However, there appears to be none other than the original studies by Horn and colleagues that have found such factors that are based on similar measures from only correctly answered items. Slater (1938) noted that latencies of incorrect and correct solutions have no common basis for comparison. These concerns were echoed by Lohman (1989), for the reason that latencies for error trials are difficult to interpret because the processes underlying them are ambiguous, more so than those underlying correct responses. Thus, it appears that Horn's Correct Decision Speed factor remains as poorly replicated as it was in 1994.

Moreover, from review of the literature it is clear that investigation of the relationship between response speed and accuracy in intelligence test items has primarily been conducted at the task level of analysis. Although this provides useful information as to the possible relationships that exist between speed and accuracy, for certain tasks, the issue could be more informatively addressed at the level of latent variables. Further, item difficulty appears to be a potentially important variable, which if left unconsidered may obscure the true relationship between speed and accuracy.

We consider that the most conceptually and methodologically appropriate way to examine any moderating effect of item difficulty on the relationship between speed of correct response and level, in intelligence tasks, is to take into account the relative difficulty of an item for people of varying ability. However, to fully consider the effect of the possibly complex relationship between speed, ability, and item difficulty, it is necessary to compare results from such an analysis with results obtained when the relationship between CDS and ability is analysed in ways similar to that previously undertaken.

Thus, our aims to address these issues can be considered four-fold and represent a “top-down” approach to examining the nature of correct decision speed and its relationship to other factors. The first aim is to investigate CDS, and the possibility that distinct factors exist based upon the broad ability assessed, using confirmatory factor-analytic methods. More specifically, correct response latencies from Gf and Gc tasks will be examined, as these are the most important main abilities and have been investigated most in the current context.

The second aim is to examine the relationships between any Correct Decision Speed factors that emerge and the abilities of Gf, Gc, and Gs, at the level of latent variables.

The third aim is to assess the moderating effect of item difficulty on any relationship between correct response speed and ability, in a manner similar to that with which a moderating influence has been found previously.

The fourth and final aim is to investigate the effect of item difficulty on the relationship between speed and accuracy, when “calibrating” level of item difficulty for level of ability. This will allow us to examine the relationship between speed of response and ability, when items are of equivalent difficulty for each ability group.

Method

Participants

The participants were 186 (129 females) undergraduate university students and people recruited from the wider community, ranging in age from 18 to 36 years ($M = 25.19$, $SD = 4.78$). Ninety-one of the participants had a high-school degree and 115 were currently enrolled as students at a university. Participants were given either 20€ or course credit for their participation.

Test materials and apparatus

The tests used constitute part of a larger study examining mental speed. The tests examined herein were four tests of Gs, four tests of Gf, and four tests of Gc. All tests were computerised using the program Inquisit (v. 1.33, Millisecond Software) and run using Windows XP, on identical computers with 19 in CRT monitors. For all tasks, response times for each item or trial were recorded to the nearest ms, from the presentation of the stimulus until a response was recorded. Custom-made keyboards, with special response keys, were used for the speed tasks.

Broad Processing Speed (Gs) tasks. All tasks were administered without a time limit. Three of the four tasks are versions of well-known marker tests of Gs, specifically the primary factor of Perceptual Speed. The exception was Symbol Comparison, which was a newly devised task, similar to the Number Comparison task.

1. Finding As. In this task, the stimuli were two syllable nouns, five to eight letters in length. The task was to press either the “yes” key if the word had an “A” in it, otherwise the “no” key. Ten of the words had an “A” in them.

2. Number Comparison (adapted from the WAIS-R; Wechsler, 1981). Stimuli were pairs of digit strings of equal length, 3 to 12 digits long, presented side by side with a set space between them. The task was to press the “same” key if the strings were identical and the “different” key otherwise. The stimuli were balanced so there were an equal number of digit strings of each length, and an equal number of identical and non-identical strings within each string-length. For the non-identical strings, one digit was different in strings of three to five digits in length; two digits in those six to eight digits long; three digits in those nine to eleven digits long; and one, two, or three digits in the longest strings. When more than one digit was different in the non-identical pairs of strings, the digits that were different were always adjacent in the string.

3. Symbol Comparison. This task was the same as the previous task except symbols from the Wingdings font, in Microsoft Windows, were used as stimuli.

4. Digit Symbol (adapted from the WAIS-R; Wechsler, 1981). A code with nine symbols linked to the numbers one through to nine was present at the top of the screen throughout. The stimulus was one of the nine symbols, and the task was to press the number key linked to the symbol.

Digit Symbol comprised 45 trials, and all other tasks consisted of 40 trials. For each task, a practice set of 12 trials preceded the start of the test. The exception was Digit Symbol, for which the practice set comprised 18 trials. Feedback, in the form of “Right” or “Wrong” stimuli, was presented after each practice trial.

The stimuli in both the practice trials and the test proper were balanced so that an equal number of each potential response was required. All trials were presented in a fixed but randomly generated order. For Digit Symbol the response keys used were in the form of a number line at the top of the keyboard. Participants were instructed to leave the index finger of their preferred hand on a “non-functional” home key at the start of each trial, before moving it to the appropriate response key after the appearance of the stimulus. This home key was positioned directly beneath the centre of the number line. For the other tasks, participants left their two index fingers on the left and right response keys relevant to the particular task. Items were presented individually on the screen after one of ten fixed, randomly presented, inter-stimulus intervals ranging from 553 to 1,612 ms, until the participant pressed one of the valid response keys. The first item was presented 1,500 ms after the participant had pressed the key indicating they were ready to start the test.

Fluid intelligence tasks. The number of items in each task is presented in Table 1. The tasks used were (a) an abbreviated version of the APM (Raven, 1962), with the items taken from the odd-numbered items in Set II; (b) the Series Test from Scale 3, Form B, Part 1 of the Cattell Culture Fair Test (CFT; Cattell & Cattell, 1963); (c) Number Series (Wilhelm, 2000), with open-ended response format; and (d) Arrow Series, from the Omnibus Screening Protocol (Roberts & Stankov, 2001), with eight response alternatives. For Number Series, the response keys were the line of number keys on the keyboard. For the other tasks, participants used the mouse to click their chosen response on the screen.

Crystallized intelligence tasks. The number of items used in each task is presented in Table 1. The tasks were a vocabulary test and three general knowledge tests, each presented with five response alternatives. The vocabulary test was taken from Form 1 of the WILDE Intelligence Test (Jäger & Althoff, 1994). The general knowledge test from the IST 2000 R (Amthauer, Brocke, Liepmann, & Beauducel, 2001) was presented as three separate tests, each being approximately equivalent in difficulty and utilising a mixture of symbolic, verbal, and numeric item content. For these tasks, the mouse was used to click the chosen response on the screen.

Procedure

The extended test battery was administered to groups of up to six participants, and lasted approximately 4 hr. The battery was divided into four sessions of approximately 50 min duration, with a break of 10 min between each session. The ability tests were evenly interspersed amongst the speed tasks in each session. Two experimenters were present throughout the entire testing session. Prior to each task, instructions were presented on screen and read verbally by an experimenter at the same time. The instructions for each task stressed to work as quickly and accurately as possible, with the emphasis on either accuracy or speed for the ability and speed tasks, respectively. For the ability tests, several examples were given on the instruction page, and the experimenter worked through these examples with the participants. After completion of the instructions, or the practice trials for the speed tasks, the experimenters then ensured that all participants understood the task requirements before participants simultaneously started the test.

Although we wished to approximate power conditions for the ability tasks as closely as possible, time limits were deemed pragmatically necessary. The time limits imposed for the ability tests are presented in Table 1, and the participants were informed of the time allowed at the beginning of each test. Based on both a pilot session with 15 people, in which no time limits were imposed, and consultation of the relevant test manuals, the time constraints for each test were formed so that they were considered relaxed enough to allow most people to complete the test within the given time (cf. Wilhelm & Schulze, 2002). Further, participants were informed of when there was 2 min till the expiry of the time limit and after this time was completed, the experimenters used their discretion and allowed a few more minutes if people were still working.

Data manipulation

The data were subjected to a number of validity checks and manipulations prior to any analyses. These steps are detailed below, in the order in which they were conducted. All checks and modifications for the tasks were made only on latencies for correct answers, and all following analyses use only these correct response latencies.

Gs Tasks. Data of single trials were first examined and any unfeasibly long latencies that were obvious outliers were discarded. This resulted in only 15 data points being discarded. Next, latencies less than 150 ms, or greater than three intraindividual *SD*s from the individuals' mean on a particular task, were discarded. For each task, individuals' mean latencies were then computed and any outliers were discarded. This resulted in only one participant's mean being discarded for two tasks. These data points were then replaced by the group mean on the particular task. Finally, the mean correct response latencies were inverted according to the formula 1,000 / mean latency in ms, as this transformation results in a less skewed distribution compared to mean RT. The resulting data was then used for all analyses.

Gf and Gc Tasks. For these tasks, the main intention was to purify the correct response latencies from guessing "contamination." In order to approximate this desired result, single trial latencies were examined and any latencies less than 1,000 ms were discarded. Further, any latencies for a single trial that were considered outliers at the lower tail of the distribution, were also discarded. An outlier was determined by a heuristic rather than by a set rule; that is, any latencies that were considerably far away from the main group of latencies at the lower end were discarded, even if it was a group of several latencies. For example, if there were several latencies on a particular trial that were 4,000 or 5,000 ms when the next latencies at the tail-end of the main group were around 10,000 ms, then the lowest were considered outliers. It is acknowledged that this method is subjective and that people can guess at any stage when examining a problem; however, this method was considered preferable to leaving in correct response latencies that appeared likely to be the result of guessing.[4]

Following this procedure, individuals' mean correct response latencies were calculated for each task. Four people had missing values for the mean latency on a task, and these were replaced by the group mean for that task. Our assumption that the time limits provided for the ability tasks were relaxed enough to ensure that most people attempted all items, was supported during the whole procedure, as very few (64 out of 29016) latencies were eliminated.[5]

---

[4] This process serves to highlight some of the considerable problems inherent within the analysis of correct decision speed. The method of encouraging participants to abandon a problem rather than guess, while going some way toward alleviating the problem, does not in any way solve it. For instance, there will still be individual differences in peoples' thresholds for being "reasonably sure that the answer is correct," thus some will provide an answer whereas others may be just as certain (or uncertain) that they have solved the problem yet abandon it. Similarly, some will continue to attempt to determine the solution when they are just as certain (or uncertain) as others, who decide to provide what they think may be the answer.

[5] For Arrow Series six latencies were removed for the last two items, and 13 latencies were discarded for the last CFT item, suggesting that these tests may contain an element of speededness.

Results

Summary statistics for all tasks are presented in Table 1, for both accuracy and latency scores. One observation that is immediately obvious from examination of Table 1 is that the mean latencies for all Gf tasks are longer than the latencies for the Gc tasks. The correlation matrix for all variables is presented in Appendix A.[6]

Table 1:
Descriptive Statistics for all Tasks

Task	No. items	Mean accuracy	*SD* accuracy	Mean latency[a]	*SD* latency[a]	Time limit[b]
Finding As	40	39.48	0.84	1.83[c]	0.25[c]	
Number Comparison	40	38.78	1.16	0.47[c]	0.08[c]	
Symbol Comparison	40	38.53	1.67	0.35[c]	0.06[c]	
Digit Symbol	45	39.07	1.19	0.81[c]	0.15[c]	
APM	15	8.18	3.08	52,361	16,700	17
CFT	13	8.74	1.54	15,770	4,025	4
Number Series	12	8.55	2.91	39,410	12,099	12
Arrow Series	12	7.15	2.20	31,338	10,290	8
Vocabulary	20	12.29	2.90	9,929	2,517	4.5
General Knowledge 1	28	15.83	3.88	10,672	2,187	7
General Knowledge 2	28	17.16	3.99	9,894	2,083	7
General Knowledge 3	28	15.22	3.49	11,014	2,217	7

Note: [a]*Statistics for Gf and Gc item latencies are for correct answers, and are presented in ms.* [b]*Presented in min.* [c]*Based on the transformed score: 1,000 / mean latency in ms.*

Confirmatory Factor Analyses and Structural Equation Modeling

As a first step in the analyses, AMOS (Arbuckle, 1999) was used to estimate measurement models for Gf and Gc, using the accuracy scores, and for Gs, using the inverted mean latencies for correct solutions, to assess the adequacy of these variables as indicators for the broad ability factors. For the Gs model, the residual terms of Number Comparison and Symbol Comparison were allowed to covary, as they would be expected to share residual variance given their similarity. The fit of these models was excellent (Gf: $\chi^2 = 1.2$, $df = 2$, $p = .550$; CFI = 1.000; RMSEA = .000. Gc: $\chi^2 = 0.9$, $df = 2$, $p = .625$; CFI = 1.000; RMSEA = .000. Gs: $\chi^2 = 0.5$, $df = 1$, $p = .465$; CFI = 1.000; RMSEA = .000). Next, all three factors were modeled together, allowing correlations among all factors, to ascertain the adequacy of the "established abilities" side of the model. The fit of this model to the data was also excellent ($\chi^2 = 50.5$, $df = 50$, $p = .453$; CFI = .999; RMSEA = .008).

[6] The means and variance-covariance matrix are also provided in Appendix A, for the purpose of reanalysis.

Two measurement models for CDS were then estimated, using the mean correct response latencies as indicators. In the first model, both the Gf and Gc indicators were specified to load on one CDS factor (χ^2 = 38.7, *df* = 20, *p* = .007; CFI = .922; RMSEA = .071). The second and alternative model specified the Gf indicators loading on a CDS_{Gf} factor and the Gc indicators loading on a CDS_{Gc} factor, with the two factors allowed to correlate (χ^2 = 27.0, *df* = 19, *p* = .105; CFI = .967; RMSEA = .048). The second model with two correlated CDS factors (*r* = .68) provided a better fit to the data than the one-factor model, as indicated by a significant Chi^2-difference test ($\Delta\chi^2$ = 11.7, Δdf = 1, *p* <.001) and a lower value for the Bayes Information Criterion (115.8 vs. 122.3). The second model was thus taken to represent the structure of the CDS variables.

In the final step, all factors, as specified in the previously estimated measurement models, were included in a structural model. Correlations among all factors were freely estimated. Although the fit of the model to the data left room for improvement (χ^2 = 247.6, *df* = 159, *p* = .000; CFI = .915; RMSEA = .055), it was deemed sufficient as we attribute the low value of the CFI to the fact that, overall, the relationships amongst the manifest variables included in the model are not that large (see Appendix A). Hence, this model is not as discrepant from the null model as if the case were otherwise. Correspondingly, only three of the

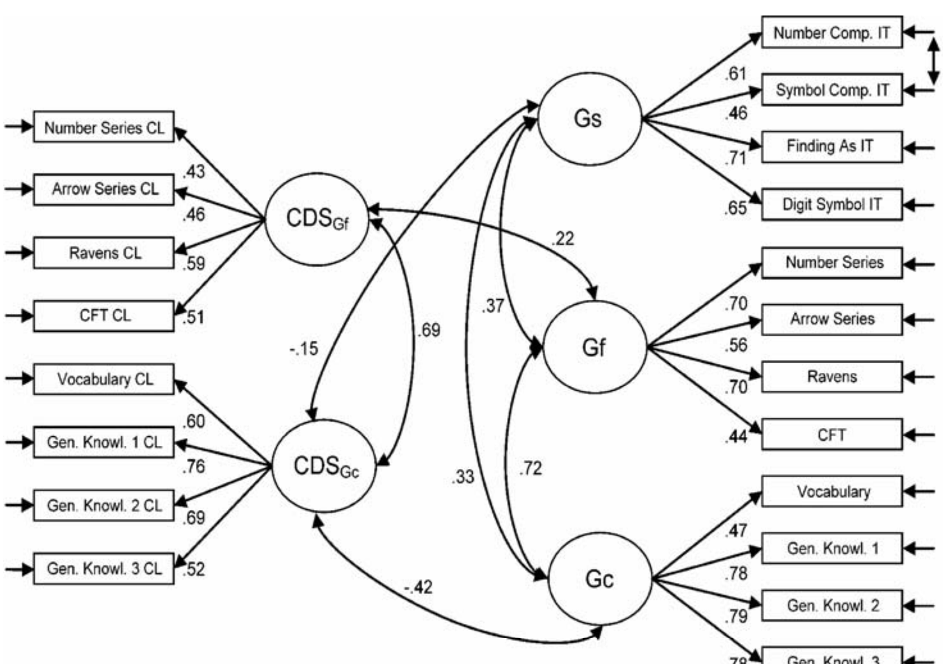

Figure 1:
Structural model of all variables. CL stands for correct latency; IT stands for inverted latency (1,000 / mean latency in ms). The covariances between the residual terms for CFT latency and accuracy (.41), and Arrow Series latency and accuracy (.25), are not depicted.

correlations between the CDS factors and the ability factors were significant; those between the CDS factors and their respective ability factors, and that between CDS_{Gc} and Gs. The non-significant paths were then dropped and the model was re-estimated. Dropping these paths did not harm the fit of the model ($\chi^2 = 252.7$, $df = 162$, $p = .000$; CFI = .913; RMSEA = .055; $\Delta\chi^2 = 5.1$, $\Delta df = 3$, $p = .165$). Based on examination of the modification indices, covariances were allowed between the respective speed and accuracy based residual terms for CFT (.41) and Arrow Series (.25). This was interpreted as reflecting the speededness of these tasks compared to the other tasks. The fit of this model was acceptable ($\chi^2 = 216.6$, $df = 160$, $p = .002$; CFI = .946; RMSEA = .044), and this final model is presented as Figure 1.

The models suggest that CDS is somehow specific to the particular type of ability tasks from which the latencies are measured, that is, Gf and Gc tasks; yet the CDS factors are highly correlated. Suggestions in the literature that there is a differential relationship between latencies and accuracies, dependent on the particular tasks that the measures are derived from, is also borne out at this level of analysis; a moderate sized negative correlation is present between CDS_{Gc} and Gc, whereas a smaller yet positive correlation is found between Gf and CDS_{Gf}. Additionally, the negative correlation evidenced between CDS_{Gc} and Gs suggests that processing speed plays a small role in the speed at which one answers Gc items. However, no speculations will be made as to why differing relationships are found for the two CDS factors with other factors because it is our belief that the analyses undertaken thus far are flawed by not adequately taking into account differences in item difficulty and ability level.

Correlations of correct response latencies and abilities, with difficulty as a moderating factor

Neubauer (1990) and Jensen (1982) suggest a moderating influence of item difficulty on the relationship between (correct) decision speed and accuracy. Our next approach assessed the effect of item difficulty on the relationship between the CDS and ability factors, using a procedure similar to Neubauer's; that is, without taking into account the relative difficulty of an item to people varying in ability.

Items were grouped according to seven levels of difficulty, which are henceforth called "difficulty bins," with difficulty defined as the proportion of people achieving a correct solution on an item. Mean correct response latency scores were then computed for each difficulty bin. This procedure was conducted separately for the 52 Gf items and the 104 Gc items. Table 2 shows the range of item difficulty associated with each difficulty bin, and descriptive statistics for the mean latencies in these bins.

It is evident from Table 2 that although the bins for Gf and Gc are close in difficulty, the latencies for the Gf items are considerably longer than response speeds on the Gc items for bins of equivalent difficulty. However, the finding of increasingly longer response speeds, averaged across individuals, with increasingly difficult items (see Jensen, 1982) holds within the Gc items and Gf items.

To assess the relationship between mean response latencies, for all defined levels of difficulty, and Gs, Gf, and Gc, factor scores were computed from the correlated factors model for these three abilities (see Results section). The correlations between these factor scores and the mean correct response latencies for all difficulty bins are presented in Table 3.

Table 2:
Descriptive statistics for item difficulty bins.

Bin	Item difficulty range[a]	No. items	*N*	*M*[b]	*SD*[b]
Fluid intelligence					
1	p >.895	7	186	11,555	3,237
2	.785 < p ≤ .895	7	185	25,322	9,156
3	.690 < p ≤ .785	7	186	32,634	9,467
4	.595 < p ≤ .690	7	186	38,786	15,128
5	.510 < p ≤ .595	8	181	52,367	21,378
6	.365 < p ≤ .510	8	180	56,618	22,805
7	p ≤ .365	8	153	67,364	32,869
Crystallized intelligence					
1	p > .845	14	186	7,869	1,772
2	.740 < p ≤ .845	15	186	8,972	1,989
3	.630 < p ≤ .740	15	186	10,493	2,832
4	.545 < p ≤ .630	15	186	11,834	3,060
5	.425 < p ≤ .545	15	186	11,968	3,007
6	.305 < p ≤ .425	15	186	12,980	4,467
7	p ≤ .305	15	176	13,508	5,250

Note: [a]proportion of people achieving a correct solution on an item. [b]correct response latencies in ms.

Table 3:
Correlations between Mean Correct Response Latencies for Difficulty Bins and Ability Factor Scores.

Factor	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6	Bin 7
	Fluid intelligence						
Gs	-.22	-.25	-.17	-.08	-.09	.02	.04
Gf	-.26	-.30	-.11	-.05	-.01	.07	.21
Gc	-.20	-.28	-.16	-.05	-.01	.05	.13
	Crystallized intelligence						
Gs	-.25	-.22	-.12	-.05	-.04	-.11	.05
Gf	-.38	-.32	-.21	-.10	-.07	-.13	.11
Gc	-.48	-.44	-.34	-.24	-.08	-.20	.01

Note: Correlations greater than .18 are significant at $p < .01$.

It can be observed from Table 3 that for all relationships between correct response latencies and the ability factor scores, there is a fairly consistent, almost linear, relationship between difficulty and the level of correlation; these results are depicted also as Figure 2, to further illustrate this trend. Mean latencies for items of the lowest difficulty show a negative correlation with Gf, Gc, and Gs, and overall, these correlations decrease as the difficulty level of the items increases, until it reaches either around zero for Gc, and positive values for Gf, for the hardest items.

Focusing first on correct response latencies for Gf, although none of the correlations are large there is a tendency for individuals higher in Gf and Gc to respond faster on easier Gf items, show no difference in speed to those less able on items of medium difficulty, and take more time to respond on the more difficult Gf items. Furthermore, those who are faster in processing speed, as indexed by Gs scores, respond quicker on Gf items at the easier levels of item difficulty. It should be noted that the pattern and size of these relationships, taking

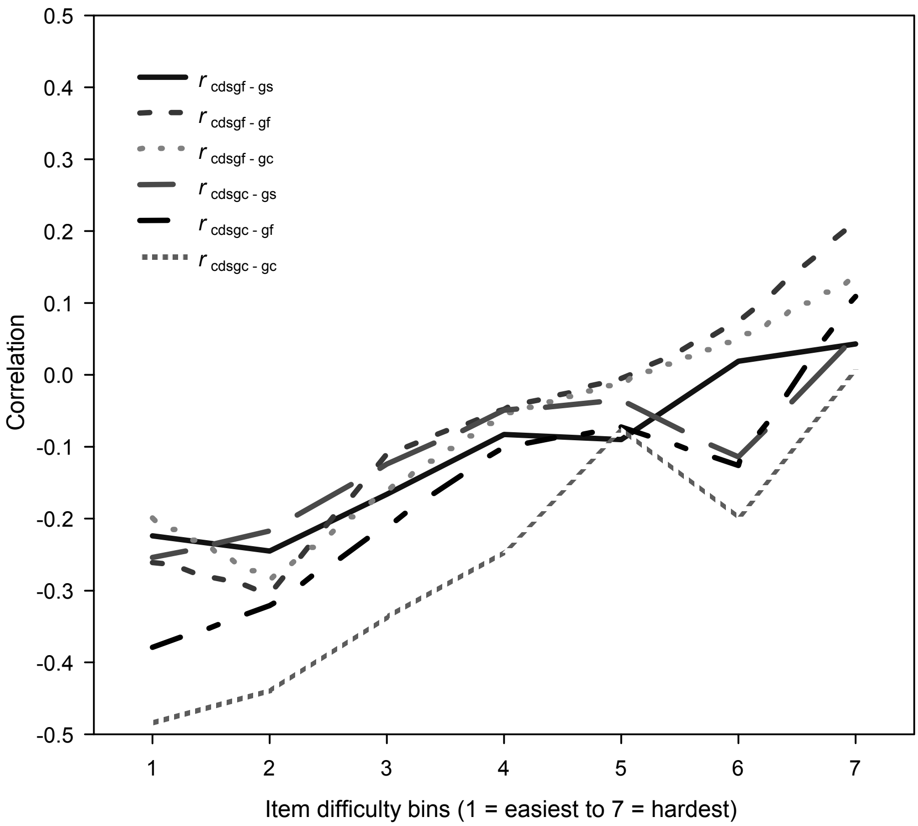

Figure 2:
Correlations between mean correct response latencies for item difficulty bins and Gs, Gf, and Gc factor scores.

into account item difficulty level, replicate Neubauer's (1990) findings, yet with a greater sample size, a greater number of tasks, and with more difficulty levels used in the current study. The finding of the relationship between Gs and the easier items can also be compared to the relationship between response speed on Ravens items and RT, found by Neubauer (see Danthiir, Wilhelm, Schulze, & Roberts, in press).

The correlations between correct response latencies on Gc items and Gs, Gf, and Gc, follow a similar pattern as the correlations evidenced for the Gf response latencies. The results for correctly answered knowledge items also suggest that those higher on Gf and Gc are faster than the less able on easy items, and this difference in speed decreases, overall, with item difficulty, until the more able people end up spending a similar amount of time on the most difficult items. Moreover, the overall magnitudes of the negative correlations are larger than the correlations evidenced with latencies on Gf items. The same pattern, yet with the relationships weaker overall, is found with the correct response latencies and Gs, suggesting that faster processing speed plays a role in fast responses on the easier Gc items.

The results thereby suggest that on Gf and Gc items, more able individuals spend different amounts of time correctly answering items than do the less able, depending on the difficulty of the items. However, as previously mentioned, these analyses are methodologically problematic, and the items composing each difficulty bin are easier for the more able people. Thus, at the current level of analysis it is still not possible to disentangle the relationships between ability, item difficulty, and the speed with which people correctly solve items. We consider that examining mean response latencies for difficulty bins that are of equivalent difficulty for groups differing in ability, will help to obtain a clearer picture of the roles of ability and item difficulty in response speeds.

The relationship between correct response latencies and ability, with item difficulty calibrated for ability

The final step was to examine the relationship between ability and correct response speeds on items of different difficulty levels, with item difficulty calibrated for ability. This allows the disentangling of ability differences and item difficulty, confounded in the previous analysis, and obviates the main methodological concerns. Because, for each task type, a similar pattern of correlations was found when investigating correct response speeds with the level of both Gf and Gc, the relationship between speed and level within the same ability was focused upon, as these relationships were the strongest. To achieve this aim, for Gf and Gc separately, the previously derived factor scores for these abilities were used to divide the sample into equally populated groups of (a) below average, (b) average, and (c) above average ability participants.

Using the seven difficulty bins previously established for Gf items, the difficulty level of all bins was defined, separately for each of the three Gf ability groups, as the mean proportion of correct responses for the items in each bin. For each ability group, four difficulty bins were then found whose difficulty corresponded approximately to the difficulty level of four (different) bins for the other groups. Thus, different items for each ability group constitute a similar level of difficulty. The same procedure was carried out for the Gc difficulty bins and ability groups. Appendix B details the number of items from each test in each difficulty bin, and the corresponding difficulty level for each group. The levels of difficulty for the Gf bins/groups were similar for the Gc bins/groups and are as follows, with the range of difficulty of the bins across groups in parentheses: *Hard* (Gf: .37 - .48, Gc: .35 - .36), *Medium* (Gf: .57 - .60, Gc: .50 - .56), *Mid-easy* (Gf: .75 - .80, Gc: .68 - .72), and *Easy* (Gf: .88 - .91, Gc: .80 - .86).

The descriptive statistics for the mean correct response latencies for the ability groups, across the four equivalent levels of difficulty, are shown in Table 4. To illustrate the differences in response latencies these results are also depicted as Figure 3.

Table 4:
Mean correct response latencies (SD) for ability groups, with item difficulty calibrated for ability.

	Gf item difficulties				*Gc item difficulties*			
Ability group[a]	Easy	Mid-easy	Medium	Hard	Easy	Mid-easy	Medium	Hard
1	12,470 (3,162)	28,820 (10,456)	33,982 (11,666)	40,409 (19,884)	8,530 (1,731)	9,711 (1,878)	11,285 (2,703)	12,068 (3,211)
2	24,845 (8,966)	32,565 (8,325)	53,384 (20,174)	56,856 (21,867)	9,457 (2,066)	11,007 (3,206)	11,983 (2,871)	14,081 (4,776)
3	31,355 (7,931)	35,861 (10,961)	57,428 (17,898)	69,627[b] (28,932)	9,186 (2,025)	10,834 (2,879)	11,628 (3583)	13,256 (3,728)

Note: Statistics reported for Gf item difficulties are with participants grouped according to Gf ability; statistics reported for Gc item difficulties are with participants grouped according to Gc ability. Response latencies are in ms. [a]1 = below average, 2 = average, 3 = above average. [b]N = 61, in all other cells N = 62.

It is evident from examination of Figure 3 that the pattern of relationships that were implied by the correlations between level of ability and mean correct response latency, when difficulty of the items was not calibrated for level of ability, do not apply when difficulty of the items is calibrated for level of ability. Specifically, the calibrated analyses do not suggest that those higher in Gf or Gc are faster on the easier items than the people lower in ability are, nor do these analyses suggest that this difference in speed of response decreases as the difficulty of the items increases.

Repeated measures ANOVAs were conducted to assess the effects of ability and item difficulty on the speed of correct responses, for both Gf and Gc items. For Gf items, there were significant differences between the ability groups, F (2, 182) = 73.39, $p < .001$, partial $\eta^2 = .45$, with the mean latencies increasing from the lowest to the highest ability group. The differences in latencies across levels of item difficulty were also significant, multivariate F (3, 180) = 265.57, $p < .001$, partial $\eta^2 = .82$, with all ability groups increasing the amount of time taken on items as the difficulty level increased. However, the increase in response latencies as the difficulty of the items increased was different across the ability groups, multivariate F (6, 360) = 9.76, $p < .001$, partial $\eta^2 = .14$, although the size of this effect is not large.

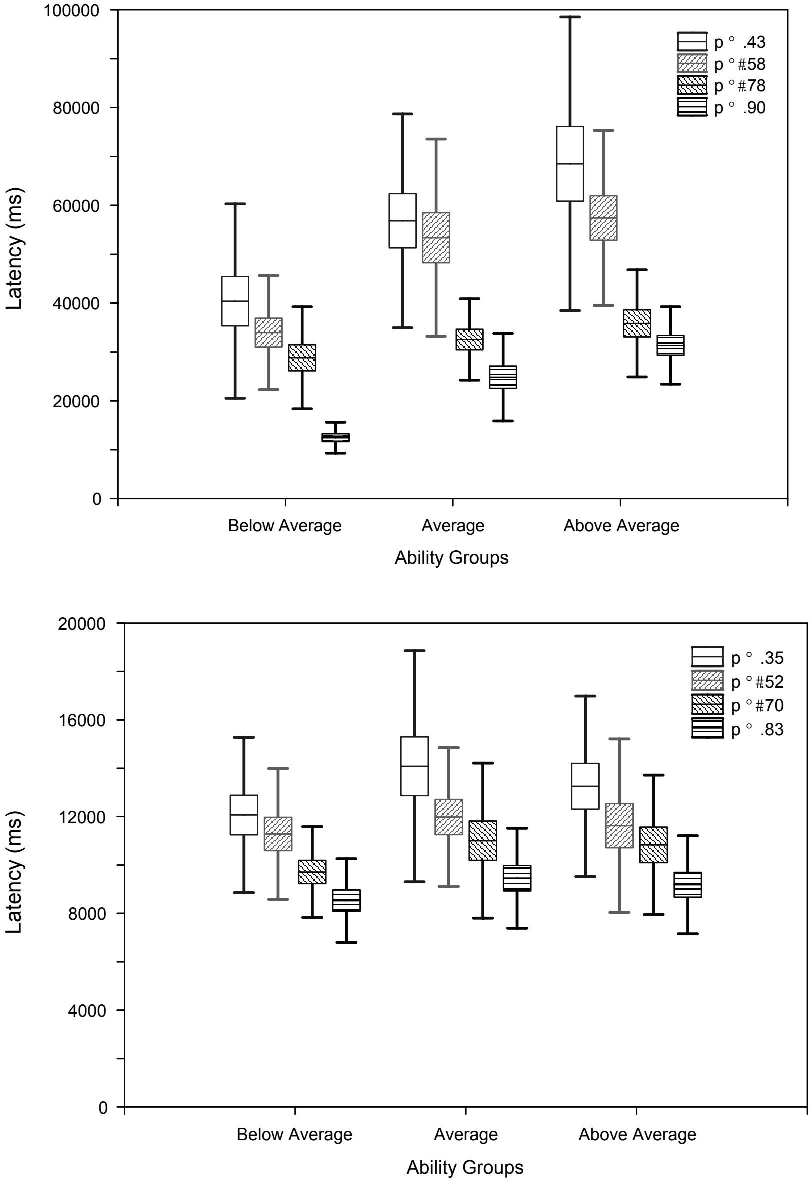

Figure 3:
(Top) Mean correct latency (± 2*SE* [box] ± *SD* [whisker]) for Gf items and ability groups, with item difficulty calibrated for ability. p is the approximate difficulty level of the bins, averaged across ability groups. (Bottom) Mean correct latency (± 2*SE* [box] ± *SD* [whisker]) for Gc items and ability groups, with item difficulty calibrated for ability. p is the approximate difficulty level of the bins, averaged across ability groups.

For Gc items, the differences in latencies across ability groups were also significant, F (2, 183) = 6.50, p = .002, partial η^2 = .07, but the effect was quite small; for these items, the average ability group took the longest time for items at all difficulty levels, and again, the below average group was the fastest at all difficulty levels. The effect of item difficulty was also significant for the Gc items, multivariate F (3, 181) = 97.77, p = <.001, partial η^2 = .62, with the trend of longer latencies as item difficulty increased also evident for each ability group. The interaction of these effects for the Gc items was not significant. Overall, the only substantial sized effects, as evidenced by the partial η^2 statistics, were the effects of Gf ability on Gf item response latencies and item difficulty on Gf and Gc item response latencies. Different relationships were thus revealed using the current approach, as compared to the prior analyses.

Discussion

The overall aim of the current study was to assess the status of a Correct Decision Speed factor or factors, and the relationship between this type of speed and other factors of ability in Gf-Gc theory. Our approach to this intention was top-down and four-fold: (a) To test measurement models of such factors using confirmatory factor-analytic methods; (b) to assess any relationships between CDS and Gf, Gc, and Gs ability factors; (c) to examine a possible moderating effect of item difficulty on any relationships found between CDS and the ability factors; and (d) to assess the extent of any relationship between correct response latencies and ability level for items of differing difficulty, with item difficulty calibrated for ability level.

Regarding these aims, the final model for CDS comprises two correlated factors, CDS_{Gf} and CDS_{Gc}, with loadings from mean correct response speeds from four Gf and Gc tasks, respectively. The establishment of factors of CDS using structural equation modeling (SEM) has not been previously reported. In the final structural model, the relationships indicated for the CDS factors with their respective accuracy based counterparts is concordant with the prior literature (e.g., Stankov & Crawford, 1997) that suggests that the magnitude and direction of the relationship between accuracy and speed is different for Gf and Gc tasks. Moreover, the magnitude of the relationship found between the factors of CDS_{Gc} and Gc is considerably higher than the typically small relationship between CDS and ability reported by Horn and colleagues. However, these results do not adequately assess the relationship between speed and ability in Gf and Gc tasks, because the moderating effect of item difficulty is not considered.

Further evaluation of the relationship between ability level and response times, for items of differing difficulty, led to more complex relationships than those suggested by the modeling of latent variables. The results clearly indicate that the nature and extent of the relationship found between level of ability and correct decision speed, is moderated by item difficulty. When item difficulty is not calibrated for ability level, the relationships between Gf and Gc and their respective correct decision speeds on differentially difficult items, evidence a fairly consistent pattern. This pattern is, on the whole, reflected to a lesser extent in the correlations between the decision speeds with the other factors. The results indicate that those higher on Gf are faster at answering easy Gf items and become slower than those less able, at the hardest difficulty level. The relationships between correct response latencies on

Gc items of varying difficulty and Gc, show a similar pattern to the Gf findings. People higher on Gc are faster at the easier levels of item difficulty, yet essentially no different from those less able, as the items become more difficult. Further, processing speed plays a role in the speed of correct responses at the two easiest levels of item difficulty, for both Gf and Gc items, with those faster at responding on these items being higher on Gs. Importantly, these results are similar to the findings of Neubauer (1990), using Ravens and RT.

However, we argue that the above approach is flawed both methodologically and conceptually, as individual differences in ability are not taken into account in the operationalisation of item difficulty. How difficult the items are for individuals, in each of the difficulty bins used, varies depending on the level of ability of the person, such that items at all difficulty levels are easier for the higher ability individuals than for the lower ability individuals. This leads to the methodological problems mentioned in the introduction, which, at this level of analysis, are likely to result in the correlations for the easier bins being more representative of the actual relationship that exists than the correlations for the more difficult bins. This is considered the cause of the differences in correlations that are found amongst the difficulty bins. Taken together, these issues result in spurious correlations between speed of correct responses and ability. Indeed, when item difficulty is calibrated for ability level, a considerably different picture emerges of the relationship between speed of correct responses to items of varying difficulty levels and peoples' level of ability.

The results of the calibrated analyses show that individuals in each ability group spend more time answering both Gf and Gc items as the difficulty of the item increases. The results also point to differences in time taken, for both Gf and Gc items, across the ability groups. Contradicting the findings when item difficulty is not calibrated for ability, it appears that people who are above average in Gf spend the greatest amount of time on Gf items at all levels of difficulty, and people below average in Gf are, in fact, the fastest in responding at all levels of difficulty. The differences in speed between the ability groups are greatest for the most difficult level of Gf items. For Gc items, with ability indexed by Gc, once again the relationship suggested when item difficulty is calibrated for ability, differs from when it is not calibrated. In the calibrated analysis, the lowest ability group is the fastest and the average ability group is the slowest, at all levels of difficulty. It is important to note, however, that the proportion of variance in the latencies that is accounted for by ability group is considerable for the Gf items (45%) and only small for the Gc items (7%). Thus, too much emphasis should not be placed on the differences in latencies amongst the ability groups, on Gc items. The proportion of variance accounted for by item difficulty is more similar for Gf and Gc items, yet still considerably greater for Gf items (82% vs. 62%).

It thus seems that the relationship between level of ability and speed of correct responses, on intelligence items, does differ between Gf and Gc tasks; however, within both Gf and Gc, the rank order of the speed of ability groups remains the same across items of different difficulty levels. The results are in line with Sternberg's (1984) findings that more able people sometimes take more time than less able on certain tasks and task components; namely, here, on Gf items.[7] However, whether the results support Sternberg's notion that resource allocation, or knowing when to work at what speed, is an important facet of intelligent behaviour,

[7] While the studies supporting this assertion, cited herein, did not consider item difficulty effects, the rationale is lacking that would suggest that the positive relationship between latency and intelligence would be reversed in direction if item difficulty, relative to ability level, were considered.

is another matter. It is important to emphasise that in the calibrated analyses, within all levels of item difficulty, the latencies for the higher ability group(s) are derived from different items than the items from which the lower ability group(s)' latencies are derived: The difficulty, uncalibrated for ability, of the items used for the higher ability group(s) is harder than the difficulty of the items used for the group(s) lower in ability. A main element related to the difficulty level of a reasoning item is the number of cognitive operations required for a correct solution, with harder items requiring more operations (see, e.g., Spilsbury, Stankov, & Roberts, 1990). Consequently, the most simplistic explanation for the finding is that the higher ability group(s) take(s) longer on items at each level of difficulty than the time taken by the group(s) lower in ability, because the items for the higher ability group(s) require more steps to solution, as opposed to the higher ability group(s) working slower per se. This explanation is also consistent with the finding of a different pattern of group differences in latencies for the Gc items, the difficulty of which is not typically related to the number of cognitive operations required to attain a correct solution.

Overall, the conclusion we draw based on consideration of all the analyses is that there is a consistent, yet differential, relationship for Gf and Gc between correct decision speed and the respective ability counterparts, when relative item difficulty is taken into account. Moreover, it is at only this level of analysis that the relationship between ability and correct response speed can be adequately assessed. The correlations that are indicated between the CDS factors and ability factors, in the SEM, result from the spurious correlations that arise between response speed and ability, when item difficulty is not calibrated for ability level. It is thus held that in the correlations that are found between the factors of CDS_{Gf} and Gf, and CDS_{Gc} with Gc and Gs, in the final model, represent an artifactual relationship due to the confounding of ability, speed, and item difficulty, as outlined.

The status of CDS as an ability is problematic for a number of reasons. First, and perhaps foremost, is that the scores appear to confound speed and ability. Second, the averaged speed scores that are used do not comprise the same number of latencies for each person, nor do the averaged speed scores necessarily reflect responses to the same items for each person, even if they are comprised of the same number of latencies. Thus, peoples' mean response latencies do not represent the same information, and to equate them is therefore of very questionable validity (cf. Lohman, 1989). Third and finally, when attempting to assess the construct, even if the measurement intention is made clear to participants[8], that is, presumably to record the maximal speed at which participants can correctly process information, there still exist individual differences in the extent to which people trade speed for accuracy, or vice versa, even if a test is untimed (Lohman, 1989). This effect is not even eradicated by giving bonuses for rapid and accurate performance (Bethell-Fox, Lohman, & Snow, 1984). Thus, it is more than likely that other non-ability variables affect the speed at which people work through intellectual tasks, such as persistence, carefulness, thoroughness, motivation, and test-taking strategy. To what extent these variables influence decision speed is unknown, however, the possibility is widely acknowledged (e.g., Jensen, 1982; Phillips & Rabbitt, 1995; Salthouse, 2000; Stankov, Boyle, & Cattell, 1995).

[8] Arguably, more often than not it seems speed measures from intelligence tasks are collected merely as incidental information, and sometimes participants are not even instructed to work quickly, nor informed that their speed of response is being recorded.

It therefore seems that CDS probably reflects a complex interplay of all the different factors we have mentioned, rather than reflecting simply some presumably physiologically-bound speed at which a person can reason, or think through, to a correct solution. However, despite the fact that it is unclear what correct decision speed in intellectual tasks represents, CDS has entered into the tables of cognitive factors and abilities, and descriptions of intelligences, that are based on Gf-Gc theory (see, e.g., Flanagan & McGrew, 1998; Horn, 1998; Horn & Hofer, 1992; Horn et al., 1981; Vanderwood et al., 2002; Woodcock, 1998).

Although we consider CDS' status should not be consolidated as an ability or abilities, this does not imply that we believe speed scores from intelligence tasks are not interesting or useful to examine further. One interpretation put forward to explain the considerably large differences in latencies amongst the ability groups on Gf items, is linked to the number of cognitive operations required to reach a correct solution. However, the result is open to a number of other explanations. One of these is that the findings reflect the suggestion that people of differing levels of ability use different strategies when answering intelligence test items. For instance, that the below average ability group was fastest at all levels of task difficulty, for both Gf and Gc items, is concordant with the finding that even on untimed tasks less able individuals follow a self-imposed deadline on complex items (Kyllonen, Lohman, & Snow, 1984). On the other hand, that the above average ability group took the longest on all Gf items could reflect a greater level of persistence and preference for accuracy over speed, perhaps due to a greater confidence in their ability to solve most items. Both these possibilities are in line with the idea that those differing in intelligence allocate their resources differently.

One avenue which may shed some light on the possible factors influencing decision speed on intelligence tasks is to place the same time limit on each item, for items of equivalent difficulty, for more and less able people. If this decreases the level of correctness of the more able people, then this would support the interpretation that the more able people took more time on Gf items due to the greater number of steps required to solve the item. Another possible avenue is to investigate the role of metacognitive factors in how much time people spend on items. For instance, several studies report the existence of a confidence factor, which appears to mediate individuals' accuracy of self-assessment on cognitive tasks (e.g., Kleitman & Stankov, 2001; Pallier et al., 2002). Recent evidence suggests that, independent of accuracy, those who are more confident of their success on intelligence test items are also faster in responding (Preckel & Freund, this volume). It seems plausible that metacognitive factors may also moderate how individuals adjust the time they spend on items as the relative difficulty increases or decreases. It would also be of interest to see if instructions emphasising either, for example, typical or maximal speed of response, changed the rank order of people in terms of CDS and corresponding ability factors.

References

1. Amthauer, R., Brocke, B., Liepmann, D. & Beauducel, A. (2001). Intelligenz-Struktur-Test [Intelligence-Structure-Test] 2000 R. Göttingen: Hogrefe.
2. Arbuckle, J. L. (1999). AMOS 4.0 users' guide. Chicago, IL: SPSS Inc.
3. Bethell-Fox, C. E., Lohman, D. F., & Snow, R. E. (1984). Adaptive reasoning: Componential and eye movement analysis of geometric analogy performance. Intelligence, 8, 179-203.

4. Carroll, J. B. (1993). Human cognitive abilities: A survey of factor-analytic studies. Cambridge: Cambridge University Press.
5. Cattell, R. B., & Cattell, A. K. S. (1963). The Cattell Culture Fair Test. Champaign, IL: IPAT.
6. Danthiir, V., Wilhelm, O., Schulze, R., & Roberts, R. D. (in press). Factor Structure and Validity of Paper-and-Pencil Measures of Mental Speed: Evidence for a Higher-Order Model? Intelligence.
7. Flanagan, D. P., & McGrew, K. S. (1998). Interpreting intelligence tests from contemporary Gf-Gc Theory: Joint confirmatory factor analysis of the WJ-R and KAIT in a non-white sample. Journal of School Psychology, 36, 151-182.
8. Frearson, W., Eysenck, H.-J., & Barrett, P. T. (1990). The Furneaux model of human problem solving: Its relationship to reaction time and intelligence. Personality and Individual Differences, 11, 239-257.
9. Horn, J. L. (1985). Remodeling old models of intelligence. In B. B. Wolman (Ed.), Handbook of intelligence: Theories, measurement, and applications (pp. 267-300). New York: John Wiley & Sons.
10. Horn, J. L. (1987). A context for understanding information processing studies of human abilities. In P. A. Vernon (Ed.), Speed of information-processing and intelligence (pp. 201-237). Norwood, NJ: Ablex Publishing.
11. Horn, J. L (1988). Thinking about human abilities. In J. R. Nesselroade & R. B. Cattell (Eds.), Handbook of multivariate experimental psychology (pp. 645-685). New York: Plenum Press.
12. Horn, J. L. (1998). A basis for research on age differences in cognitive capabilities. In J. J. McArdle & R. W. Woodcock (Eds.), Human cognitive abilities in theory and practice (pp. 57-91). Chicago: Riverside Press.
13. Horn, J. L., & Donaldson, G. (1980). Cognitive development in adulthood. In O. G. Brim & J. Kagan (Eds.), Constancy and change in human development: A volume of review essays (pp. 445-529). Cambridge, MA: Harvard University Press.
14. Horn, J. L., Donaldson, G., & Engstrom, R. (1981). Apprehension, memory, and fluid intelligence decline in adulthood. Research on Aging, 3, 33-84.
15. Horn, J. L., & Hofer, S. M. (1992). Major abilities and development in the adult period. In R. J. Sternberg & C. Berg (Eds.), Intellectual development (pp. 44-99). New York: Cambridge University Press.
16. Horn, J. L., & Noll, J. (1994). A system for understanding cognitive capabilities: A theory and the evidence on which it is based. In D. K. Detterman (Ed.), Current topics in human intelligence: Volume IV. New York: Springer-Verlag.
17. Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: Gf-Gc theory. In D. P. Flanagan & J. L. Genshaft (Eds.), Contemporary intellectual assessment: Theories, tests, and issues (pp. 53-91). New York: Guilford Press.
18. Jäger, A. O., & Althoff, K. (1994). Der WILDE-Intelligenz-Test [The WILDE Intelligence Test] (2nd rev. version). Göttingen: Hogrefe.
19. Jensen, A. R. (1982). Reaction time and psychometric g. In H.-J. Eysenck (Ed.), A model for intelligence (pp. 93-132). Berlin: Springer-Verlag.
20. Kleitman, S., & Stankov, L. (2001). Ecological and person-oriented aspects of metacognitive processes in test-taking. Journal of Applied Cognitive Psychology, 15, 321-341.
21. Kyllonen, P. C. (1985). Dimensions of information processing speed. Brooks Airforce Base, TX: Air Force Systems Command, AFHRL-TP-84-56.

22. Kyllonen, P. C., Lohman, D. F., & Snow, R. E. (1984). Effects of aptitudes, strategy training, and task facets on spatial task performance. Journal of Educational Psychology, 76, 130-145.
23. Lohman, D. F. (1989). Individual differences in errors and latencies on cognitive tasks. Learning and Individual Differences, 1, 179-202.
24. MacLennan, R. N., Jackson, D. N., & Bellantino, N. (1988). Response latencies and the computerized assessment of intelligence. Personality and Individual Differences, 9, 811–816.
25. Neubauer, A. C. (1990). Speed of information processing in the Hick paradigm and response latencies in a psychometric intelligence test. Personality and Individual Differences, 11, 147-152.
26. Pallier, G., Wilkinson, R., Danthiir, V., Kleitman, S., Knezevic, G., Stankov, L., et al. (2001). The role of individual differences in the accuracy of confidence judgments. The Journal of General Psychology, 129, 257-299.
27. Phillips, L. H., & Rabbitt, P. M. A. (1995). Impulsivity and speed-accuracy strategies in intelligence test performance. Intelligence, 21, 13-29.
28. Rafaeli, S., & Tractinsky, R. F. (1991). Time in computerised tests: a multitrait, multimethod investigation of general-knowledge and mathematical-reasoning on-line examinations. Computers in Human Behavior, 7, 215–225.
29. Raven, J. C. (1962). Advanced Progressive Matrices. London: H. K. Lewis.
30. Roberts, R. D., & Stankov, L. (1999). Individual differences in speed of mental processing and human cognitive abilities: Towards a taxonomic model. Learning and Individual Differences, 11, 1-120.
31. Roberts, R. D., & Stankov, L. (2001). Omnibus Screening Protocol. Sydney: E-ntelligent Testing Products.
32. Salthouse, T. A. (2000). Aging and measures of processing speed. Biological Psychology, 54, 35-54.
33. Slater, P. (1938). Speed of work in intelligence tests. British Journal of Psychology, 29, 55-68.
34. Spilsbury, G., Stankov, L., & Roberts, R. D. (1990). The effect of a test's difficulty on its correlation with intelligence. Personality and Individual Differences, 11, 1069-1077.
35. Stankov, L. (1994). The complexity effect phenomenon is an epiphenomenon of age-related fluid intelligence decline. Personality and Individual Differences, 16, 265-288.
36. Stankov, L., Boyle, G. J., & Cattell, R. B. (1995). Models and paradigms in personality and intelligence research. In D. K. Saklofske, & M. Zeidner (Eds.), International handbook of personality and intelligence (pp. 15-43). New York: Plenum Press.
37. Stankov, L., & Crawford, J. D. (1993). Ingredients of complexity in fluid intelligence. Learning and Individual Differences, 5, 73-111.
38. Stankov, L., & Crawford, J. (1997). Self-confidence and performance on cognitive tests. Intelligence, 25, 93-109.
39. Stankov, L., & Cregan, A. (1993). Quantitative and qualitative properties of an intelligence test: Series completion. Learning and Individual Differences, 5, 137-169.
40. Sternberg, R. J. (1981). Intelligence and nonentrenchment. Journal of Educational Psychology, 73, 1-16.
41. Sternberg, R. J. (1984). Toward a triarchic theory of human intelligence. The Behavioral and Brain Sciences, 7, 269-315.
42. Sternberg, R. J., & Davidson, J. E. (1982). The mind of the puzzler. Psychology Today, 16, 37-44.

43. Vanderwood, M. L., McGrew, K. S., Flanagan, D. P., & Keith, T. Z. (2002). The contribution of general and specific cognitive abilities to reading achievement. Learning and Individual Differences, 13, 159-188.
44. Wechsler, D. (1981). Manual for the Wechsler Adult Intelligence Scale – Revised. New York: The Psychological Corporation.
45. Wilhelm, O. (2000). Psychologie des schlussfolgernden Denkens: Differentialpsychologische Prüfung von Strukturüberlegungen [Psychology of reasoning: Testing of structural theories]. Dr Kovac: Hamburg.
46. Wilhelm, O., & Schulze, R. (2002). The relation of speeded and unspeeded reasoning with mental speed. Intelligence, 30, 537-554.
47. Woodcock, R. W. (1998). Extending Gf-Gc theory into practice. In J. J .McArdle & R. W. Woodcock (Eds.), Human cognitive abilities in theory and practice (pp. 137-156). Chicago: Riverside Press.

For appendixes see the next pages.

Appendix A:
Means, variance-covariance, and correlation matrix for all variables

Variable	1	2	3	4	5	6	7	8	9	10
1. Finding As IT	0.0631	0.0091	0.0049	0.0160	0.1216	0.0731	0.0536	0.0953	0.1010	0.1949
2. Number Comparison IT	.469	0.0059	0.0032	0.0044	0.0125	0.0230	0.0107	0.0132	-0.0043	0.0198
3. Symbol Comparison IT	.334	.711	0.0034	0.0027	0.0120	0.0209	0.0055	0.0073	0.0042	0.0127
4. Digit Symbol IT	.433	.388	.314	0.0216	0.0916	0.0523	0.1122	0.0831	0.0222	0.1130
5. Ravens	.157	.053	.067	.203	9.4765	1.1875	4.4958	2.5795	2.6086	5.0784
6. CFT	.189	.194	.233	.231	.250	2.3789	1.3737	0.9682	0.5836	1.4099
7. Number Series	.073	.048	.032	.263	.502	.306	8.4647	2.6273	2.4792	4.3414
8. Arrow Series	.173	.078	.057	.257	.381	.285	.411	4.8383	1.2935	2.2135
9. Vocabulary	.138	-.019	.025	.052	.292	.130	.293	.203	8.4319	4.1829
10. Gen. Knowledge 1	.200	.066	.056	.198	.425	.235	.384	.259	.371	15.0694
11. Gen. Knowledge 2	.241	.088	.088	.245	.434	.276	.439	.423	.403	.642
12. Gen. Knowledge 3	.188	.096	.051	.279	.423	.246	.457	.354	.372	.642
13. Ravens CL	.091	.015	-.066	.054	.247	.150	.099	.093	.084	.181
14. CFT CL	-.080	.032	-.065	.002	.075	.366	.024	-.039	.029	.074

Note: Covariances are in the upper triangle, correlations in the lower triangle, and variances are on the main diagonal. IT = inverted latency: 1,000 / mean latency in ms. CL = Correct latency in seconds.

Appendix A:
Means, variance-covariance, and correlation matrix for all variables (continued)

Variable	11	12	13	14	15	16	17	18	19	20
1. Finding As IT	0.2411	0.1651	0.3826	-0.0804	-0.6757	-0.1067	-0.0427	-0.0766	-0.0753	-0.0478
2.Number Comparison IT	0.0270	0.0259	0.0193	0.0098	-0.1433	-0.1303	0.0098	-0.0289	-0.0199	-0.0148
3. Symbol Comparison IT	0.0204	0.0103	-0.0642	-0.0153	-0.0926	-0.0759	-0.0078	-0.0173	-0.0160	-0.0108
4. Digit Symbol IT	0.1432	0.1433	0.1317	0.0013	-0.4078	0.0909	0.0147	-0.0342	-0.0893	0.0017
5. Ravens	5.3235	4.5454	12.6865	0.9232	-3.9968	1.8003	0.0272	-0.5764	-0.5211	0.1109
6. CFT	1.6954	1.3272	3.8544	2.2708	-1.4506	1.6815	0.0889	-0.1408	0.0173	0.0026
7. Number Series	5.0970	4.6478	4.7901	0.2796	-3.4472	4.9127	-0.2229	-0.9275	-0.7936	0.1762
8. Arrow Series	3.7124	2.7200	3.4300	-0.3409	-5.3516	4.6535	-0.3429	-0.8346	-0.7920	0.1881
9. Vocabulary	4.6634	3.7708	4.0964	0.3336	-2.7105	3.1725	-1.3695	-1.0400	-0.3787	-0.7778
10. Gen. Knowledge 1	9.9396	8.7063	11.7630	1.1608	-6.7269	-0.7677	-1.7096	-2.4469	-2.4743	-1.0705
11. Gen. Knowledge 2	15.8945	8.3555	3.6598	-0.4065	-8.1776	0.4625	-2.7093	-2.5086	-1.7805	-1.0818
12. Gen. Knowledge 3	.600	12.2022	8.9770	0.7408	-9.7398	2.1509	-1.5441	-2.2515	-2.5152	-0.9306
13. Ravens CL	.055	.154	278.9019	23.2891	41.9248	44.7908	11.2584	8.7029	5.2923	6.8716
14. CFT CL	-.025	.053	.346	16.1986	7.8022	10.4934	2.1814	2.1278	1.7350	1.0255

Note: Covariances are in the upper triangle, correlations in the lower triangle, and variances are on the main diagonal. IT = inverted latency: 1,000 / mean latency in ms. CL = Correct latency in seconds.

Appendix A:
Means, variance-covariance, and correlation matrix for all variables (continued)

Variable	1	2	3	4	5	6	7	8	9	10
15. Number Series CL	-.222	-.154	-.131	-.229	-.107	-.078	-.098	-.201	-.077	-.143
16. Arrow Series CL	-.041	-.164	-.127	.060	.057	.106	.164	.206	.106	-.019
17. Vocabulary CL	-.068	.050	-.053	.040	.004	.023	-.030	-.062	-.187	-.175
18. Gen. Knowledge 1 CL	-.140	-.172	-.136	-.106	-.086	-.042	-.146	-.173	-.164	-.288
19. Gen. Knowledge 2 CL	-.144	-.124	-.132	-.292	-.081	.005	-.131	-.173	-.063	-.306
20. Gen. Knowledge 3 CL	-.086	-.087	-.084	.005	.016	.001	.027	.039	-.121	-.124
Mean	1.8266	0.4645	0.3451	0.8092	8.1774	8.7366	8.5538	7.1452	12.2849	15.8333

Variable	11	12	13	14	15	16	17	18	19	20
15. Number Series CL	-.170	-.230	.207	.160	146.3894	19.1498	8.8344	9.1782	8.5468	4.9270
16. Arrow Series CL	.011	.060	.261	.253	.154	105.8784	3.4643	3.2700	3.2825	5.3554
17. Vocabulary CL	-.270	-.176	.268	.215	.290	.134	6.3336	2.5608	1.9855	1.3730
18. Gen. Knowledge 1 CL	-.288	-.295	.238	.242	.347	.145	.465	4.7821	2.3300	1.9016
19. Gen. Knowledge 2 CL	-.215	-.346	.152	.207	.339	.153	.379	.512	4.3348	1.9105
20. Gen. Knowledge 3 CL	-.122	-.120	.186	.115	.184	.235	.246	.392	.414	4.9156
Mean	17.1559	15.2151	52.3612	15.7697	39.4103	31.3384	9.9294	10.6715	9.8942	11.0138

Note: Covariances are in the upper triangle, correlations in the lower triangle, and variances are on the main diagonal. IT = inverted latency: 1,000 / mean latency in ms. CL = Correct latency in seconds.

Appendix B:
Number of items of tasks in difficulty bins, and the mean difficulty of bins for ability groups

	Difficulty bins						
	1	2	3	4	5	6	7
Fluid intelligence							
Task				No. items			
APM	0	2	1	2	3	4	3
CFT	4	1	2	2	0	2	2
Number Series	1	3	2	2	4	0	0
Arrow Series	2	1	2	1	1	2	3
Ability group[a]				Difficulty[b]			
1	.91	.75	.57	.48	.29	.28	.16
2	.97	.88	.77	.71	.57	.46	.22
3	.98	.96	.90	.80	.80	.60	.37
Crystallized intelligence							
Task				No. items			
Vocabulary	3	3	4	4	2	1	3
General Knowledge 1	5	2	1	6	6	4	4
General Knowledge 2	3	5	6	3	5	3	3
General Knowledge 3	3	5	4	2	2	7	5
Ability group[a]				Difficulty[b]			
1	.86	.68	.56	.45	.35	.25	.12
2	.93	.80	.69	.60	.51	.35	.20
3	.97	.89	.83	.72	.65	.50	.36

Note: [a]1 = below average, 2 = average, 3 = above average. [b]Expressed as the mean proportion of correct responses for items in the bin.