



Article

Modeling Mental Speed: Decomposing Response Time Distributions in Elementary Cognitive Tasks and Correlations with Working Memory Capacity and Fluid Intelligence

Florian Schmitz * and Oliver Wilhelm

Institute of Psychology, University Ulm, Albert-Einstein Allee 47, 89081 Ulm, Germany; oliver.wilhelm@uni-ulm.de

* Correspondence: florian.schmitz@uni-ulm.de; Tel.: +49-731-50-31143

Academic Editor: Paul De Boeck

Received: 18 April 2016; Accepted: 15 September 2016; Published: 14 October 2016

Abstract: Previous research has shown an inverse relation between response times in elementary cognitive tasks and intelligence, but findings are inconsistent as to which is the most informative score. We conducted a study ($N = 200$) using a battery of elementary cognitive tasks, working memory capacity (WMC) paradigms, and a test of fluid intelligence (gf). Frequently used candidate scores and model parameters derived from the response time (RT) distribution were tested. Results confirmed a clear correlation of mean RT with WMC and to a lesser degree with gf . Highly comparable correlations were obtained for alternative location measures with or without extreme value treatment. Moderate correlations were found as well for scores of RT variability, but they were not as strong as for mean RT. Additionally, there was a trend towards higher correlations for slow RT bands, as compared to faster RT bands. Clearer evidence was obtained in an ex-Gaussian decomposition of the response times: the exponential component was selectively related to WMC and gf in easy tasks, while mean response time was additionally predictive in the most complex tasks. The diffusion model parsimoniously accounted for these effects in terms of individual differences in drift rate. Finally, correlations of model parameters as trait-like dispositions were investigated across different tasks, by correlating parameters of the diffusion and the ex-Gaussian model with conventional RT and accuracy scores.

Keywords: mental speed; working memory capacity; intelligence; response time modeling; ex-Gaussian model; diffusion model

1. Introduction

Intelligence researchers have investigated the correlation of mental speed with measures of intelligence since the mid-20th century [1]. In fact, large-scale studies and meta-analyses have documented a moderate relation of intelligence with inspection time, reaction time, and other measures of speed of processing [1–3].

Reductionist theories posit that mental speed constitutes a fundamental of fluid intelligence (gf) [4]. Similarly, some accounts of working memory capacity (WMC) assume a causal role of mental speed. For instance, the time-based resource-sharing model [5,6] assumes some executive resource is responsible for processing and rehearsal. The faster processing is completed, the more time is available for the rehearsal of decaying memory traces. In turn, it was suggested that WMC constitutes a basis of gf [7,8]. However, the notion that speed plays a causal role for intelligence is not unequivocal. It was emphasized that mental speed should not be considered the basis of intelligence, but as one important factor among others [9]. Additionally, it was argued that speed requirements in measures of WMC or gf may contribute to the observed correlation [10].

Nevertheless, a moderate relation of mental speed with cognitive ability is empirically well supported in large scale population-based studies [3] and in different age groups [11]. However, a meta-analysis [1] suggests that this relation is moderated by characteristics of the speed tasks. Specifically, choice reaction time tasks show a trend towards higher correlations with increasing complexity. Early evidence in this respect dates back to the work of Hick [12], who demonstrated that response times increase with the number of response alternatives. Additionally, this increase was found to be steeper for persons with lower intelligence [13]. This implies that tasks discriminate better between persons as complexity increases [2].

Interestingly, classical studies suggest the standard deviation of the RT distribution (SD_{RT}) is slightly more correlated with gf than the mean (M_{RT}) [2]. Variability in task performance is of interest as it may indicate impaired stability of the cognitive system. It was shown that SD_{RT} is lowest in young adults and is higher in both younger children and older adults [14]. Thereby, the developmental trajectory of SD_{RT} as an inverse marker of cognitive functioning resembles that of cognitive ability. A recent meta-analysis on the basis of 27 independent samples [15] confirms a moderate relation between SD_{RT} and intelligence, but does not support the previously held notion [2] that variability is a better predictor of intelligence than mean RT. Generally, M_{RT} and SD_{RT} are highly collinear across participants (i.e., $r \approx 0.9$; [16]). Additionally, both scores are affected by extreme values, which can be expected to contribute to their correlation. Therefore, the common practice in some RT measures to scale means by the individual's variability may function as a pragmatic remedy [17]. Theoretically, more satisfactory modeling approaches are discussed below. Since M_{RT} , SD_{RT} , and gf are highly related, it was suggested that, from a psychometric perspective, a common factor could account for their relation [18].

Other lines of research suggest that not mean and variability of the RT distribution, but the slowest RT values of a person are the most predictive score of cognitive ability. In a classic study, Larson and Alderton [19] sorted responses in increasing order and investigated relations with ability for different RT bands. They found that the slower the RT, the higher the relation with ability, which is now known as the *worst performance rule* (WPR). The WPR was confirmed in many studies (see [20], for a review), both for fluid intelligence and for executive functions [21]. Additionally, the WPR was shown across different age groups [11]. However, the evidence is not unequivocal, as some studies revealed that all portions of the reaction time distribution contain similar individual difference information [22] (see also [23]). It was argued that WPR only holds for complex but not for simple tasks [24]. It should be noted that the WPR can be considered a paradigm shift in experimental research: Traditionally, researchers were interested in mean response times, and they tried to reduce the potentially biasing effects of extreme values by excluding or winsorizing outliers, by log-transforming individual RTs, or by computing their median. Of course, such treatments only make sense if they help remove contaminants but not meaningful information. In fact, some studies revealed that cognitive ability is slightly more correlated with the mean (M_{RT}) than with the median (Mdn_{RT}) RT [18,25], where the former but not the latter is known to be biased by extreme values.

A more elegant way to dissociate components that contribute to an RT distribution is to fit an explicit RT model. The most popular approaches include the ex-Gaussian model [26–28] and the diffusion model [29,30]. Additionally, growth curve modeling was recently suggested as an alternative to model task demands and WPR simultaneously [31].

The ex-Gaussian approach offers a parsimonious parameterization that effectively describes the shape of a typical RT distribution, which resembles a normal distribution but with heavy right tail. Technically, the ex-Gaussian model assumes that RT distributions can be approximated by folding a Gaussian normal distribution, with parameters μ and σ for the mean and the standard deviation, with an exponential decay function with parameter τ . The ex-Gaussian parameters are informative because they offer a bias-corrected estimate of the location of the RT distribution (μ), of its variability (σ), and of the proportion of exponential contaminants (τ). Parameter μ is conventionally interpreted as the average time required for a (correct) response in most of the trials (as an inverse indicator of

mental speed), while the parameter τ is sensitive to the proportion of occasional, but extreme, RT outliers. Therefore, the τ parameter has been interpreted as an indicator of attentional/cognitive lapses—or other interruptions in information processing [21,32]. The ex-Gaussian model is a good descriptive approximation of the RT distribution [27,28] and has been shown to be a useful tool in RT research [26,33]. Its parameters have been used to test theories (e.g., τ as an indicator of cognitive lapses [21,32], among others). However, the model itself is descriptive and lacks an explicit theoretical basis. Additionally, the ex-Gaussian model only uses the information in correct responses and cannot adequately cope with speed-accuracy trade-offs.

The diffusion model [29,34] is a better alternative in both respects since it offers a psychologically plausible yet parsimonious model of a binary decision process. In addition, it helps dissociate task-performance from speed-accuracy settings among other components. Specifically, the diffusion model (see Figure 1) decomposes a decision process into non-decision time and the actual decision process. Non-decision time (T_{er}) subsumes processes before and after the actual decision phase (encoding of stimuli and execution of the motor response). In turn, the actual decision process is characterized by a continuous sampling of information. A decision process, originating from starting point z , fluctuates over time as a function of systematic stimulus information and random noise (see gray sample path). When it hits either the lower or the upper response boundary (at 0 or a , respectively), the according response is elicited. The mean slope of the decision process across trials denotes the drift rate (v).

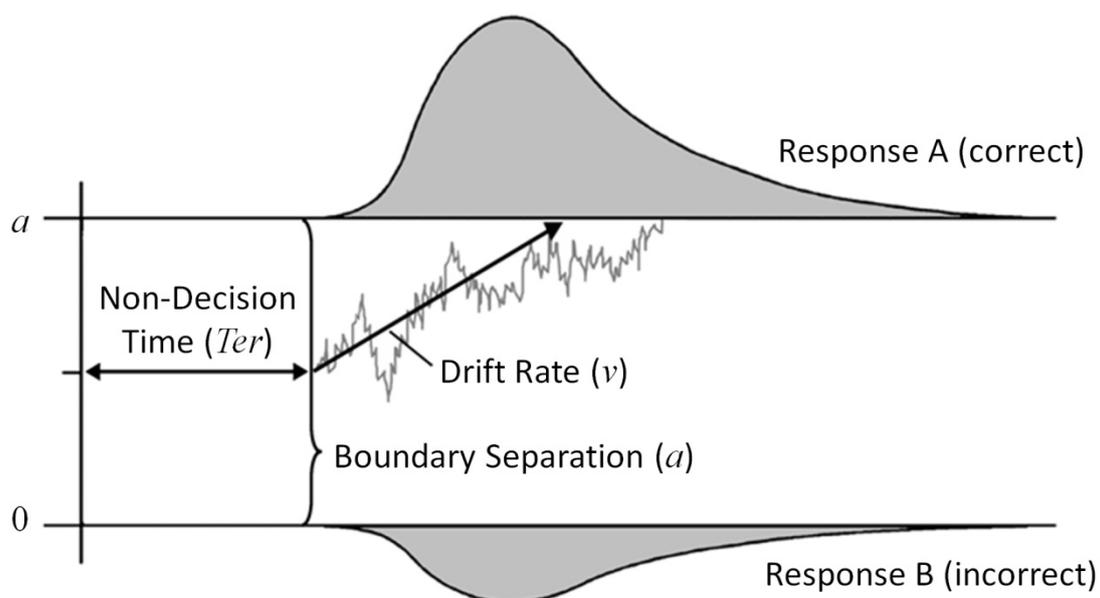


Figure 1. Diffusion model for binary decision tasks.

Thus, the diffusion model dissociates some components of the decision process. The boundary separation indexes the setting of the speed-accuracy trade-off, while the drift rate is an estimate of the efficiency of information processing (i.e., the speed of evidence accumulation for the correct response per time). An advantage over conventional RT scores is that the simultaneous estimation of parameters should result in more adequate estimates of task performance with a reduced bias of speed-accuracy setting. It was demonstrated on theoretical grounds, as well as simulations [35,36], that the diffusion model can account for a number of replicated phenomena in RT research, including the right skew of the RT distribution, the WPR, the higher ability correlations of SD_{RT} over M_{RT} and the latter over Mdn_{RT} , and the linear correlation of M_{RT} and SD_{RT} . All these effects were argued to be driven by a single latent relation between individual differences in drift rate and individual differences in general intelligence.

One challenge pertaining to parameter estimation is to cope with computational dependencies. These can arise from trade-offs between parameter estimates in the data-fitting process and would result in correlated deviations of the estimated values from the true parameter values. Such trade-offs have been demonstrated for the ex-Gaussian model [37,38] and for the diffusion model [34]. In the ex-Gaussian model, dependencies are positive for μ and σ , but negative for both Gaussian parameters with the τ parameter of the exponential component. In the diffusion model, positive correlations are frequently observed between all parameters (a , ν , and T_{er}). These can result from the by-chance occurrence of a few extra slow error response times that would bias all parameters jointly in the same direction [34]. Additionally, estimating drift rate (ν) and response caution (a) can be a challenge in case of only few errors values, as different combinations of both parameters (both jointly increasing) could account for the observed distribution of correct RT values.

Schmiedek et al. [38] suggested an elegant way of circumventing these problems of parameter dependencies, combining the parameterization of RT distributions and structural equation modeling. Specifically, they fitted the ex-Gaussian model and the diffusion model independently to the RT data obtained in eight elementary tasks. Variability in these parameter estimates can be expected to contain true parameter variance as well as estimation error/biases. These parameter estimates were entered, technically, as observed variables in a CFA model, with latent parameter factors accounting for the reliable (shared) portion of the variance in the parameter estimates. Residuals of parameters simultaneously estimated from the same data were allowed to be correlated and would reflect covariance in the deviations of parameter estimates. This approach reduced problems with parameter dependencies and with unreliability of the parameter estimates. The parameter factors, which captured the reliable portion of parameter variance, were used as predictors of WMC and *gf*.

In spite of estimating parameters from relatively simple tasks, substantial relations with ability were obtained for the τ parameter in the ex-Gaussian model and for the drift rate (ν) in the diffusion model. The observed relations with the τ parameter can be reconciled with an account on lapses of attention (or at any other processing stage) that would result in occasional delayed responses. However, Schmiedek et al. [38] argued in favor of the theoretically more satisfactory diffusion model, and they offered a simulation study showing that assuming a diffusion process as the “true” underlying mechanism generates RT data that can account for the relations of the ex-Gaussian parameters. The ability of the diffusion model to account for relations of the τ parameter or the WPR, respectively, has been confirmed in recent simulation studies [35,36].

As all elementary tasks in the Schmiedek et al. study [38] were relatively simple; the previously well-documented moderating effect of task complexity [1,2] could not be tested in their study. It can be hypothesized, though, that task complexity affects the shape and location of the RT distribution, and consequently derived model parameters. Most elementary tasks require maintaining a number of task relevant bits of information, e.g., S-R mapping rules, in working memory. Presented stimuli have to be compared with these rules until an appropriate rule is found and the response can be selected. Building on the seminal finding that search in working memory is serial and exhaustive [39,40], a linear relation of the number of task-relevant bits of information maintained in working memory and response times can be expected. In turn, this would result in a shift of the entire RT distribution towards slower RT values as complexity increases. Such an effect would be reflected in the μ parameter of the ex-Gaussian model. Additionally, variability and skew of the RT distribution may be increased, as longer processing requirements usually result in higher variability, too [34,41]. These effects would be reflected in increased σ and τ in the ex-Gaussian model, while all these effects could be parsimoniously accounted for by lower drift rate (ν) in the diffusion model [35]. In turn, individual differences associated with an increase in task complexity would be reflected in these parameters (i.e., μ additionally to τ in the ex-Gaussian model; and drift rate ν in the diffusion model).

What is the status of individual differences in these parameter estimates? How are the reliable portions of individual differences in parameter values related across models and with the conventionally computed RT and accuracy scores? First, if parameter estimates are considered

“measures of trait-like cognitive styles and abilities” ([42], p. 4), sufficiently high temporal stability is required. Further prerequisites comprise their generality (at least across a set of tasks assigned to one processing domain) and validity with other parameters and criterion variables (e.g., shown in [38]); Second, with respect to relations across the diffusion model and the ex-Gaussian model, it has been consistently shown that correlations are only moderate [37,38,43] in spite of sufficiently high (parallel test) reliability [38] and temporal stability [42]. This situation suggests that parameters of different models may, at most, simply tap different processes. In fact, most of the studies specifically addressing parameter relations across diffusion model and ex-Gaussian model may have yielded inflated relation estimates at the upper boundary. For instance, some of these studies investigated relations by means of simulating RT distributions based on parameters of the one model (with many trials and without contamination), and subsequently used the simulated RT data to estimate parameters of the other model [37,38,43]. These simulations were supplemented by analyses of moderately sized experimental datasets that were jointly fitted by both models [38,43]. In the latter case, even random fluctuations or contaminants that do not correspond with stable personality dispositions may contribute to parameter relations. However, parameter relations across models were only moderate and were not very specific in most studies. Nevertheless, their patterns appear to be comparable across studies [37,38,43]: response caution (a) is positively related with all ex-Gaussian parameters, in particular with μ and τ . The drift rate (ν) is moderately negatively correlated with μ , somewhat more negatively correlated with σ , and the strongest negative correlation is with τ . Finally, non-decision time (T_{er}) is positively related with μ but not with the other ex-Gaussian parameters.

Relations of model parameters with conventional RT and accuracy scores are hardly ever reported (but see [35,38]), although these relations may be potentially informative as they offer a first insight into which processes may be reflected in the scores conventionally reported in the literature. In fact, Schmiedek et al. [38] report correlations of model parameters with some RT scores in an appendix of their paper. Since these correlations were estimated as pairwise latent correlations on the basis of identical RT data, most correlations were mediocre to high. It could be predicted on theoretical/computational grounds that they found drift rate (ν) to be inversely correlated with RT scores, in particular with those indexing slow response times (slower quantiles) and with variability (SD_{RT}). In contrast, non-decision time (T_{er}) was most highly correlated with the faster quantiles of the RT distribution. Analogous correlations were obtained for μ and τ of the ex-Gaussian distribution, respectively. Unfortunately, the authors did not report these correlations with accuracy. Nevertheless, these findings are definitely informative with respect to the correspondence of parameters with observable scores. However, since identical RT distributions were entered, this does not distinguish between state/contamination and the stable/generalizable portions of parameters as (indicators of) trait-like dispositions (cf. [42]) that can be predicted to be considerably more moderate. Another challenge is that the correlations of parameters with observed RT scores were shown to depend on the variance in the other parameters, as demonstrated in simulation studies [35]. Therefore, observed correlations in experimental datasets can be expected to vary as a function of the variance in the other parameters. The resulting range of relationship estimates is therefore an empirical question.

To summarize, previous research has shown moderate but not unique correlations between parameters of the diffusion model and of the ex-Gaussian model [37,38,43]. Differential correlations with conventional RT scores can be predicted on theoretical grounds and for computational reasons [34,41]. However, there is only limited evidence of how the reliable portion of parameter variance (i.e., on the level of latent factors; [38]) is correlated with conventional RT scores. To arrive at a more realistic estimate of parameter correlations conceived of as trait-like dispositions, it is desirable to model correlations across non-overlapping RT data.

Goals and Hypotheses of the Present Study

This study was conducted to test correlations of RT scores and parameters in elementary cognitive tasks with WMC and *gf*. A set of conventionally employed scores and model-based parameters found

in the RT literature was investigated as predictors of cognitive ability. The design was inspired by the Schmiedek et al. [38] study, but we employed tasks with different levels of complexity to test possible moderation effects on parameter correlations. Additionally, correlations of parameters across models and with conventional RT and accuracy scores will be reported using non-overlapping raw RT data. Specifically, the following hypotheses were tested:

H1: Correlations of mental speed with WMC and gf: Building on previous research, we predicted that mental speed is moderately correlated with WMC and *gf* [1–3]. Such a correlation was predicted by theoretical accounts that postulate a contribution of speed to WMC [5,6] and *gf* [4]. Additionally, such correlations were expected to result if the cognitive measures require speedy processing as a confounding factor [10].

H2: Differential validity of RT scores: Higher validity of some RT scores over others could be postulated on theoretical grounds [35], as well as on the basis of previous findings [2,18,19,25]. Given that the slowest RT values are the most predictive of cognitive ability (i.e., the worst performance rule, [19]), validity of RT scores were expected to increase with their sensitivity to slow RT values. For instance, slower RT bands were predicted to be more highly correlated with ability than fast RT bands. Additionally, location scores that can be biased by extreme values (e.g., M_{RT}) were expected to be somewhat more highly correlated with ability than with their robust counterparts (e.g., Mdn_{RT} or $M_{\log(RT)}$) [18,25]. However, in line with the current meta-analysis, we did not predict RT variability to be consistently more highly correlated with ability than with mean RT [15]. Error scores were less frequently employed in the literature, supposedly because of their reduced variability and, consequently, their reduced reliability and validity. We still included them in this study and investigated their correlations with model parameters and cognitive ability to uncover possible speed-accuracy trade-offs.

H3: Validity of parameters sensitive to the right tail of the distribution: Relatedly, we predicted that the exponential component τ in the ex-Gaussian model and the drift rate ν in the diffusion model are correlated with ability, since both parameters are affected by the slow tail of the RT distribution [35,38].

H4: Moderation of the WPR by task complexity: Since very simple tasks do not require much mental work, hardly any differences were expected in the mean response times [2,24]. Only occasional slow outliers (e.g., attentional, cognitive, or other lapses or interruptions in information processing [21,32]) could account for the relation with ability [19,38], and correlations would be confined to the τ parameter in the ex-Gaussian model, replicating previous findings [38]. However, given that processing time increases with task-relevant information maintained in working memory [39,40], the location of the RT distribution was expected to be shifted towards slower response times. In turn, individual differences would be additionally reflected in μ in the ex-Gaussian model as task complexity increases. In the diffusion model, these effects in location and skew [34,41] with increasing task complexity should be accounted for by decreased drift rate (ν) [34,35,43].

H5: Testing model parameters as trait-like dispositions: Given that model parameters may serve as indicators of trait-like dispositions [42], they were expected to possess sufficient generality. In other words, individual differences in parameter estimates should be replicable across different tasks—at least if tasks are comparable in their cognitive requirements. Additionally, parameter estimates were expected to possess consistent relations (with other parameters and scores), even when estimated from independent experimental data. We expected model parameters to display a pattern of relations comparable to that observed in previous studies [37,38,43]. However, the magnitude of relationships was predicted to be attenuated when estimated across different task classes, even when modeling the reliable portion of parameter variance (as in [38]). Nevertheless, reducing possible confounding factors and contamination as well as computational dependencies, was considered as more adequate to arrive at better estimates of parameters as trait-like dispositions.

2. Experimental Section

2.1. Sample

The study was advertised in a local newspaper, by means of flyers, and in electronic media. Inclusion criteria were an age of 18 to 40 years old and sufficient knowledge of the German language. $N = 200$ participants ($n = 144$ female) completed the battery comprising mental speed tests, as well as measures of working memory capacity and of fluid intelligence. Testing was conducted in small groups of up to six people per testing session. Participants were 25.7 years old on average ($SD = 5.2$). The educational level of the sample was above average, as a majority of the participants ($n = 130$) had completed high school, and they were working in various occupational fields. All participants signed informed consent before participating and received compensation.

2.2. Speed Tasks

We used a computerized version of typical clerical speed tasks, namely Search, Comparison, and Substitution tasks [44]. Following a matrix construction format, all three tasks were administered with numbers, letters, and symbols as stimuli. The different classes of tasks can be expected to differ in complexity. Additionally, with three tasks (stimulus materials) per class, the measurement models for the latent ex-Gaussian and diffusion model parameter factors had sufficient degrees of freedom ($df = 15$) to be testable. The computerized administration required a few procedural changes so that individual RTs could be recorded that would be used for the RT modeling (see Figure 2).

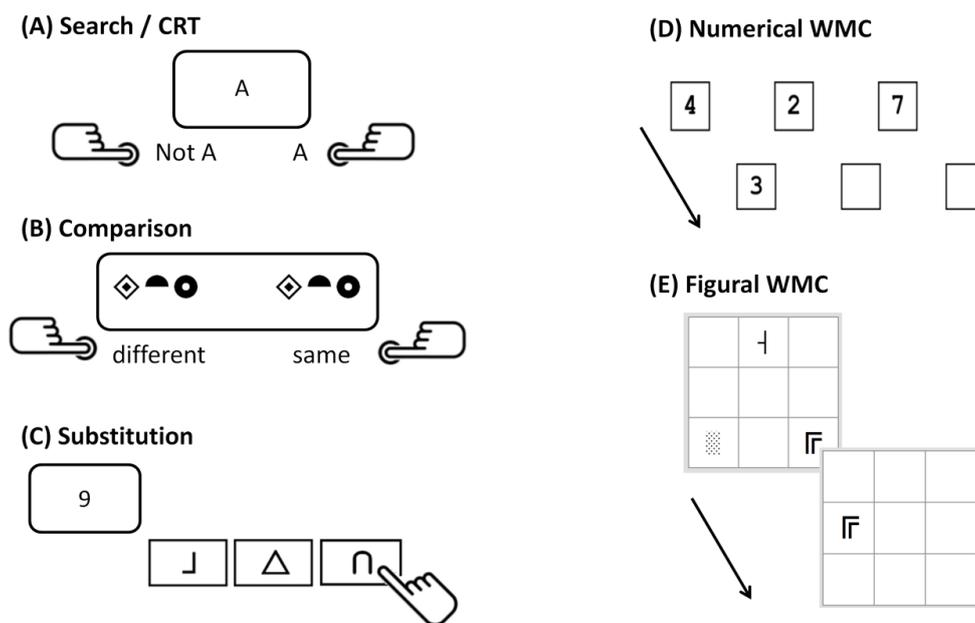


Figure 2. Illustration of mental speed tasks (A–C) and of WMC tasks (D,E).

The original paper-and-pencil Search task consisted of a printed matrix of stimuli: participants were asked to go through the matrix line by line and to mark all target stimuli (but not distractors) they saw in the given time. In the computerized version, single stimuli were presented sequentially on screen and participants indicated whether the displayed stimulus was the target or not by pressing a right or left button, respectively. In the version with letters as stimuli, the target was an 'A', in the version with numbers, it was the digit '3', and in the version with symbols as stimuli, it was a smiley among other emoticons. For each stimulus material, participants completed two blocks of 40 trials each (and two warm-up trials each that were discarded from the analyses). Note that the computerized version of the "Search" task could be formally classified as a one-bit choice reaction time (CRT) task.

In the Comparison task, two strings of three elements each were presented horizontally on the screen. Participants indicated whether the strings of stimuli were identical or different by pressing a left or a right button, respectively. In the case of a difference, only one of the elements was randomly exchanged. The presented elements could either be numbers, letters, or symbols. Again, there were two blocks with 40 trials each (plus two warm-up trials) for each stimulus domain.

In the original paper-and-pencil Substitution tasks, there is a coding table presenting how stimuli of one stimulus domain (e.g., numbers) are mapped onto the stimuli of another domain (e.g., symbols). Below this table, participants see lines of stimuli from the first domain, and they are asked to draw the corresponding stimulus from the second domain. In the computerized version, the coding table was permanently visible on screen and stimuli of the first domain appeared sequentially. We used special keyboards with keys labeled with the corresponding second stimulus, and participants were instructed to press the according key. There were three different combinations of stimulus domains in the Substitution tasks: numbers to symbols, symbols to letters, and letters to numbers. For each combination, participants completed two blocks with 30 trials each (plus two warm-up trials).

Note that the three task classes differ markedly in complexity, i.e., in the number of task-relevant bits of information that have to be maintained in working memory in order to solve the task correctly. The Search tasks require keeping just one target stimulus in mind for the duration of the entire task and to decide in each trial whether a presented stimulus matches it or not. The Comparison tasks are more demanding. They necessitate the comparison of two three-item sets in each trial. Finally, Substitution tasks are by far the most complex, since they require operating with nine arbitrary S-R mapping rules in order to select the appropriate response for the presented stimulus in each trial. The number of mapping rules exceeds what people can reliably keep in mind at one time (so they need to look up the mapping rules repeatedly).

2.3. WMC and *gf* Measures

Three recall-one-back tasks [45] with different stimuli were included in this study. This design allowed for modeling WMC as a latent factor, thereby removing problems associated with compromised reliability and task specificity. The tasks were constructed following a matrix design balancing memory load (two, three, and four stimuli) and an updating requirement (six and nine updates). In the numerical WMC task, two to four boxes appeared on a screen aligned horizontally (see Figure 2D). Each run started by presenting numbers in all boxes. Then, all numbers disappeared and one number was shown unpredictably in one of the boxes for 3000 ms. Participants were asked to type in the number shown in that box the time before and to remember the new number. After an inter-trial interval of 500 ms the next trial started. The task version with letters as stimuli was analogous, only with consonants instead of numbers. In the figural task version, all stimuli were shown at unpredictable positions in a 3×3 grid. At the beginning of each run, all stimuli appeared simultaneously (i.e., either 2, 3, or 4 stimuli) at different positions in the grid. Then, all stimuli were removed and only one of these stimuli would appear randomly in one the cells for 3000 ms. Participants were asked to indicate, by mouse-click, the cell where this stimulus was shown last and to remember the actual position of the stimulus. After an inter-trial interval of 500 ms, the next stimulus was shown until the 6 or 9 updates of the current run were completed. The tasks were scored following a partial credit scheme [46], i.e., the proportion of correct responses in all trials.

Fluid intelligence was assessed with a figural sequence reasoning test from the BEFKI [47]. The scale is comprised of 16 items in which two consecutive pictures have to be selected from a set of distractors that complete a sequence in a logical way. The individual items were aggregated into three parcels so that *gf* could be modeled as a latent factor. The paper-and-pencil version of the BEFKI was administered.

2.4. Scoring and Modeling of Response Times

We computed a number of conventional scores used in research with elementary cognitive tasks [26,28], including speed of responding, response times, quartiles, variability, and errors. Most of the scores are highly correlated, since they all use information in the RT distribution. However, given they are differentially sensitive to the shape of the RT distribution and to extreme RT values, it can be predicted that the scores are also differentially correlated to cognitive ability. Data were analyzed with R [48], using *retimes* [49] for the ex-Gaussian analyses, *EZ* for the simplified diffusion model [50], and the *psych* package [51] for psychometric analyses. *Mplus* [52] was used for structural equation modeling.

Speed of responding: Mental speed is conventionally scored as correct responses in the given time in most paper-and-pencil tests. We averaged reciprocal response times ($1/RT$) for correct responses as an approximation. Although this method is less common in laboratory research, it corresponds with standard scoring in psychometric tests. As an additional benefit, the distribution of reciprocal response times across participants is approximately normal, a requirement for some correlational analyses.

Response Time: The most frequently used score in laboratory research is the mean latency of correct responses. The arithmetic mean (M_{RT}) is frequently used, but it can be biased towards higher values by a few extreme RT values. Therefore, the mean of individually log-transformed RTs ($M_{\log(RT)}$) or the Median of the RT distribution (Mdn_{RT}) are frequently computed when a more robust estimate of the location of the distribution is desired. Distributions of mean response times across participants can be positively skewed.

Quartiles: The Worst Performance Rule was originally discovered by inspecting RT bands (i.e., quantiles of the RT distribution). Given the available trial numbers per task in the current study, we computed the mean for four quartiles (Q1–Q4).

Variability: Variability of responding may correspond with impaired stability of cognitive processes. We computed the within-person standard deviation (SD_{RT}) and the interquartile range (IQR_{RT}) of the response times as two conventional scores. The first can be biased by just a few outliers, whereas the second may be a more robust estimate of the spread of the RT distribution.

Errors: Cognitive impairments can also manifest in erroneous responses. Additionally, individuals may differ in their speed-accuracy trade-offs, either sacrificing speed for accuracy or the other way round. Errors may be informative, but they are usually rare events, thereby characterized by low variability and low reliability. Additionally, they are heavily skewed to the right. We computed the error rate as well as the probit of the error rate, which compresses the positive skew.

Additionally, we fit the ex-Gaussian and a simplified diffusion model to the RT data. Since parameter dependencies can result in trade-offs in the estimates of parameter values, we used a two-step maximum likelihood (ML) procedure to estimate the ex-Gaussian parameters [53], as is implemented in the *retimes* R package [49]: The μ and σ parameters of the normal distribution are estimated from the RT distribution in a first step; then, the exponential τ parameter is determined with help of a bootstrapping procedure. We computed parameters of the *EZ* diffusion model [50], which is a simplified closed-form expression yielding equivalents of the three most important parameters of the diffusion model, i.e., boundary separation (a), drift rate (ν), and non-decision time (T_{er}). Dependencies are less of a problem using *EZ*, since parameters are not estimated, but their respective equivalents are directly computed from moments of the RT distribution and accuracy. Additionally, a number of simulation studies converge in showing that *EZ* estimates are robust for the purpose of modeling individual differences [54,55], even when only a moderate number of trials are available [56]. Since the diffusion model is only suited for binary choice tasks, it could only be fitted to the data from the Search and Comparison tasks.

To estimate the relations between parameters as trait-like dispositions, we modeled their reliable portions of variance analogously to what was done by Schmiedek et al. [38]: first fitting models independently to each task, and then modeling the communality of these parameter estimates as latent factors, while allowing residuals of the parameters estimated from the same data to be correlated (see

Figure 4). These measurement models were used to investigate latent correlations of model parameters across task classes—within and across models and with conventional scores. For testing relations with conventional RT and accuracy scores, the reliable portion of variance was modeled as one factor accounting for the communality in the scores of one task class. In order to avoid adverse effects of random fluctuation, contamination, and computational dependency, relations were estimated across different task classes. As the diffusion model could be fitted only to the binary choice Search and Comparison tasks, there was only one cross-task correlation for each combination of parameters. The ex-Gaussian model could be fitted to all task classes; consequently, three independent correlations could be estimated across task classes. Finally, four correlations could be estimated across task classes for the parameters of diffusion model with the ex-Gaussian parameters (i.e., Search–comparison, Comparison–search, Search–substitution, and Comparison–substitution). The same combinations of task classes were also used to compute correlations with conventional scores.

2.5. Data Preparation and Descriptive Analyses

All RT data were carefully screened and a limited number of potentially invalid trials were removed prior to conducting the analyses. The two warm-up trials at the beginning of each block were removed because of re-start costs. All trials directly following an error were discarded because post-error control processes may affect RTs independent of task requirements [57] (ca. 4%, see Table 1 for details). Extreme RTs (ca. 1%) were removed by applying the liberal Tukey criterion [58], individually adjusted to person and task. Specifically, the criteria were set at three interquartile ranges below the 25th percentile (but not below 200 ms) and above the 75th percentile of the RT distribution.

As expected, descriptive statistics of task performance and corresponding model parameters were relatively homogeneous within task classes, but differed considerably across task classes (see Table 1). In general, accuracy was high and comparable across tasks, but different response times across task classes confirmed differential task complexity. The Comparison tasks were most comparable with the tasks used by Schmiedek et al. [38] in terms of response times. The Search/CRT tasks had considerable faster response times, suggesting that they were simpler. Conversely, the Substitution tasks had considerably longer response times, suggesting that they were more complex. The differences in mean RT were paralleled by a corresponding increase in the variability of RT, replicating previous findings [16,18].

Analogous results were observed in the parameters of the ex-Gaussian model. As expected, all parameters yielded larger values with slower mean response times. The marked increase in the exponential τ parameter, in addition to the standard deviation parameter σ , suggests that the increase in the conventional SD_{RT} score is in part driven by extreme values from the right tail of the RT distribution. The same holds for the M_{RT} , which appears to be biased towards higher values compared with the μ parameter.

The diffusion model could only be fitted to the Search/CRT and the Comparison tasks with binary response values. The reduced drift rates ν in the Comparison tasks relative to the Search/CRT tasks confirm their higher task difficulty. Additionally, generally increased estimates of the a parameter indicate that participants exercised higher response caution, which quite likely reflects that they perceived Comparison tasks as more difficult. Finally, non-decision time (T_{er}) was higher in the Comparison tasks, possibly indicating longer encoding time of the more complex stimuli.

In order to model fluid intelligence as a latent factor, individual items of the matrix test were aggregated into three parcels by sorting them according to a systematic ABC scheme. It was tested whether parcels were strictly parallel: loadings, intercepts, and residuals did not differ significantly across parcels. (Constraining these parameters to be equal across parcels did not deteriorate the model fit relative to the configural model, $\Delta\chi^2(6) = 10.03, p = 0.12$.)

Table 1. Descriptive statistics of RT scores and model parameters.

Tasks	Excluded		Descriptive Statistics			Ex-Gaussian Model			Diffusion Model		
	Post-Err	Extreme	M_{RT}	SD_{RT}	Accuracy	μ	σ	τ	a	ν	T_{er}
Search/CRT											
Numbers	0.04 (0.08)	0.01 (0.02)	377 (46)	78 (20)	0.97 (0.03)	307 (46)	32 (18)	70 (27)	0.09 (0.01)	0.39 (0.07)	263 (35)
Letters	0.04 (0.05)	0.02 (0.04)	386 (43)	79 (21)	0.96 (0.04)	317 (45)	37 (18)	69 (30)	0.09 (0.01)	0.38 (0.07)	272 (34)
Symbols	0.05 (0.04)	0.01 (0.02)	469 (49)	90 (19)	0.95 (0.04)	397 (44)	51 (20)	72 (26)	0.09 (0.01)	0.34 (0.06)	343 (34)
Comparison											
Numbers	0.04 (0.04)	0.00 (0.01)	871 (145)	218 (81)	0.95 (0.04)	676 (102)	83 (47)	196 (84)	0.15 (0.03)	0.23 (0.05)	561 (81)
Letters	0.05 (0.04)	0.01 (0.01)	1009 (196)	258 (101)	0.95 (0.04)	773 (129)	93 (44)	236 (105)	0.16 (0.04)	0.20 (0.04)	643 (102)
Symbols	0.07 (0.04)	0.02 (0.02)	1230 (208)	331 (110)	0.93 (0.04)	924 (139)	151 (62)	306 (136)	0.17 (0.04)	0.17 (0.03)	769 (101)
Substitution											
Num→Sym	0.02 (0.03)	0.01 (0.02)	1473 (254)	447 (111)	0.98 (0.02)	1127 (251)	253 (110)	347 (169)	—	—	—
Sym→Let	0.03 (0.03)	0.00 (0.01)	1252 (210)	403 (113)	0.97 (0.03)	940 (221)	222 (102)	312 (161)	—	—	—
Let→Num	0.03 (0.03)	0.01(0.01)	1345 (275)	420 (125)	0.97 (0.03)	1006 (254)	225 (107)	339 (171)	—	—	—

Substitution tasks (given stimuli → required responses): Num = numbers, Let = letters, Sym = Symbols as stimuli and responses, respectively. Excluded = M (SD) proportion of excluded trials. Post-Err = post-error trials, Extreme = extreme values according to liberal Tukey (1977) criterion. All statistics given are M (SD) across persons. M_{RT} = mean response time, SD_{RT} = within-person standard deviation in RT. The ex-Gaussian model decomposes RT distributions into the parameters μ and σ (M and SD of the normal distribution) and τ (as the parameter of an exponential component). The diffusion model yields parameters a (*boundary separation*), ν (*drift rate*), and T_{er} (*non-decision time*), for the binary-choice Search/*CRT* and Comparison tasks.

3. Results

3.1. Correlations of Mental Speed with WMC and Fluid Intelligence (H1)

The correlation of mental speed with cognitive ability was tested using the reciprocal response time scores (1/RT), so positive correlations were expected. In line with the notion of speed as a hierarchical construct, bifactor measurement models with nested specific factors and one task as a reference method were specified [59]. In these models, mental speed was conceptualized as a general factor capturing the communality of all speed tasks, while nested method-factors for the Search and Comparison tasks were allowed. Correlations of mental speed with WMC and *gf* were tested by allowing correlations between the corresponding latent factors (see Figure 3). All loadings of the measurement models were substantial, and the model had an excellent fit (see Appendix A, Model #1). As expected, WMC was substantially correlated with *gf*, and a correlation of comparable magnitude was observed between mental speed and WMC. A somewhat smaller, but still moderate, correlation was also obtained for mental speed with *gf*. The difference between both correlations was significant, as indicated by a decrease in model fit when both correlations of speed with WMC and with *gf* were constrained to be equal ($\chi^2(1) = 7.17, p < 0.01$).

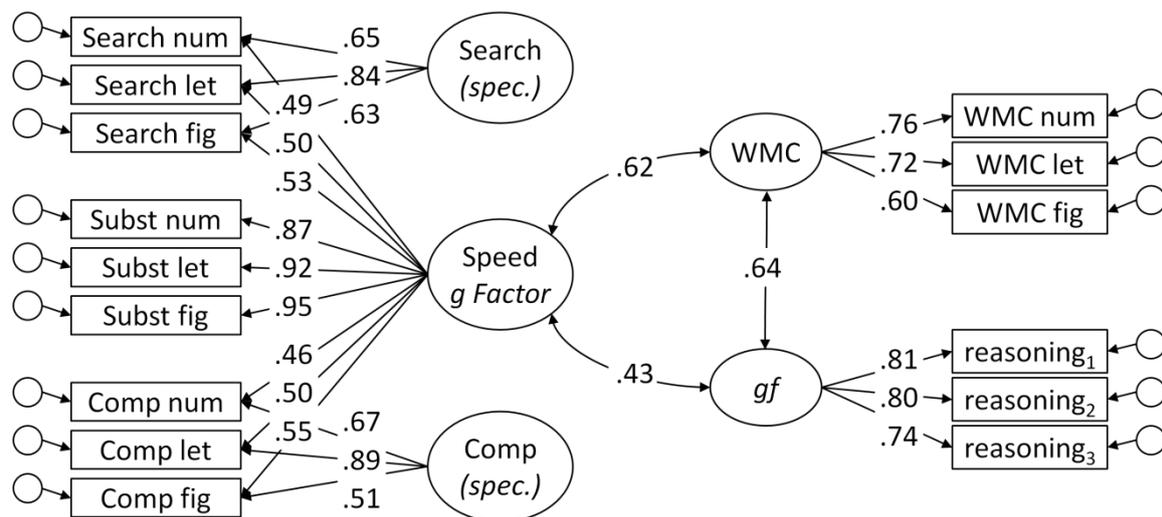


Figure 3. Correlation of mental speed with WMC and *gf*.

3.2. Differential Validity of RT Scores (H2)

Analogous analyses to the one for the speed of responding (1/RT) were conducted for all alternative mental speed scores. Information on measurement models (saturation and range of loadings) and model fit statistics are provided in Appendix A: Model #0 comprises only the measurement models of the correlates WMC and *gf* factors, while Models #1–12 additionally include a bifactor measurement model of mental speed, analogous to the one depicted in Figure 3 but for the alternative scores. The fit of the models was excellent throughout all RT location scores, and still satisfactory to good for the RT variability and error scores. The correlations of the speed *g* factor with WMC and *gf* for all alternative scores is given in Table 2. It is of note that all correlations with WMC and *gf* of RT location scores were comparable. Descriptively larger correlations were observed for the scores that reflect slower response times. However, the increase in the correlation with WMC of Q1 ($r = .59$) to Q4 ($r = .68$) was not significant, as was indicated by a non-significant decrease in model fit when the correlation of Q4 was numerically fixed to the value of Q1 ($\chi^2(1) = 2.11, p = 0.15$). Correlations of the RT variability scores were somewhat smaller than the location scores. Error scores were not consistently correlated with WMC and *gf*.

3.3. Validity of Parameters Sensitive to the Right Tail of the Distribution (H3)

Since estimates of model parameters may be biased due to random trade-offs in estimation, we modeled their correlations similarly to Schmiedek et al. [38]. Parameters were estimated independently for each speed task. These parameter estimates were entered as manifest variables in a confirmatory factor analysis, and latent “parameter” factors were specified that accounted for their communality. Residuals of the indicators were allowed to be correlated within one task (see Figure 4). Thereby, random trade-offs at estimation would be reflected in the correlation of residuals, while the latent parameter factors capture the reliable portion of the variance. This procedure allows estimation of the bias-corrected relations of the parameter factors. Further, these parameter factors were entered as correlated predictors of cognitive ability, which allows estimating their unique contribution to WMC and *gf*. We analyzed all tasks of one class simultaneously, but set up separate models for the three task classes to investigate the effects of task complexity (H4). Details concerning the measurement models and model fit are provided in Appendix A. Model fit was again highly satisfactory. The relations of the parameter factors with WMC and *gf* are displayed in Table 2. The correlations of the parameters were generally smaller compared with the conventional scores, but their pattern was relatively specific. In line with the predictions, the exponential τ parameter in the ex-Gaussian model and the drift rate ν in the diffusion model revealed moderate correlations with WMC and *gf* across the speed tasks.

Table 2. Correlations of speed scores and model parameters with WMC and *gf*.

#	Score/Parameter	WMC	<i>gf</i>				
<i>Speed</i>							
1	1/RT	0.62 ***	0.43 ***				
<i>Response Time</i>							
2	M_{RT}	−0.69 ***	−0.46 ***				
3	$M_{\log(RT)}$	−0.65 ***	−0.45 ***				
4	Mdn_{RT}	−0.67 ***	−0.45 ***				
<i>RT Quartiles</i>							
5	Q1	−0.59 ***	−0.41 ***				
6	Q2	−0.65 ***	−0.44 ***				
7	Q3	−0.69 ***	−0.46 ***				
8	Q4	−0.68 ***	−0.44 ***				
<i>RT Variability</i>							
9	SD_{RT}	−0.50 ***	−0.30 ***				
10	IQR_{RT}	−0.50 ***	−0.31 ***				
<i>Errors</i>							
11	Error Rate	−0.18 *	−0.09				
12	Probit (Error)	−0.06	−0.02				
<i>Ex-Gaussian Model</i>							
13	μ	−0.10	Comp	Subst	Search	Comp	Subst
	σ	−0.16	−0.09	−0.55 ***	−0.11	−0.02	−0.39 ***
	τ	−0.32 **	−0.02	0.01	−0.12	0.23	0.03
<i>Diffusion Model</i>							
14	a	−0.20	Comp	Subst	Search	Comp	Subst
	ν	0.35 ***	0.41 ***	—	0.15	0.29 **	—
	T_{er}	−0.19	−0.12	—	−0.10	0.11	—

Correlations are bivariate correlations between latent factors for the RT scores (1–12) and standardized regression coefficients for the model parameters entered as correlated predictors (13–14). Search = Search/CRT tasks, Comp = Comparison tasks, Subst = Substitution tasks. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

3.4. Moderation of the WPR by Task Complexity (H4)

Finally, we tested whether the mean response time becomes more predictive of cognitive ability with increasing task complexity. In fact, there was a substantial relation of μ additionally to τ in the

Substitution tasks (see Table 2). Thereby, individual differences in mean RT, in addition to differences in the proportion of extreme values, were found to be related with cognitive ability in the most complex tasks. These effects were parsimoniously reflected by individual differences in drift rate ν in the diffusion model. Drift rates tended to be more highly related with ability in the more complex Comparison tasks compared with the simpler Search tasks.

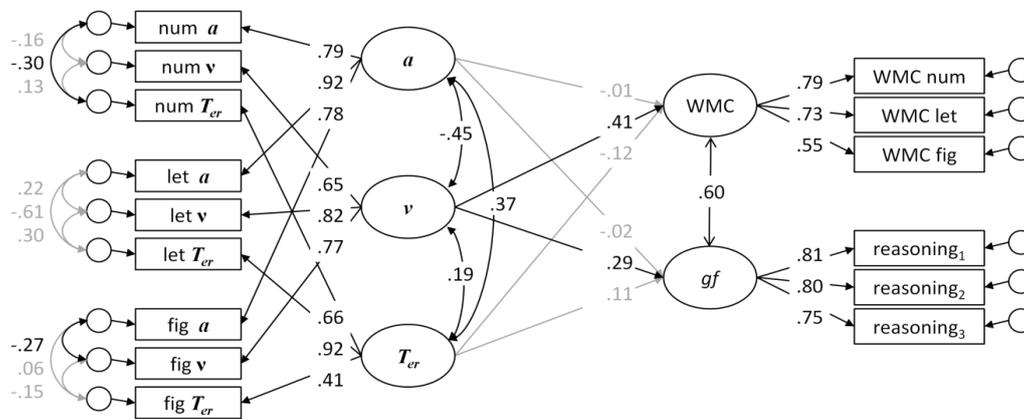


Figure 4. Parameters of the diffusion model fit to Comparison tasks as predictors of WMC and gf . Significant parameters are displayed in black font.

3.5. Testing Model Parameters as Trait-Like Dispositions (H5)

The latent correlations of parameter estimates across task classes are displayed in Figure 5 for the diffusion model (upper panel) and the ex-Gaussian model (lower panel). Gray dots correspond with independent correlation estimates across task classes. The box-and-whisker plots summarize the distribution statistics. The median correlation is highlighted in red. For the diffusion model, cross-task correlations were high and relatively specific for the drift rate (ν) and for boundary separation (a), but not for non-decision time (T_{er}). Cross-task correlations tended to be more moderate for the ex-Gaussian parameter estimates and were less specific for the respective parameters. The highest correlations were observed for the τ parameter, while the σ parameter was only moderately correlated across task classes.

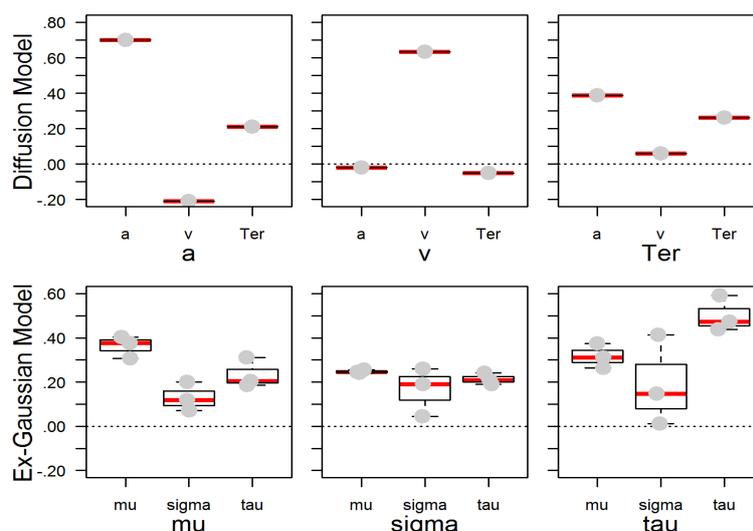


Figure 5. Cross-task correlations of latent variables for the diffusion model and the ex-Gaussian parameters.

The correlations of parameters across models and with conventional RT and accuracy scores are displayed in Figure 6. Generally, correlations across task classes were moderate, with the highest correlations in the magnitude of $r = |0.4|$. The correlations of the diffusion model with ex-Gaussian parameters are displayed in the lower left part of Figure 6. The drift rate (v) was inversely correlated with all ex-Gaussian parameters, but the magnitude of the correlation was higher for the variability parameters than for the μ parameter. The strongest correlation was observed with the τ parameter. Response caution (a) was positively correlated with μ and τ , but there was little correlation with sigma on average. Non-decision time (T_{er}) was positively correlated with μ , but there was no substantial correlation with the other parameters.

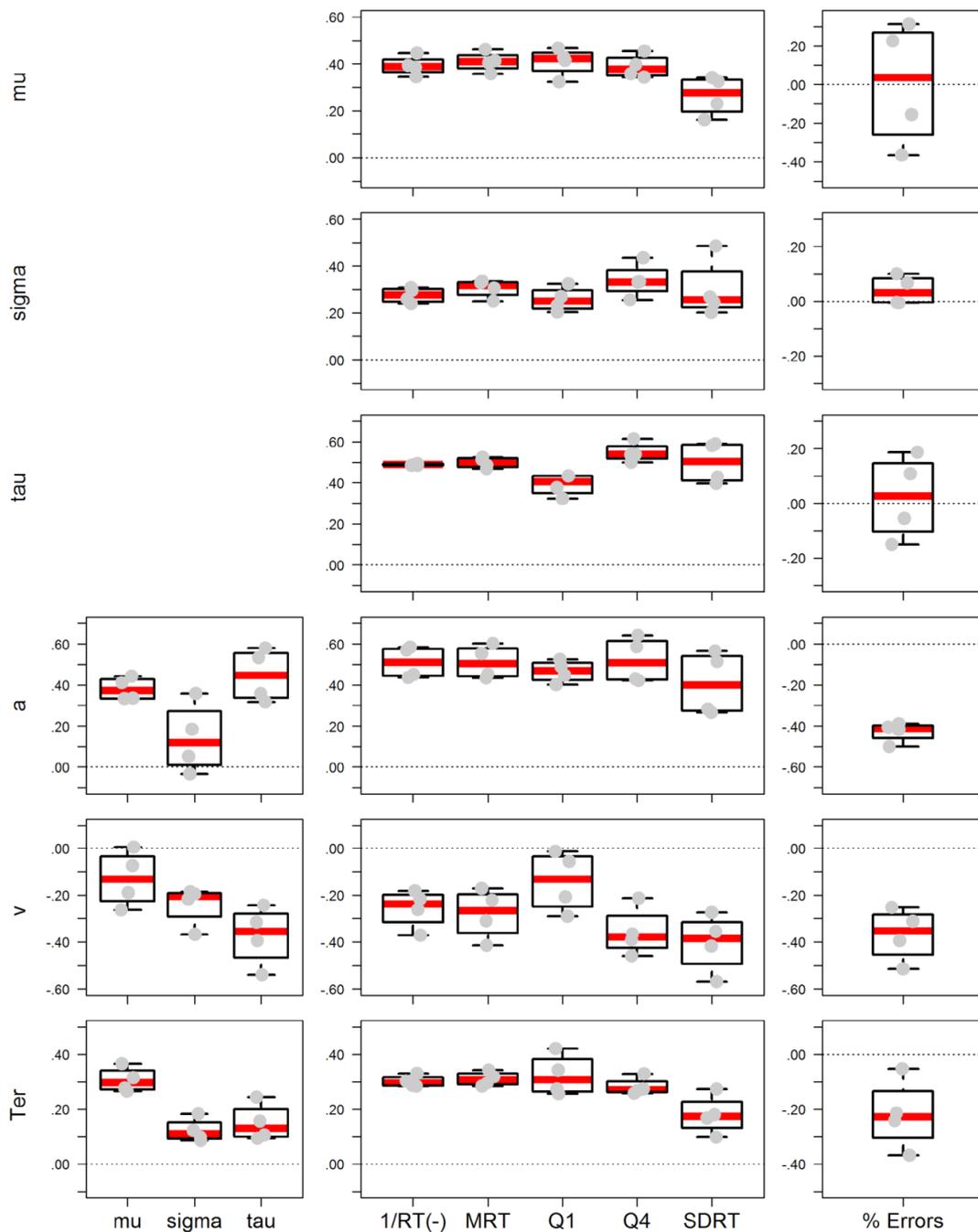


Figure 6. Correlations of latent variables for the diffusion model, the ex-Gaussian parameters, and with RT and accuracy scores across different tasks.

Correlations of the diffusion model parameters with conventional RT scores are displayed in the lower center part. Drift rates were inversely correlated with all RT scores, but correlations were considerably higher (negative) with the slow quantiles than with the fast quantiles of the RT distribution. Additionally, the negative relation with SD_{RT} exceeded the average correlation obtained for M_{RT} . Response caution (a) was positively correlated with all location scores in a comparable magnitude. Non-decision (T_{er}) time also showed a moderate positive relation with the location scores, while its relation with SD_{RT} was somewhat weaker. All diffusion model parameters were negatively correlated with the error rate in the other tasks, in particular drift rate and response caution.

Relations of the ex-Gaussian parameters were analogous. The μ parameter was moderately and comparably correlated with all RT location scores, but somewhat more weakly correlated with SD_{RT} . The σ parameter was moderately related with all RT location scores, but weaker than μ . Actually, the τ parameter showed the highest correlations with the RT scores compared with the other ex-Gaussian parameters. Its correlations were stronger with the slow quantiles compared with the fast quantiles of the RT distribution in the other tasks. Please note that the average correlations with error rate were virtually zero.

4. Discussion

We conducted this study to test the correlation of response times in elementary cognitive tasks with WMC and gf . A set of frequently used scores and model-based parameters derived from RT distributions was investigated in tasks of varying complexity.

Correlations of mental speed with WMC and gf (H1): In line with previous research [1–3], we predicted that mental speed is moderately correlated with WMC and gf . This hypothesis was confirmed in the current study: the faster participants completed elementary cognitive tasks, the higher their WMC. The same was also found for gf , although to a lesser extent. The correlative relation of mental speed with WMC can be reconciled with both theories that consider mental speed a causal constituent of WMC [5], as well as with a more technical interpretation in terms of overlapping task requirements, such as a speed requirement in WMC or in the gf tasks as a confounding factor [10]. In fact, the somewhat larger correlation of speed with WMC compared to gf could result from the faster presentation of stimuli and response deadlines in the recall-one-back tasks employed in the current study.

However, other mechanisms could also contribute to the higher correlation of speed with WMC. Given that the correlation was largely driven by the exponential component (τ) of the RT distribution, an account in terms of attentional lapses could hold: Stimuli are continuously presented in the recall-one-back task, and if participants do not pay attention for a short moment, they can miss a stimulus, thereby losing credit points. The situation is different for the figural reasoning test. Since the task is largely self-paced, participants can easily restart the solution process and still arrive at the correct response in time. Therefore, task performance in gf tests can be predicted to be less strongly correlated with attentional lapses. In fact, previous research has shown that increasing speed pressure in a reasoning task (used as a gf measure) increases the correlation of gf with WMC [60]. Further, the narrow scope of the gf factor, which is identified only by items of a single test, could contribute to the reduced correlation with the speed factor. Finally, and in line with the notion that WMC constitutes the basis of gf [7,8], indirect effects of mental speed via WMC could explain the lower correlation. At the same time, reasoning ability likely requires more than capacity of working memory alone (e.g., knowledge of rules and their application). Thereby, a speed-related mechanism will only explain some of the total variance in reasoning tasks.

Differential validity of RT scores (H2): We predicted that alternative scores derived from RT data are differentially related with cognitive ability. This second hypothesis was motivated on theoretical grounds and simulations [35], as well as on the basis of previous findings [2,18,19,25]. In line with the Worst Performance Rule [19], we predicted that scores reflecting or biased by slow RTs are more highly related with cognitive ability than those scores that are less affected by extreme values. In fact, there was a descriptive increase in the magnitude of the correlation across RT bands, although

the increase was only small in magnitude and appeared to be confined to the correlation with WMC, but not with gf . We additionally tested whether mean RT (M_{RT}), which is potentially biased by extreme RT values, is more highly correlated with cognitive ability than those location scores that are less affected by extreme RT values (Mdn_{RT} , $M_{\log(RT)}$) [18,25]. However, there was virtually no difference in their correlations. Apparently, a few outlier values did not distort the rank order of participants. This finding suggests that all of the conventional mean RT scores can be used as predictors of cognitive ability and none of them is clearly superior over the others. In addition, scores of RT variability (SD_{RT}) were not found to be more strongly related with ability than with mean RT (M_{RT}), confirming findings from a recent meta-analysis [15]. As expected, the error scores were less reliable than the RT scores. Not surprisingly, their correlations obtained with WMC and gf were low. The fit of the error score model was still good. Consequently, the probit transformation did not considerably improve the fit of the model, but slightly decreased validity of the scores. This result suggests that the correlations in error variables are largely driven by participants with more extreme error proportions.

Validity of parameters sensitive to the right tail of the distribution (H3): We predicted that the effects of extreme RT values are more clearly related to model-based parameters that either reflect the tail of the RT distribution (τ in the ex-Gaussian model) or are sensitive to the tail of the RT distribution (such as the drift rate ν in the diffusion model) [20,21,35,38]. Confirming these predictions, the relations of the ex-Gaussian τ parameter and of the diffusion model drift rate ν with WMC and gf were generally replicated across task classes.

While the pattern of observed parameter correlations was replicated, the magnitude of the correlations was only moderate in strength compared with the Schmiedek et al. study [38]. In part, this result may be due to differences in task characteristics that affect the level and distribution of parameter estimates: our task-homogeneous measurement models likely reflect, in part, task-specific parameter effects. The more heterogeneous measurement model in the Schmiedek et al. [38] study may capture communality across all paradigms, and thus the so-obtained parameter factors may reflect correlations with ability at a higher level.

Generally, parameter estimates were not as highly correlated with ability as conventional RT scores. One reason for this can be that the information in empirical data is “distributed” across parameters, whereas it is “aggregated” in the conventional scores. In that sense, reliability may be traded for theoretical clarity in the models. Additionally, some of the modeling benefits may not come into play in the current study: the error rate was generally low and there was only limited variance in the errors. Consequently, there may not have been much of a speed-accuracy trade-off that would have distorted the rank order of participants in RT scores. Therefore, the diffusion model had only a limited chance to demonstrate its inherent advantage.

Moderation of the WPR by task complexity (H4): We predicted that individual differences in simple, elementary cognitive tasks are largely reflected in the proportion of extreme values [2,24]. However, the average response time was expected to gain predictive validity as task complexity increased. Confirming these predictions, the exponential parameter τ showed moderate relations in all tasks, whereas the mean response time (μ) was correlated with ability only in the most complex task class. These effects were parsimoniously reflected by individual differences in drift rate ν in the diffusion model.

At first glance, these effects in mean RT (μ) appear to contradict previous findings of increased WPR effects (i.e., higher correlations of slow RT bands) with increasing task complexity [2,13,19]. However, it needs to be considered that WPR effects have been traditionally investigated by analyzing RT bands. In turn, individual differences in RT bands are not only determined by the skew of the individual RT distributions, but also by individual differences in their locations. We believe the ex-Gaussian decomposition is a more adequate data description, since it helps dissociate components that are confounded in RT bands. This interpretation is supported by the differential parameter correlations observed in the current study, as well as in previous research using the ex-Gaussian model [21,38].

Testing model parameters as trait-like dispositions (H5): Cross-task correlations of parameters within models were moderate to high (but far from unity). Generally, there appeared to be a better correspondence of diffusion model parameters than those of the ex-Gaussian decomposition across different task classes. It is of note that these correlations were only moderate, given that correlations were estimated across latent factors capturing the reliable portion of parameter variance and that measurement models were generally satisfactory (see Appendix A). In part, the only moderate and not very specific parameter correlations in the ex-Gaussian model may result from the fact that all parameters are estimated from the same distribution of correct RT values. The diffusion model surely profits from also using error rate for the estimation of boundary separation (a), while drift rate (ν) largely accounts for location, skew, and variability of the correct RT distribution in the absence of competing performance parameters. Some of the problems in non-decision time may result from the computational procedure in EZ [50], where non-decision time is computed as a residual after removing decision time (as function of ν and a) from M_{RT} . However, comparably moderate retest correlations were also observed using other estimation procedures [42].

Nevertheless, the pattern of parameter correlations closely replicated that previously observed in other studies [37,38,43], although relationships were generally attenuated if computed across task classes. This result was somewhat expected, as correlations across different experimental tasks remove adverse effects of random variation, contamination, and computational dependencies. In part, this result may reflect previously demonstrated moderating effects in the variation of other parameters [35] and specificity of task requirements in the different task classes. Confirming previous findings, the drift rate was more highly correlated with variability of responding (SD_{RT}) than with mean response times (M_{RT}) [35], even when correlations were estimated across non-overlapping data sets. Still then, SD_{RT} was not more highly related with ability than M_{RT} , in line with the recent meta-analysis on RT and intelligence [15]. This finding suggests that drift rate is not the only “determinant” of a high task score. It is of note for the assessment of mental speed that the correlations of the response caution (a) parameter with virtually all location scores suggests the setting of speed-accuracy trade-off contributes as well to the observed RT scores in elementary tasks. The virtual null correlations of the ex-Gaussian parameters with error rate might be due to the constraint that ex-Gaussian parameters are estimated from the distribution of correct response times only. This observation suggests that errors contain independent (possibly incremental) information to RT scores. However, the null relations could also be the result of an artifact from the joint effects of processing efficiency (which would result in a positive relation of RT and error rate) and of the speed-accuracy trade-off (which would result in a negative relation of RT and error rate). The diffusion model is better suited to dissociate both components.

Equivalence with paper-and-pencil tests: In this study, we tested different RT scoring and modeling approaches that can be employed for data obtained with computerized elementary tasks. Most of these scores meet the psychometric criteria in terms of reliability (factor saturation) and validity (of cognitive ability), and a few of them may be particularly informative from a theoretical perspective. However, with the exception of (non-corrected) mean RT (or $1/RT$) and the error rate, the computation of virtually all other scores requires individual RT data. The latter are typically not available in paper-and-pencil tests. Therefore, the equivalence of computer and paper versions cannot be tested for all scores. Fitting diffusion models is not possible in paper tests either, making it more difficult to quantify and compensate possible speed-accuracy trade-offs. In turn, the possibility of modeling task performance, thereby dissociating components of the decision process, makes chronometric tasks and their data analyses an appealing domain of their own. In this study, we demonstrated the validity of the derived scores in terms of correlations with WMC and gf . What still needs to be tested is the equivalence of paper and computerized speed tests using aggregate scores that can be computed in both versions analogously.

Structure of mental speed: In the present study, we administered typical clerical speed tasks [44]. Although all of them are conventionally considered to be elementary cognitive tasks, they differed

considerably in complexity, as was indicated by the time required to produce a correct response. Psychometrically, a bifactor model [59] best described the structure of the speed tasks, with a g factor of mental speed and task class-specific nested factors. This model confirms previous findings of mental speed as a hierarchical construct [61,62]. In this study, we additionally showed that the pattern of correlations of model-based parameters differed across task classes (with μ contributing to ability correlations only in the most complex task). This finding suggests that different cognitive operations are involved in the tasks, and that task specificity could be more than method-related variance. Still, the high communality of all tasks justifies the extraction of a g factor in a hierarchical model for psychometric purposes.

5. Limitations

The participants in the current study had an above-average level of education. Given that they were more skilled at computer work and capable of concentrating on the task, they might have experienced fewer attentional fluctuations or lapses that would result in reduced levels and variance in the τ parameter or drift rate ν . In turn, limited variance could have contributed to the reduced correlations observed, relative to the Schmiedek et al. study [38], among the other factors discussed.

The lower correlation of speed with gf relative to WMC could reflect that only one reasoning task was used, resulting in a relatively narrow ability factor. Correlations should be higher if other tasks or stimulus domains were also incorporated in the measurement model.

Finally, task complexity was not manipulated within tasks, but differed between task classes. In order to rule out the possibility that other factors drove the effects, it is desirable to replicate the current study with tasks where complexity (e.g., the number of S-R mapping rules) is manipulated within the same task class.

The moderate to high correlations of parameter estimates across tasks suggests a certain breadth of the captured characteristics. However, this does not elucidate temporal stability, another prerequisite of “trait-like dispositions” (however, see [42]).

6. Conclusions

Correlations of mental speed, WMC, and gf were found to be mediocre to large in the present study. All mean RT scores were reliable and their correlations with cognitive ability were highly comparable. RT variability scores were inversely correlated with ability, but their correlations were not as high as for mean RT. Limited evidence suggests that slower RT bands are more predictive of WMC than faster RT bands. An ex-Gaussian decomposition of the RT distributions was superior to conventional scores in showing that the correlation with ability was selectively driven by extreme values in easy tasks. With increasing task complexity, mean response time became predictive of ability, additionally to the proportion of outliers. In the diffusion model, the correlation of RT with ability was parsimoniously accounted for in terms of individual differences in drift rate. Some of the correlations of model parameters across task classes suggest parameters have some breadth, while pointing as well to the considerable effects of task-specific requirements.

Acknowledgments: The research reported in this paper was supported by grant SCHM 3235/3-1 from the German Research Foundation (DFG) accorded to F.S. and by grant WI 2667-4 from the German Research Foundation (DFG) accorded to O.W.

Author Contributions: F.S. and O.W. conceived and designed the research; F.S. conducted the study and analyzed the data. F.S. wrote a first version of the manuscript which was revised by O.W. Both authors approved the final version.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Measurement models and model fit.

#	Model	Measurement Models			Model Fit			
		Saturation	Loadings	χ^2	df	RMSEA	CFI	
0	<i>Cognitive Ability</i> WMC <i>gf</i>	0.74	0.69 (0.55–0.78)	12	8	0.048	0.991	
		0.83	0.77 (0.75–0.79)					
1	<i>Speed</i> +1/RT	0.74	0.64 (0.46–0.95)	132	81	0.056	0.975	
2	<i>Response Time</i> +M _{RT}	0.74	0.64 (0.46–0.93)	130	81	0.055	0.975	
3	+M _{log(RT)}	0.73	0.64 (0.45–0.95)	129	81	0.054	0.976	
4	+M _{dnRT}	0.70	0.61 (0.42–0.92)	122	81	0.050	0.977	
5	<i>Quartiles</i> +Q1	0.64	0.56 (0.35–0.92)	140	81	0.060	0.965	
6	+Q2	0.69	0.61 (0.43–0.94)	125	81	0.052	0.976	
7	+Q3	0.71	0.62 (0.43–0.92)	122	81	0.050	0.978	
8	+Q4	0.76	0.64 (0.47–0.88)	152	81	0.066	0.960	
9	<i>Variability</i> +SD _{RT}	0.67	0.53 (0.34–0.82)	154	81	0.067	0.948	
10	+IQR _{RT}	0.54	0.44 (0.29–0.74)	155	81	0.068	0.926	
11	<i>Errors</i> +Error Rate	0.66	0.54 (0.34–0.69)	126	81	0.052	0.953	
12	+Probit (Error)	0.64	0.54 (0.46–0.66)	112	81	0.044	0.963	
13a	<i>Ex-Gaussian Model</i> +Search/CRT	μ	0.83	0.79 (0.71–0.89)	133	71	0.066	0.959
		σ	0.50	0.48 (0.23–0.71)				
		τ	0.55	0.58 (0.53–0.66)				
13b	+Comparison	μ	0.81	0.76 (0.54–0.94)	84	71	0.030	0.991
		σ	0.55	0.54 (0.40–0.64)				
		τ	0.84	0.79 (0.69–0.92)				
13c	+Substitution	μ	0.83	0.79 (0.70–0.86)	136	71	0.068	0.958
		σ	0.61	0.55 (0.35–0.74)				
		τ	0.63	0.61 (0.58–0.64)				
14a	<i>Diffusion Model</i> +Search/CRT	<i>a</i>	0.80	0.74 (0.72–0.77)	125	71	0.062	0.959
		<i>v</i>	0.81	0.75 (0.70–0.82)				
		<i>T_{er}</i>	0.86	0.83 (0.76–0.93)				
14b	+Comparison	<i>a</i>	0.87	0.83 (0.78–0.92)	99	71	0.044	0.976
		<i>v</i>	0.79	0.75 (0.65–0.82)				
		<i>T_{er}</i>	0.70	0.66 (0.41–0.92)				

Table A1 summarizes measurement models and fit statistics for the models used to estimate latent relations of conventional speed scores and model parameters with the ability constructs reported in Table 2. All models were independently fitted to the data. Model 0 only comprised the two ability constructs WMC and *gf* that were modeled as correlated latent factors with three observed indicators each (as shown on the right side of Figures 3 and 4). Models 1–12 comprised a measurement model for the respective speed score, additionally to the measurement models of the two ability factors. All measurement models of speed scores were analogous bifactor models, comprising one general factor of mental speed that accounted for the communality of all nine speed indicators (i.e., 3 task types × 3 stimulus materials), and two nested task-type specific factors that captured the shared method variance of the Search and Comparison tasks (see left side of Figure 3). Only the general factor

was allowed to be correlated with the ability constructs, but not the method factors. Measurement models for the model parameters were fitted separately for each task type and additionally comprised the two ability factors (see Figure 4). Following the rationale described in Schmiedek et al. [38] (see also Section 2.3), all models specified three correlated latent parameter factors for each of the three model parameters. Error terms of parameters simultaneously estimated from the same data were allowed to be correlated. The first column in this table reports the factor saturation (McDonald's ω) in the indicators of the measurement models: Hierarchical ω is given for the g factor in the bifactor measurement models fitted for the conventional RT and error-based speed scores. Total ω is reported for the three-item WMC/ gf and model parameter measurement models. Additionally, the second column gives the mean (range) of item loadings on the speed g factor (Models 1–12) or on their corresponding parameter factor (Models 13a–14b). The four columns on the right side of this table report fit statistics for each of the independently fitted models.

References

1. Sheppard, L.D.; Vernon, P.A. Intelligence and speed of information-processing: A review of 50 years of research. *Personal. Individ. Differ.* **2008**, *44*, 535–551. [[CrossRef](#)]
2. Jensen, A.R. *Clocking the Mind: Mental Chronometer Individual Differences*; Elsevier: Amsterdam, The Netherlands, 2006; p. xi 272.
3. Deary, I.J.; Der, G.; Ford, G. Reaction times and intelligence differences: A population-based cohort study. *Intelligence* **2001**, *29*, 389–399. [[CrossRef](#)]
4. Eysenck, H.J. Speed of information processing, reaction time, and the theory of intelligence. In *Speed of Information-Processing and Intelligence*; Vernon, P.A., Ed.; Ablex Publishing: Westport, CT, USA, 1987; pp. 21–67.
5. Barrouillet, P.; Bernardin, S.; Portrat, S.; Vergauwe, E.; Camos, V. Time and cognitive load in working memory. *J. Exp. Psychol. Learn. Mem. Cogn.* **2007**, *33*, 570–585. [[CrossRef](#)] [[PubMed](#)]
6. Barrouillet, P.; Portrat, S.; Camos, V. On the law relating processing to storage in working memory. *Psychol. Rev.* **2011**, *118*, 175–192. [[CrossRef](#)] [[PubMed](#)]
7. Kane, M.J.; Hambrick, D.Z.; Tuholski, S.W.; Wilhelm, O.; Payne, T.W.; Engle, R.W. The Generality of Working Memory Capacity: A Latent-Variable Approach to Verbal and Visuospatial Memory Span and Reasoning. *J. Exp. Psychol. Gen.* **2004**, *133*, 189–217. [[CrossRef](#)] [[PubMed](#)]
8. Süß, H.-M.; Oberauer, K.; Wittmann, W.W.; Wilhelm, O.; Schulze, R. Working-memory capacity explains reasoning ability—And a little bit more. *Intelligence* **2002**, *30*, 261–288. [[CrossRef](#)]
9. Stankov, L.; Roberts, R.D. Mental speed is not the "basic" process of intelligence. *Personal. Individ. Differ.* **1997**, *22*, 69–84. [[CrossRef](#)]
10. Wilhelm, O.; Schulze, R. The relation of speeded and unspeeded reasoning with mental speed. *Intelligence* **2002**, *30*, 537–554. [[CrossRef](#)]
11. Fernandez, S.; Fagot, D.; Dirk, J.; de Ribaupierre, A. Generalization of the worst performance rule across the lifespan. *Intelligence* **2014**, *42*, 31–43. [[CrossRef](#)]
12. Hick, W. Information theory and intelligence tests. *Br. J. Psychol.* **1951**, *4*, 157–164. [[CrossRef](#)]
13. Roth, E. The speed of information processing and its relation to intelligence. *Z. Exp. Angew. Psychol.* **1964**, *11*, 616–622.
14. Williams, B.R.; Hulstsch, D.F.; Strauss, E.H.; Hunter, M.A.; Tannock, R. Inconsistency in reaction time across the lifespan. *Neuropsychology* **2005**, *19*, 88–96. [[CrossRef](#)] [[PubMed](#)]
15. Doebler, P.; Scheffler, B. The relationship of choice reaction time variability and intelligence: A meta-analysis. *Learn. Individ. Differ.* **2016**. [[CrossRef](#)]
16. Wagenmakers, E.-J.; Brown, S. On the linear relation between the mean and the standard deviation of a response time distribution. *Psychol. Rev.* **2007**, *114*, 830–841. [[CrossRef](#)] [[PubMed](#)]
17. Schmitz, F.; Teige-Mocigemba, S.; Voss, A.; Klauer, K.C. When scoring algorithms matter: Effects of working memory load on different IAT scores. *Br. J. Soc. Psychol.* **2013**, *52*, 103–121. [[CrossRef](#)] [[PubMed](#)]
18. Jensen, A.R. The importance of intraindividual variation in reaction time. *Personal. Individ. Differ.* **1992**, *13*, 869–881. [[CrossRef](#)]

19. Larson, G.E.; Alderton, D.L. Reaction time variability and intelligence: A 'worst performance' analysis of individual differences. *Intelligence* **1990**, *14*, 309–325. [[CrossRef](#)]
20. Coyle, T.R. A review of the worst performance rule: Evidence, theory, and alternative hypotheses. *Intelligence* **2003**, *31*, 567–587. [[CrossRef](#)]
21. Unsworth, N.; Redick, T.S.; Lakey, C.E.; Young, D.L. Lapses in sustained attention and their relation to executive control and fluid abilities: An individual differences investigation. *Intelligence* **2010**, *38*, 111–122. [[CrossRef](#)]
22. Salthouse, T.A. Relations of successive percentiles of reaction time distributions to cognitive variables and adult age. *Intelligence* **1998**, *26*, 153–166. [[CrossRef](#)]
23. Ratcliff, R.; Thapar, A.; McKoon, G. Individual differences, aging, and IQ in two-choice tasks. *Cogn. Psychol.* **2010**, *60*, 127–157. [[CrossRef](#)] [[PubMed](#)]
24. Kranzler, J.H. A test of Larson and Alderton's (1990) worst performance rule of reaction time variability. *Personal. Individ. Differ.* **1992**, *13*, 255–261. [[CrossRef](#)]
25. Baumeister, A.A. Intelligence and the "personal equation". *Intelligence* **1998**, *26*, 255–265. [[CrossRef](#)]
26. Heathcote, A.; Popiel, S.J.; Mewhort, D. Analysis of response time distributions: An example using the Stroop task. *Psychol. Bull.* **1991**, *109*, 340–347. [[CrossRef](#)]
27. Hohle, R.H. Inferred components of reaction times as functions of foreperiod duration. *J. Exp. Psychol.* **1965**, *69*, 382–386. [[CrossRef](#)] [[PubMed](#)]
28. Luce, R.D. *Response Times: Their Role in Inferring Elementary Mental Organization*; Oxford University Press: New York, NY, USA, 1986.
29. Ratcliff, R. A theory of memory retrieval. *Psychol. Rev.* **1978**, *85*, 59–108. [[CrossRef](#)]
30. Ratcliff, R.; Rouder, J.N. Modeling response times for two-choice decisions. *Psychol. Sci.* **1998**, *9*, 347–356. [[CrossRef](#)]
31. Borter, N.; Troche, S.; Dodonova, Y.; Rammsayer, T. A Latent Growth Curve (LGC) Analysis to Model Task Demands and the Worst Performance Rule Simultaneously. In Proceedings of the European Mathematical Psychology Group Meeting, Tübingen, Germany, 30 July–1 August 2014.
32. Leth-Steensen, C.; Elbaz, Z.K.; Douglas, V.I. Mean response times, variability, and skew in the responding of ADHD children: A response time distributional approach. *Acta Psychol.* **2000**, *104*, 167–190. [[CrossRef](#)]
33. Ratcliff, R.; Murdock, B.B. Retrieval processes in recognition memory. *Psychol. Rev.* **1976**, *83*, 190–214. [[CrossRef](#)]
34. Ratcliff, R.; Tuerlinckx, F. Estimating parameters of the diffusion model: Approaching to dealing with contaminant reaction and parameter variability. *Psychon. Bull. Rev.* **2002**, *9*, 438–481. [[CrossRef](#)] [[PubMed](#)]
35. Van Ravenzwaaij, D.; Brown, S.; Wagenmakers, E.-J. An integrated perspective on the relation between response speed and intelligence. *Cognition* **2011**, *119*, 381–393. [[CrossRef](#)] [[PubMed](#)]
36. Ratcliff, R.; Schmiedek, F.; McKoon, G. A diffusion model explanation of the worst performance rule for reaction time and IQ. *Intelligence* **2008**, *36*, 10–17. [[CrossRef](#)] [[PubMed](#)]
37. Spieler, D.H. Modelling age-related changes in information processing. *Eur. J. Cogn. Psychol.* **2001**, *13*, 217–234. [[CrossRef](#)]
38. Schmiedek, F.; Oberauer, K.; Wilhelm, O.; Suess, H.-M.; Wittmann, W.W. Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *J. Exp. Psychol. Gen.* **2007**, *136*, 414–429. [[CrossRef](#)] [[PubMed](#)]
39. Sternberg, S. The discovery of processing stages: Extensions of Donders' method. *Acta Psychol. Amst.* **1969**, *30*, 276–315. [[CrossRef](#)]
40. Sternberg, S. High speed scanning in human memory. *Science* **1966**, *153*, 652–654. [[CrossRef](#)] [[PubMed](#)]
41. Voss, A.; Nagler, M.; Lerche, V. Diffusion models in experimental psychology: A practical introduction. *Exp. Psychol.* **2013**, *60*, 385–402. [[CrossRef](#)] [[PubMed](#)]
42. Lerche, V.; Voss, A. Retest reliability of the parameters of the Ratcliff diffusion model. *Psychol. Res.* **2016**. [[CrossRef](#)] [[PubMed](#)]
43. Matzke, D.; Wagenmakers, E.-J. Psychological interpretation of the ex-Gaussian and shifted Wald parameters: A diffusion model analysis. *Psychon. Bull. Rev.* **2009**, *16*, 798–817. [[CrossRef](#)] [[PubMed](#)]
44. Roberts, R.D.; Stankov, L. Individual differences in speed of mental processing and human cognitive abilities: Toward a taxonomic model. *Learn. Individ. Differ.* **1999**, *11*, 1–120. [[CrossRef](#)]

45. Wilhelm, O.; Hildebrandt, A.; Oberauer, K. What is working memory capacity, and how can we measure it? *Front. Psychol.* **2013**, *4*, 433. [[CrossRef](#)] [[PubMed](#)]
46. Conway, A.R.A.; Kane, M.J.; Bunting, M.F.; Hambrick, D.Z.; Wilhelm, O.; Engle, R.W. Working memory span tasks: A methodological review and user's guide. *Psychon. Bull. Rev.* **2005**, *12*, 769–786. [[CrossRef](#)] [[PubMed](#)]
47. Wilhelm, O.; Schroeders, U.; Schipolowski, S. *Berliner Test zur Erfassung Fluidier und Kristalliner Intelligenz für Die 8. Bis 10. Jahrgangsstufe (BEFKI 8–10)*; Hogrefe: Göttingen, Germany, 2014.
48. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2016.
49. Massidda, D. Retimes: Reaction Time Analysis. R Package Version 0.1–2. Available online: <https://CRAN.R-project.org/package=retimes> (accessed on 21 March 2016).
50. Wagenmakers, E.-J.; van der Maas, H.L.J.; Grasman, R.P.P. An EZ-diffusion model for response time and accuracy. *Psychon. Bull. Rev.* **2007**, *14*, 3–22. [[CrossRef](#)] [[PubMed](#)]
51. Revelle, W. *Psych: Procedures for Personality and Psychological Research, 1.5.1.*; Northwestern University: Evanston, IL, USA, 2015.
52. Muthén, L.K.; Muthén, B.O. *Mplus User's Guide*, 7th ed.; Muthén & Muthén: Los Angeles, CA, USA, 2012.
53. Van Zandt, T. How to fit a response time distribution. *Psychon. Bull. Rev.* **2000**, *7*, 424–465. [[CrossRef](#)] [[PubMed](#)]
54. Van Ravenzwaaij, D.; Oberauer, K. How to use the diffusion model: Parameter recovery of three methods: Ez, fast-dm, and DMAT. *J. Math. Psychol.* **2009**, *53*, 463–473. [[CrossRef](#)]
55. Ratcliff, R.; Childers, R. Individual Differences and Fitting Methods for the Two-Choice Diffusion Model of Decision Making. *Decision* **2015**. [[CrossRef](#)] [[PubMed](#)]
56. Lerche, V.; Voss, A.; Nagler, M. How many trials are required for parameter estimation in diffusion modeling? A comparison of different optimization criteria. *Behav. Res. Methods* **2016**. [[CrossRef](#)] [[PubMed](#)]
57. Botvinick, M.M.; Braver, T.S.; Barch, D.M.; Carter, C.S.; Cohen, J.D. Conflict monitoring and cognitive control. *Psychol. Rev.* **2001**, *108*, 624–652. [[CrossRef](#)] [[PubMed](#)]
58. Tukey, J.W. *Exploratory Data Analysis*; Addison-Wesley Reading: Boston, MA, USA, 1977.
59. Eid, M.; Nussbeck, F.W.; Geiser, C.; Cole, D.A.; Gollwitzer, M.; Lischetzke, T. Structural equation modeling of multitrait-multimethod data: Different models for different types of methods. *Psychol. Methods* **2008**, *13*, 230–253. [[CrossRef](#)] [[PubMed](#)]
60. Chuderski, A. When are fluid intelligence and working memory isomorphic and when are they not? *Intelligence* **2013**, *41*, 244–262. [[CrossRef](#)]
61. Danthiir, V.; Wilhelm, O.; Roberts, R.D. Further evidence for a multifaceted model of mental speed: Factor structure and validity of computerized measures. *Learn. Individ. Differ.* **2012**, *22*, 324–335. [[CrossRef](#)]
62. Danthiir, V.; Wilhelm, O.; Schulze, R.; Roberts, R.D. Factor structure and validity of paper-and-pencil measures of mental speed: Evidence for a higher-order model? In *Intelligence*; Wilhelm, O., Engle, R.W., Eds.; Elsevier Science: Amsterdam, The Netherlands, 2005; Volume 33, pp. 491–514.

