Intelligence 102 (2024) 101804

Contents lists available at ScienceDirect

Intelligence

journal homepage: www.elsevier.com/locate/intell

Don't waste your time measuring intelligence: Further evidence for the validity of a three-minute speeded reasoning test

Anna-Lena Schubert [a,*], Christoph Löffler [a,b], Clara Wiebel [a], Florian Kaulhausen [c], Tanja Gabriele Baudson [d,e]

[a] Department of Psychology, University of Mainz, Mainz, Germany
[b] Department of Psychology, Heidelberg University, Heidelberg, Germany
[c] Department of Psychology, Free University of Berlin, Berlin, Germany
[d] Department of of Human Sciences, Vinzenz Pallotti University, Vallendar, Germany
[e] Institute for Globally Distributed Open Research and Education (IGDORE), Sweden

ARTICLE INFO

Keywords:
Intelligence screening
Cognitive abilities
Speeded reasoning
Validity
mini-q

ABSTRACT

The rise of large-scale collaborative panel studies has generated a need for fast, reliable, and valid assessments of cognitive abilities. In these studies, a detailed characterization of participants' cognitive abilities is often unnecessary, leading to the selection of tests based on convenience, duration, and feasibility. This often results in the use of abbreviated measures or proxies, potentially compromising their reliability and validity. Here we evaluate the mini-q (Baudson & Preckel, 2016), a three-minute speeded reasoning test, as a brief assessment of general cognitive abilities. The mini-q exhibited excellent reliability (0.96–0.99) and a substantial correlation with general cognitive abilities measured with a comprehensive test battery ($r = 0.57$; age-corrected $r = 0.50$), supporting its potential as a brief screening of cognitive abilities. Working memory capacity accounted for the majority (54%) of the association between test performance and general cognitive abilities, whereas individual differences in processing speed did not contribute to this relationship. Our results support the notion that the mini-q can be used as a brief, reliable, and valid assessment of general cognitive abilities. We therefore developed a computer-based version, ensuring its adaptability for large-scale panel studies. The paper- and computer-based versions demonstrated scalar measurement invariance and can therefore be used interchangeably. We provide norm data for young (18 to 30 years) and middle-aged (31 to 60 years) adults and provide recommendations for incorporating the mini-q in panel studies. Additionally, we address potential challenges stemming from language diversity, wide age ranges, and online testing in such studies.

1. Introduction

The rise of large-scale collaborative panel studies has generated a need for fast, reliable, and valid assessments of cognitive abilities. Traditionally, individual differences in cognitive abilities are assessed using extensive test batteries such as the Wechsler Adult Intelligence Test (WAIS-IV; Wechsler, 2008), the Intelligence Structure Test (I-S-T 2000 R; Liepmann, Beauducel, Brocke, and Amthauer, 2010), the Berlin Intelligence Structure Test (BIS; Jäger, Süß, and Beauducel, 1997), or the Armed Services Vocational Aptitude Battery (AVSAB; U.S. Department of Defense, 1984). Alternatively, an individual's cognitive ability is sometimes inferred from highly *g*-loaded fluid reasoning tests such as Raven's Advanced Progressive Matrices (APM; Raven, Court, and Raven, 1994) or the Bochum Matrices Test Advanced (BOMAT; Hossiep, Turck, and Hasella, 1999). All of these established tests possess excellent psychometric properties. Furthermore, the more extensive test batteries often allow for assessing relative strengths and weaknesses of an individual's cognitive ability profile in a very detailed manner. However, they also have in common that they take relatively long to administer: Many fluid reasoning tests require about one to two hours, and cognitive ability test batteries usually take even more than two hours to complete. Where short versions of these tests are available, they still require between 30 and 75 min of testing time (Hossiep, Turck, and Hasella, 2001; Liepmann, Beauducel, Brocke, and Nettelnstroth, 2012; Wechsler, 2011). Taken together, these established and popular cognitive ability tests yield a highly accurate and often detailed characterization of an

* Corresponding author at: Johannes Gutenberg University Mainz, Institute of Psychology, D-55099 Mainz, Germany.
E-mail address: anna-lena.schubert@uni-mainz.de (A.-L. Schubert).

https://doi.org/10.1016/j.intell.2023.101804
Received 30 January 2023; Received in revised form 15 September 2023; Accepted 21 November 2023
Available online 5 December 2023
0160-2896/

individual's cognitive abilities, yet at the cost of a relatively long administration time.

These properties of well-known cognitive ability tests are exactly opposite to the requirements for tests in the context of large-scale collaborative panel studies. As a consequence of the reproducibility crisis (Marek et al., 2022; Open Science Collaboration, 2015), large-scale studies have become more important for science, as they allow researchers to tackle important questions in large and heterogeneous samples they could never collect on their own. Thus, they contribute to ensuring the robustness and generalizability of findings, not only in cognitive abilities research (Hilger, Spinath, Troche, and Schubert, 2022), but in many different fields of psychology and cognitive neuroscience (Forscher et al., 2022). In these studies, a detailed characterization of participants' cognitive abilities is often unnecessary, and the choice of administered tasks and tests usually depends less on psychometric considerations but rather on ease, duration, and feasibility of administration. As a result, cognitive phenotypes are often assessed with abbreviated or adapted measures or even related proxies, potentially compromising the reliabilities and validities of those measures (Kievit, McCormick, Fuhrmann, Deserno, and Orben, 2022). As a field, it is therefore our duty to provide psychometric tools that allow for a short, reliable, and valid measurement of cognitive abilities by developing new measures and evaluating and improving the reliability and validity of existing brief assessments of cognitive abilities.

Recent efforts resulted in the validation of cognitive ability measures included in ongoing large-scale panel studies as well as the development of novel brief assessments for use in future large-scale studies. For example, the UK Biobank is a large longitudinal cohort study with over 500,000 enrolled participants designed to study the genetic and environmental factors contributing to well-being and the development of disease (Sudlow et al., 2015). Initially comprising a brief baseline cognitive assessment, additional tests were added during the project. Although selection and development of these cognitive tests was not guided by a cohesive theoretical framework, a recent validation study showed that a measure of general cognitive abilities based on 11 UK Biobank tests correlated highly with general cognitive abilities as measured with more established tests (Fawns-Ritchie and Deary, 2020). Moreover, cognitive abilities could also be reliably and validly measured using only the five baseline tests, taking just five minutes. Another example for a recent effort to develop brief assessments of cognitive abilities is the Pathfinder test, which is a 15-min computer-based, gamified measure of cognitive abilities that assesses fluid and crystallized abilities (Malanchini et al., 2021). A validation study with young adults revealed that a Pathfinder-derived measure of general cognitive abilities was highly correlated with more established tests of cognitive abilities and predicted academic achievement. What remains an open question for future research is how suitable Pathfinder is to measure cognitive abilities in middle-aged and older adults. A validation study of a Polish translation of the test that included a larger number of middle-aged adults revealed promising results, suggesting that the test can be easily adapted to different languages and is also suited for different age groups (Muszyński et al., 2023). A last example of a newly developed brief assessment of cognitive abilities is the TestMyBrain.org Digital Neuropsychology Toolkit (TMT DNT), which is a computer-based test battery that takes about 40 min to complete and contains 11 cognitive tasks measuring the capacity of working memory and long-term memory, attention, processing speed, executive functioning, and perceptual reasoning (Singh et al., 2021). First evidence suggests that a general higher-order factor derived from the TMT DNT captures individual differences in performance across all 11 tests. However, its use for large-scale panel studies is limited by the extended duration and lack of validation with established measures of cognitive abilities. In summary, this concise overview showcases the existence of numerous promising test collections that have been recently developed. These collections can enable a comprehensive measurement of general cognitive abilities in large-scale panel studies within brief time spans.

The goal of the present study is to introduce and validate another, highly concise assessment of cognitive abilities, originally designed for use in face-to-face assessments, for large-scale panel studies: the mini-q. The mini-q was developed as a speeded figural and verbal reasoning test that can be administered in only three minutes (Baudson and Preckel, 2016). Its development was inspired by Baddeley's (1968) short verbal reasoning test, which consists of 64 letter pairs (e.g., "AB") that are each preceded by a statement about their relation (e.g., "A is not preceded by B.") that participants have to judge as true or false. The mini-q also asks test takers to assess relations but does so using figure sequences consisting of squares, triangles, and circles instead of letter pairs. These figure sequences always show two geometric figures next to each other and one further away from the other two (e.g., a triangle next to a circle and a square that is apart from the other two figures). Statements preceding figure sequences antromorphize the relations between the presented figures (e.g., "The triangle prefers the circle."), which requires participants to match the relations from the verbal statement and the figural representation.

Within the framework of the Cattell-Horn-Carroll (CHC) theory of intelligence, the mini-q falls under the broad ability domain of fluid reasoning (Gf). This domain is characterized by "the use of deliberate and controlled mental operations to solve novel problems that cannot be performed automatically" (McGrew, 2009, p. 5). These mental operations include, among others, classification, identifying relations, and transforming information (McGrew, 2009; Schneider and McGrew, 2018), all of which are also required when one attempts to solve the mini-q. Specifically, test takers have to classify the geometric figures and map their relations by identifying which two are closer to each other than to the third. They must then compare the visual representation of these relations to the verbal statement preceding the figure sequence, requiring them to either transform the visual representation into the verbal domain or vice versa. While these mental operations fall within the purview of the broad ability Gf, we can also conceive of the mini-q as a test of the speed of reasoning. Within the framework of CHC theory, the speed of reasoning is a lower-level (narrow) ability that reflects one's capacity to complete reasoning tasks in a limited timeframe, which loads not only on Gf but also on the broad ability of processing speed (Gs; McGrew, 2009; Schneider and McGrew, 2018). Taken together, the classification of the mini-q as a speeded reasoning test aligns with CHC theory.

In initial validation studies, the mini-q showed an excellent estimate of reliability with a split-half correlation of 0.98 between the average number of correctly solved odd and even items, and a one-dimensional confirmatory factor model provided excellent fit to the data (Baudson and Preckel, 2016). Moreover, the mini-q showed good convergent validity with more established tests of cognitive abilities such as the Cattell Culture Fair Test (CFT 20-R; Weiß, 2008; $r = 0.51$) and the short screening version of the Intelligence Structure Test (Liepmann et al., 2012; $r = 0.67$). The test was only moderately related to crystallized intelligence assessed with the multiple-choice vocabulary test Mehrfachwahl-Wortschatz-Intelligenztest (MWT-B; Lehrl, 1999; $r = 0.32$), but strongly related to clerical speed measured with the unpublished Ulm Speed Battery (Schmitz and Wilhelm, 2015; $r = 0.73$). Finally, the test also predicted students' self-reported high-school grades, $r = -0.28$, and thus was able to predict educational success about as well as other fluid reasoning tests (Schmidt and Hunter, 2004). Taken together, these results from initial validation studies are very promising and suggest that the test may be a useful tool for the short, reliable, and valid assessment of cognitive abilities.

However, one major limitation of these validation studies was that much of the data was collected in selective student samples consisting of students of psychology and teaching, resulting in reduced variance in cognitive abilities. Specifically, correlations between mini-q performance with school grades ($N = 402$), the CFT 20-R ($N = 126$), and the MWT-B ($N = 47$) were estimated in samples consisting of 90 to 100% students, whereas correlations with the screening version of the

Intelligence Structure Test ($N = 50$) and clerical speed ($N = 51$) were assessed in samples consisting of only 40 to 45% students. Hence, it is currently unknown whether these initial validation results generalize to samples with different educational and occupational backgrounds. Moreover, because age is a third variable strongly related to both cognitive abilities and measures of processing speed, the large age range in non-student samples (11–57 years) may have led to an overestimation of the correlation between mini-q performance and clerical speed, although previous research has shown that the relationship between processing speed and cognitive abilities in age-heterogenous samples can usually not be accounted for by age differences (Schubert, Hagemann, Löffler, and Frischkorn, 2020). Hence, one aim of the present study was to independently validate the mini-q in a new heterogeneous sample consisting of participants with different occupational and educational backgrounds.

Another open research question is which cognitive processes contribute to test performance in the mini-q. This is a particularly important question because the short test is unlikely to be a perfect measure of general cognitive abilities, although a correlation of $r = 0.67$ with the full-scale test score of the screening version of the Intelligence Structure Test (a broad cognitive abilities test) suggests that it may be a good proxy for general cognitive abilities. Instead, performance in the mini-q is likely reflecting individual differences in different lower-level cognitive abilities such as relational reasoning, processing speed, as well as verbal and figural abilities. The test authors argue that test takers have to mentally represent figural relations and compare these relations to verbal statements (Baudson and Preckel, 2016). This requires building and maintaining temporal bindings in memory, which is an elementary process of working memory that contributes to individual differences in fluid intelligence (Chuderski, 2019; Oberauer, Süβ, Wilhelm, and Wittmann, 2008; Wilhelm, Hildebrandt, and Oberauer, 2013). Moreover, because participants have only three minutes time to complete as many items as possible, performance in the mini-q also depends on test takers' processing speed. As is the case with working memory capacity, individual differences in processing speed are known to be related to individual differences in general cognitive abilities (Frischkorn, Wilhelm, and Oberauer, 2022; Schubert and Frischkorn, 2020). The importance of processing speed for mini-q performance was also demonstrated in the initial validation studies, in which a strong ($r = 0.73$) correlation between mini-q performance and clerical speed was found. Hence, in addition to domain-specific verbal and figural abilities, performance in the mini-q is likely determined by these two domain-general abilities: the ability to build and maintain relational bindings in working memory and processing speed. However, the relative contributions of the two domain-general and domain-specific processes to mini-q performance are unknown. To evaluate the generalizability of findings from studies using the mini-q, it is crucial to know the extent to which performance in the mini-q reflects individual differences in different domain-general and domain-specific cognitive abilities. Hence, one aim of the present study was to compare the relative contributions of the two domain-general and domain-specific processes to mini-q performance.

In particular, we first evaluated to what extent the relationship between mini-q performance and general cognitive abilities could be accounted for by individual differences in working memory capacity and processing speed to better understand which and to what extent domain-general abilities underlie the correlation between cognitive abilities assessed in a three-minute speeded reasoning test and in a broad test battery. While the measurement of working memory capacity is straightforward, because many different measures of working memory capacity are suitable indicators of the underlying latent construct (Oberauer et al., 2018), the measurement of processing speed is more complicated, because an individual's overt response time is affected by many different process parameters (e.g., their speed of encoding, their speed of information uptake, their speed of response preparation and execution, and their decision cautiousness). We therefore used a mathematical model of decision making—the diffusion model (Ratcliff, 1978)—to disentangle these processes in order to generate a more process-pure estimate of processing speed (conceptualized as the speed of information uptake; see Frischkorn and Schubert, 2018). Of particular interest for the present study is the drift rate parameter v, because it provides a more process-pure measure of processing speed than the mean or median of an individual's reaction time distribution and is therefore usually more strongly related to individual differences in cognitive abilities than average reaction times (Frischkorn et al., 2022; Schubert and Frischkorn, 2020). Hence, we used participants' drift rates instead of their mean reaction times as a measure of processing speed to evaluate to what degree individual differences in processing speed and working memory capacity accounted for the relationship between mini-q performance and general cognitive abilities.

In addition, we supplemented this process-based analysis by investigating how different operation-related components required for completing a classical intelligence test battery, namely processing capacity, memory, creativity, and processing speed, contributed to test performance. If the mini-q is a valid measure of speeded reasoning, we would anticipate that both processing capacity (i.e., reasoning) and processing speed emerge as the strongest predictors of mini-q performance.

Moreover, we evaluated to what degree the two domain-specific abilities required by the mini-q test material, verbal and figural abilities, contributed to test performance to derive recommendations for using the test as a brief assessment of cognitive abilities in large-scale panel studies with participants from heterogeneous backgrounds.

2. The present study

The present study pursued five aims. First, we independently validated the mini-q in a heterogeneous community sample. For this purpose, participants from different educational and occupational backgrounds completed a broad cognitive test battery consisting of the short version of the BIS (Jäger et al., 1997) to measure their general and domain-specific cognitive abilities, four different working memory tasks to measure their working memory capacity, and three different simple two-choice reaction time tasks to measure their processing speed.

Second, we compared the relative contributions of two domain-general abilities—working memory capacity and processing speed—to mini-q performance. For this purpose, we formally tested whether individual differences in working memory capacity and/or processing speed mediated the relationship between mini-q performance and general cognitive abilities.

Third, we also tested which of the operation-related components of the BIS were most strongly related to mini-q performance, expecting to find that processing capacity and processing speed were more strongly related to mini-q performance than memory or creativity.

Fourth, we compared the relative contributions of two domain-specific abilities—verbal and figural abilities—to mini-q performance. For this purpose, we used the verbal and figural subscales of the BIS to predict mini-q performance. To assess not only the convergent but also the discriminant validity of domain-specific abilities, we also explored whether the numerical subscale of the BIS predicted mini-q performance.

Lastly, we developed and validated a computer-based version of the mini-q, compared the pen-and-paper and computer-based versions regarding measurement invariance, and provided norm tables to facilitate the test's adaptation for use in large-scale panel studies. This will enable an efficient and cost-effective computer-based in-person or online administration of the test.

Taken together, the overall goal of the present study was to validate the mini-q to evaluate its usefulness for a short, reliable, and valid assessment of cognitive abilities in large-scale panel studies. In Study 1, we further validated the existing pen-and-paper based version of the mini-q in a heterogeneous community sample, and in Study 2, we developed and validated a computer-based version of the test in a heterogeneous online sample.

3. Study 1

3.1. Methods

3.1.1. Participants

140 participants (92 females, 47 males, one diverse) between 18 and 60 years ($M = 31.93$, $SD = 14.12$) from different educational and occupational backgrounds took part in a psychometric testing session in groups of up to four participants as part of a larger multi-session study on individual differences in cognition (Löffler, Frischkorn, Hagemann, Sadus, and Schubert, 2022; Schubert, Löffler, and Hagemann, 2022; Schubert, Löffler, Hagemann, and Sadus, 2022). We recruited this sample via advertisements in local newspapers, flyers, and the departmental participant pool. Four participants went to secondary school as their highest level of education, six participants had a university entrance qualification for applied sciences, 75 participants had a general university entrance qualification, 53 participants had a university degree, and 2 participants had a PhD (see the left part of Fig. 1).

80 participants were university students, 37 were employed, 10 were self-employed, four were high-school students, three were homemakers, three were retired, two were volunteers, and one participant was unemployed (see the right part of Fig. 1). 136 participants were native German speakers, of which one person was bilingual (German and Russian). In addition to these participants, there was one individual each who could speak other languages such as French, Greek, Russian, and Hungarian as their mother tongue. 128 participants were right-handed and 12 left-handed, and 65 participants had normal and 75 participants corrected-to-normal vision.

The sample size of 140 participants yielded a power of 99% to test the hypothesis of close fit as suggested by Browne and Cudeck (1992) for the structural equation model with the highest degrees of freedom ($df = 71$), an alpha error of $\alpha = 0.05$, and a power of $1 - \beta = 0.80$, and a power of 50% for the structural equation model with the lowest degrees of freedom ($df = 12$).

3.1.2. Material

mini-q. The mini-q (Baudson and Preckel, 2016) was administered as a pen-and-paper test in groups of up to four participants. After reading the written instructions out loud, the experimenter went through two already solved sample items and then asked participants to solve four additional sample items on their own. Participants were then given the opportunity to ask questions about test instructions. Once all participants had confirmed that they had no further questions, they were given three minutes time to complete as many of the 64 items as possible.

Incorrectly solved items, skipped items, and items not completed in time were scored as errors. On average, participants correctly solved 39 items ($SD = 10.61$, range 13–64 items).

Based on the assumptions that the mini-q is a unidimensional test of speeded reasoning and that test items are essentially congeneric indicators of the underlying construct (Baudson and Preckel, 2016), we built three parcels for structural equation modeling, for which we averaged performance across every third item, starting with items one, two, and three. These parcels were transformed into z-scores for further analyses.

Cognitive abilities. Cognitive abilities were assessed with the short version of the BIS (Jäger et al., 1997). The BIS is based on the bimodal Berlin intelligence structure model, which divides general intelligence into four operation-related (processing speed, memory, creativity, processing capacity) and three content-related (verbal, numerical, figural) components (Jäger, 1982). Processing speed measures the ability to execute simple tasks quickly and accurately, while memory pertains to the ability to recollect lists and item configurations only a few minutes after learning them (episodic memory). Creativity refers to the capacity to generate a multitude of novel ideas with ease, and processing capacity incorporates inductive, deductive, and spatial reasoning.

The short version of the BIS consists of 15 subtests, each reflecting a combination of one content-related component (verbal, numerical, or figural) with one operation-related component (processing capacity, processing speed, memory, or creativity). Each content-related component is tested with five tasks across all four operation-related components, allowing specific component scores (e.g., verbal abilities) to be calculated by combining related subtests across all operation-related

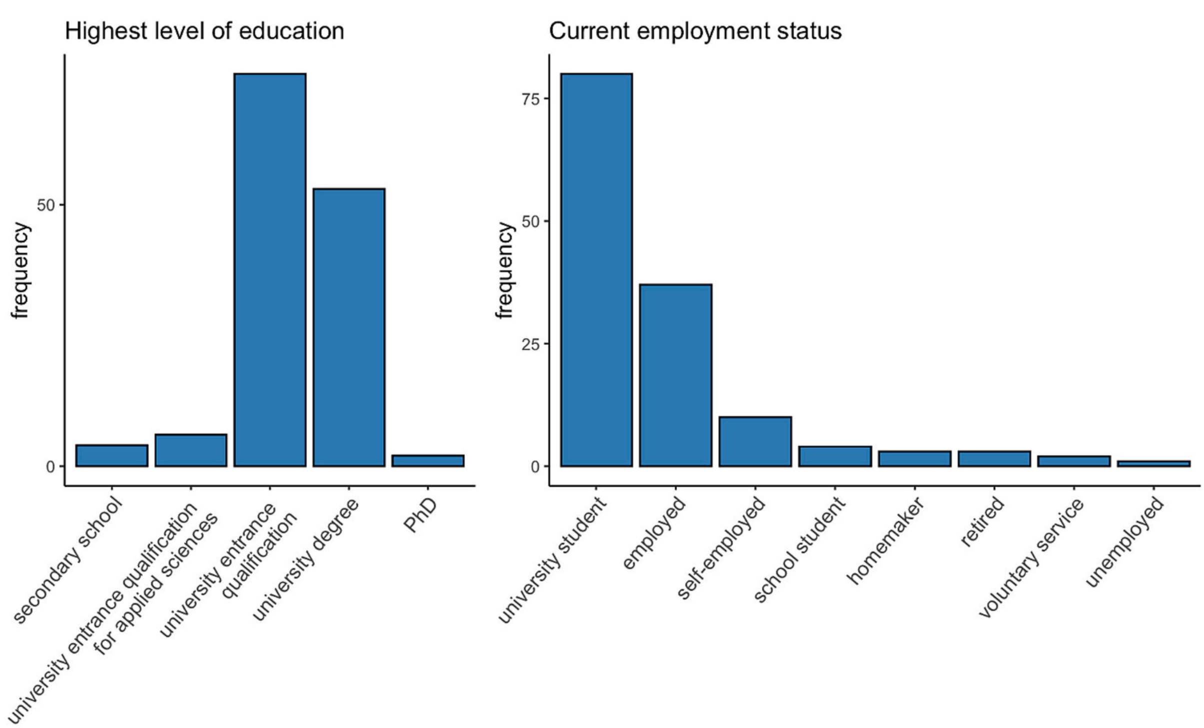

Fig. 1. Distribution of the highest level of education and occupation of participants in Study 1.

components. As a result, the scores for content- and operation-related abilities are not independent, and a measure of general intelligence is either calculated across operation- or content-related scores. Group administration allows the short version of the BIS to be completed in 45 min, making it faster and more cost-effective than, for example, the WAIS.

Participants had a mean IQ of $M = 97.09$ ($SD = 10.61$), which is likely an underestimation of their true intelligence as the BIS norming sample only contains data from senior high school students between 16 and 19 years.

Working memory capacity. Participants completed a test battery consisting of five working memory tasks. They completed the memory updating task, the operation span task, the sentence span task, and the spatial short-term memory task from the working memory test battery by Lewandowsky, Oberauer, Yang, and Ecker (2010), and all but five participants completed the location-letter binding task from the working memory test battery by Wilhelm et al. (2013). Due to a programming error, data from the spatial short-term memory task could not be used.

For each task, we calculated the proportion of correctly solved items separately for each set size and then averaged across set sizes to calculate a measure of working memory capacity. Four participants' operation span scores, three participants' sentence span scores, and one participant's binding score were removed from further analyses because they exceeded ±3 *SDs* of the sample mean.

Processing speed. Participants completed three speeded two-alternative forced choice tasks with 100 to 120 experimental trials each. In the Sternberg memory scanning (SMS) task (Sternberg, 1969), participants saw a set of five sequentially presented digits followed by a memory probe. They had to decide if the digit shown as the memory probe had been included in the previously presented memory set. In the choice reaction time (CRT) task (Schubert, Hagemann, Voss, Schankin, and Bergmann, 2015), a cross appeared in one of two squares and participants had to indicate its location by pressing the corresponding left or right key. In the Posner letter matching (PLM) task (Posner and Mitchell, 1967), participants saw a pair of letters and had to decide whether their names were identical. Participants always responded by pressing the "D" or "L" key on a standard keyboard and were instructed to respond as quickly and accurately as possible.

Two participants' behavioral data from the SMS task and three participants' behavioral data from the CRT task were removed from further analyses because they exceeded ±3 *SDs* of the sample mean. We used participants' individual reaction time distributions to estimate the drift rate parameter of the diffusion model as a measure of the speed of information uptake (see below). Two participants' parameter estimates from the SMS task, one participant's parameter estimates from the CRT task, and one participant's parameter estimates from the PLM task were removed from further analyses because they exceeded ±3 *SDs* of the sample mean.

3.1.3. Procedure

Data were collected as part of a three-session study on individual differences in cognition. Participants signed an informed consent form before taking part in the study. They received 75 € and were offered feedback about their intelligence test results as reward for their participation. The study was approved by the ethics committee of the Faculty of Behavioural and Cultural Studies at Heidelberg University. All procedures were conducted in accordance with the declaration of Helsinki.

In the first and second session, participants completed six different experimental tasks each while they sat in a dimly lit cabin and their EEG was recorded. In the third session (the psychometric testing session), participants completed the BIS, the working memory test battery, the mini-q, a highly speeded cognitive ability test (adapted from Chuderski, 2019; data not reported here), a pretzel task (unpublished; data not reported here), and the Brief Mind Wandering Three-Factor Scale (Schubert et al., 2023; data not reported here) in groups of up to four participants. Data from all tasks and tests reported in this manuscript were collected at the third session, except for the processing speed tasks and the demographic questionnaire, which participants completed during the first (SMS task, demographic questionnaire) and second (CRT task, PLM task) session.[1] The three sessions were about four months apart but had to be rescheduled for some participants due to a two-month break in data collection at the outbreak of COVID-19. One participant decided to complete the psychometric testing session via video chat due to the pandemic, which was only feasible for the pen-and-pencil tests but not the computer-based working memory tasks.

3.1.4. Data analysis

Preprocessing. We conducted an intraindividual outlier analysis of participants' reaction time distributions separately for each of the three processing speed tasks and discarded any trials with RTs faster than 150 ms or with logarithmized RTs exceeding ±3 SD of each participant's mean RT. In addition, we conducted an interindividual outlier analysis for all variables as reported above by discarding any values exceeding ±3 SD of the mean.

Diffusion modeling. The diffusion model is a cognitive measurement model which assumes that when deciding between two alternatives, individuals continuously accumulate evidence until one of two decision thresholds is reached. This information accumulation process is described by a random-walk process consisting of a constant systematic component (the drift), and normally distributed random noise (see Fig. 2).

The basic diffusion model uses a participant's response time distribution of correct and incorrect responses to estimate four parameters: (1) the drift rate parameter v, which describes the strength and direction of the information accumulation process; (2) the threshold separation parameter a, which describes the distance between decision thresholds; (3) the starting point z, which reflects the starting point of evidence accumulation and can thus indicate biases in decision making; and (4) the non-decision time parameter t_0, which reflects the speed of all non-decisional processes (e.g., encoding, response execution, etc.).

Diffusion models were fitted with *fast-dm-30* (Voss and Voss, 2007) using the Kolmogorov-Smirnov statistic for parameter estimation. For each participant and each of the three processing speed tasks, we estimated the drift rate (v), the boundary separation (a), the non-decision time (t_0), and the trial-to-trial variability of the non-decision time parameters (s_{t0}). For descriptive statistics of parameter estimates, see Table 1. To estimate the reliabilities of model parameters, we fitted the model separately to data from odd and even trials of each task and calculated Spearman-Brown corrected correlations between the model parameters estimated from odd-even split data as an estimate of reliability. We evaluated model fit through correlations between the predicted and observed values for the 25th, 50th, and 75th quantile of the reaction time distribution. The models were able to recover the shape of participants' reaction time distributions very well, as indicated by correlations between predicted and observed values of quantiles of the reaction time distribution ranging from $r = 0.88$ to $r = 0.98$. Due to the low error rates, accuracy rates were less successfully predicted by the model, in particular accuracy rates of the CRT task, $r = 0.05$, while accuracy rates of the SMS task, $r = 0.80$, and the PLM task, $r = 0.68$, were recovered much better.

Structural equation modeling. We z-standardized all variables before entering them into structural equation models. We fitted

[1] In theory, the completion of processing speed tasks during a different session than the other tests could result in an underestimation of the correlation between processing speed and the other measures due to situational factors, such as variations in fatigue or time of day. However, it is unlikely that situational factors affected the results in the present study, because individual differences in the BIS, working memory capacity, and drift rates have previously been shown to not be affected by situational factors (Danner et al., 2011; Rummel et al., 2022; Schubert et al., 2016).

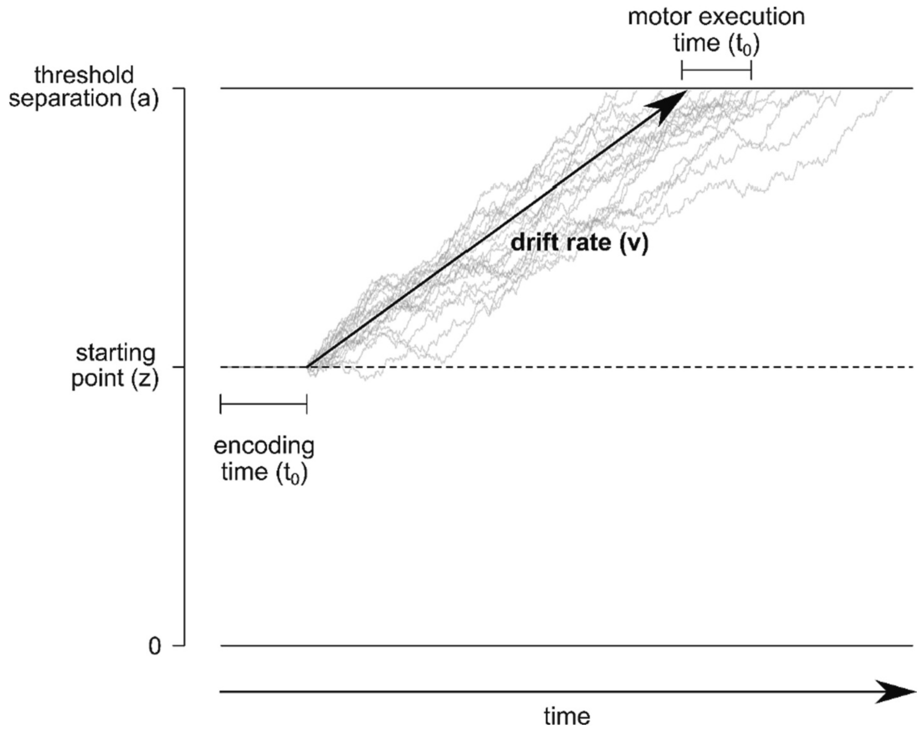

Fig. 2. Illustration of the diffusion model.

Table 1
Descriptive statistics and reliability estimates for all measures in Study 1.

Construct		*N*	*M*	*SD*	Range	Reliability
Speeded reasoning	mini-q	140	39	10.61	13–64	.99[a]/.96[b]
Cognitive abilities	BIS verbal	140	104.86	5.89	88.20–121.60	.50[b]
	BIS numerical	139	98.53	7.04	81.40–112.20	.74[b]
	BIS figural	140	97.11	6.76	77.80–112.60	.65[b]
	BIS PC	140	101.61	7.12	85.00–118.67	.75[b]
	BIS PS	140	101.14	7.15	82.00–119.00	.49[b]
	BIS M	140	98.59	7.16	83.67–116.00	.57[b]
	BIS C	140	98.15	6.97	80.33–114.00	.45[b]
Working memory capacity	Memory updating	139	0.63	0.20	0.14–0.97	.88[b]
	Operation span	135	0.78	0.13	0.42–1.00	.89[b]
	Sentence span	136	0.84	0.11	0.44–1.00	.87[b]
	Binding	133	0.86	0.11	0.52–1.00	.82[b]
Processing speed	SMS task – RT	136	0.93	0.23	0.54–1.62	.99[a]
	SMS task – v	136	2.36	0.69	1.49–4.96	.42[a]
	SMS task – a	136	1.53	0.42	0.73–2.43	.63[a]
	SMS task – t_0	136	0.61	0.17	0.32–0.84	.95[a]
	SMS task – s_{t0}	136	0.24	0.16	0.00–0.51	.57[a]
	CRT task – RT	133	0.38	0.05	0.30–0.57	.98[a]
	CRT task – v	133	6.24	1.60	2.31–10.94	.70[a]
	CRT task – a	133	0.95	0.24	0.44–1.73	.67[a]
	CRT task – t_0	133	0.30	0.04	0.18–0.49	.90[a]
	CRT task – s_{t0}	133	0.08	0.05	0.00–0.25	.65[a]
	PLM task – RT	136	0.71	0.13	0.50–1.06	.99[a]
	PLM task – v	136	3.27	0.72	0.89–4.41	.50[a]
	PLM task – a	136	1.38	0.34	0.60–2.95	.69[a]
	PLM task – t_0	136	0.49	0.07	0.33–1.23	.87[a]
	PLM task – s_{t0}	136	0.19	0.09	0.00–0.91	.46[a]

Note. The sample size *N* shows the number of participants for each measure after excluding interindividual outliers. BIS = Berlin Intelligence Structure test; PC = processing capacity; PS = processing speed; M = memory; C = creativity; SMS = Sternberg memory scanning; CRT = choice reaction time; PLM = Posner letter matching; RT = reaction time; v = drift rate; a = boundary separation; t_0 = non-decision time; s_{t0} = trial-to-trial variability of the non-decision time parameter.
[a] Reliability estimates are based on Spearman Brown corrected odd-even split correlations.
[b] Reliability estimates are based on Cronbach's α.

covariance-based structural equation models with the *R* package *lavaan* (Rosseel, 2012) using the full information maximum likelihood algorithm to account for missing data. We evaluated goodness of fit based on the comparative fit index (CFI; Bentler, 1990) and the root mean square error of approximation (RMSEA; Browne and Cudeck, 1992). Following the recommendations by Browne and Cudeck (1992) and Hu and Bentler (1999), we considered CFI values >0.90 and RMSEA values <0.08 to indicate acceptable model fit and CFI values >0.95 and RMSEA values

<0.06 to indicate good model fit. The statistical significance of model parameters was assessed with the two-sided critical ratio test.

Data and code availability. The data and code supporting the findings of the study are available in the Open Science Framework repository at https://osf.io/bpm5e/. Raw data and materials are available in the Open Science Framework repository at https://osf.io/4pvz3/.

3.2. Results

Descriptive statistics and reliability estimates for all measures are shown in Table 1 (see Fig. 3 for the distribution of the mini-q performance and Figs. S1–S3 in the supplementary material for the other distributional properties).

In the present study, we pursued four goals to validate the mini-q (see Table 2 for zero-order correlations of all variables): (1) We assessed the association between mini-q performance and general intelligence, (2) we evaluated to what degree individual differences in processing speed and working memory capacity accounted for the relationship between mini-q performance and general cognitive abilities, (3) we compared the relative contributions of operation-related abilities—processing capacity, processing speed, memory, and creativity—to mini-q performance, and (4) we compared the relative contributions of domain-specific abilities—verbal, figural, and numerical abilities—to mini-q performance.

3.2.1. How strongly is mini-q performance related to general cognitive abilities?

To address the first research question, we estimated the latent correlation between mini-q performance and general cognitive abilities using a structural equation model. The model provided an excellent account of the data, $\chi^2(14) = 12.64$, $p = .555$, CFI = 1.00, RMSEA = 0.00 (90% CI = [0.00; 0.08]). General cognitive abilities as assessed with the short version of the BIS and mini-q performance were highly correlated, $r = 0.57$ (95% CI = [0.43; 0.72]), $p < .001$, indicating that the mini-q is suitable for the brief assessment of cognitive abilities in heterogeneous samples.

Because our sample had a relatively wide age range from 18 to 60 years, we also tested whether age differences might have led to an overestimation of this correlation. If the relationship between mini-q performance and cognitive abilities became smaller after controlling for age differences, this would indicate that their relationship was at least in part due to age differences affecting performance in both tests. The resulting model, in which the correlation between mini-q performance and general cognitive abilities was controlled for age differences by regressing both variables on participants' age, also provided an excellent account of the data, $\chi^2(19) = 15.91$, $p = .664$, CFI = 1.00, RMSEA = 0.00 (90% CI = [0.00; 0.06]). Older participants performed worse than younger adults in the mini-q, $\beta = -0.40$ (95% CI = [−0.54; −0.26]), $p < .001$, and in the BIS, $\beta = -0.40$ (95% CI = [−0.57; −0.22]), $p < .001$, but the correlation between performances in the two tests was only slightly diminished after controlling for age, $r = 0.50$ (95% CI = [0.34; 0.66]), $p < .001$. Moreover, the point estimate of the correlation between mini-q performance and general cognitive abilities not corrected for age differences was still contained in the 95% confidence interval of the age-controlled correlation.

3.2.2. To what degree do processing speed and working memory capacity account for the relationship between mini-q performance and general cognitive abilities?

To address the second research question, we estimated a mediation model to formally test to what degree two domain-general abilities – working memory capacity and processing speed – accounted for the relationship between mini-q performance and general cognitive abilities (see Fig. 4).

The mediation model showed a good fit to the data, $\chi^2(71) = 98.64$, $p = .017$, CFI = 0.98, RMSEA = 0.05 (90% CI = [0.02; 0.08]). Working memory capacity, $\beta_{indirect} = 0.31$ (95% CI = [0.17; 0.45]), $p < .001$, but not processing speed, $\beta_{indirect} = 0.01$ (95% CI = [−0.15; 0.17]), $p = .899$, mediated the relationship between performance in the two tests, accounting for a total of 54% ($\beta_{ab}/\beta_{total} = 0.31/0.57 = 0.54$) of the covariance between mini-q performance and general cognitive abilities. We only found evidence for a partial mediation, as the direct effect, despite being greatly reduced ($\Delta = 0.32$), remained significant after controlling for the two domain-global abilities, $\beta = 0.25$ (95% CI = [0.03; 0.47]), $p = .033$.[2]

To compare how using drift rates instead of mean reaction times as a measure of processing speed affected the results, we estimated the same mediation model using mean response times instead of drift rates and working memory capacity as mediators.[3] This model also showed a good fit to the data, $\chi^2(71) = 104.90$, $p = .006$, CFI = 0.98, RMSEA = 0.06 (90% CI = [0.03; 0.08]). Again, only working memory capacity, $\beta_{indirect} = 0.30$ (95% CI = [0.16; 0.43]), $p < .001$, but not mean response times, $\beta_{indirect} = 0.10$ (95% CI = [−0.01; 0.20]), $p = .089$, mediated the relationship between mini-q and BIS performance. In comparison to the previous mediation model, however, we observed a full mediation, as the direct effect was no longer significant after controlling for working memory capacity and reaction times, $\beta = 0.18$ (95% CI = [−0.02; 0.38]), $p = .091$. The observed complete mediation in the model may be attributed to the fact that response time measures reflect individual differences in different aspects of information processing (e.g., not only the speed of information processing, but also the speed of encoding, the speed of response preparation and execution, and decision cautiousness), leading to a stronger effect on the relationship between the two cognitive ability measures. Moreover, response times were more strongly related to age differences than drift rates ($r = 0.46$ vs. $r = -0.32$), which may have accounted for part of the relationship between mini-q and BIS performance.

3.2.3. How do different operation-related components contribute to mini-q performance?

To address the third research question, we regressed mini-q performance simultaneously on the processing capacity, processing speed, memory, and creativity scores of the short version of the BIS (see Fig. 5). This multiple regression model provided an excellent account of the data, $\chi^2(15) = 3.67$, $p = .999$, CFI = 1.00, RMSEA = 0.00 (90% CI = [0.00; 0.00]). As predicted, both processing capacity, $\beta = 0.23$ (95% CI = [0.06 0.40]), $p < .001$, and processing speed, $\beta = 0.34$ (95% CI = [0.17; 0.51]), $p < .001$, but not memory, $\beta = 0.10$ (95% CI = [−0.07; 0.26]), $p = .255$, nor creativity, $\beta = -0.08$ (95% CI = [−0.24; 0.07]), $p = .290$, contributed to mini-q performance.

We then modified the model by removing the regression of mini-q performance on memory and creativity to test whether the contributions of these two operation-related components to mini-q performance were negligible. We found that removing the two regression paths did not impair model fit, $\Delta\chi^2(2) = 2.61$, $p = .271$, which indicates that memory and creativity do not make any specific contribution to mini-q performance. We then further modified the model by constraining the regressions weights of processing capacity and speed to be equal, which did not impair model fit, $\Delta\chi^2(1) = 0.51$, $p = .473$, thus indicating that the regression weights of processing capacity and speed did not differ significantly from each other.

[2] Note that results from this mediation analysis do not imply that working memory capacity contributes *causally* to the relationship between mini-q performance and general cognitive abilities, as many potential confounders may account for part of relationship between the three constructs (MacKinnon and Pirlott, 2015; Rohrer, Hünermund, Arslan, and Elson, 2022). We selected this model purely as a statistical technique to break down distinct and common sources of variance, not for the purpose of making causal inferences.

[3] We would like to thank an anonymous reviewer for this suggestion.

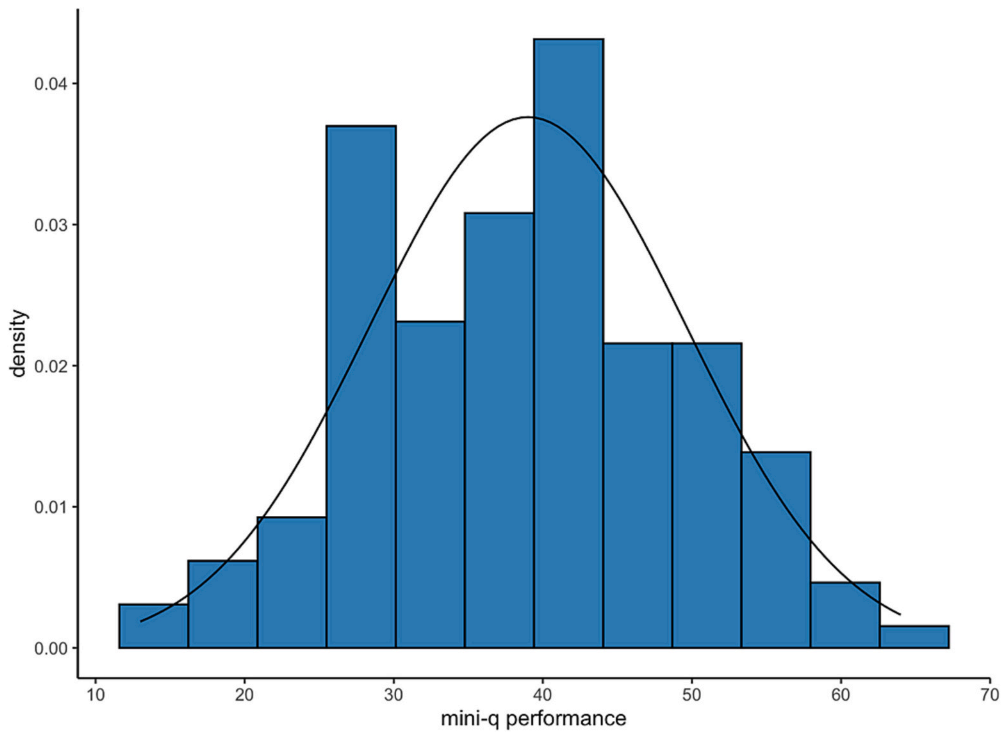

Fig. 3. Distribution of performance in the mini-q in Study 1.

3.2.4. How do verbal, figural, and numerical abilities contribute to mini-q performance?

To address the fourth research question, we regressed mini-q performance simultaneously on the verbal, figural, and numerical scores of the short version of the BIS (see Fig. 6). This multiple regression model provided an excellent account of the data, $\chi^2(12) = 9.53$, $p = .657$, CFI = 1.00, RMSEA = 0.00 (90% CI = [0.00; 0.07]). As predicted, both verbal, $\beta = 0.33$ (95% CI = [0.16; 0.50]), $p < .001$, and figural abilities, $\beta = 0.26$ (95% CI = [0.09; 0.42]), $p < .001$, contributed to mini-q performance. In contrast, numerical abilities did not contribute to mini-q performance, $\beta = 0.01$ (95% CI = [−0.16; 0.18]), $p = .909$.

We modified the model by removing the regression of mini-q performance on numerical abilities to test whether the contribution of numerical abilities to mini-q performance was negligible. We found that removing this regression path did not impair model fit, $\Delta\chi^2(1) = 0.01$, $p = .909$, which indicates that numerical abilities do not make any specific contribution to mini-q performance. We then further modified the model by constraining the regression weights of verbal and figural abilities to be equal, which did not impair model fit, $\Delta\chi^2(1) = 0.26$, $p = .880$, thus indicating that the regression weights of verbal and numerical abilities did not differ significantly from each other.

3.3. Discussion

The goal of Study 1 was to evaluate the suitability of the mini-q, a three-minute speeded reasoning test, for a brief assessment of general cognitive abilities in large-scale panel studies. The present study went beyond the initial validation studies, which collected most of their data in selective student samples, by including participants with a broad age range from different educational and occupational backgrounds.

Overall, our results suggest that the mini-q is a suitable test for the brief assessment of cognitive abilities, as it was substantially related ($r = 0.57$; age-corrected $r = 0.50$) to participants' general cognitive abilities as measured with a broad and validated test battery, the Berlin Intelligence Structure Test (Jäger et al., 1997). Although both tests are speeded, the correlation between mini-q performance and cognitive abilities remained virtually unchanged after controlling for age differences. This indicates that the substantial relationship between test performances was not severely overestimated due to age-related cognitive slowing affecting performance in either test. In consequence, we argue that the mini-q is not only suitable for a brief assessment of cognitive abilities in age-homogeneous, but also in age-heterogeneous samples, at least in an age range from 18 to 60 years. Future studies could evaluate its suitability for the assessment of cognitive abilities in adolescent and elderly samples. Our results also endorse the assumption that the mini-q is a test of speeded reasoning. This is supported by participants' performance being linked to their processing capacity and speed as indicated by the BIS. CHC theory classifies speeded reasoning as a lower-level, specific ability, falling under the broader aspects of fluid intelligence and processing speed (McGrew, 2009; Schneider and McGrew, 2018). Therefore, we recommend the mini-q as a reliable and efficient way of measuring individual differences in fluid intelligence and processing speed.

The substantial correlation between mini-q performance and general cognitive abilities was largely driven by the extent to which both measures reflected individual differences in working memory capacity. After accounting for participants' working memory capacity, their processing speed made no additional contribution to the relationship between mini-q performance and general cognitive abilities. In other words, the mini-q mostly measured general cognitive abilities to the degree that it measured working memory capacity, even though it is important to note that a substantial part of the relationship (about 46%) between the two measures could not be accounted for by either domain-general processing ability. The finding that working memory capacity accounted for a large part of the relationship between mini-q performance and general cognitive abilities is consistent with previous research that emphasized the role of establishing and maintaining relational bindings in working memory for fluid reasoning (Oberauer et al., 2008; Wilhelm et al., 2013). When solving a mini-q item, test takers build a mental representation of the object relations described in the verbal statement and maintain this mental representation in working memory to compare it to the figural display of object relations displayed in the item. This building

Table 2
Zero-order correlations between all variables in Study 1.

		mini-q	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
2	BIS_{IQ}	**0.46**	–																				
3	age	**-0.40**	**-0.33**	–																			
4	mini-q_1	**0.99**	**0.45**	**-0.40**	–																		
5	mini-q_2	**0.99**	**0.47**	**-0.40**	**0.97**	–																	
6	mini-q_3	**0.99**	**0.46**	**-0.39**	**0.97**	**0.97**	–																
7	BIS_V	**0.45**	**0.80**	**-0.25**	**0.42**	**0.46**	**0.46**	–															
8	BIS_F	**0.41**	**0.79**	**-0.35**	**0.42**	**0.41**	**0.41**	**0.48**	–														
9	BIS_N	**0.26**	**0.80**	**-0.24**	**0.26**	**0.27**	**0.25**	**0.47**	**0.38**	–													
10	BIS_{PC}	**0.41**	**0.88**	**-0.34**	**0.40**	**0.41**	**0.41**	**0.71**	**0.66**	**0.73**	–												
11	BIS_{PS}	**0.46**	**0.74**	**-0.36**	**0.45**	**0.47**	**0.45**	**0.61**	**0.57**	**0.59**	**0.50**	–											
12	BIS_M	**0.33**	**0.65**	**-0.23**	**0.32**	**0.33**	**0.32**	**0.50**	**0.55**	**0.48**	**0.43**	**0.43**	–										
13	BIS_C	0.12	**0.57**	0.06	0.11	0.13	0.13	**0.47**	**0.49**	**0.40**	**0.35**	**0.34**	0.13	–									
14	MU	**0.47**	**0.57**	**-0.42**	**0.45**	**0.48**	**0.45**	**0.40**	**0.42**	**0.50**	**0.60**	**0.42**	**0.34**	0.13	–								
15	Binding	**0.31**	**0.55**	**-0.61**	**0.29**	**0.32**	**0.29**	**0.37**	**0.53**	**0.39**	**0.53**	**0.44**	**0.34**	**0.19**	**0.70**	–							
16	OSPAN	**0.38**	**0.36**	**-0.37**	**0.37**	**0.38**	**0.37**	**0.28**	**0.29**	**0.23**	**0.38**	**0.20**	**0.23**	0.09	**0.58**	**0.50**	–						
17	SSPAN	**0.40**	**0.40**	**-0.18**	**0.39**	**0.42**	**0.39**	**0.33**	**0.25**	**0.33**	**0.41**	**0.22**	**0.27**	0.13	**0.61**	**0.35**	**0.66**	–					
18	v_{SMS}	**0.39**	**0.28**	−0.16	**0.38**	**0.39**	**0.39**	**0.22**	**0.27**	0.16	**0.27**	**0.25**	0.16	0.11	**0.43**	**0.24**	**0.29**	**0.24**	–				
19	v_{CRT}	**0.25**	**0.19**	**-0.22**	**0.25**	**0.24**	**0.27**	0.05	**0.26**	0.16	0.16	**0.22**	0.12	0.07	**0.24**	**0.19**	0.10	0.09	**0.38**	–			
20	v_{PLM}	**0.31**	**0.26**	**-0.20**	**0.29**	**0.32**	**0.30**	**0.22**	**0.18**	**0.24**	**0.27**	**0.21**	0.14	0.05	**0.23**	**0.23**	0.03	0.05	**0.34**	**0.36**	–		
21	RT_{SMS}	**-0.45**	**-0.30**	**0.43**	**-0.44**	**-0.44**	**-0.46**	**-0.29**	**-0.30**	−0.14	**-0.29**	**-0.31**	−0.10	−0.16	**-0.31**	**-0.36**	**-0.25**	−0.09	**-0.54**	**-0.37**	**-0.32**	–	
22	RT_{CRT}	**-0.35**	**-0.29**	**0.40**	**-0.34**	**-0.35**	**-0.36**	−0.15	**-0.38**	**-0.18**	**-0.21**	**-0.40**	−0.13	−0.12	**-0.31**	**-0.32**	−0.16	−0.16	**-0.35**	**-0.68**	**-0.33**	**0.43**	–
23	RT_{PLM}	**-0.44**	**-0.40**	**0.35**	**-0.41**	**-0.44**	**-0.45**	**-0.36**	**-0.35**	**-0.27**	**-0.36**	**-0.38**	**-0.19**	**-0.21**	**-0.37**	**-0.31**	−0.16	−0.16	**-0.42**	**-0.55**	**-0.56**	**0.57**	**0.67**

Note. BIS = Berlin Intelligence Structure test; mini-q = sum score of the mini-q after three minutes; BIS_{IQ} = IQ score of the BIS; mini-q_1 = first parcel of the mini-q; mini-q_2 = second parcel of the mini-q; mini-q_3 = third parcel of the mini-q; BIS_V = verbal score of the BIS; BIS_F = figural score of the BIS; BIS_N = numerical score of the BIS; BIS_{PC} = processing capacity score of the BIS; BIS_{PS} = processing speed score of the BIS; BIS_M = memory score of the BIS; BIS_C = creativity score of the BIS; MU = memory updating; OSPAN = operation span; SSPAN = sentence span; v_{SMS} = drift rate in the Sternberg memory scanning task; v_{CRT} = drift rate in the choice reaction time task; v_{PLM} = drift rate in the Posner letter matching task; RT_{SMS} = mean reaction time in the Sternberg memory scanning task; RT_{CRT} = mean reaction time in the choice reaction time task; RT_{PLM} = mean reaction time in the Posner letter matching task; Significant correlations ($p \leq .05$) are presented in bold.

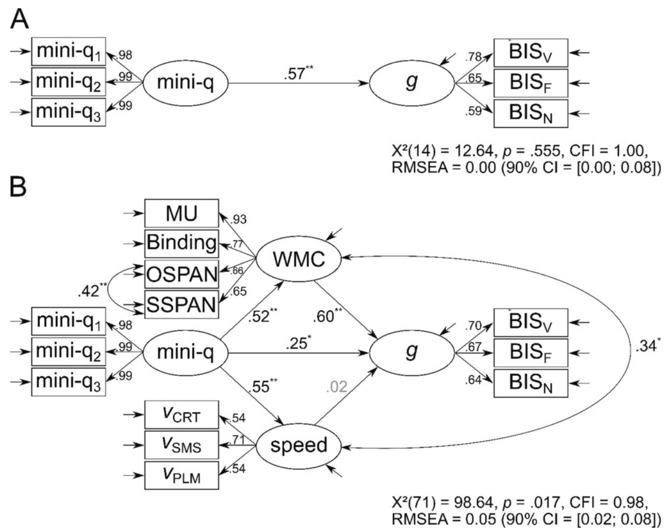

Fig. 4. Regression model (A) and mediation model (B) describing the relationship between mini-q performance and general cognitive abilities.
Note. mini-q_1 = first parcel of the mini-q; mini-q_2 = second parcel of the mini-q; mini-q_3 = third parcel of the mini-q; *g* = general cognitive abilities; BIS_V = verbal score of the Berlin Intelligence Structure test; BIS_F = figural score of the Berlin Intelligence Structure test; BIS_N = numerical score of the Berlin Intelligence Structure test; MU = memory updating; OSPAN = operation span; SSPAN = sentence span; v_{CRT} = drift rate in the choice reaction time task; v_{SMS} = drift rate in the Sternberg memory scanning task; v_{PLM} = drift rate in the Posner letter matching task. $^*p < .05$; $^{**}p < .01$.

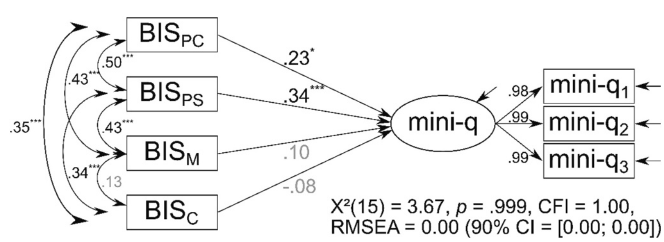

Fig. 5. Multiple regression model of mini-q performance on operation-related components.
Note. mini-q_1 = first parcel of the mini-q; mini-q_2 = second parcel of the mini-q; mini-q_3 = third parcel of the mini-q; BIS_{PC} = processing capacity score of the Berlin Intelligence Structure test; BIS_{PS} = processing speed score of the Berlin Intelligence Structure test; BIS_M = memory score of the Berlin Intelligence Structure test; BIS_C = creativity score of the Berlin Intelligence Structure test. $^*p < .05$; $^{***}p < .001$.

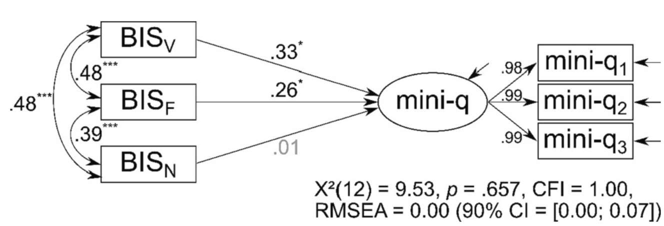

Fig. 6. Multiple regression model of mini-q performance on domain-specific abilities.
Note. mini-q_1 = first parcel of the mini-q; mini-q_2 = second parcel of the mini-q; mini-q_3 = third parcel of the mini-q; BIS_V = verbal score of the Berlin Intelligence Structure test; BIS_F = figural score of the Berlin Intelligence Structure test; BIS_N = numerical score of the Berlin Intelligence Structure test. $^*p < .05$.

of mental representations is particularly challenging for items containing passive voice and/or the negation of a relation (e.g., "The circle is not rejected by the triangle."), because these statements additionally require the rotation of mentally represented object relations. Hence, the speed with which test takers complete each item is not only and perhaps even not primarily determined by their processing speed, but rather by the efficiency with which they build, maintain, and compare sets of relational bindings in working memory. Because the ability to form even very simple relational bindings seems to be as strongly related to fluid intelligence as performance in much more complex transitive reasoning tasks (Chuderski, 2019) and because fluid intelligence is strongly related to general cognitive abilities (Carroll, 1993), the ability to build arbitrary relational bindings is a promising candidate for an elementary domain-general ability underlying the relationship between mini-q performance and general cognitive abilities.

Does this indicate that the mini-q should be considered a test of working memory capacity instead of fluid intelligence? This is a question that cannot be definitively answered based on correlational findings alone. We found overlaps between mini-q performance, general cognitive abilities, and working memory capacity, but no evidence of the mini-q being more strongly related to working memory capacity than to general cognitive abilities (see Table 2). The interpretation of these correlational findings is complicated by the well-known strong association between working memory capacity and general cognitive abilities (Frischkorn et al., 2022). To address this, we controlled for the covariance between the two constructs and found that *g* ($\beta = 0.41$, $p = .015$), rather than working memory capacity ($\beta = 0.22$, $p = .158$), was a significant predictor of mini-q performance. This is consistent with findings of the mediation analysis that a significant part of the association between mini-q performance and general cognitive abilities could not be accounted for by working memory capacity. Conceptually, it also corresponds with the mini-q's placement in the CHC hierarchy as a test of speeded reasoning rather than short-term memory (*Gsm*). Nevertheless, to causally test how much performance in the mini-q relies on one's working memory capacity, experimental studies are needed. Future studies could employ experimentally induced memory load to investigate the effects of experimental manipulations of working memory capacity on mini-q performance (Hagemann et al., 2023; Schubert et al., 2023).

Because 46% of the variance between mini-q performance and general cognitive abilities was not accounted for by working memory capacity, it is likely that further factors contributed to performance in both tests. One possible candidate domain-general ability affecting performance in both tests is attentional control (Engle, 2018; Shipstead, Harrison, and Engle, 2016), which is an umbrella term for self-regulatory abilities that facilitate the transformation and manipulation of mental representations in working memory by activating goal-relevant information and/or inhibiting goal-irrelevant information (Diamond, 2013; Friedman and Miyake, 2017; Miyake and Friedman, 2012; von Bastian et al., 2020). However, because individual differences in executive functions are strongly related to processing speed (Löffler et al., 2022), it may be that individual differences in attentional control make a comparably small contribution to the relationship between mini-q performance and general cognitive abilities as processing speed. Moreover, situational factors such as participants' fatigue, motivation, or alertness may have affected performance in both tests, as participants completed the two tests on the same day. Thus, these situational factors may also have contributed to the relationship between performances in the two tests, but likely only to a small degree, because previous research has shown that situational factors do not affect individual differences in BIS performance (Danner, Hagemann, Schankin, Hager, and Funke, 2011). To nevertheless rule out that the relationship between mini-q performance and general cognitive abilities was overestimated due to situational factors affecting both assessments, future studies should schedule separate measurement sessions for the assessment of cognitive abilities used to validate the mini-q.

Lastly, we evaluated to what degree the two domain-specific abilities required by the mini-q test material, verbal and figural abilities, contributed to test performance, and found that both abilities made a comparable contribution. In comparison, there was no evidence that any demands were placed on numerical abilities. The fact that mini-q performance is at least moderately dependent on verbal abilities has important implications for its potential use in large-scale panel studies because its reliance on verbal abilities may bias results against test takers who have a native language other than the language of the test material.

One limitation of the present study lies in the fact that, despite the inclusion of individuals from diverse educational and occupational backgrounds, the majority of participants had attained a minimum of secondary education. Hence, our sample did not include many participants from educationally disadvantaged backgrounds, which may warrant further attention in future studies.

Another limitation is that participants completed the mini-q, the cognitive abilities test, and the working memory capacity battery in the same measurement session, whereas processing speed was assessed in a separate session. Although it is well-known that individual differences in cognitive abilities, working memory capacity, and processing speed are temporally stable (i.e., that the rank order of participants remains relatively stable over time; Danner et al., 2011; Rummel, Hagemann, Steindorf, and Schubert, 2022; Schubert, Frischkorn, Hagemann, and Voss, 2016), it cannot be ruled out that relations between mini-q performance, cognitive abilities, and working memory capacity were overestimated because of situational factors (e.g., participants' fatigue or external noise) affecting all three measurements.

Taken together, our findings confirm that the mini-q can accurately measure speeded reasoning in a short, reliable, and valid manner that will be particularly useful as a short screening of cognitive abilities in large-scale panel studies. To promote its use in large-scale panel studies, we also developed a computer-based version of the mini-q implemented in lab.js (Henninger, Shevchenko, Mertens, Kieslich, and Hilbig, 2022) that can automatically administer the test, score responses, and facilitate online testing outside of the lab.[4] We validated this computer-based version that allows easier administration and scoring in a heterogeneous online sample in Study 2.

4. Study 2

The aims of Study 2 were to validate the computer-based version of the mini-q in an online testing environment and generate norm data for different age groups.

4.1. Method

4.1.1. Participants

505 participants (225 females, 272 males, 8 diverse) between 18 and 60 years (M = 31.33, SD = 8.86) from different educational and occupational backgrounds took part in an online testing session. Participants were recruited on the online platform Prolific (https://prolific.co/). The data was collected in 14 batches to ensure diverse education levels and to obtain at least 250 participants in each age group (18 to 30 and 31 to 60).

During data cleaning, 51 of the 556 participants who originally took part in the study were excluded because they reported being interrupted during the study, participating in a noisy environment, not speaking German fluently, encountering technical problems during the study, or participating more than once (in those cases, only data from the first participation was used). Additionally, a binomial test was performed for each participant, testing whether the proportion of correctly solved items to attempted (clicked) items was above 50% at a 95% chance level. Lastly, participants who prematurely skipped the mini-q items after <2.5 min and simultaneously had a proportion of correctly solved items lower than the mean – 3 SD of the sample (75.66% correct/clicked items) were excluded from analysis.

87.72% of participants reported speaking German as their native language, while the remaining 12.28% reported speaking German fluently (see Tables S1 and S2 in the supplementary materials for a description of the other languages spoken by the participants and the distribution of participant ethnicity). When asked about their highest level of education, 30.10% participants reported a high school diploma, 26.93% reported a bachelor's degree, and 20% reported a master's degree (for an overview of the highest level of education of the sample, see the left part of Fig. 7 and Table S3 in the supplementary materials). 52.87% of participants were employees and 29.50% were university students (for an overview of the current occupations of the sample, see the right part of Fig. 7 and Table S4 in the supplementary materials).

299 of the participants completed only the mini-q, and 206 participants under the age of 31 years (M = 24.74, SD = 3.25, 91 females, 113 males, 2 diverse) additionally completed the Short Form of the Hagen Matrices Test (HMT–S; Heydasch, Haubrich, and Renner, 2013).

The sample size of 645 participants (140 participants from Study 1 and 505 participants from Study 2) yielded a power of 78% to test the hypothesis of close fit as suggested by Browne and Cudeck (1992) for the structural equation model testing the assumption of configural measurement invariance (df = 4), an alpha error of $\alpha = 0.05$, and a power of $1 - \beta = 0.80$, a power of 94% for the model testing the assumption of metric measurement invariance (df = 7), and a power of 98% for the model testing the assumption of scalar measurement invariance (df = 10). The subsample of 206 participants in the younger age group yielded a power of 92% to test the hypothesis of close fit for the structural equation model estimating the latent correlation between mini-q and HMT-S performance (df = 26).

4.1.2. Material

Mini-q. The computer-based version was implemented in lab.js (Henninger et al., 2022) and designed to match the pen-and-paper version of the test. After reading the instructions, participants saw two solved sample items, then solved four more on their own. They received feedback on the solutions of the four items and then had three minutes to answer as many of the 64 items as possible. All items were on the same page. Participants clicked the left box for true or the right box for false, and the chosen box turned green upon selecting an option. They could change their answers until the time expired.

Incorrectly solved items, skipped items, and items not completed in time were scored as errors. On average, participants correctly solved 40.81 items (SD = 12.25, range 9–64 items). The median completion time for the mini-q was 4 min and 58 s (SD = 2:21), including instructions.

We again built parcels for structural equation modeling based on the assumption of unidimensionality by averaging performance across items across every third or fourth item. We built three parcels for assessing the relation between mini-q and HMT-S performance, and four parcels for reasons of identifiability for evaluating measurement invariance.

Cognitive abilities. We implemented the Short Form of the Hagen Matrices Test (HMT–S; Heydasch et al., 2013) in lab.js (Henninger et al., 2022) to measure participants' cognitive abilities. The HMT-S is a brief figural matrices test consisting of six items. These items are primarily designed to measure induction, reasoning, and fluid intelligence in accordance with the CHC model of intelligence (Schneider and McGrew, 2018). Items increase in difficulty and each item involves a 3 × 3 incomplete matrix, with one part missing that must be identified by recognizing the underlying rules of the pattern. Below the incomplete matrix, eight potential solutions to complete it are presented, but only one solution fits the pattern and is therefore correct.

After reading the instructions, participants solved two sample items

[4] We would like to thank an anonymous reviewer for this excellent suggestion.

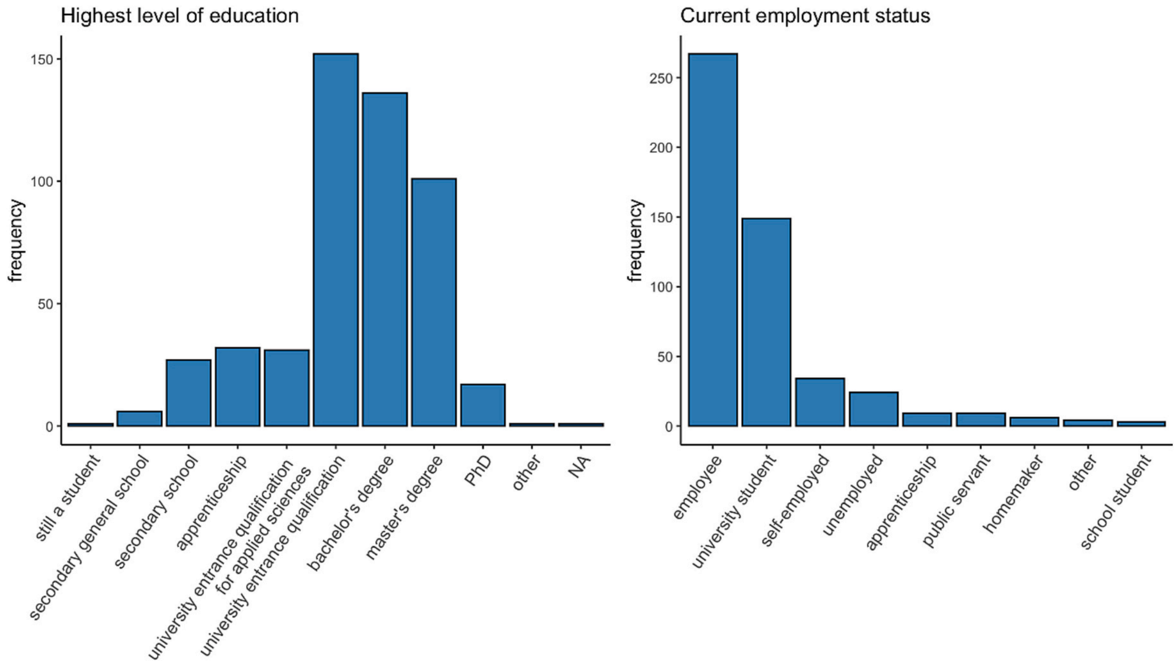

Fig. 7. Distribution of the highest level of education and occupation of participants in Study 2.

with feedback. Subsequently, they were allotted two minutes per each of the six test items, during which they had the liberty to modify their answers at their discretion, either before moving on to the next item or until the time limit had lapsed. Participants recorded their answer by clicking on the respective item, which then turned green to indicate the response.

Incorrectly solved items, skipped items, and items not completed in time were scored as errors. On average, participants correctly solved 4.48 items ($SD = 1.43$, range 0–6 items). The median completion time for the HMT-S was 5 min and 16 s ($SD = 3{:}06$), including instructions.

4.1.3. Procedure

Participants signed an informed consent form before taking part in the study. They received either £1.50 or £2.25 as a reward for their participation, depending on whether they only completed the mini-q or both the mini-q and the HMT–S. All participants completed the mini-q, and a subsample of 206 participants under the age of 31 proceeded to complete the HMT-S as well. All procedures were conducted in accordance with the declaration of Helsinki.

4.1.4. Data analysis

Structural equation modeling. We fitted covariance-based structural equation models with the *R* package *lavaan* (Rosseel, 2012), using the full information maximum likelihood estimator when models contained only continuous indicators and the diagonally weighted least squares estimator when models also contained binary indicators. We evaluated the model fits with the CFI (Bentler, 1990) and the RMSEA (Browne and Cudeck, 1992), using the same criteria as in Study 1. Measurement invariance was assessed through model comparisons using the likelihood ratio test. The statistical significance of model parameters was assessed with the two-sided critical ratio test.

Norm scores. We calculated IQ scores as norm scores separately per age group and for the overall sample. The young age group contained 254 participants between the ages of 18 and 30, while the old age group contained 251 participants between the ages of 31 and 60.

Data and code availability. The data and code supporting the findings of the study are available in the Open Science Framework repository at https://osf.io/bpm5e/.

4.2. Results

Descriptive statistics and reliability estimates for all measures are shown in Table 3 (see Fig. 8 for the distribution of the mini-q performance and Figs. S4 and S5 in the supplementary material for separate distributions per age group and the distribution of HMT-S performance). The computer-based version of the mini-q showed an excellent reliability (see Table 3). The probability of correctly solving an item decreased throughout the test, which indicates that participants rarely skipped items as instructed (see Fig. S6 in the supplementary materials for an item-wise analysis comparing data from the pen-and-paper and computer-based versions). Zero-order correlations of all variables are reported in Table 4.

4.2.1. How strongly is performance in the computer-based version of the mini-q related to cognitive abilities?

We estimated the latent correlation between mini-q performance and HMT-S performance using a structural equation model. The model provided an excellent account of the data, $\chi^2(26) = 26.12$, $p = .456$, CFI $= 1.00$, RMSEA $= 0.01$ (90% CI = [0.00; 0.06]). Cognitive abilities as assessed with the short version of the HMT and mini-q performance were moderately correlated, $r = 0.34$ (95% CI = [0.18; 0.50]), $p = .006$, indicating that the computer-based version of the mini-q is suitable for the brief assessment of fluid cognitive abilities in online studies.

4.2.2. Are the pen-and-paper and computer-based versions of the mini-q measurement invariant?

Combining the data from Studies 1 and 2, we tested the pen-and-

Table 3
Descriptive statistics and reliability estimates for all measures in Study 2.

Test	*N*	*M*	*SD*	Range	Reliability
mini-q	505	40.81	12.25	9–64	.99[a]/.96[b]
HMT-S	206	4.48	1.43	0–6	.60[b]

[a] Reliability estimates are based on Spearman Brown corrected odd-even split correlations.
[b] Reliability estimates are based on Cronbach's α. HMT-S = Short Form of the Hagen Matrices Test.

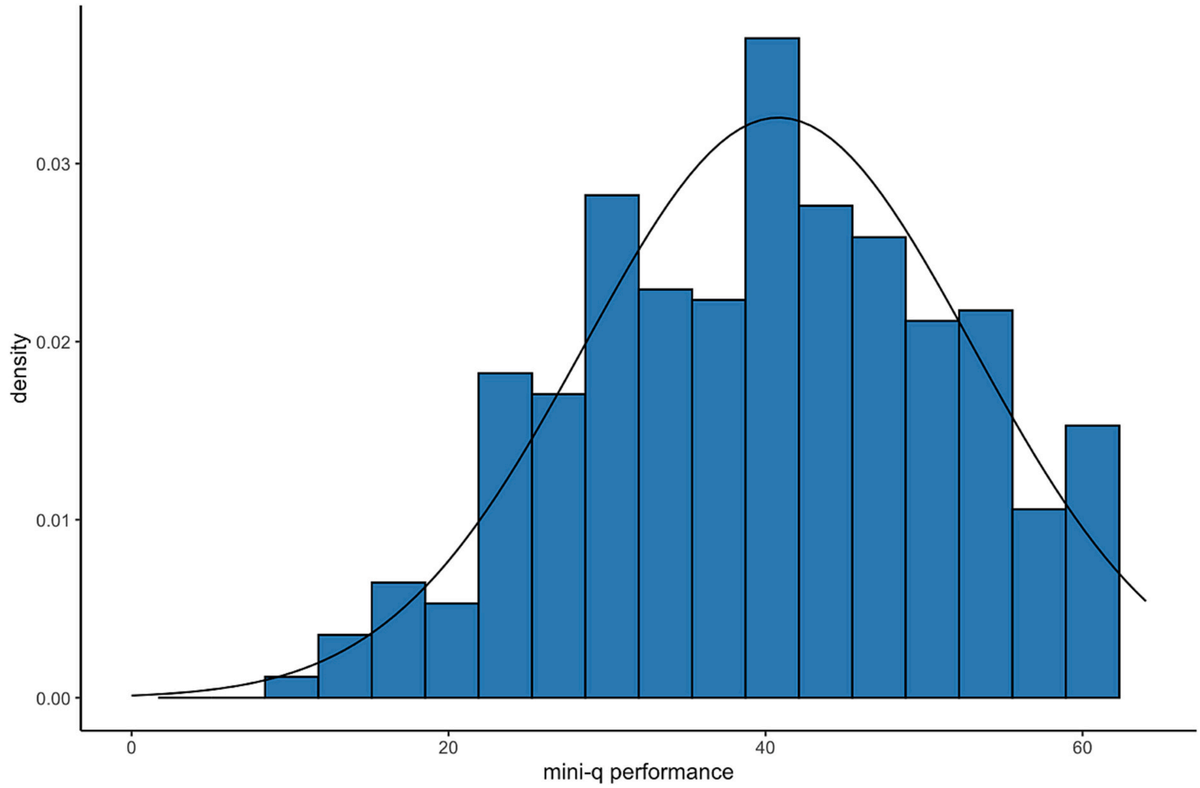

Fig. 8. Distribution of performance in the mini-q in Study 2 (online study).

Table 4
Zero-order correlations between all variables in study 2 in the HMT-S subsample (N = 206, age 18 to 30).

	mini-q	HMT-S	mini-q_1	mini-q_2	mini-q_3	HMT-S_1	HMT-S_2	HMT-S_3	HMT-S_4	HMT-S_5
mini-q	–									
HMT-S	**0.24**	–								
mini-q_1	**0.99**	**0.25**	–							
mini-q_2	**0.99**	**0.25**	**0.97**	–						
mini-q_3	**0.99**	**0.23**	**0.98**	**0.98**	–					
HMT-S_1	0.10[a]	**0.47**[a]	0.10[a]	0.11[a]	0.08[a]	–				
HMT-S_2	**0.24**[a]	**0.51**[a]	**0.24**[a]	**0.24**[a]	**0.24**[a]	**0.28**[b]	–			
HMT-S_3	0.06[a]	**0.62**[a]	0.07[a]	0.06[a]	0.06[a]	**0.18**[b]	**0.24**[b]	–		
HMT-S_4	**0.16**[a]	**0.64**[a]	**0.17**[a]	**0.16**[a]	**0.16**[a]	**0.14**[b]	**0.16**[b]	**0.29**[b]	–	
HMT-S_5	0.13[a]	**0.58**[a]	0.13[a]	**0.14**[a]	0.12[a]	0.13[b]	**0.15**[b]	**0.17**[b]	**0.29**[b]	–
HMT-S_6	**0.17**[a]	**0.63**[a]	**0.18**[a]	**0.17**[a]	**0.17**[a]	0.10[b]	**0.23**[b]	**0.25**[b]	**0.25**[b]	**0.18**[b]

Note. Bold values are significant at α = 0.05 level. mini-q = sum score of the mini-q after three minutes; HMT-S = sum score of the Hagen Matrices Test; mini-q_1 = first parcel of the mini-q; mini-q_2 = second parcel of the mini-q; mini-q_3 = third parcel of the mini-q; HMT-S_1 = binary score of the first item of the Hagen Matrices Test; HMT-S_2 = binary score of the second item of the Hagen Matrices Test; HMT-S_3 = binary score of the third item of the Hagen Matrices Test; HMT-S_4 = binary score of the fourth item of the Hagen Matrices Test; HMT-S_5 = binary score of the fifth item of the Hagen Matrices Test; HMT-S_6 = binary score of the sixth item of the Hagen Matrices Test.
[a] Point biserial correlation.
[b] phi correlation.

paper and computer-based versions on three levels of measurement invariance: Configural invariance (i.e., the same model holds for both versions), metric invariance (i.e., factor loadings are the same across both versions), and scalar invariance (i.e., intercepts and factor loadings are the same across both versions). Model comparisons favored scalar over both configural and metric invariance (see Table 5), as model fit did not deteriorate when we assumed metric instead of configural measurement invariance, $\Delta\chi^2(3) = 4.83$, $p = .184$, nor when we assumed scalar instead of metric measurement invariance, $\Delta\chi^2(3) = 2.92$, $p = .405$. Consequently, data collected with the pen-and-paper and the computer-based version of the test can simply be collapsed without accounting for the type of test administration even when one part of the data was collected in person, and the other part of the data was collected

Table 5
Measurement invariance testing.

Level	χ^2	*df*	*p*	CFI	RMSEA [90% CI][1]	AIC
Configural	30.56	4	<0.001	1.00	0.14 [0.10; 0.19]	−6981.55
Metric	35.39	7	<0.001	1.00	0.11 [0.08; 0.15]	−6982.72
Scalar	38.31	10	<0.001	1.00	0.09 [0.06; 0.12]	−6985.80

Note. The RMSEA tends to underestimate model fit when the degrees of freedom are low (Hu and Bentler, 1999; Kenny, Kaniskan, and McCoach, 2015).

online.

4.3. Norm tables

We created a joint (see Table 6) as well as separate (see Table 7) norm tables for younger (18 to 30 years) and middle-aged (31 to 60 years) adults, because there was a significant difference in performance between participants in the younger age group ($M = 42.58$, $SD = 12.53$) compared to the older age group ($M = 39.01$, $SD = 11.70$), $t(503) = 3.31$, $p = .001$, $d = 0.29$ (95% CI = [0.12; 0.47]). Across both age groups, the performance in the mini-q was weakly correlated with age at $r = -0.10$, $p = .030$, (95% CI = [−0.18; −0.01]).

4.4. Discussion

In Study 2, we developed and validated a computer-based version of the mini-q in an online testing environment and generated norm data for different age groups. Our findings confirmed that the pen-and-paper and computer-based versions exhibited scalar measurement invariance. Hence, performance differences between any individuals assessed with different versions of the mini-q can only be attributed to the individuals themselves, rather than the mode of test administration. As a result, researchers are free to use both versions interchangeably, even within a single study.

Performance in the computer-based version of the mini-q was moderately related to performance in the short version of the HMT, a validated matrix reasoning test (Heydasch et al., 2013). This correlation was observed in an age-homogeneous sample consisting of participants between 18 and 30 years of age, which supports our previous conclusion that the relationship between mini-q and BIS performance was not overestimated due to participants' broad age range in Study 1. The correlation between the mini-q and the HMT–S, observed in the present study, was only moderate (r = 0.32). However, it was comparable to the correlations between the HMT-S and verbal reasoning ($r = 0.30$) as well as figural reasoning ($r = 0.47$) measured with the IST-2000-R, as reported in the HMT-S test manual (Heydasch et al., 2013). This indicates that the limited variance of the short version of the HMT, and not the validity of the computer-based mini-q, limited this correlation, as the correlation between the HMT-S and verbal and figural reasoning factors derived from a larger test battery were of similar size. In conclusion, this finding adds further evidence for the mini-q's usefulness as a brief assessment of cognitive abilities, whether it is administered in-person or online.

Table 6
Norm table for the overall sample (age 18 to 60, N = 505).

mini-q	IQ score	mini-q	IQ score
9	61	38	97
11	63	39	98
12	65	40	99
13	66	41	100
15	68	42	101
16	70	43	103
17	71	44	104
18	72	45	105
19	73	46	106
20	75	47	108
21	76	48	109
22	77	49	110
23	78	50	111
24	79	51	112
25	81	52	114
26	82	53	115
27	83	54	116
28	84	55	117
29	86	56	119
30	87	57	120
31	88	58	121
32	89	59	122
33	90	60	124
34	92	61	125
35	93	62	126
36	94	63	127
37	95	64	128

Table 7
Norm tables for separate age groups.

Age group 18 to 30 years ($N = 254$)				Age group 31 to 60 years ($N = 251$)			
mini-q	IQ score	mini-q	IQ score	mini-q	IQ score	mini-q	IQ score
11	62	40	97	9	62	40	101
12	63	41	98	15	69	41	103
13	65	42	99	16	70	42	104
17	69	43	100	17	72	43	105
19	72	44	102	18	73	44	106
20	73	45	103	19	74	45	108
21	74	46	104	21	77	46	109
22	75	47	105	22	78	47	110
23	77	48	106	23	79	48	112
24	78	49	108	24	81	49	113
25	79	50	109	25	82	50	114
26	80	51	110	26	83	51	115
27	81	52	111	27	85	52	117
28	83	53	112	28	86	53	118
29	84	54	114	29	87	54	119
30	85	55	115	30	88	55	121
31	86	56	116	31	90	57	123
32	87	57	117	32	91	58	124
33	89	58	118	33	92	59	126
34	90	59	120	34	94	60	127
35	91	60	121	35	95	61	128
36	92	61	122	36	96	62	129
37	93	62	123	37	97	63	131
38	95	63	124	38	99	64	132
39	96	64	126	39	100		

We provided norm data for the computer-based version of the mini-q separately for young adults (18 to 30 years) and middle-aged adults (31 to 60 years). We decided to provide different norm data for young and middle-aged because we found evidence for a small effect of age group on mini-q performance ($d = 0.29$), and separated the groups at the age of 30 because processing speed, which has a large influence on test performance, peaks rather early in life (Hartshorne and Germine, 2015). These norm data can also be used for the pen-and-paper version of the test since the two test versions have scalar measurement invariance across them.

Despite aiming to recruit participants from different educational and occupational backgrounds through the online platform Prolific, the sample of Study 2 was still slightly skewed to a higher educational attainment compared to the reference group of 15- to 65-year-olds (Statistisches Bundesamt Destatis, 2023a, 2023b). 65 participants (12.87%) reported having attended secondary school as their highest educational attainment, compared to 48.57% in the reference group. 183 participants (36.24%) had university entrance qualifications as their highest level of education, compared to 20.42% in the reference group. 255 participants (50.50%) reported having attained a university degree, compared to 21.11% in the reference group. 149 participants (29.50%) were university students at the time of completion, compared to 6.18% in the reference group. Future studies may therefore need to exert more effort in recruiting participants with lower educational attainment.

5. General discussion

We assessed the utility of a pen-and-paper and computer-based version of the mini-q, a three-minute speeded reasoning test (Baudson and Preckel, 2016), in two heterogeneous samples to determine the test's usefulness for a brief assessment of general cognitive abilities in large-scale panel studies.

Across both studies and test versions, we found that test scores obtained with the mini-q were remarkably reliable, despite the test's brevity. Moreover, mini-q performance was substantially related to general cognitive abilities as measured with a broad test battery, and to fluid reasoning as measured with a matrix reasoning test. The largest part of the relationship between test performance and general cognitive abilities was accounted for by participants' working memory capacity, which suggests that the mini-q measures general cognitive abilities to the degree that it measures working memory capacity as a domain-general ability. Taken together, our findings support the applicability of the test as a short measure of cognitive abilities in large-scale panel studies and other studies with limited time available for the assessment of cognitive abilities.

6. Recommendations for using the mini-q in large-scale panel studies

Our results support the notion that both the pen-and-paper and the computer-based versions of the mini-q can be used as brief, reliable, and valid assessments of general cognitive abilities. We therefore derived a set of recommendations based on our results for using the mini-q in future studies.

First and foremost, we would like to emphasize that the mini-q should only be used as a brief screening of cognitive abilities, but that it cannot replace a more comprehensive assessment of general intelligence with a broad test battery. Researchers with limited time for assessing cognitive abilities within their overall study can take advantage of the mini-q's short administration time, which is approximately 5 min (including instructions). Thus, the mini-q is ideally suited for researchers who are interested in a brief screening of cognitive abilities or who plan to include cognitive abilities as one of several covariates in their main analyses. If more time is available and researchers are interested at measuring general cognitive abilities at the construct level, the Pathfinder test (Malanchini et al., 2021) or the TMT DNT (Singh et al., 2021), which take about 15 and 40 min, respectively, may be better alternatives, because they allow a broader assessment of general cognitive abilities and therefore the extraction of a latent factor of general cognitive abilities across different subtests.

Our second recommendation pertains to the relationship between age differences and mini-q performance. Across both studies, we found that older adults performed worse in the mini-q than younger ones, although the latent correlations between mini-q performance and general cognitive abilities remained virtually unchanged after controlling for age differences in Study 1. Nevertheless, we recommend considering participants' age distribution when using the mini-q as a brief measure of cognitive abilities. Specifically, we suggest either statistically controlling for age or employing age-specific IQ scores instead of raw scores in instances where the sample has a broad age range.

Third, we found that mini-q performance was at least moderately dependent on verbal abilities. This has important implications for its potential use in large-scale panel studies because the test's reliance on verbal abilities may bias results against test takers who have a native language other than the language of the test material. This potential bias is an important issue to consider when using the mini-q in samples with different cultural and educational backgrounds, because cognitive ability tests are only valid as long as test takers are similarly familiar with the test material (Saklofske, van de Vijver, Oakland, Mpofu, and Suzuki, 2015). In the case of the mini-q, this problem could be addressed by developing and validating different language versions of the test. Moreover, the degree of bias of different item types (e.g., items with active vs. passive voice, items with vs. without negation) could be directly assessed in culturally diverse samples through tests of differential item functioning applicable to unidimensional tests (Holland and Wainer, 2012). Given that mini-q performance depends on verbal abilities, we therefore recommend considering potential biases when samples include participants with different native languages and various levels of language proficiency.

Our last recommendation concerns the online administration of the computer-based version. In Study 2, we had to exclude several participants because they were exceptionally fast, demonstrated clear low-effort, were interrupted during the test, or completed the test in a noisy environment. We therefore emphasize that it is important to instruct participants to complete the test in a quiet environment where they are unlikely to be interrupted and either turn off their phone or set it to silent mode. Moreover, researchers should always collect data about test completion time and ask participants about any interruptions after test completion. Using these data, data collected online can then be screened using pre-defined criteria such as a minimally acceptable completion time or the proportion of correctly solved items to attempted (clicked) items exceeding chance level (for more recommendations on how to conduct online studies in cognitive psychology, see Gagné and Franzen, 2023; Uittenhove, Jeanneret, and Vergauwe, 2023).

7. Conclusions

We validated the mini-q, a three-minute speeded reasoning test (Baudson and Preckel, 2016), in a heterogeneous sample of participants from different educational and occupational backgrounds to evaluate its usefulness for a brief assessment of general cognitive abilities in large-scale panel studies. Although it only took three minutes to complete the test, participants' test performance showed an excellent reliability and was substantially related to their general cognitive abilities measured with a broad test battery, supporting the applicability of the test as a short screening of cognitive abilities. The largest part of the relationship between test performance and general cognitive abilities was accounted for by participants' working memory capacity, which suggests that the mini-q measures general cognitive abilities to the degree that it measures working memory capacity as a domain-general ability. Overall, our results support the notion that the mini-q can be used as a brief, reliable, and valid assessment of general cognitive abilities both face-to-face as well as online, but possible disadvantages of participants with different native languages should be carefully considered due to the test's dependency on verbal abilities.

Author Note

This work was supported by the German Research Foundation (DFG) [grant number SCHU 3266/1-1]. The data supporting the findings of the study are available in the Open Science Framework repository at https://osf.io/bpm5e/.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Anna-Lena Schubert reports financial support was provided by German Research Foundation.

Data availability

The data supporting the findings of the study are available in the Open Science Framework repository at https://osf.io/bpm5e/.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.intell.2023.101804.

References

Baddeley, A. D. (1968). A 3 min reasoning test based on grammatical transformation. *Psychonomic Science, 10*(10), 341–342. https://doi.org/10.3758/BF03331551 von Bastian, C. C., Blais, C., Brewer, G., Gyurkovics, M., Hedge, C., Kałamała, P., … Wiemers, E. (2020). Advancing the understanding of individual differences in attentional control: Theoretical, methodological, and analytical considerations. *PsyArXiv.*. https://doi.org/10.31234/osf.io/x3b9k

Baudson, T. G., & Preckel, F. (2016). mini-q: Intelligenzscreening in drei Minuten. *Diagnostica, 62*(3), 182–197. https://doi.org/10.1026/0012-1924/a000150

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238–246.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research, 21*(2), 230–258. https://doi.org/10.1177/0049124192021002005

Carroll, J. B. (1993). *Human Cognitive Abilities: A Survey of Factor-Analytic Studies*. Cambridge University Press. https://doi.org/10.1017/CBO9780511571312

Chuderski, A. (2019). Even a single trivial binding of information is critical for fluid intelligence. *Intelligence, 77*, Article 101396. https://doi.org/10.1016/j.intell.2019.101396

Danner, D., Hagemann, D., Schankin, A., Hager, M., & Funke, J. (2011). Beyond IQ: A latent state-trait analysis of general intelligence, dynamic decision making, and implicit learning. *Intelligence, 39*(5), 323–334. https://doi.org/10.1016/j.intell.2011.06.004

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*, 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Engle, R. W. (2018). Working memory and executive attention: A revisit. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science, 13*(2), 190–193. https://doi.org/10.1177/1745691617720478

Fawns-Ritchie, C., & Deary, I. J. (2020). Reliability and validity of the UK Biobank cognitive tests. *PLoS One, 15*(4), Article e0231627. https://doi.org/10.1371/journal.pone.0231627

Forscher, P. S., Wagenmakers, E.-J., Coles, N. A., Silan, M. A., Dutra, N., Basnight-Brown, D., & IJzerman, H.. (2022). The benefits, barriers, and risks of big-team science. *Perspectives on Psychological Science*. https://doi.org/10.1177/17456916221082970, 17456916221082970.

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex, 86*, 186–204. https://doi.org/10.1016/j.cortex.2016.04.023

Frischkorn, G. T., & Schubert, A.-L. (2018). Cognitive models in intelligence research: Advantages and recommendations for their application. *Journal of Intelligence, 6*(3), 34. https://doi.org/10.3390/jintelligence6030034

Frischkorn, G. T., Wilhelm, O., & Oberauer, K. (2022). Process-oriented intelligence research: A review from the cognitive perspective. *Intelligence, 94*, Article 101681. https://doi.org/10.1016/j.intell.2022.101681

Gagné, N., & Franzen, L. (2023). How to run behavioural experiments online: Best practice suggestions for cognitive psychology and neuroscience. *PsyArXiv Preprints, 3*(1). https://doi.org/10.5334/spo.34. Article 1.

Hagemann, D., Ihmels, M., Bast, N., Neubauer, A. B., Schankin, A., & Schubert, A.-L. (2023). Fluid intelligence is (much) more than working memory capacity: An experimental analysis. *Journal of Intelligence, 11*(4). https://doi.org/10.3390/jintelligence11040070. Article 4.

Hartshorne, J. K., & Germine, L. T. (2015). When does cognitive functioning peak? The asynchronous rise and fall of different cognitive abilities across the life span. *Psychological Science, 26*(4), 433–443. https://doi.org/10.1177/0956797614567339

Henninger, F., Shevchenko, Y., Mertens, U. K., Kieslich, P. J., & Hilbig, B. E. (2022). lab.js: A free, open, online study builder. *Behavior Research Methods, 54*(2), 556–573. https://doi.org/10.3758/s13428-019-01283-5

Heydasch, T., Haubrich, J., & Renner, K.-H. (2013). The short version of the Hagen matrices test (HMT-S): 6-item induction intelligence test. *Methods, Data, Analyses, 7*(2). https://doi.org/10.12758/mda.2013.011. Article 2.

Hilger, K., Spinath, F. M., Troche, S., & Schubert, A.-L. (2022). The biological basis of intelligence: Benchmark findings. *Intelligence, 93*, Article 101665. https://doi.org/10.1016/j.intell.2022.101665

Holland, P. W., & Wainer, H. (2012). *Differential Item Functioning*. Routledge. https://doi.org/10.4324/9780203357811

Hossiep, R., Turck, D., & Hasella, M. (1999). *BOMAT – Advanced – Bochumer Matrizentest*. Hogrefe.

Hossiep, R., Turck, D., & Hasella, M. (2001). *BOMAT – Advanced – Short version—Bochumer Matrizentest*. Hogrefe. https://www.testzentrale.de/shop/bochumer-matrizentest-43756.html#1+1.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Jäger, A. O. (1982). Mehrmodale Klassifikation von Intelligenzlestungen: Experimentell kontrollierte Weiterenwicklung eines deskriptiven Intelligenzstrukturmodells. [Multimodal classification of intelligence achievement: Experimentally controlled, further development of a descriptive intelligence structure model.]. *Diagnostica, 28*, 195–225.

Jäger, A. O., Süß, H.-M., & Beauducel, A. (1997). *Berliner Intelligenzstruktur-Test. Form 4*. Hogrefe.

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research, 44*(3), 486–507. https://doi.org/10.1177/0049124114543236

Kievit, R. A., McCormick, E. M., Fuhrmann, D., Deserno, M. K., & Orben, A. (2022). Using large, publicly available data sets to study adolescent development: Opportunities and challenges. *Current Opinion in Psychology, 44*, 303–308. https://doi.org/10.1016/j.copsyc.2021.10.003

Lehrl, S. (1999). *Mehrfachwahl-Wortschatz-Intelligenztest MWT-B (4. Aufl.)*. Spitta Verlag.

Lewandowsky, S., Oberauer, K., Yang, L.-X., & Ecker, U. K. H. (2010). A working memory test battery for MATLAB. *Behavior Research Methods, 42*(2), 571–585. https://doi.org/10.3758/BRM.42.2.571

Liepmann, D., Beauducel, A., Brocke, B., & Amthauer, R. (2010). *Intelligence Structure Test (I-S-T 2000 R)*. Hogrefe. https://www.hogrefe.com/de/shop/intelligence-structure-test.html.

Liepmann, D., Beauducel, A., Brocke, B., & Nettelnstroth, W. (2012). *IST-Screening—Intelligence Structure Test*. Hogrefe.

Löffler, C., Frischkorn, G. T., Hagemann, D., Sadus, K., & Schubert, A.-L. (2022). The common factor of executive functions measures nothing but speed of information uptake. *PsyArXiv.*. https://doi.org/10.31234/osf.io/xvdyz

MacKinnon, D. P., & Pirlott, A. G. (2015). Statistical approaches for enhancing causal interpretation of the M to Y relation in mediation analysis. *Personality and Social Psychology Review, 19*(1), 30–43. https://doi.org/10.1177/1088868314542878

Malanchini, M., Rimfeld, K., Gidziela, A., Cheesman, R., Allegrini, A. G., Shakeshaft, N., … Plomin, R. (2021). Pathfinder: A gamified measure to integrate general cognitive ability into the biological, medical, and behavioural sciences. *Molecular Psychiatry, 26*(12). https://doi.org/10.1038/s41380-021-01300-0. Article 12.

Marek, S., Tervo-Clemmens, B., Calabro, F. J., Montez, D. F., Kay, B. P., Hatoum, A. S., … Dosenbach, N. U. F. (2022). Reproducible brain-wide association studies require thousands of individuals. *Nature, 603*(7902), 654–660. https://doi.org/10.1038/s41586-022-04492-9

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence, 37*(1), 1–10. https://doi.org/10.1016/j.intell.2008.08.004

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science, 21*(1), 8–14. https://doi.org/10.1177/0963721411429458

Muszyński, M., Banasik-Jemielniak, N., Żółtak, T., Rimfeld, K., Shakeshaft, N. G., Schofield, K. L., … Pokropek, A. (2023). Moving intelligence measurement online: Adaptation and validation of the polish version of the pathfinder general cognitive ability test. *PsyArXiv.*. https://doi.org/10.31234/osf.io/tqyux

Oberauer, K., Lewandowsky, S., Awh, E., Brown, G. D. A., Conway, A., Cowan, N., … Ward, G. (2018). Benchmarks for models of short-term and working memory. *Psychological Bulletin, 144*(9), 885–958. psyh https://doi.org/10.1037/bul0000153.

Oberauer, K., Süβ, H.-M., Wilhelm, O., & Wittmann, W. W. (2008). Which working memory functions predict intelligence? *Intelligence, 36*(6), 641–652. https://doi.org/10.1016/j.intell.2008.01.007

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716. https://doi.org/10.1126/science.aac4716

Posner, M. I., & Mitchell, R. F. (1967). Chronometric analysis of classification. *Psychological Review, 74*(5), 392–409. https://doi.org/10.1037/h0024913

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*(2), 59–108. https://doi.org/10.1037/0033-295X.85.2.59

Raven, J. C., Court, J. H., & Raven, J. (1994). Manual for Raven's progressive matrices and mill hill vocabulary scales. In *Advanced Progressive Matrices*. Oxford University Press.

Rohrer, J. M., Hünermund, P., Arslan, R. C., & Elson, M. (2022). That's a lot to process! Pitfalls of popular path models. *Advances in Methods and Practices in Psychological Science, 5*(2). https://doi.org/10.1177/25152459221095827, 25152459221095828.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36.

Rummel, J., Hagemann, D., Steindorf, L., & Schubert, A.-L. (2022). How consistent is mind wandering across situations and tasks? A latent state–trait analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 48*, 1385–1399. https://doi.org/10.1037/xlm0001041

Saklofske, D. H., van de Vijver, F. J. R., Oakland, T., Mpofu, E., & Suzuki, L. A. (2015). Intelligence and culture: History and assessment. In S. Goldstein, D. Princiotta, & J. A. Naglieri (Eds.), *Handbook of Intelligence: Evolutionary Theory, Historical Perspective, and Current Concepts* (pp. 341–365). Springer. https://doi.org/10.1007/978-1-4939-1562-0_22.

Schmidt, F. L., & Hunter, J. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology, 86*(1), 162–173. https://doi.org/10.1037/0022-3514.86.1.162

Schmitz, F., & Wilhelm, O. (2015). *Modeling Mental Speed [Presentation]*. Amsterdam, NL: International Convention of Psychological Science.

Schneider, W. J., & McGrew, K. S. (2018). The Cattell–Horn–Carroll theory of cognitive abilities. In *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (4th ed., pp. 73–163). The Guilford Press.

Schubert, A.-L., & Frischkorn, G. T. (2020). Neurocognitive psychometrics of intelligence: How measurement advancements unveiled the role of mental speed in intelligence differences. *Current Directions in Psychological Science, 29*(2), 140–146. https://doi.org/10.1177/0963721419896365

Schubert, A.-L., Frischkorn, G. T., Hagemann, D., & Voss, A. (2016). Trait characteristics of diffusion model parameters. *Journal of Intelligence, 4*(3), 7. https://doi.org/10.3390/jintelligence4030007

Schubert, A.-L., Frischkorn, G. T., Sadus, K., Welhaf, M., Kane, M. J., & Rummel, J. (2023). *The Brief Mind Wandering Three-Factor Scale (BMW-3)*. PsyArxiv.

Schubert, A.-L., Hagemann, D., Löffler, C., & Frischkorn, G. T. (2020). Disentangling the effects of processing speed on the association between age differences and fluid intelligence. *Journal of Intelligence, 8*(1), 1. https://doi.org/10.3390/jintelligence8010001

Schubert, A.-L., Hagemann, D., Voss, A., Schankin, A., & Bergmann, K. (2015). Decomposing the relationship between mental speed and mental abilities. *Intelligence, 51*, 28–46. https://doi.org/10.1016/j.intell.2015.05.002

Schubert, A.-L., Löffler, C., & Hagemann, D. (2022). A neurocognitive psychometrics account of individual differences in attentional control. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0001184

Schubert, A.-L., Löffler, C., Hagemann, D., & Sadus, K. (2022). How robust is the relationship between neural processing speed and cognitive abilities? *Psychophysiology, e14165*. https://doi.org/10.1111/psyp.14165

Schubert, A.-L., Löffler, C., Sadus, K., Göttmann, J., Hein, J., Schröer, P., Teuber, A., & Hagemann, D. (2023). Working memory load affects intelligence test performance by reducing the strength of relational item bindings and impairing the filtering of irrelevant information. *Cognition, 236*, Article 105438. https://doi.org/10.1016/j.cognition.2023.105438

Shipstead, Z., Harrison, T. L., & Engle, R. W. (2016). Working memory capacity and fluid intelligence: Maintenance and disengagement. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science, 11*(6), 771–799. https://doi.org/10.1177/1745691616650647

Singh, S., Strong, R. W., Jung, L., Haofei Li, F., Grinspoon, L., Scheuer, L. S., … Germine, L. (2021). The TestMyBrain digital neuropsychology toolkit: Development and psychometric characteristics. *Journal of Clinical and Experimental Neuropsychology, 43*(8), 786–795. https://doi.org/10.1080/13803395.2021.2002269

Statistisches Bundesamt (Destatis). *Population aged 15 and over in primary residence households: Germany, years, sex, age groups, general school education (12211–0100)* [dataset]. https://www-genesis.destatis.de/genesis//online?operation=table&code=12211-0101&bypass=true&levelindex=0&levelid=1701429051952#abreadcrumb.

Statistisches Bundesamt (Destatis). *Population aged 15 and over in primary residence households: Germany, years, sex, age groups, vocational qualification attained (12211–0101)* [dataset]. https://www-genesis.destatis.de/genesis//online?operation=table&code=12211-0100&bypass=true&levelindex=0&levelid=1701429051952#abreadcrumb.

Sternberg, S. (1969). Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist, 57*(4), 421–457.

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., … Collins, R. (2015). UK biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Medicine, 12*(3), Article e1001779. https://doi.org/10.1371/journal.pmed.1001779

U.S. Department of Defense. (1984). *Test Manual for the Armed Services Vocational Aptitude Battery (DoD 1340.12AA)*. U.S. Military Entrance Processing Command.

Uittenhove, K., Jeanneret, S., & Vergauwe, E. (2023). From lab-testing to web-testing in cognitive research: Who you test is more important than how you test. *Journal of Cognition, 6*(1). https://doi.org/10.5334/joc.259. Article 1.

Voss, A., & Voss, J. (2007). Fast-dm: A free program for efficient diffusion model analysis. *Behavior Research Methods, 39*(4), 767–775. https://doi.org/10.3758/BF03192967

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale* (4th ed.). Pearson Assessment.

Wechsler, D. (2011). *Wechsler Abbreviated Scale of Intelligence* (2nd ed.). Pearson Assessment https://www.pearsonassessments.com/store/usassessments/en/Store/Professional-Assessments/Cognition-%26-Neuro/Wechsler-Abbreviated-Scale-of-Intelligence-%7C-Second-Edition/p/100000593.html.

Weiß, R. H. (2008). *Grundintelligenztest Skala 2 – Revision (CFT 20-R) mit Wortschatztest und Zahlenfolgetest – Revision (WS/ZF-R)*. Hogrefe. https://www.testzentrale.de/shop/grundintelligenztest-skala-2-revision-cft-20-r-mit-wortschatztest-ws-und-zahlenfolgentest-zf-revision-ws-zf-r-90116.html.

Wilhelm, O., Hildebrandt, A. H., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it? *Frontiers in Psychology, 4*. https://doi.org/10.3389/fpsyg.2013.00433