methods, data, analyses | 2013, pp. 183e-205e DOI: 10.12758/mda.2013.021

Translation of “The Short Version of the Hagen Matrices Test (HMT-S). A 6-Item Induction Intelligence Test” by Heydasch, Haubrich, and Renner (2013)

Abstract

The short version of the Hagen Matrices Test (HMT-S) is a free of charge online intelligence test measuring induction in reference to the CHC model of intelligence (Schneider & McGrew, 2012). The 6-item HMT-S is based on the 20-item Hagen Matrices Test (HMT; Heydasch, 2014). The internal consistency of the HMT-S in our studies was .62. The correlations with the original scale were $r = .79$ in a first study and $r = .78$ in a second one. Additionally, convergent validity was shown by correlations with the Intelligence-Structure-Test 2000 R (Liepmann, Beauducel, Brocke, & Amthauer, 2007). Associations with academic performance indicated criterion related validity. To sum it up, the HMT-S is a reliable, valid, economic, and efficient intelligence test. Free applications can be requested from the author via email.

Keywords: short version of the Hagen Matrices Test, HMT-S, HMT, intelligence, measurement

Important life outcomes are connected to intelligence, and intelligence has been found to play a key role in success at school, universities (Kuncel, Hezlett, & Ones, 2004; Kunina, Wilhelm, Formazin, Jonkmann, & Schroeders, 2007; Poropat, 2009; Rindermann & Neubauer, 2000), and at work (Harrell & Harrell, 1945; Hülsheger, Maier, & Stumpp, 2007; Hunter & Hunter, 1984; Judge, Higgins, Thoresen, & Barrick, 1999; Kuncel et al., 2004; Ng, Eby, Sorensen, & Feldman, 2005; Salgado, Anderson, Moscoso, Bertua, de Fruyt, & Rolland, 2003; Schmidt & Hunter, 1998; Ziegler, Dietl, Danay, Vogel, & Bühner, 2011; see also Strenze, 2007). Other areas such as physical health (e.g., Khandaker, Barnett, White, & Jones, 2011), divorce, and death (Roberts, Kuncel, Shiner, Caspi, & Goldberg, 2007) have also been found to be associated with intelligence.

Any authors who are planning a study in which intelligence will play an important part – as predictor, moderator, control variable (see Blickle, Kramer, & Mierke, 2010), mediator, or outcome – may first be happy about having an abundance of choices when considering the many and diverse intelligence measures that are seem to be available. However, besides the need to make decisions about the definition, theory, and chosen domains, economic considerations are also important. Therefore, two core problems may arise: First, many of the measures are too long and require participants or test administrators to invest unrealistic amounts of time; and second, the tests must be purchased, which may result in considerable costs to researchers.

Free short tests of general intelligence or specific domains are rare. As a non-commercial measure, the 20-item Hagen Matrices Test (Heydasch, 2014)[1] is one of them. Its reliability (internal consistency and retest reliability) and convergent and divergent validity have been demonstrated. In addition, associations with measures of success in school and at universities have indicated criterion-related validity. However, in some cases, with 20 tasks, the HMT may be too long, especially if there is a need to reduce costs by offering short and therefore less time-consuming measures to increase the number of people who are willing to participate. Another problem with the HMT is its difficulty: The HMT was developed for students to predict their success at universities, and as a consequence of wanting to ensure an

1 The original reference was Heydasch, Renner, Haubrich, Hilbig, and Zettler (2013), but this paper was never published.

Acknowledgments

Due to the fact that the web-links from the original paper are not valid anymore, please contact the author via email if you are looking for the tests, a preview, or other information.

Direct correspondence to

Timo Heydasch, FernUniversität in Hagen, Universitätsstr. 33, 58084 Hagen

E-mail: timo.heydasch@fernuni-hagen.de

appropriate degree of differentiation, the tasks are relatively challenging to solve. Because of these problems, the development of a shorter and easier test of intelligence was necessary. This is why we developed the short version of the Hagen Matrices Test (HMT-S) as an additional free intelligence test that is based on the HMT.

Even if there is no consensus on what intelligence tests measure or even on what the definition of the term intelligence is (Wasserman, 2012; Willis, Dumont, & Kaufman, 2011), in 1994, a group of 52 experts (including Carroll, Cattell, Eysenck, Horn, Jensen, Thorndike, and Vernon) signed a definition that was accepted as mainstream in intelligence research to achieve a minimum level of convention:

> Intelligence is a very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly, and learn from experience. It is not merely book-learning, a narrow academic skill, or test-taking smarts. Rather, it reflects a broader and deeper capability for comprehending our surroundings – "catching on", "making sense" of things, or "figuring out" what to do. (Gottfredson, 1997, p. 13)

Sternberg (2005) added that intelligence is "the capacity to learn from experience, using metacognitive processes to enhance learning, and the ability to adapt to the surrounding environment, which may require different adaptations within different social and cultural contexts" (p. 751). Thus, intelligent individuals are able to learn and to adapt, and consequently they achieve success in the end.

Concerning a theory or a hierarchical structure of different components of cognitive abilities, no consensus exists here either (Gottfredson & Saklofske, 2009). Furthermore, diverse theories and models are present (see Davidson & Kemp, 2011). The Cattell-Horn-Carroll (CHC) model of intelligence (Schneider & McGrew, 2012; see also McGrew, 1997, 2005) is a taxonomy that allows for the theoretical classification and integration of theories and measurements such as the HMT-S. The CHC model is founded on the Horn-Cattell Gf-Gc Theory (Horn & Blankson, 2012; Horn & Noll, 1997) and the Three-Stratum-Theory (Carroll, 1993, 2005) and distinguishes between three levels of abilities. On top is *g*, general intelligence (see also Jensen, 1998; Spearman, 1904). The level below *g* contains 16 broad domains including *fluid reasoning* (Gf), *short-term memory* (Gsm), *processing speed* (Gs), *psychomotor speed* (Gps), *domain-specific knowledge* (Gkn), and *visual processing* (Gv). Below this level are different specific abilities that can be understood as facets of the broader abilities they belong to. Within Gf, there is strong evidence supporting the facets *induction* (I), *general sequential reasoning* (RG; also known as deduction; see Evans, 2005), and *quantitative reasoning* (RQ).

Because of the test material (i.e., figural matrices) and the theoretical considerations of the HMT (Heydasch, 2014), the HMT-S should measure induction, "the ability to observe a phenomenon and discover the underlying principles or rules…" (Schneider & McGrew, 2012, p. 112; see also Sloman & Lagnado, 2005). Therefore, the HMT-S is also a measure of Gf because "induction is probably the core aspect of Gf" (Schneider & McGrew, 2012, p. 112; see also Carroll, 1993). In addition, Gf and *g* are connected: Schneider and McGrew (2012) pointed out that Cattell understood *g* as Gf that has accumulated over a lifetime, whereas other researchers see Gf and *g* as one and the same.

In intelligence research alternative models were proposed which differentiate cognitive abilities and test tasks regarding the content (see Süß & Beauducel, 2005). One is the Radex model (Guttman, 1965; Guttman & Levy, 1991). Another is the structure-of-intellect model (Guilford, 1967), the Berlin Model of Intelligence Structure (BIS; Jäger, 1982), and the hierarchical frame and proto model of intelligence research (Amthauer, Brocke, Liepmann, & Beauducel, 2001; Liepmann, Beauducel, Brocke, & Amthauer, 2007). The common sense behind these models involves the differentiation of verbal, numerical, and figural abilities and tasks. By taking these models into account, it can be seen that figural matrices belong to figural abilities, and therefore, the HMT-S should be classified as a figural test of induction.

The validity of the HMT-S should remain despite the shortening of the original test, and it should be comparable to the validity of the HMT. The validity of a short form is at first defined as accordance with the original test (Silverstein, 1990). Because the HMT-S is a research instrument and was therefore not designed to be diagnostic of individuals (i.e., it is not intended to be used for selection nor to determine an individual's IQ or intellectual giftedness), its validity can be calculated as the correlation between the two test forms.

The correlation and the shared variance between the HMT-S and HMT do not permit conclusions to be drawn about the construct or criterion-related validity. Therefore, additional parameters are necessary: High validity coefficients should result in indications of success in school, at university, or at work and with tests that are broadly accepted as measures of reasoning. When measures have less theoretical overlap, their validity coefficients should also be lower. The overlap decreases when other domains (e.g., Gkn or Gsm vs. Gf), different levels of intelligence (e.g., *g* vs. I), or other content (verbal or numeric vs. figural) are involved. Additionally, different sources (self-report vs. objective tests; see Cattell, 1957; Pawlik, 2006) and different methods (web-based vs. paper and pencil; see Guttman, 1965; Guttman & Levy, 1991) should show lower correlations even when they measure theoretically similar constructs. Statistical independence should be demonstrated for constructs that are not theoretically associated.

Method

To develop and validate the short form, we used the same sample used to validate the original HMT (Heydasch, 2014). The sample was randomly divided (using SPSS version 21) into two groups. The analysis of the first group (Study 1) resulted in a selection of items on the basis of item difficulty ($p > .20$), item discrimination ($r_{it} > .30$), and their validity, that is, correlations with the Intelligence-Structure-Test 2000 R (Liepmann et al., 2007; $r > .30$). The results from Study 1 were cross-validated with the second group (Study 2). For further confirmation of the results, the final HMT-S was administered in Study 3.

Participants

In all three studies, participants were students in the Bachelor of Science Psychology program at the University of Hagen (Germany). They received course credit for participating. A total of $N = 1,572$ students (75% women) with an average age of $M = 31.6$ years ($SD = 8.97$) participated. Table 1 shows the sample sizes.

Table 1 Sample Sizes, Percentages of Women, and Age Distributions

			Age						
			Percentile						
Study	*N*	Women	10%	25%	50%	75%	90%	*M*	*SD*
1	681	74%	21	24	30	37	45	31.5	8.89
2	658	74%	22	25	30	38	45	31.8	8.95
3	233	80%	21	24	30	39	46	31.6	9.28

Measures

The HMT and the HMT-S are comprised of three parts: instructions, tasks (the matrices), and the ending. The instructions explain the tasks: Participants are asked to look for patterns in the incomplete 3 x 3 matrices[2], identify the underlying rules, and select the correct answer out of eight. The solution for each task has to be marked within a time limit of 2 min (but can also be marked and sent earlier). The underlying rules are listed in the instructions: addition, subtraction, as well

2 Matrices were kindly provided by Lutz Hornke.

as spatial movement (e.g., parts of the patterns move from one side to the other or rotate). The rules are demonstrated with two examples.[3] The second part contains 20 (HMT) or six (HMT-S) matrices. If a participant marks an answer within the 2-min limit but does not press "continue", the answer will still be recorded. The running time limit is displayed as a reference. The scoring is calculated automatically during participation. Correct solutions are coded as 1, incorrect or nonsolutions (even those belonging to an early withdrawal) are coded as 0. The sum is the final score. For each participant, the individual result is presented in the final third part of the of the HMT and the HMT-S, including the number and percentage of correct solutions. Additionally in language that is easy for laypersons to understand, the results are explained. For example, that the test is not perfectly reliable and that the test results are not deterministic. In contrast to the HMT, the HMT-S does not give feedback on the test-taker's IQ.

For further validation, the following measures were used.

Intelligence-Structure-Test 2000 R (I-S-T 2000 R; Liepmann et al., 2007; see also Schmidt-Atzert, 2002; Schmidt-Atzert & Rauch, 2008). The I-S-T 2000 R is an objective performance test comprised of measures of reasoning (R), knowledge (K), and memory (M). R and K are comprised of scales of verbal, numeric, and figural content, which also contain three different types of tasks. M contains figural and verbal content. Furthermore, two independent factors representing fluid intelligence (*gf*) and crystallized intelligence (*gc*) can be calculated. Theoretically, the I-S-T 2000 R refers to the Radex model (Guttman, 1965; Guttman & Levy, 1991) and the BIS (Jäger, 1982), both of which distinguish between verbal, numeric, and figural content. At the same time, the authors refer to Cattell's (1987) R and K dimensions. The two approaches were combined by Amthauer et al. (2001) into the hierarchical frame and proto model of intelligence research, which is the fundament of the I-S-T 2000 R.

Inventory of Self-Estimated Intelligence (ISI; Rammstedt & Rammsayer, 2002). An online version of the ISI was used to measure self-estimated intelligence in accordance with the theory of multiple intelligences (Gardner, 1983): verbal comprehension, word fluency, mathematical intelligence, spatial intelligence, memory, perceptual speed, reasoning, musical intelligence, bodily-kinesthetic intelligence, interpersonal intelligence, and intrapersonal intelligence.

General self-efficiency (GSE; Schwarzer & Jerusalem, 1995). This scale measures cross-situational beliefs in handling and mastering problems and challenges.

Study-specific self-efficacy (Schiefele, Moschner, & Husstegge, 2002). In accordance with general self-efficacy scales, this questionnaire measures (adapted to the specific context of a distant educational university) students' beliefs in the

3 A preview of the HMT including the instructions, examples, and ending may be viewed online. Please email the author for an access.

extent to which they can handle and master the problems and challenges involved in studying.

General helplessness (Jerusalem & Schwarzer, 1986, 2012). In accordance with the theory of learned helplessness (Seligman, 1975), this scale measures self-estimated cross-situational helplessness due to the handling and mastery of problems and challenges.

Study-specific helplessness (Jerusalem & Schwarzer, 1986, 2012). This questionnaire measures students' expectations of the extent to which they are not able to master the problems encountered in studying.

Measures of self-concepts. General academic, mathematical, and verbal self-concept (Schiefele et al., 2002) capture the self-concepts of abilities and problem-solving competences in domains listened above.

Revised Achievement Motivation Scale (AMS-R; Lang & Fries, 2006). The AMS-R measures explicit achievement motivation using the dimensions hope of success and fear of failure, which represent approach and avoidance tendencies, respectively.

In addition, demographic variables (sex and age), the nation from which the student's school-leaving qualification was obtained, and the school-leaving qualification itself (*Allgemeine Hochschulreife* = 3, *Fachhochschulreife oder fachgebundene Hochschulreife* = 2, *Mittlerer Schulabschluss* = 1) were measured.[4] Furthermore, participants reported their high school GPA and their final grades in mathematics, English, German, biology, art, and sports. The grades reported from the Bachelor of Science in Psychology program were standardized for each course in order to take course differences into account. Afterwards, the mean was computed so that individually achieved marks were comparable even though the courses differed in difficulty. At least one grade had to be reported for this measure.

Procedure

Data collection was conducted primarily online and unproctored by EFS-Survey by QuestBack GmbH (see Buchwald, Spoden, Fleischer, & Leutner, 2013). Participants chose location (e.g., PC at home or mobile device) as well as the date of their participation. Information about the study was presented on a website (https://www.fernuni-hagen.de/psychologie/forschung/virtuelles-labor.shtml),[5] where alternative

4 "Allgemeine Hochschulreife" is the highest school-leaving qualification, which allows admission to universities; "Fachhochschulreife oder fachgebundene Hochschulreife" is a lower form, which allows admission to universities only under certain conditions; "Mittlerer Schulabschluss" is even a lower qualification, which usually does not permit university admission

5 This page has moved. The original but outdated link was (http://www.fernuni-hagen.de/psychologie/forschung/vlabor.shtml).

studies were also presented. After participants began taking the survey, additional terms of their participation were specified, and informed consent was requested.

In surveys containing the HMT or the HMT-S, basic demographic information (e.g., age and sex) was requested first, and then participants were surveyed about their self-estimated intelligence. Afterwards, the HMT or HMT-S was presented, and finally, individual test results were displayed. Because of the huge number and sometimes enormous length of some instruments[6] measures of self-efficacy, helplessness, self-concept, and achievement motivation were placed in independent online surveys rather than being presented in the same survey. Despite the length, in Study 3 these scales were integrated into the same survey between the demographic variables and the ISI.

In contrast to the online surveys, the I-S-T 2000 R was administered as a paper-and-pencil test. Participants registered for a specific date and location. Subsequently, the I-S-T 2000 R was administered and proctored in accordance with the instructions given in the manual.

At the beginning of every session, each participant had to enter a fixed 6-digit code. Therefore, every participant had a unique fixed code. These codes were used to match participants across different sessions without knowing their identities (e.g., name or university registration number). Multiple participations in a specific session could also be eliminated by using the code. Finally, on the basis of school information, the data from students who did not finish school in Germany were eliminated because their educational information might not be comparable.

Data analyses were conducted with SPSS (version 21) after the I-S-T 2000 R data were transferred. All further recoding, computing of scales, and analyses were afterwards made in SPSS.

Results

The first step was to select six matrices from the HMT to be used in the short version on the basis of their item characteristics: item difficulty, item discrimination, and correlation with the I-S-T 2000 R (see Table 2). Items 1 to 3 were selected because no exclusion criteria were met by these items (see above). By contrast, Items 4, 6, and 8 were excluded because they had low correlations with the I-S-T 2000 R ($r = .13$, $r = .05$, and $r = .08$, respectively). Moreover, Item 8 had a strikingly low discrimination value ($r_{it} = .22$). Items 5, 7, and 9 were additionally selected as unproblematic items, maintaining the order of the items. At this point, item selection was stopped because (a) an internal consistency of .64 was achieved, (b) easy,

6 The surveys included other instruments too, but these were used in other studies and are not described in this paper.

Table 2 Properties and Correlations of the HMT Items with the I-S-T 2000 R in Study 1

Item	M_t[a]	p[b]	r_{it}[b]	r_{IST}[c]
1	43	.88	.33	.21
2	30	.85	.35	.26
3	53	**.66**	**.40**	**.36**
4	57	.66	.39	.13
5	**51**	**.65**	**.45**	**.27**
6	76	.54	.30	.05
7	**53**	**.58**	**.42**	**.33**
8	82	.36	.22	.08
9	**76**	**.25**	**.35**	**.44**
10	68	.29	.43	.35
11	81	.25	.29	.03
12	66	.30	.30	.22
13	66	.20	.37	.31
14	78	.15	.24	.41
15	78	.16	.52	.16
16	60	.14	.40	.03
17	67	.18	.28	.36
18	61	.15	.51	.23
19	52	.12	.29	.13
20	41	.11	.44	.07

Note. I-S-T 2000 R = Intelligence-Structure-Test 2000 R (Liepmann et al., 2007). HMT = Hagen Matrices Test (Heydasch, 2014). M_t = Mean duration in seconds. p = difficulty. r_{it} = item discrimination. r_{IST} = correlations with the I-S-T 2000 R reasoning score. Bold items were selected for the short version of the HMT. Underlined values resulted in the exclusion of that item from the short version ($p \leq .20$, $r_{it} \leq .30$ or $r_{IST} \leq .20$).
[a]$N_1 = 606$. [b]$N_1 = 681$. [c]$N_1 = 56$.

moderate, as well as more difficult tasks were included, and (c) the mean duration of completing the test would go beyond 10 min if more items had been chosen.

The complete duration of the HMT-S (i.e., reading the instructions and solving the tasks) was 9.4 min (see Table 3). Table 4 presents the solution times for each task. Additionally, Table 4 shows the item difficulty and discrimination values. The values were nearly the same in the three studies. The first two items were solved quite quickly (with means of $28 \leq M_t \leq 44$ s), Items 3 to 5 had means of 51 to 54 s, and the sixth item took relatively longer (with a mean of $76 \leq M_t \leq 77$ s). Concerning item difficulty, the first two items were very easy ($p \geq .83$), Items 3 to 5 were moderately difficult ($.53 \leq p \leq .70$), and the last item was rather hard to solve

Table 3 Test Duration (in Minutes)

	Percentile						
	10%	25%	50%	75%	90%	M_t	SD_t
Instructions	1.8	2.4	3.3	4.6	6.1	4.5	9.07
Tasks	2.7	3.7	4.9	3.2	7.7	5.1	1.97
Total	5.0	6.4	8.3	10.8	13.5	9.4	8.88

Note. Studies 1, 2, and 3 were aggregated because there were no substantial differences. M_t = Mean duration. SD_t = Standard deviation of duration. N = 1,563.

Table 4 Mean Duration, Difficulty, and Item Discrimination

	Study								
	M_t			p			r_{it}		
Item	1[a]	2[b]	3[c]	1[d]	2[e]	3[f]	1[d]	2[e]	3[f]
1	43	44	39	.88	.88	.88	.38	.32	.34
2	30	32	28	.85	.83	.87	.42	.42	.35
3	53	53	52	.66	.65	.67	.39	.38	.25
4	51	53	51	.65	.64	.70	.47	.46	.39
5	53	54	53	.58	.53	.57	.40	.31	.29
6	76	76	77	.25	.25	.35	.23	.20	.21

Note. M_t = Mean duration in minutes. p = difficulty. r_{it} = item discrimination.
[a]N_1 = 636. [b]N_2 = 633. [c]N_3 = 229. [d]N_1 = 681. [e]N_2 = 658. [f]N_3 = 233.

($p \leq .35$). The item discriminations for Items 1 to 5 ranged from $r_{it} = .29$ to $r_{it} = .47$. The sixth and last item had a lower value, which was $r_{it} = .23$ at best.

Most frequently, four out of six tasks were solved correctly. Due to the low difficulty levels of the items, the distributions in all three studies were left skewed. The kurtosis levels were different: The distributions in Studies 1 and 2 were flat, whereas the distribution in Study 3 was normal. Table 5 also presents internal consistency values computed with the Kuder-Richardson Formula 20 (KR20; Kuder & Richardson, 1937). These were $KR20_1 = .64$, $KR20_2 = .61$, and $KR20_3 = .57$. The weighted mean was KR20 = .62 (N = 1,572).

There were sex differences (see Table 6), but the effect sizes were quite small ($d_1 = 0.16$, $d_2 = 0.25$, and $d_3 = 0.15$; $p_1 = .058$, $p_2 = .004$, and $p_3 = .354$).

Table 5 Scale Properties

Study	*N*	*M*	*SD*	Skewness	Kurtosis	KR20
1	681	3.88	1.56	-0.65	-0.31	.64
2	658	3.78	1.53	-0.55	-0.41	.61
3	233	4.03	1.47	-0.76	0.05	.57

Note. KR20 = Internal consistency calculated with the Kuder-Richardson Formula 20 (Kuder & Richardson, 1937).

Table 6 Sex Differences

Study		*N*	*M*	*SD*	*t*	*df*	*d*
1	Men	174	4.08	1.64	1.90	676	0.16
	Women	504	3.82	1.51			
2	Men	173	4.06	1.46	2.89**	316.82	0.25
	Women	483	3.69	1.53			
3	Men	47	4.21	1.53	0.93	231	0.15
	Women	186	3.99	1.46			

** $p < .01$.

The association between the HMT-S test result and age was slightly negative: Generally speaking, older participants solved fewer matrices correctly. But this connection was quite low in the first two studies and was not verified in Study 3 (r_1 = -.13, r_2 = -.12, r_3 = -.02; p_1 = .001, p_2 = .003, p_3 = .768; N_1 = 679, N_2 = 654, N_3 = 233).

The correlations between the HMT-S and HMT were r_1 = .79 ($p > .001$) in Study 1 and r_2 = .78 ($p < .001$) in Study 2. To determine what to attribute the shared variance to, we computed a hierarchical regression. As the dependent variable and a marker for intelligence, we used the I-S-T 2000 R reasoning scale (R). We entered the HMT score in the first step and the HMT score in the second step. Table 7 shows the results. Because only a small proportion of variance was explained by the HMT over and above the HMT-S (ΔR^2 = .04 in Study 1 and ΔR^2 = .13 in Study 2) compared with the variance explained by the HMT-S (R^2 = .28 in Study 1 and R^2 = .24 in Study 2), we were able to conclude that the shared variance was not due to error variance (e.g., methodological variance) but was instead due to reasoning ability.

Table 7 Hierarchical Regression and the Incremental Validity of the HMT compared with the HMT-S

	Study 1[a]				Study 2[b]			
	R	ΔR^2	*F*	β	*R*	ΔR^2	*F*	β
Step 1	.530	.281	21.1***		.494	.244	10.7**	
HMT-S				.530				.494
Step 2	.567	.041	3.2		.607	.125	6.3*	
HMT-S				.215				.072
HMT				.374				.551*

Note. AV: Reasoning score from the Intelligence-Structure-Test 2000 R (Liepmann et al., 2007). HMT = Hagen Matrices Test (Heydasch, 2014). HMT-S = Short version of the Hagen Matrices Test.
[a]$N_1 = 56$. [b]$N_2 = 35$.
* $p < .05$. ** $p < .01$. *** $p < .001$.

To determine validity in a manner that goes beyond a comparison between the HMT and HMT-S, correlations with the I-S-T 2000 R were computed (see Table 8). Because some correlation coefficients (especially with the verbal measures) were quite different, the data from Studies 1 and 2 were pooled, and the correlations were computed again. Thus, the HMT-S was correlated with reasoning at $r = .52$ and with the g_f factor at $r = .49$. Concerning the different contents, the correlations with numeric ($r = .46$) and figural ($r = .47$) reasoning were just minimally smaller compared with the one for reasoning. The correlation with verbal reasoning was moderate ($r = .30$), and so were the correlations with K ($r = .31$) and ME ($r = .24$), whereby the findings according to the content were quite comparable to the reasoning domain: The correlations with the numeric and figural scales were moderate and nearly the same ($.27 \leq r \leq .35$), and the verbal content had small correlations ($r = .13$ with K and $r = .12$ with ME). Looking deeper, it is possible to see that the correlations varied across the different types of tasks within a specific content area. This was true for the verbal and numerical types of tasks, but it was especially true for the figural types of tasks such as figure selection (two-dimensional visual discrimination; $r = .28$), dice (spatial rotation; $r = .50$), and figural matrices ($r = .22$).

The connections to the ISI (see Table 9) were nearly the same across all three studies. The HMT was positively correlated with the rather theoretically close constructs: mathematical intelligence, numerical intelligence, and reasoning. There were no additional correlations except slightly negative ones with bodily-kinesthetic intelligence, interpersonal intelligence, and intrapersonal intelligence. Medium-level correlations resulted between the HMT-S and beliefs about one's

Table 8 Convergent Validity with the I-S-T 2000 R

Domain	Content	Tasks	Study 1[a]	Study 2[b]	Study 1+2[c]
Reasoning	Total		.53***	.49**	.52***
	g_f		.50***	.47**	.49***
	Verbal		.46***	.03	.30**
		Sentence Completion	.24	.08	.16
		Analogies	.34*	-.14	.18
		Commonalities	.44***	.13	.33**
	Numeric		.42**	.53**	.46***
		Math	.31*	.42*	.35***
		Numerical Series	.37**	.42*	.39***
		Arithmetic Operators	.40**	.54***	.45***
	Figural		.47***	.46**	.47***
		Figure Selection	.25	.31	.28**
		Dice	.50***	.51**	.50***
		Matrices	.21	.24	.22*
Knowledge	Total		.24	.42*	.31**
	g_c		.15	.36*	.23*
	Verbal		.06	.24	.13
	Numeric		.26*	.36*	.30**
	Figural		.29*	.48**	.35***
Memory	Total		.36**	.07	.24*
	Verbal		.24	-.04	.12
	Figural		.34**	.16	.27*

Note. All values are Pearson correlations. I-S-T 2000 R = Intelligence-Structure-Test 2000 R (Liepmann et al., 2007); g_f = fluid intelligence factor; g_c = crystallized intelligence factor.
[a]$N_1 = 56$. [b]$N_2 = 35$. [c]$N_{12} = N_1 + N_2 = 91$.
* $p < .05$. ** $p < .01$. *** $p < .001$.

own abilities in the area of mathematical self-concept. This connection was replicated in Studies 2 and 3. At least a very small correlation was found with academic self-concept as well as a small negative one with study-specific helplessness. Looking at the dimensions of the AMS-R, fear of failure was independent, and hope for success had a tendency to be slightly positively connected.

Table 9 Divergent Validity

		Study		
		1	2	3
ISI[a]				
	Verbal Comprehension	.02	-.07	-.05
	Word Fluency	-.03	-.10*	-.04
	Mathematical Intelligence	.23***	.24***	.20**
	Spatial Intelligence	.22***	.20***	.22***
	Memory	-.04	-.03	-.01
	Perceptual Speed	.05	-.01	-.07
	Reasoning	.17***	.18***	.14*
	Musical Intelligence	-.02	-.04	-.06
	Bodily-Kinesthetic Intelligence	-.00	-.09*	-.13*
	Interpersonal Intelligence	-.04	-.16***	-.13
	Intrapersonal Intelligence	-.02	-.14***	-.05
Self-efficacy				
	General[b]	.09	.07	.08
	Study-specific[c]	.18*	.14	.01
Helplessness				
	General[d]	.14*	.00	-.02
	Study-specific[c]	-.21**	-.15*	-.10
Self-concept				
	Academic	.23***	.10	.10
	Mathematical	.28***	.27***	.32***
	Verbal	.10	-.07	-.00
AMS-R[f]				
	Hope of Success	.21***	.09	.12
	Fear of Failure	-.12	.04	-.06

Note. All values are Pearson correlations. ISI = Inventory of Self-Estimated Intelligence (Rammstedt & Rammsayer, 2002); AMS-R = Revised Achievement Motivation Scale (Lang & Fries, 2006).

[a]$N_1 = 675$, $N_2 = 654$, $N_3 = 225$. [b]$N_1 = 286$, $N_2 = 262$, $N_3 = 155$. [c]$N_1 = 176$, $N_2 = 177$, $N_3 = 155$. [d]$N_1 = 266$, $N_2 = 242$, $N_3 = 155$. [e]$N_1 = 345$, $N_2 = 317$, $N_3 = 155$. [f]$N_1 = 244$, $N_2 = 242$, $N_3 = 155$.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Associations of the HMT-S with indicators of academic success are presented in Table 10. Across all three studies, there were positive correlations between the HMT-S with school GPA as well as with students' final grade in mathematics. The results regarding the school-leaving qualification and university GPA were inconsistent, but on average, higher test scores were accompanied by more success. Only in one of the three studies were there substantial results regarding grades in biology in high school and statistics at a university. There were no correlations with English, German, art, or sports.

Discussion

The HMT-S is a flexible measure that can be implemented in any online survey (relevant technical restrictions are not known to us) and can be administered at any location, in the field, as well as in a lab setting as long as a PC or a laptop with an Internet connection is available. The HMT-S is appropriate not only in psychologi-

Table 10 Criterion-Related Validity

	Study		
	1	2	3
School Education[a]			
School-Leaving Qualification[b]	.16**	.10	.00
GPA	.14*	.16**	.16*
Mathematics	.19***	.20***	.19**
English	.03	.03	.09
German	-.01	-.02	.01
Biology	.07	.07	.16*
Art	.01	.04	.09
Sports	.03	.09	.03
University Education			
GPA[c]	.28**	.18*	.06
Grade in statistics[d]	.22	.38**	.05

Note. All values except for the one for the school-leaving qualification (Spearman correlation) are Pearson correlations. Grades were recoded: Positive correlations indicate better grades in the sense of higher test scores.
[a]$N_1 = 308$, $N_2 = 287$, $N_3 = 191$. [b]*Allgemeine Hochschulreife* = 3, *Fachhochschulreife oder fachgebundene Hochschulreife* = 2, *Mittlerer Schulabschluss* = 1. [c]$N_1 = 131$, $N_2 = 124$. $N_3 = 78$. [d]$N_1 = 70$, $N_2 = 70$, $N_3 = 67$.
* $p < .05$. ** $p < .01$. *** $p < .001$.

cal, social, and educational science but also in practical applications. Administration is not limited to psychologists. In addition, other researchers who meet the scientific standards are invited to use the test. Only implementations of individual diagnostics such as personnel selection are excludet.[7] In these cases, other measures that meet crucial standards (e.g., the DIN 33430 in Germany; DIN, 2002) need to be used.

The difficulty of the HMT-S is moderate, and therefore, the test can be applied to samples that are not characterized by very high or very low mental abilities. Nevertheless, the sample may be quite heterogeneous because, due to the different item difficulties, ceiling or floor effects can be avoided: The level of difficulty increases from task to task and, at the end, it reaches a high level such that a distinction between low and high ability levels is possible (within the limits of a measure with six dichotomous items). However, the results of Studies 1 to 3 have to be interpreted with caution because the samples were students who can be assumed to have a higher than average ability level. The difference should not be extraordinary, however, because in contrast to other universities, these students were not selected by a *numerus clausus*.

The HMT-S is extremely economical. The duration of about five minutes is only 10% to 25% of the length of common tests, which often last 20 to 45 min. The time limit of 2 min for each matrix seems to be enough in most cases to mark the solution and send it. Only for the last and most difficult matrix will the limit sometimes be reached. But on the other hand, individual participants often exploit the full 12 min.

There are effects of sex and age, but only small ones. Thus, the HMT-S seems to be quite fair with respect to sex and age, but a final evaluation cannot be made at this point because of inconsistent findings in the three studies.

Objectivity is given due to scoring as well as to interpretation. The scoring is done automatically online, the interpretation of the results are web-based and standardized, and individual scores are presented to each participant after the matrices are completed. There might be a lack of objectivity with respect to participants' free choice of taking part online at any time and any place (Gnabs, Batinic, & Hertel, 2011). The instructions, layout, and time limits are the same because of the web-programming. But if the test is administered in the field instead of in a controlled setting (e.g., a laboratory setting), certain uncontrollable factors might interfere with participation or might interrupt participants' concentration, and as

7 In the meantime, additional conditions of use are specified: (a) studies must be strictly academic without commercial interests (e.g., participants do not need to pay to participate; institutes or organizations are not allowed to sell information, products, or study results), (b) studies must run for a limited amount of time and involve a limited number of participants, (c) the test materials (e.g., pictures, graphics, texts, data files) must remain confidential, and finally, (d) the HMT and the HMT-S, respectively, will be cited according to scientific standards.

a consequence, test conditions might be unequal. On the other hand, unproctored administrations lead to higher ecological validity, which cannot be achieved in an artificial setting in a laboratory with supervision present.

According to Aiken and Marnat (2006), who accept .60 as a sufficient level of reliability, the reliability of the HMT-S can be considered reasonable with its mean internal consistency of KR20 = .62.

The validity of the HMT-S was demonstrated by its very high correlation with the long form of the HMT. Additionally, the regressions verified that the variance shared between the short and the long forms was mostly variance that could be explained by reasoning ability as measured by the I-S-T 2000 R. Therefore, the correspondence was due to differences in intelligence rather than to methodological artefacts that may emerge in unproctored web-implementations. Regressions also indicated the superiority of the long form (the HMT, which contains 20 tasks, is three times longer than the short form), but the variance of the I-S-T 2000 R was explained primarily by the HMT-S.

The associations between the HMT-S and the I-S-T 2000 R were quite different between Studies 1 and 2. It is important to take into consideration the fact that the sample sizes were relatively small ($N_1 = 56$ and $N_2 = 35$) and that fluctuations could be traced not only to imperfect reliabilities but also to sampling errors.[8] Despite the small samples, the correlations between the HMT-S and the IS-T 2000 R support the quality of the short form and qualify the HMT-S as a reasoning test. The reasoning scale from the I-S-T 2000 R reflects *induction* (I) from the CHC model of intelligence, and the numeric reasoning scale is comparable to *quantitative reasoning* (RQ). High correlations with these two domains and lower ones with domains such as knowledge and memory provide further evidence that the HMT-S is an operationalization of Gf, however, neglecting the facets of deduction and *general sequential reasoning* (RG; see Schneider & McGrew; 2012). The pattern of results reveals a possible characterization of the HMT-S as a *numerical-figural* induction test: Consequently, the associations were higher with the numerical and figural scales in contrast to the scales involving verbal content. However, Schneider and McGrew did not associate Gf with specific content, and similarly, the descriptions of *induction* (I) or *quantitative reasoning* (RQ) are not bound to figural content (even though this is not excluded). In fact, one of the 16 broad abilities is called *visual processing* (Gv), which represents the ability "…to make use of simulated mental imagery (often in conjunction with currently perceived images) to solve problems" (Schneider & McGrew, 2012, p. 129). The correlation with the dice task from the I-S-T 2000 R was strikingly high. These tasks, in which a given die has to be compared with other rotated dice, involves *visualization* (Vz), a core element of Gv: "The ability to perceive complex patterns and mentally simulate

8 Alternative explanations such as extreme values, coding errors, and incorrect scoring were explored.

how they might look when transformed (e.g., rotated, changed in size, partially obscured)" (Schneider & McGrew, 2012, p. 129).

In this context, it is remarkable that the association between the HMT-S and the figural reasoning scale is based on the dice tasks and not due to the matrices from the I-S-T 2000 R. The correlation of $r = .22$ is relatively small and indicates that the associations do not occur because the tasks are superficially the same, that is, involving figural matrices. The reason might be different principles: In contrast to the HMT-S, not all the cognitive operations are explained or introduced with examples. Additionally, there are many more operations involved in the I-S-T 2000 R matrices. We believe these matrices involve other mental abilities (e.g., divergent thinking with respect to creativity as proposed by Jäger, 1982) that are not relevant to the HMT-S and therefore account for the low correlation. Besides the theoretical classification of the I-S-T 2000 R as an objective performance test, future investigations should include other multidimensional tests such as the Berlin Intelligence Structure Tests (Jäger, Süß, & Beauducel, 1997) or the Wilde Intelligence Tests 2 (Kersting, Alsthoff, & Jäger, 2008).

The results for the ISI self-rated intelligences underline the classification of the HMT-S as a figural-numerical reasoning test: positive correlations resulted in the areas of mathematical intelligence, spatial intelligence, and reasoning. The tendency toward slightly negative correlations with bodily-kinesthetic intelligence, interpersonal intelligence, and intrapersonal intelligence might be due to an awareness of intraindividual contrasts. In conformance with our expectations associations with self-reported mathematical and academic self-efficacies, with problem-solving experiences (i.e., self-concepts), and with helplessness exist. Achievement motivation seems to be slightly connected. This is not desirable (Stern, 1911) but is difficult to avoid for intelligence tests (Conrad, 1983).

The criterion-related validity of the HMT-S was demonstrated by its associations with academic performance. Connections emerged with subjects that imply logical thinking and reasoning (i.e., mathematics as well as methods and statistics). Small effects resulted for more general indicators of academic success such as the school-leaving qualification and high school and university GPA. This might be explained by the aggregation of grades from different subjects that might require less reasoning (e.g., sports and art). For school grades, it is important to consider that (a) the period between taking the HMT-S and school was about 10 years for the (on average) 30-year-old students who probably did not correctly remember their grades in every class and (b) there is not absolute stability in intelligence. Because the samples in all three studies were not representative and were rather homogenous, further studies should include students who are able to obtain grades from school records rather than as self-reports. Additionally, students who are studying other subjects at universities might provide samples with a more balanced gender ratio.

Our conclusion about the HMT-S is generally positive despite the facts that the studies were restricted to homogenous samples and that the sample sizes for the correlations with the I-S-T 2000 R were rather small. The HMT-S is a highly economical, free, web-based intelligence test for measuring figural-numeric reasoning. Because of the high correlation with the original long version and also because of the pattern of results, which is quite comparable to the HMT, the HMT-S and the HMT can be concluded to measure the same construct. Despite containing only six tasks, the reliability of the HMT-S is acceptable. All of the evidence across the three studies pointed to construct and criterion-related validity. Thus, we hope the HMT-S will gain currency. Information and free applications can be requested from the author via email.

References

Aiken, L. R., & Groth-Marnat, G. (2006). *Psychological testing and assessment (12th ed.).* Boston, MA: Pearson.

Amthauer, R., Brocke, B., Liepmann, D., & Beauducel, A. (2001). *I-S-T 2000 R. Intelligenz-Struktur-Test 2000 R [I-S-T 2000 R. Intelligence Structure Test 2000 R].* Göttingen, Germany: Hogrefe.

Blickle, G., Kramer, J., & Mierke, J. (2010). Telephone-administered intelligence testing for research in work and organizational psychology: A comparative assessment study. *European Journal of Psychological Assessment*, *26*, 154-161. doi:10.1027/1015-5759/a000022

Buchwald, F., Spoden, C., Fleischer, J., & Leutner, D. (2013). Verzweigte Lernumgebungen und Tests mit EFS Survey 8 [Branched learning environments and tests with EFS Survey 8]. *Diagnostica*, *59*, 113-117. doi:10.1026/0012-1924/a000080

Conrad, W. (1983). Intelligenzdiagnostik [Intelligence diagnostics]. In K. Groffmann & L. Michel (Eds.), *Intelligenz- und Leistungsdiagnostik* (pp. 104-201). Göttingen, Germany: Hogrefe.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factoranalytic studies.* Cambridge, MA: University Press.

Carroll, J. B. (2005). The three-stratum theory of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, test, and issues* (2nd ed., pp. 69-76). New York, NY: Guilford.

Cattell, R. B. (1957). *Personality and motivation structure and measurement.* Oxford: World Book Co.

Cattell, R. B. (1987). *Intelligence: Its structure, growth, and action.* Amsterdam, Netherlands: Elsevier.

Davidson, J. E., & Kemp, I. A. (2011). Contemporary Models. In R. J. Sternberg & S. B. Kaufman (Eds.), *The Cambridge handbook of intelligence* (pp. 58-82). Cambridge, NY: University Press. doi:10.1017/CBO9780511977244.005

Deutsche Gesellschaft für Psychologie & Bundesverband Deutscher Psychologinnen und Psychologen (2004). *Ethische Richtlinien der DGPs und des BDP [Ethical guidelines of the DGPs and the BDP].* Retrieved from https://www.dgps.de/index.php?id=85

DIN. (2002). *DIN 33430: Anforderungen an Verfahren und deren Einsatz bei berufsbezogenen Eignungsbeurteilungen [DIN 33430: Requirements for proficiency assessment procedures and their implementation]*. Berlin, Germany: Beuth.

Evans, J. St. B. T. (2005). Deductive reasoning. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp.169-184). Cambridge, NY: University Press.

Gardner, G. (1983). *Frames of mind: The Theory of Multiple Intelligences*. New York, NY: Basic Books.

Gnambs, T., Batinic, B., & Hertel, G. (2011). Internetbasierte psychologische Diagnostik [Web-based psychological diagnostics]. In L. F. Hornke, M. Amelang, & M. Kersting (Eds.), *Verfahren zur Leistungs-, Intelligenz- und Verhaltensdiagnostik* (pp. 448-498). Göttingen, Germany: Hogrefe.

Gottfredson, L. S. (1997). Mainstream science on intelligence: An editorial with 52 signatories, history and bibliography. *Intelligence*, *24*, 13-23. doi:10.1016/S0160-2896(97)90011-8

Gottfredson, L. S., & Saklofske, D. H. (2009). Intelligence: Foundation and issues in assessment. *Canadian Psychology*, *50*, 183–195. doi:10.1037/a0016641

Guilford, J.P. (1967). *The nature of human intelligence*. New York, NY: McGraw-Hill.

Guttman, L., & Levy, S. (1991). Two structural laws for intelligence tests. *Intelligence, 15,* 79-103. doi:10.1016/0160-2896(91)90023-7

Guttman, L. (1965). A faceted definition of intelligence. In R. Eiferman (Ed.), *Studies in psychology, scripta hierosolymitana* (Vol. 14, pp. 166-181). Jerusalem, Israel: Hebrew University.

Harrell, T. W., & Harrell, M. S. (1945). Army General Classification Test scores for civilian occupations. *Educational and Psychological Measurement*, *5*, 229-239. doi:10.1177/001316444500500303

Heydasch, T. (2014). *The Hagen Matrices Test (HMT).* (Part of doctoral dissertation thesis). University of Hagen, Germany. doi:10.13140/RG.2.2.31433.75361

Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: Gf-Gc theory. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (pp. 53-91). New York, NJ: Guilford

Horn, J. L., & Blankson, A. N. (2012). Foundations for better understanding of cognitive abilities. In D. P. Flanagan und P. L. Harrison (Eds.): *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 73-98). New York, NY: Guilford Press.

Hülsheger, U. R., Maier, G. W., & Stumpp, T. (2007). Validity of General Mental Ability for the Prediction of Job Performance and Training Success in Germany: A meta-analysis. International Journal of Selection and Assessment, *15,* 3-18. doi:10.1111/j.1468-2389.2007.00363.x

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, *96*, 72-98. doi:10.1037/0033-2909.96.1.72

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

Jäger, A. O., Süß, H.-M., & Beauducel, A. (1997). Berliner Intelligenzstruktur - Test. Form 4 [Berlin Test of Intelligence Structure. Form 4]. Göttingen, Germany: Hogrefe

Jensen, A. (1998) *The g factor: The science of mental ability*. Westport, CT: Praeger.

Jerusalem, M., & Schwarzer, R. (1986). Fragebogen zur Erfassung von Hilflosigkeit [Helplessness assessment questionnaire]. In R. Schwarzer (Ed.), *Skalen zur Befindlichkeit und Persönlichkeit (Forschungsbericht 5)*. Berlin, Germany: FU.

Jerusalem, M., & Schwarzer, R. (2012). Dimensionen der Hilflosigkeit [Dimensions of helplessness]. In A. Glöckner-Rist (Ed.), *Zusammenstellung sozialwissenschaftlicher Items und Skalen* [Compilation of social science items and scales]. Bonn, Germany: GESIS.

Judge, T. A., Higgins, C. A., Thoresen, C. J., & Barrick, M. R. (1999). The Big Five personality traits, general mental ability, and career success across the life span. *Personnel Psychology*, *52*, 621-652. doi:10.1111/j.1744-6570.1999.tb00174.x

Kersting, M., Althoff, K., & Jäger, A. O. (2008). WIT-2. Der Wilde-Intelligenztest [WIT-2. The Wilde-Intelligence Test]. Göttingen, Germany: Hogrefe.

Khandaker, G. M., Barnett, J. H., White, I. R., & Jones, P. B. (2011). A quantitative meta-analysis of population-based studies of premorbid intelligence and schizophrenia. *Schizophrenia Research*, *132*, 220-227. doi:10.1016/j.schres.2011.06.017

Kuder, G. F., & Richardson, M.W. (1937). The theory of the estimation of test reliability. *Psychometrika*, *2*, 151-160. doi:10.1007/BF02288391

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2004). Academic Performance, Career Potential, Creativity, and Job Performance: Can One Construct Predict Them All? *Journal of Personality and Social Psychology*, *86*, 148-161. doi:10.1037/0022-3514.86.1.148

Kunina, O., Wilhelm, O., Formazin, M., Jonkmann, K., & Schroeders, U. (2007). Extended criteria and predictors in college admission: Exploring the structure of study success and investigating the validity of domain knowledge. *Psychology Science*, *49*, 88-114.

Lang, J. B., & Fries, S. (2006). A revised 10-item version of the Achievement Motives Scale: Psychometric properties in german-speaking samples. *European Journal of Psychological Assessment*, *22*, 216-224. doi:10.1027/1015-5759.22.3.216

Leibniz-Zentrum für Psychologische Information und Dokumentation (2012). Verzeichnis Testverfahren. Kurznamen. Langnamen. Autoren. Testrezensionen [Test Directory. Short names. Long names. Authors. Test reviews] (19th ed.). Trier, Germany: ZPID.

Liepmann, D., Beauducel, A., Brocke, B. & Amthauer, R. (2007). *Intelligenz-Struktur-Test 2000 R (I-S-T 2000 R)* [*Intelligence Structure Test 2000 R (I-S-T 2000 R)*]. Göttingen, Germany: Hogrefe.

McGrew, K. S. (1997). Analysis of the major intelligence batteries according to a proposed comprehensive Gf-Gc framework. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 151-179). New York, NY: Guilford.

McGrew, K. S., (2005). The Cattell-Horn-Carroll theory of cognitive abilities: Past, present, and future. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 136-181). New York, NY: Guilford Press.

Ng, T. W. H., Eby, L. T., Sorensen, K. L., & Feldman, D. C. (2005). Predictors of Objective and Subjective Career Success: A Meta-Analysis. *Personnel Psychology*, *58*, 367-408. doi:10.1111/j.1744-6570.2005.00515.x

Pawlik, K., (2006). Psychologische Diagnostik I: Methodische Grundlagen [Psychological diagnostics I: Methodological foundations]. In K. Pawlik (Ed.), *Handbuch Psychologie. Wissenschaft – Anwendung - Berufsfelder* (pp. 555-562.) Berlin: Springer.

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin, 135*, 322-338. doi:10.1037/a0014996

Rammstedt, B. & Rammsayer, T. (2002). Die Erfassung der selbsteingeschätzten Intelligenz: Konstruktion, teststatistische Überprüfung und erste Ergebnisse des Inventars zur selbsteingeschätzten Intelligenz (ISI) [Assessment of self-estimated intelligence: Construction, statistical testing, and first results of the inventory of self-estimated intelligence (ISI)]. *Zeitschrift für Differentielle und Diagnostische Psychologie, 23*, 435-466. doi:10.1024//0170-1789.23.4.435

Rindermann, H., & Neubauer, A. (2000). Informationsverarbeitungsgeschwindigkeit und Schulerfolg: Weisen basale Maße der Intelligenz prädiktive Validität auf? [Speed of information processing and success at school: Do basal measures of intelligence have predictive validity?]. *Diagnostica, 46*, 8-17.

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science, 2*, 313-345. doi:10.1111/j.1745-6916.2007.00047.x

Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., & de Fruyt, F. (2003). International Validity Generalization of GMA and cognitive abilities: A European community meta-analysis. *Personnel Psychology, 56*, 573-605. doi:10.1111/j.1744-6570.2003.tb00751.x

Schiefele, U., Moschner, B., & Husstegge, R. (2002). Skalenhandbuch SMILE-Projekt (unveröffentlichtes Manuskript) [Scale Manual SMILE Project (unpublished manuscript)]. Bielefeld, Germany: University.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124,* 262-274. doi:10.1037/0033-2909.124.2.262

Schmidt-Atzert, L. (2002). Tests und Tools – Intelligenz-Struktur-Test 2000 R [Tests and tools: Intelligence-Structure-Test 2000 R]. *Zeitschrift für Personalpsychologie*, *1*, 50–56.

Schmidt-Atzert, L. & Rauch, W. (2008). TBS-TK Rezension: „Intelligenz-Struktur-Test 2000 R (I-S-T 2000 R)“ [TBS-TK Review: „Intelligence Structure Test 2000 R (I-S-T 2000 R)“]. *Report Psychologie, 33*, 303-304.

Schneider, W. J., & McGrew, K. (2012). The Cattell-Horn-Carroll model of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 99-144). New York: NY: Guilford Press.

Schwarzer, R., & Jerusalem, M. (1995). Generalized Self-Efficacy scale. In J. Weinman, S. Wright, & M. Johnston (Eds.), *Measures in health psychology: A user's portfolio. Causal and control beliefs* (pp. 35-37). Windsor, England: NFER-NELSON

Seligman, M. E. P. (1975). *Helplessness: on depression, development, and death.* San Francisco, CA: W.H. Freeman.

Silverstein, A. B. (1990). Short forms of individual intelligence tests. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 2*, 3-11. doi:10.1037/1040-3590.2.1.3

Sloman, S. A., & Lagnado, D. A. (2005). The Problem of Induction. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 95-116.) Cambridge, NY: University Press.

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology, 15*, 72-101. doi:10.2307/1412159

Stern, W. (1911). Die Differentielle Psychologie in ihren methodischen Grundlagen [Differential psychology in its methodological foundations]. Leipzig, Germany: Johann Ambrosius Barth.

Sternberg, R. J. (2005). Intelligence. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 751-773.) Cambridge, NY: University Press.

Strenze, T. (2007). Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence, 35*, 401-426. doi:10.1016/j.intell.2006.09.004

Süß, H.-M., & Beauducel, A. (2005). Faceted Models of Intelligence. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligenc*e (pp. 313-332). Thousand Oaks, CA: Sage.

Wasserman, J. D. (2012). A history of intelligence assessment: The unfinished tapestry. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 3-55). New York, NY: Guilford Press.

Willis, J. O., Dumont, R., & Kaufman, A. S. (2011). Factor-analytic models of intelligence. In R. J. Sternberg & S. Kaufman (Eds.), *The Cambridge handbook of intelligence* (pp. 39-57). New York, NY: Cambridge University Press. doi:10.1017/CBO9780511977244.004

Ziegler, M., Dietl, E., Danay, E., Vogel, M., & Bühner, M. (2011). Predicting training success with general mental ability, specific ability tests, and (Un)structured interviews: A meta-analysis with unique samples. *International Journal of Selection and Assessment, 19*, 170-182. doi:10.1111/j.1468-2389.2011.00544.x