

Die Kurzform des Hagerer Matrizen-Tests (HMT-S)

Ein 6-Item Intelligenztest zum schlussfolgernden Denken

The Short Version of the Hagen Matrices Test (HMT-S)

A 6-Item Induction Intelligence Test

Timo Heydasch, Julia Haubrich, Karl-Heinz Renner

Zusammenfassung

Die Kurzform des Hagerer Matrizen-Tests (HMT-S) ist ein kostenfreier Online-Intelligenztest, der bezogen auf das CHC- Modell der Intelligenz (Schneider/McGrew 2012) Induktion bzw. Reasoning, die Fähigkeit zum schlussfolgernden Denken misst. Der 6-Item HMT-S basiert auf dem 20 Items umfassenden Hagerer Matrizen-Test (HMT; Heydasch/Renner/Haubrich/Hilbig/Zettler 2013). Die interne Konsistenz des HMT-S betrug in den durchgeführten Studien .62. Die Korrelationen des HMT-S zur Langversion betrug in einer ersten Studie $r = .79$ und $r = .78$ in einer zweiten. Zudem konnte die konvergente Validität durch Zusammenhänge u.a. zum Intelligenz-Struktur-Test 2000 R (Liepmann/Beauducel/Brocke/Amthauer 2007) belegt werden. Die Kriteriumsvalidität wurde durch Assoziationen zu akademischen Erfolgen nachgewiesen. Der HMT-S ist somit ein reliabler, valider und ökonomischer Intelligenztest, dessen Einsatz über folgende URL angefordert werden kann: <http://HMT.de.lv>

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/Renner/Haubrich/Hilbig/Zettler 2013). 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. In addition, 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 via <http://HMT.de.lv>



1 Einleitung

Bedeutende Bereiche des Lebens stehen mit Intelligenz in Zusammenhang. Sowohl für den Erfolg in der Schule oder im Studium (Rindermann/Neubauer 2000; Kuncel/Hezlett/Ones 2004; Kunina/Wilhelm/Formazin/Jonkmann/Schroeders 2007; Poropat 2009) als auch im beruflichen Kontext (Harrell/Harrell 1945; Hunter/Hunter 1984; Schmidt/Hunter 1998; Judge/Higgins/Thoresen/Barrick 1999; Salgado/Anderson/Moscoso/Bertua/de Fruyt/Rolland 2003; Kuncel/Hezlett/Ones 2004; Ng/Eby/Sorensen/Feldman 2005; Hülshager/Maier/Stumpp 2007; Ziegler/Dietl/Danay/Vogel/Bühner 2011; vgl. auch Strenze 2007) spielt Intelligenz eine zentrale Rolle. Aber auch andere Bereiche wie psychische Gesundheit (z.B. Khandaker/Barnett/White/Jones 2011), Ehescheidungen und die Lebensdauer (Roberts/Kuncel/Shiner/Caspi/Goldberg 2007) sind mit Intelligenz assoziiert.

Ist bei einer Studie die Berücksichtigung der Intelligenz angezeigt – sei es als Prädiktor, Moderator, Kontrollvariable (vgl. Blickle/Kramer/Mierke 2010), Mediator oder Kriterium – so stellt sich anscheinend zunächst die Qual der Wahl, aus dem vielfältigen Angebot von Intelligenztestverfahren (vgl. Leibniz-Zentrum für Psychologische Information und Dokumentation 2012) ein passendes Instrument auszuwählen. Neben Überlegungen zur Intelligenzdefinition, zum zugrundeliegenden Intelligenzmodell und der Festlegung des spezifischen Intelligenzbereiches spielen ökonomische Erwägungen eine Rolle. So können u.a. zwei Probleme resultieren, die dem Einsatz eines spezifischen Verfahrens entgegenstehen: 1. Die Instrumente sind zu lang und erfordern ein nicht zu realisierendes Maß an Testadministratoren- oder Teilnehmerzeit und/oder 2. handelt es sich um kommerzielle Verfahren, die teilweise zugangsbeschränkt über Verlage erworben werden müssen und für deren Einsatz ggf. erhebliche Kosten entstehen.

Kostenfreie Kurztests zur Messung der allgemeinen Intelligenz oder verschiedener Intelligenzbereiche sind Mangelware. Der 20 Items umfassende Hagerer Matrizen-Test (HMT; Heydasch/Renner/Haubrich/Hilbig/Zettler 2013) stellt als nicht-kommerzielles Online-Verfahren eine mögliche Option zur Intelligenzmessung dar. Sowohl Reliabilität (interne Konsistenz und Retestreliabilität) als auch konvergente und divergente Validität wurden belegt. Zudem sprechen Zusammenhänge zur Schulausbildung und zu universitären Leistungen für dessen Kriteriumsvalidität. Kritisch anzumerken ist jedoch, dass der HMT mit 20 Aufgaben womöglich für einige Einsatzbereiche zu zeitintensiv ist. Dies gilt insbesondere für Bereiche, in denen Kosteneinsparungen durch kurze und somit schnelle Verfahren zwingend sind bzw. die Akzeptanz von Teilnehmern und die Bereitschaft zur Teilnahme möglichst hoch sein sollen. Ein anderer Kritikpunkt bezieht sich auf die

Schwierigkeit. Da der HMT primär zur Studienerfolgsprädiktion entwickelt wurde, sind eher schwierige Matrizen enthalten, damit eine angemessene Differenzierung bei Studierenden und Studiumsinteressierten erzielt werden kann. Aus diesen Gründen war die Entwicklung eines kürzeren und weniger schwierigen Intelligenztests erforderlich. Umgesetzt wurde dies, indem wir auf Basis des HMT die Kurzversion des Hagener Matrizen-Tests, den HMT-S, entwickelten und somit nun ein weiterer kostenfreier Intelligenztest vorliegt.

Konsens über das, was Intelligenztests messen bzw. über eine Definition des Intelligenzbegriffes existieren aber nicht (Willis/Dumont/Kaufman 2011; Wasserman 2012). Im Bestreben, ein gewisses Maß an Übereinkunft bezüglich einer Definition herzustellen, unterzeichneten 1994 52 Experten (u.a. Carroll, Cattell, Eysenck, Horn, Jensen, Thorndike und Vernon) eine Definition, die sie als Mainstream in der Intelligenzforschung akzeptierten (Gottfredson 1997: 13):

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.

Unter anderem auf dieser Ausführung basierend definiert Sternberg (2005: 751) Intelligenz als „*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*“. Intelligente Personen sind also lernfähig und können sich an den jeweiligen Kontext anpassen und sind daher schließlich erfolgreich.

Bezüglich einer Theorie der Intelligenz oder dem hierarchischen Aufbau der kognitiven Fähigkeitskomponenten besteht ebenfalls kein Einvernehmen (Gottfredson/Saklofske 2009); vielmehr liegt eine Vielzahl an Theorien und Modellen vor (vgl. Davidson/Kemp 2011). Das Cattell-Horn-Carroll(CHC)-Modell der Intelligenz (Schneider/McGrew 2012; vgl. auch McGrew 1997; McGrew 2005) ist aber eine Taxonomie, die die theoretische Einordnung und Integration von Intelligenztheorien und Intelligenztests wie dem HMT-S ermöglicht. Das CHC-Modell basiert auf der Horn-Cattell Gf-Gc Theorie (Horn/Noll 1997; Horn/Blankson 2012) und der Three-Stratum-Theory (Carroll 1993; 2005) und ordnet die Intelligenzbereiche auf drei hierarchischen Ebenen an. An der Spitze steht *g*, die allgemeine Intelligenz (vgl. Spearman 1904; Jensen 1998). Die Ebene unterhalb von *g* enthält 16 breite Fähigkeitsbereiche, darunter *fluid reasoning* (Gf), *short-term memory* (Gsm), *processing speed* (Gs), *psychomotor speed* (Gps), *domain-specific knowledge* (Gkn) und *visual processing* (Gv). Unterhalb dieser Ebene werden verschiedene spezifi-

sche Fähigkeitsbereiche benannt, die als Facetten des jeweiligen breiten Fähigkeitsbereichs zu verstehen sind. Innerhalb von Gf werden beispielsweise die Bereiche *induction* (I), *general sequential reasoning* (RG, auch Deduktion genannt; vgl. Evans 2005) und *quantitative reasoning* (RQ) als gut gesicherte Facetten aufgeführt.

Ausgehend vom Aufgabenmaterial – den Matrizen – und der theoretischen Einordnung des HMT (Heydasch et al. 2013), erwarten wir, dass der HMT-S Induktion, „*the ability to observe a phenomenon and discover the underlying principles or rules...*“ (Schneider/McGrew 2012, S. 112; vgl. auch Sloman/Lagnado 2005), misst. Dabei operationalisiert der HMT-S Gf, denn: „*Induction is probably the core aspect of Gf*“ (Schneider/McGrew 2012, S. 112; siehe auch Carroll 1993). Zwischen Gf und g existiert wiederum ein Zusammenhang: Schneider und McGrew führen an, dass Cattell g als lebenslang kumuliertes Gf auslegte und dass andere Autoren Gf und g sogar als identisch betrachten.

In der Intelligenzforschung wurden von einigen Autoren alternative Modelle vorgestellt, die kognitive Fähigkeiten bzw. Aufgabenmaterial (neben anderen Merkmalen) nach Inhaltsbereichen differenzieren (vgl. Süß/Beauducel 2004). Dazu gehören das Radex-Modell (Guttman 1965; Guttman/Levy 1991), das Intelligenzstrukturmodell (Guilford 1967), das Berliner Intelligenzstrukturmodell (BIS; Jäger 1982) und das Hierarchische Rahmen- bzw. Protomodell der Intelligenzstrukturforschung (HPI; Amthauer/Brocke/Liepmann/Beauducel 2001; Liepmann/Beauducel/Brocke/Amthauer 2007). Diesen Modellen ist gemeinsam, dass verbale, numerische und figural-(bildhafte) Fähigkeiten bzw. Qualitäten des Aufgabenmaterials unterschieden werden. So werden figurale Matrizenaufgaben in der Regel dem figuralen Intelligenzbereich zugeordnet und demnach wäre der HMT-S als ein figuraler Induktionstest zu klassifizieren.

Die Validität des HMT-S sollte trotz der Testkürzung gegeben und zum HMT vergleichbar sein. Die Validität eines Kurztests ist zunächst als Übereinstimmung mit der Langversion definiert (Silverstein 1990). Da individualdiagnostische Erwägungen beim HMT-S als Forschungsinstrument nicht ins Gewicht fallen (z.B. individuelle Schätzung des Intelligenzquotienten oder Hochbegabungsdiagnostik), ist die Validität als Korrelation zwischen den Testversionen zu verstehen.

Aus der Korrelation bzw. der gemeinsamen Varianz von HMT-S und HMT kann nicht zwangsläufig auf die Konstrukt- oder Kriteriumsvalidität der Kurzversion geschlossen werden, sodass weitere Kennwerte in diesen Bereichen erforderlich sind: Es sollten sich hohe Validitätskoeffizienten zum einen zu Erfolgsindikatoren in Schule, Studium und Beruf ergeben und zum anderen zu Messwerten, die als Indikatoren schlussfolgernden Denkens allgemein akzeptiert sind. Validitätskoeffizienten sollten geringer ausfallen bei nur teilweise überlappenden Konstrukten.

Die Ähnlichkeit teilweise überlappender Konstrukte nimmt ab, wenn andere Intelligenzbereiche (z.B. Gkn oder Gsm versus Gf), unterschiedlich abstrakte Intelligenzbereiche (z.B. g versus l) oder andere Inhaltsbereiche (verbal oder numerisch versus figural) zum Vergleich herangezogen werden. Ebenfalls ist anzunehmen, dass durch unterschiedliche Datenquellen (z.B. Selbsteinschätzung versus objektiver Leistungstest, vgl. Cattell 1957; Pawlik 2006) oder durch verschiedene Erhebungsmethoden (z.B. Online-Test versus Paper-Pencil-Test; vgl. Guttman 1965; Guttman/Levy 1991) methodenbedingte Varianz die Korrelationen theoretisch nah stehender Konstrukte mindert. Unabhängigkeit sollte sich erweisen zu theoretisch nicht assoziierten Bereichen.

2 Methode

Zur Entwicklung und Validierung des HMT-S wurde die Stichprobe genutzt, die der Validierung des HMT diente (Heydasch et al. 2013). Diese Stichprobe wurde unter Verwendung der SPSS Statistiksoftware (Version 21) per Zufall in zwei annähernd gleich große Gruppen geteilt. Auf Basis der Datenanalyse der ersten Gruppe (Studie 1) wurden sechs Items aufgrund ihrer Schwierigkeiten ($p > .20$), Trennschärfen ($r_{it} > .30$) und Korrelationen mit dem Intelligenz-Struktur-Test 2000 R (Liepmann et al. 2007; $r > .30$) selektiert. Die resultierenden Kennwerte der nachträglich berechneten Kurzversion in Studie 1 wurden mit der zweiten Gruppe kreuzvalidiert (Studie 2). Zur weiteren Bestätigung der Kennwerte folgte eine dritte Studie, in der der HMT-S schließlich in der finalen Fassung vorgelegt wurde.

2.1 Stichprobe

In allen drei Studien bestanden die Stichproben aus Studierenden im Studiengang Psychologie (Bachelor of Science) der FernUniversität in Hagen, die für die Teilnahme Versuchspersonenstunden erhielten. Über alle drei Studien hinweg haben $N = 1.572$ Probanden an den Testungen und Befragungen teilgenommen. Der Anteil der Frauen betrug 75% und das Alter lag bei $M = 31.6$ Jahren ($SD = 8.97$). In Tabelle 1 werden die Stichprobengrößen, die Kennwerte für das Geschlechterverhältnis und die Altersverteilungen für die einzelnen Studien aufgeführt.

Tabelle 1 Stichprobengrößen, Anteile weiblicher Probandinnen und Kennwerte der Altersverteilung

Studie	N	Frauen	Alter					M	SD
			Percentile						
			10%	25%	50%	75%	90%		
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

2.2 Instrumente

Der HMT und der daraus entwickelte HMT-S gliedern sich in den Instruktionsteil, den Aufgabenteil und den Schlussteil. In der Instruktion wird das Aufgabenprinzip erläutert: Die Probanden sollen den regelgeleiteten Aufbau der unvollständigen figuralen 3x3 Matrizen¹ erkennen und aus einer Auswahl von acht Lösungsalternativen die korrekte identifizieren und markieren. Dazu stehen zwei Minuten pro Aufgabe zur Verfügung, wobei die Lösung auch schon vor Ablauf der Zeit abgesendet werden kann. Die Regeln, nach denen die Matrizen aufgebaut sein können, werden vorab benannt: Addition, Subtraktion sowie räumliche Verschiebungen (Elemente „bewegen“ sich von einer Seite zur anderen, rotieren etc.). Das Aufgabenprinzip wird dabei durch zwei Beispielaufgaben verdeutlicht². Der Aufgabenteil besteht beim HMT aus 20 bzw. beim HMT-S aus sechs Matrizen. Falls ein Testteilnehmer nicht über den „Weiter“-Button vor Ablauf der zweiminütigen Frist eine Lösung absendet, wird automatisch die ggf. markierte Lösung registriert und die folgende Aufgabe präsentiert. Zur zeitlichen Orientierung wird eine Uhr eingeblendet. Die Auswertung erfolgt online während der Testbearbeitung. Korrekte Lösungen werden mit 1, inkorrekte oder nicht bearbeitete Matrizen (auch Aufgaben, die nach einem Abbruch nicht bearbeitet wurden) werden mit 0 codiert und anschließend summiert. Im Schlussteil werden den Teilnehmern das individuelle Ergebnis in Form von Anzahl und Prozentsatz korrekt gelöster Aufgaben mitgeteilt. Zusätzlich wird der HMT bzw. der HMT-S für Laien verständlich inhaltlich näher skizziert und beispielsweise darauf hingewiesen, dass der Test nicht perfekt reliabel ist oder Zusammenhänge zu akademischen Erfolgen nicht deterministisch zu verstehen sind. Auf

1 Matrizen freundlich überlassen von Lutz Hornke.

2 Die Instruktion mit den Beispielaufgaben ist online einzusehen unter der URL https://ww3.uni-park.de/uc/HMT_S_Vorschau/

die Berechnung und Rückmeldung eines IQs wird beim HMT-S im Gegensatz zum HMT verzichtet.

Zur weiteren Validierung wurden folgende Instrumente herangezogen:

- *Intelligenz-Struktur-Test 2000 R* (I-S-T 2000 R; Liepmann et al. 2007; vgl. Schmidt-Atzert 2002; Schmidt-Atzert/Rauch 2008): Der I-S-T 2000 R erfasst als objektiver Leistungstest folgende Intelligenzbereiche: Schlussfolgendes Denken (Reasoning, R), Wissen (W) und Merkfähigkeit (ME). R und W werden jeweils gebildet aus Skalen mit verbalen, numerischen und figuralen Inhalten, die wiederum jeweils drei Aufgabengruppen enthalten. ME basiert auf einer verbalen und einer figuralen Aufgabengruppe. Weiterhin lassen sich die als unabhängig konzipierten Faktoren für fluide (g_f) und kristalline (g_c) Intelligenz bestimmen. Der I-S-T 2000 R bezieht sich theoretisch auf das Radex-Modell (Guttman 1965; Guttman/Levy 1991) bzw. das BIS (Jäger, 1982), indem Aufgabehalte zu unterscheiden sind nach verbaler, numerischer und figuraler Qualität. Gleichzeitig beziehen sich die Autoren mit der Konzeption der Dimensionen R und W auf Cattell (1987). Die beiden Ansätze wurden von Amthauer und Kollegen (2001) zum HPI synthetisiert, das schließlich die theoretische Grundlage des I-S-T 2000 R darstellt.
- *Inventar zur selbsteingeschätzten Intelligenz* (ISI; Rammstedt/Rammsayer 2002): Das ISI misst in einer Online-Adaption basierend auf dem Konzept multipler Intelligenzen (Gardner 1983) die selbsteingeschätzten Intelligenzen in den Bereichen verbales Verständnis, Wortflüssigkeit, mathematische Intelligenz, räumliche Intelligenz, Gedächtnisfähigkeit, Wahrnehmungsgeschwindigkeit, logisches Denken, musikalische Intelligenz, körperlich-kinästhetische Intelligenz, interpersonale Intelligenz sowie intrapersonale Intelligenz.
- *Allgemeine Selbstwirksamkeitserwartung* (SWE; Schwarzer/Jerusalem 1995): Mit der Skala SWE wird die situationsübergreifende Überzeugung erfasst, mit Problemen und Herausforderungen umgehen und diese erfolgreich bewältigen zu können.
- *Studiumspezifische Selbstwirksamkeit* (Schiefele/Moschner/Husstegge 2002): In Anlehnung an die Skala zur Messung der Allgemeinen Selbstwirksamkeitserwartung misst die für diese Untersuchung an das Fernstudium adaptierte Skala die Überzeugung, erfolgreich mit Herausforderungen im Studienkontext umgehen zu können.
- *Allgemeine Hilfflosigkeit* (Jerusalem/Schwarzer 1986, 2012): Auf Basis der Theorie der gelernten Hilfflosigkeit von Seligman (1975) erfasst diese Skala bereichsübergreifend die selbsteingeschätzte Hilfflosigkeit bei der Bewältigung von Problemen und Herausforderungen.

- *Studiumspezifische Hilfslosigkeit* (Jerusalem/Schwarzer 1986, 2012): Mit diesem Fragebogen wird die Erwartung von Studierenden gemessen, Probleme im Kontext eines Studiums nicht meistern zu können.
- *Instrumente zum Selbstkonzept*: Die drei Skalen *generelles akademisches Selbstkonzept*, *mathematisches Selbstkonzept* und *sprachliches Selbstkonzept* (Schiefele et al. 2002) erfassen für die aufgeführten Bereiche das Selbstverständnis einer Person, das sich auf vorhandene Fähigkeiten und Problemlösekompetenzen bezieht.
- *Revised Achievement Motivation Scale* (AMS-R; Lang/Fries 2006): Die AMS-R misst das explizite Leistungsmotiv mittels der Dimensionen *hope of success* und *fear of failure*, die die Annäherungs- bzw. die Vermeidungstendenz des Leistungsmotivs abbilden.

Zusätzlich wurden demographische Variablen (Geschlecht und Alter) erhoben, das Land des Schulabschlusses und der erreichte Schulabschluss (codiert mit *Allgemeine Hochschulreife* = 3, *Fachhochschulreife oder fachgebundene Hochschulreife* = 2, *Mittlerer Schulabschluss* = 1). Zudem berichteten die Teilnehmer ihre Schulabschlussnote und jeweils die letzte erreichte Note in den Fächern Mathematik, Englisch, Deutsch, Biologie, Kunst und Sport. Ergänzend gaben die Teilnehmer ihre Modulprüfungsnoten im Studiengang B.Sc. Psychologie an. Zur Berechnung eines interindividuell vergleichbaren Studienerfolgsindikators wurden die Modulabschlussnoten zunächst standardisiert und anschließend gemittelt, um für Niveauunterschiede zwischen den Modulprüfungen zu kontrollieren. Für die Berechnung des Indikators des Studienerfolgs musste mindestens eine Note angegeben worden sein.

2.3 Durchführung

Die Datenerhebung erfolgte weitestgehend online und unbeaufsichtigt mit EFS-Survey der QuestBack GmbH (vgl. Buchwald/Spoden/Fleischer/Leutner 2013). Die Probanden haben so sowohl den Ort (z.B. heimischer PC oder mobiler Laptop) als auch den Zeitpunkt der Teilnahme selbst bestimmt. Auf einer Web-Seite (<http://www.fernuni-hagen.de/psychologie/forschung/vlabor.shtml>) konnten sich die potentiellen Teilnehmer über die Teilnahmebedingungen der universitären Online-Studien informieren und aus verschiedenen zeitgleich angebotenen Projekten wählen. Nach Zugang zu einem Projekt wurden die Inhalte und Teilnahmebedingungen der jeweiligen Studie spezifiziert und die Teilnehmer erklärten anschließend ihr Einverständnis zur Teilnahme.

Bei den Hagener Matrizen-Tests wurden zu Beginn die grundlegenden demographischen Daten (z.B. Alter und Geschlecht) und darauf folgend die selbst-ingeschätzten Intelligenzen mit dem ISI erhoben. Im Anschluss wurde die jeweilige Version des Hagener Matrizen-Tests vorgelegt. Zuletzt wurden den Teilnehmern die individuellen Ergebnisse des HMT bzw. des HMT-S zurückgemeldet. Aufgrund der großen Anzahl und teilweise enormen Länge der vorgelegten Instrumente³ wurden bei Studie 1 und 2 die Verfahren zur Messung der Selbstwirksamkeit, zur Hilfflo-sigkeit, zum Selbstkonzept und zum Leistungsmotiv nicht im gleichen, sondern in unabhängigen Online-Projekten integriert. In Studie 3 wurden diese Instrumente trotz der resultierenden Länge in das Projekt des HMT-S integriert und zwischen den demographischen Variablen und dem ISI positioniert.

Der I-S-T 2000 R wurde abweichend zu den Online-Erhebungen als Paper-Pencil-Verfahren durchgeführt⁴. Studierende haben sich dafür für vorgegebene Termine und Orte registriert. Die Testungen wurden anschließend entsprechend den Anweisungen im Manual umgesetzt.

Bei jeder Sitzung, in der Fragebögen oder Tests bearbeitet wurden, erstellten die Teilnehmer regelgeleitet einen sechsstelligen individuellen Pseudonymisierungscodes, der immer nach denselben Vorgaben gebildet wurde. So wurde jeder Person ein bestimmter Code zugeordnet. Anhand der Pseudonymisierungscodes konnten die Daten verschiedener Sitzungen einer Person zusammengeführt werden, ohne die Identität der Teilnehmer (z.B. Name oder Matrikelnummer) zu erfahren. Mehrfachteilnahmen konnten ebenfalls durch die Pseudonymisierungscodes identifiziert und ausgeschlossen werden. Letztlich wurden auch Daten zum Schulerfolg von Teilnehmern mit nicht deutschem Schulabschluss von der Analyse ausgeschlossen, da sowohl Schulausbildung als auch Schulnotencodierung nicht unbedingt vergleichbar sind mit dem deutschen System bzw. Probleme der Übertragbarkeit existierten.

Die Datenauswertung erfolgte mit der SPSS Statistiksoftware (Version 21), wobei die Daten zum I-S-T 2000 R zunächst als Rohdaten von den Antwortbögen übertragen wurden. Unter Verwendung von SPSS-Syntaxen wurden anschließend die vorzunehmenden Recodierungen, die Berechnungen der Skalenwerte sowie die weiteren Analysen durchgeführt.

- 3 Weitere Instrumente waren enthalten, die zur Klärung anderer Fragestellungen eingesetzt wurden.
- 4 Der I-S-T 2000 R konnte aus ökonomischen Gründen in Studie 3 leider nicht mehr eingesetzt werden.

3 Ergebnisse

Die im ersten Schritt vorgenommene Selektion von sechs Matrizen des HMT für die Kurzversion wurde anhand der Eigenschaften der Items bezüglich Schwierigkeiten, Trennschärfen und den Korrelationen zur I-S-T 2000 R Reasoning-Skala (vgl. Tabelle 2) durchgeführt. Die Items 1 bis 3 konnten übernommen werden, da kein Ausschlusskriterium zur Selektion zutraf. Die Items 4, 6 und 8 wurden dagegen aufgrund der geringen Korrelationen mit dem I-S-T 2000 R ($r = .13$, $r = .05$ bzw. $r = .08$) nicht übernommen. Zudem fiel Item 8 durch eine zu geringe Trennschärfe auf ($r_{it} = .22$). Zusätzlich zu den bereits selektierten Items 1 bis 3 wurden noch die unproblematischen Items 5, 7 und 9 ausgewählt, wobei die Itemreihenfolge beibehalten wurde. An diesem Punkt wurde die Itemselektion beendet, da a) eine interne Konsistenz von .64 erreicht wurde, b) sowohl einfache und moderat schwierige Aufgaben als auch ein eher schwierigeres Item enthalten waren und zudem c) die mittlere Gesamtbearbeitungsdauer für den Kurztest bei weiteren zusätzlichen Items über zehn Minuten gestiegen wäre.

Die Gesamtdauer des HMT-S lag durchschnittlich bei 9.4 Minuten (vgl. Tabelle 3). Dabei teilt sich die Zeit auf die Instruktion und die eigentliche Aufgabenbearbeitung auf.

In Tabelle 4 sind die Lösungszeiten für die einzelnen Items aufgeführt. Zusätzlich enthält Tabelle 4 die Schwierigkeiten und Trennschärfen der Items. Es zeigen sich weitestgehend Übereinstimmungen der Kennwerte in den drei Studien. Die ersten beiden Items werden relativ schnell gelöst (im Durchschnitt $28 \leq M_t \leq 44$ Sekunden), Items 3 bis 5 werden im Durchschnitt nach 51 bis 54 Sekunden gelöst und das sechste Item relativ spät (im Durchschnitt $76 \leq M_t \leq 77$ Sekunden). Zur Schwierigkeit lässt sich feststellen, dass die ersten beiden Items sehr einfach sind (mit Lösungswahrscheinlichkeiten von $p \geq .83$), Items 3 bis 5 weisen mittlere Schwierigkeiten auf ($.53 \leq p \leq .70$) und das letzte Item ist eher schwierig ($p \leq .35$). Die Trennschärfen der Items 1 bis 5 liegen im Bereich von $r_{it} = .29$ bis $r_{it} = .47$. Das sechste Item bzw. die letzte Aufgabe fällt von diesem Niveau etwas ab. Die Trennschärfe dieses Items erreicht im günstigsten Fall von Studie 1 einen Wert von $r_{it} = .23$.

Beim HMT-S werden am häufigsten 4 von 6 Aufgaben richtig gelöst (vgl. Tabelle 5). In allen drei Studien ist die Verteilung der Lösungshäufigkeiten bedingt durch die relative Leichtigkeit linksschief bzw. rechtssteil. Bei der Wölbung zeigt sich ein nicht einheitliches Bild. Während die Verteilungen der Studien 1 und 2 flachgipflig sind, ist die Verteilung in Studie 3 normalgipflig. In Tabelle 5 sind zudem die Indikatoren der internen Konsistenz angeführt, die auf Basis der Kuder-

Tabelle 2 Eigenschaften und I-S-T 2000 R-Korrelationen der HMT-Items in Studie 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	<u>.13</u>
5	51	.65	.45	.27
6	76	.54	<u>.30</u>	<u>.05</u>
7	53	.58	.42	.33
8	82	.36	<u>.22</u>	<u>.08</u>
9	76	.25	.35	.44
10	68	.29	.43	.35
11	81	.25	<u>.29</u>	<u>.03</u>
12	66	.30	<u>.30</u>	.22
13	66	<u>.20</u>	.37	.31
14	78	<u>.15</u>	<u>.24</u>	.41
15	78	<u>.16</u>	.52	<u>.16</u>
16	60	<u>.14</u>	.40	<u>.03</u>
17	67	<u>.18</u>	<u>.28</u>	.36
18	61	<u>.15</u>	.51	.23
19	52	<u>.12</u>	<u>.29</u>	<u>.13</u>
20	41	<u>.11</u>	.44	<u>.07</u>

I-S-T 2000 R = Intelligenz-Struktur-Test 2000 R (Liepmann et al. 2007). HMT = Hagener Matrizen-Test (Heydasch et al. 2013). M_t = Mittlere Bearbeitungsdauer in Sekunden. p = Schwierigkeit. r_{it} = Trennschärfe. r_{IST} = Itemkorrelation mit dem I-S-T 2000 R Reasoning-Gesamtwert. Fett formatierte Items wurden für die Kurzform selektiert. Unterstrichene Kennwerte führten zum Ausschluss des Items für die Kurzversion ($p \leq .20$, $r_{it} \leq .30$ oder $r_{IST} \leq .20$).

^a $N_t = 606$. ^b $N_j = 681$. ^c $N_j = 56$.

Tabelle 3 Test- und Bearbeitungsdauer (in Minuten)

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

Aggregation über alle drei Studien, da keine wesentlichen Unterschiede festzustellen waren. M_t = Mittlere Bearbeitungsdauer. SD_t = Standardabweichung der Bearbeitungsdauer.

$N = 1.563$.

Tabelle 4 Mittlere Lösungsdauer, Schwierigkeiten und Trennschärfen der Items

Item	M_t			p			r_{it}		
	Studie			Studie			Studie		
	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

M_t = Mittlere Bearbeitungsdauer in Sekunden. p = Schwierigkeit. r_{it} = Trennschärfe.

^a N_1 = 636. ^b N_2 = 633. ^c N_3 = 229. ^d N_1 = 681. ^e N_2 = 658. ^f N_3 = 233.

Tabelle 5 Skaleneigenschaften

Studie	N	M	SD	Schiefe	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

KR20 = Interne Konsistenz auf Basis der Kuder-Richardson Formel 20 (Kuder/Richardson 1937).

Richardson Formel 20 (KR20; Kuder/Richardson 1937) berechnet wurde. Diese schwankt über die Studien hinweg von $KR20_1 = .64$, über $KR20_2 = .61$ bis $KR20_3 = .57$. Wird die interne Konsistenz über alle drei Studien hinweg ermittelt, resultiert ein Wert von $KR20 = .62$ ($N = 1.572$).

Festzustellen sind Geschlechtsunterschiede (vgl. Tabelle 6). Dieser Effekt ist aber eher klein ($d_1 = 0.16$, $d_2 = 0.25$ und $d_3 = 0.15$; $p_1 = .058$, $p_2 = .004$ und $p_3 = .354$).

Die Beziehung der HMT-S-Testergebnisse zum Alter ist leicht negativ: Ältere Teilnehmer lösen tendenziell weniger Aufgaben. Dieser Zusammenhang ist aber in den ersten beiden Studien niedrig und in Studie 3 nicht nachweisbar ($r_1 = -.13$, $r_2 = -.12$, $r_3 = -.02$; $p_1 = .001$, $p_2 = .003$, $p_3 = .768$; $N_1 = 679$, $N_2 = 654$ bzw. $N_3 = 233$).

Die Korrelationen des HMT-S mit der Langversion betragen $r_1 = .79$ ($p < .001$) in Studie 1 und $r_2 = .78$ in ($p < .001$) Studie 2. Zur inhaltlichen Bestimmung der in der Langversion und der Kurzversion geteilten Varianz wurde für Studie 1 und 2 jeweils eine hierarchische Regression berechnet. Als inhaltlicher Marker und

Tabelle 6 Geschlechtsunterschiede

Studie		<i>N</i>	<i>M</i>	<i>SD</i>	<i>t</i>	<i>df</i>	<i>d</i>
1	Männer	174	4.08	1.64	1.90	676	0.16
	Frauen	504	3.82	1.51			
2	Männer	173	4.06	1.46	2.89**	316.82	0.25
	Frauen	483	3.69	1.53			
3	Männer	47	4.21	1.53	0.93	231	0.15
	Frauen	186	3.99	1.46			

** $p < .01$.

Tabelle 7 Hierarchische Regressionen zur inkrementellen Validität des HMT zum HMT-S

	Studie 1 ^a				Studie 2 ^b			
	<i>R</i>	ΔR^2	<i>F</i>	β	<i>R</i>	ΔR^2	<i>F</i>	β
Schritt 1	.530	.281	21.1***		.494	.244	10.7**	
HMT-S				.530				.494
Schritt 2	.567	.041	3.2		.607	.125	6.3*	
HMT-S				.215				.072
HMT				.374				.551*

AV: Reasoning-Gesamtwert des Intelligenz-Struktur-Test 2000 R (Liepmann et al. 2007). HMT = Hagener Matrizen-Test (Heydasch et al. 2013). HMT-S = Kurzform des Hagener Matrizen-Test.

^a $N_1 = 56$. ^b $N_2 = 35$. * $p < .05$. ** $p < .01$. *** $p < .001$.

abhängige Variable diente der Reasoning-Gesamtwert des I-S-T 2000 R. Im ersten Schritt wurde der Testwert des HMT-S als Prädiktor eingegeben und im zweiten Schritt der Testwert des HMT. In Tabelle 7 sind die Ergebnisse der hierarchischen Regressionen festgehalten. Da die Varianzanteile, die über den HMT-S durch den HMT erklärt werden ($\Delta R^2 = .04$ in Studie 1 bzw. $\Delta R^2 = .13$ in Studie 2) im Vergleich zu den Anteilen, die der HMT-S aufklärt ($R^2 = .28$ in Studie 1 bzw. $R^2 = .24$ in Studie 2) relativ klein sind, kann geschlossen werden, dass nicht Fehlervarianz (z.B. Methodenvarianz), sondern die Varianzanteile, die für die Reasoning-Fähigkeit stehen, geteilt werden.

Neben der Validitätsbestimmung durch den Vergleich mit der Langversion wurden Korrelationen zum I-S-T 2000 R berechnet (siehe Tabelle 8). Da sich Kennwerte insbesondere bei den verbalen Intelligenzmaßen von Studie 1 zu Stu-

Tabelle 8 Konvergente Validität zum I-S-T 2000 R

Bereich	Inhalt	Aufgaben	Studie		
			1 ^a	2 ^b	1+2 ^c
Reasoning	gesamt		.53***	.49**	.52***
	<i>g_f</i>		.50***	.47**	.49***
	V		.46***	.03	.30**
		SE	.24	.08	.16
		AN	.34*	-.14	.18
		GE	.44***	.13	.33**
	N		.42**	.53**	.46***
		RE	.31*	.42*	.35***
		ZR	.37**	.42*	.39***
		RZ	.40**	.54***	.45***
	F		.47***	.46**	.47***
		FA	.25	.31	.28**
	WÜ	.50***	.51**	.50***	
	MA	.21	.24	.22*	
Wissen	gesamt		.24	.42*	.31**
	<i>g_c</i>		.15	.36*	.23*
	V		.06	.24	.13
	N		.26*	.36*	.30**
	F		.29*	.48**	.35***
Merkfähigkeit	gesamt		.36**	.07	.24*
	V		.24	-.04	.12
	F		.34**	.16	.27*

I-S-T 2000 R = Intelligenz-Struktur-Test 2000 R (Liepmann et al. 2007); *g_f* = fluide Intelligenz; V = Verbaler Inhaltsbereich; N = Numerischer Inhaltsbereich; F = Figuraler Inhaltsbereich; *g_c* = kristalline Intelligenz; SE = Satzergänzungen; AN = Analogien; GE = Gemeinsamkeiten; RE = Rechenaufgaben; ZR = Zahlenreihen; RZ = Rechenzeichen; FA = Figurenauswahl; WÜ = Würfelaufgaben; MA = Matrizen.

^a*N*₁ = 56. ^b*N*₂ = 35. ^c*N*₁₂ = *N*₁ + *N*₂ = 91. * *p* < .05. ** *p* < .01. *** *p* < .001.

die 2 deutlich unterschieden, wurden zur besseren Interpretierbarkeit die beiden Stichproben zusammengefasst und die Korrelationen des HMT-S zum I-S-T 2000 R erneut berechnet. So korreliert der HMT-S mit dem Reasoning-Gesamtwert mit $r = .52$ und mit g_f mit $r = .49$. Auf Ebene der verschiedenen Inhaltbereiche liegen die Korrelationen für die Skalen zum numerischen und zum figuralen Reasoning nur knapp unterhalb des Niveaus der Gesamtskala ($r = .46$ bzw. $r = .47$). Der HMT-S

korreliert hingegen mit der verbalen Reasoning-Skala lediglich moderat ($r = .30$). Die Skalen W und ME korrelieren mit dem HMT-S auf mittlerem Niveau ($r = .31$ bzw. $r = .24$). Auf der Ebene der unterschiedlichen Inhaltsbereiche lässt sich ein Zusammenhangsmuster finden, das dem innerhalb des Reasoning-Bereiches vergleichbar ist: die numerischen und die figuralen Skalen korrelieren in etwa gleich hoch auf moderatem Niveau mit dem HMT-S ($.27 \leq r \leq .35$), wohingegen die Skalen mit verbalem Aufgabenmaterial nur schwach mit dem HMT-S korrelieren ($r = .13$ bei W bzw. $r = .12$ bei ME). Betrachtet man die Zusammenhänge mit den einzelnen Aufgabengruppen, so wird deutlich, dass die Korrelationen innerhalb eines Inhaltsbereichs heterogen ausfallen. Dies betrifft zwar sowohl die verbalen als auch die numerischen Aufgabengruppen, aber insbesondere die figuralen Aufgaben Figurenauswahl ($r = .28$), Würfelaufgaben ($r = .50$) und Matrizen ($r = .22$) unterscheiden sich hinsichtlich des jeweiligen Zusammenhangs zum HMT-S deutlich.

Die Assoziationen zum ISI (vgl. Tabelle 9) sind über alle drei Studien weitestgehend äquivalent. Positive Zusammenhänge zeigt der HMT-S zu den theoretisch eher nahen Konstrukten mathematische Intelligenz, numerische Intelligenz und logischem Denken. Weitere Korrelationen sind nicht festzustellen, bis auf leicht negative Korrelationen zur körperlich-kinästhetischen, interpersonalen und intrapersonalen Intelligenz. Mittelhohe Korrelationen ergaben sich bei der Prüfung von Zusammenhängen des HMT-S zu Fähigkeitsüberzeugungen im Bereich des mathematischen Selbstkonzepts. Dieser Zusammenhang konnte sowohl in Studie 2 als auch in Studie 3 repliziert werden. Zumindest tendenziell ist ein positiver Zusammenhang zum akademischen Selbstkonzept wie auch eine negative Assoziation zur studiumspezifischen Hilflosigkeit zu identifizieren. Betrachtet man die Korrelationen zu den Dimensionen der AMS-R, so zeigt sich Unabhängigkeit zur Dimension *fear of failure* und allenfalls eine leichte Tendenz eines positiven Zusammenhangs zur Dimension *hope of success*.

Zusammenhänge des HMT-S zu akademischen Erfolgen der Teilnehmer sind in Tabelle 10 aufgeführt. Positive, über alle drei Studien hinweg bestätigte Zusammenhänge bestehen zur Schulabschlussnote sowie zur letzten erhaltenen Note in Mathematik. Uneinheitlich ist die Befundlage bezüglich des Schulabschlussniveaus und der bisher erreichten Durchschnittsnote im Studium. Tendenziell gehen bessere Testwerte mit höherer Schulbildung einher und bessere Testteilnehmer sind erfolgreicher im Studium. Lediglich in jeweils einer der drei Studien resultierten substantielle Ergebnisse bezogen auf das Schulfach Biologie sowie zum Studienfach Methoden und Statistik. Zu den Schulnoten der Fächer Englisch, Deutsch, Kunst und Sport bestehen keine Zusammenhänge.

Tabelle 9 Divergente Validität

	Studie		
	1	2	3
ISI^a			
Verbales Verständnis	.02	-.07	-.05
Wortflüssigkeit	-.03	-.10*	-.04
Mathematische Intelligenz	.23***	.24***	.20**
Räumliche Intelligenz	.22***	.20***	.22***
Gedächtnisfähigkeit	-.04	-.03	-.01
Wahrnehmungsgeschwindigkeit	.05	-.01	-.07
Logisches Denken	.17***	.18***	.14*
Musikalische Intelligenz	-.02	-.04	-.06
Körperlich-kinästhetisch Intelligenz	-.00	-.09*	-.13*
Interpersonale Intelligenz	-.04	-.16***	-.13
Intrapersonale Intelligenz	-.02	-.14***	-.05
SWE			
Allgemein ^b	.09	.07	.08
Studienspezifisch ^c	.18*	.14	.01
Hilflosigkeit			
Allgemein ^d	-.14*	.00	-.02
Studienspezifisch ^c	-.21**	-.15*	-.10
Selbstkonzept^e			
Akademisch	.23***	.10	.10
Mathematisch	.28***	.27***	.32***
Verbal	.10	-.07	-.00
AMS-R^f			
<i>hope of success</i>	.21***	.09	.12
<i>fear of failure</i>	-.12	.04	-.06

ISI = Inventar zur selbsteingeschätzten Intelligenz (Rammstedt/Rammseyer 2002); SWE = Allgemeine Selbstwirksamkeitserwartung (Schwarzer/Jerusalem 1995); AMS-R = Revised Achievement Motivation Scale (Lang/Fries 2006).
^aN₁ = 675, N₂ = 654, N₃ = 225. ^bN₁ = 286, N₂ = 262, N₃ = 155. ^cN₁ = 176, N₂ = 177, N₃ = 155. ^dN₁ = 266, N₂ = 242, N₃ = 155. ^eN₁ = 345, N₂ = 317, N₃ = 155. ^fN₁ = 244, N₂ = 242, N₃ = 155. *p < .05. **p < .01. ***p < .001.

Tabelle 10 Kriteriumsvalidität

	Studie		
	1	2	3
Schulbildung^a			
Schulabschlussniveau ^b	.16**	.10	.00
Abschlussnote	.14*	.16**	.16*
Mathematik	.19***	.20***	.19**
Englisch	.03	.03	.09
Deutsch	-.01	-.02	.01
Biologie	.07	.07	.16*
Kunst	.01	.04	.09
Sport	.03	.09	.03
Studium			
Durchschnittsnote ^c	.28**	.18*	.06
Note in Methoden und Statistik ^d	.22	.38**	.05

Noten wurden invertiert, sodass positive Korrelationen bedeuten, dass bessere Testleistungen mit höherer akademischer Leistung assoziiert sind.

^a $N_1 = 308$, $N_2 = 287$, $N_3 = 191$. ^bAllgemeine Hochschulreife = 3, Fachhochschulreife oder fachgebundene Hochschulreife = 2, Mittlerer Schulabschluss = 1 (Spearman-Korrelation). ^c $N_1 = 131$, $N_2 = 124$, $N_3 = 78$. ^d $N_1 = 70$, $N_2 = 70$, $N_3 = 67$. * $p < .05$. ** $p < .01$. *** $p < .001$.

4 Diskussion

Der HMT-S ist ein flexibles Instrument, das bei Online-Erhebungen implementiert (uns sind keine wesentlichen Einschränkungen bei der technischen Umsetzung bekannt) und an beliebigen Orten sowohl im Feld als auch im Labor eingesetzt werden kann, sofern ein internetfähiger PC oder Laptop und Internetzugang zur Verfügung stehen. Dabei ist der HMT-S nicht nur in der psychologischen, sozialwissenschaftlichen oder pädagogischen Forschung einsetzbar, sondern ebenfalls in verschiedenen Praxisfeldern. Die Testanwendung ist nicht auf Psychologinnen und Psychologen beschränkt. Auch anderen nach wissenschaftlichen Standards arbeitenden Personengruppen steht die Nutzung offen. Ausgenommen sind allerdings Anwendungen, bei denen auf Grundlage des HMT-S eine Einzelfalldiagnostik beispielsweise zu Selektionszwecken (z.B. Personalauswahl) erfolgen soll. In diesen Fällen muss auf alternative Verfahren zurückgegriffen werden, die notwendige Gütemaßstäbe erfüllen, wie sie z.B. mit der DIN 33430 (DIN, 2002) definiert sind.

Die Schwierigkeit des HMT-S ist moderat und der Test kann somit Stichproben vorgelegt werden, die weder durch extrem geringe noch durch sehr hohe kognitive Leistungsfähigkeit gekennzeichnet sind. Dennoch kann die Stichprobe durchaus heterogen sein, denn durch die unterschiedlichen Aufgabenschwierigkeiten werden sowohl Boden- als auch Deckeneffekte vermieden: Die Lösungswahrscheinlichkeit nimmt von Aufgabe zu Aufgabe zu und die letzte Aufgabe ist schließlich so schwierig, dass Differenzierungen auf niedrigerem wie auch auf höherem Fähigkeitsniveau möglich sind (im Rahmen eines Instrumentes mit sechs dichotomen Items). Die Ergebnisse zur Schwierigkeit des HMT-S sind allerdings unter Vorbehalt zu interpretieren. Es handelte sich bei den Probanden der vorgestellten Studien um Studierende und folglich ist ein höheres kognitives Fähigkeitsniveau anzunehmen als für die Gesamtpopulation. Dieser Unterschied sollte aber nicht gravierend sein, da die Studierenden nicht wie in anderen Psychologiestudiengängen durch einen Numerus Clausus selektiert wurden.

Der HMT-S ist äußerst ökonomisch. Die Bearbeitungsdauer liegt mit fünf Minuten bei 10% bis 25% bisheriger Matrizen-Tests, die üblicherweise 20 bis 45 Minuten dauern. Die zeitliche Limitierung von zwei Minuten für die Bearbeitung einer einzelnen Aufgabe erscheint dabei in den meisten Fällen ausreichend, um eine Lösungsalternative zu markieren und abzuschicken. Lediglich bei der schwierigeren sechsten und letzten Matrix werden die zwei Minuten Zeitlimit annähernd erreicht. Im Einzelfall kann es aber durchaus der Fall sein, dass die zwölf Minuten Gesamtzeit zur Bearbeitung ausgeschöpft werden⁵.

Effekte von Geschlecht und Alter der Teilnehmer auf die Messung sind vorhanden aber gering. Somit deutet sich an, dass der HMT-S bezüglich Geschlecht und Alter ein relativ fairer Test ist. Eine abschließende Bewertung der Befundlage ist aber aufgrund inkonsistenter Befunde zwischen den drei Studien noch nicht möglich.

Die Objektivität sowohl der Auswertung als auch der Interpretation ist gegeben. Die Auswertung erfolgt automatisiert online und die Interpretation erfolgt einheitlich web-basiert und wird den Testteilnehmern als Rückmeldung der individuellen Ergebnisse im Anschluss an die Testaufgaben präsentiert. Die Durchführungsobjektivität ist hingegen nicht unbedingt gegeben (Gnambs/Batinic/Hertel 2011). Die Instruktion, die Darstellung des Tests und die Zeitnahme bei den Aufgaben sind durch die Web-Programmierung objektiv. Wenn jedoch der Test außerhalb einer kontrollierten (Labor-)Umgebung durchgeführt wird, ist die Durch-

5 Dies ist aber nicht zwangsläufig an die Leistungsstärke gebunden. Eine minimale Spearman-Korrelation zwischen dem HMT-S und der Bearbeitungszeit belegt ($r = .08, p = .001, N = 1.563$), dass – wenn überhaupt – eher bessere Teilnehmer näher am Zeitlimit sind und länger an den Aufgaben arbeiten.

führungsobjektivität gefährdet und es kann zu Störeinflüssen kommen, die z.B. zu einer Unterbrechung der Konzentration oder gar der Testbearbeitung führen. Eine unbeaufsichtigte Feldtestung bietet aber die Möglichkeit, ein hohes Maß an ökologischer Validität zu erreichen, denn die Künstlichkeit einer Laborumgebung mit anwesender Versuchsleitung als Aufsicht ist dann nicht gegeben.

Die Messgenauigkeit des HMT-S ist, nach Maßstäben von Aiken und Groth-Marnat (2006), die für gruppenstatistische Untersuchungen einen Wert von .60 als ausreichend ansehen, mit einer mittleren internen Konsistenz von $KR_{20} = .62$ akzeptabel.

Die Validität des HMT-S wird zunächst durch die sehr hohen Korrelationen zur Langversion HMT belegt. Zudem wurde durch die Regressionsanalysen nachgewiesen, dass es sich bei der geteilten Varianz zwischen Lang- und Kurzversion größtenteils um denjenigen Varianzanteil handelt, der die unterschiedlichen Reasoning-Fähigkeiten der Probanden gemessen durch den I-S-T 2000 R erklärt. Damit ist die Übereinstimmung von Kurz- und Langversion weitestgehend auf Intelligenzunterschiede der Probanden zurückzuführen und stellt nicht etwa methodische Artefakte dar, die durch die identische Durchführungsmethodik im unbeaufsichtigten Online-Assessment bei HMT und HMT-S verursacht wurden. Die Regressionsanalysen sprechen zwar für die Überlegenheit der Langversion (immerhin ist diese mit 20 Items mehr als drei Mal so lang), jedoch wird ein Großteil der Varianz des I-S-T 2000 R schon alleine durch den HMT-S aufgeklärt.

Bei den Zusammenhängen des HMT-S mit dem I-S-T 2000 R fallen die Unterschiede in den Koeffizienten zwischen Studie 1 und 2 auf. Hier darf nicht außer Acht gelassen werden, dass die Stichproben eher klein sind ($N_1 = 56$ bzw. $N_2 = 35$) und Schwankungen nicht (ausschließlich) auf die Messungenauigkeit, sondern wohl auf Stichprobenfehler zurückzuführen sind⁶. Die Korrelationen zwischen HMT-S und den verschiedenen Dimensionen des I-S-T 2000 R sind trotz der (auch noch nach der Zusammenfassung) relativ geringen Größe der Stichprobe weitere Belege für die Güte der Kurzversion und bestätigen, dass der HMT-S in der Tat ein Reasoning-Test ist und die Fähigkeit zum schlussfolgernden Denken misst. Inhaltlich bildet die Reasoning-Skala des I-S-T 2000 R im CHC-Modell den Bereich *induction* (I) ab und die Skala numerischen Reasonings ist der Facette *quantitative reasoning* (RQ) zuzuordnen. Die hohen Korrelationen des HMT-S zu diesen Bereichen und geringere Zusammenhänge zu den Bereichen wie Wissen und Merkfähigkeit sind Belege, dass der HMT-S in der Tat eine Operationalisierung von Gf ist, allerdings unter Vernachlässigung deduktiver Fähigkeiten bzw. des *general sequential*

6 Alternative Erklärungen der Unterschiede wie Extremwerte, Fehler bei der Codierung oder bei der Berechnung der Skalenwerte wurden ausgeschlossen.

reasoning (RG, siehe Schneider/McGrew 2012). Es lässt sich dabei in den Daten ein Muster erkennen, welches die Bezeichnung des HMT-S als *numerisch-figuralen* Induktions-/Reasoning-Test rechtfertigt: Die Zusammenhänge sind durchgehend höher für numerische und figurale Skalen im Kontrast zu den Skalen mit verbalen Inhalten. Schneider und McGrew (2012) sehen jedoch Gf als nicht inhaltlich gebunden an, genauso wenig lässt sich aus der Beschreibung der Bereiche *induction* (I) und *quantitative reasoning* (RQ) auf figurale Inhalte schließen (auch wenn dies nicht ausgeschlossen wird). Vielmehr ist auf Ebene der 16 breiten Fähigkeitsbereiche der Faktor *visual processing* (Gv) vorgesehen, der die Fähigkeit repräsentiert „... to make use of simulated mental imagery (often in conjunction with currently perceived images) to solve problems“ (Schneider/McGrew 2012: 129). Die auffällig hohe Korrelation zwischen dem HMT-S und den Würfelaufgaben des I-S-T 2000 R, bei denen Würfel auf Übereinstimmung verglichen werden müssen mit vorgegebenen, gedrehten und gekippten Würfeln, ist vom Aufgabentyp als *visualization* (Vz) zu klassifizieren, dem Kernbereich von Gv (Schneider/McGrew 2012: 129): „The ability to perceive complex patterns and mentally simulate how they might look when transformed (e.g., rotated, changed in size, partially obscured)“.

In diesem Zusammenhang ist bemerkenswert, dass die Assoziation zwischen HMT-S und der figuralen Reasoning-Skala im Wesentlichen auf die Würfelaufgaben und nicht auf die Matrizenaufgaben des I-S-T 2000 R zurückzuführen ist. Diese Korrelation fällt mit $r = .22$ relativ gering aus und spricht gegen die Annahme, dass Zusammenhänge auf oberflächlich betrachtet gleiches Aufgabenmaterial – eben figurale Matrizen – zurückzuführen wären. Dies ist möglicherweise auf Unterschiede in den Aufgabenprinzipien zurückzuführen: Im Gegensatz zum HMT-S werden die in den Matrizen enthaltenen mentalen Operationen nicht vollständig genannt oder an Beispielen verdeutlicht. Hinzu kommt, dass insgesamt mehr Regeln bei den I-S-T 2000 R Matrizen zur Anwendung kommen. Nach unserer Auffassung werden damit weitere Intelligenzkomponenten angesprochen (divergentes Denken im Sinne von Einfallsreichtum nach Jäger 1982), die beim HMT-S irrelevant sind und daher die geringe Höhe der Korrelation erklären. Neben der Überprüfung der theoretischen Einordnung einzig durch den I-S-T 2000 R als objektiven Leistungstest sollten zukünftig im Prozess einer fortgeführten Validierung Bezüge zu anderen mehrdimensionalen Testverfahren wie dem Berliner-Intelligenzstruktur-Test (Jäger/Süß/Beauducel 1997) oder dem Wilde-Intelligenz-Test 2 (Kersting/Althoff/Jäger 2008) empirisch geprüft werden.

Auch die Resultate zu den subjektiven Intelligenzselbsteinschätzungen mittels ISI bestätigen die Einordnung des HMT-S als figural-numerischen Reasoning-Test: positive Korrelationen resultierten in den Bereichen mathematische Intelli-

genz, räumliche Intelligenz und logisches Denken. Die Tendenz, dass der HMT-S leicht negativ assoziiert ist mit der körperlich-kinästhetischen, mit der interpersonellen und mit der intrapersonalen Intelligenz könnte auf intraindividuelle Kontrastierung zurückzuführen sein. Erwartungskonform resultierten zudem Zusammenhänge zu den mathematischen und akademischen Bereichen der im Selbstbericht erhobenen Problemlöseerwartungen (SWE), zu denen der Problemlöseerfahrungen (abgebildet in den Selbstkonzepten) und zu denen der Hilflosigkeit. Zur Leistungsmotivation zeichnet sich ein geringer Zusammenhang ab. Diese sind zwar unerwünscht (Stern, 1911), jedoch bei Intelligenztestungen kaum zu vermeiden (vgl. Conrad 1983).

Die Kriteriumsvalidität des HMT-S wird durch die Zusammenhänge zu akademischen Erfolgsmaßen belegt. Es zeichnen sich Zusammenhänge zu Fächern ab, in denen logisches und schlussfolgerndes Denken gefordert ist (im Speziellen sind dies Mathematik bzw. Methoden und Statistik). Geringe Effekte resultieren bezüglich der allgemeineren Maße des akademischen Erfolgs, d.h. für Schulabschlussniveau, Schulabschlussnote und Durchschnittsnote im Studium. Erklärbar ist dies durch die Aggregation von Noten verschiedener Fächer, die unterschiedliche Anforderungen stellen und für die logisch-schlussfolgerndes Denken weniger relevant ist (z.B. für die Schulfächer Sport oder Kunst). Es darf bei den Korrelationen mit den Schulnoten nicht außer Acht gelassen werden, dass a) das zeitliche Intervall zwischen der Testteilnahme am HMT-S und dem Schulabschluss bei den im Durchschnitt über dreißigjährigen Studierenden über zehn Jahre zurück liegt und die Schulnoten nicht in allen Fällen korrekt erinnert werden und b) keine perfekte Stabilität des erfassten Konstrukts vorliegt. Da die Stichproben der drei berichteten Studien nicht repräsentativ und relativ homogen sind, sollten zur weiteren Validierung Untersuchungen z.B. mit Schülerinnen und Schülern stattfinden, deren aktuelle Noten ggf. objektiv und nicht im Selbstbericht erfasst werden könnten. Auch Studierende anderer Fachrichtungen könnten zukünftige Stichproben bilden, wobei insgesamt eine ausgewogenere Geschlechterverteilung anzustreben wäre.

Unser Fazit zum HMT-S fällt auch unter Berücksichtigung der Einschränkungen der Studien aufgrund relativ homogener und bezüglich der Zusammenhangsmaße mit dem I-S-T 2000 R eher kleinen Stichproben insgesamt positiv aus. Der HMT-S ist ein höchst ökonomischer, kostenloser Online-Intelligenztest, der figural-numerisches Reasoning erfasst. Da nicht nur die Korrelation zur Originalversion sehr hoch, sondern auch das Muster der Validitätskoeffizienten mit dem des HMT vergleichbar ist, kann davon ausgegangen werden, dass der HMT-S den HMT auch auf inhaltlicher Ebene abbildet. Die Reliabilität des HMT-S ist trotz der niedrigen Anzahl von insgesamt nur sechs Aufgaben zufriedenstellend und für Grup-

penuntersuchungen geeignet. Die in den drei Studien gesammelten Belege für die Konstrukt- und die Kriteriumsvalidität sind vielversprechend und wir erhoffen uns zukünftig eine weite Verbreitung und Etablierung des HMT-S. Angefordert werden kann der HMT-S über die URL <http://HMT.de.lv>⁷, über die auch weitere Informationen zugänglich sind.

Literatur

- Aiken, L. R. und G. Groth-Marnat, 2006: *Psychological Testing and Assessment* (12. Auflage). Boston, MA: Pearson Education.
- Amthauer, R., B. Brocke, D. Liepmann und A. Beauducel, 2001: I-S-T 2000 R. Intelligenz-Struktur-Test 2000 R. Göttingen: Hogrefe.
- Blickle, G., J. Kramer und J. Mierke, 2010: Telephone-Administered Intelligence Testing for Research in Work and Organizational Psychology. A Comparative Assessment Study. *European Journal of Psychological Assessment* 26: 154-161.
- Buchwald, F., C. Spoden, J. Fleischer und D. Leutner, 2013: Verzweigte Lernumgebungen und Tests mit EFS Survey 8. *Diagnostica* 59: 113-117.
- Conrad, W., 1983: Intelligenzdiagnostik. S. 104-201 in: K.-J. Groffmann und L. Michel (Hg.): *Intelligenz- und Leistungsdiagnostik, Enzyklopädie der Psychologie (Themenbereich B, Serie II, Band 2)*. Göttingen: Hogrefe.
- Carroll, J. B., 1993: *Human Cognitive Abilities: A Survey of Factoranalytic Studies*. Cambridge, NY: University Press.
- Carroll, J. B., 2005: The Three-Stratum Theory of Cognitive Abilities. S. 69-76 in: D. P. Flanagan und L. Harrison (Hg.): *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (2. Auflage). New York, NY: Guilford Press.
- Cattell, R. B., 1957: *Personality and Motivation Structure and Measurement*. New York, NY: World Book.
- Cattell, R. B., 1987: *Intelligence: Its Structure, Growth and Action*. Amsterdam: Elsevier.
- Davidson, J. E. und I. A. Kemp, 2011: Contemporary Models. S. 58-82 in: R. J. Sternberg und S. B. Kaufman (Hg.): *The Cambridge Handbook of Intelligence*. Cambridge, NY: University Press.
- Deutsche Gesellschaft für Psychologie & Bundesverband Deutscher Psychologinnen und Psychologen, 2004: *Ethische Richtlinien der DGPs und des BDP*. <http://www.dgps.de/dgps/aufgaben/003.php>
- DIN, 2002: DIN 33430: *Anforderungen an Verfahren und deren Einsatz bei berufsbezogenen Eignungsbeurteilungen*. Berlin: Beuth.
- Evans, J. St. B. T., 2005: Deductive Reasoning. S. 169-184 in: K. J. Holyoak und R. G. Morrison (Hg.): *The Cambridge Handbook of Thinking and Reasoning*. Cambridge, NY: University Press.
- Gardner, G., 1983: *Frames of Mind: The Theory of Multiple Intelligences*. New York, NY: Basic Books.
- Gnams, T., B. Batinic und G. Hertel, 2011: Internetbasierte psychologische Diagnostik. S. 448-498 in: L. F. Hornke, M. Amelang und M. Kersting (Hg.): *Verfahren zur Leistungs-*

7 Bei der URL handelt es sich um eine Weiterleitung zu http://www.fernuni-hagen.de/psychologie/mde/forschung/hot_project.shtml

- Intelligenz- und Verhaltensdiagnostik, Enzyklopädie der Psychologie, Psychologische Diagnostik (Bd. II/3). Göttingen: Hogrefe.
- Guilford, J. P., 1967: *The Nature of Intelligence*. New York, NY: McGraw-Hill.
- Guttman, L., 1965: A Faceted Definition of Intelligence. S. 166-181 in: R. Eiferman (Hg.): *Studies in Psychology, scripta hierosolymitana* (Vol. 14). Jerusalem: Hebrew University.
- Guttman, L. und S. Levy, 1991: Two Structural Laws for Intelligence Tests. *Intelligence* 15: 79-103.
- Gottfredson, L. S., 1997: Mainstream Science on Intelligence: An Editorial with 52 Signatories, History, and Bibliography. *Intelligence* 24: 13-23.
- Gottfredson, L. S. und D. H. Saklofske, 2009: Intelligence: Foundations and Issues in Assessment. *Canadian Psychology* 50: 183-195.
- Harrell, T. W. und M. S. Harrell, 1945: Army General Classification Test Scores for Civilian Occupations. *Educational and Psychological Measurement* 5: 229-239.
- Heydasch, T., K.-H. Renner, J. Haubrich, B. Hilbig und I. Zettler, 2013: A Free Web-Based Intelligence Test: The Hagen Matrices Test (HMT). Manuskript in Vorbereitung.
- Horn, J. L. und A. N. Blankson, 2012: Foundations for Better Understanding of Cognitive Abilities. S. 73-98 in: D. Flanagan und P. Harrison (Hg.): *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (3. Auflage). New York, NY: Guilford Press.
- Horn, J. L. und J. Noll, 1997: Human Cognitive Capabilities: Gf-Gc Theory. S. 53-91 in: D. P. Flanagan, J. L. Genshaft und P. L. Harrison (Hg.): *Contemporary Intellectual Assessment: Theories, Tests, and Issues*. New York, NY: Guilford Press.
- Hülsheger, U. R., G. W. Maier und T. Stumpp, 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.
- Hunter, J. E. und R. F. Hunter, 1984: Validity and Utility of Alternative Predictors of Job Performance. *Psychological Bulletin* 96: 72-98.
- Jäger, A. O., 1982: Mehrmodale Klassifikation von Intelligenzleistungen: Experimentell kontrollierte Weiterentwicklung eines deskriptiven Intelligenzstrukturmodells. *Diagnostica* 28: 195-225.
- Jäger, A. O., H.-M. Süß und A. Beauducel, 1997: *Berliner Intelligenzstruktur - Test. Form 4*. Göttingen: Hogrefe.
- Jensen, A., 1998: *The g Factor: The Science of Mental Ability*. Westport, CT: Praeger.
- Jerusalem, M. und R. Schwarzer, 1986: Fragebogen zur Erfassung von Hilflosigkeit. In: R. Schwarzer (Hg.): *Skalen zur Befindlichkeit und Persönlichkeit* (Forschungsbericht 5). Berlin FU Berlin, Institut für Psychologie, Abt. Pädagogische Psychologie.
- Jerusalem, M. und R. Schwarzer, 2012: Dimensionen der Hilflosigkeit. In: A. Glöckner-Rist (Hg.): *Zusammenstellung sozialwissenschaftlicher Items und Skalen*. ZIS Version 15.00. Bonn: GESIS.
- Judge, T. A., C. A. Higgins, C. J. Thoresen und M. R. Barrick, 1999: The Big Five Personality Traits, General Mental Ability, and Career Success Across the Life Span. *Personnel Psychology* 52: 621-652.
- Kersting, M., K. Althoff und A. O. Jäger, 2008: *WIT-2. Der Wilde-Intelligenztest*. Göttingen: Hogrefe.
- Khandaker, G. M., J. H. Barnett, I. R. White und P. B. Jones, 2011: A Quantitative Meta-Analysis of Population-Based Studies of Premorbid Intelligence and Schizophrenia. *Schizophrenia Research* 132: 220-227.
- Kuder, G. F. und M. W. Richardson, 1937: The Theory of the Estimation of Test Reliability. *Psychometrika* 2: 151-160.

- Kuncel, N. R., S. A. Hezlett und D. S. Ones, 2004: Academic Performance, Career Potential, Creativity, and Job Performance: Can One Construct Predict Them All? *Journal of Personality and Social Psychology* 86: 148-161.
- Kunina, O., O. Wilhelm, M. Formazin, K. Jonkmann und U. Schroeders, 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. W. B. und S. Fries, 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.
- Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID) (Hg.), 2012: Verzeichnis Testverfahren. Kurznamen. Langnamen. Autoren. Testrezensionen (19., aktualisierte Auflage). Trier: ZPID.
- Liepmann, D., A. Beauducel, B. Brocke und R. Amthauer, 2007: Intelligenz-Struktur-Test 2000 R (I-S-T 2000 R). Manual (2. Auflage). Göttingen: Hogrefe.
- McGrew, K. S., 1997: Analysis of the Major Intelligence Batteries According to a Proposed Comprehensive Gf-Gc Framework. S. 151-179 in: D. P. Flanagan, J. L. Genshaft und P. L. Harrison (Hg.): *Contemporary Intellectual Assessment: Theories, Tests, and Issues*. New York: Guilford Press.
- McGrew, K. S., 2005: The Cattell Horn Carroll (CHC) Theory of Cognitive Abilities. Past, Present and Future. S. 136-202 in: D. P. Flanagan und P. L. Harrison (Hg.): *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (2. Auflage). New York, NY: Guilford Press.
- Ng, T. W. H., L. T. Eby, K. L. Sorensen und D. C. Feldman, 2005: Predictors of Objective and Subjective Career Success: A Meta-Analysis. *Personnel Psychology* 58: 367-408.
- Pawlik, K., 2006: Psychologische Diagnostik I: Methodische Grundlagen. S. 555-562 in: K. Pawlik (Hg.): *Handbuch Psychologie. Wissenschaft-Anwendung-Berufsfelder*. Berlin: Springer.
- Poropat, A. E., 2009: A Meta-Analysis of the Five-Factor Model of Personality and Academic Performance. *Psychological Bulletin* 135: 322-338.
- Rammstedt, B. und T. Rammsayer, 2002: Die Erfassung der selbsteingeschätzten Intelligenz: Konstruktion, teststatistische Überprüfung und erste Ergebnisse des Inventars zur selbsteingeschätzten Intelligenz (ISI). *Zeitschrift für Differentielle und Diagnostische Psychologie* 23: 435-466.
- Rindermann, H. und A. Neubauer, 2000: Informationsverarbeitungsgeschwindigkeit und Schulerfolg: Weisen basale Maße der Intelligenz prädiktive Validität auf? *Diagnostica* 46: 8-17.
- Roberts, B. W., N. R. Kuncel, R. Shiner, A. Caspi und L.R. Goldberg, 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.
- Salgado, J. F., N. Anderson, S. Moscoso, C. Bertua und F. de Fruyt, 2003: International Validity Generalization of GMA and Cognitive Abilities: A European Community Meta-Analysis. *Personnel Psychology* 56: 573-605.
- Schiefele, U., B. Moschner und R. Husstegge, 2002: *Skalenhandbuch SMILE-Projekt* (unveröffentlichtes Manuskript). Bielefeld: University.
- Schmidt, F. L. und J. E. Hunter, 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.

- Schmidt-Atzert, L., 2002. Intelligenz-Struktur-Test 2000 R (Testrezension). *Zeitschrift für Personalpsychologie* 1: 50-56.
- Schmidt-Atzert, L. und W. Rauch, 2008: Intelligenz-Struktur-Test 2000 R (I-S-T 2000 R). 2., erweiterte und überarbeitete Auflage [TBS-TK Rezension]. *Report Psychologie* 33: 303-304.
- Schneider, W. J. und K. S. McGrew, 2012: The Cattell-Horn-Carroll Model of Intelligence. S. 99-144 in: D. Flanagan und P. Harrison (Hg.): *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (3. Auflage). New York, NY: Guilford Press.
- Schwarzer, R. und M. Jerusalem, 1995: Generalized Self-Efficacy Scale. S. 35-37 in: J. Weinman, S. Wright und M. Johnston (Hg.): *Measures in Health Psychology: A User's Portfolio. Causal and Control Beliefs*. Windsor: NFER-Nelson.
- Seligman, M. E. P., 1975: *Helplessness: On Depression, Development and Death*. San Francisco, CA: Freeman.
- Silverstein, A. B., 1990: Short Forms of Individual Intelligence Tests. *Psychological Assessment: A Journal of Consulting and Clinical Psychology* 2: 3-11.
- Sloman, S. A. und D. A. Lagnado, 2005: The Problem of Induction. S. 95-116 in: K. J. Holyoak und R. G. Morrison (Hg.): *The Cambridge Handbook of Thinking and Reasoning*. Cambridge, NY: University Press.
- Spearman, C., 1904: The Proof and Measurement of Association between Two Things. *The American Journal of Psychology* 15: 72-101.
- Stern, W., 1911: *Die Differentielle Psychologie in ihren methodischen Grundlagen*. Leipzig: Johann Ambrosius Barth.
- Sternberg, R. J., 2005: Intelligence. S. 751-773 in: K. J. Holyoak und R. G. Morrison (Hg.): *The Cambridge Handbook of Thinking and Reasoning*. Cambridge, NY: University Press.
- Strenze, T., 2007: Intelligence and Socioeconomic Success: A Meta-Analytic Review of Longitudinal Research. *Intelligence* 35: 401-426.
- Süß, H.-M. und A. Beauducel, 2004: Faceted Models of Intelligence. S. 313-332 in: O. Wilhelm und R. W. Engle (Hg.): *Handbook of Understanding and Measuring Intelligence*. Thousand Oaks, CA: Sage.
- Wasserman, J. D., 2012: A History of Intelligence Assessment: The Unfinished Tapestry. S. 3-55 in: D. P. Flanagan und P. L. Harrison (Hg.): *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (3. Auflage). Cambridge, NY: University Press.
- Willis, J. O., R. Dumont und A. S. Kaufman, 2011: Factor-Analytic Models of Intelligence. S. 39-57 in: R. J. Sternberg und S. B. Kaufman (Hg.): *The Cambridge Handbook of Intelligence*. Cambridge, NY: University Press.
- Ziegler, M., E. Dietl, E. Danay, M. Vogel und M. Bühner, 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.

Anschrift der Autoren	Timo Heydasch FernUniversität in Hagen Universitätsstr. 33 58084 Hagen Timo.Heydasch@FernUni-Hagen.de
Ko-Autor/-in	Julia Haubrich FernUniversität in Hagen Karl-Heinz Renner FernUniversität in Hagen