



AKTUELLE THEMEN UND ENTWICKLUNGEN IN DER MANAGEMENTDIAGNOSTIK

Ein Werkstattbericht der Projektgruppe
Eignungsdiagnostik im TOP- Managementdiagnostik

Forum Assessment e.V.

Werkstattbericht am 21. Mai 2026

DIE PROJEKTGRUPPE

Eignungsdiagnostik im Top-Management



In der Projektgruppe tauschen sich Praktiker:Innen zu aktuellen Themen und den speziellen Herausforderungen im Arbeitsfeld der Eignungsdiagnostik im Top-Management aus.

AGENDA

- 01.** Zwischen Ritterschlag und Damoklesschwert:
Diagnostik als (un)geplante Intervention
Constanze Kuptsch und Silvana von Hayn
- 02.** KI als Assistenz in der Managementdiagnostik
Richard Hossiep & Olaf Ringelband
- 03.** Update zur aktuellen Debatte über Prädiktoren der
Managementdiagnostik
Rüdiger Hossiep & Dieter Hasselmann



TN-Unterlage in der APP



<https://www.youtube.com/watch?v=OwCojQCAqyg>

3

THEMENBLOCK

Update zur aktuellen Debatte
über Prädiktoren der
Managementdiagnostik

REFERENTINNEN



Dr. Dieter Hasselmann

Diplom-Psychologe

CEO hsp consulting GmbH

Institut für Management Diagnostik

Beratungsschwerpunkt ist das Besetzen
von Positionen im oberen Management

Mitautor u. a. der Standards für Diagnostik
im Top-Management sowie der aktuellen
AC- und Interview Standards



Dr. Rüdiger Hossiep

Diplom-Psychologe

seit 1984 im Training, Coaching und der
Potenzialklärung gemeinsam mit Führungskräften

seit 1990 an der Ruhr-Universität Bochum

Mitherausgeber der Buchreihe „Praxis der
Personalpsychologie - HRM kompakt“

Mitglied im Wissenschaftlichen Beirat des Forum
Assessment

für seine herausragenden Arbeiten auf dem Gebiet
der angewandten psychologischen Diagnostik mit
dem Alfred-Binet-Preis ausgezeichnet

DIE WISSENSCHAFTLICHE DEBATTE VALIDITÄT KOGNITIVER LEISTUNGSTESTS – METAANALYTISCHE TAXIERUNGEN (1/5)

- Ausgangspunkt: Schmidt & Hunter (1998, .51)
- Revidierte Methodik von Sackett et al. (2022, .31) // (2024, .22 bzw. .38 (für objektive Maße))
- Replik von Ones und Viswesvaran (2023):
 - Varianzeinschränkungen nur bei prädiktiven Studien, nicht aber bei konkurrenten Studien anzunehmen, sei auf Grund der Datenlage nicht haltbar
 - Das Vorgehen von Sackett et al. sei zu konservativ:
 - **“Absence of evidence in narrative study descriptions is not evidence of absence”** (S. 359)
 - Bei der Vielzahl an möglichen direkten und indirekten Varianzeinschränkungen in Prädiktor und Kriterium werde eher zu wenig als zu viel korrigiert:
 - **“Any single form of range restriction correction is likely to be dwarfed by the multiple types of range restriction that we fail to correct in empirical estimations of operational validity.”** (S. 360)

Ones & Viswesvaran, 2023

VALIDITÄT KOGNITIVER LEISTUNGSTESTS – METAANALYTISCHE TAXIERUNGEN (2/5)

- 2025 folgte ein weiterer Grundsatzartikel zum Thema, der mit einem Zitat von Schmidt (2002, S. 187) beginnt und auf diese Weise einen – für wissenschaftliche Auseinandersetzungen – ungewohnt scharfen Ton anschlägt:

“If we want to remain a science-based field, we cannot reject what we know to be true in favor of what we would like to be true.”

- Sie führen an, dass das Vorgehen von Sackett et al. in diversen Artikel kritisiert wurde, damit lediglich eine „**line of literature**“ (S. 12) darstelle und **bei weitem kein „scientific consensus** on the ‘lower than previously thought’ validity of cognitive tests” (S. 11) **bestehe**:

*“Sackett et al.’s .31 estimate **rests on methodological and analytical choices, various assumptions, and propositions that have been questioned**” (S. 3)*

Kulikowski et al., 2025

VALIDITÄT KOGNITIVER LEISTUNGSTESTS – METAANALYTISCHE TAXIERUNGEN (3/5)

- Sacketts Schätzer sei vor allem deshalb so niedrig, da von ihnen angenommen werde, dass es in konkurrenten Designs keine Range Restrictions geben könne. Jedoch sei das Gegenteil der Fall:

*“empirical evidence has shown that the **degree of RR** (and observed validity) **is similar between predictive and concurrent designs**, particularly for cognitive ability tests” (S. 5)*

*“**cognitive ability test scores** tend to be the **most restricted** among major selection procedures” (S. 5)*

- Weiterhin kritisieren sie, dass sich Sackett et al. **nur auf eine Operationalisierung des Kriteriums „Job Performance“** stützen: Supervisory Rating der Job Performance. **Dies greife zu kurz**, da es diverse Arten gibt, um Leistung zu operationalisieren. So bestehen **substanzielle Zusammenhänge** zu beispielsweise Job Knowledge Tests und Arbeitsproben, **die auch dann hoch ausfallen** (.56 und .49), **wenn die konservativen Annahmen von Sackett et al. verwendet werden** (bare bones .49 und .43)

Kulikowski et al., 2025

VALIDITÄT KOGNITIVER LEISTUNGSTESTS – METAANALYTISCHE TAXIERUNGEN (3/5)

- Auch Bobko et al. kritisieren den Ansatz von Sackett scharf:

*“The conservative estimation approach in Sackett et al. (2022) **has several major logical and statistical flaws** (comparing correlations that have been differentially corrected for RR; method-construct confounds, etc.). The approach is **both conceptually and statistically flawed** and, if adopted, would have important implications for organizational research and practice” (S. 9)*

- Sie gehen auf Grundlage einer umfangreichen Reanalyse der Daten von Salgado et al. (2003) von einer prädiktiven Validität in folgender Höhe aus:

*„close to **0.50, or possibly even in the 0.60’s**” (S. 9)*

- Sie schlagen deshalb vor, sich (z. B.) in Metaanalysen **unbedingt tiefer mit den jeweiligen Daten** und Einzelstudien zu befassen und **nicht einfach konservativ vorzugehen**. Sie nennen dieses Vorgehen „**considered estimation**“ (anstatt „conservative estimation“), um die Validität so akkurat und realistisch wie möglich zu bestimmen.

Bobko et al., 2025

VALIDITÄT KOGNITIVER LEISTUNGSTESTS – METAANALYTISCHE TAXIERUNGEN (4/5)

- Ones und Viswesvaran empfehlen für eine **möglichst akkurate Schätzung der Validität die Metaanalyse von Hunter et al. (2006)**, in der eine umfassende Analyse hinsichtlich der vorliegenden kumulativen Varianzeinschränkungen vorgenommen wurde (siehe nächste Folie)
- Hieraus folgt, dass die prädiktive Validität von kognitiver Leistungsfähigkeit noch deutlich höher als gemeinhin angenommen sein könnte – **insbesondere auf Ebene des Managements**

*“Many researchers maintain that being conservative is good science, but **conservative estimates are by definition biased estimates**. We believe it is more appropriate **to aim for unbiased estimates because the research goal is to maximize the accuracy of the final estimates**”*

(Viswesvaran et al., 1996, S. 567, zitiert in Ones & Viswesvaran, 2023, S. 363).

Ones & Viswesvaran, 2023

VALIDITÄT KOGNITIVER LEISTUNGSTESTS – METAANALYTISCHE TAXIERUNGEN (5/5)

Table 1. Validity of cognitive ability for overall job performance from the US Employment Service Database

	<i>k</i>	Observed validities		Bare bones meta-analysis validities (only sampling error addressed)		Operational validities (direct range restriction corrected for) ^g		Operational validities (indirect range restriction corrected for) ^{gh}	
		Mean <i>r</i>	<i>SD_r</i>	Mean <i>r</i>	<i>SD_ρ</i>	Mean <i>ρ</i>	<i>SD_ρ</i>	Mean <i>ρ</i>	<i>SD_ρ</i>
Professional and managerial jobs ^a	36	0.32	0.16	0.32	0.08	0.58	0.09	0.74	0.04
Complex setting up jobs ^b	17	0.32	0.13	0.32	0.01	0.58	0.02	0.73	0.00
Technician and skilled jobs ^c	151	0.26	0.15	0.26	0.08	0.49	0.09	0.66	0.05
Semiskilled jobs ^d	201	0.21	0.13	0.21	0.01	0.40	0.02	0.56	0.00
Unskilled jobs ^e	20	0.13	0.14	0.13	0.05	0.25	0.06	0.39	0.06
Non-manufacturing jobs ^f	204	0.30	0.15	0.30	0.07	0.55	0.07	0.71	0.04

(Daten aus Hunter, Schmidt und Le, 2006 – Ones und Viswesvaran unterstreichen die Korrektheit des methodischen Vorgehens)

Ones & Viswesvaran, 2023, S. 363

VALIDITÄT: KOMBINATION VON INTELLIGENZ UND PERSÖNLICHKEIT

- Intelligenz ist der wichtigste Prädiktor für Erfolg (für „important life outcomes“) – aber fluide Intelligenz korreliert negativ mit dem Alter
- Wie kann es sein, dass dennoch die meisten wichtigen Positionen in Unternehmen und Politik mit Personen um die 60 besetzt werden?
- Hypothese: Eine Kombination aus Intelligenz (kristalline und fluide), Persönlichkeit im engeren Sinne (Big Five) und im weiteren Sinne (emotional intelligence, financial literacy, moral reasoning, resistance to sunk cost bias, cognitive flexibility, cognitive empathy, need for cognition) kann als Erklärungsansatz dienen.
- Daten stützen diesen Erklärungsversuch:

*“Accordingly, individuals best suited for **high-stakes leadership, judgment, or executive roles are likely to be between 55 and 60** – and unlikely to **be younger than 40 or older than 65.**” (S. 13)*

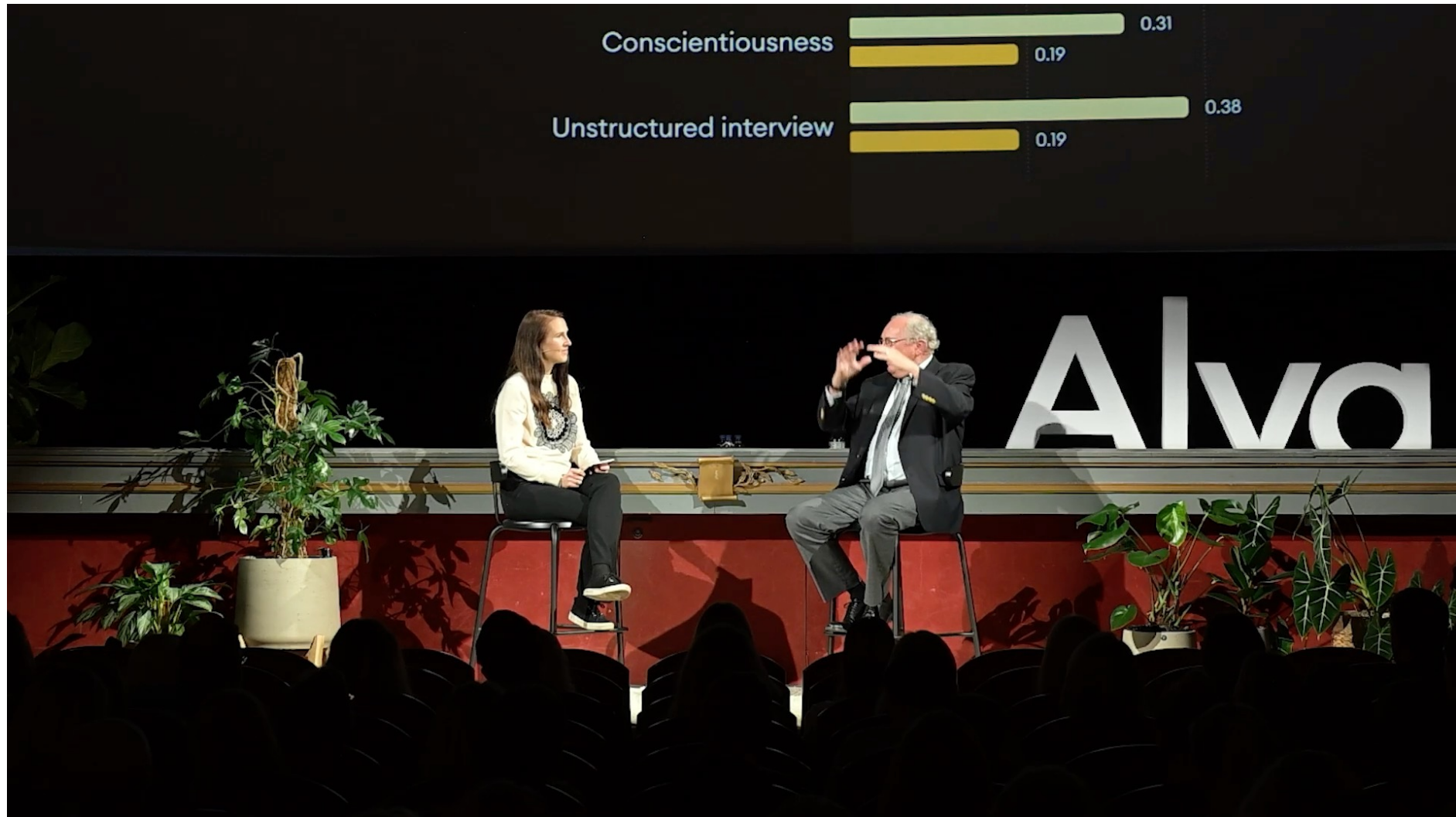
Gignac & Zajenowski, 2025



3

THEMENBLOCK

Ein praxisbezogener Blick
auf diese Debatte



<https://www.youtube.com/watch?v=OwCojQCAqyg>

„INTELLIGENZ IST ZENTRALER PRÄDIKTOR“

- spätestens seit der Meta-Analyse von Schmidt & Hunter (1998) zählt es zu den zentralen Grundsätzen der Eignungsdiagnostik, dass (GMA-)Intelligenztestskalen positionsübergreifend zu den Methoden mit der höchsten prädiktiven Validität zählen
- dies gilt, nach Auffassung zahlreicher namhafter Autoren, uneingeschränkt oder sogar besonders auch für die Eignungsdiagnostik im (Top-)Management
- überraschend ist die absolute Höhe der Validitätsschätzungen nahe am theoretischen Maximum (bei Reliabilitäten von 0,9 von GMA-Skalen und 0,5 - 0,6 für typische Kriterien)
- die hohen Werte haben z. T. zu einer Fokussierung auf den Prädiktor „Intelligenz“ geführt, die aus praktischer Sicht nicht zielführend ist

Schmidt, Oh & Shaffer, 2016 Selection procedures/predictors	Operational validity (<i>r</i>)
1. GMA tests ^a	.65
2. Integrity tests ^b	.46
3. Employment interviews (structured) ^c	.58
4. Employment interviews (unstructured) ^d	.58
5. Interests ^e	.31
6. Phone-based interviews (structured) ^f	.46
7. Conscientiousness ^g	.22
8. Reference checks ^h	.26
9. Openness to Experience ⁱ	.04
10. Biographical data ^j	.35
11. Job experience (years) ^k	.16
12. Personality-based EI ^l	.32

„Intelligenztests für Manager: ein Muss“

„Auch wenn so mancher es nicht recht glauben mag, es hilft einem Unternehmen durchaus, wenn an der Spitze Menschen stehen, die nicht nur stark auftreten, sondern auch intelligent sind“.

„Intelligenz braucht man zum Überleben in der Vuca-Welt.“

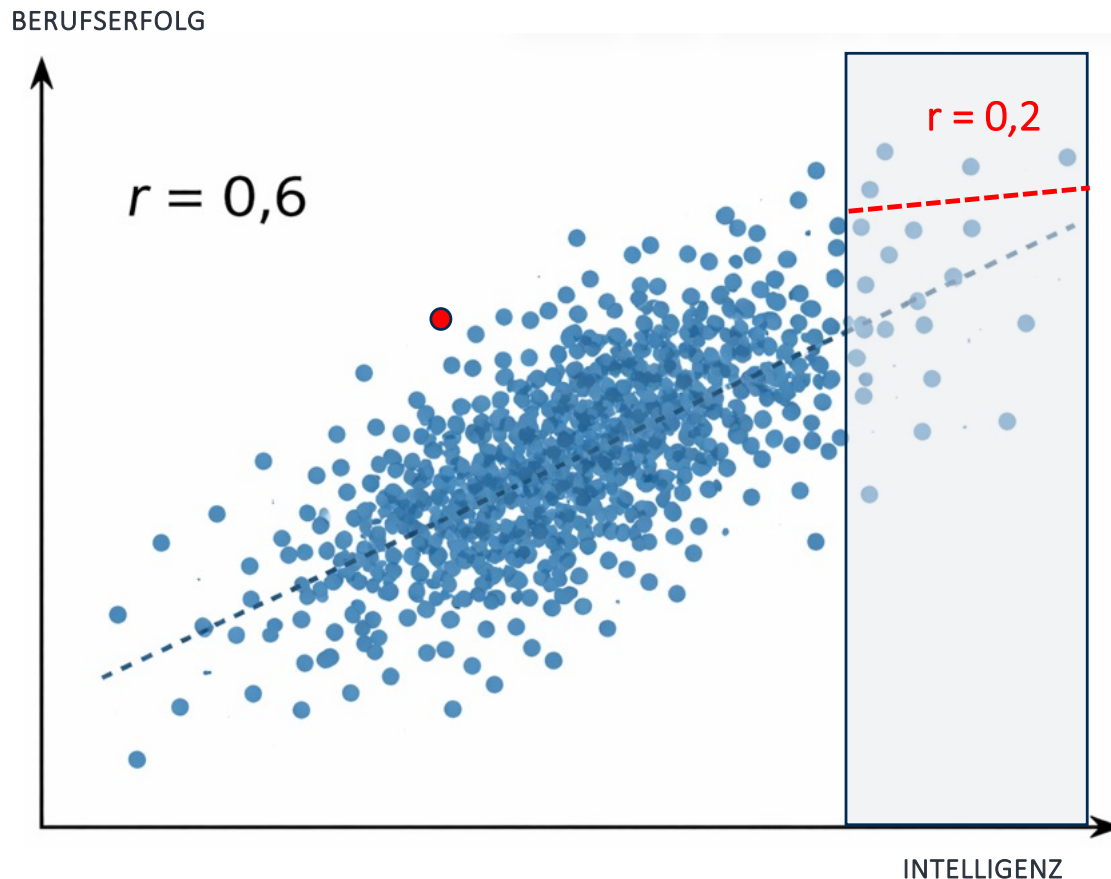
KONTROVERS IST DER UMFANG DER VARIANZAUFKLÄRUNG

Frage	Schmidt & Hunter	Sackett	Ones
Was soll geschätzt werden?	„wahre“ Validität	beobachtbare Validität	„wahre“ Validität
Umgang mit Range Restriction	stark korrigieren	vorsichtig / reduziert	korrigieren, aber präziser
Ergebnis für GMA	hoch (>.51)	moderat (~.22–.31)	eher höher als Sackett

- Konsens besteht, dass Fähigkeiten wie
 - logisch-analytisch schlussfolgerndes Denken
 - Lernfähigkeit
 - Problemlösefähigkeit Mustererkennung
 - Verarbeitungsgeschwindigkeit
 zentrale anforderungsrelevante Faktoren bzw. notwendige Erfolgsvorausfaktoren sind
- Warum ist die Größe des Kennwertes überhaupt praktisch relevant?
 - für Auswahl der (notwendigen) Verfahrenselemente
 - für das Festlegen von Gewichtungsfaktoren in der Ergebnisfindung
- Es ist wichtig zu beachten, was die Fragestellung ist:
 - will ich abschätzen, wie stark Managementenerfolg von Intelligenz abhängt (wie Schmidt & Hunter)
 - oder wie groß der prognostische Beitrag von GMA-Testwerten im konkreten Anwendungsfall ist

DIE BEDEUTUNG DER VARIANZEINSCHRÄNKUNG

Schmidt & Hunter rechnen die Varianzeinschränkungen aus den Korrelationen heraus, weil sie den „wahren Zusammenhang“ schätzen wollen



In der Managementdiagnostik, hat man/frau es jedoch mit Gruppen von Teilnehmenden zu tun, bei denen bzgl. des Prädiktors (GMA-Skalen) eine erhebliche Varianzeinschränkung besteht.

In der Konsequenz wird der konkrete prognostische Validitätsbeitrag dieses Prädiktors für die Auswahlentscheidung im praktischen Anwendungsfall deutlich geringer sein.

„GMA-INTELLIGENZTEST-SCORES SIND KEIN HINREICHENDER PRÄDIKTOR“

- **Erfahrung ist wichtig - weniger die Dauer, aber die Qualität!**

GMA-Testskalen messen fluide Intelligenz, im Management haben jedoch auch kristalline Intelligenzfaktoren eine große Bedeutung (auch in der Vuca-Welt).

Manager:innen erwerben erfahrungsabhängig

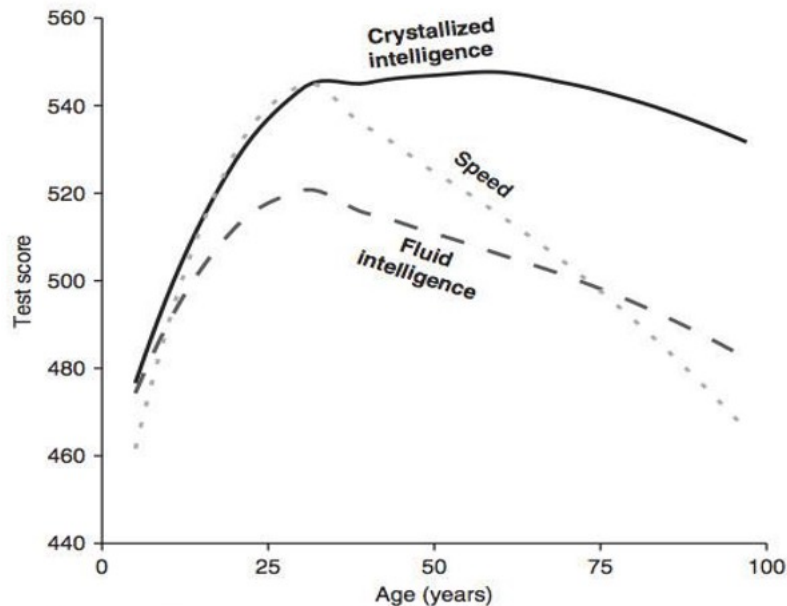
- Handlungswissen (zu möglichen Vorgehensweisen, Methoden etc.)
- Wissen zur Mustererkennung
- Systemkenntnisse zum Funktionieren der Organisationen
- Heuristiken, z. B. Umschalten von strategischer Analyse zum handlungsorientierten iterativen Vorgehen

In der Bearbeitung entsprechender Fallstudien oder mittels geeigneter Interviewfragen lassen sich diese Aspekte beurteilen und auch die Anwendung auf neue bzw. unbekannte Situationen beobachten.

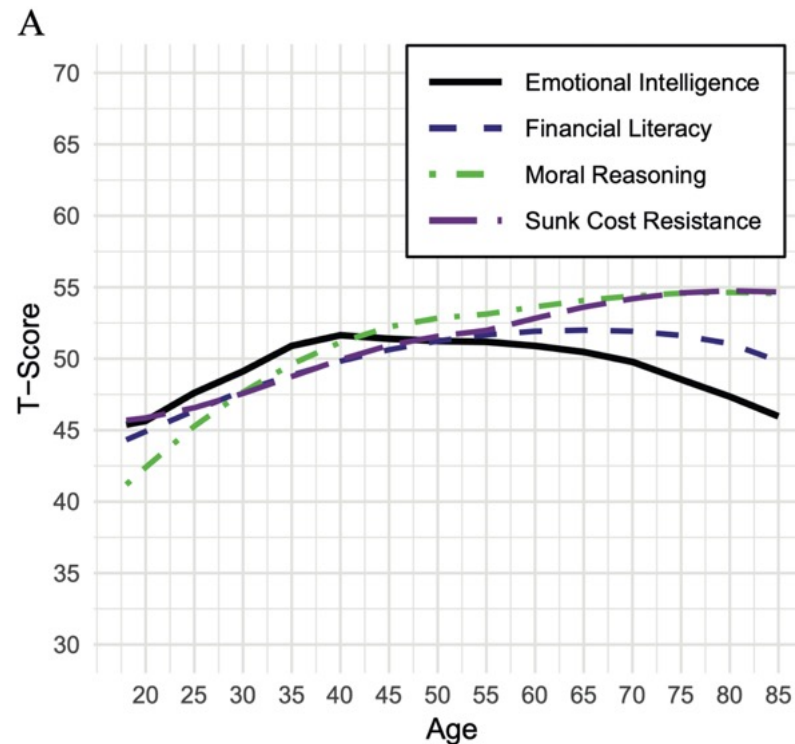
GMA-Testscores allein sind nicht geeignet, die Leistung in Situationen vorherzusagen, in denen diese Faktoren relevant sind. Sie haben jedoch umso größeren Wert, wenn es darum geht, Potenzial einzuschätzen.



WANN WIRD DIE HÖCHSTE LEISTUNGSFÄHIGKEIT ERREICHT?



▲ Figure 2.4 Change in cognitive abilities with age, taken from a study testing over 6,000 people of all ages (the data for different ages come from different people). 'Fluid' skills (dashed line) decline far more steeply than 'crystallized' ones (solid line). Tests of speed (dotted line) show a particularly pronounced decline. (Figure adapted with permission from Tucker-Drob, 2009.)



Güllich et al, 2025

In Top-Managementpositionen liegt der Peak der Leistungsfähigkeit eher im Bereich von 40 - 55 Jahren, abhängig von den jeweils kritischen Erfahrungsfeldern.

Ältere CEO's erzielen z. B. bessere Ergebnisse bzgl. finanzieller Stabilität & Performance, jüngere dagegen bzgl. mehr Innovationskraft und einer höheren Risikoneigung.

CHANCEN UND PRAKTISCHE HERAUSFORDERUNGEN

Chancen der Nutzung von GMA-Skalen

- Messung ist wesentlich genauer als mit Einschätzungen durch Beobachter ($r=0,9$ vs. $r=0,5-0,6$)
- fluide Intelligenzleistungen lassen sich (kontextunabhängig) messen, um Anpassungs- / Lernfähigkeit einzuschätzen
- Defizite bzgl. der kognitiven Leistungsfähigkeit sind auf Basis von Testergebnissen leichter zu argumentieren

Herausforderungen der Nutzung von GMA-Skalen

- diese sind für erfahrene Manager:innen wenig augenscheinlich (auch nicht, wenn sie betriebswirtschaftliche Inhalte nutzen)
- sie sind für Testpersonen aus dem oberen und Top-Management potenziell statusbedrohend
- Managementdiagnostik ist Individualdiagnostik, d. h. es sind Mindestergebnisanforderungen zu definieren, was schwierig ist
- schwierig ist auch, wie mit unplausibel schlechten Ergebnissen (bis zum Totalversagen) umgegangen werden soll



PRAKTISCHE EMPFEHLUNGEN

- fluide Intelligenzfaktoren i. S. von GMA-Skalen sollten als relevante Prädiktoren der Managementdiagnostik, aber nicht als der dominante Prädiktor angesehen werden
- die Bedeutung ist umso größer bei Potenzialfragestellungen und Aufgaben, für die es wenig relevantes Erfahrungswissen gibt (z. B. Transformation in eine Arbeitswelt der Zukunft)
- kristalline Intelligenzfaktoren sind mindestens ebenso relevant
- Intelligenzfaktoren sollten stets (auch) mit Fallstudien und speziellen Interviewteilen erhoben werden
- Fallstudien und Interviewnutzung bieten Interpretationsvorteile, da qualifizierte Beobachter:innen gut ein Unterschreiten von Mindestanforderungen feststellen können
- GMA-Testskalen liefern als (zusätzliche) Methode ebenfalls einen wichtigen Informationsgewinn mit hoher Messgenauigkeit
- für unplausibles Testversagen sollten die Ursachen exploriert werden, ggf. ist auch eine Nachtestung sinnvoll
- Mindestwerte für spezifische GMA-Skalen sollten empirisch abgesichert werden



LITERATURLISTE

- Bobko, P., Roth, P. L., Huy, L., Oh, I. S., & Salgado, J. (2025). The need for “considered estimation” versus “conservative estimation” when ranking or comparing predictors of job performance. *International Journal of Selection and Assessment*, 33(1), Article e12489
- Gignac, G. E., & Zajenkowski, M. (2025). Humans peak in midlife: A combined cognitive and personality trait perspective. *Intelligence*, 113, 101961.
- Hunter, J. E., Schmidt, F. L., & Le, H. (2006). Implications of direct and indirect range restriction for meta-analysis methods and findings. *Journal of Applied Psychology*, 91, 594-612. <https://doi.org/10.1037/0021-9010.91.3.594>
- Kulikowski, K., Cucina, J. M., Hayes, T. L., Oh, I. S., Ones, D., & Viswesvaran, C. (2025). In defence of cognitive ability testing: Affirming the evidence for its continued use in personnel selection and admission decisions. *Journal of Occupational and Organizational Psychology*, 98(3), e70048. <https://doi.org/10.1111/joop.70048>
- Ones, D. S. & Viswesvaran, C. (2023). A response to speculations about concurrent validities in selection: Implications for cognitive ability. *Industrial and Organizational Psychology* 16, 358–365. <https://doi.org/10.1017/iop.2023.43>
- Sackett, P. R., Zhang, C., Berry, C. M., & Lievens, F. (2022). Revisiting meta-analytic estimates of validity in personnel selection: Addressing systematic overcorrection for restriction of range. *Journal of Applied Psychology*, 107(11), 2040–2068. <https://doi.org/10.1037/apl0000994>
- Sackett, P. R., Demeke, S., Bazian, I. M., Griebie, A. M., Priest, R., & Kuncel, N. R. (2024). A contemporary look at the relationship between general cognitive ability and job performance. *Journal of Applied Psychology*, 109(5), 687–713. <https://doi.org/10.1037/apl0001159>
- Schmidt, F. L. (2002). The role of general cognitive ability and job performance: Why there cannot be a debate. *Human Performance*, 15(1–2), 187–210. <https://doi.org/10.1080/08959285.2002.9668091>
- 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. <https://doi.org/10.1037/0033-2909.124.2.262>
- Viswesvaran, C., Ones, D. S., & Schmidt, F. L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology*, 81(5), 557–574. <https://doi.org/10.1037/0021-9010.81.5.557>

VIELEN DANK FÜR IHRE AUFMERKSAMKEIT

Organistenweg 3, 21614 Buxtehude



04161 / 5589942



kontakt@forum-assessment.de



Forum Assessment e.V.

vertreten durch den Vorstand

Niklas Becker, Wiebke Goertz, Rüdiger Fruhner, Patrick Wiederhake