IBBW Institut für Bildungsanalysen Baden-Württemberg

ITB Consulting GmbH

Metaanalyse zur prognostischen Validität

Wie gut sagen Tests Studienerfolg vorher?

Eine Metaanalyse für Tests im deutschsprachigen Raum

Schult, J., Hofmann, A. & Stegt, S. (im Druck)

Fachspezifische Studieneignungstests haben in Deutschland schon eine gewisse Tradition

Sind im Vergleich zu den USA aber immer noch „jung“

USA

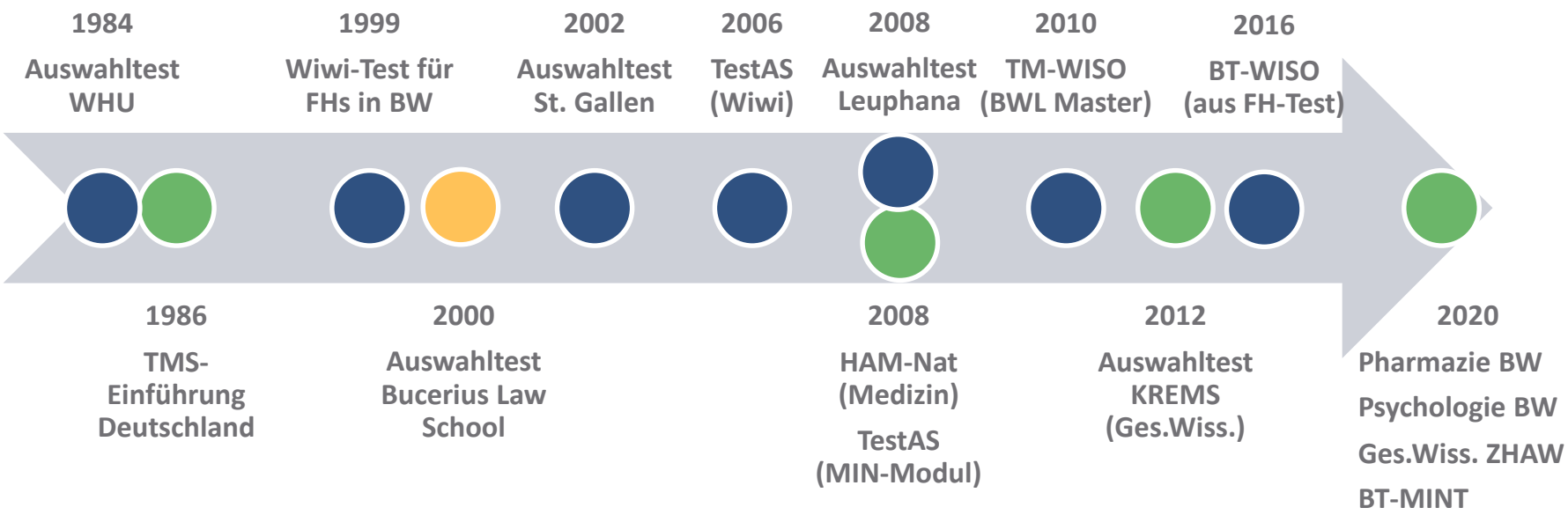
- Scholastic Aptitude Test: seit 1901
- General Record Examination: seit 1936
- Graduate Management Admission Test: seit 1953
- Pharmaceutical College Admission Test: seit 1974

Deutschland

- Test für Medizinische Studiengänge: Entwicklung seit 1977, Einführung 1986
- Test für den Studiengang Pharmazie: entwickelt 1979 (nie eingesetzt)
- HAM-Nat: seit 2008
- Test für Masterstudiengänge in Wirtschafts- und Sozialwissenschaften: seit 2010

Bei Auswahltests dominieren Medizin/ Gesundheit- und Wirtschaftswissenschaften

Bei Wirtschaftswissenschaften viele Tests für einzelne Hochschulen



Andere Studienfelder: fachspezifische Studierfähigkeitstests für Beratungszwecke

Die Tests im deutschsprachigen Raum wurden von Beginn an wissenschaftlich evaluiert

Anhand der Gütekriterien der psychologischen Diagnostik

- Objektivität
- Reliabilität (Messgenauigkeit)
- Validität
  - Inhaltsvalidität
  - Konstruktvalidität
  - **Prognostische Validität**
- Fairness
- Akzeptanz

Beim Gütekriterium prognostische Validität geht es um die Vorhersage von Erfolg

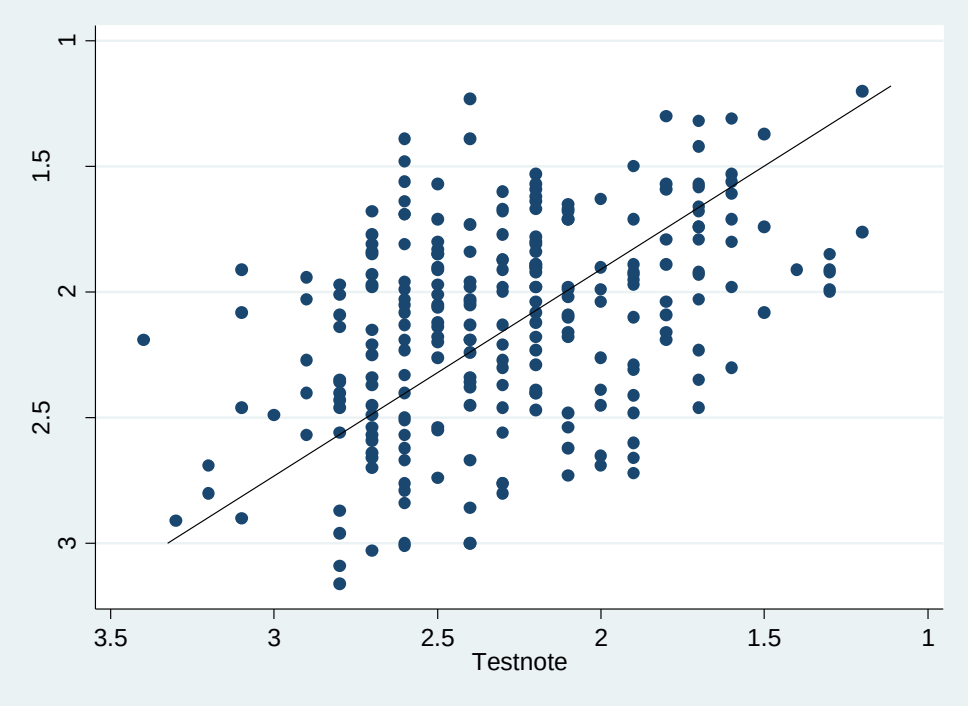
Problem: Wie lässt Erfolg sich zuverlässig „messen“?

- **Studienabbruch / Studienabschluss**
- **Studium in Regelstudienzeit**
- **Studiennoten (Zwischennoten / Endnote)**
- **Beruflicher Erfolg / Gehalt**
- **Zufriedenheit**

- Durchschnittsnoten werden bei der Erfassung von Studienerfolg am häufigsten berücksichtigt.
- Andere Studienerfolgskriterien lassen sich schwer erfassen oder haben nicht genügend Varianz; daher reicht die Zahl der Studien noch nicht für eine metaanalytische Betrachtung aus.

Die Prognosekraft wird häufig als Korrelation zwischen Prädiktor und Kriterium angegeben

Mit verschiedenen statistischen Korrekturen



Probleme/Korrekturen

- Varianzeinschränkung
- Eingeschränkte Reliabilität des Kriteriums
- Eingeschränkte Reliabilität des Prädiktors
- Multiple Regression

Die Metaanalyse von Hell et al. (2007) fasste die Ergebnisse der Studien 1980-2002 zusammen

Schwerpunkt im Bereich Medizin, alle Tests von einem Institut

9 Publikationen im DACH-Raum ab 1980 (aktuellste Publikation: 2002)

Studienfach	K	N	Test
Medizin	25	43218	Test für Medizinische Studiengänge / EMS
Wirtschaftswissenschaften	7	441	Auswahltest WHU / Reihungstest FH Vorarlberg
Informatik	1	67	SFT-MATH
Ingenieurwissenschaften	1	161	SFT-IW
Pharmazie	1	180	Test für den Studiengang Pharmazie
mehrere Studienfächer	1	1024	SFT-IW / SFT-MATH / SFT-REWI / SFT-WIWI
Gesamt	36	45091	

- $r_{unkorrigiert} = .32$; $r_{korrigiert} = .48$
- Schulnoten: $r_{korrigiert} = .53$ (Trapmann, Hell, Weigand & Schuler, 2007)
- Geschätzte inkrementelle Validität von SFTs über Schulnoten: 6,6 - 10,2% mehr Varianz wird aufgeklärt (Hell, Trapmann & Schuler, 2008)

Wie ist die Prognosekraft heute, und gelingt die Prognose auch außerhalb der Medizin?

Eine Fortschreibung und Erweiterung der Metaanalyse war notwendig

- Das Schulsystem hat sich verändert
- Das Hochschulsystem hat sich verändert (Bologna-Reform)
- Die Tests haben sich verändert (Erfahrungswerte, Testkonzepterweiterungen)
- Digitalisierung im Bildungsbereich
- Differenzierung im Bildungsbereich
- …

„Die Studien in der Metaanalyse sind doch fast alle älter als 20 Jahre, gilt das überhaupt heute noch? Außerdem wurde ja fast nur der TMS/EMS untersucht für Medizin…“

Die aktuelle Metaanalyse ist eine Fortsetzung und Erweiterung auf andere Fächer

Zudem soll die inkrementelle Validität von SFTs untersucht werden

Kriterien für Studienauswahl:

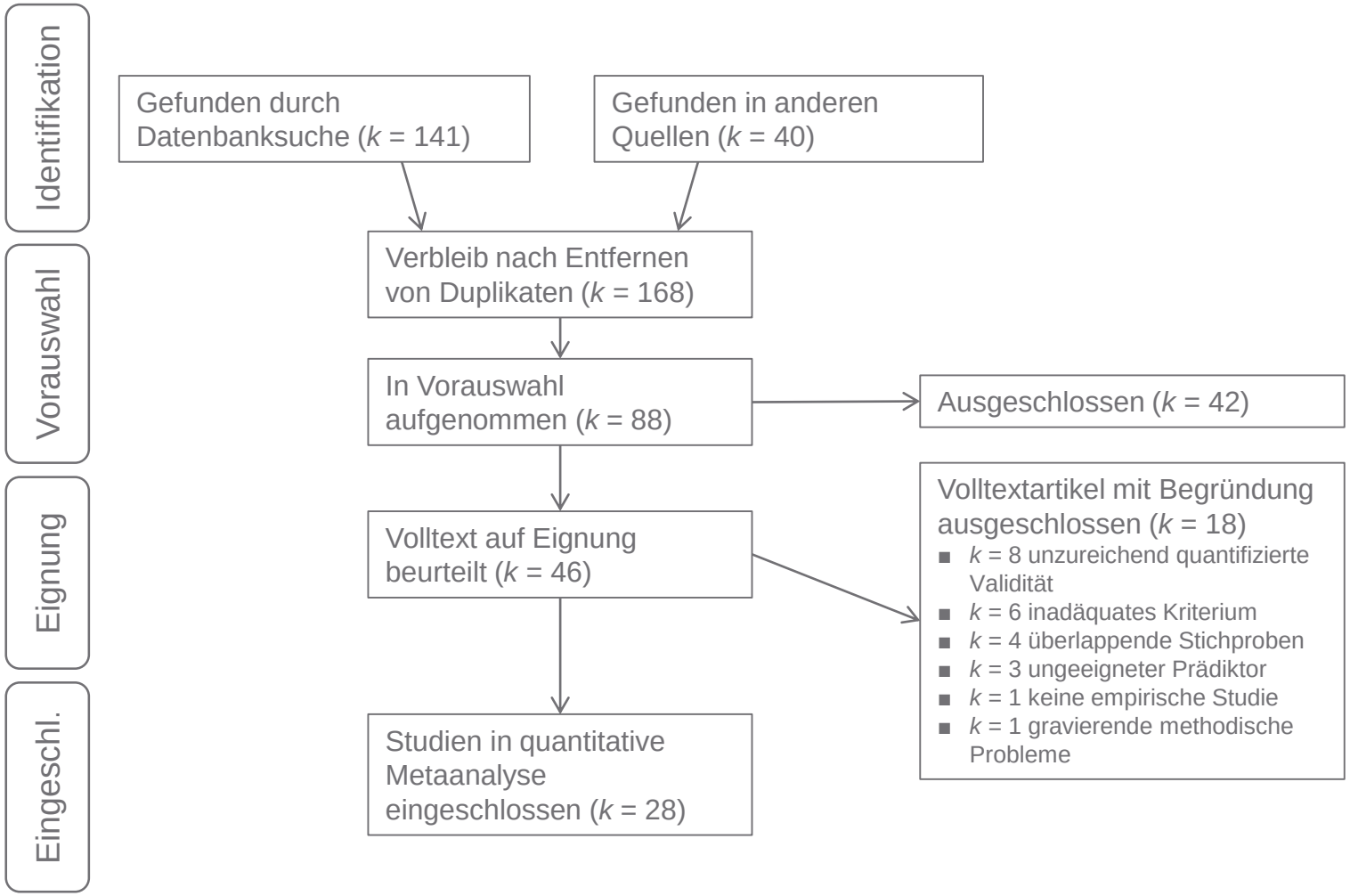
- Verwendung eines fachspezifischen Studierfähigkeitstests
- Längsschnittdesign
- Berichtete Korrelation zwischen Testleistung und Studiennoten
- Publikation nach 2005 (nicht in Metaanalyse von Hell et al. enthalten)
- Testung nach 1990
- Stichprobe besteht aus Studierenden von tertiären Bildungseinrichtungen des DACH-Raums

Betrachtete Moderatoren:

- Studienfeld
- Art des (angestrebten) Abschlusses
- Publikationsjahr
- Heterogenität der Stichprobe (bzgl. Zahl der betrachteten Hochschulen)
- Art der Veröffentlichung
- Selektionsfunktion
- Art des Kriteriums
- Jahre zwischen Testung und Kriteriumserfassung
- Art der Hochschule
- Art des verwendeten Tests

Vorgehen: Nach der Literaturrecherche wurden insgesamt 28 Studien für die Analyse verwendet

Gesucht wurde über PsycInfo, PSYNDEX und Google Scholar



Ergebnis 1: Fachspezifische Studierfähigkeitstests sind prognostisch valide

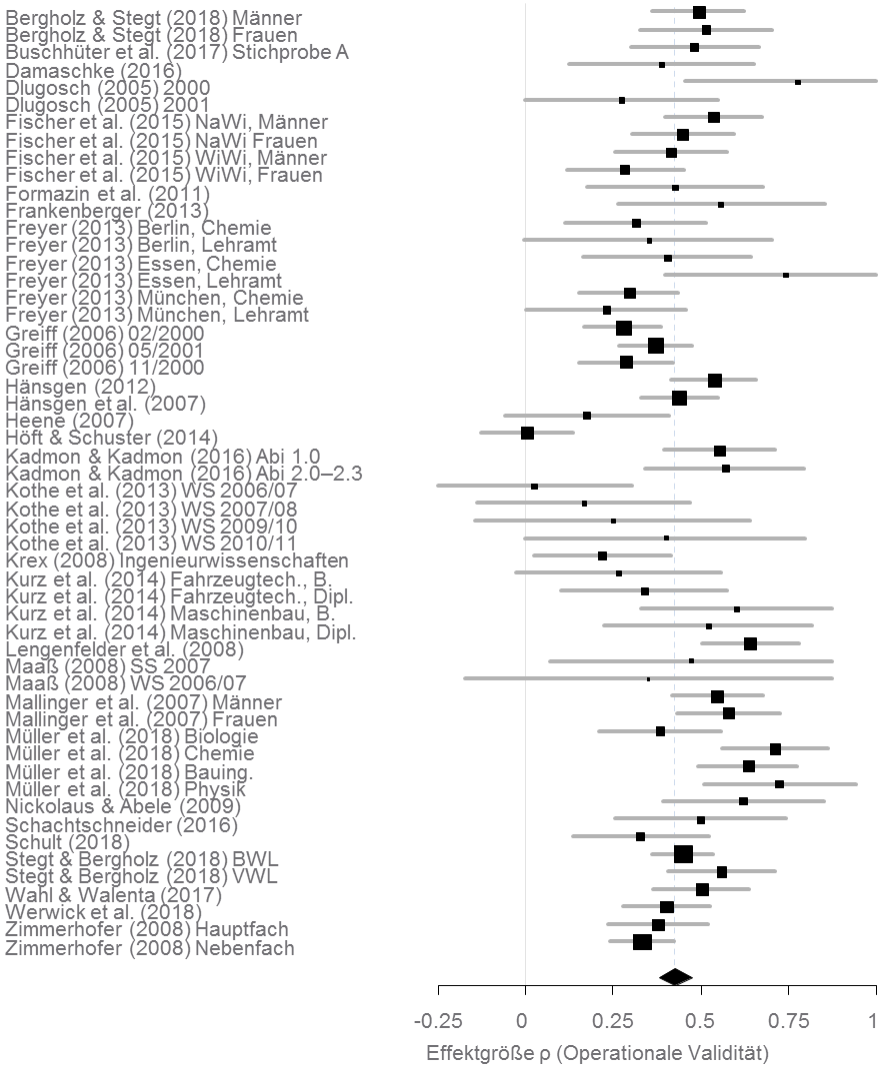
Die Studienergebnisse von Hell et al. (2007) wurden repliziert

- K = 58
- N = 8410
- unkorrigierte mittlere Korrelation:

  r = .36

- Korrelation nach Korrektur*:

  ρ = .43

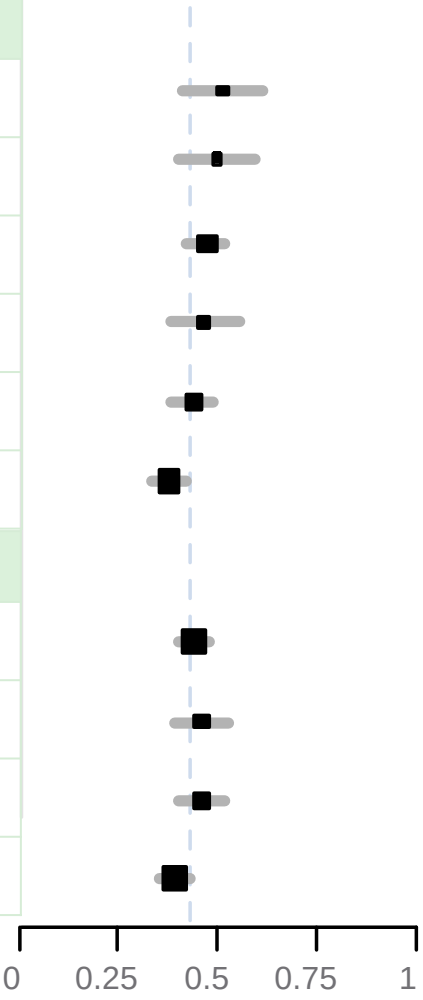


* Korrektur für die Einschränkung der Reliabilität des Kriteriums Studiennoten und die selektionsbedingte Varianzeinschränkung

Ergebnis 2: Auch außerhalb der Medizin gelingt eine gute Prognose des Studienerfolgs

…und bewährte Verfahren wie TMS sind immer noch valide

Studienfeld ($p < .005$)	**k**	**n**	**ρ**
Sozialwissenschaften	4	447	.510
Rechtswissenschaften	4	733	.498
Medizin	12	2109	.471
Ingenieurwissenschaften	6	467	.465
Naturwissenschaften	13	1243	.437
Wirtschaftswissenschaften	10	2678	.376
Abschlussart ($p < .109$)			
Bachelor	24	3163	.441
Master	3	936	.457
Staatsexamen	14	1619	.457
Diplom	13	2692	.390



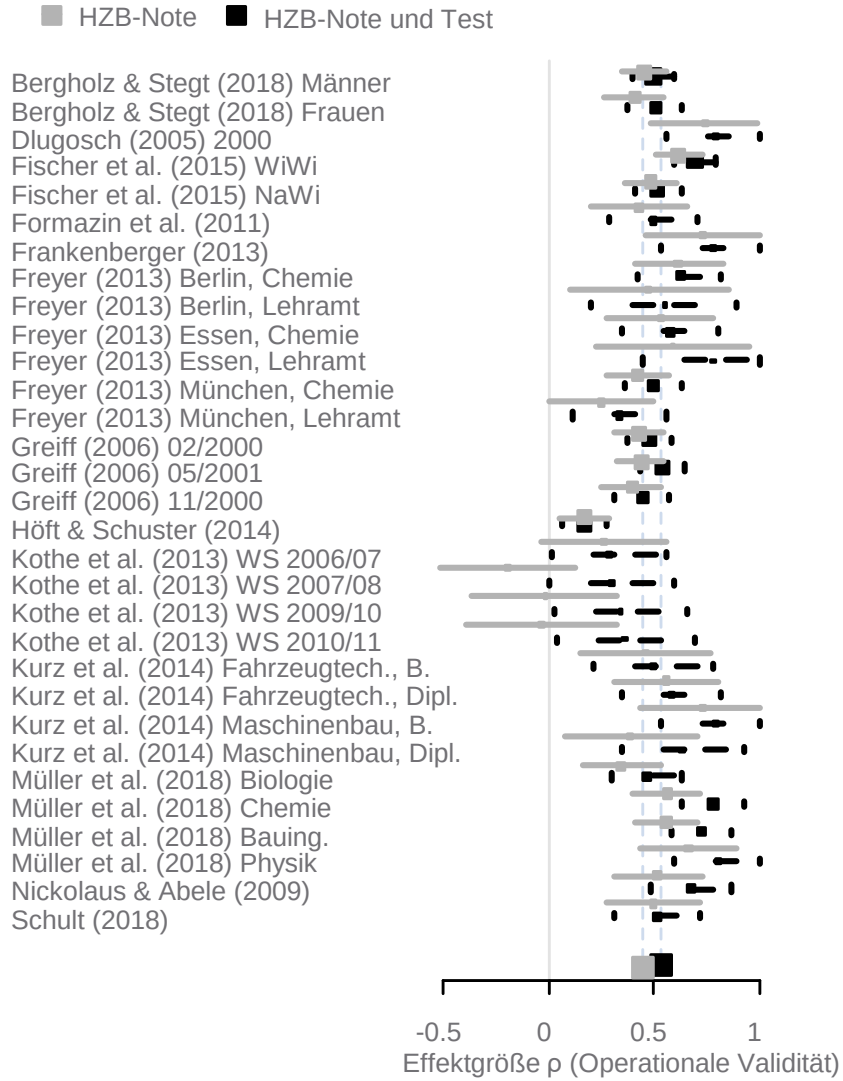
Ergebnis 3: Studierfähigkeitstests verbessern die Prognose im Vergleich zur Abiturnote allein

Das beobachtbare Inkrement liegt im geschätzten Bereich von Hell et al.

Inkrementelle Validität:

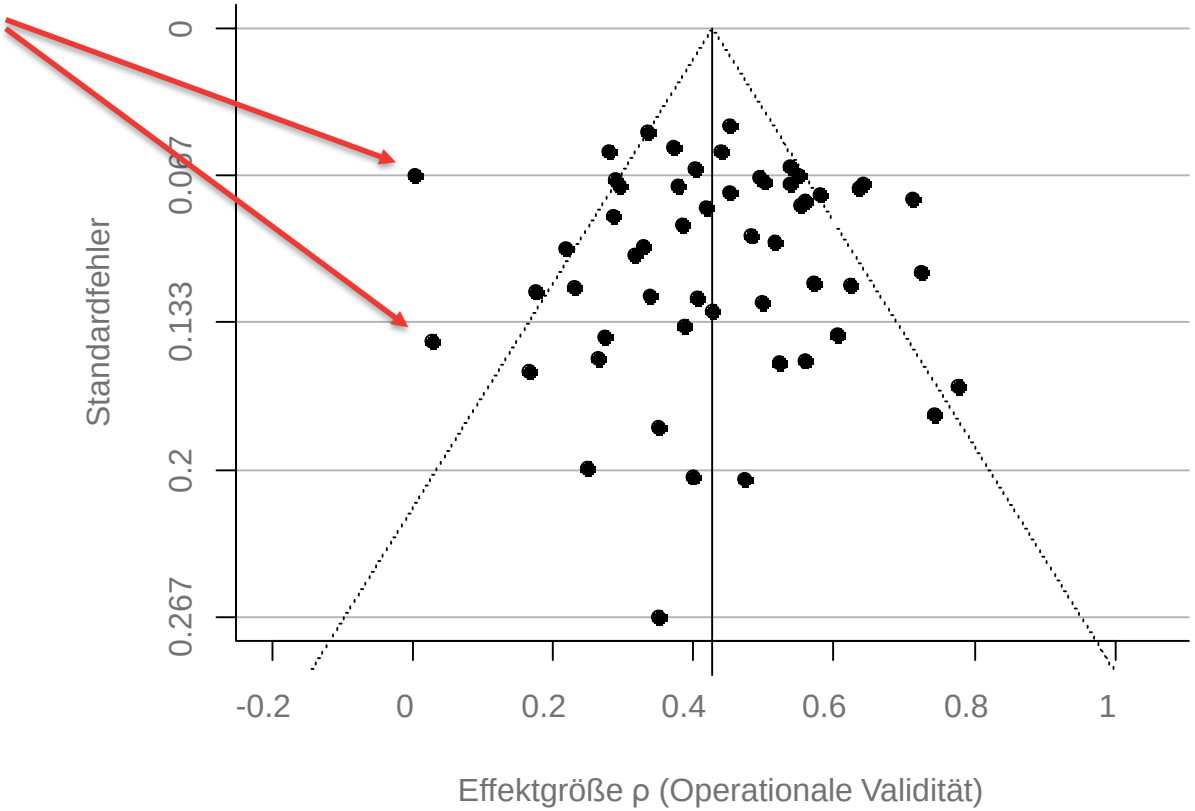
$\Delta R^2 = .530^2 - .447^2 = \mathbf{.081}$

mit k = 31 und n = 3999



Problem 1: Nicht jeder wissenschaftlich entwickelte Test ist automatisch valide (leider)

→ Jeder Test sollte wissenschaftlich evaluiert werden



Problem 2: Je nach Stichprobe findet man für einen Test verschiedene Validitäten!

→ Evaluation mit mehreren Stichproben/Jahrgänge/Hochschulen

Beispiel Auswahltest der Bucerius Law School

Stichprobe	Kriterium	N	ρ
Studienjahr 2000	LL.B.	63	.78
Studienjahr 2001	Zwischennote	91	.27
Frauen 2005-2012	LL.B.	201	.52
Männer 2005-2012	LL.B.	378	.49

Beispiel ING-Test an der FH Esslingen

Kriterium	N	ρ
B.Sc. Fahrzeugtechnik	42	.27
Diplom Fahrzeugtechnik	64	.34
B.Sc. Maschinenbau	49	.60
Diplom Maschinenbau	41	.52

Problem 3: Für viele Tests gibt es keine veröffentlichten Evaluationsdaten

→ Evaluationsstudien anregen/einfordern/durchführen

- Kein Bewusstsein für die Notwendigkeit von Evaluation
- Kein (politisches) Interesse an Evaluation
- Evaluation scheitert am Thema Datenschutz
- Schwache Evaluationsergebnisse werden nicht veröffentlicht
- Ablehnung durch Fachzeitschrift („nicht interessant“, „nicht generalisierbar“)
- Evaluationsergebnisse nicht metaanalytisch aggregierbar/fehlende Infos

Einige Moderatoren hatten einen signifikanten Einfluss auf die prognostische Validität

- Prognosezeitraum: Kurzfristiger Erfolg lässt sich leichter vorhersagen als langfristiger.
- Publikationsjahr: Neuere Studien weisen höhere prognostische Validitäten aus.
- Publikationsart: In Fachzeitschriften sind die Validitäten höher als in anderen Medien.
- Hochschulart: Höhere Validitäten in reinen Universitätsstichproben

Keinen Einfluss hatte die Testart: „klassischer Studierfähigkeitstest vs. andere/Mischtests“

Neben fachspezifischen Problemlöseaufgaben können auch fachunspezifische Problemlöseaufgaben, Kenntnisabfragen, Essays sinnvoll sein.

→ Es gibt nicht das eine beste Konzept, wichtig scheinen vielmehr eine Entwicklung des Konzepts auf Basis einer Anforderungsanalyse, die sorgfältige Entwicklung der Aufgaben, die Durchführung unter kontrollierten Bedingungen und die Evaluation nach psychologischen Gütekriterien.

Zusammenfassung der Ergebnisse: Fachspezifische Studierfähigkeitstests…

- …ermöglichen weiterhin eine gute Studienerfolgsprognose, in ähnlicher Höhe wie die Abiturnote.
- …ermöglichen eine Erhöhung der Prognosekraft über die Abiturnote hinaus.
- …funktionieren in vielen Studienfeldern und mit unterschiedlichen Konzepten.
- …sind keine Selbstläufer: Sie funktionieren nicht „automatisch“ und müssen daher evaluiert werden, am besten mit mehreren Stichproben.

Kontakt

IBBW
Institut für Bildungsanalysen
Baden-Württemberg

ITB Consulting GmbH
Koblenzer Straße 77
53177 Bonn

Tel.: +49 228 820 90 - 0
Fax: +49 228 820 90 - 38

www.itb-consulting.de

Dr. Johannes Schult

Telefon: +49 711-6642-2108

E-Mail: johannes.schult@ibbw.kv.bwl.de

Arvid Hofmann

Telefon: +49 228-82090-52

E-Mail: arvid.hofmann@itb-consulting.de

Dr. Stephan Stegt

Telefon: +49 228-82090-14

E-Mail: stephan.stegt@itb-consulting.de