
MASTERARBEIT

Frau (B.A.)
Isabelle Einhorn

**Ein Überblick zu Reliabilitäts-
und Validitätsangaben des
Test d2 – Aufmerksamkeits-
und Belastungs-Test in Ab-
hängigkeit von Eigen- und
Fremdpublikationen**

Mittweida, 2016

MASTERARBEIT

Ein Überblick zu Reliabilitäts- und Validitätsangaben des Test d2 – Aufmerksamkeits- und Belastungs-Test in Ab- hängigkeit von Eigen- und Fremdpublikationen

Autor:

Frau (B.A.)

Isabelle Einhorn

Studiengang:

Betriebswirtschaft

Seminargruppe:

BW13w1-M

Erstprüfer:

Prof. Dr. rer. pol. Ulla Meister

Zweitprüfer:

Dipl.-Psychologe Frank Schumann

Einreichung:

Mittweida, 12. Dezember 2016

Verteidigung/Bewertung:

Mittweida, 2016

MASTERTHESIS

An overview of reliability and validity of the Test d2 – Test of Attention depending on internal and external publications

author:

Ms. (B.A.)

Isabelle Elhorn

course of studies:

Business economics

seminar group:

BW13w1-M

first examiner:

Prof. Dr. rer. pol. Ulla Meister

second examiner:

Dipl.-Psychologe Frank Schumann

submission:

Mittweida, 12. December 2016

defence/ evaluation:

Mittweida, 2016

Bibliografische Beschreibung:

Einhorn, Isabelle:

Ein Überblick zu Reliabilitäts- und Validitätsangaben des Test d2 – Aufmerksamkeits- und Belastungs-Test in Abhängigkeit von Eigen- und Fremdpublikationen. - 2016.- 152, 102, 36 S.

Mittweida, Hochschule Mittweida, Fakultät Wirtschaftswissenschaften, Masterarbeit, 2016

Referat:

Der Test d2 – Aufmerksamkeits- und Belastungs-Test wird vielseitig eingesetzt, aktuell vor allem in der Arbeits- und Organisationspsychologie, um geeignete Bewerber für Unternehmen zu selektieren. Diese Masterarbeit beschäftigt sich unter anderem mit den Gütekriterien Reliabilität und Validität des Tests d2 in den Eigen- und Fremdpublikationen, welche miteinander verglichen werden. Des Weiteren werden die Unterschiede zwischen den spezifischen Anwendungsgebieten im Hinblick auf die Zeitentwicklung sowie die Stichprobengrößen in beiden Publikationsarten untersucht. Für diese wissenschaftliche Arbeit wurden 124 Artikel aus acht unterschiedlichen Fachzeitschriften im Zeitraum zwischen 1962 und 2015 genutzt.

Inhaltsverzeichnis

Inhaltsverzeichnis.....	I
Abbildungsverzeichnis.....	III
Tabellenverzeichnis.....	IV
Abkürzungsverzeichnis.....	V
1 Einleitung	1
1.1 <i>Problem- und Zielstellung</i>	1
1.2 <i>Vorgehensweise</i>	4
2 Theoretische Betrachtung.....	5
2.1 <i>Einsatz psychologischer Tests für die Personalauswahl</i>	5
2.1.1 Heterogenität in der Personalauswahl.....	5
2.1.2 Gütekriterien	7
2.1.3 Testauswahl von allgemeinen Leistungstests	13
2.2 <i>Test d2 – Aufmerksamkeits- und Belastungs-Test</i>	16
2.2.1 Begriffserklärung von Aufmerksamkeit und Konzentration	16
2.2.2 Funktionsweise des Tests d2 – Aufmerksamkeits- und Belastungs-Test.....	23
2.2.3 Kennwerte des Tests d2 – Aufmerksamkeits- und Belastungs-Test.....	26
3 Fragestellung und Hypothesenbildung.....	30
4 Untersuchungsdesign	37
4.1 <i>Definitionsphase</i>	37
4.1.1 Untersuchungsablauf	38
4.1.2 Stichprobenbeschreibung der untersuchten Eigen- und Fremdpublikationen	42
4.1.3 Stichprobenbeschreibung der Publikationszeiträume von Eigen- und Fremdpublikationen.....	44
4.2 <i>Untersuchungsmethoden</i>	46
4.2.1 Reliabilitäts-Methoden.....	46
4.2.2 Validitätskriterien.....	48
5 Ergebnisdarstellung	49

5.1	<i>Analyse der Anwendungsbereiche bei der Nutzung des Test d2 – Aufmerksamkeits- und Belastungs-Test bei den Eigen- und Fremdpublikationen</i>	49
5.2	<i>Analyse der Reliabilität des Test d2 – Aufmerksamkeits- und Belastungs-Test bei Eigen- und Fremdpublikationen</i>	54
5.3	<i>Analyse der Validität des Test d2 – Aufmerksamkeits- und Belastungs-Test bei Eigen- und Fremdpublikationen</i>	59
5.4	<i>Analyse des durchschnittlichen Stichprobenumfangs bei Eigen- und Fremdpublikationen des Tests d2 – Aufmerksamkeits- und Belastungs-Test</i>	65
5.5	<i>Reliabilitätsanalyse in Bezug auf die Stichprobenumfänge</i>	69
5.6	<i>Validitätsanalyse in Bezug auf die Stichprobenumfänge</i>	72
6	Diskussion der Untersuchungsergebnisse	76
6.1	<i>Diskussion der Ergebnisse zu den Anwendungsgebieten bei der Nutzung des Tests d2 von Eigen- und Fremdpublikationen</i>	77
6.2	<i>Diskussion der Gütekriterien Reliabilität und Validität bei den in Eigen- und Fremdpublikationen</i>	81
6.3	<i>Diskussion des durchschnittlichen Stichprobenumfangs bei Eigen- und Fremdpublikationen</i>	89
6.4	<i>Diskussion der Gütekriterien Reliabilität und Validität mit Bezug auf die Stichprobenumfänge von Eigen- und Fremdpublikationen</i>	93
6.5	<i>Diskussion der Untersuchungsmethode</i>	98
7	Zusammenfassung und Ausblick	100
	Literaturverzeichnis	103
	Anlagenverzeichnis	110
	Selbstständigkeitserklärung	146

Abbildungsverzeichnis

Abbildung 1: Vollständige markierte Musterzeile des FAIR in der Testform A (Quelle: Moosbrugger; Reiß (2004), S. 106).....	14
Abbildung 2: Testausschnitt aus dem Test d2-R (Quelle: Schmidt-Atzert; Amelang (2012), S. 193).....	25
Abbildung 3: Anzahl der erscheinenden Eigen- und Fremdpublikationen im betrachteten Publikationszeitraum (Quelle: Eigene Darstellung).....	44
Abbildung 4: Diagramm für die Zuordnung der einzelnen Eigenpublikationen zu den Anwendungsbereichen (Quelle: Eigene Darstellung)	50
Abbildung 5: Diagramm für die Zuordnung der einzelnen Fremdpublikationen zu den Anwendungsbereichen (Quelle: Eigene Darstellung)	51
Abbildung 6: Durchschnittliche Stichprobengröße des Tests d2 – Aufmerksamkeits- und Belastungs-Test von Eigen- und Fremdpublikationen (Quelle: Eigene Darstellung)	67

Tabellenverzeichnis

Tabelle 1: Häufigkeiten der einzelnen Zeichen im Block und im Test (Quelle: Eigene Darstellung in Anlehnung an Brickenkamp et al. (2010), S. 26).....	25
Tabelle 2: Publikationen für Reliabilitäts- und Validitätsuntersuchungen (Quelle: Eigene Darstellung).....	40
Tabelle 3: Anzahl der Publikationen in den verschiedenen Zeitschriften (Quelle: Eigene Darstellung).....	43
Tabelle 4: Anzahl der Eigen- und Fremdpublikationen in den unterschiedlichen Anwendungsbereichen (Quelle: Eigene Darstellung).....	53
Tabelle 5: Retest-Reliabilitäten nach der Test-Retest-Methode bei den Eigenpublikationen (Quelle: Eigene Darstellung)	54
Tabelle 6: Retest-Reliabilitäten nach der Test-Retest-Methode bei den Fremdpublikationen (Quelle: Eigene Darstellung)	56
Tabelle 7: Korrelation der d2-Kennwerte mit anderen Konzentrationstests bei Eigenpublikationen (Quelle: Eigene Darstellung).....	59
Tabelle 8: Korrelation der d2-Kennwerte mit Intelligenztest bei Eigenpublikationen (Quelle: Eigene Darstellung)	60
Tabelle 9: Korrelation der d2-Kennwerte mit anderen Konzentrationstests bei Fremdpublikationen (Quelle: Eigene Darstellung)	61
Tabelle 10: Korrelation der d2-Kennwerte mit Intelligenz bei Fremdpublikationen (Quelle: Eigene Darstellung).....	62
Tabelle 11: Durchschnittliche Stichprobengrößen der Eigen- und Fremdpublikationen in den einzelnen Anwendungsbereichen (Quelle: Eigene Darstellung).....	68

Abkürzungsverzeichnis

AF	Auslassungsfehler
AV	Allgemeines Verständnis
AW	Allgemeines Wissen
BE	Bilder ergänzen
BO	Bilder ordnen
bspw.	beispielsweise
BZO	Bearbeitete Zielobjekte
bzw.	beziehungsweise
d. h.	das heißt
DKT-K	Differentieller Konzentrationstest für Kinder
et al.	et alii (und andere)
F	Fehlerrohwert
F%	Fehlerprozentwert
FL	Figuren legen
GF	Gemeinsamkeiten finden
GZ	Gesamtzahl
GZ-F	Gesamtmenge minus Gesamtfehlerzahl
HAWIE	Hamburg-WECHSLER-Intelligenztest für Erwachsene
IST 2000	Intelligenz-Struktur-Test 2000
IST 70	Intelligenz-Struktur-Test 70
KL	Konzentrationsleistung
KLT	Konzentrations-Leistungs-Test
KKT	Komplexer Konzentrationstest
LPS	Leistungsprüfsystem
MIT	Mannheimer Intelligenztest
MT	Mosaik-Test
N	Stichprobe
p	Irrtumswahrscheinlichkeit
r	Pearson product- moment correlation (Korrelationskoeffizient)
r^{GZ}	Pearson product- moment correlation (Korrelationskoeffizient) von der Gesamtzahl

r^{GZ-F}	Pearson product- moment correlation (Korrelationskoeffizient) von der Gesamtmenge minus Gesamtfehlerzahl
r^{F%}	Pearson product- moment correlation (Korrelationskoeffizient) von den Fehlerprozenten
r^{SB}	Pearson product- moment correlation (Korrelationskoeffizient) von der Schwankungsbreite
RD	Rechnerisches Denken
Rev.T.	Revisions-Test
SB	Schwankungsbreite
SPM	Standard Progressive Matrices Test
u.a.	unter anderem
VF	Verwechslungsfehler
Vgl.	vergleiche
z.B.	zum Beispiel
ZN	Zahlen nachsprechen
ZS	Zahlen-Symbol-Test
≤	kleiner gleich
≥	größer gleich
*	p<0,05

1 Einleitung

1.1 Problem- und Zielstellung

„Es sind immer die einfachsten Ideen, die außergewöhnliche Erfolge haben.“¹ Auch der Test d2 – Aufmerksamkeits- und Belastungs-Test, ist ein einfaches Instrument in der psychologischen Diagnostik und erzielt eine große Wirkung. Der Test wird als allgemeiner Leistungstest bezeichnet² und findet in der „Verkehrspsychologie“, „Pädagogischen Psychologie“, „Neuropsychologie“, „Klinischen und Medizinischen Psychologie“ sowie „Arbeits- und Organisationspsychologie“ Anwendung.³ Der Test d2 erfasst die „[...] Konzentration bei der Selektion relevanter Reize oder bei einer Aufgabe, die selektive Aufmerksamkeit verlangt [...]“⁴ und ermöglicht das Messen der konzentrierten Aufmerksamkeit. Dieses Verfahren eignet sich als Selektionsinstrument für Unternehmen bei der Personalauswahl, denn es wird zunehmend schwerer qualifiziertes Fachpersonal aus einem heterogenen Bewerberpool auszuwählen.⁵ Besondere Anwendung findet der Test in Tätigkeiten, welche eine erhöhte Konzentrationsfähigkeit verlangen, wie bspw. bei Fluglotsen oder beim Korrigieren von Klassenarbeiten im Lehrerberuf.⁶ Bei den deutschen Testanwendern ist der Test d2 durch die einfache Handhabung, die Zeitökonomie und die Akzeptanz unter allen Konzentrationstests sehr beliebt und findet nach wie vor seine Berechtigung zur Anwendung.⁷ Jedoch sagen diese, auf die Anwendung bezogenen Vorzüge, nichts über die Qualität des Instrumentes aus. In diesem Zusammenhang wird Kritik

¹ LeoTolstoi.

² Vgl. Brickenkamp; Schmidt-Atzert; Liepmann (2010), S. 9 f.

³ Vgl. ebenda, S. 26 ff.

⁴ Ebenda, S. 11.

⁵ Vgl. Kersting (2011), S. 23.

⁶ Vgl. Brickenkamp; Schmidt-Atzert; Liepmann (2010), S. 29.

⁷ Vgl. Oswald; Hagen (1997), S. 88.

von Fremdautoren an einigen Kennwerten bzw. der Reliabilität und Validität geäußert.⁸ Dabei stellt sich die Frage, ob der Test d2 von dem Testautor geschönt ist, indem bspw. nicht genug Nachforschungen zu den Gütekriterien angestellt wurden und somit die Popularität des Tests ohne tiefgründige Nachweise gestiegen ist.⁹ In dieser Masterarbeit liegen die Schwerpunkte im Vergleich zwischen den Eigen- und Fremdpublikationen hinsichtlich der Anwendungsgebiete, der Reliabilitäts- und Validitätsangaben, der durchschnittlichen Stichprobengrößen sowie der Reliabilitäts- und Validitätsangaben in Bezug auf den Stichprobenumfang. Die verschiedenen Werte der genannten Gütekriterien, basierend auf messtheoretischen Grundlagen, geben dabei eine Aussage über die Güte des Tests und werden miteinander verglichen. Bei der Reliabilitätsschätzung werden zwei Methoden (Retest-Reliabilität sowie Split-Half-Methode) aufgegriffen. Die konvergente und diskriminante Validität wird bei der Konstruktvalidität eingesetzt. Vorrangig wird festgestellt, ob bzw. wie sich der Test d2 im Laufe der Zeit in den unterschiedlichen Anwendungsgebieten entwickelt hat und ob es gravierende Unterschiede in den Eigen- und Fremdpublikationen bezüglich der Reliabilität und Validität sowie den Stichprobengrößen gibt. Entscheidend bei der Untersuchung des durchschnittlichen Stichprobenumfangs und der Stichprobengröße mit den Reliabilitäten bzw. Validitäten ist die Anzahl der Probanden.¹⁰ Im Einzelnen sollen nachfolgende Fragen unter Bezugnahme auf die Schwerpunkte der Arbeit gezielt betrachtet werden: Wie ist der Test d2 aufgebaut?, Was misst der Test d2?, Welche Rolle spielen die Anwendungsbereiche in den Eigen- und Fremdpublikationen?, Welche Unterschiede gibt es bei den Reliabilitäten sowie Validitäten in den Eigen- und Fremdpublikationen?, In welcher Beziehung stehen die Stichprobengrößen mit den Ergebnissen der Gütekriterien in den Eigen- und Fremdpublikationen? Für diese Untersuchung werden acht unterschiedliche Fachzeitschriften im Zeitraum von 1962 bis 2015 verwendet. Insgesamt sind 136 Artikel ausgewertet, davon werden 124 genutzt. Zwölf Veröffentlichungen sind in

⁸ Vgl. Heinrich (1973), S. 118 ff.; Oehlschlägel; Moosbrugger (1991), S. 42 ff.; Fay (1995), S. 292 ff.; Daseking; Putz (2015), S. 323 f.

⁹ Vgl. Heinrich (1973), S. 118 ff.; Oehlschlägel; Moosbrugger (1991), S. 42 ff..

¹⁰ Vgl. Hartig; Frey; Jude (2012), S. 161; Schmidt-Atztert (2012), S. 141 ff.; Härtler (2014), S. 119.

dieser wissenschaftlichen Arbeit auszuschließen, da über den Test d2 keine weiteren Aussagen getroffen werden und sie somit für die Untersuchung nicht relevant sind. Bei dem Vergleich zwischen den Reliabilitäten und Validitäten ist wichtig, dass keine Modifikationen bei dem Test d2 vorgenommen wurden. Mit diesem Kriterium ist die Auswahl an Publikationen der Eigen- und Fremdautoren gesunken. Für die Vergleiche der Anwendungsgebiete stehen wesentlich mehr Artikel zur Verfügung. Auch bei der Untersuchung der durchschnittlichen Stichprobengrößen in Eigen- und Fremdpublikationen werden mehr Artikel verwendet als bei der Untersuchung von Stichprobengrößen in Verbindung mit den Reliabilitäten bzw. Validitäten.

Empfehlenswert sind große Stichprobengrößen von über $N = 100$.¹¹ In den meisten Fällen wird eine kleine Stichprobengröße, aufgrund der Schwierigkeit der Stichprobengewinnung oder zu hoher Ausgaben für die Untersuchung ermittelt.¹² Ein weiterer Grund für kleine Stichprobenumfänge sind die Anwendungen der Methoden von Reliabilitäten und Validitäten. Beispielsweise bei der Retest-Reliabilitätsmethode muss der Proband zu zwei unterschiedlichen Zeitpunkten zur Verfügung stehen, dadurch kommt es vor, dass die Versuchsteilnehmer zum zweiten Termin nicht erscheinen.¹³ Auch bei den Validitäten wird eine kleine Stichprobengröße vorausgesagt.¹⁴ Somit ist es für diese wissenschaftliche Arbeit interessant, ob es bei den Eigen und Fremdautoren Unterschiede bezüglich des (durchschnittlichen) Stichprobenumfangs gibt. Diese Untersuchung wird mit Reliabilitäts- und Validitätsanalysen sowie dem arithmetischen Mittel der Stichprobengrößen durchgeführt.

¹¹ Vgl. Boomsma (1985), S. 240; Homburg; Klarmann (2006), S. 733.

¹² Vgl. Härtler (2014), S. 119.

¹³ Vgl. Schmidt-Atzert; Amelang (2012), S. 141; Bühner (2011), S. 61.

¹⁴ Vgl. Schmidt-Atzert; Amelang (2012), S. 160.

1.2 Vorgehensweise

Für die Bearbeitung dieser Masterthesis „Ein Überblick zu Reliabilitäts- und Validitätsangaben des Test d2 – Aufmerksamkeits- und Belastungs-Test in Abhängigkeit von Eigen- und Fremdpublikationen“ ist diese Arbeit insgesamt in sieben Kapitel eingeteilt.

Das erste Kapitel führt zum wissenschaftlichen Thema hin und stellt die Problem- sowie Zielstellung dar.

Im zweiten Kapitel werden die theoretischen Vorbetrachtungen getroffen, die nötig sind, um die folgenden Zusammenhänge zu verstehen. Es werden die Heterogenität der Personalauswahl dargestellt, Begrifflichkeiten zu den Gütekriterien geklärt, die Testauswahl von allgemeinen Leistungstests vorgestellt, Begriffe der Aufmerksamkeit und Konzentration erklärt sowie die Funktionsweise und die Kennwerte des Tests d2 aufgezeigt.

Das dritte Kapitel enthält die Fragestellungen und die dazugehörigen Hypothesen, welche sich daraus ableiten lassen.

Kapitel vier gibt Auskunft über das Untersuchungsdesign. Dabei wird der Untersuchungsablauf beschrieben, indem insbesondere auf die Auswahl und Selektion der verwendeten Fachzeitschriften sowie die Zuordnung der enthaltenen Artikel zu den Eigen- und Fremdpublikationen und den verschiedenen Anwendungsgebieten eingegangen wird. In diesem Zusammenhang erfolgt auch die Stichprobenbeschreibung für die Eigen- und Fremdpublikationen.

Im fünften Kapitel werden die Ergebnisse der unterschiedlichen Fragestellungen analysiert und die Ergebnisse der Eigen- und Fremdpublikationen verglichen.

Das sechste Kapitel gibt Auskunft wie die entsprechenden Ergebnisse zu interpretieren sind. Es werden ebenfalls die Hypothesen der einzelnen Fragestellungen bestätigt bzw. abgelehnt.

Kapitel sieben bildet einen inhaltlichen Abschluss der Masterthesis, indem diese wissenschaftliche Arbeit zusammengefasst und ein Ausblick gegeben wird.

2 Theoretische Betrachtung

2.1 Einsatz psychologischer Tests für die Personalauswahl

2.1.1 Heterogenität in der Personalauswahl

Für Unternehmen wird es immer schwieriger gut ausgebildetes Personal zu finden. Aus einem heterogenen Bewerberpool wird eine Auswahl des Personals getroffen. Mit dem demografischen Wandel, der misslungenen Bildungsplanung sowie -politik wird die Auswahl an Bewerbern schwerer und somit können es sich Unternehmen nicht mehr leisten Potenziale zu übersehen.¹⁵ Die Nachfrage für qualifiziertes Fachpersonal steigt und kann nicht bedient werden, wo hingegen für gering qualifizierte Arbeitskräfte kein großer Bedarf auf dem Arbeitsmarkt besteht. Auch die Einstellungen der Arbeitnehmer zur Arbeitstätigkeit hat sich im Laufe der letzten Jahre geändert.¹⁶

Um potenzielle Arbeitskräfte zu finden, werden in der Findungsphase der Bewerber Diagnostik und psychologische Tests angewendet. Der Personalmangel hat zur Folge, dass die Vorselektion der Bewerber nicht mehr streng ausfällt und dadurch die Gruppe der Vorauswahl bunter wird.¹⁷ Laut dem Autor Kersting (2011) ist die Heterogenität der Globalisierung geschuldet, „[...] da einerseits innerhalb Deutschlands der Anteil der Personen mit Migrationshintergrund in der jungen Generation erheblich steigt und andererseits global aufgestellte Organisationen ihr Personal international rekrutieren.“¹⁸

Das Ziel eines Unternehmens zur Sicherung der personellen Ressourcen ist bspw.

¹⁵ Vgl. Kersting (2011), S. 23.

¹⁶ Vgl. ebenda, S. 23.

¹⁷ Vgl. ebenda, S. 23.

¹⁸ Ebenda, S. 24.

der Aufbau einer Arbeitgebermarke, das Personalmarketing, die Personal- und Organisationsentwicklung.¹⁹

Bei der Personalauswahl sind oft die Bewerbungsunterlagen und der erste Eindruck für Führungskräfte entscheidend, dabei entstehen Hypothesen, die mit einer Eignungsbeurteilung auf ihre Haltbarkeit mit verschiedenen Instrumenten geprüft werden. Damit subjektive Emotionen ausgeblendet werden, sind die Verfahren regelgeleitet.²⁰ Verfahren der Eignungsdiagnostik sowie Tests finden immer mehr Bedeutung bei der Bewerberauswahl, denn der Lebenslauf und die Zeugnisse verlieren zunehmend an Aussagekraft.²¹

Mittels psychologischer Tests besteht die Möglichkeit geeignetes Personal für Unternehmen zu selektieren. Jedoch kommt es darauf an, die richtigen Verfahren einzusetzen, um den geeigneten Bewerber für die freie Stelle bzw. für die vorgesehene Tätigkeit zu finden.

Personalentscheidungen sind kostenprächtigt, vor allem wenn diese Entscheidungen falsch getroffen werden.²² Um dies zu vermeiden, werden zunehmend leistungsfähige und kosteneffiziente Verfahren entwickelt, die bei neuen oder auch vorhandenen Mitarbeitern eingesetzt werden.²³ Die Testverfahren sind auf unterschiedlicher Art klassifizierbar, so wird beispielsweise eine Grobklassifizierung zwischen Leistungs- und Persönlichkeitstests²⁴ vorgenommen.²⁵

Seit dem Jahr 2002 gibt es für „Anforderungen an Verfahren und deren Einsatz bei berufsbezogenen Eignungsbeurteilungen“ eine Norm (DIN 33430), die „[...] Quali-

¹⁹ Vgl. Kersting (2011), S. 23.

²⁰ Vgl. Hornke (2004), S. 14.

²¹ Vgl. Sarges; Wottawa (2004), Einführung S. VIII.

²² Vgl. Kersting (2004), S. 55.

²³ Vgl. Sarges; Wottawa (2004), Einführung, S. VII.

²⁴ Bei Leistungstests wird das maximale Verhalten (z.B. im Merkmalsbereich der Intelligenz) und bei Persönlichkeitstests wird das typische Verhalten (z.B. im Merkmalsbereich der Extraversion) erfasst. [Vgl. ebenda, Einführung, S. VIII.]

²⁵ Vgl. ebenda, Einführung, S. VIII.

tätsstandards zum Inhalt und Ablauf von *Personalauswahlverfahren* und zur *Qualität der Auftragnehmer* [...]“²⁶ beinhaltet. Diese Norm trägt einen großen Nutzen, denn sie verbindet die Fachwelt mit der Praxis. Für Arbeitgeber ist es somit einfacher qualifiziertes Personal zu finden, die zum Unternehmen und auf die ausgeschriebene Stelle passen. Aufgrund der unterschiedlichen berufsrelevanten Fähigkeiten, Fertigkeiten und Kenntnisse, aber auch der unterschiedlichen Lernerfolge am Arbeitsplatz, in der Ausbildung oder in Weiterbildungsmaßnahmen sowie der unterschiedlichen beruflichen Leistungen der Menschen müssen diese Unterschiede mittels einer Eignungsbeurteilung in der Berufswelt abgebildet werden.²⁷ Somit ist es möglich aus dem heterogenen Bewerberpool den passenden Mitarbeiter einer Organisation auszuwählen und in seinen individuellen Potenzialen zu stärken sowie richtig einzusetzen.²⁸ Diese DIN ist fach- sowie arbeitsfeldbezogen und orientiert sich an die verschiedenen Berufsgruppen. Mit der Entstehung dieser Norm besteht eine Chance aus der wissenschaftlich-universitären Fachwelt herauszukommen und in wissenschaftlich-praktikable Erkenntnisse der betrieblichen Praxis umzusetzen.²⁹ Die berufsbezogenen Eignungsbeurteilungen sind für die Organisationen, aufgrund der heterogenen Bewerber, ein wichtiges Handwerkszeug, um nicht nur den „richtigen“ Bewerber zu finden, sondern sich auch erfolgreich im Wettbewerb mit anderen Organisationen zu positionieren.³⁰ Der Test d2 ist beispielsweise ein Instrument, das die Verfahrenshinweise nach DIN 33430 beachtet.³¹

2.1.2 Gütekriterien

Die Evaluation von Messungen orientiert sich an den Gütekriterien Objektivität, Reliabilität (Messgenauigkeit) und Validität (Gültigkeit), Skalierung, Normierung (Ei-

²⁶ Reinmann (2005), S. 17.

²⁷ Vgl. Kersting (2004), S. 55 f.

²⁸ Vgl. ebenda, S. 56.

²⁹ Vgl. Reinmann (2005), S. 18.

³⁰ Vgl. Kersing, (2004), S. 56.

³¹ Vgl. Brickenkamp; Schmidt-Atzert; Liepmann (2010), S. 86.

chung), Testökonomie, Nützlichkeit, Zumutbarkeit, Unverfälschbarkeit und Fairness. Mit der Erfüllung der Testgütekriterien werden wissenschaftlich psychologische Tests empirisch überprüft.³² Die drei Hauptgütekriterien (Objektivität, Reliabilität und Validität) werden in diesem Abschnitt näher erläutert, wobei die Objektivität für diese wissenschaftliche Arbeit keine Relevanz besitzt und nur der Vollständigkeit halber aufgeführt wird. Im Fokus steht die Validität und die Reliabilität, die für die Unterscheidung der Eigen- und Fremdpublikationen einen großen Teil dieser Arbeit einnehmen.

Objektivität

Um für ein Unternehmen passende Bewerber auszuwählen werden häufig psychologische Tests eingesetzt. Die Organisationen sind an objektiv eingeschätzten psychologischen Tests interessiert, d.h. das Hauptkriterium Objektivität bezieht sich auf die Frage, ob diese möglicherweise durch die Person, welche die Messungen durchführt, beeinflusst oder verfälscht wird. Somit lassen Messungen mit einer hohen Objektivität wenig Spielraum für subjektive Entscheidungen des Testleiters. Nachfolgend werden die drei Kategorien Durchführungsobjektivität, die Auswertungsobjektivität sowie die Interpretationsobjektivität betrachtet.³³ Eine hohe Durchführungsobjektivität ist gegeben, wenn für jeden Bewerber völlig identische Untersuchungsvoraussetzungen geschaffen werden. In den meisten Fällen erhalten die Bewerber eine standardisierte Instruktion und bearbeiten Fragebögen oder Leistungstests selbstständig. Zulässige Hilfsmittel oder Zeitvorgaben sind verbindlich geregelt.³⁴

Nach der Datenerhebung folgt die Ergebnisauswertung, hierbei hat der Untersuchungsleiter erneut mehr oder weniger Spielraum. Bei der Auswertungsobjektivität ist es ähnlich wie bei der Durchführungsobjektivität, je geringer der Freiraum für die anleitende Person, desto höher die Auswertungsobjektivität.³⁵ „Generell gilt, dass Testverfahren und insbesondere computergestützte Instrumente meist eine höhere

³² Vgl. Moosburgger; Kelava (2012), S. 8.

³³ Vgl. Kanning (2002), S. 76.

³⁴ Vgl. Kanning (2002), S. 76; Schmidt-Atzert; Amelang (2012), S. 133 f.

³⁵ Vgl. Kanning (2002), S. 77; Schmidt-Atzert; Amelang (2012), S. 135 f.

Auswertungsobjektivität erreichen als Interviews oder auch die Auswertung von Bewerbungsunterlagen.“³⁶

Die Interpretationsobjektivität eines Verfahrens schließt sich der Durchführungs- und Anwendungsobjektivität an. Hierbei wird entschieden, welche inhaltliche Bedeutung einem speziellen Zahlenwert entspricht. Die Interpretationsobjektivität ist hoch, wenn sich die Interpretation automatisch aus dem Zahlenwert ergibt, ohne eine Einwirkung des Testleiters. Für eine einfache Interpretation des Untersuchungsleiters werden den Tests Manuale beigelegt, die u.a. Interpretationsbeispiele aufweisen.³⁷

Validität

Die Validität wird laut den Autoren Moosbrugger und Kelava (2012) wie folgt definiert: „Ein Test gilt dann als valide („gültig“), wenn er das Merkmal, das er messen soll, auch wirklich misst und nicht irgendein anderes.“³⁸

Dieses Gütekriterium ist der Objektivität und der Reliabilität als übergeordnet anzusehen, denn misst der Test bspw. etwas Anderes als vorgesehen, so sind Objektivität und Reliabilität nicht mehr relevant. Jedoch ist zu beachten, dass aufgrund der Komplexität die Validität schwierig zu bestimmen ist.³⁹ Bei der Validität werden die drei Arten Inhaltsvalidität, Konstruktvalidität und Kriteriumsvalidität unterschieden.⁴⁰ „Unter Inhaltsvalidität versteht man, inwieweit ein Test oder ein Testitem das zu messende Merkmal repräsentativ erfasst.“⁴¹ Es wird von einem Repräsentationsschluss ausgegangen, welcher aus den Testitems der repräsentativen Stichprobe

³⁶ Kanning (2002), S. 77.

³⁷ Vgl. ebenda, S. 77 f.

³⁸ Moosbrugger; Kelava (2012), S. 13.

³⁹ Vgl. Hartig; Frey; Jude, (2012), S. 144.

⁴⁰ Vgl. Bühner (2005), S. 88; Brickenkamp.; Schmidt-Atzert; Liepmann (2010), S. 49 ff.

⁴¹ Moosbrugger.; Kelava (2012), S. 15.

des Itemuniversums⁴² darstellt, um das zu erfassende Merkmal zu bestimmen.⁴³ Aufgrund der hypothetischen Betrachtung wird die Inhaltsvalidität nicht bei jedem Test berücksichtigt.⁴⁴ Die Inhaltsvalidität wird nicht durch Koeffizienten bestimmt, sondern mittels logischer und fachlicher Überlegungen.⁴⁵ Bei der Inhaltsvalidität wird dieses subjektive Urteilen von Fachexperten übernommen.⁴⁶

Die Konstruktvalidität gibt wieder, inwieweit die erhaltenen Messwerte der Testverfahren in der psychologischen Theorie geeignet sind, um das interessierende Eignungsmerkmal widerzuspiegeln.⁴⁷ Hierbei werden drei Arten unterschieden. Es gibt die konvergente Validität, die diskriminante Validität und die faktorielle Validität. Um bei den drei Kategorien Aussagen treffen zu können, werden Koeffizienten bestimmt.⁴⁸ „Je mehr Koeffizienten bzw. Wissen zur Konstruktvalidität vorliegen, desto besser kann die Konstruktvalidität beurteilt werden.“⁴⁹

Die konvergente Validität ist eine möglichst hohe Korrelation zwischen unterschiedlichen Messinstrumenten, die das gleiche Merkmal erfassen.⁵⁰

Bei der diskriminanten Validität wird der Zusammenhang zwischen zwei Ausprägungen mit unterschiedlichen Eignungsmerkmalen erfasst.⁵¹ Hierbei sollten die Korrelationen unterschiedlicher Merkmale gering sein, um das zu ermittelnde Konstrukt als rein zu betrachten.⁵²

⁴² Ein Itemuniversum ist in Wirklichkeit nicht bzw. selten vorhanden, vielmehr soll es als Hilfestellung für die gedankliche Konstruktion sein, um das Merkmal zu messen. Dabei wird sich vorgestellt, dass es möglich ist alle Items bspw. zur Konzentration zu erfassen, um daraus eine repräsentative Stichprobe zu erzielen. [Vgl. Schmidt-Atzert; Amelang (2012), S. 145.]

⁴³ Vgl. Moosbrugger.; Kelava (2012), S. 15.

⁴⁴ Vgl. Schmidt-Atzert; Amelang (2012), S. 145.

⁴⁵ Vgl. Bühner (2005), S. 88; Moosbrugger; Kelava (2012), S. 15.

⁴⁶ Vgl. Kanning (2002), S. 70.

⁴⁷ Vgl. Kanning (2002), S. 74; Bühner (2005), S. 89.

⁴⁸ Vgl. Bühner (2005), S. 89.

⁴⁹ Ebenda, S. 89.

⁵⁰ Vgl. Bühner (2011), S. 64; Moosbrugger; Kelava (2012), S.17; Schermelleh-Engel.; Schweizer (2012), S. 346.

⁵¹ Vgl. Bühner (2005), S. 90.

⁵² Vgl. Bühner (2011), S. 64; Schermelleh-Engel; Schweizer (2012), S. 346.

Die faktorielle Validität findet Anwendung bei einer Untersuchung zwischen unterschiedlichen Tests, die mittels einer Faktorenanalyse durchgeführt werden. Die faktorielle Validität fasst homogene, konstruktnahe Inhaltsbereiche zusammen und trennt inhaltsfremde Bereiche.⁵³

„Ein Test weist Kriteriumsvalidität auf, wenn vom Verhalten der Testperson innerhalb der Testsituation erfolgreich auf ein „Kriterium“, nämlich auf ein Verhalten außerhalb der Testsituation, geschlossen werden kann. Die Enge dieser Beziehung ist das Ausmaß an Kriteriumsvalidität (Korrelationsschluss).“⁵⁴

Die häufigste Art der Validität für standardisierte Messverfahren ist die Kriteriumsvalidität.⁵⁵ Dieses Verfahren bezieht sich vor allem auf die praktische Anwendung der Tests für die Prognose von Verhalten und Erleben.⁵⁶ Für die Berechnung der kriteriumsbezogenen Validität wird auch hier wieder der Zusammenhang zwischen einem Eignungsmerkmal (Prädiktor) sowie einem Kriterium, welches den gewünschten Messgegenstand ausdrückt, bestimmt.⁵⁷ Es ist zu beachten, dass bei der Berechnung der Kriteriumsvalidität nicht nur Kriterien einzusetzen sind, die einen positiven Zusammenhang, beispielsweise bei neuen Tests aufweisen, sondern dass es auch Kriterien gibt bei denen von vornherein kein Zusammenhang erwartet wird.⁵⁸

Reliabilität

Die Reliabilität zeigt an wie messgenau die Testergebnisse sind.⁵⁹ Mit der Messgenauigkeit eines Messinstrumentes wird die Zuverlässigkeit eines Verfahrens ermittelt.⁶⁰ Dieses Gütekriterium ist oft Voraussetzung der Objektivität. Dabei gilt, je

⁵³ Vgl. Bühner (2011), S. 64.

⁵⁴ Moosburger; Kelava (2012), S. 18.

⁵⁵ Vgl. Kanning (2002), S. 71.

⁵⁶ Vgl. Moosburgger; Kelava (2012), S. 18.

⁵⁷ Vgl. Bühner (2005), S. 90.

⁵⁸ Vgl. Kanning (2002), S. 73.

⁵⁹ Vgl. Schermelleh-Engel; Werner (2012), S. 120.

⁶⁰ Vgl. Kanning (2002), S. 66.

objektiver die Verhaltensmerkmale gesammelt werden, desto größer ist die Reliabilität.⁶¹ Die Definition der Messgenauigkeit lautet wie folgt: „Ein Test ist dann reliabel (zuverlässig), wenn er das Merkmal, das er misst, exakt, d.h. ohne Messfehler, misst.“⁶² Aus dieser Definition ist zu entnehmen, dass es hierbei nicht darauf ankommt, ob der Inhalt des Tests das geforderte Merkmal misst oder doch etwas anderes, denn dies ist die Aufgabe der Validität.⁶³ Mittels der Messgenauigkeit ist es möglich zu bestimmen, in welchem Testverfahren Messfehler vorliegen. Bei der Entwicklung eines Messinstrumentes wird die Reliabilität empirisch untersucht und mit Hilfe eines Zahlenwertes (Reliabilitätskoeffizient) ausgedrückt.⁶⁴ In der Theorie ist dieser Zahlenwert genau definiert. Er liegt zwischen Null und Eins. Wird ein Verfahren mit dem Reliabilitätskoeffizienten Null bewertet, ist dieses Instrument als unzuverlässig und nicht nutzbar anzusehen. Zeigt die Reliabilität die Zahl Eins auf, ist dieses Instrument maximal zuverlässig, jedoch ist dieser auftretende Fall unwahrscheinlich. Der Reliabilitätskoeffizient wird nicht nur für Entscheidungen genommen, ob der Test in der Praxis eingesetzt wird, sondern ist auch hilfreich bei Auswertungen und Interpretationen. Hierbei gilt: Je höher der Koeffizient ist, desto zuverlässiger ist die Messung des Tests. Der Vorteil für den Auswerter besteht darin, dass die Messergebnisse verwendbar sind.⁶⁵

Um die Korrelationskoeffizienten der Reliabilität zu bestimmen, gibt es drei Schätzmethoden. Dazu gehören die innere Konsistenz bzw. aufgewertete Halbierungskorrelation, die Retest-Korrelation (oder Stabilität) und die Paralleltestkorrelation. Bei der aufgewerteten Halbierungskorrelation wird der Test in zwei gleiche Testhälften geteilt, die erhaltenen Ergebnisse der Testhälften werden miteinander korreliert. Zusätzlich wird die Testlänge in Betracht gezogen, denn vereinfacht gilt, je länger der Test ist, desto gleichmäßiger nimmt die Reliabilität zu. Die innere Konsistenz betrachtet jedes Item als einzelnen Testteil und berücksichtigt den Zusammenhang

⁶¹ Vgl. Bühner (2005), S. 87.

⁶² Moosburgger; Kelava (2012), S. 11.

⁶³ Vgl. Schermelleh-Engel; Werner (2012), S. 120.

⁶⁴ Vgl. Kanning (2002), S. 67.

⁶⁵ Vgl. ebenda, S. 67.

zwischen den Items sowie die Testlänge.⁶⁶ Dabei steht sie für „[...] die Messgenauigkeit des Tests zu einem bestimmten Messzeitpunkt gemessenen Messwerts.“⁶⁷ Die zweite Schätzmethode ist die Retest-Reliabilität oder auch Stabilität genannt. Dabei wird der Test zu zwei verschiedenen Testzeitpunkten durchgeführt, die erhaltenen Ergebnisse werden miteinander korreliert. Es gilt darauf zu achten, dass sich die Zeitintervalle zwischen den beiden Testungen variieren und somit die Korrelation vom Zeitintervall abhängig ist. Des Weiteren sind bei einem zu kurz gewählten Intervall Erinnerungs- und Übungseffekte möglich, die das Ergebnis beeinflussen. Aufgrund dieser Ereignisse steht die Retest-Korrelation für Merkmalsstabilität.⁶⁸ Die dritte Schätzmethode ist die Paralleltestkorrelation. Bei dieser Methode werden von zwei Tests, die gleiche Eigenschaften oder Fähigkeiten mit unterschiedlichen Items erfasst, um die Korrelation zu ermitteln. Diese Methode zeigt auf, inwieweit die Testergebnisse invariant gegenüber einer Variation der Itemauswahl sowie zeitlicher Abweichung sind. Dieser Kennwert ist der Bedingungsstabilität zuzuordnen.⁶⁹

2.1.3 Testauswahl von allgemeinen Leistungstests

In diesen Testverfahren wird zwischen Persönlichkeits- und Leistungstests unterschieden. Um den Inhalt einzugrenzen, werden lediglich die „Allgemeinen Leistungstests“ betrachtet, da sich der Test d2 in die Kategorie der „Allgemeinen Leistungstests“ einordnen lässt.⁷⁰

Unter allgemeine Leistungstests verstehen die Autoren Schmidt-Atzert und Amelang (2012), dass diese Tests allgemeine Voraussetzungen für das Erbringen von kognitiven Leistungen erfasst. Typisch für allgemeine Leistungstests sind Aufmerksamkeits- und Konzentrationstests, denn ohne eine gewisse Aufmerksamkeit und

⁶⁶ Vgl. Bühner (2011), S. 60 f.

⁶⁷ Ebenda, S. 61.

⁶⁸ Vgl. ebenda, S. 61.

⁶⁹ Vgl. ebenda, S. 61.

⁷⁰ Vgl. Brickenkamp.; Schmidt-Atzert; Liepmann (2010), S. 9.

Konzentration ist es schwierig kognitive Leistungen zu erbringen.⁷¹ Beispiele für dieses Verfahren ist der Test d2 (siehe Abschnitt 1.2) oder der FAIR (Frankfurter Aufmerksamkeits-Inventar) bzw. der FAKT-II (Frankfurter Adaptiver Konzentrationsleistungs-Test II), die sich miteinander vergleichen lassen, da sie zu den Konzentrations-tests mit Suchaufgaben gehören. Es werden kritische Reize unter ähnlichen Reizen gesucht, wobei die Unterscheidung in der Art der Zeichen (Figuren unterschiedlicher Art, z.B. FAIR und FAKT-II; bestimmte Buchstaben, welche mit Strichen versehen sind, z.B. Test d2) liegt.⁷²

Der FAIR besteht aus zwei Seiten mit je 16 Zeilen und je 20 Testitems, welche zufallsverteilt sind und die gerichtete Aufmerksamkeit eines Probanden (im Alter zwischen neun bis 72 Jahren) erfasst.⁷³ Die Aufgabe besteht aus der genauen und schnellen Diskrimination visuell ähnlicher Zeichen. Hierbei werden vier Zeichen dargestellt, die aus einem Kreis oder einem Quadrat bestehen und in der Mitte mit zwei oder drei Punkten gekennzeichnet sind. Es gibt zwei Zielitems und zwei Nicht-Zielitems, die eine zeilenweise Anordnung haben.⁷⁴ Mittels einem Markierungsprinzip muss der Proband mit einem Stift die Figuren durchgängig von links nach rechts markieren. Wird ein Zielitem vom Versuchsteilnehmer wahrgenommen ist eine zackenförmige Linie nach oben zu ziehen (siehe Abbildung 1).⁷⁵

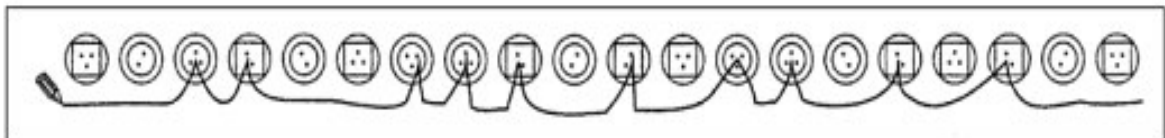


Abbildung 1: Vollständige markierte Musterzeile des FAIR in der Testform A (Quelle: Moosbrugger; Reiß (2004), S. 106)

Die reine Testzeit des Paper-Pencil-Tests beträgt sechs Minuten, dabei werden nach den ersten drei Minuten die Teilnehmer vom Testleiter aufgefordert von der

⁷¹ Vgl. Schmidt-Atzert; Amelang (2012), S. 186.

⁷² Vgl. ebenda, S. 198 f.

⁷³ Vgl. Moosbrugger; Oehlschlägel (1996), S. 325.

⁷⁴ Vgl. Moosbrugger; Reiß (2004), S. 103.

⁷⁵ Vgl. ebenda, S. 105.

ersten auf die zweite Seite zu wechseln, die ebenfalls nach drei Minuten endet.⁷⁶ Die Unterschiede zum Test d2 liegen in der optischen Darstellung des Tests. Es gibt ein vollständiges Markierungsprinzip, jedoch gestaltet sich dieses vor allem bei der Auswertung als kompliziert. Die Anordnung, die Häufigkeit und die Übergänge der Items sind besser als beim Test d2. Es gibt theoretische Konstruktionsprinzipien, die besser durchdacht sind. Jedoch sind die Stichproben zur Reliabilitätsüberprüfung beim FAIR klein, um eine exakte Aussage über die Reliabilität zu treffen. Bis jetzt hat der FAIR wenig Validitätsnachweise und findet wenig Anwendung in der Praxis im Gegensatz zum Test d2.⁷⁷ Der Test d2 besitzt eine hohe Reliabilität, interne Konsistenz und Stabilität, korreliert mit Intelligenz- und Konzentrationstests und ist von verbalen/ numerischen Fähigkeiten unabhängig.⁷⁸ Die Normierung des Tests ist umfassend, die in der neunten Auflage des Test d2 erneuert wurde.⁷⁹ Der Nachteil des Tests d2 liegt darin, dass die Validität nicht für die Daueraufmerksamkeit gilt, die Kennwerte stark von der Quantität beeinflusst werden und es kein vollständiges Markierungsprinzip gibt.⁸⁰

Der FAKT-II findet vor allem Anwendung in der Eignungsdiagnostik der Arbeits- und Betriebspsychologie, in sonstige Bereiche der Angewandten Psychologie sowie der Grundlagenforschung. Dieser Test wird bei der adaptiven Ermittlung individueller Konzentrationsfähigkeit und des Konzentrationsverlaufs bei Erwachsenen eingesetzt. Der FAKT-II ist eine Computerversion zum FAIR. Dabei werden die Größen Leistung, Genauigkeit und Homogenität untersucht. Aufgrund dessen, dass beim FAKT-II ein computerbasierter Algorithmus benutzt wird, wird das Schwierigkeitsniveau der Items an die individuelle Konzentrationsleistung jedes Teilnehmers angepasst. Somit erfolgt die Vorgabe der Items schneller, je höher die Konzentrationsleistung des Einzelnen ist. Mit diesem Test wird zwischen den interindividuellen un-

⁷⁶ Vgl. Moosbrugger; Reiß (2004), S. 106 f.

⁷⁷ Vgl. Hagemeister; Westhoff (1998), S. 229.

⁷⁸ Vgl. ebenda, S. 229.

⁷⁹ Vgl. Brickenkamp; Schmidt-Atzert; Liepmann (2010), S. 5.

⁸⁰ Vgl. Hagemeister; Westhoff (1998), S. 229.

terschiedlichen Leistungsstufen gleiche Beanspruchung geschaffen. Bei entsprechender Messgenauigkeit ist es möglich, dass das Programm des Tests früher beendet wird sowie optional auch Übungseffekte eliminiert.⁸¹

Bei der Anwendung des Tests erscheinen auf dem Bildschirm verschiedene Items (Zielitems und Nicht-Zielitems), die von den Probanden per Tastendruck entscheiden, ob es das Zielitem (welches via Zufallsprinzip erfolgt) ist oder nicht. Die Items entsprechen dem FAIR. Auch bei diesem Test existieren unterschiedliche Testformen der Itemvorgaben, die eine experimentelle Abstufung des situativen Belastungsgrades des Teilnehmers ermöglichen.⁸² Beim FAKT-II ist die Altersobergrenze (acht bis 55 Jahre) im Gegensatz zum Test d2 (neun bis 60 Jahre) niedriger gewählt. Die Auswertung für diesen Test sollte ein ausgebildeter Testleiter übernehmen, damit bei der Interpretation fehlende Testwerte berücksichtigt werden. Der FAKT-II ist gegenüber dem Geschwindigkeits-Genauigkeits-Dilemma unempfindlich. Wie auch beim Test d2 sind mit Übungseffekten zu rechnen. Des Weiteren ist der Nutzen adaptiver gegenüber nicht adaptiver Varianten nachzuweisen. Die Kriteriumsvalidität und Retest-Reliabilität werden in ihrer Ausprägung als mangelhaft eingestuft.⁸³ Der Vorteil des FAKT-II gegenüber dem Test d2 liegt in der computergestützten Version. Aufgrund der vielen Überschneidungen, werden der FAIR und FAKT-II als Alternative zum Test d2 gesehen.⁸⁴

2.2 Test d2 – Aufmerksamkeits- und Belastungs-Test

2.2.1 Begriffserklärung von Aufmerksamkeit und Konzentration

Der Test d2 passt in die Kategorie der allgemeinen Leistungstests (siehe Abschnitt 2.1.3). Diese Tests zeichnen sich dadurch aus, dass sie keinen intellektuellen Anspruch haben und die gestellten Aufgaben in einer gewissen Zeit mit einem bestimmten Tempo zu bearbeiten sind. Gerade der Test d2 verlangt keine Basis- oder

⁸¹ Vgl. Moosburgger; Goldhammer (2004), S. 331.

⁸² Vgl. ebenda, S. 332.

⁸³ Vgl. Weis; Nuerk (2010), S. 220 f.

⁸⁴ Vgl. Schmidt-Atzert; Amelang (2012), S. 199.

Grundbildung des Probanden wie bspw. das Rechnen. Vor allem findet er in der Praxis eine große Beliebtheit. Dieser Test wird nicht nur in der Kinder- und Jugendpsychiatrie häufig angewendet, sondern ist auch im Bereich Personalmanagement hilfreich.⁸⁵ In der deutschen Wirtschaft zählt dieser zu einem der drei am häufigsten angewendeten psychischen Testverfahren. Der Test stellt Anforderungen an die Aufmerksamkeit sowie Konzentrationsfähigkeit und wird bspw. bei der Beratung, Auslese und Platzierung von Personal für Berufe oder Tätigkeitsbereiche eingesetzt.⁸⁶ Ergibt die Anforderungsanalyse, dass eine Konzentrationsfähigkeit für den vorgesehenen Beruf oder die Tätigkeit erforderlich ist, ist es möglich mithilfe diesen Tests individuelle Ausprägungen zu prüfen.⁸⁷ Der Test d2 wird eingesetzt, wenn konzentrierte visuelle Aufmerksamkeit gefordert wird, somit sind „[...] kontinuierlich bestimmte visuelle Reize unter ähnlichen Reizen zu entdecken [...]“⁸⁸.

Die Gründe für die große Beliebtheit in der Wirtschaftspraxis liegen vor allem darin, dass dieser Test eine breite Absicherung der Testgütekriterien liefert, unkompliziert in der Handhabung ist und wenig Zeit sowie Material in Anspruch nimmt.⁸⁹ Voraussetzung bei diesem Test ist, dass der Teilnehmer in der Lage ist visuelle Reize wahrzunehmen sowie zu unterscheiden. Zu den bekanntesten allgemeinen Leistungstests gehören Durchstreichtests, Rechen- oder auch Sortiertests.⁹⁰ In diesem Abschnitt werden die Begrifflichkeiten Aufmerksamkeit und Konzentration herausgearbeitet, da sich beide Begriffe voneinander unterscheiden und somit nicht gleichzusetzen sind. Des Weiteren dienen sie dem Verständnis zur Herstellung von Sachzusammenhängen.

⁸⁵ Vgl. Brickenkamp; Schmidt-Atzert; Liepmann (2010), S. 8.

⁸⁶ Vgl. Brickenkamp (2002), S. 265.

⁸⁷ Vgl. Brickenkamp; Schmidt-Atzert; Liepmann (2010), S. 29.

⁸⁸ Brickenkamp; Schmidt-Atzert; Liepmann (2010), S. 29.

⁸⁹ Vgl. Wiese; Kroj (1972), S. 90; Oswald; Hagen (1997), S.88; Brickenkamp (2002), S. 269; Daseking; Putz (2015), S. 324.

⁹⁰ Vgl. Brickenkamp; Schmidt-Atzert; Liepmann (2010), S. 9 f.

Aufmerksamkeit

Aus der Sicht der psychologischen Wissenschaft hat der Begriff Aufmerksamkeit eine Vielzahl von Modellvorstellungen sowie Theorien.⁹¹ Der Begriff Aufmerksamkeit ist von vielen Autoren unterschiedlich definiert, jedoch gibt es sowohl bei den älteren als auch bei den modernen Definitionen eine Kernaussage. Harley (1992) weist diesbezüglich der Begrifflichkeiten die Eigenschaften zu, Eindrücke zu selektieren und wenn es erforderlich ist, sich über einen längeren Zeitraum darauf zu konzentrieren, um die daraus gewonnenen Erfahrungen und Erkenntnisse weiter zu verarbeiten.⁹² Wird die Definition von den Autoren Schmidt-Atzert et al. (2004) betrachtet, gibt es eine klare Trennung zwischen Aufmerksamkeit und Konzentration, so ist die Aufmerksamkeit die Wahrnehmung vom selektiven Betrachten relevanter Reize und Informationen.⁹³

In neueren neuropsychologischen Modellvorstellungen wird sich an den weiterentwickelten Modellen der Aufmerksamkeit von Posner und Boies (1971) angelehnt.⁹⁴ Auch in diesen Modellvorstellungen werden mehrere Komponenten der Aufmerksamkeit unterschieden. Dabei werden drei gemeinsame Bereiche aus den verschiedenen Modellen herausgefiltert (Cohen, 1993; Kausler, 1991; La Rue 1992; Parasuraman, 1998; Prosiegel, 1998; Sturm, 1997; van Zomeren und Brouwer, 1994):⁹⁵

- Vigilanz (Daueraufmerksamkeit),
- Selektive Aufmerksamkeit,
- Geteilte Aufmerksamkeit.

⁹¹ Vgl. Goldhammer; Mossbrugger (2006), S. 16.

⁹² Vgl. zitiert nach Heubrock; Petermann (2001), S. 11.

⁹³ Vgl. Schmidt-Atzert; Büttner; Bühner (2004), S. 5.

⁹⁴ Vgl. Heubrock; Petermann (2001), S. 18.

⁹⁵ Vgl. zitiert nach Heubrock; Petermann (2001), S. 19.

„Unter *Vigilanz* wird ein Aktivierungsgrad verstanden, der einen Menschen in eine Lage versetzt, eine Veränderung in einer Reihe von gleichförmigen Reizen wahrzunehmen.“⁹⁶ Das Hauptmerkmal der *Vigilanz* ist die Fähigkeit einen gleichmäßig hoher Aktivierungsgrad über einen längeren Zeitraum aufrechtzuerhalten, z.B. zählt die Daueraufmerksamkeit bei der Qualitätskontrolle von Waren, um die fehlerhaften Produkte auszusortieren oder auch Fluglotsen am Tower, die den Leuchtpunkt auf einem Monitor beobachten und bei einer drohenden Kollision eingreifen. Anhand der zwei verschiedenen Beispiele, die jeweils unterschiedliche kognitive Anforderungen haben, gibt es bei der *Vigilanz* Unterteilungen. Die eine ist die einfache *Vigilanz* und die andere die *Vigilanz* mit zusätzlichen kognitiven Anforderungen.⁹⁷

Die selektive oder auch fokussierte Aufmerksamkeit ist, dass aus einer Vielzahl von gleichzeitig auftretenden Reize nur ein einziger bestimmter Reiz betrachtet wird. Als Beispiel wird eine Cocktailparty aufgeführt. Denn hierbei besteht die Möglichkeit, auch in einer großen Gruppe, auf die direkte Konversation seines Gesprächspartners zu achten, ohne sich dabei von anderen Gästen bzw. Gesprächen irritieren zu lassen.⁹⁸

Bei der geteilten Aufmerksamkeit stehen vor allem die Routineanforderungen im Mittelpunkt, d.h. es wird das Vermögen beschrieben, um zwei oder drei Aufgaben gleichzeitig durchzuführen und somit die Aufmerksamkeit zwischen den zwei oder drei Aufgaben aufzuteilen. Als Beispiel ist das Autofahren zu nennen. Der Autofahrer beachtet den Verkehr und unterhält sich auch mit dem Beifahrer. Jedoch ist dies situationsabhängig, weil in einer gefährlichen Verkehrssituation bricht der Autofahrer das Gespräch mit dem Beifahrer ab und widmet sich nur dem Verkehr.⁹⁹

⁹⁶ Heubrock; Petermann (2001), S. 19.

⁹⁷ Vgl. ebenda, S. 19.

⁹⁸ Vgl. ebenda, S. 19.

⁹⁹ Vgl. ebenda, S. 20 f.

Es gibt noch eine weitere Aufmerksamkeitskomponente, welche als Alertness bezeichnet wird. Damit ist eine allgemeine Reaktionsbereitschaft oder kurzfristige Aktivierung der Aufmerksamkeit gemeint.¹⁰⁰ Hierbei werden ebenfalls zwei Bedingungen unterschieden. Zum einen gibt es die phasische Alertness, dabei „[...] geht dem visuellen Reiz ein akustischer Warnreiz voraus, der die Alertness [...]“¹⁰¹ anhebt. Zum anderen gibt es die tonische Alertness, hierbei ist „[...] die Fähigkeit zur schnellen Aufmerksamkeitsaktivierung ohne vorherigen Warnreiz (Aufgabenbezogene Reaktionsgeschwindigkeit) [...]“¹⁰² zu verstehen.

Die verschiedenen Formen der Aufmerksamkeit werden in Tests gemessen, die eine Gemeinsamkeit haben. Diese liegt darin, dass die Probanden die kritischen Reize schnell und genau entdecken. Die Aufmerksamkeitstests unterscheiden sich in der Verwendung der kritischen Reize sowie im Angebot der Bedingungen, denn entscheidend für die Messung der Form der Aufmerksamkeit sind die Bedingungen.¹⁰³

Die Aufmerksamkeitsdiagnostik erscheint als Status-, Eigenschafts- oder Prozessdiagnostik. Verwendung des Eigenschaftskonzeptes der Aufmerksamkeit findet hauptsächlich in der Fragestellung der „Arbeits- und Organisationspsychologie“ und der pädagogischen Psychologie statt.¹⁰⁴ Das Eigenschaftskonzept basiert auf der Annahme, dass eine Person in einem bestimmten Zeitkontingent stabile sowie konsistente Aufmerksamkeitsleistungen¹⁰⁵ aufzeigt.¹⁰⁶

Der Test d2 zeigt eine Aufmerksamkeitsleistung auf, da Probanden dabei bestimmte Reize (d's mit zwei Strichen; siehe Abschnitt 2.2.2) von ähnlichen Reizen unter-

¹⁰⁰ Vgl. Heubrock; Petermann (2001), S. 20.

¹⁰¹ Schmidt-Atzert; Amelang (2012), S. 187.

¹⁰² Heubrock; Petermann (2001), S. 20.

¹⁰³ Vgl. Schmidt-Atzert; Amelang (2012), S. 187.

¹⁰⁴ Vgl. Moosbrugger; F (2006), S. 84.

¹⁰⁵ „Die Aufmerksamkeitsleistung stellt den am häufigsten diagnostisch verwendeten Aufmerksamkeitskennwert dar und basiert in der Regel auf dem physikalischen Leistungsbegriff, d. h. dem Verhältnis von Arbeitsmenge und benötigter Zeit.“ [Moosbrugger; Goldhammer (2006), S. 85.]

¹⁰⁶ Vgl. ebenda, S. 84.

scheiden müssen. Dieser Test wird der selektiven Aufmerksamkeit zugeordnet. Jedoch unterscheidet sich der Test d2 von den klassischen Aufmerksamkeitstests, denn er fordert auch die Konzentration. Dies ist daran zu erkennen, dass die üblichen Aufmerksamkeitstests Reize auf einen Bildschirm projizieren, sodass die Testteilnehmer entscheiden, ob ein Zielreiz oder irrelevanter Reiz aufgezeigt wird. Hierbei fehlt der schnelle, selbst bestimmte Wechsel zur nächsten Aufgabe.¹⁰⁷

Konzentration

Konzentration ist aus dem lateinischen Wort *concedere* abgeleitet und bedeutet „[...] die Fähigkeit der Zentrierung physischer, psychischer und geistiger Kräfte (Aufmerksamkeit) auf einen bestimmten Gegenstand.“¹⁰⁸

Ebenfalls wie bei der Aufmerksamkeit gibt es unterschiedliche Definitionen bei der Konzentration. Vor allem ist die Verschiedenheit des Begriffes durch die Art der Beziehung zwischen Aufmerksamkeit und Konzentration geschuldet,¹⁰⁹ denn die Konzentration wird als eine besondere Form der Aufmerksamkeit gesehen. So definieren bspw. die Autoren Brickenkamp und Karl (1986) die Konzentration als „[...] eine spezifische Art von Aufmerksamkeit, eine Steigerungsform, die sich im Wesentlichen durch Willkürlichkeit und stärkere Fokussierung [...] auszeichnet.“¹¹⁰ Der Autor Rapp (1982) versteht die Konzentration „[...] als besondere Intensitätsform der Aufmerksamkeit [...]“¹¹¹ Die Autoren Westhoff und Hagemeister (2005) definieren „[...] die Konzentration als die absichtsvolle nicht automatisierte Koordination von Handlungsteilen und deren kontrollierte Ausführung.“¹¹² Es ist festzustellen, dass die Konzentration bei allen drei Autoren eine absichtsvolle Handlung sowie Koordination und eine Steigerungsform der Aufmerksamkeit ist. Die Konzentration erscheint

¹⁰⁷ Vgl. Brickenkamp; Schmidt-Atzert; Liepmann (2010), S. 11.

¹⁰⁸ Leitner; Ortner (2004), S. 124.

¹⁰⁹ Vgl. Goldhammer; Moosbrugger (2006), S. 30.

¹¹⁰ Brickenkamp; Karl (1986), S. 195.

¹¹¹ Rapp (1982), S. 24.

¹¹² Westhoff; Hagemeister (2005), S. 20.

nur im Zusammenhang mit Arbeit. Somit geht einer Messung der Konzentration immer ein Arbeitsauftrag einher.¹¹³ Die Unterschiede in der Konzentrationsfähigkeit spiegeln sich in der Schnelligkeit sowie der Genauigkeit bei der Bearbeitung der großen Menge von einfachen fortlaufend gestellten Aufgaben wider.¹¹⁴

Laut der Definition von Westhoff und Hagemeister (2005) erfüllt der Test d2 alle Bedingungen für einen Konzentrationstest:

„Die Leistung eines hirngesunden Probanden erzielt durch (mündliche oder manuelle) Reaktion auf mehr oder weniger einfache (Bilder alltäglicher Gegenstände oder abstrakte Zeichen; Zahlen, Buchstaben oder andere) Reize, die er klar und eindeutig wahrnehmen kann und auf die er eine einfache zu erinnernde Regel anzuwenden hat, indem er absichtsvoll Teilhandlungen so schnell wie möglich bei (sehr niedriger bis sehr hoher) Geübtheit in der Ausführung dieses Tests korrekt koordiniert, kann abgebildet werden in die (sehr niedrige bis sehr hohe) Geschwindigkeit konzentrierten Handelns und den (sehr niedrigen bis sehr hohen) Anteil an Konzentrationsfehlern.“¹¹⁵

Die aufgeführten Merkmale eines Konzentrationstests werden mit dem Test d2 verglichen:¹¹⁶

- die Reaktion auf einfache Reize,
- klares und eindeutiges Wahrnehmen der Reize,
- einfache zu erinnernde Regel, die anwendbar ist,
- diese Regel verlangt absichtsvolle Teilhandlungen schnellst möglich korrekt zu koordinieren¹¹⁷ und
- die Leistung ist mittels der Schnelligkeit sowie der Fehlerrate zu beschreiben.

¹¹³ Vgl. Westhoff; Hagemeister (2005), S. 16; Brickenkamp; Schmidt-Atzert; Liepmann (2010), S. 10.

¹¹⁴ Vgl. Brickenkamp; Schmidt-Atzert; Liepmann (2010), S. 10.

¹¹⁵ Westhoff; Hagemeister (2005), S. 39 f.

¹¹⁶ Vgl. Brickenkamp; Schmidt-Atzert; Liepmann (2010), S. 10.

¹¹⁷ Die Koordination von Teilleistungen ist bspw. bei dem Test d2 das Herausfiltern eines Zeichens, die Prüfung, ob dieses der Regel entspricht (d mit zwei Strichen) und anschließend bei einem richtigen Ergebnis die motorische Reaktion (Durchstreichen) erfolgt. [Vgl. Brickenkamp; Schmidt-Atzert; Liepmann (2010), S. 10.]

Es wurde aufgezeigt, dass der Test d2 alle Kriterien eines Konzentrationstests erfüllt und somit ein Konzentrationstest ist.

Überschneiden sich die Aufmerksamkeitsleistungen mit der Konzentration, wird von konzentrierter Aufmerksamkeit gesprochen. Dabei ist die Aufmerksamkeit über einen längeren Zeitraum unter erschwerten Bedingungen, wie Zeitdruck, Störreize oder Lärm, aufrecht zu erhalten.¹¹⁸ Demzufolge ist der Test d2 auch der konzentrierten Aufmerksamkeit zuzuordnen.¹¹⁹ „Damit gleichbedeutend ist die Feststellung, dass er Konzentration bei der Selektion relevanter Reize oder bei einer Aufgabe, die selektive Aufmerksamkeit verlangt, erfasst.“¹²⁰

2.2.2 Funktionsweise des Tests d2 – Aufmerksamkeits- und Belastungs-Test

In einer dreijährigen Entwicklungszeit erschien der Test d2 erstmals im Jahr 1962 und wurde von dem Psychologieprofessor Rolf Brickenkamp entworfen.¹²¹ Seit dem letzten halben Jahrhundert werden immer wieder neue Konstrukte und Modelle vorgestellt, die für die Einordnung von Tests, welche aus schnell zu bearbeitenden sowie kognitiv wenig anspruchsvollen Aufgaben bestehen, behilflich sind. Denn es gibt nicht das eine theoretische Modell für Tests, sondern mehrere.¹²²

Der Test d2 lässt sich zur damaligen Zeit dem Bourdon-Test¹²³ einordnen, in dessen folgten weitere Durchstreichtests. Es ist jedoch zu beachten, dass die Einordnung des Tests d2 in einen Durchstreichtest anhand des Aufgabenmerkmals nur eine Klassifizierung ist und nicht eine konzeptuelle Einordnung. Denn die Einordnung lässt offen, was der Test misst.¹²⁴ Im deutschsprachigen Raum wird der Test d2 als Konzentrationstest aufgeführt, da er alle Merkmale (siehe Definition Konzentration)

¹¹⁸ Vgl. Schmidt-Atzert; Krumm; Bühner (2008), S. 61.

¹¹⁹ Vgl. Brickenkamp; Schmidt-Atzert; Liepmann (2010), S. 11.

¹²⁰ Ebenda, S. 11.

¹²¹ Vgl. Brickenkamp; Schmidt-Atzert; Liepmann (2010), S. 16.

¹²² Vgl. ebenda, S. 9.

¹²³ Beim Bourdon-Test sind gewisse Buchstaben in einem Text zu finden, die durchgestrichen werden. Die eigentliche Aufgabe in solchen Tests besteht darin Zeichen oder Objekte unter ähnlichen Zeichen oder Objekten zu suchen. [Vgl. ebenda (2010), S. 9.]

¹²⁴ Vgl. ebenda, S. 9.

dafür erfüllt. Aufgrund dessen, dass Aufmerksamkeit als Konzentrationsfähigkeit ein komplexes Konstrukt ist, wird er der selektiven oder der konzentrierten Aufmerksamkeit zugeordnet.¹²⁵

Dieses Instrument wurde im Institut für Sicherheit im Bergbau, Industrie und Verkehr des technischen Überwachungs-Vereins Essen entwickelt.¹²⁶ Das Verfahren ist ursprünglich für Aufgaben in der „Verkehrspsychologie“ konstruiert und fand Anwendung in Medizinisch-Psychologischen Instituten der Technischen Überwachungsvereine.¹²⁷ Vor allem sollte der Test zur Beurteilung der Krafftahreignung behilflich sein,¹²⁸ wie bspw. beim Beraten und/oder der Eignungsauslese bei Fahrlehrer-Anwärtern, Beratung von Berufskrafftahrern oder mehrfachen Versagen von Personen bei Führerscheinprüfungen.¹²⁹ In der neusten Auflage (2010) gibt es fünf Anwendungsgebiete („Verkehrspsychologie“, „Pädagogische Psychologie“, „Klinische und Medizinische Psychologie“, „Neuropsychologie“ und „Arbeits- und Organisationspsychologie“) des Tests d2.¹³⁰

Die Testvorlage des Instruments enthält insgesamt 789 Zeichen. Diese bestehen jeweils aus den Buchstaben d oder p und sind mit ein bis vier Strichen versehen. Es gibt insgesamt 13 unterschiedliche Zeichen, davon treten drei als Zielobjekte auf (d mit zwei Strichen), denn die Aufgabe des Tests d2 besteht darin alle Zielitems (d mit zwei Strichen) im Test zu finden und entsprechend durchzustreichen.¹³¹ In der Abbildung 2 ist ein Ausschnitt des Test d2-R dargestellt, der vereinzelte Testitems veranschaulicht.

¹²⁵ Vgl. Brickenkamp; Schmidt-Atzert; Liepmann (2010), S. 14.

¹²⁶ Vgl. ebenda, S. 15.

¹²⁷ Vgl. Brickenkamp (2002), S.265.

¹²⁸ Vgl. Brickenkamp; Schmidt-Atzert; Liepmann (2010), S. 15.

¹²⁹ Vgl. Brickenkamp (2002), S.265.

¹³⁰ Vgl. Brickenkamp; Schmidt-Atzert; Liepmann (2010), S. 26 ff.

¹³¹ Vgl. ebenda, S. 26.

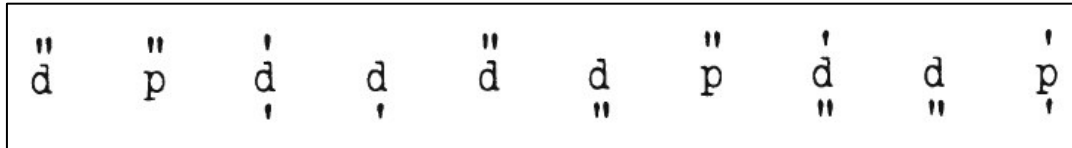


Abbildung 2: Testausschnitt aus dem Test d2-R (Quelle: Schmidt-Atzert; Amelang (2012), S. 193)

Im Test gibt es 14 Zeilen, die von den Probanden bearbeitet werden. Jede Zeile besteht aus 57 Zeichen. Ein Block setzt sich aus drei Zeilen (ein Block umfasst 171 Zeichen) zusammen, welche sich wiederholen. Zum Beispiel entsprechen die Zeilen eins bis drei der Zeile vier bis sechs etc.. In einem Block sind 77 Zielobjekte und 94 Distraktoren¹³² vorhanden. Bei einer vollständigen und fehlerfreien Bearbeitung werden vom Probanden 359 Zielobjekte durchgestrichen. Es ist zu beachten, dass in diesem Test die erste sowie die letzte Zeile nicht gewertet werden. Somit sind die vier mittleren Blöcke relevant, d. h. das bei einer vollständigen und richtigen Bearbeitung des Tests das Ergebnis 308 Zeichen beinhaltet. Die Bearbeitungszeit des gesamten Tests beträgt ohne Pause vier Minuten und 40 Sekunden. Dies entspricht 20 Sekunden pro Zeile.¹³³

In der folgenden Tabelle 1 sind die Häufigkeiten der einzelnen Zeichen, die im Block und im Test insgesamt vorkommen, noch einmal aufgeführt.

Tabelle 1: Häufigkeiten der einzelnen Zeichen im Block und im Test (Quelle: Eigene Darstellung in Anlehnung an Brickenkamp et al. (2010), S. 26)

Zeichen:	2d2	2d1	1d2	2d0	0d2	1d1	1d0	0d1	2p0	0p2	1p0	0p1	1p1
f (Block):	8	6	4	25	28	24	5	9	24	19	9	5	5
f (Test):	32	24	16	100	112	96	20	36	96	76	36	20	20

Anmerkung: Die Zeichen geben Hinweise auf die Anzahl der Striche über bzw. unter dem Buchstaben, bspw. 2d2 entspricht zwei Striche über dem Buchstaben d und zwei Striche unter dem d. Der f

¹³² Beim Test d2 sind Distraktoren die Verwechslungsfehler, welche sich in drei Kategorien einteilen lassen. [Vgl. Brickenkamp; Schmidt-Atzert; Liepmann (2010), S. 23.]

¹³³ Vgl. ebenda, S. 26.

(Block) ist die Häufigkeit eines Zeichens innerhalb eines Blockes, so besteht ein Block aus drei Zeilen. Der f (Test) gibt die Häufigkeit der Zeichen des gesamten Tests an (ausgenommen sind die Zeilen eins sowie 14).¹³⁴

Der Test d2 erscheint simpel, dennoch sind auf drei Grundvoraussetzungen, wie hinreichendes Instruktionsverständnis, motorische Funktionalität der Schreibhand sowie hinreichendes Sehvermögen zu achten.¹³⁵

Für die Testdurchführung werden eine Kurzanleitung mit Testbogen und zwei Kugelschreiber benötigt. Werden andere Stifte eingesetzt, müssen diese die Markierungen auf den Durchschreibbogen sichtbar durchlassen.¹³⁶

2.2.3 Kennwerte des Tests d2 – Aufmerksamkeits- und Belastungs-Test

Die Ergebnisse des Durchstreichtests Test d2-R¹³⁷ werden mit Hilfe eines separaten Durchschlagbogens ermittelt. Wie schon im vorherigen Abschnitt erwähnt, wird die erste und die letzte Zeile des Tests nicht gewertet, da der Proband bei der ersten Zeile zu früh oder zu spät startet. Die letzte Zeile wird bei der Auswertung ausgeschlossen, da die Versuchspersonen bei einer Gruppenuntersuchung, trotz des Stoppsignals, unbemerkt weiterarbeiten können.¹³⁸ Um den Test ordnungsgemäß auszuwerten, werden folgende fünf Kennwerte aus dem Test d2-R ermittelt und ein Bezug zu den alten Kennzahlen des Test d2 hergestellt:¹³⁹

- Konzentrationsleistung (KL),
- Bearbeitete Zielobjekte (BZO),
- Fehlerprozent (F%),
- Konzentrationsverlauf,

¹³⁴ Vgl. Brickenkamp; Schmidt-Atzert; Liepmann (2010), S. 26.

¹³⁵ Vgl. ebenda, S. 29.

¹³⁶ Vgl. ebenda, S. 30.

¹³⁷ „Das zugeführte R bedeutet so viel wie Revision. Genauer gesagt handelt es sich dabei nicht nur um kleinere Verbesserungen, sondern um eine umfassende und *grundlegende* Neufassung. Sie löst sich von einigen überholten Konzepten und basiert konsequent auf neueren wissenschaftlichen Erkenntnissen.“ [Ebenda, S. 5.]

¹³⁸ Vgl. Schmidt-Atzert; Amelang (2012), S. 194.

¹³⁹ Vgl. Brickenkamp; Schmidt-Atzert; Liepmann (2010), S. 32.

- Arbeitsstil.

Die Konzentrationsleistungsfähigkeit wird durch den Konzentrationsleistungswert (KL) ausgedrückt. Bei der Testbearbeitung ist dieser Wert vom Arbeitstempo sowie der Genauigkeit abhängig.¹⁴⁰ Für die meisten Testpersonen stellt die Anweisung des Tests ein Konflikt dar, da sich bei einer Erhöhung des Arbeitstempos in der Regel mehr Fehler einschleichen oder beim Versuch fehlerfrei zu arbeiten das Arbeitstempo langsamer wird.¹⁴¹ Mit diesem Wert wird das Arbeitsverhalten des Testteilnehmers beschrieben. Der Konzentrationswert wird aus der Anzahl entdeckter Zielobjekte (durchgestrichene Buchstaben mit Strich(en)) minus der Anzahl Verwechslungsfehler ermittelt.¹⁴² Im Idealfall wird der Kennwert der Konzentrationsleistung immer gleich ausfallen, unabhängig vom individuellen Tempo und individueller Genauigkeit.¹⁴³ Bei der alten Konzentrationsleistung wurden zwei Werte die Konzentrationsleistung (KL) und die Gesamtmenge minus Gesamtfehlerzahl (GZ-F) angegeben. Der neue KL-Wert entspricht dem alten KL-Wert, der GZ-F-Wert ist eliminiert.¹⁴⁴

Bei der neuen Version des Test d2-R wird die Anzahl der bearbeitenden Zielobjekte (BZO¹⁴⁵) angegeben, anstatt wie beim alten Test die Gesamtzahl der bearbeitenden Zeichen GZ (Zielobjekt sowie Distraktoren). Somit ersetzt die Kennzahl BZO den Wert GZ (Gesamtzahl).¹⁴⁶ Die BZO haben im Vergleich zur GZ folgende drei Vorteile:¹⁴⁷

¹⁴⁰ Vgl. Brickenkmap; Schmidt-Atzert; Liepmann (2010), S. 32.

¹⁴¹ Vgl. ebenda, S. 19.

¹⁴² Vgl. ebenda, S. 32.

¹⁴³ Vgl. ebenda, S. 21.

¹⁴⁴ Vgl. ebenda, S. 19.

¹⁴⁵ Die BZO ergeben sich aus der Anzahl der durchgestrichenen sowie der übersehenen Zielobjekte. [Vgl. ebenda, S. 32.]

¹⁴⁶ Vgl. ebenda, S. 21.

¹⁴⁷ Vgl. ebenda, S. 21.

- einfache und schnelle Ermittlung des Wertes,
- Anfallen von kleineren Zahlen beim Addieren,
- BZO werden zur Berechnung der KL benötigt.

Die Fehlerprozente (F%) ermitteln die Genauigkeit der Testbearbeitung, d. h. sie geben die Anzahl der Fehler bezogen auf die Anzahl bearbeitender Zielobjekte an, diese werden durch folgende Formel ($F\% = 100 \times (AF + VF) / BZO$) über alle Blöcke addiert.¹⁴⁸ Die Anzahl der Fehler bestehen aus Auslassungsfehlern (AF) und Verwechslungsfehlern (VF). Diese helfen bei der Auswertung von der Arbeitsweise des Probanden. Mittels der Studie von Romainczyk (2008) wurde herausgefunden, dass Verwechslungsfehler bei den Versuchsteilnehmern deutlich höher ausfallen, wenn diese ihren Fokus auf Schnelligkeit im Test d2 legen.¹⁴⁹ Der F%-Wert entspricht dem alten F%-Wert.¹⁵⁰

Der Konzentrationsverlauf ist die Konstanz bzw. Variabilität der Testleistung. Es wird bei Bedarf des Test d2-R die Konzentrationsleistung von vier aufeinanderfolgenden Abschnitten ausgewertet. Vor allem, wenn Auffälligkeiten bei den KL-Werten auftreten, besteht die Möglichkeit, diese auf große Unterschiede zu prüfen, um weitere Schritte einzuleiten.¹⁵¹ Im alten Test d2 ist der genannte Wert zeilenweise abgebildet worden, um das Arbeitstempo darzustellen. Mittels der unterschiedlichen Schwierigkeitsgrade der Zeilen, sind bei dem alten KL-Wert Schwankungen beim Arbeitstempo aufgetreten. Des Weiteren wurde ein Vergleich zwischen den höchsten und niedrigsten GZ-Werten ermittelt, um die sogenannte Schwankungsbreite (SB) zu erzielen, dennoch stellte dieser sich als weniger reliabel heraus.¹⁵²

Der Arbeitsstil ist das Verhältnis zwischen Schnelligkeit und Sorgfältigkeit bei der Testbearbeitung.¹⁵³ „Beschrieben wird das Arbeitsverhalten (oder der Arbeitsstil)

¹⁴⁸ Vgl., Brickenkmap; Schmidt-Atzert; Liepmann (2010), S. 32.

¹⁴⁹ Vgl. ebenda, S. 22 f.

¹⁵⁰ Vgl. ebenda, S. 19.

¹⁵¹ Vgl. ebenda, S. 22.

¹⁵² Vgl. ebenda, S. 19.

¹⁵³ Vgl. ebenda, S. 32.

bei der Testbearbeitung (z.B. „sehr schnell und dabei durchschnittlich genau“).¹⁵⁴ Je nach gewählter Bearbeitungsstrategie (schnell oder sorgfältig) des Probanden ist es möglich unterschiedlich schnell oder fehlerfrei zu arbeiten. Die Werte von BZO bzw. GZ und F% werden in eine Grafik eingetragen, um so nicht interpretativ, sondern deskriptiv das Arbeitsverhalten der Probanden zu beschreiben.¹⁵⁵

¹⁵⁴ Ebenda, S. 22.

¹⁵⁵ Vgl. Brickenkmap; Schmidt-Atzert; Liepmann (2010), S. 22.

3 Fragestellung und Hypothesenbildung

Das Hauptanliegen dieser Arbeit ist der Vergleich von Eigen- und Fremdpublikationen, die den Test d2 nutzen, in Hinblick auf die verschiedenen Anwendungsgebiete des Tests, die für den Test ermittelten Gütekriterien Reliabilität und Validität sowie die Stichprobengröße. Durch die Untersuchung der folgenden Fragestellungen und Hypothesen sollen aussagefähige Ergebnisse zur Wichtigkeit und Veränderung des Tests d2 erzielt werden. Die Fragestellungen und Forschungshypothesen ergeben sich aus der Fachliteratur.

Fragestellung 1:

Unterscheiden sich die Anwendungsgebiete bei der Nutzung des Tests d2 – Aufmerksamkeits- und Belastungs-Test von Eigen- und Fremdpublikationen?

Das ursprüngliche Anwendungsgebiet für den Test d2 – Aufmerksamkeits- und Belastungs-Test ist die „Verkehrspsychologie“. Dieses Instrument ist im Institut für Sicherheit im Bergbau, Industrie und Verkehr des technischen Überwachungs-Vereins Essen entwickelt worden.¹⁵⁶ Das Verfahren wurde für Aufgaben in der „Verkehrspsychologie“ konstruiert und fand Anwendung in Medizinisch-Psychologischen Instituten der Technischen Überwachungsvereine.¹⁵⁷ Vor allem sollte der Test d2 der Beurteilung der Kraftfahreignung behilflich sein,¹⁵⁸ wie bspw. beim Beraten und/oder der Eignungsauslese bei Fahrlehrer-Anwärtern, der Beratung von Berufskraftfahrern oder bei mehrfachem Versagen von Personen bei Führerscheineprüfungen.¹⁵⁹ In der neusten Auflage (2010) gibt es fünf Anwendungsgebiete des Tests d2 („Verkehrspsychologie“, „Pädagogische Psychologie“, „Klinische und Me-

¹⁵⁶ Vgl. Brickenkamp; Schmidt-Atzert; Liepmann (2010), S. 15.

¹⁵⁷ Vgl. Brickenkamp (2002), S.265.

¹⁵⁸ Vgl. Brickenkamp; Schmidt-Atzert; Liepmann (2010), S. 15.

¹⁵⁹ Vgl. Brickenkamp (2002), S.265.

dizinische Psychologie“, „Neuropsychologie“ sowie „Arbeits- und Organisationspsychologie“).¹⁶⁰ Diese zeigen, dass der Test in verschiedenen Bereichen der Psychologie eingesetzt wird und sich im Laufe der Zeit als einer der am häufigsten angewandten allgemeinen Leistungstests etabliert hat.¹⁶¹ In dem Zeitraum von 1962 bis 2007 liegen den Autoren Brickenkamp et al. (2010) über 500 Publikationen vor, welche in Fachzeitschriften veröffentlicht wurden.¹⁶²

Aufgrund der verschiedenen Anwendungsbereiche des Tests d2, die unterschiedlichen Praxisfelder der Anwender¹⁶³ sowie der ständigen Weiterentwicklung durch die Verbesserungsvorschläge bzw. Kritiken von Eigen- und Fremdautoren¹⁶⁴, wird in dieser Arbeit erwartet, dass sich der Fokus des ursprünglichen Anwendungsgebietes „Verkehrspsychologie“ für den Test d2 bei den Eigen- und Fremdpublikationen geändert hat.

Für diese Untersuchung ist zu beachten, dass die Fremdautoren dominierender auftreten und auch in unterschiedlichen Fachgebieten tätig sind als die Eigenautoren. Dadurch werden von den Fremdautoren mehr Anwendungsgebiete bedient bzw. es wird weitgehender geforscht als bei den Eigenautoren. Davon ausgehend wird angenommen, dass sich die Anzahl der Anwendungsgebiete zwischen Eigen- und Fremdpublikationen unterscheidet.

H₀: Die Domänen der Verwendung des Tests d2 – Aufmerksamkeits- und Belastungs-Test haben sich in der betrachteten Publikationszeit bei den Eigen- und Fremdpublikationen nicht geändert.

¹⁶⁰ Vgl. Brickenkamp; Schmidt-Atzert; Liepmann (2010), S. 26 ff.

¹⁶¹ Vgl. Schorr (1995), S. 7; Steck (1997), S. 273; zitiert nach Schmidt-Atzert, Amelang (2012), S. 5.

¹⁶² Vgl. Brickenkamp; Schmidt-Atzert; Liepmann (2010), S. 8.

¹⁶³ Vgl. Schorr (1995), S. 6; Steck (1997), S. 273.

¹⁶⁴ Vgl.; Heinrich (1973), S. 118 ff.; Zaworka; Hand (1981), S. 41; Rauchfleisch (1983), S. 247 ff.; Westhoff; Kluck (1983), S. 310 ff.; Eser (1991), S. 137 ff.; Moosbrugger; Oehlschlägel (1991), S.42 ff.; Fay (1995), S. 293 ff.; Steffens (2004), S. 167; Krumm; Schmidt-Atzert; Michalczyk; Danthiir (2008), S. 207 ff.; Brickenkamp; Schmidt-Atzert; Liepmann (2010), S. 16; Funsch, K.; Martin, A. B.; Halder-Sinn, P. (2011), S. 14.

H₁: Die Domänen der Verwendung des Tests d2 – Aufmerksamkeits- und Belastungs-Test haben sich in der betrachteten Publikationszeit bei den Eigen- und Fremdpublikationen geändert.

H₀: Die Fremdpublikationen weisen nicht mehr Anwendungsgebiete auf als die Eigenpublikationen.

H₂: Die Fremdpublikationen weisen mehr Anwendungsgebiete auf als die Eigenpublikationen.

Fragestellung 2:

Unterscheiden sich die Gütekriterien Reliabilität und Validität ohne und mit den spezifischen Verwendungen des Tests d2 – Aufmerksamkeits- und Belastungs-Test in Eigen- und Fremdpublikationen?

Messungsgenauigkeit und Ungültigkeit von Tests treten bei wissenschaftlichen Untersuchungen auf. Um dies zu vermeiden, werden diese auf Reliabilität und Validität geprüft.¹⁶⁵ Die Untersuchung von Heinrich (1973) weist bspw. andere Reliabilitätskoeffizienten auf als bei Brickenkamp.¹⁶⁶ Bei einer Studie von Westhoff und Kluck (1984) hat sich ergeben, dass es einen Zusammenhang zwischen der Intelligenz und der Konzentration gibt, was Brickenkamp vorher bestritten hat.¹⁶⁷ Eine andere Untersuchung von Eser (1987) hat die Reliabilitäts- und Validitätsaspekte des Tests d2 bei verhaltensgestörten Kindern und Jugendlichen untersucht, weil in diesem Anwendungsgebiet wenige Studien vorliegen.¹⁶⁸ Daher wird in dieser Arbeit erwartet, dass es Unterschiede zwischen den publizierten Gütekriterien Reliabilität und Validität mit sowie ohne Betrachtung der spezifischen Verwendungen des Tests d2-Aufmerksamkeits- und Belastungs-Test in Eigen- und Fremdpublikationen gibt.

¹⁶⁵ Vgl. Moosbrugger; Kelava (2012), S. 11 ff.

¹⁶⁶ Vgl. Heinrich (1973), S. 121 f.

¹⁶⁷ Vgl. Westhoff; Kluck (1984), S. 175 f.

¹⁶⁸ Vgl. Eser (1987), S. 74 ff.

H₀: Es gibt keine Unterschiede in den dokumentierten Angaben des Gütekriteriums Reliabilität ohne Betrachtung der spezifischen Verwendungen des Tests d2 – Aufmerksamkeits- und Belastungs-Test in Eigen- und Fremdpublikationen.

H₁: Es gibt Unterschiede in den dokumentierten Angaben des Gütekriteriums Reliabilität ohne Betrachtung der spezifischen Verwendungen des Test d2 – Aufmerksamkeits- und Belastungs-Test in Eigen- und Fremdpublikationen.

H₀: Es gibt keine Unterschiede in den dokumentierten Angaben des Gütekriteriums Reliabilität bei spezifischen Verwendungen des Tests d2 – Aufmerksamkeits- und Belastungs-Test in Eigen- und Fremdpublikationen.

H₂: Es gibt Unterschiede in den dokumentierten Angaben des Gütekriteriums Reliabilität bei spezifischen Verwendungen des Tests d2 – Aufmerksamkeits- und Belastungs-Test in Eigen- und Fremdpublikationen.

H₀: Es gibt keine Unterschiede in den dokumentierten Angaben des Gütekriteriums Validität ohne Betrachtung der spezifischen Verwendungen des Tests d2 – Aufmerksamkeits- und Belastungs-Test in Eigen- und Fremdpublikationen.

H₃: Es gibt Unterschiede in den dokumentierten Angaben des Gütekriteriums Validität ohne Betrachtung der spezifischen Verwendungen des Tests d2 – Aufmerksamkeits- und Belastungs-Test in Eigen- und Fremdpublikationen.

H₀: Es gibt keine Unterschiede in den dokumentierten Angaben des Gütekriteriums Validität bei spezifischen Verwendungen des Tests d2 – Aufmerksamkeits- und Belastungs-Test in Eigen- und Fremdpublikationen.

H₄: Es gibt Unterschiede in den dokumentierten Angaben des Gütekriteriums Validität bei spezifischen Verwendungen des Tests d2 – Aufmerksamkeits- und Belastungs-Test in Eigen- und Fremdpublikationen.

Fragestellung 3:

Unterscheiden sich beim Test d2 – Aufmerksamkeits- und Belastungs-Test die Eigen- und Fremdpublikationen im durchschnittlichen Stichprobenumfang?

Die Stichprobengröße ist bei einigen Untersuchungen wichtig, um das zu erforschende Problem leichter zu erkennen. Es kommt darauf an, in welchem Umfang etwas gemessen wird. Denn wird die Stichprobengröße zu klein gewählt, ist es möglich, dass das Forschungsproblem unerkannt bleibt. Je höher die Stichprobengröße ist, desto besser werden die Forschungsprobleme ermittelt.¹⁶⁹ Es gibt Empfehlungen, wie groß ein Stichprobenumfang sein sollte, um möglichst genaue Ergebnisse einer Untersuchung zu erzielen. Stichprobengrößen unter 100 Fällen sind kritisch zu betrachten.¹⁷⁰ Vor allem bei Anwendungen, die bspw. die Komplexität eines analytischen Modells und deren Güte messen, sollte mit Bedacht gewählt werden.¹⁷¹ Demnach liegen Stichprobenumfänge von 200 bis 250 Fällen im idealen Bereich, um eine verbesserte Ergebnisqualität zu erlangen und sollten laut Empfehlung angestrebt werden.¹⁷²

Jedoch sind in den ausgewählten Publikationen unterschiedliche Stichprobengrößen festzustellen. Bei zu kleinen Stichprobenumfängen besteht die Möglichkeit, dass der zu untersuchende Bereich zu breit gewählt wird und die Resultate nicht brauchbar sind. Oftmals ist eine große Stichprobengröße nicht umzusetzen, aufgrund der Schwierigkeit vom Auffinden bestimmter Probanden, welche in speziellen Untersuchungen (z.B. bei verhaltensgestörten Kindern und Jugendlichen (Eser, 1987)) zur Verfügung stehen.¹⁷³ Auch die Untersuchungskriterien führen zu unterschiedlichen bzw. meist zu kleinen Stichprobengrößen,¹⁷⁴ wie bspw. bei der Retest-Reliabilität. Dazu muss ein Proband den Test zu zwei unterschiedlichen Zeitpunkten

¹⁶⁹ Rumsey (2008), S. 77 ff.

¹⁷⁰ Boomsma (1985), S. 240.

¹⁷¹ Vgl. Hu; Bentler (1998), S. 446 f.; Homburg; Klarmann (2006), S. 733.

¹⁷² Vgl. Homburg; Klarmann (2006), S. 733.

¹⁷³ Vgl. Härtler (2014), S. 119.

¹⁷⁴ Vgl. Schmidt-Atzert, Amelang (2012), S. 141 ff.

durchführen.¹⁷⁵ Ein weiterer Grund für einen kleinen Stichprobenumfang ist, dass die Gewinnung eines einzelnen Beobachtungswertes zu teuer ist.¹⁷⁶ Davon ausgehend wird erwartet, dass es Unterschiede in der durchschnittlichen Stichprobengröße und auch bei der Betrachtung der unterschiedlichen Anwendungsgebiete beim Test d2 in Eigen- und Fremdpublikationen gibt.

H₀: Die Fremdpublikationen haben keinen größeren durchschnittlichen Stichprobenumfang als die Eigenpublikationen.

H₁: Die Fremdpublikationen haben einen größeren durchschnittlichen Stichprobenumfang als die Eigenpublikationen.

H₀: Die Fremdpublikationen haben keinen größeren durchschnittlichen Stichprobenumfang in den verschiedenen Anwendungsgebieten als die Eigenpublikationen.

H₂: Die Fremdpublikationen haben einen größeren durchschnittlichen Stichprobenumfang in den verschiedenen Anwendungsgebieten als die Eigenpublikationen.

Fragestellung 4:

Unterscheiden sich die Reliabilitäts- und Validitätsangaben für den Test d2 – Aufmerksamkeits- und Belastungs-Test bei Eigen- und Fremdpublikationen bezogen auf unterschiedliche Stichprobenumfänge?

Bei Fragestellung 3 wurde schon erläutert, dass eine höhere Stichprobengröße besser geeignet ist, um Forschungsprobleme zu erkennen. Um die Reliabilität und Validität zu beurteilen, ist insbesondere die Abhängigkeit des Stichprobenumfangs und der Modellkomplexität zu beachten.¹⁷⁷ „Bei großen Stichproben und geringer Mo-

¹⁷⁵ Vgl. Bühner (2011), S. 61.

¹⁷⁶ Vgl. Schermelleh-Engel; Werner (2012), S. 123; Härtler (2014), S. 119.

¹⁷⁷ Vgl. Homburg; Baumgartner (1995), S. 172.

dellkomplexität beispielsweise sollten die Anforderungen an die Kriterien höher ausfallen.“¹⁷⁸ Die Stichprobengröße beeinflusst die Reliabilitäten als Korrelationskoeffizienten. Es ist darauf zu achten, dass die Stichprobenumfänge groß sind. Vor allem bei der Schätzung der Retest-Reliabilität sind öfter Stichprobengrößen unter $N=100$ zu finden.¹⁷⁹ Dabei kann es vorkommen, dass der Test ungenau misst und somit die einzelne Messung unzuverlässig ist.¹⁸⁰ Die Reliabilität und die Validität stehen im Zusammenhang, denn es gilt: „Je reliabler ein Test ist, desto höher *kann* dessen Validität ausfallen.“¹⁸¹ Die Autoren Schmidt-Atzert und Amelang (2012) teilen mit, dass die Stichprobengrößen bei der Validität oft sehr gering sind und weisen darauf hin, dass die Validitätskoeffizienten der Untersuchungen nach unten oder oben vom tatsächlichen Wert verschoben werden können.¹⁸² Es wird angenommen, dass größere Stichprobenumfänge bessere Reliabilitäts- und Validitätsangaben im Test d_2 bei Eigen- und Fremdpublikationen aufweisen.

H₀: Die Fremdpublikationen weisen keine höheren Reliabilitäten aus, aufgrund ihrer größeren Stichproben als die Eigenpublikationen.

H₁: Die Fremdpublikationen weisen höhere Reliabilitäten aus, aufgrund ihrer größeren Stichproben als die Eigenpublikationen.

H₀: Die Fremdpublikationen weisen keine höheren Validitäten aus, aufgrund ihrer größeren Stichprobe als die Eigenpublikationen.

H₂: Die Fremdpublikationen weisen höhere Validitäten aus, aufgrund ihrer größeren Stichprobe als die Eigenpublikationen.

¹⁷⁸ Himme (2007), S. 385.

¹⁷⁹ Vgl. Schmidt-Atzert; Amelang (2012), S. 141 f.

¹⁸⁰ Vgl. Schermelleh-Engel; Werner (2012), S. 123; Schmidt-Atzert; Amelang (2012), S. 141 f.

¹⁸¹ Schmidt-Atzert; Amelang (2012), S. 154.

¹⁸² Vgl. ebenda, S. 160.

4 Untersuchungsdesign

4.1 Definitionsphase

In der Praxis der diagnostischen Testverfahren im Bereich der Leistungstests wird der Test d2 von deutschen Psychologinnen und Psychologen häufig angewendet. Eine Synopse zu Testpräferenzen von drei unterschiedlichen Untersuchungen in verschiedenen Jahren (Grubitzsch und Rexillus (1978), Schorr (1995), Steck (1997)) hat ergeben, dass der Test d2 unter den 20 aufgeführten Tests Platz sieben belegt.¹⁸³ Bei einer deutschlandweiten Umfrage von Schorr (1995) an Psychologen hat sich gezeigt, dass der Test d2 am dritt häufigsten unter den handelsüblichen Testverfahren genutzt wird.¹⁸⁴

Im Mittelpunkt der vorliegenden Untersuchung steht die Analyse der Gütekriterien Reliabilität und Validität des Tests d2 in Eigen- und Fremdpublikationen sowie in den unterschiedlichen Anwendungsbereichen. Es gibt viele Untersuchungen zum Test d2 vom Testautor sowie von Co- und Fremdautoren. In Anbetracht vieler weiterer Verfahren im Bereich der Leistungstests, die ebenfalls zur Anwendung kommen, wie bspw. der Hamburger-Wechsler-Intelligenztest für Erwachsene (HAWIE), der Intelligenz-Struktur-Test (IST) oder der Benton-Test,¹⁸⁵ stellt sich die Frage, ob der Test d2, aufgrund seines langen Bestehens (seit 1962) noch eine große Rolle für die diagnostische Psychologie besonders für die „Arbeits- und Organisationspsychologie“, einnimmt bzw. in welchen Anwendungsgebieten er besonders häufig vertreten ist. Des Weiteren ist fraglich, wie sich die Entwicklungen bzw. Unterschiede der Testergebnisse zu den Gütekriterien Reliabilität und Validität in den Eigen- und Fremdpublikationen verhalten. Aufgrund der langen Existenz des Tests d2

¹⁸³ Vgl. Steck (1997), S. 273 ff.

¹⁸⁴ Vgl. Schorr (1995), S. 7.

¹⁸⁵ Vgl. Schorr (1995), S. 7; Steck (1997), S. 273.

ist unklar, ob die damaligen Untersuchungen zum Test d2 schon exakte Werte ermittelt haben. Die Stichprobengrößen in den einzelnen Publikationen sind unterschiedlich. Es bleibt zu untersuchen, ob diese die Reliabilität oder Validität beeinflussen und dies sich auch bei den Ergebnissen der Eigen- und Fremdpublikationen auswirkt. Mittels der großen Auswahl an Artikeln liegt der Fokus auf insgesamt acht Fachzeitschriften im Zeitraum von 1962 bis 2015.

4.1.1 Untersuchungsablauf

Damit die Hypothesen dieser Arbeit beantwortet werden, sind für diese Untersuchungen ausgewählte Artikel des Tests d2 zu analysieren.

Die Publikationen sind nach Eigen- und Fremdautoren geordnet. Unter Eigenautoren versteht die Verfasserin den Testautor (Brickenkamp) sowie Coautoren, die mit dem Testautor zusammen zum Test d2 publiziert haben, wie bspw. Schmidt-Atzert; Bühner oder Liepmann. Fremdautoren (bspw. Westhoff; Moosbrugger oder Heinrich) sind diese, welche nicht mit den Test- bzw. Coautoren publiziert haben. Häufig üben Fremdautoren Kritik am Test d2 oder erbringen neue Forschungsergebnisse. Die Artikel werden aus den folgenden acht unterschiedlichen Fachzeitschriften, welche einen wirtschaftlichen Hintergrund aufzeigen, im Publikationszeitraum von 1962 bis 2015 verwendet:

- Diagnostica – Zeitschrift für Psychologische Diagnostik und Differentielle Psychologie,
- Zeitschrift für Arbeits- und Organisationspsychologie A&O,
- Zeitschrift für Personalpsychologie,
- Zeitschrift für experimentelle und angewandte Psychologie,
- Zeitschrift für experimentelle Psychologie,
- Zeitschrift für Differentielle und Diagnostische Psychologie,
- Psychologische Beiträge; Vierteljahresschrift für alle Gebiete der Psychologie und
- Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie.

Die herausgesuchten Publikationen sind in der Universitätsbibliothek Chemnitz und in der Sächsischen Landesbibliothek – Staats- und Universitätsbibliothek Dresden

vorhanden. Von der Zeitschrift „*Diagnostica – Zeitschrift für Psychologische Diagnostik und Differentielle Psychologie*“ sind die Jahrgänge 1962 bis 1965 nicht zu finden, ebenso wie die Jahrgänge 1962 bis 1994 der „*Zeitschrift für experimentelle Psychologie*“. Zur Anwendung kommen die Artikel, in denen der Test d2 als Kontrollinstrument für andere Tests, Untersuchungsinstrument für Konstrukte wie Konzentration und Aufmerksamkeit, Leistungsinstrument in Befragungen sowie für „Testbezogene Untersuchungen“ benutzt wird. Insgesamt sind 136 Publikationen mit dem Test d2 aus den vierteljährlich erscheinenden Fachzeitschriften im betrachteten Zeitraum herausgesucht worden. Für die Untersuchung werden 124 Artikel verwendet. Davon gibt es fünf Artikel (Brickenkamp; Rump, 1966; Rauchfleisch, 1983; Eser, 1987; Brickenkamp, 1991; Schmidt-Atzert; Ising, 1997), die je zwei bis drei Untersuchungen enthalten, welche für die Reliabilitäts- bzw. Validitätsuntersuchungen genutzt werden (siehe Tabellen 5; 6; 7; 8; 10; A-6 und A7). Zwölf Veröffentlichungen sind aus der Untersuchung auszuschließen, da über den Test d2 keine weiteren Aussagen getroffen werden. Eine Auflistung der verwendeten Publikationen aus den jeweiligen Zeitschriften mit Erscheinungsjahr, Autor und der Einordnung von Eigen- oder Fremdpublikationen ist im Anhang (siehe Anhang Tabelle A-1) zu finden. Des Weiteren werden die Artikel in die einzelnen und erwähnten Anwendungsbereiche eingeteilt. Die Kategorie „Testbezogene Untersuchungen“ ist von der Verfasserin hinzugefügt. Hierbei wird sich auf Artikel bezogen, die reine Testuntersuchungen beinhalten und sich nicht in den anderen Themengebieten einordnen lassen.

Wenige der betrachteten Publikationen befassen sich speziell mit der Validität und Reliabilität des Tests d2 und werden somit zur Untersuchung dieser Gütekriterien einbezogen, weil viele Publikationen inhaltlich andere Problemstellungen behandeln bzw. den Test d2 für die Reliabilität bzw. Validität modifizieren. Für allgemeinere Untersuchungen wie die der Stichprobengrößen oder der Anwendungsgebiete und Zeiträume werden dagegen mehr Artikel berücksichtigt, da hierbei die Spezifität der Artikel außer Acht gelassen wird. Um die Reliabilitäts- und Validitätsergebnisse besser vergleichen zu können, handelt es sich dabei ausschließlich um Publikationen, welche keine Modifikation des Tests d2 beinhalten. In der Tabelle 2 werden die Artikel aufgezeigt, die für die Untersuchungen der Reliabilität und Validität genutzt werden.

**Tabelle 2: Publikationen für Reliabilitäts- und Validitätsuntersuchungen
(Quelle: Eigene Darstellung)**

Jahr	Fachzeitschrift	Autor	Artikel
1966	Diagnostica	Brickenkamp, R.; Rump, G.	Die Stabilität des Aufmerksamkeits- Belastungs-Tests (Test d2) über längere Zeitabschnitte
1971	Diagnostica	Dingel, W.	Mitteilung über Zuverlässigkeitsas- pekte des Test d2
1971	Diagnostica	Marschner, G.	Der Revisions-Test als allgemeiner Leistungstest
1972	Zeitschrift für exper- imentelle und an- gewandte Psycho- logie	Wiese, W.; Kroj, G.	Untersuchung über den Zusam- menhang zwischen Intelligenz (Wechsler) und Konzentrationsfä- higkeit (Test d2 nach Brickenkamp)
1973	Diagnostica	Sommer, G.	Die Problematik der Erfassung von "Konzentration", dargestellt am KLT
1983	Diagnostica	Rauchfleisch, U.	Zur Reliabilität und Validität des Aufmerksamkeits- Belastungs- Tests (Test d2) bei Patienten mit hirndiffusen Psychosyndrom und neurotischen Störungen
1983	Diagnostica	Kühnlein, I.; Hahl- weg, K.; Müller, G.; Sevenitz, B.	Überprüfung der SPM von Raven für den psychiatrischen Bereich und Entwicklungen von Kurzformen
1987	Diagnostica	Eser, K.-H.	Reliabilitäts- und Validitätsaspekte des Aufmerksamkeits- Belastungs- Tests (Test d2) bei verhaltensge- störten Kindern und Jugendlichen
1991	Diagnostica	Brickenkamp, R.	Fehlinterpretationen von Testleis- tungen? Anmerkung zum Beitrag "Konzentrationsleistung ohne Kon- zentration?"
1994	Zeitschrift für Ar- beits- und Organi- sationspsychologie	Schmidt, J.U.	Zur Validität von Schulkenntnistests

1997	Zeitschrift für Differentielle und Diagnostische Psychologie	Schmidt-Atzert, L.; Ising, M.	Ein Beitrag zur Konstruktvalidität von d2 und Revisionstest
2000	Psychologische Beiträge; Vierteljahresschrift für alle Gebiete der Psychologie	Merten, T.	Die Computerversion des d2 und die Frage der Transferäquivalenz
2003	Diagnostica	Westhoff, K.; Graubner, J.	Konstruktion eines Komplexen Konzentrationstests
2008	Zeitschrift für Differentielle und Diagnostische Psychologie	Krumm, S.; Schmidt-Atzert, L.; Michalczyk, K. M.; Danthiir, V.	Speeded Paper-Pencil Sustained Attention and Mental Speed Tests - Can Performances Be Discriminated?
2011	Diagnostica	Funsch, K.; Arias Martin, B.; Halder-Sinn, P.	Differenzieller Konzentrationstest für Kinder (DKT-K)
2012	Zeitschrift für Differentielle und Diagnostische Psychologie	Krumm, S.; Schmidt-Atzert, L.; Schmidt, S; Zenses, E.-M.; Stenzel, M.	Attention Tests in Different Stimulus Presentation Modes - A Facet Model of Performance in Attention Tests
2015	Diagnostica	Becker, N.; Preckel, F.; Karbach, J.; Raffel, N.; Spinath, F. M.	Die Matrizenkonstruktionsaufgabe

Insgesamt kommen 17 Publikationen für die Untersuchungen der Reliabilität und Validität in Betracht. Die Aufteilung in die verschiedenen Anwendungsbereiche der einzelnen Artikel wird in den Ergebnisanalysen vorgenommen (siehe Kapitel 5). Bei der Untersuchung für die durchschnittliche Stichprobengröße werden von den ausgewählten 124 Artikeln nur 90 verwendet (siehe Tabellen A-2 und A-3). Dafür wurden folgende Publikationen ausgeschlossen: Artikel, die den Test d2 nur mit anderen Tests erwähnen, wo er als Vergleich für bzw. mit anderen Tests zur Verfügung steht oder der Test d2 eingesetzt wird, aber kein Zusammenhang mit der je-

weiligen Untersuchung entsteht und somit keine weiteren Daten zum Test d2 vorliegen.

Um die durchschnittlichen Stichprobengrößen besser einzustufen, werden die Vorgaben der Autoren Boomsma (1985) und Homburg; Klarmann (2006) beachtet. Hierbei gilt, dass Stichprobenumfänge unter 100 Fällen, aufgrund der beeinträchtigten Ergebnisse, nicht zu empfehlen sind.¹⁸⁶ Für eine Verbesserung der Ergebnisqualität wird der Bereich von 200 bis 250 Fällen bei den Stichprobengrößen angegeben.¹⁸⁷ Bei den folgenden Auswertungen werden die Empfehlungen der Stichprobengrößen beachtet.

4.1.2 Stichprobenbeschreibung der untersuchten Eigen- und Fremdpublikationen

In den letzten 53 Jahren erscheint der Test d2 in den genannten Zeitschriften regelmäßig. Die Tabelle 3 zeigt die Anzahl der Gesamtpublikationen sowie die Anzahl der Eigen- und Fremdpublikationen in den einzelnen Fachzeitschriften für den betrachteten Zeitraum auf. Am häufigsten sind die Artikel (Gesamtanzahl 38 Artikel) in der Fachzeitschrift „Diagnostica – Zeitschrift für Psychologische Diagnostik und Differentielle Psychologie“ zu finden. Auch bei den Eigen- und Fremdpublikationen enthält diese Zeitschrift die meisten Veröffentlichungen mit dem Test d2 – Aufmerksamkeits- und Belastungs-Test (Eigenpublikation: sieben Artikel; Fremdpublikationen: 31 Artikel). Die wenigsten Artikel sind in der „Zeitschrift für experimentelle Psychologie“ erschienen, Gesamtanzahl zwei Artikel, die beide den Fremdpublikationen zugeordnet sind. Insgesamt ist eindeutig erkennbar, dass die 105 Fremdpublikationen dominieren. Die Eigenpublikationen weisen 19 Artikel in den vergangenen 53 Jahren auf.

¹⁸⁶ Vgl. Boomsma (1985), S. 240.

¹⁸⁷ Vgl. Homburg; Klarmann (2006), S. 733.

**Tabelle 3: Anzahl der Publikationen in den verschiedenen Zeitschriften
(Quelle: Eigene Darstellung)**

Name der Zeitschrift	Anzahl der Eigenpublikationen	Anzahl der Fremdpublikationen	Gesamtanzahl der Publikationen
Diagnostica – Zeitschrift für Psychologische Diagnostik und Differentielle Psychologie	7	31	38
Zeitschrift für Arbeits- und Organisationspsychologie A&O	2	18	20
Zeitschrift für experimentelle und angewandte Psychologie	1	19	20
Zeitschrift für Differentielle und Diagnostische Psychologie	6	10	16
Psychologische Beiträge; Vierteljahresschrift für alle Gebiete der Psychologie	0	12	12
Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie	0	9	9
Zeitschrift für Personalpsychologie	3	4	7
Zeitschrift für experimentelle Psychologie	0	2	2
Gesamt	19	105	124

4.1.3 Stichprobenbeschreibung der Publikationszeiträume von Eigen- und Fremdpublikationen

Anhand des selbst erstellten Diagramms (siehe Abbildung 3) wird mit gestapelten Säulen die Anzahl der erschienenen Eigen- und Fremdpublikationen im Zeitraum von 1962 bis 2015 dargestellt. Hierbei wurden 124 Artikel betrachtet, ohne Rücksicht auf Reliabilität und Validität zu nehmen. Wie in Abschnitt 4.1.1 beschrieben, wurden die Artikel ausgesucht, welche den Test d2 als Kontrollinstrument, Prüfung und Neuerungen der Ergebnisse sowie zu kritischen Betrachtungen beinhalten. Betrachtet man die Gesamtzahl der Publikationen, fällt ein Anstieg der Publikationen ab den 80er Jahren auf, wohingegen im letzten Jahrzehnt die Anzahl der Publikationen leicht zurückgegangen ist. Die größte Artikelanzahl in einem Jahr liegt bei sechs Veröffentlichungen. Es zeigt sich im Vergleich, dass die Fremdpublikationen dominieren und regelmäßig in den Jahrzehnten erscheinen.

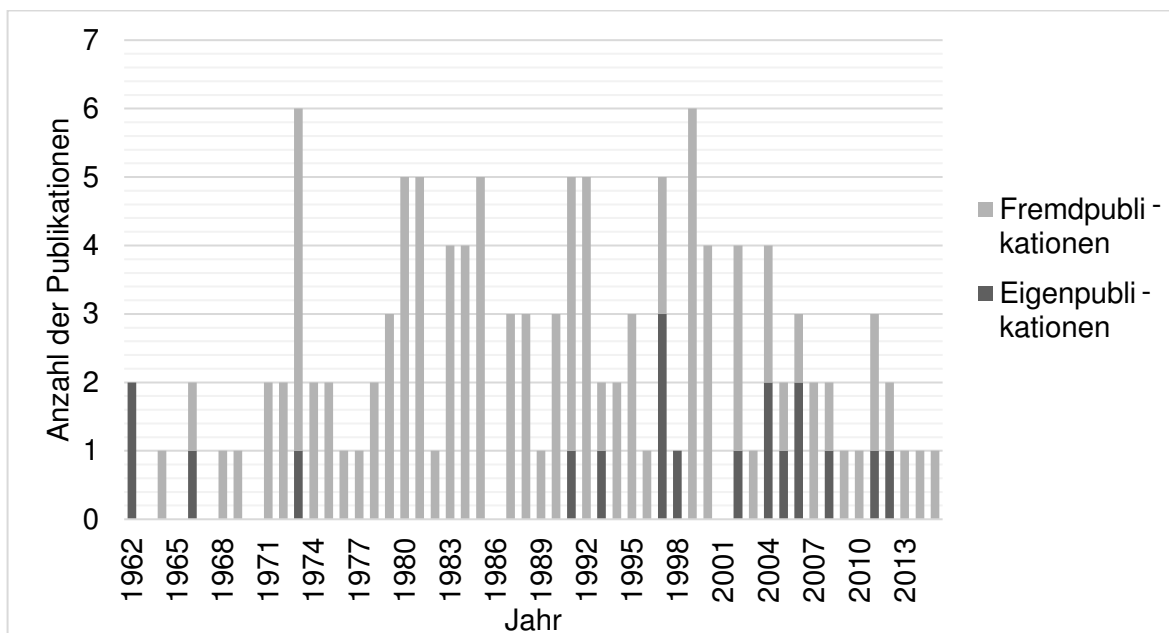


Abbildung 3: Anzahl der erscheinenden Eigen- und Fremdpublikationen im betrachteten Publikationszeitraum (Quelle: Eigene Darstellung)

In den Jahren von 1962 bis 2015 sind insgesamt 19 Eigenpublikationen in den untersuchten Zeitschriften veröffentlicht. Die meisten Eigenpublikationen sind in den Jahren 1962; 1997; 2004 und 2006 gefunden worden.

Im Zeitraum von 1974 bis 1990 sind keine Eigenpublikationen abgebildet. Die Verfasserin möchte darauf hinweisen, dass jedoch zwischen den Jahren 1972 und

1990 vier Neuauflagen (4. Auflage, 1972; 5. Auflage, 1975; 6. Auflage, 1978; 7. Auflage, 1981) des Tests d2 veröffentlicht wurden. Der Abbildung 3 ist weiterhin zu entnehmen, dass es in den 90er und 2000er Jahren mit 14 Artikeln mehr Eigenpublikationen gibt als in den vorherigen Jahren.

Wird die Summe der Artikel in den einzelnen Jahrzehnten betrachtet, steigt die Anzahl der Eigenpublikationen besonders in den Jahren von 1990 bis 1999 (sechs Artikel im Jahrzehnt) und von 2000 bis 2009 (sieben Artikel im Jahrzehnt) an. Das ist u.a. ein Indiz dafür, dass der Test nach wie vor angewendet wird und Verbesserungen vorgenommen sowie neue Forschungsansätze veröffentlicht werden. So gehen z.B. die Autoren Schmidt-Atzert et al. (2004), aufgrund der Kritik von Oehlschlägel und Moosbrugger (1991), dass Probanden systematisch an den Test d2 herangehen können, um sich Vorteile zu verschaffen bzw. ein besseres Ergebnis zu erzielen¹⁸⁸, der Frage von Simulation und Dissimulation beim Test d2 nach.¹⁸⁹ Auch der ursprüngliche Ansatz von Brickenkamp, dass beim Test d2 eine Unabhängigkeit von intellektuellen Faktoren besteht,¹⁹⁰ wurde von Westhoff und Kluck (1983) falsifiziert.¹⁹¹ Es wurde herausgefunden, dass der Test d2 nicht als Intelligenztest gilt, sondern ein Konstrukt beschreibt, das die Intelligenz kausal bedingt.¹⁹²

Im Vergleich zu den Eigenpublikationen sind die Fremdpublikationen des Tests d2 häufiger in den untersuchten Zeitschriften zu finden. Wird mit der Summe der 105 Fremdautoren veröffentlichten Artikel ein Durchschnitt der letzten 53 Jahre ermittelt, ist theoretisch in jedem halben Jahr ein Artikel mit dem Test d2 herausgekommen. Seit dem Jahr 1962 sind in nur acht Jahren (1962 – 1963; 1965; 1967; 1970; 1986; 1998; 2001) keine Fremdpublikationen erschienen (siehe Abbildung 3). Des Weiteren ist in der Abbildung 3 dargestellt, dass die meisten Fremdpublikationen in den Jahren von 1980 bis 2000 veröffentlicht wurden. Viele dieser Artikel äußern Kritik

¹⁸⁸ Vgl. Oehlschlägel; Moosbrugger (1991), S. 43 f.

¹⁸⁹ Vgl. Schmidt-Atzert; Bühner; Rischen; Warkentin (2004), S. 124 ff.

¹⁹⁰ Vgl. Brickenkamp (1978), S. 20.

¹⁹¹ Vgl. Westhoff; Kluck (1983), S. 312 ff.

¹⁹² Vgl. Schweizer (2005), S. 45 f.

bezüglich des Tests d2, indem sie einige Behauptungen von Brickenkamp falsifizieren¹⁹³ oder sogar die Kennwerte anzweifeln.¹⁹⁴ Andere Artikel beinhalten Verbesserungsvorschläge¹⁹⁵ oder weisen darauf hin, dass der Test d2 neu überarbeitet werden muss.¹⁹⁶ Der Test d2 wird auch als Kontrollinstrument für andere Tests eingesetzt.¹⁹⁷ Ab 2002 bis 2015 nimmt die Anzahl der Publikationen ab, jedoch bleiben sie stetig präsent.

4.2 Untersuchungsmethoden

In diesem Abschnitt werden die angewandten Untersuchungsmethoden bei der Reliabilität und Validität sowie die dazugehörigen Techniken näher beschrieben. Des Weiteren werden die Beurteilungsrichtlinien für Testkennwerte der Reliabilitäten und Validitäten aufgezeigt.

4.2.1 Reliabilitäts-Methoden

Bei den Fragestellungen zwei sowie vier mit den dazugehörigen Hypothesen zum Vergleich der Gütekriterien Reliabilität und Validität in Eigen- und Fremdpublikationen werden die nach den Anwendungsbereichen geordneten Publikationen genutzt. Jedoch sind hierbei die Artikel selektiert, die die Reliabilität und Validität des Tests untersuchen. Bei der Analyse der Reliabilitäten werden nur Artikel verwendet, die eine Retest-Reliabilitäts-Methode oder eine Split-Half-Reliabilität beim Test d2 aufweisen. Die Retest-Reliabilität ist am meisten vorhanden. Bei dieser Methode wird ein Testverfahren zweimal an den gleichen Versuchspersonen durchgeführt. Anschließend findet die Berechnung für die Korrelation der Testwerte aus beiden Durchgängen statt.¹⁹⁸ „Die Retest-Reliabilität eines Testverfahrens ist hoch, wenn

¹⁹³ Vgl. Wiese; Kroj (1972), S. 697 f.; Westhoff; Kluck (1983), S. 310 ff.

¹⁹⁴ Vgl. Heinrich (1973), 124; Oehlschlägel; Moosbrugger (1991), S. 43 ff.

¹⁹⁵ Westhoff; Kluck (1983), S. 310 ff.; Oehlschlaegel; Moosbrugger (1991), S. 43 ff.

¹⁹⁶ Vgl. Heinrich (1973), S. 121 ff.; Eser (1991), S. 137; Fay (1995), S. 293 ff.

¹⁹⁷ Vgl. Hasselhorn; Körkel (1984), S. 289 f.; Leikin; Paz-Baruch; Leikin (2014), S. 122 f.

¹⁹⁸ Vgl. Schermelleh-Engel; Werner (2012), S. 123.

zwei Messungen mit diesem Test zu verschiedenen Messzeitpunkten hoch miteinander korrelieren.“¹⁹⁹

Die Split-Half-Reliabilität oder auch Testhalbierungsreliabilität halbiert den Test nach der Durchführung äquivalent. Somit entstehen für jeden Versuchsteilnehmer zwei Testwerte. Anschließend wird die Korrelation der beiden Testhälften ermittelt.²⁰⁰ Es ist zu beachten, dass nur die Reliabilität eines halben Tests gemessen wird. Mit der Erhöhung der Itemzahl nimmt die Reliabilität des Tests zu.²⁰¹ Um die Reliabilität für den Gesamttest zu ermitteln, wird mithilfe der Spearman-Brown-Formel korrigiert.²⁰² Es gibt für die Split-Half-Reliabilität mehrere Techniken,²⁰³ jedoch wird sich nur auf zwei konzentriert, weil diese für die Untersuchung relevant sind. In den analysierenden Artikeln wird die Aufteilung in die erste und zweite Testhälfte und die Odd-even-Methode verwendet. Bei der Aufteilung in die erste und zweite Testhälfte wird der Test in zwei Hälften unterteilt. Dabei spielt es keine Rolle, ob die Items gerade oder ungerade sind. Die Odd-even-Methode dagegen unterteilt die Testhälften in ungerade und gerade Items.²⁰⁴

Korrelationskoeffizienten, die geringer als 0,80 sind, werden als schwache Zusammenhänge interpretiert, solche zwischen 0,80 und 0,90 als mittlere und über 0,90 als sehr starke.²⁰⁵ Die Reliabilität wird oft von der Stichprobengröße beeinflusst. Kleine Stichprobengrößen erlauben meist nur eine grobe Schätzung der Reliabilitätskorrelationen.²⁰⁶ Aufgrund dieser Zusammenhänge zwischen der Reliabilität und den Stichprobenumfängen, findet eine genauere Betrachtung in dieser Arbeit statt. Die Ergebnisse der Reliabilität setzen sich aus sieben Publikationen zusammen (siehe Tabellen 5; 6; A-4; A-5; A-6; A-7).

¹⁹⁹ Schermelleh-Engel; Werner (2012), S. 123.

²⁰⁰ Schmidt-Atzert; Amelang (2012), S. 48.

²⁰¹ Vgl. ebenda, S. 48.

²⁰² Vgl. Schermelleh-Engel; Werner (2012), S. 128.

²⁰³ Vgl. Schmidt-Atzert; Amelang (2012), S. 48.

²⁰⁴ Vgl. ebenda, S. 48.

²⁰⁵ Vgl. Fisseni (2004), S. 80.

²⁰⁶ Vgl. Schmidt-Atzert; Amelang (2012), S. 141.

4.2.2 Validitätskriterien

Bei der Validität gibt es ebenfalls verschiedene Merkmale, wie die Inhalts-, Kriteriums- und Konstruktvalidität. Um den Inhalt der Arbeit einzugrenzen, wird die Konstruktvalidität der Eigen- und Fremdpublikationen auf Unterschiede untersucht. Ein weiterer Grund für die Anwendung der Konstruktvalidität ist, dass es dazu mehr Vergleichsmöglichkeiten beim Test d2 in den Eigen- und Fremdpublikationen gibt als zur Kriteriumsvalidität. Bei Konzentrationstests ist es relativ schwierig ein nahe liegendes Kriterium zu finden, da die Leistungen bei den gestellten Aufgaben auch von anderen Faktoren abhängig sind.²⁰⁷

Für eine bessere Interpretation der Korrelationskoeffizienten gibt es eine ungefähre Richtlinie bei der Validität, wonach bei einem Wert, der geringer als 0,40 ist, bei Werten zwischen 0,40 und 0,60 ein mittlerer und über 0,60 ein hoher Zusammenhang vorliegt.²⁰⁸

Bei den Validitätskoeffizienten gilt es zu beachten, dass diese Werte allein mit einer Betrachtung auf das absolute Ergebnis, wenig über einen Test aussagen. Vielmehr ist zu beachten, inwieweit ein Test einen Beitrag zur Beantwortung einer Fragestellung leistet sowie in welcher Situation dieser angewendet wird.²⁰⁹

Die Stichprobengrößen spielen auch bei der Validität eine wichtige Rolle, denn oft sind diese zu klein und beeinflussen dadurch die Validitätsergebnisse.²¹⁰ Aufgrund dieses Zusammenhangs werden zusätzlich die Validitäten mit den Stichprobengrößen bei den Eigen- und Fremdpublikationen verglichen.

Für die Untersuchung der Validitäten von Eigen- und Fremdpublikationen sind insgesamt elf Artikel relevant (siehe Tabellen 7; 8; 9 und 10).

²⁰⁷ Vgl. Brickenkamp; Schmidt-Atzert; Liepmann (2010), S. 60.

²⁰⁸ Vgl. Fisseni (2004), S. 80.

²⁰⁹ Vgl. Bühner, S. 81.

²¹⁰ Vgl. Schmidt-Atzert; Amelang (2012), S. 160.

5 Ergebnisdarstellung

Die Fragestellungen mit deren abgeleitete Arbeitshypothesen beachtend, stellt dieses Kapitel die erhobenen Daten und die damit gewonnenen Ergebnisse in systematischer Reihenfolge dar und liefert die Grundlage für die abschließende Diskussion.

5.1 Analyse der Anwendungsbereiche bei der Nutzung des Test d2 – Aufmerksamkeits- und Belastungs-Test bei den Eigen- und Fremdpublikationen

In diesem Abschnitt wird analysiert, ob sich die Domänen der Verwendung des Tests d2 in der betrachteten Publikationszeit bei den Eigen- und Fremdpublikationen verändert haben. Dabei werden die Veröffentlichungen der Eigen- und Fremdautoren den einzelnen Anwendungsgebieten zugeordnet und aufsteigend nach dem Jahr gelistet (siehe Anhang Tabellen A-2 und A-3), um diese auszuwerten und miteinander zu vergleichen.

In den Abbildungen 4 und 5 sind die Anwendungsgebiete der Eigen- und Fremdpublikationen nach den Jahren aufgeführt. Es ist in diesen Abbildungen abzulesen, wieviel Artikel den unterschiedlichen Anwendungsbereichen im betrachteten Zeitraum entsprechen.

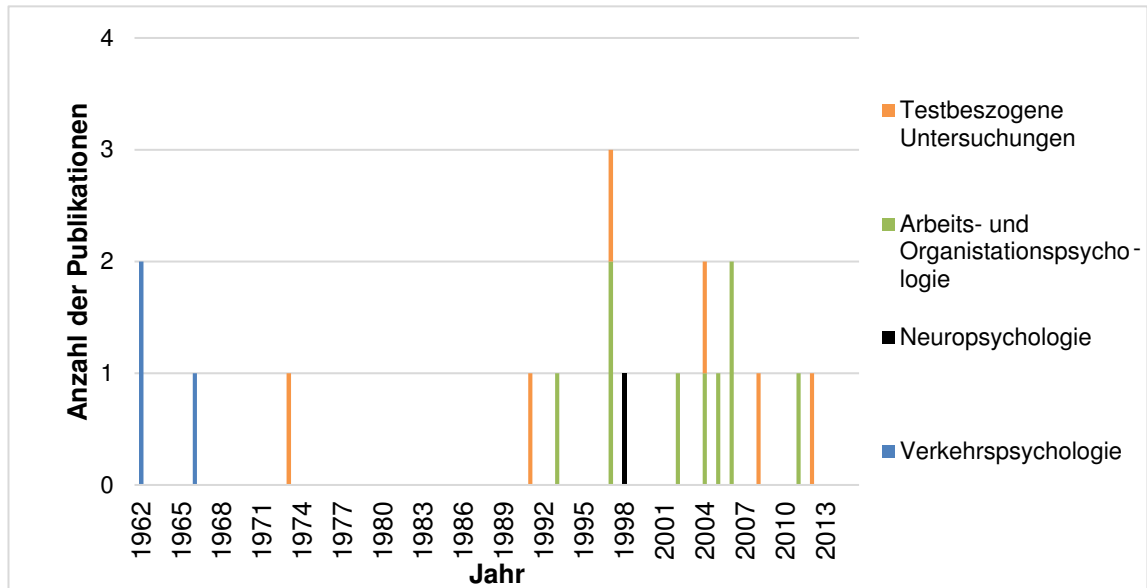


Abbildung 4: Diagramm für die Zuordnung der einzelnen Eigenpublikationen zu den Anwendungsbereichen (Quelle: Eigene Darstellung)

Zuerst werden die Anwendungsbereiche des Tests d2 der Eigenpublikationen betrachtet. Dabei fällt auf, dass sich die Literatur der Eigenautoren überwiegend der Kategorie „Arbeits- und Organisationspsychologie“ (Artikelanzahl: neun) widmet. Die Unterteilung „Testbezogene Untersuchungen“ (Artikelanzahl: sechs) ist am zweit häufigsten vertreten. Die Anwendungsgebiete von „Verkehrspsychologie“ (drei Artikel) und „Neuropsychologie“ (ein Artikel) nehmen einen geringen Teil ein. Werden die Zeiträume hinzugezogen fällt auf, dass die „Verkehrspsychologie“ bei den Eigenpublikationen nur in den Jahren 1962 und 1966 eine Rolle spielt, denn in den darauffolgenden Jahren ist keine Einordnung der „Verkehrspsychologie“ zu finden. Im Bereich der „Arbeits- und Organisationspsychologie“ sind vor allem seit Anfang der 90er Jahre vermehrt Eigenpublikationen mit dem Test d2 zu finden. Die erste Publikation in diesem Anwendungsbereich ist im Jahr 1993 erschienen, gefolgt von den Jahren 1997, 2002, 2004, 2005 und 2006. Die aktuellste Publikation in diesem Anwendungsgebiet erschien im Jahr 2011. Artikel, welche sich mit der Kategorie „Testbezogene Untersuchungen“ beschäftigten, treten seit dem Jahr 1973 sporadisch auf. Es ist festzustellen, dass in dieser Rubrik der zweite Artikel erst wieder im Jahr 1991 und der Letzte im Jahr 2012 erschien.

Die Eigenautoren veröffentlichten für das Anwendungsgebiet der „Neuropsychologie“ insgesamt einen Artikel, welcher sich auf das Jahr 1998 datieren lässt. Werden die letzten zehn Jahre (2005 bis 2015) betrachtet, fällt auf, dass vor allem

die „Arbeits- und Organisationspsychologie“ (vier Artikel) sowie die „Testbezogenen Untersuchungen“ (zwei Artikel) für die Eigenpublikationen interessant waren. Die unterschiedlichen Anwendungsbereiche im betrachteten Publikationszeitraum der Fremdveröffentlichungen sind in der Abbildung 5 dargestellt.

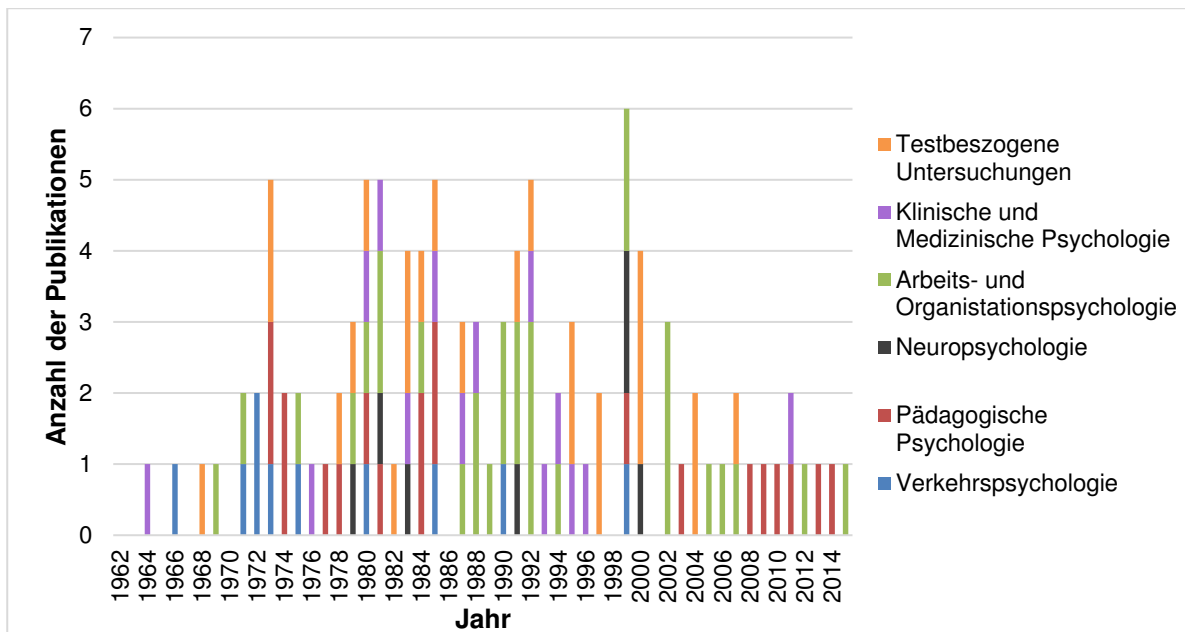


Abbildung 5: Diagramm für die Zuordnung der einzelnen Fremdpublikationen zu den Anwendungsbereichen (Quelle: Eigene Darstellung)

Bei der Zuordnung der Anwendungsbereiche für die Fremdpublikationen werden alle vorgesehenen Anwendungsbereiche abgebildet. Es ist auffällig, dass auch bei den Fremdpublikationen die Anzahl an Veröffentlichungen im Bereich „Arbeits- und Organisationspsychologie“ (Artikelanzahl: 30) dominiert, gefolgt von den „Testbezogenen Untersuchungen“ (Artikelanzahl: 24) und der „Pädagogischen Psychologie“ (Artikelanzahl: 20). Die Kategorien „Klinische und Medizinische Psychologie“ (14 Artikel), „Verkehrspsychologie“ (zehn Artikel) und „Neuropsychologie“ (sieben Artikel) sind im Vergleich schwächer vertreten.

Der Test d2 wird bei der „Arbeits- und Organisationspsychologie“ von den Fremdpublikationen seit 1969 verwendet. Verstärkt sind Artikel in diesem Anwendungsbereich in den Jahren von 1987 bis 1994 und von 1999 bis 2007 zu finden. Die jüngsten Veröffentlichungen sind in den Jahren 2012 und 2015 erschienen. Seit dem Jahr 1968 sind die testbezogenen Untersuchungen in Erscheinung getreten. Vermehrt werden Artikel von den Fremdautoren in den Jahren von 1978 bis

1987 in dieser Kategorie verfasst. Danach erschienen Veröffentlichungen zu diesem Themengebiet in den Jahren von 1991 bis 2007.

Im Gegensatz zu den Eigenpublikationen sind Fremdpublikationen im Anwendungsbereich „Pädagogische Psychologie“ zu finden. Diese treten besonders in den Jahren von 1977 bis 1985 und von 2008 bis 2014 auf.

In der „Klinischen und Medizinischen Psychologie“ wurde der Test d2 in frühen Jahren (1964) verstärkt angewendet. Mehrere Untersuchungen mit dem Test d2 zu diesem Anwendungsbereich sind von 1980 bis 1988 und von 1992 bis 1996 aufgetreten. Die aktuellste Fremdpublikation der Klinischen und Medizinischen Psychologie erschien im Jahr 2011.

Bei der Betrachtung des Anwendungsgebietes „Verkehrspsychologie“ ist im Laufe der Zeit ein Rückgang bei den Publikationen zu verzeichnen. Der Fokus der „Verkehrspsychologie“ liegt in Anfang bis Mitte der 70er Jahre. Danach erschienen Veröffentlichungen zur „Verkehrspsychologie“ vereinzelt bis zum Jahr 1999.

Seit Ende der 70er, Anfang der 80er Jahre erschienen Fremdpublikationen zur „Neuropsychologie“, die den Test d2 nutzen. Lückenhaft sind Anfang der 90er Jahre bis zum Jahr 2000 Veröffentlichungen im genannten Bereich zu finden.

Bei der Betrachtung der letzten zehn Jahre (2005 bis 2015) des Diagramms (siehe Abbildung 5) fällt auf, dass vor allem die Anwendungsgebiete „Pädagogische Psychologie“ (sechs Artikel) und die „Arbeits- und Organisationspsychologie“ (fünf Artikel) von Bedeutung für den Test d2 bei den Fremdautoren sind.

In den vorherigen Abbildungen 4 und 5 sind die Anwendungsbereiche des Tests d2 im betrachteten Publikationszeitraum für die Eigen- und Fremdpublikationen dargestellt. In der Tabelle 4 wird die Anzahl der Eigen- und Fremdpublikationen, die den Test d2 verwenden, in den einzelnen Anwendungsgebieten noch einmal gegenübergestellt.

Bei den Eigenpublikationen ist festzustellen, dass von den ursprünglich sechs Anwendungsbereichen die Artikel der Eigenautoren in vier Anwendungsbereiche („Verkehrspsychologie“, „Neuropsychologie“, „Arbeits- und Organisationspsychologie“ sowie „Testbezogene Untersuchungen“) eingeteilt werden. Zu den Anwendungsbereichen „Klinische und Medizinische Psychologie“ sowie „Pädagogische Psychologie“ sind, anders als bei den Fremdpublikationen, keine Zuordnungen möglich. So werden in den Fremdpublikationen alle sechs Anwendungsgebiete des

Tests d2 aufgegriffen. Die Anzahl der Eigenveröffentlichungen sind in jeder Kategorie bedeutend weniger als die der Fremdpublikationen. Die Gesamtanzahl der Eigenpublikationen beschränkt sich auf 19 im Gegensatz zu den 105 Fremdpublikationen, welche in den letzten 53 Jahren erschienen sind (siehe Tabelle 4).

Tabelle 4: Anzahl der Eigen- und Fremdpublikationen in den unterschiedlichen Anwendungsbereichen (Quelle: Eigene Darstellung)

Anwendungsgebiete	Eigenpublikationen	Fremdpublikationen
Verkehrspsychologie	3	10
Pädagogische Psychologie	0	20
Neuropsychologie	1	7
Arbeits- und Organisationspsychologie	9	30
Klinische und Medizinische Psychologie	0	14
Testbezogene Untersuchungen	6	24
Gesamtanzahl der Publikationen	19	105

Die meisten Publikationen erscheinen in der Kategorie „Arbeits- und Organisationspsychologie“ (neun Eigenpublikationen und 30 Fremdpublikationen). Im Anwendungsbereich „Neuropsychologie“ wird der Test d2 am wenigsten eingesetzt. Werden die Eigen- und Fremdpublikationen in der genannten Kategorie betrachtet, sind eine Eigenpublikation und sieben Fremdpublikationen zu erfassen. Allein die Fremdpublikationen veröffentlichen im Bereich der „Pädagogische Psychologie“ 20 und in dem der „Klinischen und Medizinischen Psychologie“ 14 Artikel. Die „Verkehrspsychologie“ ist bei beiden Publikationsarten vorhanden, jedoch wurden bei den Fremdauforen (Artikelanzahl: zehn) mehr Untersuchungen in diesem Themengebiet durchgeführt als bei den Eigenautoren (Artikelanzahl: drei).

In der Gesamtanzahl wurden 30 Artikel in der Kategorie „Testbezogene Untersuchungen“ verfasst. Davon sind sechs Veröffentlichungen den Eigenautoren und 24 Publikationen den Fremdauforen zuzuordnen.

5.2 Analyse der Reliabilität des Test d2 – Aufmerksamkeits- und Belastungs-Test bei Eigen- und Fremdpublikationen

Für die Betrachtung der Reliabilitäten ohne die spezifischen Verwendungen und mit den spezifischen Verwendungen des Test d2 – Aufmerksamkeits- und Belastungs-Test werden ebenfalls die Eigen- und Fremdpublikationen miteinander verglichen. Bei der Reliabilität werden die Ergebnisse von der Retest- Reliabilität, der Aufteilung in die erste und zweite Testhälfte und der Odd-even-Technik einzeln betrachtet, da diese nicht miteinander vergleichbar sind.

Zuerst werden die Reliabilitätsergebnisse der Eigenpublikationen und anschließend die der Fremdpublikationen untersucht. Die Tabelle 5 zeigt die Retest-Reliabilitätskoeffizienten der Eigenpublikationen. Die aufgelisteten Kennwerte des Tests d2 beziehen sich dabei auf die ältere Version des Test d2 – Aufmerksamkeits- und Belastungstest.

Tabelle 5: Retest-Reliabilitäten nach der Test-Retest-Methode bei den Eigenpublikationen (Quelle: Eigene Darstellung)

Jahr	Anwendungsgebiet	Autor(en) ¹	Stichprobe	Zeitintervall	Stabilitätskoeffizienten (Retest-Reliabilität) ²			
					GZ	GZ-F	F%	SB
1966	Verkehrspsychologie	Brickenkamp, R; Rump, G.	Berufskraftfahrer (N= 59)	12 Monate	0,89*	0,92*	0,61*	0,23*
1966	Verkehrspsychologie	Brickenkamp, R; Rump, G.	Kraftfahrer (N= 53)	23 Monate	0,71*	0,88*	0,51*	0,28*
1991	Testbezogene Untersuchungen	Brickenkamp, R.	Kinder (N= 38)	3 Monate		0,88*		
1991	Testbezogene Untersuchungen	Brickenkamp, R.	Jugendliche (N= 37)	3 Monate		0,72*		

1991	Testbezogene Untersuchungen	Brickenkamp, R.	Erwachsene (N= 31)	3 Monate		0,90*		
------	-----------------------------	-----------------	--------------------	----------	--	-------	--	--

Anmerkung: * $p < 0,05$ (einseitig); 1 Die Untersuchungen von Brickenkamp; Rump (1966) sind aus der gleichen Publikation entnommen; Die Untersuchungen von Brickenkamp (1991) sind aus der gleichen Publikation entnommen; 2 Produktmoment-Korrelationen, Rangkorrelationen für F% und SB außer bei Brickenkamp; Rump (1966).

Bei genauer Betrachtung der Kennwerte liegen bei der Anzahl der bearbeiteten Zeichen minus Fehler (GZ-F) die meisten Ergebnisse vor. Die anderen Kennwerte werden nur bei den Untersuchungen von Brickenkamp und Rump (1966) mitgeteilt. Die Retest-Reliabilitäten für den Kennwert GZ-F liegen in einem Bereich zwischen $r = 0,72$ bis $0,92$. Die Werte von Brickenkamp und Rump (1966) bei den Buskraftfahrern sowie Kraftfahrern und den Erwachsenen (Brickenkamp, 1991) liegen im Bereich von $r = 0,88$ bis $0,90$. Hierbei liegt die Zeitspanne der Untersuchungen zwischen drei Monaten und fast zwei Jahren. Der Reliabilitätskoeffizient des Kennwertes der Anzahl der bearbeiteten Zeichen (GZ) ist bei den Untersuchungen geringer als bei GZ-F. Bei 23 Monaten liegt die Retest-Reliabilität für den GZ-Wert bei $r = 0,71$ und bei zwölf Monaten wurde ein Wert von $r = 0,89$ ermittelt. Für den Fehlerprozentwert (F%) zeigen sich Werte von $r = 0,51$ (Zeitintervall: 23 Monate) und $r = 0,61$ (Zeitintervall: zwölf Monate). Wie auch bei dem GZ-Wert ist mit einem größeren Zeitintervall ein Abfall im Koeffizienten zu erkennen. Die Schwankungsbreite (SB) zeigt die geringsten Reliabilitätskoeffizienten von $r = 0,23$ bis $r = 0,28$ auf.

Die Anwendungsbereiche der einzelnen Untersuchungen sind der „Verkehrspsychologie“ und den „Testbezogenen Untersuchungen“ zuzuordnen. Somit sind bei den Eigenpublikationen die Retest-Reliabilitäten in zwei verschiedene Anwendungsgebiete ermittelt worden.

Die Tabelle A-4 im Anhang zeigt ein Ergebnis der Reliabilitätsschätzungen nach Aufteilung in die erste und zweite Testhälfte bei den Eigenpublikationen des Tests d2. Bei der Betrachtung ist nur eine Untersuchung von den Autoren Krumm et al. (2008) in den Artikeln zu finden, welche sich auf den Wert GZ bezieht. Der Reliabilitätskoeffizient beträgt $r = 0,97$ und liegt somit über die Reliabilitätsschätzungen von

Brickenkamp et al. (2010), denn dieser Wert liegt bei $r = 0,89$.²¹¹ Der Test ist dem Anwendungsbereich „Testbezogene Untersuchung“ zugeordnet.

Bei der Odd-even-Methode zeigt sich eine ähnliche Situation wie bei der Reliabilitätsschätzung nach Aufteilung in die erste und zweite Testhälfte (siehe Anhang Tabelle A-5). Es wird ebenfalls im Bereich der „Testbezogenen Untersuchungen“ ein Reliabilitätskoeffizient für den Wert GZ mit einem Betrag von $r = 0,93$ ermittelt.

Die Retest-Reliabilitäten der Fremdpublikationen zeigt die Tabelle 6 auf. Es werden insgesamt drei Untersuchungen in verschiedenen Zeitintervallen dargestellt. Der GZ-Wert bei der Untersuchung von Dingel (1971) weist nach 24 Monaten einen Reliabilitätskoeffizienten von $r = 0,56$ auf. In den Untersuchungen von Eser (1987) liegen die Werte dagegen bei $r = 0,75$ und $0,24$. Dabei ist zu beachten, dass sich die Zeitintervalle unter einem Jahr bewegen. Demnach sind die Ergebnisse von Dingel (1971) bei einem höheren Zeitintervall zuverlässiger als die Werte von Eser (1987).

Tabelle 6: Retest-Reliabilitäten nach der Test-Retest-Methode bei den Fremdpublikationen (Quelle: Eigene Darstellung)

Jahr	Anwendungsbereich	Autor(en) ¹	Stichprobe	Zeitintervall	Stabilitätskoeffizienten (Retest-Reliabilität) ²			
					GZ	GZ-F	F%(F)	SB
1971	Verkehrspsychologie	Dingel, W.	Teilnehmer Kraftfahruntersuchung N = 70	24 Monate	0,56*	0,71*	(0,46*)	
1987	Klinische und Medizinische Psychologie	Eser, K.- H.	Kinder und Jugendliche mit Verhaltensstörungen (N=22)	4 Monate	0,75*	0,73*	0,61*	0,52*
1987	Klinische und Medizinische Psychologie	Eser, K.- H.	Kinder und Jugendliche mit Verhaltensstörungen (N=18)	11 Monate	0,24	0,50*	0,37	0,49*

Anmerkung: * $p < 0,05$ (einseitig). 1 Die Untersuchungen von Eser (1987) sind aus der gleichen Publikation entnommen; 2 Produktmoment-Korrelationen, Rangkorrelationen für F% und SB außer bei Dingel (1971).

²¹¹ Vgl. Brickenkamp; Schmidt-Atzert; Liepmann (2010), S. 45.

Für die GZ-F-Werte liegen die Retest-Reliabilitäten im Bereich von $r = 0,50$ bis $0,73$. Es fällt auf, dass der geringste Wert bei einem Zeitintervall von elf Monaten auftritt. Somit ist dieser Wert kleiner als bei einem Intervall von 24 Monaten ($r = 0,71$) sowie von vier Monaten ($r = 0,73$). Die Reliabilitätskoeffizienten für den F%-Wert bei Eser (1987) unterscheiden sich deutlich. Der Wert $r = 0,61$ mit einem Zeitintervall von vier Monaten ist höher als der Wert $r = 0,37$ mit elf Monaten. Bei Dingel (1971) wird nicht die Retest-Reliabilität für den F%-Wert angegeben, sondern für den Kennwert F. Dieser liegt über dem Ergebnis von Eser bei einem Intervall von elf Monaten. Die Werte für die Schwankungsbreite (SB) bei Eser (1987) liegen zwischen $r = 0,52$ und $r = 0,49$.

Die untersuchenden zwei Veröffentlichungen sind der „Verkehrspsychologie“ sowie der „Klinischen und Medizinischen Psychologie“ zuzuordnen.

Bei der Reliabilitätsschätzung nach Aufteilung in die erste und zweite Testhälfte wurde wie bei den Eigenpublikationen eine Veröffentlichung bei den Fremdautoren von Rauchfleisch im Jahr 1983 gefunden. Es sind hohe Korrelationen der Werte GZ ($r = 0,97$) und F ($r = 0,95$) ermittelt worden (siehe Anhang Tabelle A-6). Dieser Artikel entspricht dem Anwendungsbereich „Klinische und Medizinische Psychologie“. Eser (1987) verwendet ebenfalls bei seiner Untersuchung die Odd-even-Methode. Es ist festzustellen, dass die Werte ($r^{GZ} = 0,93$, $r^{GZ-F} = 0,93$, $r^{F\%} = 0,96$ und $r^{SB} = 0,95$) höher sind als bei der Retest-Reliabilität (siehe Anhang Tabelle A-7). Der Anwendungsbereich bleibt der gleiche, wie bereits erwähnt wurde.

Im Vergleich der Reliabilitäten ohne Beachtung der spezifischen Verwendungen des Tests d2 zwischen Eigen- und Fremdpublikationen ist auffallend, dass die Retest-Reliabilitäten der Eigenpublikationen, vor allem bei den Kennwerten GZ (von $r = 0,72$ bis $r = 0,92$), GZ-F (von $r = 0,71$ bis $r = 0,89$) und F% (von $r = 0,51$ bis $r = 0,61$), deutlich höher sind als die Fremdpublikationen mit Koeffizienten bei dem GZ-Wert (von $r = 0,24$ bis $r = 0,75$), bei dem GZ-F-Kennwert (von $r = 0,50$ und $r = 0,73$) sowie bei dem F%-Kennwert (von $r = 0,37$ bis $r = 0,61$). Die Korrelationen der Retest-Reliabilität sind bei dem SB-Kennwert der Fremdpublikationen höher (zwischen $r = 0,49$ und $r = 0,52$) als bei den Eigenpublikationen (von $r = 0,23$ bis $0,28$). Die Eigenpublikationen liegen zum größten Teil in einem mittleren bis hohen Korrelationsbereich. Im Gegensatz dazu weisen die Ergebnisse der Fremdpublikationen einen geringen

Zusammenhang auf. Ein weiterer Unterschied bei einem Vergleich ohne die Berücksichtigung auf die spezifische Verwendung ist, dass die Stichproben heterogen sind und dies einen enormen Einfluss auf die Ergebnisse hat. Hierbei werden bspw. Erwachsene mit Kindern und Jugendliche mit Verhaltensstörungen verglichen.

Bei der Untersuchung von den spezifischen Verwendungen in Eigen- und Fremdpublikationen stellt sich heraus, dass bei der Retest-Reliabilität nur ein Anwendungsgebiet, die „Verkehrspsychologie“, verglichen wird. Das sind die Untersuchungen von Brickenkamp, Rump (1966) und Dingel (1971). Brickenkamps und Rumps Publikation enthält zwei Stichproben mit unterschiedlichen Zeitintervallen (zwölf und 24 Monate). Mit der Analyse der zwei aufgezeigten Untersuchungen (siehe Tabelle 5) spricht ein geringeres Zeitintervall für höhere Retest-Reliabilitäten, wie es auch die Autoren Schmidt-Atzert und Amelang (2012) berichten.²¹² Denn die Untersuchung mit einem Zeitintervall von zwölf Monaten bei Brickenkamp und Rump (1966) zeigt eindeutig höhere Korrelationswerte auf als die anderen Untersuchungen mit 23 Monaten von Brickenkamp (1991) bzw. mit 24 Monaten von Dingel (1971) (siehe Tabellen 5 und 6). Für den Vergleich wird die Untersuchung von Brickenkamp und Rump (1966) mit dem Zeitintervall von 24 Monaten aufgegriffen, da die Veröffentlichung von Dingel (1971) (siehe Tabelle 6) ein ähnliches Zeitintervall von 23 Monaten aufzeigt. Die Koeffizienten, welche im Bereich der „Verkehrspsychologie“ ermittelt wurden, von GZ und GZ-F bei der Eigenpublikation sind deutlich höher als bei der Fremdpublikation. Der F%-Wert korreliert jedoch bei der Fremdpublikation höher.

Dieser Vergleich hat gezeigt, dass auch bei einer einzelnen Betrachtung der spezifischen Verwendungen die Korrelationen der Eigenpublikationen höher ausfallen. Für die Reliabilitätsschätzung nach Aufteilung in die erste und zweite Testhälfte weisen beide Publikationsformen einen hohen Zusammenhang auf. Der Koeffizient für den GZ-Wert ist trotz der unterschiedlichen Anwendungsbereiche („Testbezogene Untersuchungen“ sowie „Klinische und Medizinische Psychologie“) bei den Eigen- und Fremdautoren gleich ($r = 0,97$).

Bei der Odd-even-Methode sind die Werte des GZ-Wertes ($r = 0,93$) der Eigen- und

²¹² Vgl. Schmidt-Atzert; Amelang (2012), S. 137.

Fremdautoren gleich bei unterschiedlichen spezifischen Anwendungen („Testbezogene Untersuchungen“ sowie „Klinische und Medizinische Psychologie“). Der Vergleich bei der Reliabilitätsschätzung nach Aufteilung in die erste und zweite Testhälfte und Odd-even-Methode zwischen den spezifischen Anwendungen findet, aufgrund der unterschiedlichen Anwendungsbereiche nicht statt.

5.3 Analyse der Validität des Test d2 – Aufmerksamkeits- und Belastungs-Test bei Eigen- und Fremdpublikationen

Dieser Abschnitt enthält die Untersuchung der Validität beim Test d2 ohne die spezifischen Verwendungen sowie mit den spezifischen Verwendungen bei den Eigen- und Fremdpublikationen. Die Validitätsuntersuchungen werden eingeschränkt, indem sich die Analysen auf die Konstruktvalidität konzentrieren. In den folgenden Tabellen liegt das Hauptaugenmerk auf der konvergenten und diskriminanten Validität, also der Korrelationen der Test d2-Kennwerte mit anderen Konzentrations- und Intelligenztests.

Beim Ausschuchen der Artikel mit untersuchter Konstruktvalidität in Verbindung mit dem Test d2 wird darauf geachtet, dass der Test d2, aufgrund der Verfälschung der Endergebnisse, nicht modifiziert verwendet wurde. Die Korrelation der Test d2-Kennwerte mit einem anderen Konzentrationstest bei Eigenpublikation ist in Tabelle 7 dargestellt und gehört zur konvergenten Validität.

Tabelle 7: Korrelation der d2-Kennwerte mit anderen Konzentrationstests bei Eigenpublikationen (Quelle: Eigene Darstellung)

Jahr	Anwendungsgebiet	Autor(en)	Stichprobe	Konzentrationstest und Kennwert	d2-Kennwert		
					GZ	GZ-F (KL)	F% (F)
1997	Arbeits- und Organisationspsychologie	Schmidt-Atzert, L.; Ising, M.	Bewerber (N= 524)	Rev.T. Gesamtzahl	0,30*	0,22*	
				Fehlerwert	-0,11*	0,00	

Anmerkung: Rev.T. = Revisions-Test (Marschner, 1972); *p < 0,05 (einseitig).

Zur konvergenten Validität wurde nur ein Artikel gefunden, der keine Ergebnisse von modifizierten Tests d2 aufweist. Die Mengenleistung (GZ) beim Test d2 korreliert mit der Mengenleistung (GM) des Revision-Tests (1997) gering ($r = 0,30$). Die Korrelation beim GZ-F-Kennwert liegt ebenfalls im niedrigen Bereich ($r = 0,22$) und fällt noch geringer aus als beim GZ-Kennwert. Die Korrelationen des GZ- und GZ-F-Wertes mit dem Fehlerwert des Revisions-Tests ($r^{GZ} = -0,11$; $r^{GZ-F} = 0,00$) ist gering. Bei F% sind keine Angaben vorhanden. Der verwendete Artikel ist der „Arbeits- und Organisationspsychologie“ zugeordnet.

Bei den Korrelationen der d2-Kennwerte mit Intelligenztests der Eigenpublikationen ist ebenfalls eine Untersuchung zu ermitteln (siehe Tabelle 8), die die diskriminante Validität betrachtet. Die Korrelationen zwischen dem Test d2 und dem Intelligenz-Struktur-Test 70 (IST 70) weisen geringe Werte auf. Der GZ-Wert korreliert geringer mit dem IST 70 Gesamtwert als der GZ-F-Wert. Es sind ebenfalls keine Angaben für den Fehlerprozentwert genannt. Diese Untersuchung ist dem Anwendungsgebiet „Arbeits- und Organisationspsychologie“ zugeordnet.

Tabelle 8: Korrelation der d2-Kennwerte mit Intelligenztest bei Eigenpublikationen
(Quelle: Eigene Darstellung)

Jahr	Anwendungsgebiet	Autor(en)	Stichprobe	Konzentrations-test und Kennwert	d2-Kennwert		
					GZ	GZ-F (KL)	F% (F)
1997	Arbeits- und Organisationspsychologie	Schmidt-Atzert, L.; Ising, M.	Bewerber (N= 1558)	IST 70 Gesamtwert	0,14*	0,17*	

Anmerkung: IST 70= Intelligenz-Struktur-Test 70 (Amthauer, 1970), * $p < 0,05$ (einseitig).

Damit auch in diesem Abschnitt auf den Vergleich der Eigen- und Fremdpublikationen zwischen den verschiedenen Arten der Validitäten eingegangen wird, sind in den folgenden Tabellen 9 und 10 die Validitäten der Fremdveröffentlichungen dargestellt.

Bei den Kennwerten des Tests d2 mit anderen Konzentrationstests, wie Revisions-Test, Konzentrations-Leistungs-Test, Komplexer Konzentrationstest und Differentieller Konzentrationstest für Kinder, der Fremdpublikationen (siehe Tabelle 9) korrelieren die Tempowerte von GZ mit anderen Tempowerten im geringen bis mittleren

Bereich. Das höchste Ergebnis ($r = 0,72$) zeigt die Korrelation zwischen dem GZ-Wert und der Mengenleistung des Revisions-Tests (Rev. T.) auf. Die GZ-F-Werte korrelieren ebenfalls geringfügig bis mittelmäßig mit den Werten der Leistungsmengen ($r = 0,34$ und $r = 0,60$). Der Fehlerprozentwert (F%) zeigt mit den Fehlerwerten geringe bis mittlere Zusammenhänge der vorliegenden Tests auf. Der Komplexe Konzentrations-Test (KKT) korreliert mit dem Kennwert GZ-F ($r = 0,60$) und der Differentielle Konzentrationstest für Kinder (DKT-K) erzielt den höchsten Zusammenhang zwischen dem F%- Wert und der Leistungsgüte ($r = -0,45$) (siehe Tabelle 9).

Tabelle 9: Korrelation der d2-Kennwerte mit anderen Konzentrationstests bei Fremdpublikationen (Quelle: Eigene Darstellung)

Jahr	Anwendungsgebiet	Autor(en)	Stichprobe	Konzentrationstest und Kennwert	d2-Kennwert		
					GZ	GZ-F (KL)	F% (F)
1971	Arbeits- und Organisationspsychologie	Marschner, G.	Erwachsene (N= 60)	Rev. T. Mengenleistung	0,72*		
1973	Testbezogene Untersuchung	Sommer, G.	Studenten (N= 50)	KLT Leistungsmenge		0,34*	
2003	Pädagogische Psychologie	Westhoff, K.; Graubner, J.	Schüler der 8. Klasse (N= 70)	KKT-Gesamt GZ-F		0,60*	0,07
				KKT-Gesamt F%		0,04	0,30*
2011	Klinische und Medizinische Psychologie	Funsch, K.; Arias Martin, B.; Halder-Sinn, P.	Vor- und Grundschüler (N= 68)	DKT-K Leistungstempo	-0,47*		0,14
				Leistungsgüte	0,05		-0,45*

Anmerkung: Rev. T.= Revisions-Test (Marschner, 1971); KLT= Konzentrations-Leistungs-Test (Dücker et al., 1965); KKT= Komplexer Konzentrationstest (Westhoff et al., 2003); DKT-K= Differentieller Konzentrationstest für Kinder (Funsch et al. 2011); * $p < 0,05$ (einseitig).

Werden die vier Untersuchungen in den Anwendungsbereichen der Tabelle 9 betrachtet, fallen die verschiedenen Bereiche, „Arbeits- und Organisationspsychologie“, „Testbezogene Untersuchungen“, „Pädagogische Psychologie“ sowie „Klinische und Medizinische Psychologie“, auf.

Bei den Korrelationen zwischen den d2-Kennwerten und Intelligenztests, wie Hamburg-WECHSLER-Intelligenztest für Erwachsene (HAWIE), Standard Progressive Matrices Test (SPM), Mannheimer Intelligenztest (MIT), Leistungsprüfsystem (LPS), der Fremdpublikationen (siehe Tabelle 10) sind im Vergleich zur vorherigen Tabelle 9 mehr Untersuchungen aufzuzeigen. Die GZ-Werte korrelieren gering bis mittelmäßig mit den anderen Tests. Ausnahmen bilden der SPM von Raven des Autors Merten (2000), welcher hoch mit den Kennwerten GZ ($r= 0,52$) und GZ-F ($r= 0,58$) des Tests d2 korreliert und der Untertest Zahlen-Symbol-Test (ZS) des HAWIE von dem Autor Rauchfleisch (1983) mit Korrelationen für den GZ-Wert von $r= 0,60$ und für den GZ-F-Wert von $r= 0,67$. Die meisten Zusammenhänge von GZ-F (KL) lassen sich im geringen Bereich mit Werten zwischen $r= 0,18$ bis $r= 0,58$ einordnen. Für F% (F) sind Werte von $r= 0,05$ bis $r= -0,48$ im geringen bis mittleren Korrelationsbereich zu finden. Die Ergebnisse des SB-Wertes sind gering korreliert und werden aufgrund der Eliminierung in der aktuellsten Ausgabe vom Test d2-R vernachlässigt.

Tabelle 10: Korrelation der d2-Kennwerte mit Intelligenz bei Fremdpublikationen (Quelle: Eigene Darstellung)

Jahr	Anwendungsgebiet	Autor(en)	Stichprobe	Intelligenztest	GZ	GZ-F (KL)	F% (F)	SB
1972	Arbeits- und Organisationspsychologie	Wiese, W.; Kroj, G..	Erwachsene (N=102)	HAWIE (Verbalteil)	0,20	0,26	-0,35	0,06
				HAWIE (Handlungsteil)	0,32	0,35	-0,21	0,05
1983	Neuropsychologie	Kühnlein, I.; Hahlweg, K.; Müller,	Patienten (N=194)	Raven SPM		0,341*	0,350*	

		G.; Seve- nitz, B.						
1983	Klini- sche und Me- dizini- sche Psycho- logie	Rauch- fleisch, U.	Patienten mit hirn- diffusem Psycho- syndrom und neu- rotischen Störun- gen (N= 90)	HAWIE				
				AW	0,37*	0,41*	(0,26*)	0,07
				AV	0,19*	0,18*	(0,05)	-0,01
				ZN	0,25*	0,38*	(0,46*)	0,26*
				RD	0,38*	0,45*	(0,35*)	0,14
				GF	0,21*	0,28*	(0,06)	0,09
				ZS	0,60*	0,67*	(0,19*)	-0,04
				BO	0,14	0,26*	(0,33*)	0,33*
				BE	0,35*	0,46*	(0,39*)	0,18*
				MT	0,42*	0,49*	(0,25*)	0,00
				FL	0,23*	0,27*	(0,05)	0,02
1994	Arbeits- und Or- ganisa- tions- psycho- logie	Schmidt, J U.	Bewerber (N= 1030)	MIT	0,29*		(-0,26*)	
2000	Neu- ropsy- cholo- gie	Merten, T.	Neurolo- gische Patienten (N=74)	Raven SPM Richtige	0,52	0,58 (0,60)	-0,48	
2015	Arbeits- und Or- ganisa- tions- psycho- logie	Becker, N.; Pre- ckel, F.; Karbach, J.; Raffel, N.; Spi- nath, F.M.	Studen- ten (N= 318)	Matri- zenkon- strukti- onsauf- gabe; LPS		(0,29)* ¹		

Anmerkung: HAWIE= Hamburg-WECHSLER-Intelligenztest für Erwachsene (Wechsler, 1965); SPM= Standard Progressive Matrices Test (Raven, 1938); MIT= Mannheimer Intelligenztest (Conrad et al., 1971); AW= Allgemeines Wissen; AV= Allgemeines Verständnis; ZN= Zahlen nachsprechen;

*RD= Rechnerisches Denken; GF= Gemeinsamkeiten finden; ZS= Zahlen-Symbol-Test; BO= Bilder ordnen; BE= Bilder ergänzen; MT= Mosaik-Test; FL= Figuren legen; LPS= Leistungsprüfsystem (Horn, 1983); * $p < 0,05$ (einseitig); * $^1p < 0,05$ (zweiseitig).*

Die „Arbeits- und Organisationspsychologie“, die „Neuropsychologie“ sowie die „Klinische und Medizinische Psychologie“ sind die Anwendungsgebiete der erschlossenen Artikel.

Im Mittelpunkt des Vergleichs von Eigen- und Fremdpublikationen der Konstruktvalidität sind die Korrelationen zwischen den Test d2-Kennwerten und den Konzentrationstests sowie die Korrelationen zwischen den Test d2-Kennwerten und den Intelligenztests.

Bei näherer Analyse der einzelnen Tabellen in den Eigen- und Fremdpublikationen fällt auf, dass die Eigenpublikation der Korrelationen von d2-Kennwerten mit anderen Konzentrationstests weniger Artikel enthalten als die Fremdpublikationen (siehe Tabellen 7 und 9). Für den Vergleich der Eigen- und Fremdpublikationen der konvergenten Validität sind die gleichen Tests (Test d2 und Revisions-Test) herangezogen wurden. Die Korrelationen der Fremdveröffentlichungen sind größer als die ausgewiesenen Ergebnisse der Eigenpublikationen. Besonders auffällig sind die Korrelationswerte von $r^{GZ} = 0,72$ und $r^{GZ-F} = 0,60$ bei den Fremdpublikationen, im Gegensatz zu den höchsten Zusammenhängen der Eigenpublikationen, welche bei dem GZ-Wert von $r = 0,30$ und bei dem GZ-F-Wert bei $r = 0,22$ liegen. Aufgrund der fehlenden Werte bei dem Artikel der Eigenautoren sind Fehlerprozentwerte nicht miteinander gegenüberzustellen.

Werden die Ergebnisse der konvergenten Validität mit Betrachtung der spezifischen Verwendung von Eigen- und Fremdpublikationen verglichen, sind Artikel mit der Zuordnung der „Arbeits- und Organisationspsychologie“ zu verwenden, da bei diesem Anwendungsgebiet beide Publikationsformen Ergebnisse aufweisen. Die Korrelation des GZ-Kennwertes liegt bei den Fremdautoren bei $r = 0,72$. Wird dieses Ergebnis mit dem der Eigenpublikation ($r^{GZ} = 0,30$) verglichen, liegen die Werte hier unter dem Ergebnis der Fremdpublikation. Im genannten Anwendungsgebiet werden keine weiteren Vergleiche angestellt, denn die übrigen Korrelationen der Kennwerte der Fremdpublikation fehlen.

Der Vergleich zwischen den Eigen- und Fremdpublikationen der diskriminanten Validität verdeutlicht, dass die meisten Untersuchungen in beiden Publikationsbereichen geringe Korrelationen aufzeigen. Hierbei waren keine gleichen Tests bei den Untersuchungen der Eigen- und Fremdautoren gegenüberzustellen. Die Fremdpublikationen weisen neben den niedrigen Zusammenhängen auch einige mittlere Korrelationsbereiche auf. Die Ergebnisse der Eigenpublikationen sind deutlich geringer als bei den Fremdpublikationen.

Die meisten Untersuchungen fanden wie bei den Analysen zur konvergenten Validität in der „Arbeits- und Organisationspsychologie“ statt. Dabei ist in diesem Anwendungsbereich ersichtlich, dass die Werte der Fremdpublikationen sich ähneln und bei den GZ-Werten im Korrelationsbereich von $r = 0,20$ bis $0,32$ liegen. Der GZ-F-Kennwert zeigt ähnliche Korrelationen auf (von $r = 0,26$ bis $r = 0,35$). Bei der Gegenüberstellung liegt der GZ-Wert der Eigenpublikation bei $r = 0,14$ und beim GZ-F-Wert bei $r = 0,17$. Die Korrelationen sind geringer als bei den Fremdpublikationen, dennoch nähert sich der GZ-Wert den Ergebnissen der Fremdpublikationen. Der Kennwert GZ-F ist etwas weiter von den Werten der Fremdpublikationen entfernt.

Bei den Angaben ist jedoch der Stichprobenumfang zu beachten, der hierbei außer Acht gelassen wird, da dies ebenfalls Auswirkung auf die Validität nimmt und erst im Abschnitt 5.6 analysiert und im Abschnitt 6.4 interpretiert wird.

5.4 Analyse des durchschnittlichen Stichprobenumfangs bei Eigen- und Fremdpublikationen des Tests d2 – Aufmerksamkeits- und Belastungs-Test

Der Stichprobenumfang spielt für die Reliabilitäts- und Validitätsanalysen eine wichtige Rolle, denn dieser ist neben anderen Faktoren auch ausschlaggebend für die

Koeffizienten der Reliabilität bzw. Validität.²¹³ Es gilt, umso größer der Stichprobenumfang (N) ist, desto genauer sind die Ergebnisse eines Tests.²¹⁴ In diesem Abschnitt wird die durchschnittliche Stichprobengröße der Eigen- und Fremdpublikationen, die den Test d2 enthalten, analysiert. Wie im Kapitel 3 unter der Fragestellung 3 berichtet, ist bei einigen Untersuchungen die Stichprobengröße entscheidend, um das Forschungsproblem besser zu erkennen.²¹⁵ Um festzustellen, ob sich die Eigenpublikationen von den Fremdpublikationen in dieser Beziehung unterscheiden, werden die durchschnittlichen Stichprobenumfänge untersucht. Erwähnenswert ist, dass von den 124 ausgesuchten Artikeln 90 Veröffentlichungen für die genannte Analyse in den Eigen- und Fremdpublikationen verwendet werden. Artikel, die den Test d2 nur mit anderen Tests erwähnen, wo er als Vergleich für bzw. mit anderen Tests zur Verfügung steht oder der Test d2 eingesetzt wird, aber es kein Zusammenhang mit der jeweiligen Untersuchung gibt und somit keine weiteren Daten zum Test d2 vorliegen, werden bei dieser Untersuchung ausgeschlossen.

Für die durchschnittliche Stichprobengröße wird das arithmetische Mittel in unterschiedlichen Konstellationen berechnet. Die errechneten Durchschnittswerte der Stichprobengrößen werden danach bewertet, ob diese über oder unter 100 Probanden erfasst.

Insgesamt stehen 15 Veröffentlichungen der Eigenautoren (siehe Anhang Tabelle A-8) und 75 Publikationen der Fremdautoren (siehe Anhang Tabelle A-9) zur Verfügung. In der Abbildung 6 ist die durchschnittliche Stichprobengröße des Tests d2 in Eigen- und Fremdpublikationen dargestellt. Es ist festzustellen, dass die Eigenpublikationen ein größeres arithmetisches Mittel des Stichprobenumfangs bilden als die Fremdpublikationen. Im Durchschnitt werden bei den Veröffentlichungen der Eigenautoren 364 und der Fremdpublikationen 233 Versuchsteilnehmer erfasst. Dies ergibt bei den Eigenpublikationen durchschnittlich 131 Probanden mehr als bei den Fremdpublikationen.

²¹³ Vgl. Schmidt-Atzert; Amelang (2012), S. 141; 160.

²¹⁴ Vgl. Härtler (2014), S. 119.

²¹⁵ Vgl. Rumsey (2008), S. 77 ff.

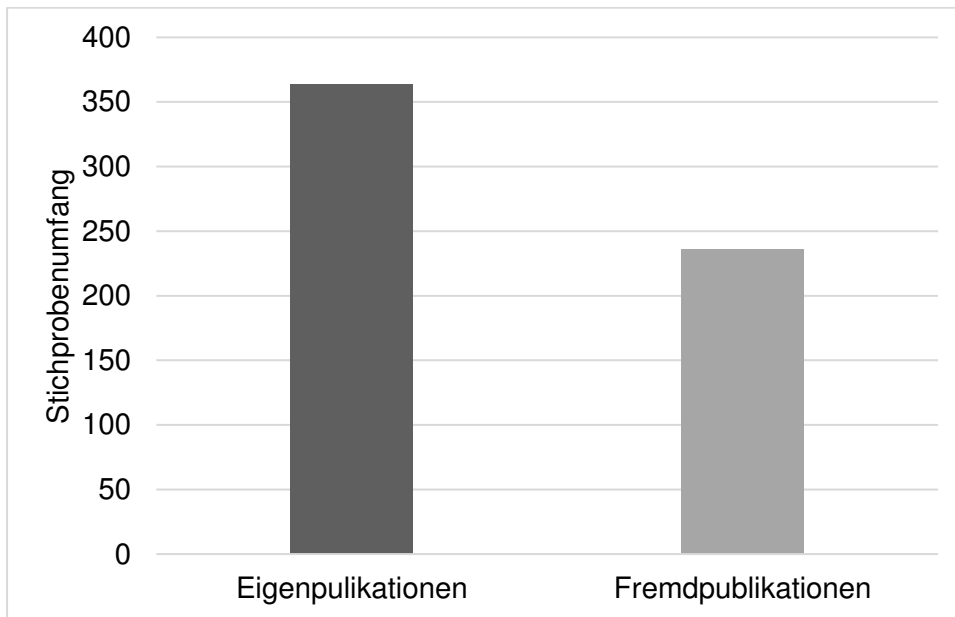


Abbildung 6: Durchschnittliche Stichprobengröße des Tests d2 – Aufmerksamkeits- und Belastungs-Test von Eigen- und Fremdpublikationen (Quelle: Eigene Darstellung)

Die Eigen- und Fremdveröffentlichungen liegen im bzw. über dem Bereich des empfohlenen Stichprobenumfangs (von 200 bis 250 Fällen), indem eine verbesserte Ergebnisqualität auftritt.²¹⁶

Die durchschnittlichen Stichprobengrößen in den Anwendungsbereichen des Tests d2 unterscheiden sich zwischen den Eigen- und Fremdpublikationen (siehe Tabelle 11). Die Tabelle 11 zeigt auf, dass die durchschnittlichen Stichprobengrößen in den einzelnen Anwendungsgebieten der Eigen- und Fremdpublikationen den Bereich unter 100 Fällen nicht unterschreiten und somit genauere Ergebnisse ermitteln. Die höchsten Werte liegen im Bereich der „Arbeits- und Organisationspsychologie“, sowohl bei den Eigenpublikationen (N= 604) als auch bei den Fremdpublikationen (N= 304). Beide durchschnittlichen Stichprobenumfänge liegen über den Stichprobenbereich von $N > 250$, der eine bessere Ergebnisqualität aufzeigt. In der „Klinischen und Medizinischen Psychologie“ sind nur Ergebnisse bei den Fremdpublikationen vorhanden. Das Ergebnis liegt unter dem Idealbereich der Stichprobengrößen, jedoch ist die durchschnittliche Stichprobengröße von $N= 107$ akzeptabel. Dies ist das geringste arithmetische Mittel der Fremdpublikationen. Die Eigenpublikationen zeigen in diesem Anwendungsgebiet sowie bei der „Pädagogischen Psychologie“

²¹⁶ Vgl. Homburg; Klarmann (2006), S. 733.

keine Ergebnisse auf. Bei der „Neuropsychologie“ ist die durchschnittliche Stichprobengröße der Eigenpublikation (N= 378) höher als bei den Fremdpublikationen (N= 175). Die Fremdpublikationen in der „Pädagogischen Psychologie“ zeigen einen hohen Wert von N= 293 auf. Dieses Ergebnis liegt im empfohlenen Bereich, um eine verbesserte Qualität der Ergebnisse zu erhalten. Werden die arithmetischen Mittel der Eigen- und Fremdpublikationen im Anwendungsbereich „Testbezogene Untersuchungen“ verglichen, ist bei den Fremdpublikationen ein höherer Wert dargestellt als bei den Eigenpublikationen. Das Ergebnis der Eigenautoren in der „Verkehrspsychologie“ ist im unteren akzeptablen Bereich (N= 104), im Gegensatz zum Wert der Fremdpublikationen (N=160), welcher höher als der Eigenpublikationen ist, sich dennoch nicht im empfohlenen Idealbereich befindet.²¹⁷

Tabelle 11: Durchschnittliche Stichprobengrößen der Eigen- und Fremdpublikationen in den einzelnen Anwendungsbereichen (Quelle: Eigene Darstellung)

Anwendungsbereiche	Durchschnittliche Stichprobengrößen der Eigenpublikationen	Durchschnittliche Stichprobengrößen der Fremdpublikationen
Arbeits- und Organisationspsychologie	628	304
Klinische und Medizinische Psychologie	-	107
Neuropsychologie	378	175
Pädagogische Psychologie	-	293
Testbezogene Untersuchungen	136	230
Verkehrspsychologie	104	160

²¹⁷ Vgl. ebenda, S. 733.

Wird das ursprüngliche Anwendungsgebiet „Verkehrspsychologie“ mit den anderen Anwendungsbereichen verglichen, ist bei den Eigenpublikationen der durchschnittliche Stichprobenumfang ($N=104$) am geringsten. Bei den Fremdautoren ist in diesem Anwendungsgebiet der durchschnittliche Stichprobenumfang um 56 Versuchsproubanden größer.

Im Vergleich sind die durchschnittlichen Stichprobengrößen bei den Fremdpublikationen in vier von sechs Kategorien größer als bei den Veröffentlichungen der Eigenautoren. Es gibt insgesamt zwei Anwendungsbereiche der Eigenpublikationen („Arbeits- und Organisationspsychologie“ sowie „Neuropsychologie“) und drei Anwendungsbereiche der Fremdpublikationen („Arbeits- und Organisationspsychologie“, „Pädagogische Psychologie“ und „Testbezogene Untersuchungen“), in denen der durchschnittliche Stichprobenwert den Empfehlungen entspricht bzw. über dem Empfehlungsbereich von 200 bis 250 Fällen liegt. Wie schon erwähnt sind bei der „Arbeits- und Organisationspsychologie“ die größten Stichprobengrößen vorzufinden, welche eine Verbesserung der Ergebnisqualität aufzeigen. Hierbei sind die der Eigenpublikationen größer als die der Fremdpublikationen, genauso wie im Bereich der „Neuropsychologie“.

Die durchschnittlichen Stichprobengrößen der Eigenpublikationen sind insgesamt größer als die Fremdpublikationen und lassen laut Literatur ein Forschungsproblem besser erkennen. Werden die durchschnittlichen Stichprobenumfänge in den einzelnen Anwendungsbereichen des Test d2 betrachtet, ist festzustellen, dass die Fremdpublikationen in allen Anwendungsbereichen mit einer durchschnittlichen Stichprobengröße von über $N \geq 100$ vorhanden sind. Diese Werte sind ebenfalls bei den Eigenpublikationen dargestellt, dennoch weisen diese nicht alle Anwendungsbereiche auf.

5.5 Reliabilitätsanalyse in Bezug auf die Stichprobenumfänge

Zur Prüfung der Reliabilität in Bezug der Stichprobenumfänge in den Eigen- und Fremdpublikationen wurden wie in Abschnitt 5.2 die unterschiedlichen Methoden der Reliabilität und dazu die Stichprobenumfänge fokussiert.

Die unterschiedlichen Methoden der Reliabilitäten werden auch in diesem Abschnitt

wieder getrennt voneinander betrachtet. Die Ergebnisse in den einzelnen Tabellen vom Abschnitt 5.2 werden ebenfalls übernommen. Bei den Retest-Reliabilitäten der Eigenpublikationen ist ersichtlich, dass die Stichprobengrößen sehr gering sind. In jeder Kategorie der Tabelle 5 sind die Stichprobengrößen unter $N=100$. Die größte Stichprobe der Eigenpublikationen liegt bei $N= 59$ und die kleinste bei $N= 31$. Die Reliabilitätskoeffizienten für die GZ- sowie die GZ-F-Werte nehmen mit sinkender Stichprobengröße ab (siehe Tabelle 5). Jedoch gibt es einen Korrelationskoeffizienten des GZ-F-Wertes ($r= 0,90$), welcher mit einem geringen Stichprobenumfang ($N= 31$) dem Ergebnis ($r= 0,92$) mit der größten Stichprobe ($N= 59$) ähnelt. Bei dem F%-Wert steigen die Ergebnisse mit zunehmender Stichprobengröße, anders als bei den SB-Werten. Diese zeigen Werte auf, indem mit sinkender Stichprobengröße der Wert geringfügig steigt. Bei $N= 59$ zeigt sich ein Retest-Reliabilitätskoeffizient von $r= 0,23$ und bei $N= 53$ ergibt sich $r= 0,28$. Im Gegensatz zu den GZ und GZ-F Kennwerten korrelieren die F% und SB-Werte gering.

Wird die Reliabilitätsschätzung nach der Aufteilung in die erste und in die zweite Testhälfte der Eigenpublikation betrachtet (siehe Anhang Tabelle A-4), ist das Ergebnis beim GZ-Wert ($r= 0,97$) höher als bei der Retest-Methode. Auch die Stichprobengröße ($N= 199$) ist höher als bei den anderen aufgezeigten Methoden der Reliabilität der Eigenautoren. Der angegebene Stichprobenumfang entspricht den Empfehlungen von $N \geq 100$.

In der Tabelle A-5 im Anhang ist bei der Odd-even-Methode ersichtlich, dass weniger Versuchsteilnehmer als bei der vorherigen Methode vorzufinden sind, dennoch ist die Stichprobengröße mit $N= 150$ im empfohlenen Bereich der Reliabilitätsbestimmungen. Der Korrelationswert von GZ ($r= 0,93$) der Eigenpublikation im Vergleich zur Aufteilung in die erste und zweite Testhälfte hat ebenfalls abgenommen.

Die Retest-Reliabilitäts-Methode bei den Fremdpublikationen weist auch Stichprobengrößen unter $N= 100$ auf. Die Spanne des Stichprobenumfangs beträgt von $N= 18$ bis $N= 70$ (siehe Tabelle 6) und ist damit größer als bei den Eigenpublikationen. Die dazugehörigen Ergebnisse bei GZ sind geringer als bei den Eigenpublikationen. Ein Wert von Eser (1987) besitzt bei geringer Stichprobengröße ($N= 22$) die höchste Korrelation ($r= 0,75$) der aufgezeigten Ergebnisse in Tabelle 6. Die anderen Ergeb-

nisse des GZ-Wertes weisen niedrige Korrelationen zu den kleinen Stichprobengrößen auf. Auffällig ist, dass die größte Stichprobengröße (N=70) von Dingel (1971) bei diesen aufgeführten Werten eine niedrigere Korrelation ($r = 0,56$) aufzeigt als die Untersuchung mit einem Stichprobenumfang von N= 22 mit $r = 0,75$. Ähnlich wie bei den GZ-Kennwerten stellen sich die Ergebnisse in Verbindung mit den Stichprobenumfängen bei dem d2-Kennwert GZ-F dar. Bei dem größten Stichprobenumfang N= 70 ist eine Korrelation von $r = 0,71$ erzielt worden. Die Untersuchung bei der nächst kleineren Stichprobengröße (N= 22) erzielt einen geringfügig größeren Korrelationswert von $r = 0,73$. Bei der kleinsten Stichprobengröße (N= 18) von Eser (1987) ist das Ergebnis am geringsten ($r = 0,50$). Der größte Korrelationswert der F%-Werte ($r = 0,61$) liegt bei einer Stichprobengröße von N= 22 vor, wohingegen bei der kleinsten Stichprobe (N= 18) die niedrigste Korrelation gefunden wurde ($r = 0,37$). Die Reliabilitätskorrelation des Wertes F von $r = 0,46$ tritt bei einem Stichprobenumfang von N= 70 auf. Bei den SB-Werten sinken die Korrelationswerte mit dem Stichprobenumfang. Wie bei den anderen Werten auch, gilt es neben den heterogenen Stichprobenumfängen auf die Intervalle zu achten, denn diese haben ebenfalls einen Einfluss auf die Korrelationskoeffizienten der Reliabilitäten.

Mit der Betrachtung der Korrelationskoeffizienten bei der Retest-Reliabilität in Abhängigkeit des Stichprobenumfangs ist insgesamt zu sagen, dass nicht die größte Stichprobe auch den höchsten Korrelationswert aufzeigt.

Mittels der Testhalbierungsmethode, nach der Aufteilung der ersten und zweiten Testhälfte der Fremdpublikationen, ergibt sich bei einer Stichprobengröße N= 90 ein hoher Korrelationswert von 0,97. Auch der F%-Wert weist einen hohen Zusammenhang auf (siehe Anhang Tabelle A-6).

Die Odd-even-Methode enthält eine etwas geringere Stichprobengröße (N= 85) als die Methode nach der Aufteilung in die erste und zweite Testhälfte, dennoch sind die d2-Kennwerte im hohen Korrelationsbereich (siehe Anhang Tabelle A-7).

Es ist zu beachten, dass die Stichprobenumfänge der letzten beiden genannten Methoden unter den Empfehlungen von $N \geq 100$ Fällen liegen und somit als geringe

Stichprobenumfänge gelten, die als problematisch gesehen werden, um aussagefähige Ergebnisse zu erzielen.²¹⁸

Im Vergleich zwischen den Eigen- und Fremdpublikationen ohne die Betrachtung der spezifischen Einordnung in die Anwendungsgebiete ist ersichtlich, dass die Veröffentlichungen der Eigenautoren mehr Stichprobenumfänge aufweisen, die größer sind als die der Fremdautoren. Den größten Stichprobenumfang bei der Retest-Methode weist die Fremdpublikationen auf mit $N=70$, jedoch erzielt diese keinen großen Korrelationskoeffizienten. Die Eigenpublikationen berichten höhere Korrelationskoeffizienten bei den d^2 -Kennwerten GZ und GZ-F als die Fremdpublikationen. Bei den $F\%$ -Werten sind die Ergebnisse der Eigenpublikationen mit einem größeren Stichprobenumfang ebenfalls höher als bei den Fremdpublikationen. Die Reliabilitätskoeffizienten für die SB-Kennwerte sind bei den Fremdpublikationen trotz niedriger Stichprobengröße höher als bei den Eigenpublikationen.

Auffallend ist, dass nach der Aufteilung der ersten und zweiten Testhälfte der Eigen- sowie der Fremdpublikation bei unterschiedlichen Stichprobenumfängen der Korrelationskoeffizient beim Kennwert GZ gleich auftritt.

Auch der Korrelationskoeffizient der Eigenpublikationen bei der Odd-even-Methode weist einen hohen Korrelationswert (GZ: $r=0,93$) bei einer Stichprobe von $N=199$ auf. Ähnlich wie bei der Aufteilung der ersten und zweiten Testhälfte sind die Korrelationskoeffizienten, trotz einer unterschiedlichen Stichprobengröße (Eigenpublikation: $N=150$; Fremdpublikation: $N=85$), beim Wert GZ der Eigen- und Fremdautoren gleich.

5.6 Validitätsanalyse in Bezug auf die Stichprobenumfänge

Die Werte der Validitätsuntersuchungen werden aus dem Abschnitt 5.3 übernommen. Die Stichprobengrößen der Eigen- und Fremdpublikationen bei den untersuchten Validitätskoeffizienten sind höher als bei den Reliabilitätsuntersuchungen. Die Korrelationen der Validitäten sind im geringen bis hohen Bereich angesiedelt.

²¹⁸ Vgl. Homburg; Klarmann (2006), S. 733.

Beginnend mit der konvergenten Validität der Eigenpublikationen ist, wie schon erwähnt (siehe Abschnitt 5.3), nur ein Wert in Tabelle 7 aufgegriffen. Es ist zu sagen, dass die Validitätskoeffizienten nicht mit steigender Stichprobengröße bei der konvergenten Validität (siehe Tabelle 7) zunehmen. Es gibt eine hohe Stichprobengröße, die geringe Korrelationen bei den Kennwerten zwischen dem Test d2 und dem Revisions-Test aufzeigen. Korrelationen zwischen Intelligenztests und dem Konzentrationstest Test d2 sind ebenfalls mit einer Veröffentlichung der Eigenautoren vorhanden. Die diskriminante Validität dokumentiert niedrige Korrelationen bei einer hohen Stichprobengröße, die über dem Empfehlungsbereich liegt (siehe Tabelle 8).

Die Stichprobenumfänge der Fremdpublikationen der konvergenten Validität treten zwischen dem Bereich von $N= 50$ bis $N= 70$ auf. Somit erzielen die vier Untersuchungen (siehe Tabelle 9) den Mindeststichprobenumfang von $N \geq 100$ nicht. Die geringen Stichprobenumfänge besitzen Korrelationswerte, die unterschiedliche Zusammenhänge ausweisen. Somit ist ein hoher Validitätskoeffizient ($r= 0,72$) bei dem Kennwert GZ mit einer geringen Stichprobengröße ($N= 60$) zu verzeichnen. Der GZ-F-Kennwert bei der Untersuchung zwischen dem KKT und dem Test d2 weist einen hohen Zusammenhang bei einer geringen Stichprobengröße ($N= 70$) auf. Der F%-Wert korreliert mit dem Fehlerwert des KKT im niedrigen Bereich signifikant, die anderen beiden Werte werden nicht beachtet, aufgrund der nicht vorhandenen Signifikanz. Die niedrigste Stichprobengröße bei der Untersuchung von Sommer (1973) erzielt einen GZ-F-Wert von $r= 0,34$, welcher bei einem kleinen Stichprobenumfang zu erwarten ist. Ein mittlerer Zusammenhang der Validitätskoeffizienten besteht zwischen dem Test d2 und dem DKT-K des Leistungstempos und dem GZ-Wert sowie zwischen der Leistungsgüte und dem F%-Wert.

Interessant ist bei der konvergenten Validität der Vergleich zwischen den Eigen- und Fremdpublikationen. Die Ergebnisse der Eigenpublikationen fallen mit einer größeren Stichprobengröße deutlich kleiner aus als die Ergebnisse der Fremdpublikationen. Erwähnenswert ist der Vergleich bei den Eigen- und Fremdpublikationen zwischen den d2-Kennwerten und den Kennwerten des Revisions-Tests. Hierbei ist nur der Kennwert GZ gegenüberzustellen. Dieser weist in den ausgewählten Artikeln große Unterschiede auf. Der höhere Kennwert von $r= 0,72$ tritt bei den Fremdautoren mit 60 Versuchspersonen auf. Im Gegensatz dazu gibt es eine Studie bei

den Eigenpublikationen, die bei einer Stichprobengröße von $N = 524$ einen GZ-Wert von $r = 0,30$ aufzeigt. Ein weiterer Wert einer anderen Studie der Fremdautoren mit dem Test d2 und dem KKT zeigt ebenfalls bei einer geringen Stichprobe ($N = 70$) einen hohen Validitätskoeffizienten bei dem Kennwert GZ-F ($r = 0,60$) auf. Mit der Gegenüberstellung ist festzustellen, dass die Eigenpublikation bei der Korrelation der d2-Kennwerte mit einem anderen Konzentrationstest einen höheren Stichprobenumfang aufweist als die Fremdpublikationen. Die Validitätskorrelationen sind jedoch kleiner als bei den Fremdpublikationen. Bei den Fremdpublikationen ist darauf zu achten, dass die Validitätsuntersuchungen geringere Stichprobengrößen aufzeigen und somit die Möglichkeit besteht, dass sich die ermittelten Werte verschieben.²¹⁹

Die Fremdpublikationen bei der diskriminanten Validität weisen deutlich höhere Stichprobenumfänge aus als bei der konvergenten Validität. Hierbei liegt die Spanne der Stichprobengrößen bei $N = 74$ bis $N = 1030$.

Die meisten Stichprobengrößen der Fremdpublikationen liegen im bzw. über den empfohlenen Stichprobenumfang, sodass hier die Ergebnisse ebenfalls relativ genau sind. Die Stichprobengrößen bei den Korrelationen der d2-Kennwerte mit anderen Intelligenztests der Fremdpublikationen sind von sechs Untersuchungen zwei Stichproben unter $N = 100$. Es gibt zwei Validitätsuntersuchungen, die im Bereich zwischen $N \geq 100$ und $N < 200$ und zwei Untersuchungen, die über $N > 250$ liegen. Die höchsten Werte der Korrelationen sind bei den Stichprobenumfängen zu finden, welche eine kleine Stichprobengröße enthalten (siehe Tabelle 10 mit den folgenden Tests: Raven SPM von Merten, 2000, $N = 74$ und HAWIE von Rauchfleisch, 1983, $N = 90$). Werden die anderen Werte der Korrelationen betrachtet, weisen diese größtenteils einen niedrigen Zusammenhang auf.

Werden bei den Fremdpublikationen Untersuchungen verglichen, welche die gleichen Tests verwenden, fällt darunter der Raven SPM. Hierbei sind die Ergebnisse von Merten (2000) mit denen von Kühnlein et al. (1983) gegenüberzustellen. Die Untersuchung von Merten (2000) enthält eine deutlich geringere Stichprobengröße

²¹⁹ Vgl. Schmidt-Atzert; Amelang (2012), S. 160.

(N= 74) mit mittleren Korrelationswerten ($r^{GZ} = 0,52$; $r^{GZ-F} = 0,58$) als die von Kühnlein et al. (1983). Die Stichprobengröße von Kühnlein et al. (1983) ist wesentlich größer (N= 194) und erzielt bei GZ-F eine Korrelation von $r = 0,341$. Aufgrund der höheren Stichprobengröße ist die Untersuchung von Kühnlein et al. (1983) zu präferieren. Ähnlich wie die Gegenüberstellung des Raven SPM sind Unterschiede zwischen den Untersuchungen des HAWIE von Wiese und Kroj (1972) (N= 102) sowie von Rauchfleisch (1983) (N= 90) zu erkennen, die Korrelationen zwischen dem HAWIE und dem Test d2 ermitteln. Dabei sind die Ergebnisse von den Autoren Wiese und Kroj (1972), hinsichtlich ihrer Stichprobengröße zu bevorzugen.

Bei den genannten Vergleichen (Raven SPM und HAWIE) ist jedoch nicht außer Acht zu lassen, dass nicht nur die Stichprobengröße mit den unterschiedlichen Werten in Verbindung steht, sondern auch die Heterogenität zwischen der Probandenauswahl. Denn die Untersuchungen mit der jeweils geringeren Stichprobengröße beziehen sich auf neurologische Patienten bzw. Patienten mit hirndiffusem Psychosyndrom und neurotischen Störungen.

Der Vergleich zwischen den Eigen- und Fremdpublikationen bei den Korrelationen der d2-Kennwerte mit den Intelligenztests verhält sich ähnlich wie bei den Korrelationen der d2-Kennwerte mit den Konzentrationstests. Die Tabelle 8 zeigt Korrelationswerte zwischen dem IST 70 und dem Test d2 auf. Die Stichprobengröße ist bei einer geringen Korrelation hoch. Der hohe Stichprobenumfang lässt eine relativ genaue Schätzung zu. Die Fremdpublikationen weisen höhere Zusammenhänge auf, die größtenteils im geringen Korrelationsbereich liegen, wie es bei der diskriminanten Validität vorgesehen ist.²²⁰

²²⁰ Vgl. Bühner (2011), S. 64.

6 Diskussion der Untersuchungsergebnisse

In diesem Kapitel soll herausgefunden werden, ob der Test d2 in der heutigen Zeit noch ein aussagekräftiges und häufig genutztes Testinstrument ist. Dabei wurden die unterschiedlichen Anwendungsgebiete, die Reliabilitäten und Validitäten, der durchschnittliche Stichprobenumfang sowie die Reliabilitäten und Validitäten in Verbindung mit den Stichprobengrößen in Eigen- und Fremdpublikationen untersucht. Es gibt zehn Testgütekriterien, die von psychologischen Tests zu erfüllen sind. Dazu zählen Objektivität, Reliabilität, Validität, Skalierung, Normierung, Testökonomie, Nützlichkeit, Zumutbarkeit, Unverfälschbarkeit und Fairness.²²¹ Nicht auf alle diese Kriterien konnte im Rahmen dieser Arbeit eingegangen werden und somit beschränkt sich die Untersuchung auf die Reliabilität (Retest-Reliabilität und Split-Half-Methode) und Validität (Konstruktvalidität). Bei diesem Vergleich zwischen Eigen- und Fremdpublikationen spielen auch die Anwendungsgebiete des Tests d2 eine Rolle, um eine Veränderung des Tests d2 in den Domänen herauszufinden. Auch die Stichprobengröße der Eigen- und Fremdpublikationen wurde untersucht, um zu prüfen, ob die Werte der Gütekriterien willkürlich sind.

In diesem Kapitel sollen die erhobenen Ergebnisse auf der Basis wissenschaftlicher Erkenntnisse und in Orientierung an die gebildeten *Hypothesen*, welche im Kapitel 2 erläutert wurden, interpretiert werden. Zudem werden Schlussfolgerungen zu den Ergebnissen der Unterschiede bei den Anwendungsgebieten, der Gütekriterien Reliabilität und Validität, der durchschnittlichen Stichprobengröße sowie der Reliabilität und Validität in Verbindung mit den Stichprobengrößen des Tests d2 – Aufmerksamkeits- und Belastungs-Test in Eigen- und Fremdpublikationen gezogen und diese kritisch hinterfragt.

²²¹ Vgl. Kubinger (2003), S. 195 ff.

6.1 Diskussion der Ergebnisse zu den Anwendungsgebieten bei der Nutzung des Tests d2 von Eigen- und Fremdpublikationen

Fragestellung 1:

Unterscheiden sich die Anwendungsgebiete bei der Nutzung des Tests d2 – Aufmerksamkeits- und Belastungs-Test von Eigen- und Fremdpublikationen?

H₁: Die Domänen der Verwendung des Tests d2 – Aufmerksamkeits- und Belastungs-Test haben sich in der betrachteten Publikationszeit bei den Eigen- und Fremdpublikationen geändert.

Für die Untersuchung der ersten Arbeitshypothese wurden die Anwendungsgebiete des Tests d2 – Aufmerksamkeits- und Belastungs-Test von Eigen- und Fremdpublikationen in dem betrachteten Publikationszeitraum geprüft. Die Artikel wurden händisch ausgezählt und den Eigen- und Fremdpublikationen sowie den einzelnen Anwendungsbereichen zugeordnet (siehe Anhang Tabellen A-2 und A-3). Die Auszählung der Artikel in den unterschiedlichen Zeitschriften hat gezeigt, dass nicht alle Anwendungsgebiete bei den Publikationen benötigt werden. Im betrachteten Publikationszeitraum gibt es Jahre, in denen keine Veröffentlichungen des Tests d2 erscheinen. Vor allem bei den Eigenpublikationen ist es ersichtlich, dass nicht alle Anwendungsgebiete bedient werden und die betrachteten Zeiträume lückenhaft sind (siehe Abbildung 4). Wird die „Verkehrspsychologie“ bei den Eigenpublikationen betrachtet, sind nur Artikel in den Jahren 1962 und 1966 vorhanden. Dominierend sind die Anwendungsgebiete „Arbeits- und Organisationspsychologie“ sowie „Testbezogene Untersuchungen“. Veröffentlichungen der „Arbeits- und Organisationspsychologie“ sind seit Anfang der 90er Jahre bis zum Jahr 2011 vermehrt aufgetreten. Die „Testbezogenen Untersuchungen“ treten seit 1973 bis zum Jahr 2012 sporadisch auf. Auch die Fremdpublikationen zeigen vermehrt die Dominanz der „Arbeits- und Organisationspsychologie“ auf. Das Anwendungsgebiet „Verkehrspsychologie“ ist bei den Fremdautoren ebenfalls gering vertreten.

Es ist festzustellen, dass das ursprüngliche Anwendungsgebiet der „Verkehrspsychologie“ beim Test d2 bei den Eigen- und Fremdpublikationen gegenwärtig keine

große Rolle mehr spielt. Der Test d2 wird sowohl bei den Eigen- als auch Fremdauf-toren dominierend im Anwendungsbereich „Arbeits- und Organisationspsychologie“ eingesetzt, welcher seit Ende der 80er, Anfang der 90er Jahre vermehrt bei den Fremdauf-toren sowie ab Anfang der 2000er Jahre von Eigen- und Fremdauf-toren in den untersuchenden Zeitschriften auftritt (siehe Abbildungen 4 und 5). Dabei liegen die aktuelleren Publikationen in diesem Anwendungsbereich bei den Fremdauf-toren vor. Aufgrund der aufgezeigten Ergebnisse bei den Änderungen der Anwendungs-bereiche im betrachteten Publikationszeitraum in den Eigen- sowie Fremdpublikati-onen wird die Arbeitshypothese 1 angenommen. Die Ergebnisse entsprechen zum Teil einer Studie von Roth und Herzberg (2008), die belegt, dass die „Arbeits- und Organisationspsychologie“ hinter der „Klinischen Psychologie“ und der „Pädagogi-schen Psychologie“ eines der größten Arbeitsfelder der befragten Psychologen ein-nimmt.²²² Bei den Fremdpublikationen ist die „Pädagogische Psychologie“ am dritt-häufigsten verwendet worden. Im Vergleich zu den Ergebnissen von Roth und Herz-berg (2008) steht das Arbeitsfeld „Pädagogische Psychologie“ mit 19 % auf dem zweiten Rang bei der Befragung von Testeinsätzen.²²³ Eine hohe Beliebtheit des Tests d2 in der „Arbeits- und Organisationspsychologie“ zeigt schon relativ früh die Untersuchung von Brambring (1983) auf. Hierbei wird festgestellt, dass der Test d2 in deutschen Unternehmen Anklang findet.²²⁴ Bei einer anderen Untersuchung von Steck (1997) wurden die Praxisfelder der befragten Psychologen aufgegriffen. Da-bei hat sich ergeben, dass die Psychotherapie sowie die „Klinische und Medizini-sche Psychologie“ bei mehrfach Nennung bevorzugte Kategorien sind. Die „Arbeits- und Organisationspsychologie“ wird immerhin als fünft häufigstes Praxisfeld ange-geben.²²⁵ Anhand der genannten Studien zeigt sich somit, dass die „Arbeits- und Organisationspsychologie“ Zuspruch bei den Psychologen findet, was mit dem Er-gebnis dieser Untersuchung übereinstimmt.

²²² Vgl. zitiert nach Schmidt-Atzert; Amelang (2012), S. 185.

²²³ Vgl. ebenda, S. 185.

²²⁴ Vgl. Brambring (1983), S. 445 ff.

²²⁵ Vgl. Steck (1997), S. 273.

In der deutschen Wirtschaft gehört der Test d2 zu den drei häufigsten angewendeten psychologischen Tests.²²⁶ Dies spiegelt ebenfalls die Änderung des Anwendungsgebietes in dieser Untersuchung wider. Ein Grund für die vielseitige Nutzung des Tests d2 in anderen Anwendungsgebieten wie der „Pädagogischen Psychologie“, der „Klinischen und Medizinischen Psychologie“, der „Neuropsychologie“ und der „Arbeits- und Organisationspsychologie“, ist der ökonomische Vorteil, den der Test d2 mit sich bringt und er bei vielen Anwendern beliebt ist.²²⁷

H₂: Die Fremdpublikationen weisen mehr Anwendungsgebiete auf als die Eigenpublikationen.

Die Fremdpublikationen weisen im vorgesehenen Publikationszeitraum mehr Artikel als bei den Eigenpublikationen auf. Der Abschnitt 5.1 gibt zu den einzelnen Ergebnissen in den Anwendungsbereichen der Eigen- und Fremdpublikationen Auskunft. Es ist deutlich dargestellt, dass die Fremdpublikationen alle ausgewiesenen Anwendungsgebiete erfüllen und diese in jedem Bereich mit der Anzahl der Artikel dominieren, wohingegen die Eigenpublikationen nur vier Anwendungsgebiete aufzeigen (siehe Tabelle 4). Folglich wird die Arbeitshypothese 2 angenommen.

Bei der Auswertung ist zu beachten, dass nur ein Teil der psychologischen Fachzeitschriften ausgewertet wurde. Demnach fehlen Artikel in den einzelnen Anwendungsgebieten der Eigenpublikation, die spezifischer bspw. die „Klinische und Medizinische Psychologie“ oder die „Pädagogische Psychologie“ behandeln und deswegen in speziellen Fachzeitschriften veröffentlicht werden. Es ist hinzuzufügen, dass von dem Testautor und den Coautoren seit der Entstehung des Tests d2 Neuauflagen des Manuals erschienen sind, um den Test d2 – Aufmerksamkeits- und Belastungs-Test zu verbessern und auf einem aktuellen Stand zu halten. Die Manuale wurden in den folgenden Jahren veröffentlicht: 1962 (1. Auflage), 1967 (2. Auflage), 1970 (3. Auflage), 1972 (4. Auflage), 1975 (5. Auflage), 1978 (6. Auflage), 1981 (7. Auflage), 1994 (8. Auflage), 2002 (9. Auflage) und 2010 (10. Auflage).²²⁸

²²⁶ Vgl. Brickenkamp (2002), S. 265.

²²⁷ Vgl. Oswald; Hagen (1997), S. 88.

²²⁸ Vgl. Oehlschlägel; Moosbrugger (1991), S. 43 ff.; Brickenkamp; Schmidt-Atzert; Liepmann (2010), S. 16.

Mit der genaueren Betrachtung der Jahreszahlen der Auflagen des Manuals vom Test d2 werden diese in regelmäßigen Abständen veröffentlicht. Dies ist vermutlich auch einer der Gründe für die wenigen Veröffentlichungen der Eigenautoren. Das Manual beinhaltet meist aufwendige Untersuchungen und Normierungen des Tests d2.²²⁹ Besonders im Zeitraum von 1974 bis 1990 sind keine Eigenpublikationen (siehe Abbildung 4) in den untersuchten Zeitschriften zu finden, dafür sind im genannten Zeitfenster drei Manuale entstanden.

Die Tatsache, dass die Fremdautoren in ihren Artikeln auch auf Fehler im Test d2 hinweisen,²³⁰ hat vermutlich einen Anreiz für die Eigenautoren, den Test d2 ständig zu überprüfen sowie zu erneuern und dies in die Manuale dementsprechend zu überarbeiten. Die Fremdautoren haben in den untersuchten 53 Jahren und bezogen auf die betrachteten Zeitschriften 86 Publikationen mehr veröffentlicht als die Eigenautoren. Dabei ist zu beachten, dass es mehr Fremdautoren gibt als die Eigenautoren. Des Weiteren sind die Fremdautoren in unterschiedlichen Fachgebieten tätig und können dadurch mehr Anwendungsgebiete bedienen bzw. weitgehender forschen als die Eigenautoren. Mit diesem eindeutigen Ergebnis ist ebenfalls festzustellen, dass sich die Eigen- und Fremdpublikationen ergänzen, um den Test d2 auf einem aktuellen Stand zu halten und diesen trotz vieler weiterer erscheinender Tests nicht in Vergessenheit geraten zu lassen. Anhand dieser Forschungen wird deutlich, dass der Test d2 einen hohen Stellenwert bei den Forschern und Anwendern aufzeigt.

²²⁹ Vgl. Brickenkamp; Schmidt-Atzert; Liepmann (2010), S. 16.

²³⁰ Vgl. Heinrich (1973), S. 120; Oehlschlägel; Moosbrugger (1991), S. 43 ff.; Fay (1995), S. 294 f.

6.2 Diskussion der Gütekriterien Reliabilität und Validität bei den in Eigen- und Fremdpublikationen

Fragestellung 2:

Unterscheiden sich die Gütekriterien Reliabilität und Validität ohne und mit den spezifischen Verwendungen des Tests d2 – Aufmerksamkeits- und Belastungs-Test in Eigen- und Fremdpublikationen?

H₁: Es gibt Unterschiede in den dokumentierten Angaben des Gütekriteriums Reliabilität ohne Betrachtung der spezifischen Verwendungen des Tests d2 – Aufmerksamkeits- und Belastungs-Test in Eigen- und Fremdpublikationen.

Zur Prüfung, ob die Reliabilitäten ohne Betrachtung der spezifischen Verwendungen sich bei den Eigen- und Fremdpublikationen unterscheiden sind die Ergebnisse der Tabellen 5, 6, A-4, A-5, A-6 und A-7 verglichen worden.

Dabei hat sich herausgestellt, dass sich die Ergebnisse bei der Retest-Reliabilität der Eigenpublikationen deutlich von den Fremdpublikationen unterscheiden. Die Werte der Eigenpublikationen sind in den gesamten Kennwerten, außer dem SB-Wert, reliabler als die der Fremdpublikationen. Von den fünf dargestellten Untersuchungen der Eigenpublikationen in der Tabelle 5 sind jeweils zwei Werte den Kennwerten GZ, F% und SB zuzuordnen, die sich bei der Retest-Reliabilität im geringen bis mittleren Korrelationsbereich ($r = 0,23$ bis $0,89$) befinden.²³¹ Im Vergleich dazu sind die Werte der Gesamtleistung (GZ-F) deutlich reliabler und weisen höhere Korrelationen auf als die anderen angegebenen Kennwerte. Der Kennwert GZ-F liegt bei zwei Ergebnissen im hohen Korrelationsbereich ($r = 0,92$ und $0,90$), zwei Koeffizienten liegen im mittleren ($r = 0,88$) und ein Koeffizient liegt im geringen Bereich ($r = 0,72$).

Die Kennwerte der Retest-Reliabilität der Fremdpublikationen (siehe Tabelle 6) liegen alle im niedrigen Korrelationsbereich (von $r = 0,24$ bis $0,75$).²³² Die meisten Koeffizienten sind bei dem Kennwert GZ-F, wie bei den Eigenpublikationen, höher als

²³¹ Vgl. Fisseni (2004), S. 80.

²³² Vgl. ebenda, S. 80.

bei den anderen genannten Kennwerten. Dennoch besteht bei diesen Ergebnissen ein geringer Zusammenhang.²³³

Mit diesen Betrachtungen ist zu erkennen, dass die Ergebnisse der Eigenpublikationen sich deutlich von den Fremdpublikationen bei der Retest-Reliabilität unterscheiden. Vor allem die Gesamt- und die Mengenleistungen sind bei den Eigenpublikationen stabil und zeigen Werte im mittleren bis hohen Korrelationsbereich auf, um die Anforderungen an einen Test zu bestehen.²³⁴

Werden daraufhin die Retest-Intervalle der Untersuchungen von Eigen- und Fremdpublikationen beachtet, sind diese jeweils unterschiedlich (siehe Tabellen 5 und 6). In der Literatur wird angegeben, dass die Reliabilitätsschätzungen von den Intervallen abhängig sein können.²³⁵ Dennoch bewegen sich bei dem Kennwert GZ-F der Eigenpublikationen vier von fünf Untersuchungen mit Intervallen von drei bis 23 Monaten im mittleren bis hohen Korrelationsbereich der Reliabilität und die Werte liegen nah beieinander, was für ein stabiles Item spricht.²³⁶ Der Kennwert GZ-F der Fremdpublikationen ist in einem Retest-Intervall von vier bis 24 Monaten angegeben, der jedoch eine deutlich geringere Korrelation aufzeigt. Auch hier ist es auffällig, dass die Gesamtleistung (GZ-F) bei Dingel (1971) und Eser (1987), trotz des unterschiedlichen Zeitintervalls (24 sowie vier Monate) und der heterogenen Stichproben, ähnliche Werte aufweist. Werden die GZ-F-Werte ($r = 0,72$) der Jugendlichen von Brickenkamp (1991) mit den GZ-F-Werten ($r = 0,73$) der Kinder und Jugendlichen mit Verhaltensstörungen von Eser (1987) verglichen, ist bei einem ähnlichen Zeitintervall von drei bzw. vier Monaten das Ergebnis fast gleich. Bis auf einen ähnlichen GZ-F-Wert der Eigen- und Fremdpublikationen sind die Ergebnisse ungleich.

Die erheblichen Unterschiede zwischen den ermittelten Retest-Reliabilitäten der Ei-

²³³ Vgl. Fisseni (2004, S. 80).

²³⁴ Vgl. Schmidt-Atzert; Amelang, (2012), S. 141.

²³⁵ Vgl. Bühner (2011), S. 61; Schermelleh-Engel; Werner (2012), S. 125.; Schmidt-Atzert; Amelang (2012), S. 141.

²³⁶ Vgl. Schmidt-Atzert; Amelang (2012), S. 141.

gen- und Fremdpublikationen sind wahrscheinlich der Stichprobengröße geschuldet²³⁷ (siehe Abschnitt 5.5), da die Intervalle der Eigen- und Fremdpublikationen nah beieinanderliegen, aber unterschiedliche Werte aufzeigen.

Bei der Testhalbierungsmethode (Reliabilitätsschätzung nach Aufteilung in die erste und zweite Testhälfte und der Odd-even-Methode) sind keine Unterschiede zwischen den Ergebnissen der Eigen- und Fremdaufgaben festzustellen. Beide weisen hohe Reliabilitäten auf. Bei der Testhalbierungsmethode ist zu beachten, dass vor allem bei den Speed-Tests, zu denen der Test d2 auch eingeordnet wird, wegen seiner hohen Ähnlichkeit in den Testaufgaben, es triviale Weise zu hohen Reliabilitätsschätzungen führt. Aus diesem Grund wird eher eine Retest-Methode empfohlen, da die Reproduzierbarkeit der Ergebnisse bei einer Testwiederholung am aussagefähigsten ist.²³⁸

Dieser Empfehlung folgend, dass zwischen den Eigen- und Fremdpublikationen bei der Betrachtung der Retest-Reliabilität deutliche Unterschiede bestehen, wird die Arbeitshypothese 1 angenommen, obwohl die Testhalbierungsmethode keinen Unterschied bei den Reliabilitätskoeffizienten aufzeigt.

Bei einer genauen Betrachtung der Ergebnisse der Retest-Reliabilität fällt auf, dass vor allem der Kennwert GZ-F bei den Eigen- und Fremdpublikationen höhere Korrelationskoeffizienten aufzeigt als die anderen dargestellten Kennwerte. Das zeigt, dass dieser Kennwert im Gegensatz zu den anderen reliabler ist.

H₂: Es gibt Unterschiede in den dokumentierten Angaben des Gütekriteriums Reliabilität bei spezifischen Verwendungen des Test d2 – Aufmerksamkeits- und Belastungs-Test in Eigen- und Fremdpublikationen.

Um Unterschiede zwischen den Reliabilitäten in den spezifischen Verwendungen bei den Eigen- und Fremdpublikationen festzustellen, sind gleiche Anwendungsfelder verglichen worden. Dabei kam jedoch nur ein Anwendungsgebiet bei den Retest-Reliabilitäten in Frage, die „Verkehrspsychologie“ (siehe Tabellen 5 und 6). Die anderen Methoden zeigen keine gleichen Anwendungsgebiete auf. Die Ergebnisse

²³⁷ Vgl. Moosbrugger (2012), S. 116; Schmidt-Atzert; Amelang (2012), S. 141.

²³⁸ Vgl. Bühner (2011), S. 157.

der Eigenpublikationen weisen auch bei dieser spezifischen Verwendung höhere Reliabilitätskoeffizienten auf als die Fremdpublikationen. Die Arbeitshypothese 2 wird somit bestätigt.

Bei der Berechnung der Retest-Reliabilitäten ist darauf zu achten, dass konstante wahre Werte der Versuchsteilnehmer und konstante Fehlervarianzen als Voraussetzung gegeben sind. Die gleiche Fehlervarianz ist für Testdurchgänge mit standardisierten Testdurchführungen einsichtig, sodass die Messfehler nicht kleiner oder größer werden. Im Gegensatz dazu sind die wahren Werte sensibel, die sich zwischen der ersten und zweiten Testwiederholung verändern können. Dabei kommt es auf die Art der Veränderung an, welche die Reliabilitätskoeffizienten verfälschen.²³⁹ Ein Hinweis für die unterschiedlichen Werte beim Vergleich der Eigen- und Fremdpublikationen sind bspw. die Zeitintervalle. Bei den Eigenpublikationen der „Verkehrspsychologie“ ist zu erkennen, dass die Ergebnisse, die in einem kürzeren Zeitintervall erfasst werden, höher sind als bei den Ergebnissen mit einem längeren Zeitintervall (siehe Tabellen 5 und 6). Dabei ist die Wahl der Retest-Intervalle entscheidend, wie die Ergebnisse ausfallen.²⁴⁰

Die Länge der Zeitintervalle spielen bei den systematischen²⁴¹ und unsystematischen Veränderungen der wahren Werte²⁴² eine Rolle. Es ist darauf zu achten, dass mit einem zu kurz gewählten Abstand der Testwiederholungen die unsystematischen Merkmalsveränderungen abnehmen und größere Übungseffekte die Ergebnisse verfälschen.²⁴³ Mit einem längeren Retest-Intervall ist das Risiko der Erinne-

²³⁹ Vgl. Schermelleh-Engel, K.; Werner, C. S. (2012), S. 123.

²⁴⁰ Vgl. Bühner (2011), S. 61; Schermelleh-Engel, K.; Werner, C. S. (2012), S. 125.; Schmidt-Atzert (2012), S. 141.

²⁴¹ Systematische Veränderungen der wahren Werte sind bei Retest-Reliabilitäten unkritisch, da sie bei den getesteten Personen gleich ausfallen. Mittels des Subtrahierens oder Addierens eines konstanten Betrags aller Ergebnisse verändern sich die Korrelationen zwischen den Testwiederholungen nicht. [Vgl. Schermelleh-Engel; Werner (2012), S. 123.]

²⁴² Bei der unsystematischen Veränderung der wahren Werte verändern sich die Werte zwischen den Messzeitpunkten für verschiedene Personen unterschiedlich, somit reduziert sich die Korrelation zwischen der ersten und zweiten Testwiederholung. [Vgl. ebenda, S. 123.]

²⁴³ Vgl. Schermelleh-Engel, K.; Werner, C. S. (2012), S. 125.

rungseffekte geringer, jedoch ist es möglich, dass die unsystematischen Merkmalsveränderungen potenziell störender werden.²⁴⁴

Bei dem Vergleich der Ergebnisse von Brickenkamp (1966) und Dingel (1971) (siehe Tabellen 5 und 6) mit einem fast gleichen Zeitintervall von 23 und 24 Monaten mit ähnelnden Versuchspersonen gibt Dingel (1971) für die niedrigeren Werte als Grund die „...hohe psychische Belastung bei Kraftfahreignungsuntersuchungen...“²⁴⁵ an. Um weitere Aussagen über mögliche Ursachen unterschiedlicher Retest-Reliabilitäten zu treffen, ist es nötig mehr Publikationen von Eigen- und Fremdautoren zu untersuchen, die die gleichen spezifischen Anwendungsfelder aufzeigen.

H₃: Es gibt Unterschiede in den dokumentierten Angaben des Gütekriteriums Validität ohne Betrachtung der spezifischen Verwendungen des Test d2 – Aufmerksamkeits- und Belastungs-Test in Eigen- und Fremdpublikationen.

Mit der Betrachtung der Unterschiede zwischen den dokumentierten Angaben des Gütekriteriums Validität ohne die spezifischen Verwendungen des Tests d2 – Aufmerksamkeits- und Belastungs-Test in Eigen- und Fremdpublikationen wird deutlich, dass die Validität des Test d2 bei den Fremdautoren in den vorliegenden Zeitschriften häufiger gemessen wird als bei den Eigenautoren. Bei den Publikationen der Eigenautoren ist jeweils eine Validitätsprüfung der konvergenten und diskriminanten Validität eingegangen im Gegensatz zu den Fremdautoren, die vier bzw. sechs Untersuchungen aufweisen.

Die Tabellen 7 und 9 zeigen die konvergenten Validitäten auf. Dabei sind für konstruktnahe Tests hohe Zusammenhänge zu erwarten.²⁴⁶

Die Ergebnisse von Schmidt-Atzert und Amelang (1997) korrelieren bei der Untersuchung von Test d2 und Revisions-Test im geringen Bereich. Im Vergleich dazu sind die Ergebnisse der Fremdautoren im geringen bis hohen Korrelationsbereich vertreten. Die meisten Korrelationskoeffizienten weisen einen mittleren bis hohen

²⁴⁴ Vgl. Schermelleh-Engel, K.; Werner, C. S. (2012), S. 125.

²⁴⁵ Dingel (1971), S. 86.

²⁴⁶ Vgl. Bühner (2011), S. 64.

Zusammenhang bei den Tempowerten auf.

Bei der diskriminanten Validität sind niedrige Zusammenhänge wünschenswert, da das rein zu messende Konstrukt der Konzentration, zu erfassen ist.²⁴⁷

Die Korrelation der d2-Kennwerte mit anderen Intelligenztests (siehe Tabellen 8 und 10) hat bei der Eigenpublikation gezeigt, dass geringe Korrelationen im Bereich zwischen $r = 0,14$ und $0,17$ erzielt werden. Dies verhält sich bei den Fremdpublikationen ähnlich. Hierbei sind die meisten Korrelationswerte im niedrigen Bereich zu verzeichnen, reichen aber etwas weiter von $r = 0,20$ bis $0,58$. Wie bereits schon erwähnt, ist die mittlere Korrelation von dem GZ- und GZ-F-Kennwert ($r^{\text{GZ}} = 0,52$ und $r^{\text{GZ-F}} = 0,58$) ein unerwünschtes Ergebnis.

Trotz dieser zwei auffälligen Ergebnisse weisen die Fremdautoren bei mehreren Untersuchungen der diskriminanten Validität niedrige Korrelationsbereiche auf. Mit dem Ergebnis, dass die Fremdautoren bei der konvergenten und diskriminanten Validität mehr Publikationen aufzeigen und sich vor allem die Ergebnisse der konvergenten Validität deutlich von den Eigenautoren unterscheiden, wird die Arbeitshypothese 3 angenommen.

Bei der konvergenten Validitätsprüfung sind Korrelationen bei dem Kennwert GZ von $r = 0,33$ bis $0,72$ und bei GZ-F von $0,34$ bis $0,60$ zu erkennen. Die meisten Ergebnisse werden bei dem Gesamtleistungswert abgebildet. Die höchsten Korrelationen zeigen die GZ- und GZ-F-Werte bei den Fremdpublikationen auf. Die Fehlerprozentwerte (F%) sind im geringen Korrelationsbereich. Dennoch besteht kein Vergleichswert von den Eigenpublikationen, um dazu Aussagen zu treffen, ob diese Ergebnisse valider sind als die Ergebnisse der Eigenautoren.

Brickenkamp et al. (2010) beschreiben im zehnten Testmanual, dass die Korrelationswerte von GZ mit Tempowerten, dem GZ-F-Wert mit der Anzahl an Richtigen und dem Fehlerprozentwert (F%) mit Fehlerwerten anderer Konzentrationstests einen hohen Zusammenhang aufweisen.²⁴⁸ Dies spiegelt sich auch in den hier dargestellten Werten wider. Die höchsten Korrelationswerte liegen zwischen dem Test

²⁴⁷ Vgl. Bühner (2011), S. 64.

²⁴⁸ Vgl. Brickenkamp; Schmidt-Atzert; Liepmann (2010), S. 52.

d2 und dem Revisions-Test sowie dem Test d2 und dem KKT (Komplexer Konzentrations-Test) vor. Dabei ist auffallend, dass die genannten Konzentrationstests keine Durchstreichtests, sondern ein Rechentest (Revisions-Test) und ein komplexer Konzentrationstest sind. Letzterer setzt sich aus verschiedenen Aufgabenarten zusammen (Rechenaufgaben, Durchstreichen von Zeichen, Buchstabenfolge von Wörtern in umgekehrter Reihenfolge aufschreiben). Auch Brickenkamp et al. (2010) erwähnen diesbezüglich, dass hohe Korrelationswerte bei anderen Konzentrationstests auftauchen, die keine Durchstreichtests sind.²⁴⁹ Den Autoren zufolge ist das ein Indiz dafür, dass der Test d2 von der Fähigkeit Konzentration abhängt und nicht „[...] nur vom Geschick im Umgang mit einem bestimmten Aufgabenmaterial.“²⁵⁰

Im Vergleich dazu sind die anderen angegebenen Ergebnisse der Eigen- und Fremdpublikationen im niedrigen bis mittleren Bereich (siehe Tabellen 8 und 10). Die Eigenpublikationen weisen geringere Korrelationen als die Fremdpublikationen auf. Es ist zu sagen, dass die Fremdpublikationen einen schwachen Zusammenhang zwischen der Konzentration und der Intelligenz herstellen. Die Eigenpublikation ermittelt ebenfalls einen schwachen Zusammenhang zwischen Konzentration und Intelligenz, jedoch ist dieser noch schwächer als bei den Fremdpublikationen. Auch die Autoren Wiese und Kroj (1972) sowie Westhoff und Kluck (1983) haben den schwachen Zusammenhang zwischen Konzentrationsleistungen und Intelligenzaspekten festgestellt.²⁵¹ Somit wird die Aussage von Brickenkamp (1978) widerlegt, dass es keinen Zusammenhang zwischen Intelligenz und Konzentration gibt.²⁵²

Die konvergente Validität muss deutlich höher sein als die diskriminante Validität, welches sich bei den Untersuchungen bestätigt hat.²⁵³ Besonders fallen bei dem

²⁴⁹ Vgl. Brickenkamp; Schmidt-Atzert; Liepmann (2010), S. 52 f.

²⁵⁰ Ebenda, S. 53.

²⁵¹ Vgl. Wiese; Kroj (1972), S. 697 f.; Westhoff; Kluck (1983), S. 310.

²⁵² Vgl. Brickenkamp (1978), S. 8.

²⁵³ Vgl. Schmidt-Atzert; Amelang (2012), S. 149.

Vergleich die unterschiedliche Stichprobengröße sowie die Heterogenität der Stichproben auf.²⁵⁴ Diese können Gründe sein, weshalb sich die Werte in den Eigen- und Fremdpublikationen unterscheiden. Erwähnenswert sind dazu die Ergebnisse der Eigenpublikation. Die Problematik mit dem Stichprobenumfang wird bei der Fragestellung 4 mit der Arbeitshypothese 2 genauer erläutert (siehe Abschnitt 6.4).

Bühner (2011) merkt an, dass bei der Beurteilung der Validität nicht nur die Validitätskoeffizienten in Betracht gezogen werden sollen, sondern auch der Beitrag den ein Test zu einer Fragestellung bereithält sowie die Situation, in der ein Test Anwendung findet.²⁵⁵ Die Untersuchungsergebnisse in Bezug auf andere Konzentrations- bzw. Intelligenztests zeigen auf, dass der Test d2 Validität besitzt und das misst, was er messen soll.

H₄: Es gibt Unterschiede in den dokumentierten Angaben des Gütekriteriums Validität bei spezifischen Verwendungen des Test d2 – Aufmerksamkeits- und Belastungs-Test in Eigen- und Fremdpublikationen.

Die Unterschiede der dokumentierten Angaben des Gütekriteriums Validität bei spezifischen Verwendungen des Test d2 – Aufmerksamkeits- und Belastungs-Test sind gering, jedoch vorhanden. Bei der Validitätsuntersuchung wird zwischen den Korrelationen der d2-Kennwerte mit anderen Konzentrationstests und Intelligenztests unterschieden. Für die spezifische Verwendung des Test d2 bei der konvergenten sowie diskriminanten Validität hat sich ergeben, dass die „Arbeits- und Organisationspsychologie“ bei beiden Publikationsformen auftritt und für den Vergleich geeignet ist.

Der Vergleich der Validität bei den Korrelationen zwischen den d2-Kennwerten und anderen Konzentrationstests der Eigen- und Fremdpublikationen in der „Arbeits- und Organisationspsychologie“ basiert nur auf dem GZ-Wert, da bei den anderen d2-Kennwerten keine Korrelationen der Fremdpublikationen ermittelt wurden. Es hat sich herausgestellt, dass der Kennwert GZ der Fremdpublikationen ($r = 0,72$) höher ausfällt als bei den Eigenveröffentlichungen ($r = 0,30$). Beide Korrelationen

²⁵⁴ Vgl. Schmidt-Atzert; Amelang (2012), S. 160.

²⁵⁵ Vgl. Fisseni (2004), S. 81.

sind mit den gleichen Tests (Test d2 und Revisions-Test) entstanden.²⁵⁶

Die Ergebnisse für diskriminante Validität sind zwischen beiden Publikationsbereichen im Bereich „Arbeits- und Organisationspsychologie“ unterschiedlich. Auch die Anzahl der Publikationen ist unterschiedlich. Somit weisen die Fremdpublikationen in diesem Anwendungsbereich drei Publikationen auf, die Eigenpublikationen hingegen nur eine. Die GZ-Werte der Fremdpublikationen erzielen Korrelationen zwischen $r = 0,20$ und $0,29$, der GZ-F-Wert erreicht Ergebnisse zwischen $r = 0,26$ bis $0,35$.

Die aufgezeigten Kennwerte GZ ($r = 0,14$) und GZ-F ($r = 0,17$) der Eigenpublikationen sind im Gegensatz niedriger. Ein Vergleich zwischen den Korrelationen für die Fehlerwerte ist aufgrund der fehlenden Angaben nicht möglich.

Die Arbeitshypothese 4 ist durch den Unterschied zwischen den Eigen- und Fremdpublikationen zu bestätigen.

Die Analyse der vorherigen Hypothese, welche ausführlich diskutiert wurde, zeigt keine überraschenden Ergebnisse auf. Dies ist noch einmal ein kleiner Ausschnitt aus der größeren Datensammlung der Validitäten. Es zeigt, dass selbst in den einzelnen Anwendungsgebieten Unterschiede zwischen Eigen- und Fremdpublikationen nachweisbar sind.

Auch hier ist zu beachten, dass die Zusammenhänge zwischen den verwendeten Tests bei der konvergenten Validität größer sein sollten als bei der diskriminanten. Um nicht nur Aussagen zu dem Anwendungsgebiet „Arbeits- und Organisationspsychologie“ zu treffen, sind weitere Untersuchungen aus anderen Fachzeitschriften bzw. anderer Literatur notwendig, die Aufschluss über die eventuellen Unterschiede zwischen Eigen- und Fremdpublikationen geben.

6.3 Diskussion des durchschnittlichen Stichprobenumfangs bei Eigen- und Fremdpublikationen

Fragestellung 3:

²⁵⁶ Vgl. Marschner (1971), S. 14 ff.; Schmidt-Atzert; Ising (1997), S. 241 ff.

Unterscheiden sich beim Test d2 – Aufmerksamkeits- und Belastungs-Test die Eigen- und Fremdpublikationen im durchschnittlichen Stichprobenumfang?

H₁: Die Fremdpublikationen haben einen größeren durchschnittlichen Stichprobenumfang als die Eigenpublikationen.

Aus den zu untersuchenden Artikel in den unterschiedlichen Fachzeitschriften ist der durchschnittliche Stichprobenumfang aus den Eigen- und Fremdpublikationen ermittelt worden. Mit der Untersuchung der Unterschiede beim Test d2 der Eigen- und Fremdpublikationen des durchschnittlichen Stichprobenumfangs hat sich herausgestellt, dass die Eigenpublikationen ein deutlich höheres arithmetisches Mittel im Stichprobenumfang erzielen als die Fremdpublikationen (siehe Abbildung 6).

In beiden Publikationsbereichen ist ersichtlich, dass die durchschnittlichen Stichprobengrößen den Empfehlungen der Literatur entsprechen und somit für eine bessere Ergebnisqualität sorgen.²⁵⁷ Folglich ist damit bewiesen, dass es Unterschiede beim Test d2 in den Eigen- und Fremdpublikationen im durchschnittlichen Stichprobenumfang gibt, jedoch sind die durchschnittlichen Stichprobengrößen der Eigenpublikationen größer als die Fremdpublikationen. Somit wird die Arbeitshypothese 1 nicht bestätigt und die Nullhypothese angenommen.

Die Vermutung, dass die durchschnittlichen Stichprobenumfänge bei den Fremdpublikationen größer sind als bei den Eigenpublikationen, liegt zum einen an der größeren Anzahl der Veröffentlichungen (siehe Tabelle 2) und zum anderen an dem Nachweis zur Änderung von Kennzahlen sowie Anregungen zu Verbesserungen des Test d2 der Fremdautoren.²⁵⁸ Des Weiteren bietet die Berechnung des arithmetischen Mittels einen weiteren Anlass zur kritischen Betrachtung der Ergebnisse. Dieses lässt sich leicht von sehr kleinen bzw. großen Werten beeinflussen.²⁵⁹ Werden die durchschnittlichen Stichprobenumfänge der Eigen- und Fremdpublikationen genau betrachtet, fällt auf, dass die Veröffentlichungen der Fremdautoren

²⁵⁷ Vgl. Homburg; Klarmann (2006), S. 733.

²⁵⁸ Vgl. Wiese; Kroj (1972), S. 690 ff.; Heinrich (1973), S. 118 ff.; Janssen (1975), S. 226 ff.; Westhoff; Kluck (1983), S. 311 ff.; Hasselhorn; Körkel (1984), S. 288 ff.; Oehlschlägel; Moosbrugger (1991), S. 43 ff.; Fay (1992), S. 303 ff.; Fay (1995), S. 293 ff.

²⁵⁹ Vgl. Rumsey (2015), S. 64.

viele Untersuchungen vorzeigen, die eine kleine Stichprobengröße von $N < 100$ sowie größere Stichprobenumfänge von $N \geq 1000$ aufzeigen, welches sich auf die gesamte durchschnittliche Stichprobengröße auswirkt.

In der Literatur wird angegeben, dass sich die Höhe der ermittelten Daten proportional auf deren Genauigkeit auswirkt.²⁶⁰

Die Stichprobengrößen bei den Eigen- und Fremdpublikationen, die im Idealbereich liegen, kommen aus unterschiedlichen Anwendungsgebieten (siehe Anhang Tabellen A-8 und A-9). Die durchschnittlichen Stichprobengrößen der Eigenpublikationen sind insgesamt größer als die Fremdpublikationen und lassen mittels der größeren durchschnittlichen Stichprobengröße laut Literatur ein Forschungsproblem besser erkennen.²⁶¹

H₂: Die Fremdpublikationen haben einen größeren durchschnittlichen Stichprobenumfang in den verschiedenen Anwendungsgebieten als die Eigenpublikationen.

Es gibt ebenfalls Unterschiede der durchschnittlichen Stichprobengröße in den einzelnen Anwendungsgebieten des Tests d2. Dabei sind die Fremdpublikationen in den untersuchenden arithmetischen Mitteln der Stichprobengrößen führend, denn zwei von sechs spezifischen Anwendungsgebieten werden bei den Eigenveröffentlichungen nicht aufgezeigt und zwei Anwendungsgebiete sind kleiner als die Fremdpublikationen (siehe Tabelle 11). Dadurch wird die Arbeitshypothese 2 angenommen. Bei den analysierten Artikeln ist weiterhin aufgefallen, dass im Durchschnitt die empfohlene Mindeststichprobengröße von $N \geq 100$ bei Eigen- und Fremdpublikationen vorliegt.²⁶²

Wird auch hier wieder das ursprüngliche Anwendungsgebiet „Verkehrspsychologie“ mit den anderen Anwendungsbereichen verglichen, fällt auf, dass vor allem bei den Eigenpublikationen der durchschnittliche Stichprobenumfang ($N=104$) am geringsten ist. Bei den Fremdautoren ist in diesem Anwendungsgebiet der durchschnittliche

²⁶⁰ Vgl. ebenda, S. 204.

²⁶¹ Vgl. ebenda, S. 204.

²⁶² Vgl. Boomsma (1985), S. 240.

Stichprobenumfang um 56 Versuchsprobanden größer.

Es gibt insgesamt zwei Anwendungsbereiche der Eigenpublikationen („Arbeits- und Organisationspsychologie“ sowie „Neuropsychologie“) und drei Anwendungsbereiche der Fremdpublikationen („Arbeits- und Organisationspsychologie“, „Pädagogische Psychologie“ und „Testbezogene Untersuchungen“), in denen der durchschnittliche Stichprobenwert den Empfehlungen entspricht bzw. über dem Empfehlungsbereich von 200 bis 250 Fällen liegt. Wie schon erwähnt, sind bei der „Arbeits- und Organisationspsychologie“ die größten durchschnittlichen Stichprobengrößen vorzufinden. Jedoch sind die der Eigenpublikationen in diesem Anwendungsbereich größer als die der Fremdpublikationen.

Im Anwendungsbereich der „Neuropsychologie“ ist von den Eigenautoren lediglich eine Veröffentlichung mit einer hohen Stichprobengröße zu finden (siehe Anhang Tabelle A-8). Der Vergleich mit den Fremdautoren zeigt eine größere Anzahl an Fremdpublikationen im Bereich der „Neuropsychologie“. Diese weisen durchschnittlich geringere Stichprobengrößen im Bereich zwischen $N= 40$ bis $N= 372$ auf, woraus zwangsläufig eine Verringerung des arithmetischen Mittels resultiert (siehe Anhang Tabelle A-9).

Mit den durchschnittlichen Stichprobengrößen der einzelnen Anwendungsgebiete von Eigen- und Fremdpublikationen wurde dargestellt, dass die Artikel der Fremdautoren größere durchschnittliche Stichprobengrößen aufzeigen und somit für eine bessere Qualität der Ergebnisse geeignet ist.

Mittels dieser Untersuchung ist eine differenziertere Betrachtung möglich. Dabei fällt vor allem auf, dass die Anwendungsgebiete „Arbeits- und Organisationspsychologie“ sowie „Neuropsychologie“ bei den Eigenpublikationen hohe durchschnittliche Stichprobengrößen aufweisen. Die anderen Anwendungsgebiete lassen bei den Eigenpublikationen eine geringere durchschnittliche Stichprobengröße als bei den Fremdpublikationen erkennen. Das Ergebnis des Wertes für das arithmetische Mittel der Stichprobengröße wird durch sehr hohe Stichproben in wenigen Bereichen gegenüber geringeren Stichproben in vielen Bereichen verfälscht (siehe Tabellen 11, A-8 und A-9). Dadurch wird deutlich, weshalb unter anderem der Wert der Eigenpublikationen bei der ersten Hypothese höher als bei den Fremdpublikationen ausfällt. Wie bereits erwähnt lässt sich der Mittelwert von sehr kleinen oder großen

Daten beeinflussen²⁶³ und wirkt sich somit auf die gesamte durchschnittliche Stichprobengröße aus. Mit der Arbeitshypothese 2 wird detaillierter aufgezeigt, wie sich das arithmetische Mittel der Stichprobenumfänge zusammensetzt, wodurch sich andere Perspektiven ergeben. Es wird dargestellt, dass die Fremdpublikationen in vier von sechs verschiedenen Anwendungsgebieten eine größere durchschnittliche Stichprobengröße ausweisen.

6.4 Diskussion der Gütekriterien Reliabilität und Validität mit Bezug auf die Stichprobenumfänge von Eigen- und Fremdpublikationen

Fragestellung 4:

Unterscheiden sich die Reliabilitäts- und Validitätsangaben im Test d2 – Aufmerksamkeits- und Belastungs-Test bei Eigen- und Fremdpublikationen bezogen auf unterschiedliche Stichprobenumfänge?

H₁: Die Fremdpublikationen weisen höhere Reliabilitäten aus, aufgrund ihrer größeren Stichproben als die Eigenpublikationen.

Bei der Untersuchung der Reliabilitäten der Fragestellung 2 mit der Arbeitshypothese 1 (siehe Abschnitt 5.2) wurden die Ergebnisse der Eigen- und Fremdpublikationen ohne Berücksichtigung auf die Stichprobengrößen untersucht. In diesem Abschnitt stehen die Stichprobengrößen mit den Reliabilitäten im Mittelpunkt, jedoch werden hier die spezifischen Anwendungsgebiete außer Acht gelassen. Unter der Fragestellung 4 im Kapitel 3 wird darauf hingewiesen, dass die Untersuchungen zu den Reliabilitäten, vor allem den Retest-Reliabilitäten, oft zu geringe Stichprobenumfänge aufzeigen und dadurch die Korrelationskoeffizienten beeinträchtigen.²⁶⁴ In beiden Publikationsbereichen bei der Retest-Reliabilität ist es der Fall, dass die

²⁶³ Vgl. Rumsey (2015), S. 204.

²⁶⁴ Vgl. Schmidt-Atzert; Amelang (2012), S. 141.

Stichprobengrößen gering sind und nicht der empfohlenen Mindeststichprobengröße von $N \geq 100$ des Autors Boomsma (1985) entsprechen.²⁶⁵ Die Fremdpublikationen der Retest-Reliabilitäten zeigen bei zwei Untersuchungen Stichprobengrößen von $N= 18$ und $N= 22$ auf (siehe Tabelle 6). Dies ist im Vergleich zu den aufgezeigten Stichprobengrößen der Eigenpublikationen (im Bereich von $N= 31$ bis $N= 59$) sehr gering (siehe Tabelle 5). Werden dabei die Ergebnisse genauer betrachtet, fällt auf, dass sich der Korrelationskoeffizient für den GZ-Wert, bei einem fast gleichen Intervall, bei den Jugendlichen ($N= 37$) von Brickenkamp (1991) und bei den Kindern und Jugendlichen mit Verhaltensstörungen ($N= 22$) von Eser (1987) ähnelt. Dabei hat Brickenkamp (1991) Kinder und Jugendliche getrennt voneinander betrachtet im Gegensatz zu Eser (1987), der die Kinder und Jugendliche mit Verhaltensstörungen als eine Gruppe zusammengefügt hat. Es ist möglich, dass das Ergebnis der Retest- Reliabilitäten durch die unterschiedliche Verwendung der Probandenstichproben beeinflusst wird. Auch die Autoren Schmidt-Atzert und Ising (1997) erwähnen das in ihrer Untersuchung als Ursache für unterschiedliche Ergebnisse.²⁶⁶

Trotz der angenommenen geringen Korrelationen mittels der geringen Stichprobengrößen sind die meisten Retest-Reliabilitäten bei den GZ- und GZ-F-Kennwerten der Eigenpublikationen im mittleren bis hohen Korrelationsbereich. Das bedeutet, dass die Eigenpublikationen trotz geringer Stichprobengröße reliabler sind als die Fremdpublikationen.

Werden die Fehlerwerte und die Ergebnisse des SB-Wertes in Betracht gezogen ist ersichtlich, dass diese geringer sind. Diese Ergebnisse sind normal, da bei Schnelligkeits-Tests, wie es der Test d2 auch ist,²⁶⁷ die Fehlerwerte bzw. Fehlerprozentwerte geringer sind als die Tempowerte.²⁶⁸ Da die SB-Werte sehr niedrige Werte aufzeigen, lässt sich vermuten, dass sie nicht stabil genug sind. Das haben auch Brickenkamp et al. (2010) erkannt und eliminierten daraufhin diesen Wert im Test

²⁶⁵ Vgl. Boomsma (1985), S. 240.

²⁶⁶ Vgl. Schmidt-Atzert; Ising (1997), S. 248.

²⁶⁷ Vgl. Brickenkamp, Schmidt-Atzert, Llepmann (2010), S. 14.

²⁶⁸ Vgl. Bühner (2011), S. 159.

d2-R.²⁶⁹

Aufgrund der geringeren Korrelationskoeffizienten mittels der geringen Stichprobengrößen bei den Fremdpublikationen ist es möglich, dass die Stichprobengröße die Ergebnisse bei den Fremdautoren mehr beeinträchtigt als bei den Eigenautoren.²⁷⁰ Weitere Gründe für die unterschiedlichen Ergebnisse sind bspw. nicht nur die heterogenen Stichprobengrößen, sondern auch die unterschiedlichen Längen der Reliabilitäts-Intervalle. „Für die Retest-Reliabilität gilt, dass sie grundsätzlich **hoch** ausfallen soll, wenn das **Retest-Intervall** kurz ist und das gemessene Merkmal als stabil gilt.“²⁷¹

Werden die Reliabilitäten der Split-Half-Methode der Eigen- und Fremdpublikationen betrachtet, sind bei beiden Methoden die Reliabilitäten gleich, trotz eines hohen Unterschiedes in den Stichprobengrößen. Dabei sind die Stichprobengrößen der Eigenpublikationen höher und weisen Stichprobengrößen über $N \geq 100$ auf (siehe Anhang Tabellen A-4 und A-5), im Gegensatz zu den Fremdpublikationen ($N= 90$; $N= 85$) (siehe Anhang Tabellen A-6 und A-7).

Die Split-Half-Methode oder Halbierungsmethode kommt den instrumentellen Messgenauigkeiten am nächsten, vorausgesetzt die Homogenität und die Anzahl der Items in zwei gleiche Hälften des Tests sind gegeben.²⁷²

Dennoch fallen die Retest-Reliabilitäten, aufgrund der größeren Anzahl der Publikationen, sowohl bei den Eigen- als auch Fremdautoren, mehr ins Gewicht. Es hat sich herausgestellt, dass die Fremdpublikationen keine höheren Reliabilitäten als die Eigenpublikationen aufweisen, weder bei der Retest-Methode noch bei der Aufteilung der ersten und zweiten Testhälfte oder Odd-even-Methode. Es zeigt sich, dass die Stichprobengrößen keinen großen Einfluss auf die Ergebnisse der Reliabilitäten haben, denn bspw. trotz einer größeren Stichprobenanzahl von $N= 70$ der Fremdpublikationen (siehe Tabelle 6) hat sich keine größere Korrelation bei den d2-Kennwerten ($r^{GZ} = 0,56$; $r^{GZ-F} = 0,71$; $r^F = 0,46$) ergeben. Dies lässt sich ebenfalls mit

²⁶⁹ Vgl. Brickenkamp, Schmidt-Atzert; Liepmann (2010), S. 46.

²⁷⁰ Vgl. Schmidt-Atzert; Amelang (2012), S. 141 f.

²⁷¹ Ebenda, S. 141.

²⁷² Vgl. Schmidt-Atzert; Amelang (2012), S. 140.

einer kleineren Stichprobengröße ($N=53$) der Eigenpublikationen vergleichen, welche ein ähnliches Zeitintervall aufzeigt und größere Ergebnisse ($r^{GZ} = 0,71$; $r^{GZ-F} = 0,88$; $r^{F\%} = 0,51$; $r^{SB} = 0,28$) darstellt.

Zu vermuten ist, dass die Reliabilitäten von dem Zeitintervall sowie den heterogenen Stichproben abhängig sind.²⁷³

Somit wird Arbeitshypothese 1 abgelehnt und die Nullhypothese bestätigt.

H₂: Die Fremdpublikationen weisen höhere Validitäten aus, aufgrund ihrer größeren Stichprobe als die Eigenpublikationen.

Die Validitäten der Eigen- und Fremdpublikationen wurden mittels der Stichprobengrößen betrachtet. Dabei hat sich herausgestellt, dass sich die Validität mittels eines größeren Stichprobenumfangs nicht erhöhen, wie bspw. bei den Untersuchungen der Eigenpublikationen (siehe Tabellen 7 und 8). Die höchsten Validitäten mit einer geringen Stichprobengröße sind den Fremdpublikationen nachzuweisen (siehe Tabellen 9 und 10). Die Eigenpublikationen zeigen trotz einer großen Stichprobe geringe Validitätskoeffizienten auf. Es ist bei der Validitätsuntersuchung zu beachten, dass nur die Konstruktvalidität eine Rolle spielt, da hierfür vergleichbare Daten zwischen Eigen- und Fremdpublikationen gefunden wurden. Aufgrund der höheren Validitätskoeffizienten, jedoch geringen Stichprobengrößen bei den Fremdpublikationen im Vergleich zu den Eigenpublikationen wird die Arbeitshypothese 2 abgelehnt und die Nullhypothese bestätigt. Es ist zu beachten, dass zwischen der konvergenten und diskriminanten Validität zu unterscheiden ist, denn bei der diskriminanten Validität sind niedrige Zusammenhänge erwünscht. Bei den konvergenten Validitäten werden hohe Zusammenhänge erwartet. Weitere Einflussfaktoren für die erhöhte Validität sind bspw. die Reliabilität des Tests oder eine gemeinsame Methodenvarianz.²⁷⁴

Stichprobenumfänge bei Validitätsuntersuchungen fallen in der Regel klein aus. Dadurch kann es zu einer falschen Einschätzung der Validitäten von Tests in der

²⁷³ Vgl. Schermelleh-Engel; Werner (2012), S. 125; Schmidt-Atzert; Amelang (2012), S.141.

²⁷⁴ Vgl. Schmidt-Atzert; Amelang (2012), S. 154.

Population kommen. Es gilt, je kleiner eine Stichprobengröße ist, desto mehr können sich die Validitätskoeffizienten nach unten oder oben verschieben.²⁷⁵ Diese genannten Unterschiede treten zwischen den Korrelationen der d2-Kennwerte und anderen Konzentrationstests auf. Den passenden Beweis dafür liefern die zwei unterschiedlichen Untersuchungen zwischen dem Test d2 und dem Revisions-Test von Eigen- und Fremdautoren. Bei dem Vergleich der Untersuchungen der Autoren Schmidt-Atzert und Ising (1997) sowie Marschner (1971) werden die Ergebnisse aufgrund der unterschiedlichen Stichprobengrößen deutlich. Die Studie in Bezug auf eine höhere Stichprobengröße von Schmidt-Atzert und Ising ($N= 524$; $r= 0,30$) weisen einen deutlich geringeren Korrelationswert der Mengenleistungen (GZ) zwischen dem Test d2 und dem Revisions-Test auf als bei Marschner (1971; $N= 60$; $r= 0,72$) (siehe Tabellen 7 und 9).²⁷⁶ Vor allem mit dieser Untersuchung wird deutlich, wie sehr die Stichprobenumfänge die Validitätskoeffizienten beeinflussen.

Mit der Betrachtung der konvergenten Validität der Fremdpublikationen ist somit auffällig, dass diese eine geringe Stichprobengröße aufzeigen und somit auch die Ergebnisse vorsichtig betrachtet werden müssen. Es gibt keine Untersuchung bei den Fremdpublikationen, welche die konvergente Validität ermittelt, indem der Stichprobenumfang $N \geq 100$ ist (siehe Tabelle 9).

Werden die Stichprobengrößen der diskriminanten Validität betrachtet, sind hierbei die Stichprobenumfänge der Fremdpublikationen meistens $N \geq 100$ (siehe Tabelle 10). Der Stichprobenumfang der Eigenpublikation ist auch hier sehr hoch ($N \geq 1000$). Mittels der unterschiedlichen Stichprobengrößen der Fremdpublikationen ist erkennbar, dass sich diese im geringen bis mittleren Bereich aufhalten. Der höhere Korrelationswert von $r^{GZ} = 0,52$ bzw. $r^{GZ-F} = 0,58$, mit einer kleinen Stichprobengröße, ist bei der diskriminanten Validität eine Ausnahme.

Für die diskriminante Validität ist es im legitimen Bereich, dass die Zusammenhänge im Vergleich zur konvergenten Validität klein ausfallen.²⁷⁷

Bei der konvergenten und diskriminanten Validität gibt es Stichprobenunterschiede.

²⁷⁵ Vgl. Hartig; Frey; Jude (2012), S. 161; Schmidt-Atzert; Amelang (2012), S. 160.

²⁷⁶ Vgl. Marschner (1971), S. 19; Schmidt-Atzert; Ising (1997), S. 241 ff.

²⁷⁷ Vgl. Hartig; Frey; Jude (2012), S. 158.

Sowohl bei der Eigenpublikation, als auch bei den Fremdpublikationen ist der Stichprobenumfang bei der konvergenten Validität deutlich kleiner als bei der diskriminieren.

Es ist zu empfehlen, weitere Untersuchungen mit größeren Stichprobengrößen durchzuführen, vor allem bei den Fremdpublikationen, welche die konvergente Validität prüfen, um eine bessere Aussage zur Validität zu treffen und im Vergleich von Eigen- und Fremdpublikationen gegenüberzustellen zu können. Auch bei den Eigenpublikationen ist es wünschenswert mehr Publikationen hinzuzuziehen, um einen eindeutigeren Beweis für die Unterschiede aufzuzeigen.

6.5 Diskussion der Untersuchungsmethode

Ein Problem zeigte sich darin, dass der Erfolg der Datenerhebung von den ausgewählten Fachzeitschriften, welche nicht vollständig zusammengetragen werden konnten, abhing. Somit ist die Vollständigkeit der Datenerhebung nicht ausnahmslos gesichert. Bei genauer Betrachtung der Daten der Eigenpublikationen fällt auf, dass zwei Anwendungsbereiche nicht mit dem des Test d2 nachweisbar sind.

Ein weiteres Problem zeigt sich bei der Auswahl von Reliabilitäts- und Validitätsergebnissen, denn es musste darauf geachtet werden, dass die Eigen- und Fremdaufgaben keine Modifikation am Test d2 – Aufmerksamkeits- und Belastungs-Test vorgenommen haben. Publikationen, welche dieses Kriterium erfüllten, konnten in die Datensammlung aufgenommen werden. Aus diesem Grund nimmt die Datenauswahl einen geringeren Umfang ein, als ursprünglich angesetzt. Infolge der geringeren Datenerhebung ist eine eindeutigere Aussage zu den Eigen- und Fremdpublikationen in Bezug auf die Reliabilität und Validität nicht möglich. Bei der Validitätsprüfung ist nur die Konstruktvalidität, aufgrund der nicht vorhandenen Untersuchungen bzw. Modifikationen beim Test d2, in Erscheinung getreten.

Des Weiteren wurde von den Fachzeitschriften eine Auswahl getroffen, um den Inhalt dieser Arbeit einzugrenzen.

Sowohl bei den Eigenpublikationen der Validitäten, als auch bei den Reliabilitäten (Odd-even-Methode und der Aufteilung der ersten und zweiten Testhälfte) beider Publikationsformen ist anzumerken, dass nur jeweils eine Untersuchung genutzt

wurde. Auch hier ist es erstrebenswert mehr Publikationen einzubinden, um die Ergebnisse von Eigen- und Fremdautoren besser gegenüberzustellen.

Es gilt auch den Vergleich der Reliabilitäten und Validitäten mit der spezifischen Verwendung kritisch zu betrachten. Hierbei wurde bei der Reliabilität nur das Anwendungsgebiet „Verkehrspsychologie“ der Eigen- und Fremdpublikationen gegenübergestellt und bei der Validität war nur der Vergleich zur „Arbeits- und Organisationspsychologie“ beider Publikationsformen möglich.

Des Weiteren ist festzustellen, dass weniger Publikationen seit den 2000er Jahren zu finden sind. Das liegt möglicherweise daran, dass der Test d2 in der Wissenschaft sowie Praxis etabliert ist und es in der Vergangenheit viele Forschungsansätze gab. Das ist vor allem beim Vergleich der Retest-Reliabilitäten zu erkennen, denn hierbei war es nur möglich mit älteren Publikationen zu arbeiten. Die jüngste Untersuchung wurde aus dem Jahr 1991 der Eigenpublikationen ermittelt. Die Eigenpublikationen der Split-Half-Methode sind dagegen jünger (Jahre: 2008 und 2012). Werden dabei die Werte der Reliabilitäten und Validitäten aus dem Manual des Test d2 von Bickenkamp et al. (2010) verglichen, sind viele Ergebnisse wesentlich jünger. Dennoch ist resümierend zu sagen, dass die meisten Ergebnisse des Manuals aus unveröffentlichten Diplomarbeiten bestehen oder die Ergebnisse aus den einzelnen Untersuchungen nicht explizit in den Fachzeitschriften zu finden waren, um daraus einen Vergleich anzustellen.²⁷⁸

²⁷⁸ Vgl. Bickenkamp; Schmidt-Atzert; Liepmann (2010), S.47 ff.

7 Zusammenfassung und Ausblick

Vor allem bei der Auswahl von Mitarbeitern in Unternehmen ist es vorteilhaft geeignete Instrumente auszusuchen, welche aussagekräftige und bearbeitbare Ergebnisse liefern.²⁷⁹ Durch die große Vielfalt an Tests, welche auf dem Markt zur Verfügung stehen, ist es für Testanwender schwierig ein geeignetes Verfahren zu finden. Die Betrachtungen des Tests d2 zeigen, dass dieser sich aufgrund seiner ökonomischen Anwendung, einfachen Handhabung und seines geringen Zeitaufwandes als beliebtes Untersuchungs- und Selektionsinstrument u.a. für Unternehmen etabliert hat. Die unterschiedlichen Anwendungsgebiete zeigen, dass der Test d2 vielseitig einsetzbar ist. Im Fokus steht besonders die „Arbeits- und Organisationspsychologie“. Der Test d2 ist in der Wissenschaft und Praxis bekannt und anerkannt, so dass er häufig als Kontrollinstrument eingesetzt wird, um neue Tests zu prüfen oder neue wissenschaftliche Erkenntnisse zu testen. Zudem können auf der Basis dieses Tests neue Forschungsansätze generiert werden. Auf Grundlage dieser Hinweise nehmen die Testautoren Modifikationen in den neusten Auflagen des Test d2 vor.²⁸⁰ Die Möglichkeit, die alten und neuen Werte gegenüberzustellen, schafft bei der Auswertung der Ergebnisse von Eigen- und Fremdpublikationen die Voraussetzung für einen besseren Vergleich.

Das Ziel dieser wissenschaftlichen Arbeit liegt in der Untersuchung von Unterschieden der Reliabilität und Validität des Test d2 – Aufmerksamkeits- und Belastungstest von Eigen- und Fremdpublikationen. Um dies zu untersuchen, werden Artikel unterschiedlicher Fachzeitschriften ausgewählt und den Eigen- sowie Fremdautoren zugeordnet. Somit sind Aussagen zur Thematik zu treffen. Aufgrund der kritischen Betrachtungsweisen von Fremdpublikationen gilt es dabei herauszufinden, ob diese tendenziell bessere Ergebnisse für die Reliabilität und Validität aufweisen als die Eigenpublikationen. Die Eigenpublikationen bieten meist den Anreiz, dass

²⁷⁹ Vgl. Kanning; Pöttker; Klinge (2008), S. 38 f.

²⁸⁰ Vgl. Brickenmap; Schmidt-Atzert; Liepmann (2010), S. 16.

diese gut dargestellt werden, so dass der Test d2 auch zur Anwendung kommt und genutzt wird. In der angegebenen untersuchten Literatur sind von 1962 bis 2015 insgesamt 124 Artikel veröffentlicht. Die meisten Publikationen sind von Fremdautoren, die den Test d2 z.B. für Untersuchungen verwenden, um die Erkenntnisse des Testautors zu prüfen, als Kontrollinstrument für andere Tests einsetzen oder als Testinstrument für die Praxis nutzen.²⁸¹ Bei den Eigenpublikationen wird der Test d2 vor allem für neue Erkenntnisse eingesetzt, aber auch zur Korrektur von Testergebnissen und zur Prüfung von Ergebnissen der Fremdautoren, die den Test kritisch sehen.²⁸² In den Artikeln ist festzustellen, dass sich der Testautor bzw. die Coautoren zu Kritiken oder Änderungsvorschlägen von Fremdautoren äußern.²⁸³

Die Ergebnisse der Untersuchung haben zum größten Teil gezeigt, dass es Unterschiede bei den Anwendungsgebieten, der Reliabilität und Validität, den durchschnittlichen Stichprobengrößen sowie bei der Betrachtung von Reliabilitäten und Validitäten in Verbindung mit der Stichprobengröße von Eigen- und Fremdpublikationen gibt.

Um einen ersten Eindruck zu bekommen, ob sich die Eigen- und Fremdpublikationen unterscheiden, werden die Häufigkeiten der unterschiedlichen Anwendungsgebiete betrachtet sowie der zeitliche Rahmen in denen der Test d2 zur Anwendung bzw. zur Untersuchung verschiedener Studien genommen wurde. Hierbei wird herausgefunden, dass sich der Ursprung des Anwendungsgebietes „Verkehrspsychologie“ des Tests d2 verändert hat. Ein weiterer Aspekt für die Untersuchung sind die Kennwerte von Reliabilitäten und Validitäten, die meistens die Fremdautoren bei den Eigenpublikationen in Frage stellen.²⁸⁴

Die Fremdpublikationen sind bei dieser Untersuchung der Unterschiede zwischen

²⁸¹ Vgl. Heinrich (1973), S. 118 ff.; Janssen (1975), S. 226 ff.; Westhoff; Kluck (1983), S. 311 ff. Hasselhorn; Körkel (1984); Oehlschlägel; Moosbrugger (1991), S. 43 ff.; Fay (1992), S. 303 ff.; Fay (1995), S. 293 ff., S. 288 ff.

²⁸² Vgl. Brickenkamp (1973), S. 125 ff.; Brickenkamp (1991), S. 52 ff.; Schmidt-Atzert; Ising (1997), S. 242 ff.; Schmidt-Atzert; Bühner; Rischen; Warkentin (2004), S. 124 ff.

²⁸³ Vgl. Brickenkamp (1991), S.52 ff.; Schmidt-Atzert; Bühner (1997), S. 314 ff.; Schmidt- Atzert; Ising (1997), S. 241 ff.; Schmidt-Atzert; Bühner; Rischen; Warkentin (2004), S. 126 ff. Schmidt-Atzert; Bühner; Enders (2006), S. 33 ff.; Krumm; Schmidt-Atzert; Bracht; Ochs (2011), S. 117 ff.

²⁸⁴ Vgl. Fay (1995), S. 293 ff.; Daseking; Putz (2015), S. 324.

den Eigen- und Fremdpublikationen dominierend aufgetreten. Bei den Unterschieden der Reliabilitätsuntersuchungen hat sich herausgestellt, dass die Eigenpublikationen reliabler sind als die Fremdpublikationen. Im Gegensatz dazu werden bei den Validitätsuntersuchungen mehr Fremdpublikationen als Eigenpublikationen ermittelt. Dadurch wird festgestellt, dass die Artikel der Fremdautoren vor allem in der konvergenten Validität höhere Ergebnisse aufzeigen. Die größere Anzahl der Fremdpublikationen spricht für den Test, da Fremdautoren neue Anregungen bzw. Verbesserungsvorschläge für den Test d2 einbringen. Das zeigt sich ebenfalls im Manual des Tests d2 – Revision von Brickenkamp et al. (2010), denn darin sind überwiegend Untersuchungen der Reliabilität und Validität von Fremdautoren enthalten.²⁸⁵

Es wäre zu begrüßen, dass sich weitere Studien mit dieser Thematik auseinandersetzen, welche Bezug auf mehr und aktuellere Publikationen nehmen. Daraus ließen sich aussagekräftigere Vergleiche zwischen Eigen- und Fremdpublikationen hinsichtlich der Reliabilität und Validität ableiten.

²⁸⁵ Vgl. Brickenkamp; Schmidt-Atzert; Liepmann (2010), S. 47 ff.

Literaturverzeichnis

- Becker, N.; Preckel, F.; Karbach, J.; Raffel, N.; Spinath, F. M. (2015): Die Matrizenkonstruktionsaufgabe, in: *Diagnostica*, 61 (1), S. 22-33.
- Boomsma, A. (1985): Nonconvergence, Improper Solutions, and Starting Values in Lisrel Maximum Likelihood Estimation, in: *Psychometrika*, 50 (2), S. 229-242.
- Brickenkamp, R. (1962): Test d2. Aufmerksamkeits- und Belastungs-Test, 1. Auflage - Göttingen: Hogrefe 1962.
- Brickenkamp, R. (1967): Test d2. Aufmerksamkeits- und Belastungs-Test, 2., erweiterte und verbesserte Auflage - Göttingen: Hogrefe 1967.
- Brickenkamp, R. (1970): Test d2. Aufmerksamkeits- und Belastungs-Test, 3. Auflage - Göttingen: Hogrefe 1970.
- Brickenkamp, R. (1972): Test d2. Aufmerksamkeits- und Belastungs-Test, 4. Auflage - Göttingen: Hogrefe 1972.
- Brickenkamp, R. (1973): Anmerkungen zur Interpretierbarkeit von Leistungsschwankungen im Test d2. Eine Stellungnahme zu Bemerkungen von Hans Ch. Heinrich, in: *Diagnostica*, 19 (2), S. 125-131.
- Brickenkamp, R. (1975): Test d2. Aufmerksamkeits- und Belastungs-Test, 5., erweiterte Auflage - Göttingen: Hogrefe 1975.
- Brickenkamp, R. (1978): Test d2. Aufmerksamkeits- und Belastungs-Test, 6. Auflage - Göttingen: Hogrefe 1978.
- Brickenkamp, R. (1981): Test d2. Aufmerksamkeits- und Belastungs-Test, 7. Auflage - Göttingen: Hogrefe 1981.
- Brickenkamp, R. (1991): Fehlinterpretation von Testleistungen? Anmerkungen zum Beitrag „Konzentrationsleistung ohne Konzentration?“, in: *Diagnostica*, 37 (1), S. 52-57.
- Brickenkamp, R. (1994): Test d2. Aufmerksamkeits- und Belastungs-Test, 8., erweiterte und neu gestaltete Auflage - Göttingen: Hogrefe 1994.
- Brickenkamp, R. (2002): d2. Aufmerksamkeits-Belastungs- Test, in: Sarges, W.; Wottawa, H. (Hrsg.): *Handbuch wirtschaftspsychologischer Testverfahren*, Bd. 1 Personalpsychologische Instrumente, S. 265-270. 2., überarbeitete Auflage - Lengerich: Pabst Science Publishers 2004.

- Brickenkamp, R. (2002): Test d2. Aufmerksamkeits- und Belastungs-Test, 9., vollständig überarbeitete und neunormierte Auflage - Göttingen: Hogrefe 2002.
- Brickenkamp, R.; Karl, G. (1986): Geräte zur Messung von Aufmerksamkeit, Konzentration und Vigilanz, in: Brickenkamp, R. (Hrsg.): Handbuch apparativer Verfahren in der Psychologie, S. 195-211. Göttingen: Hogrefe, 1986.
- Brickenkamp, R.; Rump, G. (1966): Die Stabilität des Aufmerksamkeits-Belastungs-Tests (Test d2) über längere Zeitabschnitte, in: Diagnostica, 12 (1), S. 17-24.
- Brickenkamp, R.; Schmidt-Atzert, L.; Liepmann, D. (2010): Test d2-Revision. Aufmerksamkeits- und Konzentrationstest - Göttingen: Hogrefe, 2010.
- Bühner, M. (2005): Gütekriterien von Verhaltensbeobachtungen, in: Westhoff, K.; Hellfrisch, L. J.; Hornke, L. F.; Kubinger, K. D.; Lang, F.; Moosbrugger, H., Püschel, A.; Reimann, G. (Hrsg.): Grundwissen für die berufsbezogene Eignungsbeurteilung nach DIN 33430, S. 81-92. 2., überarbeitete Auflage - Lengerich: Pabst Science Publishers 2005.
- Bühner, M. (2011). Einführung in die Test- und Fragebogenkonstruktion, 3., aktualisierte Auflage - München: Pearson Studium.
- Daseking, M.; Putz, D. (2015): TBS-TK Rezension: „Test d2-R. Aufmerksamkeits- und Konzentrationstest“, in: Report Psychologie 40 (7/8), S. 323-324.
- Dingel, W. (1971): Mitteilung über Zuverlässigkeitsaspekte des Test d2, in: Diagnostica, 17, S. 84-87.
- Eser, K. H. (1991): Zur prognostischen Validität des Aufmerksamkeits- Belastungs-Tests (Test d2) bei lernbehinderten und verhaltensauffälligen Jugendlichen in der Berufsvorbereitung, in: Zeitschrift für Arbeits- und Organisationspsychologie A&O, 35 (3), S. 137 -140.
- Eser, K.-H. (1987): Reliabilitäts- und Validitätsaspekte des Aufmerksamkeits- und Belastungs-Tests (Test d 2) bei verhaltensgestörten Kindern und Jugendlichen, in: Diagnostica, 33 (1), S. 74-80.
- Fay, E. (1992): Über die Übbarkeit der Leistung in einem Durchstreichverfahren zur Messung der Konzentrationsfähigkeit, in: Diagnostica, 38 (4), S. 301-311.
- Fay, E. (1995): Aufmerksamkeits-Belastungs-Test (d2), in: Diagnostica, 41 (3), S. 291-296.
- Fisseni, H.-J. (2004): Lehrbuch der psychischen Diagnostik, 3., überarbeitet und erweiterte Auflage, Göttingen: Hogrefe 2004.
- Funsch, K.; Arias Martín, B.; Halder-Sinn, P. (2011): Differentieller Konzentrationstest für Kinder (DKT-K): Entwicklung und empirische Überprüfung eines computergestützten Verfahrens zur Erfassung der längerfristigen Konzentrations-

- fähigkeit, in: *Diagnostica*, 57 (1), S. 2-16.
- Goldhammer, F.; Mossbrugger, H. (2006): Aufmerksamkeit, in: Schweizer, K. (Hrsg.): *Leistung und Leistungsdiagnostik*, S. 16-32. Heidelberg: Springer 2006.
- Grubitzsch, S.; Rexillus, G. (1978): *Testtheorie – Testpraxis*. Reinbek bei Hamburg: Rowohlt 1978.
- Hagemeister, C.; Westhoff, K. (1998): FAIR Frankfurter Aufmerksamkeits-Inventar, in: *Diagnostica*, 44 (4), S. 225-229.
- Hartig, J.; Frey, A.; Jude, N. (2012): Validität, in: Moosbrugger, H.; Kelava, A. (Hrsg.): *Testtheorie und Fragebogenkonstruktion*, S. 143-171. 2., aktualisierte Auflage - Heidelberg: Springer 2012.
- Härtler, G. (2014): *Statistisch gesichert und trotzdem falsch? – Vom (Un-)Wesen statistischer Schlüsse* - Berlin, Heidelberg: Springer 2014.
- Hasselhorn, M; Körkel, J. (1984): Zur differentiellen Bedeutung metakognitiver Komponenten für das Verstehen und das Behalten von Texten, in: *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 16 (4), S. 283-296.
- Heinrich, H.Ch. (1973): Einige Anmerkungen zum d2-Durchstreichtest nach Brickenkamp, in: *Diagnostica*, 19 (2), S. 118-124.
- Heubrock, D.; Petermann, F. (2001): *Aufmerksamkeitsdiagnostik, Kompendien Psychologische Diagnostik*, Bd. 2, Göttingen: Hogrefe 2001.
- Himme, A. (2007): Gütekriterien der Messung: Reliabilität, Validität und Generalisierbarkeit, in: Albers, S.; Klapper, D.; Konradt, U.; Walter, A.; Wolf, J. (Hrsg.): *Methodik der empirischen Forschung*, S. 375-390, 2. Auflage, Wiesbaden: Springer 2007.
- Homburg, C.; Baumgartner, H. (1995): Beurteilung von Kausalmodellen. Bestandsaufnahme und Anwendungsempfehlungen, in: *Marketing: Zeitschrift für Forschung und Praxis*, 17 (3), 162-176.
- Homburg, C.; Klarmann, M. (2006): Die Kausalanalyse in der empirischen betriebswirtschaftlichen Forschung: Problemfelder und Anwendungsempfehlungen, in: *Die Betriebswirtschaft*, 66 (6), S. 727-748.
- Hornke, L. F. (2004): Normen, Standards, Richtlinien auch für die Personalarbeit, in: Hornke, L. F.; Winterfeld, U. (Hrsg.): *Eignungsbeurteilung auf dem Prüfstand: DIN 33430 zur Qualitätssicherung*, S. 9-25. Heidelberg: Spektrum Akademischer Verlag 2004.

- Hu, L.-T.; Bentler, P. M. (1998): Fit Indices in Covariance Structure Modeling: Sensitivity to Underparametrized Model Misspecification, in: *Psychological Methods*, 3 (4), S. 424–453.
- Janssen, J. P. (1975): Motivation Analysis Test (MAT) ein nicht verfälschbarer Test?, in: *Zeitschrift für experimentelle und angewandte Psychologie*, 22 (2), S. 218-240.
- Kanning, U. P. (2002): Grundlagen psychologischer Diagnostik, in: Holling, H.; Kanning, U. P. (Hrsg.): *Handbuch personaldiagnostischer Instrumente*, S. 47-92, Göttingen: Hogrefe 2002.
- Kanning, U. P.; Pöttker, J.; Klinge, K. (2008): *Personalauswahl: Leitfaden für die Praxis*, Stuttgart: Schäffer-Poeschel 2008.
- Kersting, M. (2004): Kosten und Nutzen beruflicher Eignungsbeurteilungen, in: Hornke, L. F.; Winterfeld, U. (Hrsg.): *Eignungsbeurteilung auf dem Prüfstand: DIN 33430 zur Qualitätssicherung*, S. 55-77, Heidelberg: Spektrum Akademischer Verlag 2004.
- Kersting, M. (2011): Schlaraffenland ist abgebrannt – warum sich die Bewerberauswahl ändern muss, in: *Wirtschaftspsychologie aktuell*, 2011 (4), S. 23-26.
- Krumm, S.; Schmidt-Atzert, L.; Bracht, M.; Ochs, L. (2011): Coordination as a Crucial Component of Performance on a Sustained Attention Test - Evidence From the Digit-Symbol Substitution Test, in: *Zeitschrift für Differentielle und Diagnostische Psychologie*, 32 (3), S. 117-128.
- Krumm, S.; Schmidt-Atzert, L.; Michalczyk, K. M.; Danthiir, V. (2008): Speeded Paper-Pencil Sustained Attention and Mental Speed Tests - Can Performances Be Discriminated?, in: *Zeitschrift für Differentielle und Diagnostische Psychologie*, 29 (4), S. 205-216.
- Krumm, S.; Schmidt-Atzert, L.; Schmidt, S.; Zenses, E.-M.; Stenzel, M. (2012): Attention Tests in Different Stimulus Presentation Modes - A Facet Model of Performance in Attention Tests, in: *Zeitschrift für Differentielle und Diagnostische Psychologie*, 33 (3), S. 146-159.
- Kubinger, K.D. (2003). Gütekriterien. in: Kubinger, K. D.; Jäger, R.S. (Hrsg.): *Schlüsselbegriffe der Psychologischen Diagnostik*, S. 195-204, Weinheim: Beltz PVU 2003.
- Kühnlein, I.; Hahlweg, K.; Müller, G.; Sevenitz, B. (1983): Überprüfung der SPM von Raven für den psychiatrischen Bereich und Entwicklungen von Kurzformen, in: *Diagnostica*, 29 (3), S. 236-246.
- Leikin, R.; Paz-Baruch, N.; Leikin, M. (2014): Cognitive Characteristics of Students with Superior Performance in Mathematics, in: *Zeitschrift für Differentielle und Diagnostische Psychologie*, 35 (3), S. 119 - 129.

- Leitner, W.; Ortner, R. (2004): Aufmerksamkeit und Konzentration im Blickfeld von Bildung und Erziehung, in: Korte, P. (Hrsg.): Kontinuität, Krise und Zukunft der Bildung: Analyse und Perspektiven, S. 123-134. Münster: Lit 2004.
- Marschner, G. (1971): Revisions-Test (Rev.T.) als allgemeiner Leistungstest. Beiträge zur Standardisierung und Validierung des Revisions-Testes nach B. Stender, in: Diagnostica, 17 (1), S. 14-26.
- Merten, T. (2000): Die Computerversion des d2 und die Frage der Transferäquivalenz, in: Psychologische Beiträge; Vierteljahresschrift für alle Gebiete der Psychologie 42 (4), S. 572-789.
- Moosbrugger, H. (2012): Klassische Testtheorie (KKT), in: Moosbrugger, H.; Kelava, A. (Hrsg.): Testtheorie und Fragebogenkonstruktion, S. 103-117. 2., aktualisierte Auflage – Heidelberg: Springer 2012.
- Moosbrugger, H.; Goldhammer, F. (2004): FAKT-II. Frankfurter Adaptiver Konzentrationsleistungs-Test II, in: Sarges, W.; Wottawa, H. (Hrsg.): Handbuch wirtschaftspsychologischer Testverfahren, Bd. 1 Personalpsychologische Instrumente, S. 331-335. 2., überarbeitete Auflage - Lengerich: Pabst Science Publishers 2004.
- Moosbrugger, H.; Goldhammer, F. (2006): Aufmerksamkeits- und Konzentrationsdiagnostik, in: Schweizer, K. (Hrsg.): Leistung und Leistungsdiagnostik, S. 83-102. Heidelberg: Springer 2006.
- Moosbrugger, H.; Kelava, A. (2012): Qualitätsanforderungen an einen psychologischen Test (Testgütekriterien), in: Moosbrugger, H.; Kelava, A. (Hrsg.): Testtheorie und Fragebogenkonstruktion, S. 7-26. 2., aktualisierte Auflage – Heidelberg: Springer 2012.
- Moosbrugger, H.; Oehlschlägel, J. (1996): FAIR. Frankfurter Aufmerksamkeits-Inventar, in: Sarges, W.; Wottawa, H. (Hrsg.): Handbuch wirtschaftspsychologischer Testverfahren, Bd. 1 Personalpsychologische Instrumente, S. 325-329. 2., überarbeitete Auflage - Lengerich: Pabst Science Publishers 2004.
- Moosbrugger, H.; Reiß, S. (2004): Das Frankfurter Aufmerksamkeits-Inventar FAIR, in: Büttner G.; Schmidt-Atzert, L. (Hrsg.): Diagnostik von Konzentration und Aufmerksamkeit, Jahrbuch der pädagogisch-psychologischen Diagnostik, Tests und Trends, Neue Folge Bd. 3, S. 103-118. Göttingen: Hogrefe 2004.
- Oehlschlägel, J.; Moosbrugger, H. (1991): Konzentrationsleistung ohne Konzentration? Zur Schätzung wahrer Leistungswerte im Aufmerksamkeits-Belastungs-Test d2, in: Diagnostica, 37 (1), S. 42-51.
- Oswald, W. D.; Hagen, B. (1997): Test d2 – Aufmerksamkeits- und Belastungs-Test, in: Zeitschrift für Differentielle und Diagnostische Psychologie, 18 (1/2), S. 87-89.

- Posner, M. I.; Boies, S. J. (1971): Components of attention, in: *Psychological Review*, 78 (5), S. 391 – 408.
- Rapp, G. (1982): *Aufmerksamkeit und Konzentration. Erklärungsmodelle, Störungen, Handlungsmöglichkeiten*. Bad Heilbrunn.: Klinkhardt 1982.
- Rauchfleisch, U. (1983): Zur Reliabilität und Validität des Aufmerksamkeits- Belastungs-Tests (Test d2) bei Patienten mit hirndiffusen Psychosyndrom und neurotischen Störungen, in: *Diagnostica*, 29 (3), S. 247-255.
- Reimann, G. (2005): Einführung in die DIN 33430, in: Westhoff, K.; Hellfritsch, L. J.; Hornke, L. F.; Kubinger, K. D.; Lang, F.; Moosbrugger, H., Püschel, A.; Reimann, G. (Hrsg.): *Grundwissen für die berufsbezogene Eignungsbeurteilung nach DIN 33430*, S. 17-38. 2., überarbeitete Auflage – Lengerich: Pabst Science Publishers 2005.
- Rumsey, D. (2015): *Statistik für Dummies*, 3., aktualisierte Auflage – Weinheim: Wiley-VCH 2015.
- Rumsey, D. (2008): *Weiterführende Statistik für Dummies*, 1. Auflage – Weinheim: Wiley-VCH 2008.
- Sarges, W.; Wottatwa, H. (2004): Einführung, in: Sarges, W.; Wottawa, H. (Hrsg.): *Handbuch wirtschaftspsychologischer Testverfahren*, Bd. 1 Personalpsychologische Instrumente, Einführung S. VII-XII. 2., überarbeitete Auflage – Lengerich: Pabst Science Publishers 2004.
- Schermelleh-Engel, K.; Schweizer, K. (2012): Multitrait-Multimethod-Analysen, in: Moosburger, H.; Kelava, A. (Hrsg.): *Testtheorie und Fragebogenkonstruktion*, S. 345-362. 2., aktualisierte Auflage - Heidelberg: Springer 2012.
- Schermelleh-Engel, K.; Werner, C. S. (2012): Methoden der Reliabilitätsbestimmung, in: Moosburger, H.; Kelava, A. (Hrsg.): *Testtheorie und Fragebogenkonstruktion*, S. 120-141. 2., aktualisierte Auflage - Heidelberg: Springer 2012.
- Schmidt-Atzert, L.; Amelang, M. (2012): *Psychologische Diagnostik*, 5., überarbeitete Auflage - Heidelberg: Springer 2012.
- Schmidt-Atzert, L.; Bühner, M.; Rischen, S.; Warkentin, V. (2004): Erkennen von Simulation und Dissimulation im Test d2, in: *Diagnostica*, 50 (3), S. 124-133.
- Schmidt-Atzert, L.; Büttner, G.; Bühner, M. (2004): Theoretische Aspekte von Aufmerksamkeits/Konzentrationsdiagnostik, in: Büttner, G.; Schmidt-Atzert, L. (Hrsg.): *Diagnostik von Konzentration und Aufmerksamkeit*, Jahrbuch der pädagogisch-psychologischen Diagnostik, Tests und Trends, Neue Folge Bd. 3, S. 3-23. Göttingen: Hogrefe 2004.
- Schmidt-Atzert, L.; Ising, M. (1997): Ein Beitrag zur Konstruktvalidität von d2 und

- Revisionstest, in: Zeitschrift für Differentielle und Diagnostische Psychologie, 18 (4), S. 241-250.
- Schmidt-Atzert, L.; Krumm, S.; Bühner, M. (2008): Aufmerksamkeitsdiagnostik: Ableitung eines Strukturmodells und systematische Einordnung von Tests, in: Zeitschrift für Neuropsychologie, 19 (2), S. 59-82.
- Schmidt, J. U. (1994): Zur Validität von Schulkenntnistests, in: Zeitschrift für Arbeits- und Organisationspsychologie A&O, 38 (3), S. 112-118.
- Schorr, A. (1995): Stand und Perspektiven diagnostischer Verfahren in der Praxis. Ergebnisse einer repräsentativen Befragung westdeutscher Psychologen, in: Diagnostica, 41 (1), S. 3-20.
- Schweizer, K. (2005): An overview of research into the cognitive basis of intelligence, in: Journal of Individual Differences, 26 (1), S. 43-51.
- Sommer, G. (1973): Die Problematik der Erfassung von „Konzentration“, dargestellt am KLT, in: Diagnostica, 19 (1), S. 62-75.
- Steck, P. (1997): Psychologische Testverfahren in der Praxis: Ergebnisse einer Umfrage unter Testanwendern, in: Diagnostica, 43 (3), S. 267 – 284.
- Steffens, M. C. (2004): Is the Implicit Association Test Immune to Faking?, in: Zeitschrift für experimentelle Psychologie, 51 (3), S. 165-179.
- Weis, S.; Nuerk, H.-C. (2010). TBS-TK Rezension: FAKT-II. Frankfurter Adaptiver Konzentrationsleistungs-Test, in: Report Psychologie, 36 (5), S. 219-221.
- Westhoff, K.; Graubner, J. (2003): Konstruktion eines Komplexen Konzentrations-tests, in: Diagnostica, 49 (3), S. 110-119.
- Westhoff, K.; Hagemeyer, C. (2005): Konzentrationsdiagnostik, Lengerich: Pabst Science Publishers 2005.
- Westhoff, K.; Kluck, M.-L. (1983): Zusammenhang zwischen Intelligenz und Konzentration, in: Diagnostica, 29 (4), S. 310-319.
- Westhoff, K.; Kluck, M.-L. (1984): Ansätze einer Theorie konzentrativer Leistungen, in: Diagnostica, 30 (3), S. 167-183.
- Wiese, W.; Kroj, G. (1972): Untersuchung über den Zusammenhang zwischen Intelligenz (Wechsler) und Konzentrationsfähigkeit (Test d2 nach Brickenkamp), in: Zeitschrift für experimentelle und angewandte Psychologie, 19 (4), S. 690-699.
- Zaworka, W.; Hand, I. (1981): Die "Anankastische Persönlichkeit" - Fakt oder Fiktion?. Experimentelle Diagnostik der Zwangsneurose, in: Zeitschrift für Diagnostische und Differentielle Psychologie, 2 (1), S. 31-54.

Anlagenverzeichnis

Teil 1 – Zuordnung der Artikel

Tabelle A-1: Liste der Test d2 Literatur (Quelle: Eigene Darstellung)

Teil 2 – Zuordnung der Artikel in Eigen- und Fremdpublikationen mit den dazugehörigen Anwendungsgebieten

Tabelle A- 2: Einordnung der Artikel in den unterschiedlichen Anwendungsgebieten, Jahren und für die Verwendung der Stichprobengrößen der Eigenpublikationen (Quelle: Eigene Darstellung)

Tabelle A-3: Einordnung der Artikel in den unterschiedlichen Anwendungsgebieten, Jahren und für die Verwendung der Stichprobengrößen der Fremdpublikationen (Quelle: Eigene Darstellung)

Teil 3 – Tabellen der Reliabilitätsanalyse

Tabelle A-4: Reliabilitätsschätzungen nach Aufteilung in die erste und zweite Testhälfte bei den Eigenpublikationen (Quelle: Eigene Darstellung)

Tabelle A-5: Reliabilitätsschätzungen nach der Odd-even-Methode bei den Eigenpublikationen (Quelle: Eigene Darstellung)

Tabelle A-6: Reliabilitätsschätzungen nach Aufteilung in die erste und zweite Testhälfte bei den Fremdpublikationen (Quelle: Eigene Darstellung)

Tabelle A-7: Reliabilitätsschätzungen nach der Odd-even-Methode bei den Fremdpublikationen (Quelle: Eigene Darstellung)

Teil 4- Durchschnittliche Stichprobengrößen in Eigen- und Fremdpublikationen

Tabelle A-8: Durchschnittliche Stichprobengrößen in den spezifischen Verwendungen der Eigenpublikationen (Quelle: Eigene Darstellung)

Tabelle A-9: Durchschnittliche Stichprobengrößen in den spezifischen Verwendungen der Fremdpublikationen (Quelle: Eigene Darstellung)

Tabelle A-1: Liste der Test d2 Literatur (Quelle: Eigene Darstellung)

Zeitschrift	Jahr	Titel	Autor	Eigenpublikation	Fremdpublikation
Diagnostica	1966	Die Stabilität des Aufmerksamkeits-Belastungs-Tests (Test d2) über längere Zeitabschnitte	Brickenkamp, R.; Rump, G.	x	
Diagnostica	1971	der Revisions-Test als allgemeiner Leistungstest	Marschner, G.		x
Diagnostica	1971	Mitteilung über Zuverlässigkeitsaspekte des Test d2	Dingel, W.		x
Diagnostica	1973	Die Problematik der Erfassung von "Konzentration", dargestellt am KLT	Sommer, G.		x
Diagnostica	1973	Einige Anmerkungen zum d2-Durchstreichtest nach Brickenkamp	Heinrich, H. Ch.		x
Diagnostica	1973	Anmerkungen zur Interpretierbarkeit von Leistungsschwankungen im Test d2 - Eine Stellungnahme zu Bemerkungen von Hannes Ch. Heinrich	Brickenkamp, R.	x	
Diagnostica	1978	Standardisierung einer Skala zur Erfassung der sensorischen Suggestibilität (indirekte Variante) an Schülern der Sekundarstufe II	Gheorghin, V. A.; Grimm, H.-J.; Hodapp, V.		x
Diagnostica	1979	Validierung einer Testbatterie zur Erfassung hirnorganischer Schädigungen	Hahlweg, K.		x
Diagnostica	1979	Zur diagnostischen Relevanz des Zahlen-Verbindungstests (ZVT)	Kallenbach, K.; Oswald, W. D.		x

Diagnostica	1981	Die Bewährung des Benton-Test in der psychiatrischen Klinik; I. Aufgabenanalyse und Bestimmung der inneren Konsistenz; II. Konstruktvalidität und differentielle Gültigkeit	Hahlweg, K.; Kühnlein, I.		x
Diagnostica	1981	Entwicklung und Erprobung eines neuartigen Aufgabentyps zur Erfassung des räumlichen Vorstellungsvermögens	Stumpf, H.; Fay, E.		x
Diagnostica	1983	Überprüfung der SPM von Raven für den psychiatrischen Bereich und Entwicklungen von Kurzformen	Kühnlein, I.; Hahlweg, K.; Müller, G.; Sevenitz, B.		x
Diagnostica	1983	Zur Reliabilität und Validität des Aufmerksamkeits-Belastungs-Tests (Test d2) bei Patienten mit hirndiffusen Psychosyndrom und neurotischen Störungen	Rauchfleisch, U.		x
Diagnostica	1983	Zusammenhang zwischen Intelligenz und Konzentration	Westhoff, K.; Kluck, M.-L.		x
Diagnostica	1984	Ansätze einer Theorie konzentrativer Leistungen	Westhoff, K.; Kluck, M.-L.		x
Diagnostica	1985	Kreuzvalidierung eines Meßverfahrens zur Erfassung von Kontrolllängsten paranoid-schizophrener Patienten in psychologischen Untersuchungen	Sorgatz, H.; Barth, G.		x
Diagnostica	1985	Eine erste Prüfung einer Konzentrationstheorie	Westhoff, K.		x
Diagnostica	1987	Reliabilitäts- und Validitätsaspekte des Aufmerksamkeits-Belastungs-Tests	Eser, K.-H.		x

		(Test d2) bei verhaltensgestörten Kindern und Jugendlichen			
Diagnostica	1987	Neue Befunde zur Reliabilität, Validität und Normierung der "Schlauchfiguren"	Stumpf, H.; Fay, E.		x
Diagnostica	1988	Eine erweiterte Prüfung einer Konzentrationstheorie	Westhoff, K.; Lemme, M.		x
Diagnostica	1989	Übungsabhängigkeit von Leistungen in Konzentrationstests	Westhoff, K.		x
Diagnostica	1990	Effekte der Übung in der Bearbeitung von Konzentrationstests	Westhoff, K.; Dewald, D.		x
Diagnostica	1990	Computergestützte Leistungsdiagnostik bei verkehrspsychologischen Eignungsuntersuchungen	Bukasa, B.; Kisser, R.; Wenniger, U.		x
Diagnostica	1991	Konzentrationsleistung ohne Konzentration? Zur Schätzung wahrer Leistungswerte im Aufmerksamkeits-Belastungs-Test d2	Oehlschlägel, J.; Moos- burgger, H.		x
Diagnostica	1991	Fehlinterpretationen von Testleistungen? Anmerkung zum Beitrag "Konzentrationsleistung ohne Konzentration?"	Brickenkamp, R.	x	
Diagnostica	1992	Reliabilität von Fehlern in Konzentrationstests	Westhoff, K.; Hagemeister, C.		x
Diagnostica	1992	Aus der Arbeit des Testkoloratoriums Alters-Konzentrations-Test (AKT)	Westhoff, K.	-	-

Diagnostica	1992	Über die Übbarkeit der Leistung in einem Durchstreichverfahren zur Messung der Konzentrationsfähigkeit	Fay, E.		x
Diagnostica	1993	Der Allgemeine Büroarbeitstest (ABAT)- mehr als ein Bürottest?	Schmidt, J. U.	-	-
Diagnostica	1995	Stand und Perspektiven diagnostischer Verfahren in der Praxis. Ergebnisse einer repräsentativen Befragung westdeutscher Psychologen	Schorr, A.		x
Diagnostica	1995	Aufmerksamkeits-Belastungs-Test (d2)	Fay, E.		x
Diagnostica	1996	Die Prüfung der Dauerkonzentration mit einer Apparateversion des Pauli-Tests	Steck, P.		x
Diagnostica	1997	Aus der Arbeit des Testkoloratoriums Psychologische Testverfahren in der Praxis Ergebnisse einer Umfrage unter Testanwendern	Steck, P.		x
Diagnostica	1997	Komponenten von Konzentrationsleistungen: Eine Untersuchung mit systematisch variierten Durchstreich-Varianten	Schmidt-Atzert, L.; Bühner, M.	x	
Diagnostica	1998	Fehlertypen im Aufmerksamkeits-Belastungs-Test d2	Schmidt-Atzert, L.; Bühner, M.	x	
Diagnostica	1998	Testinformation FAIR Frankfurter Aufmerksamkeits-Inventar	Moosburger, H.; Oehlschlägel, J.	-	-
Diagnostica	1999	Testinformation	Rollett, B.	-	-

Diagnostica	1999	Normierungsstudie zum verbalen Lern- und Merkfähigkeitstest (VLMT)	Lux, S.; Helmstaedter, C.; Eger, C. E.	-	-
Diagnostica	2003	Konstruktion eines Konzentrationstests	Westhoff, K.; Graubner, J.		x
Diagnostica	2004	Erkennen von Simulation und Dissimulation im Test d2	Schmidt-Atzert, L.; Bühner, M.; Rischen, S.; Warkentin, V.	x	
Diagnostica	2006	Messen Konzentrationstests Konzentration?	Schmidt-Atzert, L.; Bühner, M.; Enders, P.	x	
Diagnostica	2011	Differenzieller Konzentrationstest für Kinder (DKT-K)	Funsch, K.; Arias Martin, B.; Halder-Sinn, P.		x
Diagnostica	2015	Die Matrizenkonstruktionsaufgabe	Becker, N.; Preckel, F.; Karbach, J.; Raffel, N.; Spinath, F.M.		x
Zeitschrift für Arbeits- und Organisationspsychologie	1962	Experimentelle Untersuchung zur Beeinflussbarkeit des Reaktionsverhaltens	Brickenkamp, R.	x	
Zeitschrift für Arbeits- und Organisationspsychologie	1964	Orientierungs-Reaktionsprobe nach Poppelreuter - Untersuchungen zur Validität	Kunkel, E.		x
Zeitschrift für Arbeits- und Organisationspsychologie	1966	Validitätsbestimmung durch Gruppenvergleiche bei Fahrtauglichkeitsuntersuchungen	Kunkel, E.		x

Zeitschrift für Arbeits- und Organisationspsychologie	1969	Weitere Untersuchungen zum Zweifaktorenmodell der Leistungsmotivation	Bäumler, G.; Dvorak, H.-P.		x
Zeitschrift für Arbeits- und Organisationspsychologie	1972	Der Intelligenzabbau nach WECHSLER in Bezug zur psychisch-funktionalen Leistungsfähigkeit- geprüft am Wiener Determinationsgerät und im Aufmerksamkeits-Belastungs-Test d2-, zum Alter und zur Alkoholdelinquenz im Straßenverkehr	Kroj, G.		x
Zeitschrift für Arbeits- und Organisationspsychologie	1974	Lehraussagen zur Konzentrationsfähigkeit bei sechsjährigen Kindern	Kleber, E. W.		x
Zeitschrift für Arbeits- und Organisationspsychologie	1975	Faktorenanalytischer Vergleich gebräuchlicher deutschsprachiger Intelligenztests	Steck, P.; Rausche, A.; Steege, F.W.		x
Zeitschrift für Arbeits- und Organisationspsychologie	1976	Sequenzielle Entschädigungsstrategie zur psychologischen Diagnose	Eich, F.; Klug, J.; Wetzel, H.		x
Zeitschrift für Arbeits- und Organisationspsychologie	1979	Der Einfluss von Variablentransformationen nach McCall auf die Faktorenstruktur von Intelligenz- und Leistungsvariablen	Fay, E.; Prystav, G.		x
Zeitschrift für Arbeits- und Organisationspsychologie	1980	Die Wirkung sogenannter depressiver Denkhemmungen auf die Leistung in Intelligenztests	Steck, P.		x

Zeitschrift für Arbeits- und Organisationspsychologie	1980	Zur Validität und Reliabilität des KFP-30, der FPI-Skalen "Aggressivität, Erregbarkeit und Dominanz" in einer echten Prüfungssituation	Janssen, J. P.; Reitzner, Chr.		x
Zeitschrift für Arbeits- und Organisationspsychologie	1980	Entwicklung und Ersterprobung eines maschinenlesbaren Konzentrationstests im Rahmen der "Test für medizinische Studiengänge" (TMS)	Fay, E.; Stumpf, H.		x
Zeitschrift für Arbeits- und Organisationspsychologie	1985	Leistungsniveau und Leistungsverlauf im Daueraufmerksamkeitstest als Eignungskriterien bei Fahr-, Steuer- und Überwachungstätigkeiten	Brunner, A.; Dvorak, H.		x
Zeitschrift für Arbeits- und Organisationspsychologie	1988	Über die praktischen Auswirkungen eines Wechsels der diagnostischen Entscheidungsgrundlage	Fay, E.; Klieme, E.		x
Zeitschrift für Arbeits- und Organisationspsychologie	1990	Personenauswahl aus der Sicht der Bewerber: Zum Erleben eignungsdiagnostischer Situationen	Schuler, H.		x
Zeitschrift für Arbeits- und Organisationspsychologie	1991	Zur prognostischen Validität des Aufmerksamkeits- Belastungs-Tests (Test d2) bei lernbehinderten und verhaltensauffälligen Jugendlichen in der Berufsvorbereitung	Eser, K.-H.		x
Zeitschrift für Arbeits- und Organisationspsychologie	1993	Die Vorhersage des Ausbildungserfolgs bei verschiedenen Berufsgruppen durch Leitungstests	Schmidt-Atzert, L.; Deter, B.	x	

Zeitschrift für Arbeits- und Organisationspsychologie	1994	Zur Validität von Schulkenntnistests	Schmidt, J. U.		x
Zeitschrift für Arbeits- und Organisationspsychologie	1999	Ambulante Datenerhebung in der Arbeitspsychologie - der Nutzen von PC-Tagebüchern für die Erfassung von Arbeitsbelastung, Erleben und Verhalten	Rau, R.; Triemer, A.		x
Zeitschrift für Arbeits- und Organisationspsychologie	2000	Rezension des "Intelligenz-Struktur-Test 2000" von Amthauer, R.; Brocke, B.; Liepmann, D.; Beauducel, A.	Kersting, M.		x
Zeitschrift für Personalpsychologie	2002	Verbessert Vorerfahrung die Leistung im Assessment Center?	Kelbetz, G.; Schuler, H.		x
Zeitschrift für Personalpsychologie	2002	Tests und Tools. Intelligenz-Struktur-Test 2000 R	Schmidt-Atzert, L.	x	
Zeitschrift für Personalpsychologie	2002	Wie kann man Geübtheit in Konzentrationstests erkennbar machen?	Hagemeister, C.; Scholz, A.; Westhoff, K.		x
Zeitschrift für Personalpsychologie	2004	Prädikation von Ausbildungserfolg: Allgemeine Intelligenz (g) oder spezifische kognitive Fähigkeiten?	Schmidt-Atzert, L.; Deter, B.; Jaeckel, S.;	x	
Zeitschrift für Personalpsychologie	2005	Tests und Tools - Der Objektive Leistungsmotivations-Test (OLMT)	Schmidt-Atzert, L.; Brandstätter, V.	x	
Zeitschrift für Personalpsychologie	2006	Arbeitsprobe zur berufsbezogenen Intelligenz - Büro- und kaufmännische Tätigkeiten (AZUBI-BK)	Fisseni, H.-J-		x

Zeitschrift für Personalpsychologie	2012	Dual Goals for Speed and Accuracy on the same Performance Task - Can they prevent Speed-Accuracy Trade-offs?	Wetter, O. E.; Wegge, J.; Jonas, K.; Schmidt, K.-H.		x
Zeitschrift für experimentelle und angewandte Psychologie	1962	Bewährungsstudie über die Aussagekraft von Leistungstests zum Problem der Kraftfahreignung	Brickenkamp, R.	x	
Zeitschrift für experimentelle und angewandte Psychologie	1968	Faktorenanalytische Validierung der Handschrift mit besonderer Berücksichtigung projektiver Methoden	Lockowandt, O.		x
Zeitschrift für experimentelle und angewandte Psychologie	1972	Untersuchung über den Zusammenhang zwischen Intelligenz (Wechsler) und Konzentrationsfähigkeit (Test d2 nach Brickenkamp)	Wiese, W.; Kroj, G.		x
Zeitschrift für experimentelle und angewandte Psychologie	1973	Bemerkungen zur Untersuchung über den Zusammenhang zwischen Intelligenz (WECHSLER) und Konzentrationsfähigkeit (Test d2 nach Brickenkamp) von W. Wiese und G. Kroj	Jäger, R.		x
Zeitschrift für experimentelle und angewandte Psychologie	1973	Über das Verhalten vorschulpflichtiger Kinder unter fremdgesetzter Aufgabenstellung bei der Bearbeitung von optischem Material	Kleber, E. W.		x
Zeitschrift für experimentelle und angewandte Psychologie	1975	Motivation Analysis Test (MAT) ein nicht verfälschbarer Test?	Janssen, J. P.		x

Zeitschrift für experimentelle und angewandte Psychologie	1977	Eine Analyse des Verweigerungsproblems mit Beinahe-Verweigerern	Effler, M.; Böhmke, W.		x
Zeitschrift für experimentelle und angewandte Psychologie	1978	Tageszeit, Extraversion und Leistungen im Test d2	Westhoff, K.; Krull-Wittkopf, R.		x
Zeitschrift für experimentelle und angewandte Psychologie	1980	Die Wirkung von erhöhter Zielsetzung und der Anwesenheit anderer auf die Mengen- und Güteleistung bei drei Arten von Aufgaben	Schneider, K.; Albers, H.; Hover, D.; Sommer, D.		x
Zeitschrift für experimentelle und angewandte Psychologie	1981	Zum Lernen verbalen Materials unterschiedlicher semantischer Eigenschaften	von Eye, A.; Krampen, G.		x
Zeitschrift für experimentelle und angewandte Psychologie	1982	Hilfsbereitschaft in Abhängigkeit von Aufgaben- und sozialer Struktur der Situation. Ein Experiment zum Inventionsverhalten	Abele, A.		x
Zeitschrift für experimentelle und angewandte Psychologie	1983	Zur Wahl und Prüfung statistischer Hypothesen in psychologischen Untersuchungen	Hager, W.; Westermann, R.		x
Zeitschrift für experimentelle und angewandte Psychologie	1984	Funktionale Hilfslosigkeit oder Diskrepanzreduktion?	Bossong, B.		x

Zeitschrift für experimentelle und angewandte Psychologie	1984	Die Wirkung elektrischer Felder auf Leistung und Befindlichkeiten gesunder Probanden	Hansk, G.; Riedl, H.; Möse, J. R.; Fischer, G.		x
Zeitschrift für experimentelle und angewandte Psychologie	1988	Frustrationsbedingter Blutdruckanstieg und dessen Abbau durch aggressive Reaktion	Huber, H. P.; Hauke, D.; Gramer, M.		x
Zeitschrift für experimentelle und angewandte Psychologie	1991	Stimmung und Leistungs-niveau bei einfachen Aufgaben	Klauer, K. C.; Siemer, M.; Stöber, J.		x
Zeitschrift für experimentelle und angewandte Psychologie	1992	Psychologische Variablen und Erfolg interpersonaler Koordination bei motorischen Aufgabenforderungen	Müller, G. F.		x
Zeitschrift für experimentelle und angewandte Psychologie	1992	Emotionsinduktion ohne Bewusstheit und Verhaltensänderung ohne Absicht	Sokolowski, K.		x
Zeitschrift für experimentelle und angewandte Psychologie	1993	Die Bedeutung des Kontextes für das episodische Gedächtnis bei jüngeren und älteren Menschen	Preußler, W.		x
Zeitschrift für experimentelle und angewandte Psychologie	1994	Effekte natürlicher Stim-mungsveränderungen auf Diskriminationsleistung und Antworttendenz bei der Mustererkennung	Otto, J. H.; Hänze, M.		x
Zeitschrift für experimentelle Psychologie	2003	Design-Effects in Prospective and Retrospective Memory for Actions	Jahn, P.; En-gelkamp, J.	-	-

Zeitschrift für experimentelle Psychologie	2004	Is the implicit association test immune to faking?	Steffens, M. C.		x
Zeitschrift für experimentelle Psychologie	2007	Assessing automatic activation of valence	Stahl, C.; Degner, J.		x
Zeitschrift für experimentelle Psychologie	2011	Exaggeration is harder than understatement, but practice makes perfect!	Röhner, J.; Schröder-Abé, M.; Schütz, A.	-	-
Zeitschrift für experimentelle Psychologie	2012	Effects of monetary incentives on task switching	Kleinsorge, T.; Rinkenauer, G.	-	-
Zeitschrift für experimentelle Psychologie	2015	Instruction-Based Approach-Avoidance Effects changing stimulus evaluation via the mere instruction to approach or avoid stimuli	von Dessel, P.; de Houwer, J.; Gast, A.; Tucker Smith, C.	-	-
Zeitschrift für Differentielle und Diagnostische Psychologie	1981	Zur Taxonomie fehlerhaften Arbeitens in psychologischen Leistungstest	Goeters, K. - M.		x
Zeitschrift für Differentielle und Diagnostische Psychologie	1981	Die "Anankastische Persönlichkeit" - Fakt oder Fiktion? Experimentelle Diagnostik der Zwangsneurose	Zaworka, W.; Hand, I.		x
Zeitschrift für Differentielle und Diagnostische Psychologie	1988	Berufseignungsdiagnostik	Schuler, H.	-	-

Zeitschrift für Differentielle und Diagnostische Psychologie	1989	Die Bearbeitung von Persönlichkeitsfragebogen in der Eignungswahl: Ein Indikator für Verfälschung?	Maschke, P.	-	-
Zeitschrift für Differentielle und Diagnostische Psychologie	1992	Die Auswirkung von Depressivität auf Intelligenzleistung, Konzentrationsfähigkeit und feinmotorische Geschicklichkeit	Brenner, B.; Formann, A. K.; Fretska, E.-M.		x
Zeitschrift für Differentielle und Diagnostische Psychologie	1997	Test d2-Aufmerksamkeits-Belastungs-Test	Oswald, W. D.; Hagen, B. / Replik von Brickenkamp, R.	x	x
Zeitschrift für Differentielle und Diagnostische Psychologie	1997	Ein Beitrag zur Konstruktvalidität von d2 und Revisi-onstest	Schmidt-Atzert, L.; Ising, M.	x	
Zeitschrift für Differentielle und Diagnostische Psychologie	1999	Testangst bei Fahreignungsbegutachtungen: Die Angst-Leistung-Relation	Strohbeck-Külmer, P.		x
Zeitschrift für Differentielle und Diagnostische Psychologie	1999	Konventionelle und computergestützte Durchführung von Leistungstests: Benton-Test	Merten, T.		x
Zeitschrift für Differentielle und Diagnostische Psychologie	1999	Der Raumvorstellungstest "Schnitte" Evaluation des Verfahrens an mathematisch besonders befähigten Schülerinnen und Schülern	Quaiser-Pohl, C.; Lehmann, W.		x

Zeitschrift für Differentielle und Diagnostische Psychologie	2000	Aktuelle Aktiviertheit und selektive Aufmerksamkeit: Ein Betrag zur Hypothese von der umgekehrt u-förmigen Beziehung zwischen Aspekten der Aktiviertheit und Leistungsmerkmalen	Imhof, M.		x
Zeitschrift für Differentielle und Diagnostische Psychologie	2003	Testrezesion: FAIR	Kubinger, K. D.; Moosbrugger, H.; Oehlschlägel, J.; Wagner, M. M.	-	-
Zeitschrift für Differentielle und Diagnostische Psychologie	2004	Kann die Konfundierung von Konzentrationsleistung und Aktivierung durch adaptives Testen mit dem FAKT vermieden werden?	Frey, A.; Moosbrugger, H.		x
Zeitschrift für Differentielle und Diagnostische Psychologie	2005	An Overview of Research into the Cognitive Basis of Intelligence	Schweizer, K.	-	-
Zeitschrift für Differentielle und Diagnostische Psychologie	2006	Cognitive Abilities and Their Interplay - Reasoning, Crystallized Intelligence, Working Memory - Components, and Sustained Attention	Buehner, M.; Krumm, S.; Ziegler, M.; Pluecken, T.	x	
Zeitschrift für Differentielle und Diagnostische Psychologie	2008	Speeded Paper-Pencil Sustained Attention and Mental Speed Tests – Can Performances Be Discriminated?	Krumm, S.; Schmidt-Atzert, L.; Michalczyk, K. M.; Danthiir, V.	x	
Zeitschrift für Differentielle und Diagnostische Psychologie	2011	Coordination as a Crucial Component of Performance on a Sustained Attention Test - Evidence From the Digit-Symbol Substitution Test	Krumm, S.; schmidt-Atzert, L.; Bracht, M.; Ochs. L.	x	

Zeitschrift für Differentielle und Diagnostische Psychologie	2012	Attention Tests in Different Stimulus Presentation Modes - A Facet Model of Performance in Attention Tests	Krumm, S.; Schmidt-Atzert, L.; Schmidt, S.; Zenses, E.-M.; Stenzel, N.	x	
Zeitschrift für Differentielle und Diagnostische Psychologie	2014	Cognitive Characteristics of Students with Superior Performance in Mathematics	Leikin, R.; Paz-Baruch, N.; Leikin, M.		x
Psychologische Beiträge; Vierteljahresschrift für alle Gebiete der Psychologie	1974	Zur Konstruktion differentieller Kategoriensysteme von Lehrern	Kleiter, E.		x
Psychologische Beiträge; Vierteljahresschrift für alle Gebiete der Psychologie	1980	Zum Einfluss eines Orientierungskontextes auf das freie reproduzieren und Wiedererkennen	Humpert, W.; Schwarze, H.		x
Psychologische Beiträge; Vierteljahresschrift für alle Gebiete der Psychologie	1987	Zur Kurzzeitspeicherung von Kraft- Zeit-Parametern: Untersuchungen mit dem Ruder- und Fahrradergometer zur motorischen Kodierung	Janssen, J. P.; Stoll, H.; Volkens, K.		x
Psychologische Beiträge; Vierteljahresschrift für alle Gebiete der Psychologie	1991	Das Problem der Reaktionszeiten: Was geschieht in Systemen, wenn viel Zeit vergeht?	Hajos, A.		x

Psychologische Beiträge; Vierteljahresschrift für alle Gebiete der Psychologie	1995	Zusammenhänge zwischen Testosteronkonzentration im Speichel und Leistung in einem Aufmerksamkeitstest	Müller, M. J.; Netter, P.		x
Psychologische Beiträge; Vierteljahresschrift für alle Gebiete der Psychologie	1999	Der Einfluß von Vergleichsergebnissen und Selbsteinschätzungen	Grau, I.; Mielke, R.		x
Psychologische Beiträge; Vierteljahresschrift für alle Gebiete der Psychologie	1999	Word association responses after traumatic brain injury	Merten, T.		x
Psychologische Beiträge; Vierteljahresschrift für alle Gebiete der Psychologie	2000	Die Computerversion des d2 und die Frage der Transferäquivalenz	Merten, T.		x
Psychologische Beiträge; Vierteljahresschrift für alle Gebiete der Psychologie	2002	Performance of vigilance during a hyperbaric chamber dive related to preferred coping strategies: A Pilot Study	Dietmaier, G.; Anegg, U.		x

Psychologische Beiträge; Vierteljahresschrift für alle Gebiete der Psychologie/ Psychology Science quarterly	2005	Accuracy, latency and confidence in abstract reasoning: the influence of fear of failure and gender	Preckel, F.; Freund, P. A.		x
Psychologische Beiträge; Vierteljahresschrift für alle Gebiete der Psychologie/ Psychology Science quarterly	2007	Self-estimates of attention performance	Mengelkamp, C.; Jäger, R. S.		x
Psychologische Beiträge; Vierteljahresschrift für alle Gebiete der Psychologie/ Psychology Science quarterly	2008	Deficits fine motor skill as an important factor in the identification of gifted underachievers in primary school	Stoeger, H.; Ziegler, A.; Martzog, P.		x
Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie	1973	Einfluß von Belastung auf die Feinmotorik von Schülern	Kiener, F.; Hugow, K.		x
Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie	1984	Zur differentiellen Bedeutung metakognitiver Komponenten für das Verstehen und das Behalten von Texten	Hasselhorn, M.; Körkel, J.		x

Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie	1985	Der Einfluß des Schulbesuchs auf Intelligenzleistungen im Grundschulalter	Merz, F.; Renner, H.; Ehlers, Th.		x
Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie	1985	Exploratorische Analysen zu Komponenten des Schulerfolgs	Schneider, W.; Bös, K.		x
Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie	2000	Testbesprechung – Intelligenz-Struktur-Test 2000 (I-S-T 2000)	Sirsch, U.		x
Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie	2009	Testbesprechung - BUEGA. Basisdiagnostik Umschriebener Entwicklungsstörungen im Grundschulalter.	Renner, G.		x
Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie	2010	Aufmerksamkeit als Mediator der Beziehung zwischen Fernsehverhalten und Schulleistung	Schüttenhelm, R.; Ennemoser, M.; Schneider, W.		x
Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie	2011	Testbesprechung - Intelligence and Development Scales (IDS). Intelligenz und Entwicklungsskalen für Kinder von 5–10 Jahren.	Koch, H.; Kastner-Koller, U.; Deimann, P.		x

Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie	2013	Testbesprechung - ZAREKI-R, TEDI-MATH und TeDDy-PC im Vergleich	Pletschko, T.		x
--	------	---	---------------	--	---

Tabelle A- 2: Einordnung der Artikel in den unterschiedlichen Anwendungsgebieten, Jahren und für die Verwendung der Stichprobengrößen der Eigenpublikationen (Quelle: Eigene Darstellung)

Eigenpublikationen				
Jahr	Anwendungsbereiche	Zeitschrift	Autor(en)	Verwendung des Artikels bei durchschnittlicher Stichprobengröße
1993	Arbeits- und Organisationspsychologie	Zeitschrift für Arbeits- und Organisationspsychologie	Schmidt-Atzert, L.; Deter, B.	x
1997	Arbeits- und Organisationspsychologie	Diagnostica	Schmidt-Atzert, L.; Bühner, M.	x
1997	Arbeits- und Organisationspsychologie	Zeitschrift für Differentielle und Diagnostische Psychologie	Schmidt-Atzert, L.; Ising, M.	x
2002	Arbeits- und Organisationspsychologie	Zeitschrift für Personalpsychologie	Schmidt-Atzert, L.	
2004	Arbeits- und Organisationspsychologie	Zeitschrift für Personalpsychologie	Schmidt-Atzert, L.; Deter, B.; Jaeckel, S.;	x
2005	Arbeits- und Organisationspsychologie	Zeitschrift für Personalpsychologie	Schmidt-Atzert, L.; Brandstätter, V.	
2006	Arbeits- und Organisationspsychologie	Diagnostica	Schmidt-Atzert, L.; Bühner, M.; Enders, P.	x
2006	Arbeits- und Organisationspsychologie	Zeitschrift für Differentielle und Diagnostische Psychologie	Buehner, M.; Krumm, S.; Ziegler, M.; Pluecken, T.	x

2011	Arbeits- und Organisationspsychologie	Zeitschrift für Differentielle und Diagnostische Psychologie	Krumm, S.; Schmidt-Atzert, L.; Bracht, M.; Ochs, L.	x
1998	Neuropsychologie	Diagnostica	Schmidt-Atzert, L.; Bühner, M.	x
1973	Testbezogene Untersuchungen	Diagnostica	Brickenkamp, R.	
1991	Testbezogene Untersuchungen	Diagnostica	Brickenkamp, R.	x
1997	Testbezogene Untersuchungen	Zeitschrift für Differentielle und Diagnostische Psychologie	Oswald, W. D.; Hagen, B./ Replik von Brickenkamp, R.	
2004	Testbezogene Untersuchungen	Diagnostica	Schmidt-Atzert, L.; Bühner, M.; Rischen, S.; Warkentin, V.	x
2008	Testbezogene Untersuchungen	Zeitschrift für Differentielle und Diagnostische Psychologie	Krumm, S.; Schmidt-Atzert, L.; Michalczyk, K. M.; Danthiir, V.	x
2012	Testbezogene Untersuchungen	Zeitschrift für Differentielle und Diagnostische Psychologie	Krumm, S.; Schmidt-Atzert, L.; Schmidt, S.; Zenses, E.-M.; Stenzel, N.	x
1962	Verkehrspsychologie	Zeitschrift für Arbeits- und Organisationspsychologie	Brickenkamp, R.	x
1962	Verkehrspsychologie	Zeitschrift für experimentelle und angewandte Psychologie	Brickenkamp, R.	x
1966	Verkehrspsychologie	Diagnostica	Brickenkamp, R.; Rump, G.	x

Tabelle A-3: Einordnung der Artikel in den unterschiedlichen Anwendungsgebieten, Jahren und für die Verwendung der Stichprobengrößen der Fremdpublikationen (Quelle: Eigene Darstellung)

Fremdpublikationen				
Jahr	Anwendungsbereiche	Zeitschrift	Autor(en)	Verwendung des Artikels

				bei durchschnittlicher Stichprobengröße
1969	Arbeits- und Organisationspsychologie	Zeitschrift für Arbeits- und Organisationspsychologie	Bäumler, G.; Dvorak, H.-P.	x
1971	Arbeits- und Organisationspsychologie	Diagnostica	Marschner, G.	x
1975	Arbeits- und Organisationspsychologie	Zeitschrift für Arbeits- und Organisationspsychologie	Steck, P.; Rausche, A.; Steege, F.W.	x
1979	Arbeits- und Organisationspsychologie	Zeitschrift für Arbeits- und Organisationspsychologie	Fay, E.; Prystav, G.	x
1980	Arbeits- und Organisationspsychologie	Zeitschrift für Arbeits- und Organisationspsychologie	Fay, E.; Stumpf, H.	
1981	Arbeits- und Organisationspsychologie	Diagnostica	Stumpf, H.; Fay, E.	x
1981	Arbeits- und Organisationspsychologie	Zeitschrift für Differentielle und Diagnostische Psychologie	Goeters, K. - M.	
1984	Arbeits- und Organisationspsychologie	Zeitschrift für experimentelle und angewandte Psychologie	Hansk, G.; Riedl, H.; Möse, J. R.; Fischer, G.	x
1987	Arbeits- und Organisationspsychologie	Diagnostica	Stumpf, H.; Fay, E.	x
1988	Arbeits- und Organisationspsychologie	Zeitschrift für Arbeits- und Organisationspsychologie	Fay, E.; Klieme, E.	
1988	Arbeits- und Organisationspsychologie	Diagnostica	Westhoff, K.; Lemme, M.	x

1989	Arbeits- und Organisationspsychologie	Diagnostica	Westhoff, K.	x
1990	Arbeits- und Organisationspsychologie	Diagnostica	Westhoff, K.; Dewald, D.	x
1990	Arbeits- und Organisationspsychologie	Zeitschrift für Arbeits- und Organisationspsychologie	Schuler, H.	
1991	Arbeits- und Organisationspsychologie	Zeitschrift für Arbeits- und Organisationspsychologie	Eser, K.-H.	x
1991	Arbeits- und Organisationspsychologie	Zeitschrift für experimentelle und angewandte Psychologie	Klauer, K. C.; Siemer, M.; Stöber, J.	
1992	Arbeits- und Organisationspsychologie	Diagnostica	Westhoff, K.; Hagemeister, C.	x
1992	Arbeits- und Organisationspsychologie	Zeitschrift für experimentelle und angewandte Psychologie	Müller, G. F.	x
1992	Arbeits- und Organisationspsychologie	Diagnostica	Fay, E.	x
1994	Arbeits- und Organisationspsychologie	Zeitschrift für Arbeits- und Organisationspsychologie	Schmidt, J. U.	x
1999	Arbeits- und Organisationspsychologie	Zeitschrift für Arbeits- und Organisationspsychologie	Rau, R.; Triemer, A.	
1999	Arbeits- und Organisationspsychologie	Psychologische Beiträge; Vierteljahresschrift für alle Gebiete der Psychologie	Grau, I.; Mielke, R.	x
2002	Arbeits- und Organisationspsychologie	Zeitschrift für Personalpsychologie	Kelbetz, G.; Schuler, H.	x

2002	Arbeits- und Organisationspsychologie	Zeitschrift für Personalpsychologie	Hagemeister, C.; Scholz, A.; Westhoff, K.	
2002	Arbeits- und Organisationspsychologie	Psychologische Beiträge; Vierteljahresschrift für alle Gebiete der Psychologie	Dietmaier, G.; Anegg, U.	x
2005	Arbeits- und Organisationspsychologie	Psychologische Beiträge; Vierteljahresschrift für alle Gebiete der Psychologie	Preckel, F.; Freund, P. A.	x
2006	Arbeits- und Organisationspsychologie	Zeitschrift für Personalpsychologie	Fisseni, H.-J.	
2007	Arbeits- und Organisationspsychologie	Psychologische Beiträge; Vierteljahresschrift für alle Gebiete der Psychologie	Mengelkamp, C.; Jäger, R. S.	x
2012	Arbeits- und Organisationspsychologie	Zeitschrift für Personalpsychologie	Wetter, O. E.; Wegge, J.; Jonas, K.; Schmidt, K.-H.	
2015	Arbeits- und Organisationspsychologie	Diagnostica	Becker, N.; Preckel, F.; Karbach, J.; Raffel, N.; Spinath, F. M.	x
1964	Klinische und Medizinische Psychologie	Zeitschrift für Arbeits- und Organisationspsychologie	Kunkel, E.	
1976	Klinische und Medizinische Psychologie	Zeitschrift für Arbeits- und Organisationspsychologie	Eich, F.; Klug, J.; Wetzel, H.	

1980	Klinische und Medizinische Psychologie	Zeitschrift für Arbeits- und Organisationspsychologie	Steck, P.	x
1981	Klinische und Medizinische Psychologie	Zeitschrift für Differentielle und Diagnostische Psychologie	Zaworka, W.; Hand, I.	x
1983	Klinische und Medizinische Psychologie	Diagnostica	Rauchfleisch, U.	x
1985	Klinische und Medizinische Psychologie	Diagnostica	Sorgatz, H.; Barth, G.	x
1987	Klinische und Medizinische Psychologie	Diagnostica	Eser, K.-H.	x
1988	Klinische und Medizinische Psychologie	Zeitschrift für experimentelle und angewandte Psychologie	Huber, H. P.; Hauke, D.; Gramer, M.	x
1992	Klinische und Medizinische Psychologie	Zeitschrift für Differentielle und Diagnostische Psychologie	Brenner, B.; Formann, A. K.; Fretska, E.- M.	
1993	Klinische und Medizinische Psychologie	Zeitschrift für experimentelle und angewandte Psychologie	Preußler, W.	x
1994	Klinische und Medizinische Psychologie	Zeitschrift für experimentelle und angewandte Psychologie	Otto, J. H.; Hänze, M.	x
1995	Klinische und Medizinische Psychologie	Psychologische Beiträge; Vierteljahresschrift für alle Gebiete der Psychologie	Müller, M. J.; Netter, P.	x
1996	Klinische und Medizinische Psychologie	Diagnostica	Steck, P.	x

2011	Klinische und Medizinische Psychologie	Diagnostica	Funsch, K.; Arias Martin, B.; Halder-Sinn, P.	x
1979	Neuropsychologie	Diagnostica	Hahlweg, K.	x
1981	Neuropsychologie	Diagnostica	Hahlweg, K.; Kühnlein, I.	x
1983	Neuropsychologie	Diagnostica	Kühnlein, I.; Hahlweg, K.; Müller, G.; Sevenitz, B.	x
1991	Neuropsychologie	Psychologische Beiträge; Vierteljahresschrift für alle Gebiete der Psychologie	Hajos, A.	x
1999	Neuropsychologie	Zeitschrift für Differentielle und Diagnostische Psychologie	Merten, T.	x
1999	Neuropsychologie	Psychologische Beiträge; Vierteljahresschrift für alle Gebiete der Psychologie	Merten, T.	x
2000	Neuropsychologie	Psychologische Beiträge; Vierteljahresschrift für alle Gebiete der Psychologie	Merten, T.	x
1973	Pädagogische Psychologie	Zeitschrift für experimentelle und angewandte Psychologie	Kleber, E. W.	x
1973	Pädagogische Psychologie	Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie	Kiener, F.; Hugow, K.	

1974	Pädagogische Psychologie	Zeitschrift für Arbeits- und Organisationspsychologie	Kleber, E. W.	
1974	Pädagogische Psychologie	Psychologische Beiträge; Vierteljahresschrift für alle Gebiete der Psychologie	Kleiter, E.	
1977	Pädagogische Psychologie	Zeitschrift für experimentelle und angewandte Psychologie	Effler, M.; Böhmke, W.	x
1978	Pädagogische Psychologie	Diagnostica	Gheorghiu, V. A.; Grimm, H.-J.; Hodapp, V.	x
1980	Pädagogische Psychologie	Zeitschrift für experimentelle und angewandte Psychologie	Schneider, K.; Albers, H.; Hoyer, D.; Sommer, D.	x
1981	Pädagogische Psychologie	Zeitschrift für experimentelle und angewandte Psychologie	von Eye, A.; Krampen, G.	x
1984	Pädagogische Psychologie	Zeitschrift für experimentelle und angewandte Psychologie	Bossong, B.	x
1984	Pädagogische Psychologie	Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie	Hasselhorn, M.; Körkel, J.	x
1985	Pädagogische Psychologie	Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie	Merz, F.; Renner, H.; Ehlers, T.	x

1985	Pädagogische Psychologie	Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie	Schneider, W.; Bös, K.	x
1999	Pädagogische Psychologie	Zeitschrift für Differentielle und Diagnostische Psychologie	Quaiser-Pohl, C.; Lehmann, W.	x
2003	Pädagogische Psychologie	Diagnostica	Westhoff, K.; Graubner, J.	x
2008	Pädagogische Psychologie	Psychologische Beiträge; Vierteljahresschrift für alle Gebiete der Psychologie	Stoeger, H.; Ziegler, A.; Martzog, P.	x
2009	Pädagogische Psychologie	Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie	Renner, G.	
2010	Pädagogische Psychologie	Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie	Schüttenhelm, R.; Ennemoser, M.; Schneider, W.	x
2011	Pädagogische Psychologie	Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie	Koch, H.; Kastner-Koller, U.; Deimann, P.	
2013	Pädagogische Psychologie	Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie	Pletschko, T.	x
2014	Pädagogische Psychologie	Zeitschrift für Differentielle und Diagnostische Psychologie	Leikin, R.; Paz-Baruch, N.; Leikin, M.	x

1968	Testbezogene Untersuchungen	Zeitschrift für experimentelle und angewandte Psychologie	Lockowandt, O.	x
1973	Testbezogene Untersuchungen	Diagnostica	Sommer, G.	x
1973	Testbezogene Untersuchungen	Diagnostica	Heinrich, H. Ch.	
1978	Testbezogene Untersuchungen	Zeitschrift für experimentelle und angewandte Psychologie	Westhoff, K.; Krull-Wittkopf, R.	x
1979	Testbezogene Untersuchungen	Diagnostica	Kallenbach, K.; Oswald, W. D.	x
1980	Testbezogene Untersuchungen	Psychologische Beiträge; Vierteljahresschrift für alle Gebiete der Psychologie	Humpert, W.; Schwarze, H.	
1982	Testbezogene Untersuchungen	Zeitschrift für experimentelle und angewandte Psychologie	Abele, A.	
1983	Testbezogene Untersuchungen	Diagnostica	Westhoff, K.; Kluck, M.-L.	x
1983	Testbezogene Untersuchungen	Zeitschrift für experimentelle und angewandte Psychologie	Hager, W.; Westermann, R.	
1984	Testbezogene Untersuchungen	Diagnostica	Westhoff, K.; Kluck, M.-L.	
1985	Testbezogene Untersuchungen	Diagnostica	Westhoff, K.	x

1987	Testbezogene Untersuchungen	Psychologische Beiträge; Vierteljahresschrift für alle Gebiete der Psychologie	Janssen, J. P.; Stoll, H.; Volkens, K.	x
1991	Arbeits- und Organisationspsychologie	Diagnostica	Oehlschlägel, J.; Moosbrugger, H.	
1992	Testbezogene Untersuchungen	Zeitschrift für experimentelle und angewandte Psychologie	Sokolowski, K.	x
1995	Testbezogene Untersuchungen	Diagnostica	Fay, E.	
1995	Testbezogene Untersuchungen	Diagnostica	Schorr, A.	
1997	Testbezogene Untersuchungen	Diagnostica	Steck, P.	
1997	Testbezogene Untersuchungen	Zeitschrift für Differentielle und Diagnostische Psychologie	Oswald, W. D.; Hagen, B.	
2000	Testbezogene Untersuchungen	Zeitschrift für Arbeits- und Organisationspsychologie	Kersting, M.	
2000	Testbezogene Untersuchungen	Zeitschrift für Differentielle und Diagnostische Psychologie	Imhof, M.	x
2000	Testbezogene Untersuchungen	Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie	Sirsch, U.	x
2004	Testbezogene Untersuchungen	Zeitschrift für experimentelle Psychologie	Steffens, M. C.	x

2004	Testbezogene Untersuchungen	Zeitschrift für Differentielle und Diagnostische Psychologie	Frey, A.; Moosbrugger, H.	
2007	Testbezogene Untersuchungen	Zeitschrift für experimentelle Psychologie	Stahl, C.; Degner, J.	x
1966	Verkehrspsychologie	Zeitschrift für Arbeits- und Organisationspsychologie	Kunkel, E.	
1971	Verkehrspsychologie	Diagnostica	Dingel, W.	x
1972	Verkehrspsychologie	Zeitschrift für Arbeits- und Organisationspsychologie	Kroj, G.	x
1972	Verkehrspsychologie	Zeitschrift für experimentelle und angewandte Psychologie	Wiese, W.; Kroj, G.	x
1973	Verkehrspsychologie	Zeitschrift für experimentelle und angewandte Psychologie	Jäger, R.	
1975	Verkehrspsychologie	Zeitschrift für experimentelle und angewandte Psychologie	Janssen, J. P.	x
1980	Verkehrspsychologie	Zeitschrift für Arbeits- und Organisationspsychologie	Janssen, J. P.; Reitzner, Chr.	x
1985	Verkehrspsychologie	Zeitschrift für Arbeits- und Organisationspsychologie	Brunner, A.; Dvorak, H.	x
1990	Verkehrspsychologie	Diagnostica	Bukasa, B.; Kisser, R.; Wenninger, U.	x

1999	Verkehrspsychologie	Zeitschrift für Differentielle und Diagnostische Psychologie	Strohbeck-Külmer, P.	x
------	---------------------	--	----------------------	---

Tabelle A-4: Reliabilitätsschätzungen nach Aufteilung in die erste und zweite Testhälfte bei den Eigenpublikationen (Quelle: Eigene Darstellung)

Jahr	Anwendungsgebiet	Autor(en)	Stichprobe	GZ	GZ-F	KL	F%	SB
2008	Testbezogene Untersuchungen	Krumm, S.; Schmidt-Atzert, L.; Michalczyk, K. M.; Danthiir, V.	Studenten (N=199)	0,97*				

Anmerkung: * $p < 0,05$ (einseitig).

Tabelle A-5: Reliabilitätsschätzungen nach der Odd-even-Methode bei den Eigenpublikationen (Quelle: Eigene Darstellung)

Jahr	Anwendungsgebiet	Autor(en)	Stichprobe	GZ	GZ-F	KL	F%	SB
2012	Testbezogene Untersuchungen	Krumm, S.; Schmidt-Atzert, L. Schmidt, S.; Zenses, E.- M.; Stenzel, M.	Studenten (N=150)	0,93				

Tabelle A-6: Reliabilitätsschätzungen nach Aufteilung in die erste und zweite Testhälfte bei den Fremdpublikationen (Quelle: Eigene Darstellung)

Jahr	Anwendungsgebiet	Autor(en)	Stichprobe	GZ	GZ-F	KL	F	SB
1983	Klinische und Medizinische Psychologie	Rauchfleisch, U.	Patienten mit hirndiffusem Psychosyndrom und neurotischen Störungen (N= 90)	0,97*			0,95*	

Anmerkung: * $p < 0,05$ (einseitig).

Tabelle A-7: Reliabilitätsschätzungen nach der Odd-even-Methode bei den Fremdpublikationen (Quelle: Eigene Darstellung)

Jahr	Anwendungsgebiet	Autor(en)	Stichprobe	GZ	GZ-F	KL	F%	SB
1987	Klinische und Medizinische Psychologie	Eser, K.-H.	Kinder und Jugendliche mit Verhaltensstörungen (N= 85)	0,93*	0,93*		0,96*	0,95*

Anmerkung: * $p < 0,05$ (einseitig).

Tabelle A-8: Durchschnittliche Stichprobengrößen in den spezifischen Verwendungen der Eigenpublikationen (Quelle: Eigene Darstellung)

Durchschnittliche Stichprobengröße				
Jahr	Verkehrspsychologie	Neuropsychologie	Arbeits- und Organisationspsychologie	Testbezogene Untersuchungen
1962	28			
1962	172			
1963				
1964				
1965				
1966	112			
1967				
1968				
1969				
1970				
1971				
1972				
1973				
1974				
1975				
1976				
1977				
1978				
1979				
1980				
1981				
1982				
1983				
1984				
1985				
1986				
1987				
1988				
1989				
1990				

1991				98
1992				
1993			1758	
1994				
1995				
1996				
1997			121	
1997			1041	
1998	378			
1999				
2000				
2001				
2002				
2003				
2004			646	94
2005				
2006			207	
2006			121	
2007				
2008				199
2009				
2010				
2011			331	
2012				153
2013				
2014				
2015				
Mittelwert			364	

Tabelle A-9: Durchschnittliche Stichprobengrößen in den spezifischen Verwendungen der Fremdpublikationen (Quelle: Eigene Darstellung)

Durchschnittliche Stichprobengröße						
Jahr	Verkehrpsychologie	Pädagogische Psychologie	Neuropsychologie	Arbeits- und Organisationspsychologie	Klinische und Medizinische Psychologie	Testbezogene Untersuchungen
1962						
1963						
1964						
1965						
1966						

1967						
1968						100
1969				108		
1970						
1971	70			60		
1972	179					
1972	102					
1973		93				50
1974						
1975	100			329		
1976						
1977		1568				
1978		300				271
1979			258	873		214
1980	200	144			217	
1981		117	148	1235	95	
1982						
1983			194		90	195
1984		47		50		
1984		48				
1985	200	487			72	105
1985		87				
1986						
1987				1304	85	32
1988				108	224	
1989				15		
1990	248			75		
1991			66	50		
1992				75		30
1992				54		
1992				10		
1993					72	
1994				1030	13	
1995					64	
1996					177	
1997						
1997						
1998						
1999	181	236	372	160		
1999			40			
2000			148			380
2000						1303
2001						
2002				170		

2002			134		
2003	85				
2004					48
2005			103		
2006					
2007			134		40
2008	128				
2009					
2010	60				
2011				68	
2012					
2013	796				
2014	200				
2015			318		
Mittelwert	233				

Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Stellen, die wörtlich oder sinngemäß aus Quellen entnommen wurden, sind als solche kenntlich gemacht.

Diese Arbeit wurde in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegt.

Gelenau, den 12. Dezember 2016

Isabelle Einhorn