



universität
wien

Diplomarbeit

Titel der Arbeit

Reliability Generalization-Studien: Ein systematischer Review
zweiter Ordnung

Verfasser

Philipp Simbrunner

Angestrebter akademischer Grad

Magister der Naturwissenschaften (Mag. rer. nat.)

Wien, im Januar 2012

Studienkennzahl: 298

Studienrichtung: Psychologie

Betreuer: Assistenzprof. Privatdoz. MMag. DDDr. Martin Voracek

Inhaltsverzeichnis

1. Theoretischer Hintergrund	1
1.1. Reliabilität.....	1
1.1.1. Bedeutung der Reliabilität	2
1.1.2. Bestimmung der Reliabilität.....	2
1.1.3. Reliabilitätsbeeinflussende Faktoren.....	5
1.2. Reliability Generalization	6
1.2.1. Hintergrund und Begründung.....	6
1.2.2. Zielsetzung	7
1.2.3. Vorgehensweise.....	7
1.2.4. Kritische Punkte	9
1.3. Ziel dieser Diplomarbeit	10
2. Methode.....	13
2.1. Literatursuche	13
2.2. Datenerhebung	19
2.2.1. Kodierschema auf Studienebene	20
2.2.2. Kodierschema auf Skalenebene.....	21
2.2.3. Kodierschema für Meta-Analysen.....	22
2.3. Auswertung.....	25
3. Ergebnisse	27
3.1. Leitfrage A: „Das Forschungsfeld“	27
3.2. Leitfrage B: „Literatursuche und Induktion“	29
3.3. Leitfrage C: „Ermittlung typischer Werte“	31
3.4. Leitfrage D1: „Quellen der Variabilität der Reliabilität“	36
3.5. Leitfrage D2: Meta-Analysen zu Quellen der Variabilität	37

4.	Diskussion	39
4.1.	Leitfrage A: „Das Forschungsfeld“	39
4.2.	Leitfrage B: „Literatursuche und Induktion“	40
4.3.	Leitfrage C: „Ermittlung typischer Werte“	41
4.4.	Leitfrage D1: „Quellen der Variabilität der Reliabilität“	42
4.5.	Leitfrage D2: Meta-Analysen zu Quellen der Variabilität.....	43
4.6.	Einschränkungen dieser Arbeit	44
4.7.	Kritik an der Methode der Reliability Generalization.....	45
4.8.	Fazit	46
4.9.	Forschungsausblick	47
5.	Anhang.....	49
5.1.	Daten auf Studienebene mit Legende.....	49
5.2.	Daten auf Skalenebene mit Legende	55
5.3.	Daten zu Meta-Analysen mit Forest Plots.....	73
5.3.1.	Alter	73
5.3.2.	Altershomogenität	79
5.3.3.	Ethnie	83
5.3.4.	Geschlecht	90
5.3.5.	Klinische Stichprobe	98
5.3.6.	Standardabweichung des Score.....	102
5.3.7.	Sprache.....	105
5.3.8.	Stichprobengröße	110
5.4.	Autoren nach Anzahl der Publikationen	117
5.5.	Fachzeitschriften nach Anzahl der Publikationen.....	118
5.6.	Datenbanken nach Nutzungshäufigkeit.....	119
6.	Literaturverzeichnis	121
7.	Abstract.....	133
8.	Eidesstattliche Erklärung	135
9.	Curriculum Vitae	136

1. Theoretischer Hintergrund

1.1. *Reliabilität*

„Mit *Reliabilität* (Zuverlässigkeit) ist das Ausmaß gemeint, wie genau der Test misst, was er misst (egal, was er misst)“ (Rost, 2004, S. 33). Das Konzept der Reliabilität bezieht sich also auf die Genauigkeit einer Messung. Sie zählt, neben der Objektivität und der Validität, zu den drei Hauptgütekriterien eines psychologischen Messinstruments (Lienert & Raatz, 1994). Objektivität stellt eine Voraussetzung für Reliabilität dar und meint die Unabhängigkeit einer Messung von Testsituation, Auswertung und Ergebnisinterpretation (Moosbrugger & Kelava, 2007). „Mit Validität [oder Gültigkeit] ist gemeint, inwieweit der Test das misst, was er messen soll“ (Rost, 2004, S. 33). Eine ausführliche Beschreibung dieser und anderer Gütekriterien findet sich zum Beispiel bei Kubinger (2006). Die hier vorliegende Arbeit beschäftigt sich mit der Reliabilität, weshalb auf eine tiefgehende Diskussion anderer Gütekriterien verzichtet wird. Eine anschauliche Darstellung der Beziehungen zwischen den genannten Hauptgütekriterien bieten beispielsweise Lienert und Raatz (1994).

Die Grundannahme der Klassischen Testtheorie besagt, dass der beobachtete Wert (X) sich aus dem wahren Wert (T) und einem Messfehler (E) zusammensetzt (Formann, 2004). Dieser Messfehler beschreibt unsystematische Einflüsse (ist also kein systematischer Fehler im Sinne eines Bias), sodass er im Mittel einer Population den Wert Null annimmt und darüber hinaus nicht mit dem wahren Wert (T) korreliert ist (Rost, 1994). Die Definition der Reliabilität eines Tests erfolgt in der Klassischen Testtheorie über die Varianzen dieser drei Komponenten und lautet $Rel(X) = \text{Var}(T)/\text{Var}(X)$. Sie ergibt sich also aus dem Quotient der Varianz der wahren Werte und der Varianz der beobachteten Werte, letztere entspricht der Summe der Varianz der wahren Werte und der Varianz der Messfehler (Bühner, 2004; Lienert & Raatz, 1994; Rost, 2004). „Die Reliabilität gibt [also] den Anteil der Varianz der wahren Werte an der Varianz der beobachteten Werte an“ (Bühner, 2004, S. 115). Durch Subtraktion des Wertes des Reliabilitätskoeffizienten von Eins und der Multiplikation dieses Wertes mit 100 kann der Prozentsatz der Varianz der beobachteten Werte ermittelt werden, welcher auf Messfehler zurückzuführen ist. Eine Reliabilität von .70 bedeutet somit, dass 30 Prozent dieser Varianz durch Messfehler hervorgerufen werden (Meyer, 2004).

1.1.1. Bedeutung der Reliabilität

Reliabilität stellt eine Voraussetzung für Validität dar (z.B. Lienert & Raatz, 1994; Moosbrugger & Kelava, 2007). Dies ist logisch nachvollziehbar, da ohne Zuverlässigkeit auch keine Gültigkeit gegeben sein kann. Thompson (2003a) beispielsweise bemerkt, dass „perfectly unreliable scores measure nothing The only time that perfectly unreliable scores could conceivably be valid is if someone was designing a test intended consistently to measure nothing“ (p. 6). Eine ausführliche Darstellung der Beziehung zwischen Reliabilität und Validität bietet beispielsweise Meyer (2004).

Wie oben dargestellt, liegt der durch unsystematische Einflüsse bedingte Anteil der Variabilität einer Messung umso höher, je niedriger die Reliabilität ist. Die Bestimmung einer statistischen Signifikanz oder einer praktischen Bedeutsamkeit mittels Effektstärkenmaße auf der Basis von Daten, welche stark durch unsystematische Schwankungen beeinflusst sind, ist problematisch. Thompson (2003a) merkt hierzu an, dass „at the extreme, perfectly unreliable scores are perfectly random and cannot yield statistically, or practically ... significant results“ (p. 8). Nach Bühner (2004) kann eine Korrelation zwischen zwei mittels psychologischer Messinstrumente ermittelten Werten maximal der Quadratwurzel aus dem Produkt der beiden Reliabilitäten entsprechen. Eine ausführliche Besprechung von Möglichkeiten zur Anpassung von Effektstärken hinsichtlich der entsprechenden Reliabilität bietet der Artikel von Baugh (2002).

Die Bestimmung von Konfidenzintervallen für mittels psychologischer Messinstrumente gewonnene Werte erfolgt unter anderem mit Hilfe der Reliabilität (siehe Bühner, 2004 auch für eine ausführliche Betrachtung; Kubinger, 2006). Der Interpretation von Testwerten unter Berücksichtigung der entsprechenden Messgenauigkeit kommt insbesondere – aber nicht nur – in der Einzelfalldiagnostik eine besondere Bedeutung zu (Schermelleh-Engel & Werner, 2007).

1.1.2. Bestimmung der Reliabilität

Zur Bestimmung der Zuverlässigkeit einer Messung können mehrere Methoden unterschieden werden, welche statistisch gesehen auf Korrelationen beruhen (Bühner, 2004). Alle Methoden stellen eine Schätzung der Reliabilität dar. Diese kann nicht direkt berechnet werden, da weder die Messfehlervarianz noch die Varianz der wahren Werte bekannt sind

(Schermelleh-Engel & Werner, 2007). Nachfolgend werden die wichtigsten Methoden kurz dargestellt.

Bei der *Retestmethode* wird ein Messinstrument zu zwei verschiedenen Zeitpunkten derselben Stichprobe dargeboten. Die so gewonnenen Daten der beiden Zeitpunkte werden sodann korreliert (Bühner, 2004). Zu beachten ist hierbei, dass Veränderungen zwischen den beiden Zeitpunkten die Rohwerte und damit die Reliabilität beeinflussen können. Diese sind insbesondere Durchführungsbedingungen sowie Lern- oder Übungeffekte (Schermelleh-Engel & Werner, 2007). Reliabilitätskoeffizienten, welche auf der Retestmethode beruhen, geben auch einen Hinweis auf die zeitliche Stabilität des zu messenden Merkmals (Lienert & Raatz, 1994). Bei der *Paralleltestmethode* werden zwei parallele Formen eines Verfahrens derselben Stichprobe vorgegeben, die Korrelation der beiden Formen entspricht sodann der Reliabilität (Bühner, 2004). Abhängig vom Zeitintervall zwischen den beiden Vorgaben sind unterschiedliche Umstände zu beachten. Bei einem längeren Zeitintervall können ähnliche Probleme wie bei der Retestmethode auftreten. Werden beide Formen sehr kurz oder unmittelbar hintereinander dargeboten, sind situationsbedingte individuelle Faktoren (Konzentration, Müdigkeit) und Transfereffekte zu berücksichtigen (Schermelleh-Engel & Werner, 2007). Bei der Methode der *inneren Konsistenzanalyse* wird ein Instrument einmalig vorgegeben. Mit dieser Methode gewonnene Koeffizienten berücksichtigen weder intraindividuelle noch situative Bedingungen und lassen auch keine Rückschlüsse auf die zeitliche Stabilität eines Merkmals zu (Lienert & Raatz, 1994). Es können zwei Ansätze unterschieden werden. Bei der Testhalbierungsmethode werden die Items des Verfahrens in zwei gleich große Gruppen geteilt und mittels Korrelation und anschließender testlängenkorrigierender Aufwertung wird der Reliabilitätskoeffizient ermittelt. Die Aufteilung der Items erfolgt beispielsweise durch Zufall oder es wird eine Gruppe mit ungeraden und eine mit geraden Reihungsnummern gebildet (Bühner, 2004). Problematisch kann die Bildung von zwei vergleichbaren Gruppen sein. Bei dem zweiten Ansatz der inneren Konsistenzanalyse werden alle einzelnen Items als eigene Testteile angesehen. Die innere Zusammenhangsstruktur der Items dient sodann als Reliabilitätsschätzung im Sinne der inneren Konsistenz. Der bekannteste Koeffizient dieser Methode ist Cronbach's Alpha (Schermelleh-Engel & Werner, 2007). Eine weitere Methode der Reliabilitätsschätzung bilden Maße zur Charakterisierung des Ausmaßes von Übereinstimmungen in der Beurteilung oder Bewertung von Sachverhalten durch zwei oder mehr Personen (Atteslander, 2010). Hierbei

sind zwei Formen zu unterscheiden. Interrater Agreement meint die Übereinstimmung der Bewertung eines Sachverhaltes durch zwei oder mehr Beurteiler und wird in Prozentsätzen oder Cohen's Kappa-Koeffizienten ausgedrückt. Interrater-Reliabilität bezieht sich auf das Ausmaß, in welchem Sachverhalte von zwei oder mehr Beurteilern in derselben Reihenfolge geordnet werden. Interrater-Reliabilität wird beispielsweise in Korrelationen ausgedrückt (Goodwin, 2001).

Reliabilitätskoeffizienten, welche nach der Methode der inneren Konsistenzanalyse berechnet wurden, sind die am häufigsten berichteten Koeffizienten. Nach Hogan, Benjamin und Brezinski (2000) wurde Cronbach's Alpha in mehr als der Hälfte (66.5%) aller von ihnen untersuchten Artikel berichtet, nach der Kuder-Richardson Formel 20 (KR-20) ermittelte Koeffizienten wurden von zweieinhalb Prozent und Koeffizienten nach der Testhalbierungsmethode von etwa vier Prozent der Autorinnen und Autoren berichtet. Retest-Reliabilitäteskoeffizienten wurden von 19 Prozent angegeben und ein Prozent berichtete Kennwerte zur Urteilerübereinstimmung.

Alle Methoden der Reliabilitätsschätzung geben Auskunft über die Unzuverlässigkeit einer Messung, also den Anteil der Messfehlervarianz an der Gesamtvarianz. Zu beachten ist allerdings, dass unterschiedliche Methoden unterschiedliche Arten von Messfehlern untersuchen und somit der resultierende Reliabilitätskennwert von der gewählten Berechnungsmethode abhängig ist (Kieffer, Reese & Vacha-Haase, 2010; Thompson & Vacha-Haase, 2000). Werden beispielsweise mittels zweier verschiedener Methoden zwei Reliabilitätskoeffizienten in der Höhe von .90 ermittelt, so sind diese beiden zehn Prozent Messfehlervarianz „actually *separate and cumulative*. For these data a better reliability estimate might be .80!“ (Thompson, 2003b, p. 45). Einen modernen Ansatz innerhalb der Klassischen Testtheorie bildet die Generalisierbarkeitstheorie (oder G-Theory). Die bisher dargestellten Methoden führen die beobachtete Gesamtvarianz auf nur zwei Quellen zurück, nämlich auf Personen und einen unspezifischen Messfehler. Die G-Theory ermöglicht eine Teilung der Gesamtvarianz in eine beliebige Anzahl an Komponenten (Moosbrugger, 2007). Sie beruht auf dem statistischen Verfahren der Varianzanalyse (ANOVA), somit wird die simultane Betrachtung unterschiedlicher Messfehlerquellen, beispielsweise verschiedener Beobachter und verschiedener Messzeitpunkte, möglich. Darüber hinaus können Interaktionseffekte zwischen Quellen identifiziert werden (siehe Thompson, 2003b für eine ausführliche Einführung sowie ein anschauliches Anwendungsbeispiel).

1.1.3. Reliabilitätsbeeinflussende Faktoren

Neben den für die jeweilige Schätzmethode spezifischen Einflussfaktoren (beispielsweise das gewählte Zeitintervall zwischen zwei Testvorgaben bei der Retestmethode) sind noch weitere Faktoren bekannt. Die sinnvolle Interpretation von Reliabilitätskoeffizienten aus der Gruppe der inneren Konsistenzanalyse setzt voraus, dass alle Items des entsprechenden Tests von allen Personen der Stichprobe beantwortet wurden. Somit stellt die verfügbare Bearbeitungszeit einen Einflussfaktor dar (Liener & Raatz, 1994). Tests mit gleichen oder ähnlichen Items liefern meist reliablere Daten als solche mit sehr unterschiedlichen Einzelaufgaben. Die Reliabilität steigt meist mit der Anzahl der Items eines Verfahrens. Homogenitätsgrad und Testlänge beeinflussen also ebenfalls die Reliabilität (Bühner, 2004). Eine ausführliche Diskussion zum Thema der Beeinflussung der Reliabilität durch den Zufall, also durch Erraten der korrekten Antwort eines Items, bieten Lienert und Raatz (1994). Nach ihnen liefern offene Antwortformate im Allgemeinen reliablere Daten als Multiple-Choice Formate oder geschlossene dichotome Aufgaben. Einen weiteren bedeutenden Einflussfaktor stellt die Testwertstreuung der Stichprobe dar. Die Reliabilität ist per definitionem (siehe oben) direkt von der Varianz der wahren Werte abhängig. Nimmt diese Varianz – im Extremfall – den Wert Null ($\text{Var}(T) = 0$) an, so resultiert auch eine Reliabilität von Null, da $\text{Rel}(X) = \text{Var}(T)/\text{Var}(X) = 0/\text{Var}(X) = 0$. Die Reliabilität fällt somit in homogenen Gruppen niedriger aus, als in heterogenen Gruppen (z.B. Formann, 2004; Rost, 2004). Der Reliabilitätskoeffizient für Daten, welche anhand einer Stichprobe von hundert Studierenden der Psychologie in Wien im Alter von 20 – 25 gewonnen wurden, wird vermutlich niedriger ausfallen, als der Koeffizient für Daten, welche mit demselben Verfahren anhand einer zufällig gezogenen Stichprobe von hundert 20 – 25jährigen Wienerinnen und Wienern gewonnen wurden. Nicht etwa weil das Verfahren in der ersten Gruppen ungenauer messen würde, sondern weil die Streuung des interessierenden Merkmals vermutlich in der zweiten Gruppe höher sein wird (in Anlehnung an Schermelleh-Engel & Werner, 2007). Damit wird deutlich, dass die Reliabilität ein stichprobenabhängiges Maß ist und keine feste, testspezifische Größe darstellt. Rost (2004) merkt hierzu an, dass „die Reliabilität ... ein Merkmal des Tests *in einer Population* [ist] Es ist kein Merkmal, das dem Testinstrument wie ein Gütesiegel anhaftet...“ (S. 376).

1.2. Reliability Generalization

1.2.1. Hintergrund und Begründung

Wie eben dargestellt, ist Reliabilität keine feste Eigenschaft eines Tests. Sie sollte somit immer für die jeweiligen Daten eigens berechnet werden. Verschiedene Studien haben jedoch gezeigt, dass in weniger als der Hälfte aller analysierten Artikel Reliabilitätskoeffizienten, für die dort gewonnenen Daten, berechnet wurden. Somit berichtet die Mehrheit der Autorinnen und Autoren überhaupt keinen Wert für die Reliabilität der Daten oder bezieht sich auf die im entsprechenden Testhandbuch oder in einer anderen Studie berichteten Reliabilitätskoeffizienten (Meier & Davis, 1990; Vacha-Haase, Ness, Nilsson & Reetz, 1999; Willson, 1980). Die Bezugnahme auf einen andernorts berichteten Reliabilitätskoeffizienten als Mittel zur Beurteilung der eigenen Daten hinsichtlich ihrer Zuverlässigkeit wurde von Vacha-Haase, Kogan und Thompson (2000) als Reliability Induction bezeichnet. Eine gerechtfertigte Reliability Induction würde voraussetzen, dass beide Stichproben sowie deren Testwertvarianzen, also jene aus welcher der Reliabilitätskoeffizient stammt und jene, für welche dieser Wert induziert wurde, tatsächlich vergleichbar sind. In lediglich 20 Prozent der von Vacha-Haase et al. (2000) untersuchten Studien wurden Stichprobenmerkmale präsentiert, welche einen derartigen Vergleich ermöglichen und von diesen schien „in many instances the induction ... starkly implausible“ (p. 520). Deditius-Island und Caruso (2002) erweiterten den Begriff der Reliability Induction auf jene Situationen, in welchen gänzlich auf die Besprechung des Themas der Reliabilität verzichtet wird. Dieser Umstand kann entweder bedeuten, dass die Autorin/der Autor sich nicht mit der Reliabilität der eigenen Daten auseinandersetzt oder sie/er auf Grund andernorts berichteter Koeffizienten Reliabilität als gegeben ansieht, „with the former indicating poor research and the latter indicating that a reliability induction has taken place“ (p. 732). Shields und Caruso (2004) beschrieben später die explizite Bezugnahme auf nicht eigens berechnete Koeffizienten als Induktion by report und die implizite Bezugnahme als Induktion by omission. In verschiedene Reliability Induction-Studien konnte gezeigt werden, dass die Induktion von Reliabilitätskoeffizienten meist nicht gerechtfertigt scheint (Vacha-Haase et al., 2000; Vassar, Ridge & Hill, 2008).

1.2.2. Zielsetzung

Viele Forscherinnen und Forscher scheinen sich nicht ausreichend darüber bewusst zu sein, dass Reliabilität keine feste testspezifische Größe ist. Der Grund hierfür könnten unter anderem sprachliche Formulierungen wie „die Reliabilität des Tests“ sein. Insbesondere Bruce Thompson und Tammi Vacha-Haase setzen sich dafür ein, Reliabilität als Eigenschaft von Daten zu begreifen und damit einhergehend auch für die Verwendung einer korrekten Sprach- und Schreibweise, wie beispielsweise Scorereliabilität oder Reliabilität der Daten (Thompson, 1994; Thompson, 2003a, 2003b; Thompson & Vacha-Haase, 2000; Vacha-Haase, 1998). Diese Bewusstseinsbildung unter Wissenschaftlerinnen und Wissenschaftler stellt bereits die erste Zielsetzung der von Vacha-Haase (1998) vorgestellten Methode der Reliability Generalization dar. Neben dieser implizit formulierten Zielsetzung, soll diese Methode nach Vacha-Haase (1998) folgende explizit genannten Fragestellungen beantworten:

- Welche Höhe nimmt der Reliabilitätskoeffizient eines gegebenen Tests über mehrere Testvorgaben hinweg betrachtet typischer Weise an?
- In welchem Bereich bewegen sich diese Koeffizienten?
- Können Quellen der Variabilität der Reliabilität identifiziert werden und welche sind diese?

1.2.3. Vorgehensweise

Die Methode der Reliability Generalization wurde von Vacha-Haase (1998) als offener Vorschlag präsentiert, das heißt, es wurden keine exakten Vorgaben und Vorschriften bezüglich der genauen Vorgehensweise und der zu verwendenden statistisch-analytischen Verfahren gemacht (Thompson, 2002; Vacha-Haase, Henson & Caruso, 2002). Dies ist unter anderem deshalb sinnvoll, da es sich bei dieser Methode um einen meta-analytischen Ansatz handelt. Hierbei bilden die Reliabilitätskoeffizienten aus bereits durchgeführten Testvorgaben die Analyseeinheit und die dazugehörigen berichteten Stichprobenmerkmale bilden die jeweils zu untersuchenden potentiellen Quellen der Variabilität. Thompson und Vacha-Haase (2000) merken hierzu an, „ that ‚RG‘ studies are a meta-analytic characterization of what is hoped is a population of previous reports. We may not like the ingredients that go into making this sausage, but the ‚RG‘ chef can only work with the ingredients provided by the literature.“ (p. 184). Da die Wahl der statistischen Analyseverfahren von der jeweiligen Datenlage abhängig ist, ist der Verzicht auf die Darstellung von festen Durchführungsregeln sinnvoll

und nachvollziehbar. Darüber hinaus ist zu beachten, dass diese Methode als meta-analytischer *Ansatz* präsentiert wurde. Eine Reliability Generalization-Studie basiert auf Primärdaten und stellt damit immer einen (systematischen) Review dar. Eine Meta-Analyse, also die Anwendung entsprechender statistischer Verfahren, wird nicht notwendigerweise in einer derartigen Studie durchgeführt. Eine detailliertere Diskussion der Unterschiede zwischen Reviews und Meta-Analysen bietet Cooper (2010). Ausführliche Informationen zum Thema Meta-Analysen beziehungsweise systematischen Reviews finden sich beispielsweise bei Borenstein, Hedges, Higgins und Rothstein (2009) beziehungsweise bei Petticrew und Roberts (2006).

Die Durchführung einer Reliability Generalization-Studie sieht (in Anlehnung an Henson und Thompson, 2002) folgende Punkte vor: Zuerst wird ein Test ausgewählt, für welchen diese Analyse durchgeführt werden soll. Anschließend wird eine umfangreiche Literatursuche durchgeführt, um eine möglichst große Zahl an potentiell relevanten Primärstudien zu erhalten. Anhand eines zu entwickelnden Kodierschemas werden sodann die Reliabilitätskoeffizienten sowie interessierende Studienmerkmale (beispielsweise Stichprobeneigenschaften wie Alter oder Geschlecht) erhoben. Schließlich wird die Charakterisierung eines typischen Reliabilitätskoeffizienten sowie dessen Variabilität präsentiert – etwa mittels Lage- und Streuungsmaße und/oder einer grafischen Darstellung der Verteilung (siehe beispielsweise Vacha-Haase, 1998; Yin & Fan, 2000). Zur Bestimmung des Einflusses der zuvor kodierten Studienmerkmale auf die Variabilität der Reliabilität können unterschiedliche statistische Verfahren zur Anwendung kommen. Vacha-Haase (1998) verwendete multiple Regressionsanalysen und die Methode der kanonischen Korrelation, andere Forscherinnen und Forscher entschieden sich beispielsweise für die Berechnung von bivariaten Korrelationen (z.B. Viswesvaran & Ones, 2000) oder nutzten die Möglichkeit von Mittelwertvergleichen (z.B. Rouse, 2007). Empfehlungen zur Durchführung von Reliability Generalization-Studien finden sich zum Beispiel bei Henson und Thompson (2002). Demnach sollte die Suche nach Primärquellen weder zu breit, noch zu eng angelegt sein. Im ersten Fall wäre die Anzahl der gefundenen, aber irrelevanten Studien möglicherweise zu groß. Im zweiten Fall könnten eventuell potentiell relevante Quellen nicht identifiziert werden. Insbesondere aber weisen Henson und Thompson (2002) auf die Bedeutung der Organisation und der Planung des Forschungsprozesses hin. Im Rahmen einer Reliability Generalization-Studie ist mit einer großen Zahl an Primärstudien und Informationen zu

rechnen, wodurch eine transparente und lückenlose Dokumentation des gesamten Forschungsprozesses wichtig sei, vor allem dann, wenn mehrere Forscherinnen und Forscher an einem derartigen Projekt beteiligt sind.

1.2.4. Kritische Punkte

Aus der Abhängigkeit von berichteten Daten aus Primärquellen ergeben sich zwei Einschränkungen bei der Durchführung einer Reliability Generalization-Studie. Beide Punkte wurden weiter oben bereits genauer erörtert und seien daher an dieser Stelle nur kurz erwähnt. Erstens können nur jene Reliabilitätskoeffizienten in die Analyse einfließen, welche für die jeweils eigenen Daten von den Autorinnen und Autoren berichtet wurden. Zweitens können nur jene Studienmerkmale als potentielle Einflussfaktoren berücksichtigt werden, für welche entsprechende Angaben in den jeweiligen Studien gemacht wurden.

Vacha-Haase (1998) diskutierte das Thema Gewichtung in ihrer Präsentation der Reliability Generalization-Methode nicht. Ergebnisse, welche anhand größerer Stichproben gewonnen wurden, sind nach Yin & Fan (2000) statistisch stabiler als jene, welche aus kleineren Stichproben stammen. In der von ihnen durchgeführten Reliability Generalization-Studie wurden alle Reliabilitätskoeffizienten entsprechend ihrer zu Grunde liegenden Stichprobengröße gewichtet. Verschiedene Methoden zur Gewichtung von unterschiedlichen Arten von Reliabilitätskoeffizienten, beispielsweise mittels der Verwendung des Kehrwertes der entsprechenden Varianz der Reliabilitätsschätzung, finden sich bei Botella, Suero und Gambaro (2010). Caruso (2000) führte Berechnungen in der von ihm veröffentlichten Reliability Generalization-Studie sowohl mit gewichteten als auch mit ungewichteten Koeffizienten durch. Er konnte allerdings keine Unterschiede hinsichtlich der resultierenden Erkenntnisse feststellen. Ob eine Gewichtung vorgenommen werden soll oder nicht, stellt also den dritten kritischen Punkt in der Durchführung einer Reliability Generalization-Studie dar.

Ein weiterer, vierter, kritischer Punkt betrifft die Verteilung und Skalierung von Reliabilitätskoeffizienten. Durch Anwendung der Fisher-z-Transformation können nach Caruso, Witkiewitz, Belcourt-Dittloff und Gottlieb (2001) diese an eine Normalverteilung angenähert werden, welche eventuell auf Grund der Begrenzung des Wertbereichs von Reliabilitätskoeffizienten nicht gegeben sein kann. Beretvas und Pastor (2003) weisen darauf hin, dass die in Studien berichteten Koeffizienten meist Werte deutlich über Null und im Regelfall über .60 annehmen. Damit steigt die Wahrscheinlichkeit einer Nicht-

Normalverteilung. Darüber hinaus ist nach Sawilowsky (2000a) kein Intervallskalenniveau anzunehmen, da Reliabilitätskoeffizienten Korrelationen sind. Thompson und Vacha-Haase (2000) argumentieren allerdings, dass, obwohl formal betrachtet es sich um Korrelationen handelt, Reliabilitätskoeffizienten inhaltlich ein Verhältnis von Varianzen ausdrücken. Sie stellen somit eine quadrierte Maßeinheit dar, sind direkt als Prozentsätze zu interpretieren und sind damit als intervallskaliert anzusehen. Aus diesem Grund sprechen sie sich gegen die Anwendung von Prozeduren zur Transformation der Koeffizienten aus und bemerken, dass es notwendig ist zu „distinguish *what* we are estimating from *how* we formulate the estimates.“ (p. 185). Ob Koeffizienten vor weiteren statistischen Analysen transformiert werden oder nicht, ist also eine weitere Entscheidung, welche getroffen werden muss. Verschiedene Arten der Transformation für unterschiedliche Typen von Koeffizienten, etwa die z -Transformation nach Fisher oder die Transformation nach T_i , werden bei Botella et al. (2010) präsentiert.

Der fünfte und letzte kritische Punkt bezieht sich auf die Darstellung der Charakterisierung eines typischen Reliabilitätskoeffizienten und dessen Variabilität. Henson und Thompson (2002) sprechen sich für die Präsentation der Ergebnisse in Form von Konfidenzintervallen aus, da diese die Darstellung der wichtigsten Informationen in einer anschaulichen Weise ermöglichen. Eine ausführliche Anleitung zur Berechnung von Konfidenzintervallen für verschiedene Arten von Reliabilitätskoeffizienten findet sich bei Fan und Thompson (2001). Eine weitere Anwendungsmöglichkeit von Konfidenzintervallen wird von Bonnett (2010) dargestellt. Er schlägt deren Einsatz zur Bestimmung von Mittelwertsunterschieden hinsichtlich der Reliabilitätskoeffizienten zwischen potentiell relevanten Einflussgrößen vor, beispielsweise zur Berechnung von Unterschieden zwischen Frauen und Männern.

1.3. Ziel dieser Diplomarbeit

Die Methode der Reliability Generalization bildet einen neuen Forschungsansatz und wurde von Vacha-Haase im Jahr 1998 als offener Vorschlag ohne vorgeschriebene Durchführungsregeln präsentiert. Diese Diplomarbeit soll einen Überblick über bisher durchgeführte Reliability Generalization-Studien geben, eine Auseinandersetzung mit methodischen Aspekten der Durchführung solcher Forschungsleistungen bieten und die wichtigsten Erkenntnisse aus ebendiesen Studien präsentieren. Die vorliegende Arbeit orientiert sich entlang folgender a priori formulierter Leitfragen:

Leitfrage A: „Das Forschungsfeld“. Die Akteure dieses Forschungsfeldes sollen skizziert werden. Welche Forscherinnen und Forscher sind in diesem Feld aktiv? Insbesondere soll der Frage nachgegangen werden, ob einige wenige Personen in diesem Bereich forschen oder viele verschiedene. In welchen Fachzeitschriften erscheinen Reliability Generalization-Studien? Von besonderem Interesse ist hierbei die Frage, ob die Mehrzahl der Publikationen in jener Zeitschrift erscheint, welche die erste Reliability Generalization-Studie 1998 veröffentlichte – dies war die Zeitschrift *Educational and Psychological Measurement*. Wie entwickelte sich die Publikationstätigkeit im Zeitverlauf? Und schließlich wurde die Frage nach der Motivation der Forschungsgruppen zur Anwendung beziehungsweise Durchführung dieser Methode gestellt.

Leitfrage B: „Literatursuche und Induktion“. Da Reliability Generalization-Studien auf Daten aus Primärquellen angewiesen sind, stellt sich die Frage, wie nach diesen gesucht wurde. Wie viele Literaturdatenbanken wurden jeweils konsultiert? Welche waren diese? Zusätzlich soll gezeigt werden, wie viele relevante Primärstudien identifiziert und wie viele davon tatsächlich genutzt werden konnten. Das Ausmaß des Phänomens der oben beschriebenen Induktion soll ebenfalls dargestellt werden. Schließlich wird im Rahmen dieser Leitfrage noch untersucht, ob von Autorinnen und Autoren das Thema publication bias (z.B. Rosenberg, 2005) angesprochen wird.

Leitfrage C: „Ermittlung typischer Werte“. Hier soll gezeigt werden, wie aus Primärquellen gewonnenen Reliabilitätskoeffizienten behandelt werden, also ob für diese eine Form von Transformation (z.B. z-Transformation nach Fisher) angewandt sowie ob ein Verfahren zur Gewichtung eingesetzt wird. Die ermittelten typischen Reliabilitätskoeffizienten sowie deren Variabilität soll dargestellt werden. Diese Sekundärdaten werden wiederum deskriptivstatistisch charakterisiert, wodurch ein typisch-typischer Koeffizient (also auf Tertiärdatenniveau) ermittelt wird. Zusätzlich wird untersucht, ob Autorinnen und Autoren für die Präsentation der ermittelten typischen Koeffizienten als Darstellungsform Konfidenzintervalle wählen.

Leitfrage D: „Quellen der Variabilität der Reliabilität“. Diese vierte und letzte Leitfrage wird in zwei Fragestellungen (D1 und D2) unterteilt. D1 beschäftigt sich mit Methoden zur Analyse des Einflusses verschiedener Faktoren auf die Höhe der Reliabilitätskoeffizienten. Vacha-Haase wählte 1998 für die Demonstration der ersten Reliability Generalization-Studie zwei unterschiedliche inferenzstatistische Methoden. Es soll

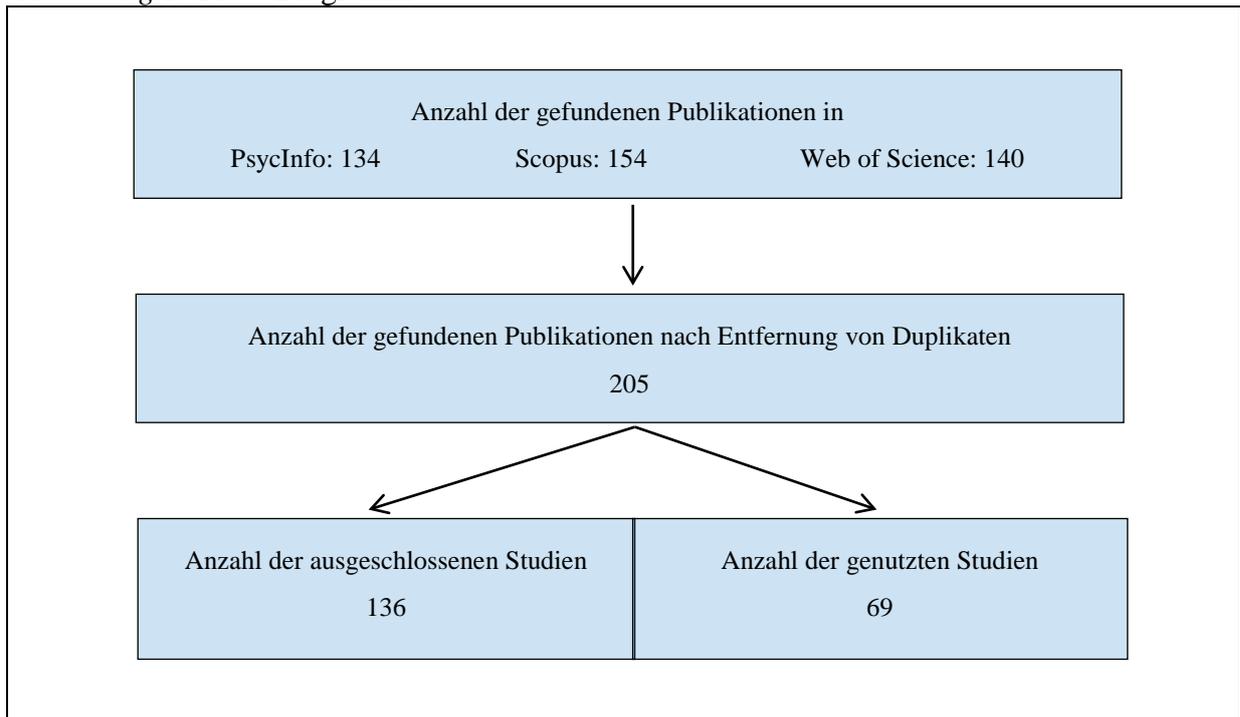
gezeigt werden, welche weiteren Methoden Forscherinnen und Forscher wählten und wie häufig diese angewandt wurden. Im Rahmen der Frage D2 sollen Erkenntnisse bezüglich der Quellen der Variabilität der Reliabilität generalisiert werden. Die Effekte besonders häufig untersuchter potentieller Einflussgrößen werden mittels Meta-Analysen untersucht. Dies stellt den Versuch dar, mögliche systematische Einflussfaktoren (unabhängig von spezifischen psychologischen Messinstrumenten) zu identifizieren.

2. Methode

2.1. Literatursuche

Eine zusammenfassende grafische Übersicht des Literatursuchprozesses findet sich in nachstehender Abbildung 1.

Abbildung 1: Flow-Diagramm



Zur Identifikation von durchgeführten Reliability Generalization-Studien wurden die wissenschaftlichen Datenbanken PsycInfo, Scopus und Web of Science konsultiert. Es wurde nach Publikationen gesucht, welche sich mittels der Suchanfrage „reliability generalization“ identifizieren lassen. Zusätzlich wurde in allen drei genannten Datenbanken mittels cited reference-Suche nach Publikationen gesucht, welche die erste veröffentlichte Reliability Generalization-Studie von Vacha-Haase aus dem Jahr 1998 zitieren. Es wurde eine Beschränkung der Suchergebnisse auf Publikationen in englischer Sprache gesetzt. Als weitere Einschränkung wurde ein zeitlicher Rahmen von 1998 bis Ende März 2011 gewählt. Die vier in Tabelle 1 letztgenannten Publikationen von Graham, Diebels und Barnow (2011), Kieffer und MacDonald (2011), Vassar, Knaup, Hale, W. und Hale, H. (2011) sowie Wheeler, Vassar, Worley und Barnes (2011) bilden somit die aktuellsten Studien, welche in der

vorliegenden Arbeit berücksichtigt wurden. Im Zuge einer erneuten Literatursuche zum Zeitpunkt der Fertigstellung dieser Diplomarbeit im Jänner 2012 konnten fünf weitere Reliability Generalization-Studien identifiziert werden, welche zwischen den beiden Suchzeitpunkten erschienen waren.

Nachfolgend finden sich die, für jede genutzte Datenbank spezifischen, Suchstrings:

- PsycInfo: 1 "reliability generalization".mp. [mp=title, abstract, heading word, table of contents, key concepts, original title, tests & measures] (86)
2 limit 1 to (english language and yr="1998 -Current") (83)
3 Vacha-Haase.cu. and Reliability generalization: Exploring variance in measurement error affecting score reliability across studies.cv. and "1998".cb. (110)
4 limit 3 to (english language and yr="1998 -Current") (108)
5 2 or 4 (134)
- Scopus: ((TITLE-ABS-KEY("reliability generalization") AND PUBYEAR AFT 1997) OR (REF(vacha-haase 1998) AND PUBYEAR AFT 1997)) AND (LIMIT-TO(LANGUAGE, "English"))
- Web of Science: Topic=("reliability generalization") [kombiniert über "OR" mit] Cited Author=(vachahaase) AND Cited Year=(1998) Timespan=1998-2011. Databases=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH.

Mit Hilfe dieser Suchstrategie konnten 134 Publikationen über PsycInfo, 154 über Scopus und 140 mittels Web of Science gefunden werden. Nach Entfernung von Duplikaten ergaben sich in Summe 205 verschiedene Publikationen. Ziel des Literatursuchprozesses war es, jene Publikationen zu finden, welche die von Vacha-Haase (1998) vorgeschlagene Methode der Reliability Generalization aufgegriffen und umgesetzt haben. Somit ergab sich als Einschlusskriterium die Verfolgung der dort angeführten Zielsetzungen. Diese sind: a) die Ermittlung eines typischen Reliabilitätskoeffizienten, b) die Charakterisierung der Variabilität dessen sowie c) die Identifikation möglicher Quellen der Variabilität der Reliabilität (Vacha-Haase, 1998). Zusätzlich wurde die Konzentration der Untersuchung auf ein konkretes Konstrukt (zum Beispiel Viswesvaran & Ones, 2000) und/oder ein bestimmtes Messinstrument und/oder eine Gruppe von Messinstrumenten (zum Beispiel Henson, Kogan, & Vacha-Haase, 2001) als Einschlusskriterium festgelegt. Bei der von Vacha-Haase (1998) vorgeschlagenen Methode handelt es sich um einen meta-analytischen Ansatz. Der

protokollierte Versuch einer umfangreichen Identifizierung relevanter Primärstudien und/oder Daten wurde somit als drittes Einschlusskriterium definiert.

Entsprechend dieser Kriterien wurden 132 (64.4%) der 205 Publikationen nicht eingeschlossen. Von den verbliebenen 73 Publikationen wurden vier ausgeschlossen. Die Dissertationen von Campbell (2003) und Rexrode (2005) wurden bereits als Fachartikel veröffentlicht und entsprechen Campbell, Pulos, Hogan und Murry (2005) beziehungsweise Rexrode, Petersen und O'Toole (2008). Die Dissertation von Shields (2003) über drei Messinstrumente ist in drei eigenständigen Publikationen erschienen und entspricht den Fachartikeln von Shields und Caruso (2003), Shields und Caruso (2004) sowie Shields, Howell, Potter und Weiss (2007). Der Kongressbeitrag von Rothrock, Barocas und Victorson (2004) ist ebenfalls als Fachartikel erschienen und entspricht Victorson, Barocas, Song und Cella (2008). Die somit verbliebenen 69 Publikationen wurden für die vorliegende Arbeit verwendet. Tabelle 1 bietet eine nach Veröffentlichungsjahr sortierte Übersicht über diese Studien sowie die jeweils untersuchten Instrumente.

Tabelle 1: Übersicht zu identifizierten Reliability Generalization-Studien

Autorinnen und Autoren	Untersuchte Instrumente/Konzepte
Vacha-Haase (1998)	Bem Sex Role Inventory
Caruso (2000)	NEO Personality Inventory
	„Big Five“ Persönlichkeitsfaktoren anhand von Daten aus den Manualen folgender Instrumente: Adjective Checklist Army Background and Life Experiences California Psychological Inventory Basic Personality Inventory Comrey Personality Scales Eysenck Personality Inventory Global Personality Inventory Goldberg Five Factor Markers Goldberg's International Personality Item Pool Guilford-Zimmerman Temperament Survey
Viswesvaran & Ones (2000)	Hogan Personality Inventory Interpersonal Style Inventory Jackson Personality Inventory Minnesota Multiphasic Personality Inventory Myers-Briggs Type Indicator Multidimensional Personality Questionnaire NEO-Personality Inventory Occupational Personality Questionnaire PDI Personality Inventory Personal Characteristics Inventory Personality Research Form Singer Loomis Personality Inventory 16 Personality Factors Wonderlic's Comprehensive Personality Profile
Yin & Fan (2000)	Beck Depression Inventory
Capraro, Capraro & Henson (2001)	Mathematics Anxiety Rating Scale
Caruso & Edwards (2001)	Junior Eysenck Personality Questionnaire
Caruso, Witkiewitz, Belcourt-Dittloff & Gottlieb (2001)	Eysenck Personality Questionnaire
Henson, Kogan & Vacha-Haase (2001)	Teacher Efficacy Scale Science Teaching Efficacy Belief Instrument Teacher Locus of Control Responsability for Student Achievement
Vacha-Haase, Kogan, Tani & Woodall (2001)	Minnesota Multiphasic Personality Inventory: klinische Skalen

Vacha-Haase, Tani, Kogan, Woodall & Thompson (2001)	Minnesota Multiphasic Personality Inventory: Lie, Infrequency und Correction
Barnes, Harp & Jung (2002)	State-Trait-Anxiety-Inventory
Beretvas, Meyers & Leite (2002)	Marlowe-Crowne Social Desirability Scale
Capraro & Capraro (2002)	Myers-Briggs Type Indicator
Deditius-Island & Caruso (2002)	Sensation Seeking Scale Form V
Hanson, Curry & Bandalos (2002)	Working Alliance Inventory
Henson & Hwang (2002)	Learning Style Inventory
Kieffer & Reese (2002)	Geriatric Depression Scale
Lane, White & Henson (2002)	Coopersmith Self-Esteem Inventory
Nilsson, Schmidt & Meek (2002)	Career Decision-Making Self-Efficacy Scale
	Adult Attachment Scale
	Bell Object Relations Inventory
Reese, Kieffer & Briggs (2002)	Inventory of Parent and Peer Attachment
	Parental Attachment Questionnaire
	Parental Bonding Inventory
Thompson & Cook (2002)	LibQual
Wallace & Wheeler (2002)	Life Satisfaction Index
Gronnerod (2003)	Rohrschach Test
Shields & Caruso (2003)	Alcohol Use Disorder Identification Test
Wilder & Sudweeks (2003)	Behavior Assessment System for Children
Youngstrom & Green (2003)	Differentail Emotions Scale
Kieffer, Cronin & Fister (2004)	Alcohol Expectancy Questionnaire
O'Rourke (2004)	Center for Epidemiologic Studies-Depression Scale
Shields & Caruso (2004)	CAGE Questionnaire
Campbell, Pulos, Hogan & Murry (2005)	Hare Psychopathy Checklists
De Ayala, Vonderharr-Carlson & Kim (2005)	Beck Anxiety Inventory
Mji & Alkhateeb (2005)	Conceptions of Mathematics Questionnaire
Ross, Blackburn & Forbes (2005)	Patterns of Adaptive Learning Survey
Ryngala, Shields & Caruso (2005)	Revised Children's Manifest Anxiety Scale
Zangaro & Soeken (2005)	

Dunn, Smith & Montoya (2006)	Multicultural Counseling Inventory Multicultural Awareness Knowledge and Skills Survey Multicultural Counseling Knowledge and Awareness Scale Cross-Cultural Counseling Competence Inventory-Revised
Graham, Liu & Jeziorski (2006)	Dyadic Adjustment Scale
Hellman, Fuqua & Worley (2006)	Survey of Perceived Organizational Support
Leach, Henson, Odom & Cagle (2006)	Self-Description Questionnaire
Bachner & O'Rourke (2007)	Zarit Burden Interview
Li & Bagger (2007)	Balanced Inventory of Desirable Responding
Miller, Shields, Campfield, Wallace & Weiss (2007)	MacAndrew Alcoholism Scale Addiction Acknowledgement Scale Addiction Potential Scale
Payne, Youngcourt & Beaubien (2007)	mehrere Verfahren zu Goal Orientation
Rouse (2007)	Minnesota Multiphasic Personality Inventory-2: Personality Psychopathology Five
Shields, Howell, Potter & Weiss (2007)	Michigan Alcoholism Screening Test
Beretvas, Suizzo, Durham & Yarnell (2008)	Internality-Externality Scale Nowicki and Strickland's Internality-Externality Scale
Hellman, Muilenburg-Trevino & Worley (2008)	Just World Scale Global Belief in a Just World Scale
Pearson (2008)	Eating Disorder Inventory Eating Attitudes Test Bulimia Test
Rexrode, Petersen & O'Toole (2008)	Ways of Coping Scale Adolescent Alcohol Involvement Scale Rutgers Alcohol Problem Index
Shields, Campfield, Miller, Howell, Wallace & Weiss (2008)	Young Adult Alcohol Problem Screening Test Problem-Oriented Screening Instrument for Teenagers Personal Experience Screening Questionnaire Drug Use Screening Inventory
Vassar (2008a)	Satisfaction with Life Scale
Vassar (2008b)	South Oaks Gambling Screen
Vassar & Crosby (2008)	UCLA Loneliness Scale
Victorson, Barocas, Song & Cella (2008)	Functional Assessment of Cance Therapy-General

Graham & Christiansen (2009)	Love Attitude Scale Passionate Love Scale Triangular Love Scale Rubin
Huynh, Howell & Benet-Martinez (2009)	Dominant und Nondominant Culture Scales von: General Ethnicity Questionnaire-Abridged Stephenson Multigroup Acculturation Scale Vancouver Index of Acculturation
Lopez-Pina, Sanchez-Meca & Rosa-Alcazar (2009)	Hamilton Rating Scale fpr Depression
Miller, Byrne, Rutherford & Hansen (2009)	Perceptions of Organizational Politics Scale
Miller, Woodson, Howell & Shields (2009)	Substance Abuse Subtle Screening Inventory
Nassen (2009)	Child Behavior Checklist
Barlow & Zangaro (2010)	Anticipated Turnover Scale
Bornmann, Mutz & Daniel (2010)	Peer-review Prozess
Ha, Lee & Puig (2010)	Frost Multidimensional Perfectionism Scale
Mahapoonyanont, Krahamwong, Kochakornjarupong & Rachasong (2010)	Robert H. Ennis critical thinking concept
Vassar & Bradley (2010)	Life Orientation Test Locke-Wallace Marital Adjustment Test Kansas Marital Satisfaction Scale Quality of Marriage Index
Graham, Diebels & Barnow (2011)	Relationship Assessment Scale Marital Opinion Questionnaire Semantic Differential Scale Couple Satisfaction Index
Kieffer & MacDonald (2011)	Ways of Coping Questionnaire
Vassar, Knaup, Hale & Hale (2011)	Impact of Events Scale
Wheeler, Vassar, Worley & Barnes (2011)	Maslach Burnout Inventory

2.2. Datenerhebung

Zur Erhebung der interessierenden Daten kamen in Summe drei Kodierschemata zur Anwendung, welche nachfolgend beschrieben werden. Die für diese Diplomarbeit relevanten Daten beziehen sich einerseits auf Eigenschaften der zu analysierenden Studien, andererseits

auf in diesen Studien untersuchte Skalen. Die Anzahl der Studien war zu Beginn der Datenerhebung bekannt und somit auch die Anzahl der zu erwartenden Datenpunkte auf Studienebene. Die Anzahl der jeweils untersuchten Skalen konnte zu diesem Zeitpunkt ebenso wenig geschätzt werden, wie die Anzahl der für die meta-analytischen Berechnungen zu kodierenden Effektstärken. Aus arbeitsökonomischen Gründen sowie auf Grund der Tatsache, dass sich die zu erwartenden Datenpunkte auf unterschiedlichen Ebenen (Studien-beziehungsweise Skalenebene) bewegen, wurde die Entscheidung für die Verwendung von drei eigenständigen Schemata – und damit gegen die Entwicklung eines einzelnen Kodierschemas – getroffen.

2.2.1. Kodierschema auf Studienebene

Leitfrage A: „Das Forschungsfeld“: Die Autorinnen bzw. Autoren, das(die) untersuchte(n) Instrument(e), das Erscheinungsjahr sowie die Zeitschrift, in welcher die Publikation veröffentlicht wurde, wurden notiert. Es wurde versucht, die Forschungsmotivation der jeweiligen Forscherinnen und Forscher zu erfassen. Nach einer ersten Durchsicht verfügbarer Studien konnten drei Kategorien der Motivation identifiziert werden. Diese waren a) die Vorstellung der Reliability Generalization-Methode, b) Forschungsmotive entsprechend den von Vacha-Haase (1998) präsentierten Zielsetzungen der dort vorgestellten Reliability Generalization und damit einhergehend ein primäres Interesse an der Anwendung ebendieser Methode und schließlich c) ein primäres Interesse an dem untersuchten Instrument und dessen Bedeutung für die Praxis, woraus sich die Begründung für die Durchführung einer derartigen Studie ergibt. Hierbei ist zu beachten, dass diese drei Kategorien sich nicht notwendigerweise gegenseitig ausschließen. Es werden hier also streng genommen nicht die Kriterien einer qualitativen Inhaltsanalyse (z.B. Atteslander, 2010) erfüllt. Konnte mehr als eine passende Kategorie identifiziert werden, wurde das (nach subjektiver Lesart) dominantere Motiv zugewiesen.

Leitfrage B: „Literatursuche und Induktion“: Um den Prozess der Literatursuche abzubilden, wurden folgende Daten kodiert: Welche Datenbanken zur Literatursuche genutzt wurden sowie deren Anzahl je Studie, die Anzahl der identifizierten relevanten Primärstudien, die Anzahl jener relevanten Primärstudien, welche für die jeweilige Untersuchung genutzt werden konnten sowie die Anzahl jener Primärstudien welche ausgeschlossen werden mussten. Bei Letzteren wurde unterschieden zwischen jenen, welche nicht in die jeweilige

Untersuchung einfließen konnten, da a) das Thema Reliabilität nicht angesprochen wurde und b) jenen, welche Aussagen zum Thema Reliabilität getroffen hatten, aber nicht auf Basis der eigenen Daten. Diese Unterteilung entspricht der Definition der beiden Arten von Induktion nach Shields und Caruso (2004), wonach a) als Induktion by omission und b) als Induktion by report bezeichnet wird. Darüber hinaus wurde notiert, ob sich die Autorinnen und Autoren mit dem Thema publication bias auseinandersetzten.

Leitfrage C: „Ermittlung typischer Werte“: Es wurde kodiert, ob vor der Ermittlung der typischen Reliabilitätskoeffizienten und deren Variabilitäten sowie der Analyse potentieller Einflussgrößen auf die Variabilitäten der Reliabilitäten ein Verfahren zur Transformation der Koeffizienten angewandt und/oder ob eine Art von Gewichtung vorgenommen wurde. Die Verwendung von Konfidenzintervallen zur Ergebnisdarstellung wurde ebenfalls notiert.

Leitfrage D1: „Quellen der Variabilität der Reliabilität“: Die zur Identifikation potentieller Einflussgrößen auf die Höhe der Reliabilitätskoeffizienten angewandten statistischen Verfahren wurden ebenfalls auf Studienebene kodiert.

2.2.2. Kodierschema auf Skalenebene

Leitfrage B: „Ermittlung typischer Werte“: Die Art des Reliabilitätskoeffizienten wurde notiert, wobei sämtliche Koeffizienten, welche aus inneren Konsistenzanalysen stammen, zu einer Gruppe zusammengefasst wurden. Weitere Gruppen waren: Reliabilitätskoeffizienten aus der Retestmethode, Interrater-Reliabilität, Cohen's Kappa sowie die Gruppe „gemischt“ für jene Berechnungen, bei welchen keine Unterscheidung zwischen verschiedenen Arten von Reliabilitätskoeffizienten getroffen wurde. Die berichteten typischen Reliabilitätskoeffizienten (Mittelwerte und/oder Mediane), deren Variabilität (Standardabweichungen und/oder Minima und Maxima und/oder Ranges) sowie die Anzahl der den jeweiligen Berechnungen zu Grunde liegenden Koeffizienten aus Primärstudien wurden kodiert. Wenn keine numerischen Ergebnisse der Berechnungen angeführt wurden, wurden aus den präsentierten grafischen Darstellungen die interessierenden Informationen unter Zuhilfenahme eines Lineals abgelesen. Wurden Berechnung für Fixed-Effect- und Random-Effects-Modelle durchgeführt, wurden die Daten für das Fixed-Effect-Modell kodiert. Bei getrennt durchgeführten Analysen mit sowohl gewichteten als auch ungewichteten Koeffizienten, wurden die gewichteten Koeffizienten notiert. Zur

Gewährleistung der Zuordnung und Orientierung wurden die Bezeichnung der Skala und/oder des Instruments, die erste Autorin bzw. der erste Autor sowie das Erscheinungsjahr der Publikation notiert.

2.2.3. Kodierschema für Meta-Analysen

Leitfrage D2: Es wurden acht einzelne Meta-Analysen zu häufig untersuchten, potentiellen Quellen der Variabilität der Reliabilität durchgeführt. In Tabelle 2 werden diese in alphabetischer Reihenfolge aufgelistet. Inhaltlich vergleichbare Sachverhalte wurden in den einzelnen Reliability Generalization-Studien teilweise in deutlich unterschiedlicher Art und Weise kodiert. Ebenfalls in Tabelle 2 werden jene Varianten der Kodierungen angeführt, welche im Rahmen der hier vorliegenden Arbeit als inhaltlich je einem Themenkreis zugehörig betrachtet wurden sowie die Anzahl jener Studien, in welchen die jeweilige Art der Kodierung gewählt wurde.

Tabelle 2: Fragestellungen der Meta-Analysen

Thema	Richtung der Fragestellung	Art der Kodierung in Primärstudien	N
Alter	Je höher das Durchschnittsalter der Stichprobe ist, desto höher/niedriger ist der Reliabilitätskoeffizient.	Alter in Jahren	29
		gemischt/Volksschule/Oberstufe/Erwachsene ^a	2
		Erwachsene vs. Jugendliche	1
		Jugendliche/Erwachsene/Senioren ^b	1
		unter 16 vs. Ältere	1
		34 bis 56 vs. 57 bis 75	1
Altersheterogenität	Je altersheterogener die Stichprobe ist, desto höher/niedriger ist der Reliabilitätskoeffizient.	Standardabweichung des Alters	5
		homogene vs. gemischte Gruppe	4
		Schüler vs. andere	3
		Jugendliche vs. andere	2
		Erwachsene vs. andere	1
Ethnie	Je mehr Menschen europäischer Abstammung in der Stichprobe sind, desto höher/niedriger ist der Reliabilitätskoeffizient.	Prozentsatz Weiße bzw. Minderheiten	18
		Weiße vs. Schwarze	1
		bis 50% Minderheit vs. bis 50% Weiße	1
		gemischt/Weiß/Minderheit/International ^c	1
		europäische Abstammung vs. andere	1
Geschlecht	Je mehr Frauen in der Stichprobe sind, desto höher/niedriger ist der Reliabilitätskoeffizient.	Prozentsatz Frauen bzw. Männer	30
		weiblich vs. männlich	2
		weiblich/männlich/gemischt ^d	2

Klinische Stichprobe	Je mehr Personen mit klinisch-psychologischen Auffälligkeiten in der Stichprobe vertreten sind, desto höher/niedriger ist der Reliabilitätskoeffizient.	klinische Stichprobe vs. andere Patientenstichprobe vs. andere medizinischer Befund vorhanden vs. andere	12 2 2
SD des Scores	Je höher die Standardabweichung des mittleren Scores der Stichprobe ist, desto höher/niedriger ist der Reliabilitätskoeffizient.	Standardabweichung des Score	23
Sprache	Je mehr Personen in der Stichprobe das vorgegebene Verfahren in englischer Sprache bearbeiten, desto höher/niedriger ist der Reliabilitätskoeffizient.	Englisch vs. andere Englisch/Hebräisch/europäisch/asiatisch	23 1
Stichproben- größe	Je größer der Stichprobenumfang ist, desto höher/niedriger ist der Reliabilitätskoeffizient.	Stichprobengröße	32

Anmerkung: ^a Berechnung von r aus den Mittelwerten, Standardabweichungen und N für Volks- und Oberstufenschüler versus Erwachsene

^b Berechnung von r aus η^2

^c Berechnung von r aus Mittelwerten, Standardabweichungen und N für Weiß versus Minderheit

^d Berechnung von r aus Mittelwerten, Standardabweichungen und N für männlich versus weiblich

Es wurden wieder jeweils die Bezeichnung der Skala und/oder des Instruments, die erste Autorin bzw. der erste Autor, das Erscheinungsjahr der Publikation sowie die Anzahl der den jeweiligen Berechnungen zu Grunde liegenden Koeffizienten aus Primärstudien notiert. Die meta-analytischen Berechnungen wurden ausschließlich für Koeffizienten aus inneren Konsistenzanalysen durchgeführt. Als Effektstärkenmaß wurde r gewählt. Andere als in r angeführte Maße wurden mit Hilfe nachfolgend genannter Strategien (vergleiche z.B. Cooper, 2010; Lipsey & Wilson, 2001) entsprechend transformiert. In Klammer finden sich die Studien, bei welchen die jeweilige Vorgehensweise zur Anwendung kam.

- Berechnung aus t -Werten und den dazugehörigen Freiheitsgraden (Beretvas et al., 2002; Beretvas et al., 2008; Caruso & Edwards, 2001; Caruso et al., 2001; Rexrode et al., 2008; Wallace & Wheeler, 2002).

- Berechnung aus Mittelwerten, Standardabweichungen und Stichprobengrößen (Lane et al., 2002; Leach et al., 2006; Lopez-Pina et al., 2009 berichteten keine Standardabweichungen, aber Konfidenzintervalle für die einzelnen Gruppenmittelwerte, woraus die Standardabweichungen berechnet werden konnten; Pearson, 2008).
- Berechnung aus Cohen's d (Rouse, 2007).
- Berechnung aus Strukturkoeffizienten (Capraro, M. M. et al., 2001; Bachner & O'Rourke; 2007; Graham et al., 2006; Henson & Hwang; 2002; Kieffer & MacDonald, 2001; Leach et al., 2006; O'Rourke, 2004).

Die ersten drei der bisher genannten Berechnungsmethoden wurden mit Hilfe des unter <http://www.polyu.edu.hk/mm/effectsizafaqs/calculator/calculator.html> frei verfügbaren Effektstärkenrechners von Professor Paul D. Ellis der Hong Kong Polytechnic University durchgeführt. Die vierte Methode beruht auf einer Umformung der Formel zur Berechnung von Strukturkoeffizienten (Courville & Thompson, 2001 auch für weiterführende Informationen). Folgende weitere Strategien wurden angewandt:

Yin & Fan (2000) berichteten ausschließlich η^2 . Hierbei handelt es sich um ein Bestimmtheitsmaß, welches analog zu r^2 interpretiert werden kann (Bortz & Döring, 2006). Es wurden die Quadratwurzeln aus η^2 berechnet und die Richtungen dieser Effekte konnten dem Fließtext entnommen werden. Leach et al. (2006) berichteten teilweise ausschließlich das Bestimmtheitsmaß r^2 . Hier wurden ebenfalls die Quadratwurzeln berechnet und die Richtungen der Effekte dem Text entnommen. Eine ähnliche Vorgehensweise zur Berechnung der Effektstärke r wurde bei Lopez-Pina et al. (2009) gewählt. Hier wurden teilweise ausschließlich angepasste R^2 aus einfachen Regressionsanalysen berichtet. Dieses konnte durch Umformen der Formel zur Ermittlung des angepassten R^2 in eine einfache bivariate Korrelation transformiert werden. Wallace und Wheeler (2002) berichteten für nicht signifikante Ergebnisse bivariater Korrelationsanalysen keine numerischen Daten. Hier wurde die Annahme $r = 0$ getroffen. Bachner und O'Rourke (2007) untersuchten den Einfluss der Sprache, in welcher das Instrument vorgegeben wurde, in drei Gruppen (Hebräisch, europäische und asiatische Sprachen jeweils kontrastiert mit Englisch). Aus den Effektstärken der drei einzelnen Berechnungen wurde ein gewichtetes Mittel gebildet. Hanson et al. (2002) teilten die Variable Stichprobengröße in zwei Subgruppen, eine allgemeine Analyse wurde

nicht durchgeführt. Aus den angegebenen beiden bivariaten Korrelationen wurde ebenfalls ein gewichtetes Mittel gebildet.

Wurden keine Angaben zur Stichprobengröße getätigt (also der den Analysen zu Grunde liegenden Anzahl der Primärdaten), wurde die Anzahl der in die Ermittlung des typischen Reliabilitätskoeffizienten eingeflossenen Primärdaten angenommen (Kieffer et al., 2004 und O'Rourke, 2004). Hellman et al. (2006) berichteten den Stichprobenumfang für lediglich eine Korrelationsanalyse. Diese Anzahl wurde für alle anderen Analysen angenommen.

Kieffer und MacDonald (2011) berichteten für ihre durchgeführten multiplen Regressionsanalysen das Bestimmtheitsmaß R^2 sowie das angepasste R^2 . Durch Umformen der Formel zur Berechnung des angepassten R^2 (Backhaus, Erichson, Plinke & Weiber, 2011, S. 76) und Einsetzen der bekannten Variablen, konnten die Stichprobengrößen ermittelt werden (siehe Formel 1).

Formel 1

$$N = \frac{R^2 + k - aR^2 * k - aR^2}{R^2 - aR^2}$$

Anmerkung: R^2 = Bestimmtheitsmaß; k = Zahl der Studien; a = Zahl der Prädiktorvariablen

2.3. Auswertung

Zu den Kennzahlen der Literatursuchprozesse sowie zu den ermittelten typischen Reliabilitätskoeffizienten und deren Variabilitäten wurden Lage- und Streuungsmaße berechnet. Für sämtliche anderen Variablen wurden Häufigkeiten gebildet. Als elektronische Hilfsmittel für Berechnungen und grafische Darstellungen wurden MS Excel sowie SPSS (Statistical Package for the Social Sciences) Version 19 von IBM verwendet.

Die acht einzelnen Meta-Analysen wurden mittels des Programms CMA (Comprehensive Meta-Analysis) Version 2 von Biostat durchgeführt. Aus den Analysen wurden elf Effektstärken ausgeschlossen, da diese Eigenschaften aufwiesen, welche die – von dem verwendeten Programm festgelegten – Voraussetzung verletzten. Auf Grund zu geringer Fallzahlen ($n < 4$) wurden bei Reese et al. (2002) vier, bei Shields et al. (2008) zwei und bei Nassen (2009) vier Effektstärken ausgeschlossen. O'Rourke (2004) berichtet einen Korrelationskoeffizient von 1.0. Dieser musste ebenfalls ausgeschlossen werden, da CMA diesen als fehlerhaften Wert ansieht.

3. Ergebnisse

3.1. Leitfrage A: „Das Forschungsfeld“

Tabelle 3: Autoren nach Publikationen

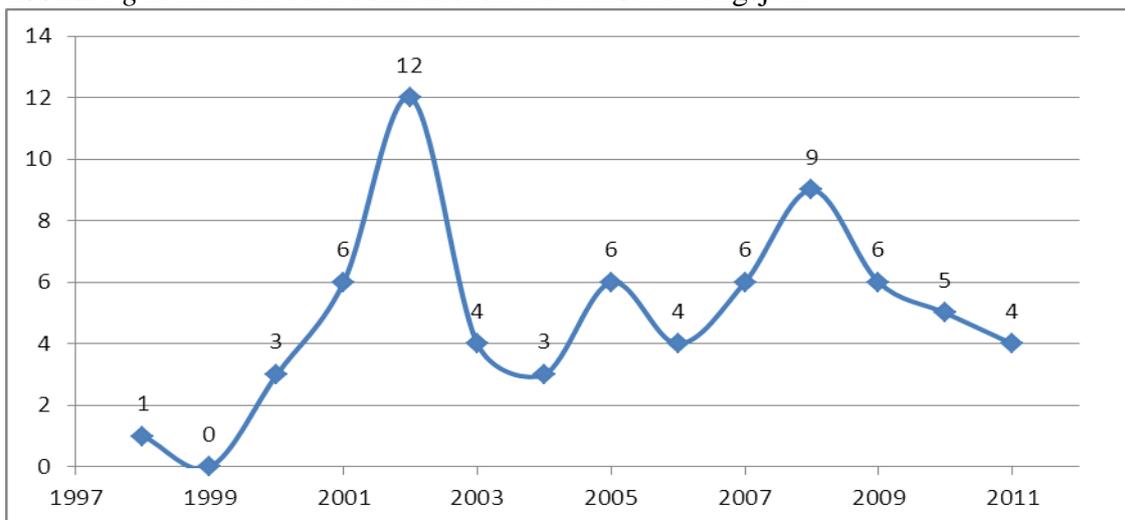
Caruso	7	Beretvas	2
Shields	7	Campfield	2
Vassar	6	Capraro, M. M.	2
Henson	5	Capraro, R. M.	2
Howell	4	Hellman	2
Kieffer	4	O'Rourke	2
Vacha-Haase	4	Reese	2
Graham,	3	Tani	2
Kogan	3	Thompson	2
Miller, C. S.	3	Weiss	2
Wallace	3	Woodall	2
Worley	3	Zangaro	2
Barnes	2		

Mit Hilfe der angewandten Suchstrategie konnten 69 Reliability Generalization-Studien identifiziert werden. Der Artikel von Zangaro und Soeken (2005) konnte leider nicht im Volltext erworben werden. Alle weiteren Ergebnisdarstellungen ab Leitfrage B beziehen sich somit auf die im Volltext verfügbaren 68 Studien. Die 69

gefundenen Publikationen wurden von insgesamt 133 Forscherinnen und Forschern verfasst. Von diesen waren 25 Personen an mehr als einer Publikation beteiligt (siehe Tabelle 3). Eine vollständige, nach Anzahl der Publikationen geordnete, Liste der Autorinnen und Autoren findet sich im Anhang.

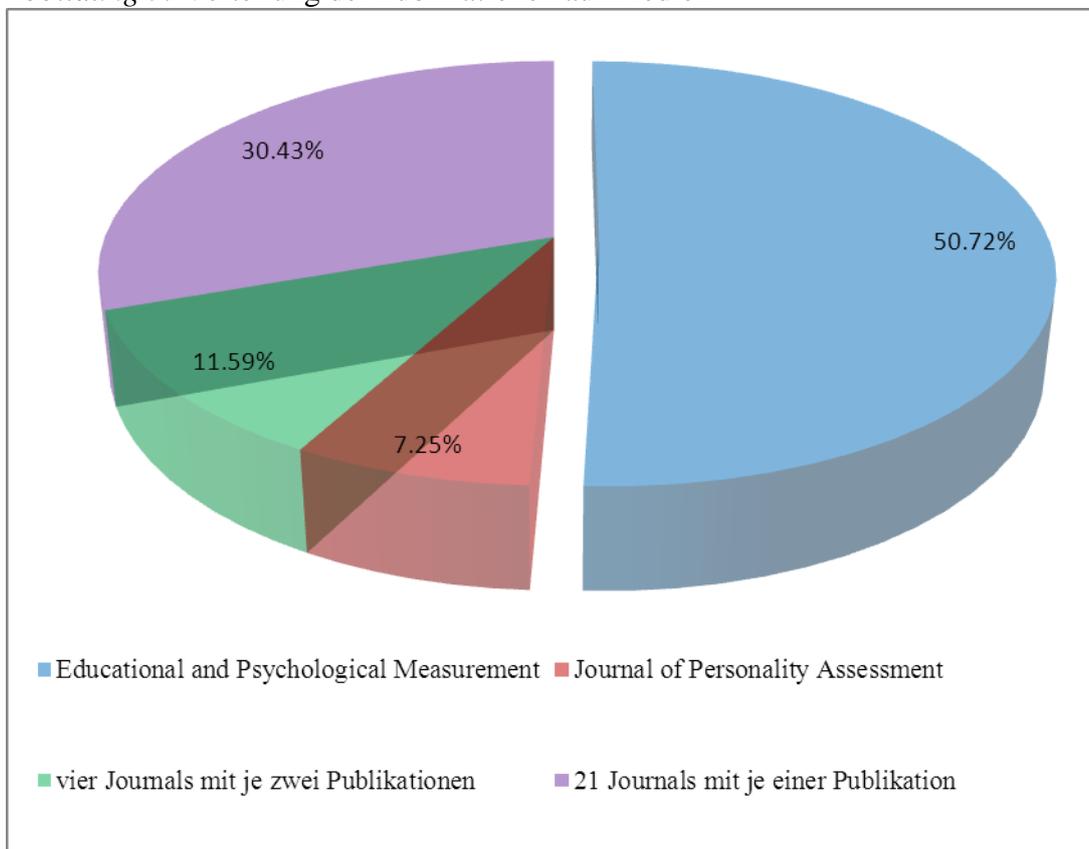
Abbildung 2 zeigt die Anzahl der publizierten Reliability Generalization-Studien im Zeitverlauf. Zu beachten ist, dass sich die Daten für das Jahr 2011 auf Veröffentlichungen bis März 2011 beziehen.

Abbildung 2: Anzahl der Publikationen nach Erscheinungsjahr



Die gefundenen Studien wurden in insgesamt 27 verschiedenen Medien veröffentlicht. 35 Reliability Generalization-Studien sind in der Zeitschrift Educational and Psychological Measurement erschienen. Somit entfallen 50.72% der Publikationen auf diese Zeitschrift. Im Journal of Personality Assessment erschienen fünf Fachartikel. Je zwei Studien erschienen in den Periodika Dissertation Abstracts International: Section B, Psychological Reports, South African Journal of Psychology und Substance Use and Misuse. Abbildung 3 zeigt die prozentuale Verteilung der Publikationen und weist insbesondere auf die Verdichtung der Publikationstätigkeit auf einige wenige Medien hin.

Abbildung 3: Verteilung der Publikationen auf Medien



Bei dem Versuch einer Charakterisierung der Forschungsmotivation der Autorinnen und Autoren zeigte sich folgendes Bild: Bei 34 der 68 Studien scheint das Interesse an der Durchführung der Methode der Reliability Generalization im Vordergrund zu stehen. Für 31 der 68 Forschungsgruppen scheint der Fokus auf der Bedeutung des jeweils untersuchten

Instruments für die praktische Anwendung zu liegen. Dreimal konnte als vordergründiges Motiv die Präsentation dieser Methode identifiziert werden. Vacha-Haase stellte im Jahr 1998 das Konzept der Reliability Generalization zum ersten Mal vor. Miller, B. K. et al. (2009) präsentierten diese Methode für die Anwendung im Bereich wirtschaftswissenschaftlicher Fragestellungen. Bornmann et al. (2010) erweiterten das Spektrum der Anwendungsmöglichkeiten auf die Erforschung von Peer-Review Prozessen.

3.2. Leitfrage B: „Literatursuche und Induktion“

Zur Identifikation relevanter Primärstudien wurden von den einzelnen Autorinnen und Autoren meist eine oder zwei Datenbanken genutzt. 23 beziehungsweise 20 Forschungsgruppen wählten diese Vorgehensweise. Drei, vier oder fünf Datenbanken wurden von jeweils acht, sechs beziehungsweise vier Forschungsgruppen konsultiert. Zweimal wurden sieben Datenbanken genutzt und in einer Studie wurde in zehn unterschiedliche Datenbanken nach relevanten Primärstudien gesucht. In vier Studien wurde keine Literatursuche mittels Datenbanken durchgeführt. Die Analysen von Viswesvaran und Ones (2000) basieren auf Daten aus 28 verschiedenen technischen Handbüchern von unterschiedlichen Persönlichkeitsinventaren. Thompson und Cook (2002) untersuchten einen online-basierten Fragebogen zur Messung der Zufriedenheit mit der Qualität von Universitätsbibliotheken in den USA und Kanada. Sämtliche Daten wurden von den 43 teilnehmenden Universitäten zur Verfügung gestellt. Ryngala et al. (2005) konnten die Daten eines Angstfragebogens nutzen, welche in 80 Schulbezirken in den USA gewonnen wurden und Hellman et al. (2006) verwendeten Daten aus einer zuvor bereits durchgeführten Meta-Analyse.

Abbildung 4 zeigt den Prozentsatz jener Studien, in welchen die zehn am häufigsten genutzten Datenbanken verwendet wurden. Eine vollständige Liste aller 28 Datenbanken und deren Nutzungshäufigkeiten findet sich im Anhang. Die drei am häufigsten konsultierten Datenbanken waren PsycInfo, Eric und PubMed/Medline. Sie wurden von 56, 25 beziehungsweise 18 der 68 Forschungsgruppen genutzt. Academic Search Elite wurde in sechs und Social Science Citation Index in fünf Studien verwendet. PsycLit und Sociological Abstracts wurden je viermal und Cumulative Index of Nursing and Allied Health Literature, EBSCO und PsycFirst wurden je dreimal genutzt. Zehn weitere Datenbanken wurden je zweimal verwendet und acht weitere je einmal.

Abbildung 4: Datenbanken nach Nutzungshäufigkeit (Top 10) in Prozent der Studien

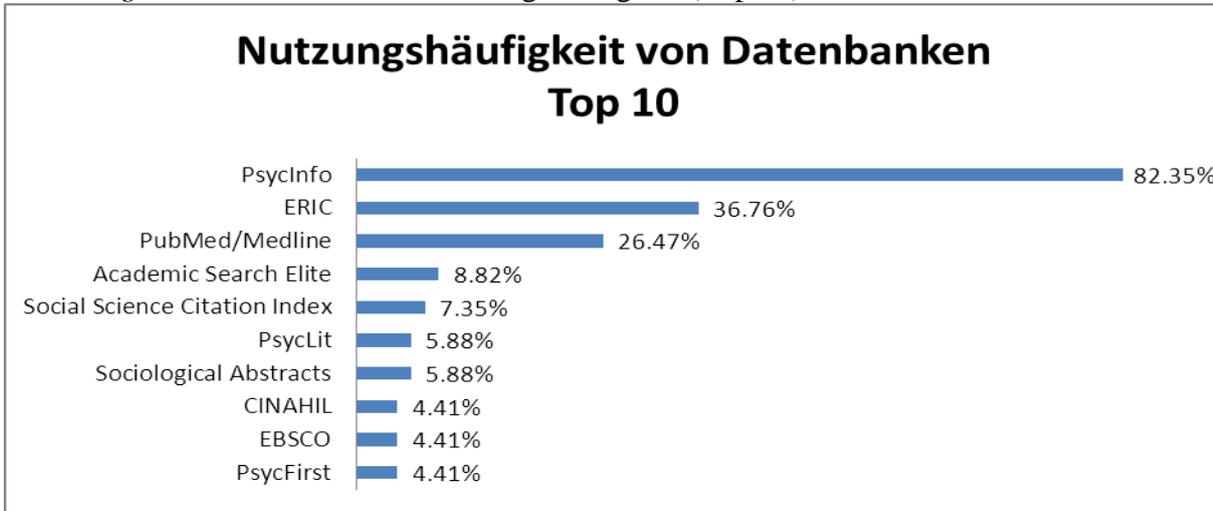


Tabelle 4 gibt einen Überblick über Daten zum Prozess der Literatursuche. Von den 64 Studien, in welchen mittels Datenbanken nach Primärstudien gesucht wurde, wurden in 61 Angaben zur Anzahl der gefundenen Studien gemacht. Der Median der gefundenen Studien liegt bei 206. Mji und Alkhateeb (2005) berichteten mit fünf Studien die kleinste Anzahl, Nassen (2009) mit 2107 die Größte. Ebenfalls 61-mal wurden Angaben zur Anzahl der für die jeweilige Untersuchung genutzten Studien gemacht. Der Median der genutzten Studien liegt bei 47. Graham et al. (2011) berichteten mit 395 die größte Anzahl. Youngstrom und Green (2003) konnten keine der mittels Datenbanksuche gefundenen Studien nutzen, da keine relevanten Daten berichtet wurden. Durch direkte Kontaktaufnahme mit Forscherinnen und Forschern konnten allerdings verwertbare Daten gewonnen werden. In 51 Studien wurden Angaben zur Anzahl jener Studien gemacht, welche keine Aussagen zum Thema Reliabilität getätigt hatten (Induktion by omission). Aus diesem Grund mussten Youngstrom und Green (2003) 100 Prozent der Studien und Nilsson et al. (2002) vier Prozent ausschließen. Der Median liegt bei 54 Prozent. Angaben zur Anzahl jener Studien, welche Aussagen zu Reliabilität auf Basis nicht eigens gewonnener Daten getroffen hatten (Induktion by report), wurden in 41 der 64 Studien getätigt. Wilder et al. (2003) mussten aus diesem Grund 85 Prozent der Studien und Reese et al. (2002) zwei Prozent ausschließen. Der Median liegt bei 17 Prozent.

Tabelle 4: Literatursuche und Induktion

	<i>N</i>	<i>1. Quartil</i>	<i>Median</i>	<i>3. Quartil</i>	<i>Min</i>	<i>Max</i>
gefundene Primärstudien	61	79.5	206	341	5	2107
genutzte Primärstudien	61	22	47	77	0	395
% genutzte Primärstudien	61	9.5%	28%	41%	0%	100%
% Induktion – by omission	51	37%	54%	67%	4%	100%
% Induktion – by report	41	8.5%	17%	25%	2%	85%

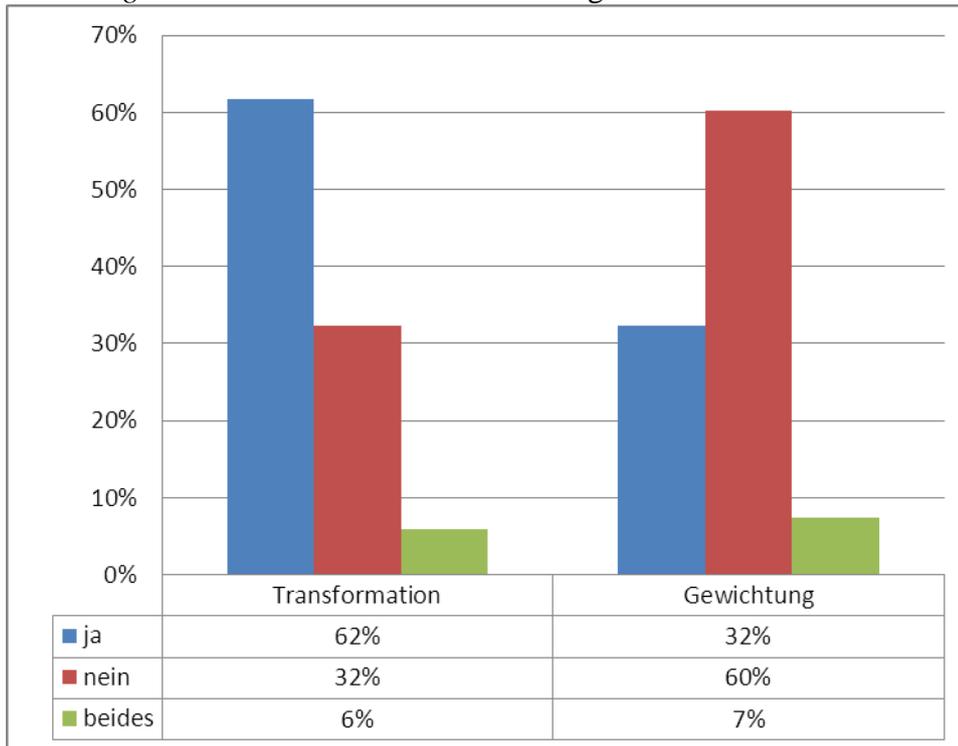
Das Thema publication bias wurde in 39 Studien nicht angesprochen. In 20 Publikationen wurde dieses Phänomen thematisiert und von neun Forschungsgruppen wurden entsprechende fail-safe-*N* Berechnungen zur Schätzung des Ausmaßes der dadurch entstehenden Verzerrung der Ergebnisse (z.B. Lipsey & Wilson, 2001) durchgeführt.

3.3. Leitfrage C: „Ermittlung typischer Werte“

In allen untersuchten Reliability Generalization-Studien wurde der Versuch der Charakterisierung von typischen Reliabilitätskoeffizienten sowie deren Variabilitäten vorgenommen. Einzige Ausnahme bildet die Arbeit von Leach et al. (2006). In Summe wurden 401-mal entsprechende Berechnungen präsentiert. Eine detaillierte Liste dazu findet sich im Anhang.

Abbildung 5 zeigt die prozentuale Verteilung der Verwendung von Verfahren zur Transformation beziehungsweise zur Gewichtung von Koeffizienten vor der Durchführung weiterer Berechnungen und Analysen.

Abbildung 5: Transformation und Gewichtung



22 Forschungsgruppen führten keine Transformation durch. 42 wandten irgendeine Art der Transformation an, wobei hier nicht nach verschiedenen Möglichkeiten der Transformation unterschieden wurde. Autorinnen und Autoren von vier Studien führten Berechnungen sowohl mit untransformierten als auch mit transformierten Koeffizienten durch. Parallele Berechnungen mit untransformierten Koeffizienten, mit quadrierten Koeffizienten sowie mit z -transformierten Koeffizienten nach Fisher wurden von Caruso et al. (2001) sowie von Shields und Caruso (2003 und 2004) durchgeführt. Es konnten keine Unterschiede hinsichtlich der Ergebnisse festgestellt werden. Die in diesen Artikeln präsentierten Ergebnisse basieren auf untransformierten Reliabilitätskoeffizienten. Ryngala et al. (2005) führten ebenfalls zusätzliche Analysen mit z -transformierten Koeffizienten nach Fisher durch. Vier von fünf mit transformierten Koeffizienten durchgeführte Regressionsanalysen unterschieden sich nicht von jenen, welche mit untransformierten Koeffizienten durchgeführt wurden. Auch in dieser Studie wurden sodann alle Ergebnisse basierend auf Letzteren dargestellt.

41 Forschungsgruppen wandten keine Art der Gewichtung an. 22 führten irgendeine Art der Gewichtung durch, wobei auch hier wieder nicht nach verschiedenen Möglichkeiten

unterschieden wurde. Fünf Autorinnen und Autoren führten Berechnungen mit sowohl gewichteten als auch ungewichteten Koeffizienten durch. Es konnten keine signifikanten Unterschiede in den Ergebnissen festgestellt werden. Weiterführende Analysen wurden sodann mit ungewichteten Koeffizienten durchgeführt (Caruso, 2000; Miller, C. S. et al., 2007; Shields et al., 2007 und 2008; Viswesvaran & Ones, 2000).

Die Tabellen 5, 6 und 7 bieten eine Zusammenfassung der Ergebnisse zu den ermittelten typischen Reliabilitätskoeffizienten sowie deren Variabilitäten. In Tabelle 8 finden sich, für die Gruppen der inneren Konsistenz und der zeitlichen Stabilität getrennt dargestellt, die Häufigkeiten der ermittelten Mittelwerte und der Standardabweichungen in Zehnerbeziehungweise Fünferschritten. Autorinnen und Autoren von 25 Studien entschieden sich bei der Ergebnisdarstellung der typischen Reliabilitätskoeffizienten für die Verwendung von Konfidenzintervallen. In 43 Studien wurde diese Möglichkeit nicht gewählt.

280 der 401 Charakterisierungen beziehen sich auf mittels innerer Konsistenzanalyse gewonnene Koeffizienten. 279-mal wurden Angaben zur Anzahl der, den Berechnungen zu Grunde liegenden, Reliabilitätskoeffizienten getätigt. Der Median liegt bei 30. Mittelwerte wurden 267-mal und Standardabweichungen 216-mal berechnet. Der Median der Mittelwerte beträgt .79, jener der Standardabweichungen .08. Mediane wurden 76-mal und Ranges 170-mal präsentiert. Der Median der Mediane liegt bei .79, jener der Ranges bei .31. Viermal wurden Berechnungen zum Koeffiziententyp KR-21 präsentiert. Der Median der Anzahl der Koeffizienten liegt bei 70.5, der Median der Mittelwerte bei .72. Auf Grund der speziellen Berechnungsmethode (z.B. Henson et al., 2001) wurde dieser Typ als eigene Gruppe geführt.

Tabelle 5: Lage- und Streuungsmaße zu Koeffizienten der inneren Konsistenz

	<i>N</i>	<i>MW</i>	<i>SD</i>	<i>MD</i>	<i>1. Quartil</i>	<i>3. Quartil</i>	<i>Range</i>
Anzahl der Koeffizienten	279	40.92	48.60	30	13	50	368
Mittelwerte	267	.79	.11	.79	.73	.87	.75
Standardabweichungen	216	.08	.04	.08	.05	.10	.25
Mediane	76	.76	.14	.79	.68	.86	.68
Range	170	.32	.19	.31	.18	.46	1.04

58 der 401 Charakterisierungen beziehen sich auf Koeffizienten, welche das Konzept der zeitlichen Stabilität beschreiben. 57-mal wurden Angaben zur Anzahl der den

Berechnungen zu Grunde liegenden Reliabilitätskoeffizienten getätigt. Der Median liegt bei 4. Mittelwerte wurden 53-mal und Standardabweichungen 44-mal berechnet. Der Median der Mittelwerte beträgt .71, jener der Standardabweichungen .11. Mediane wurden siebenmal und Ranges achtmal präsentiert. Der Median der Mediane liegt bei .56, jener der Ranges bei .46.

Tabelle 6: Lage- und Streuungsmaße zu Koeffizienten der zeitlichen Stabilität

	<i>N</i>	<i>MW</i>	<i>SD</i>	<i>MD</i>	<i>1. Quartil</i>	<i>3. Quartil</i>	<i>Range</i>
Anzahl der Koeffizienten	57	24.14	48.50	4	3	20	220
Mittelwerte	53	.71	.10	.71	.66	.78	.52
Standardabweichungen	44	.11	.06	.11	.06	.15	.22
Mediane	7	.60	.15	.56	.52	.68	.48
Range	8	.42	.18	.46	.26	.59	.53

55 der 401 Charakterisierungen stammen aus der gemischten Gruppen, in der Reliabilitätskoeffizienten, welche unterschiedliche Konzepte (beispielsweise innere Konsistenz und zeitliche Stabilität) beschreiben, gemeinsam betrachtet wurden. 55-mal wurden Angaben zur Anzahl der den Berechnungen zu Grunde liegenden Reliabilitätskoeffizienten getätigt. Der Median liegt bei 25. Mittelwerte wurden 53-mal und Standardabweichungen ebenfalls 53-mal berechnet. Der Median der Mittelwerte beträgt .80, jener der Standardabweichungen .09. Mediane wurden 15-mal und Ranges 52-mal präsentiert. Der Median der Mediane liegt bei .77, jener der Ranges bei .43.

Tabelle 7: Lage- und Streuungsmaße zu Koeffizienten aus der gemischten Gruppe

	<i>N</i>	<i>MW</i>	<i>SD</i>	<i>MD</i>	<i>1. Quartil</i>	<i>3. Quartil</i>	<i>Range</i>
Anzahl der Koeffizienten	55	33.84	31.72	25	8	49	162
Mittelwerte	53	.78	.08	.80	.72	.84	.26
Standardabweichungen	53	.09	.05	.09	.04	.13	.20
Mediane	15	.76	.07	.77	.71	.73	.27
Range	52	.40	.27	.43	.13	.68	.93

Dreimal wurden Berechnungen auf Basis von Koeffizienten des Konzepts der Interrater-Reliabilität durchgeführt. Der Median der Anzahl der Koeffizienten liegt bei 18, der

Median der Mittelwerte bei .79. Cohen's Kappa bildete in einer Studie die Berechnungsbasis ($n = 26$, $\kappa = .17$).

Der kleinste je berechnete Mittelwert wurde für die Subskala Subtle Attributes (SAT) des Substance Abuse Subtle Screening Inventory (SASSI) von Miller, C. S. et al. (2009) gefunden und beträgt .23. Der größte je berechnete Mittelwert liegt bei .98 und wurde von Reese et al. (2002) für das Inventory of Parent and Peer Attachment (IPPA), von Graham und Christiansen (2009) für die Triangular Love Scale (TLS) sowie von Graham et al. (2011) für die Semantic Differential Scale (SMD) gefunden. Die kleinste Standardabweichung wurde von Nilsson et al. (2002) für die Subskala Goal Selection der Kurzversion der Career Decision-Making Self-Efficacy Scale (CDMSE-SF) gefunden und beträgt .00. Die größte je berechnete Standardabweichung beträgt .26 und wurde von Miller, C. S. et al. (2009) für die Subskala Symptoms des Substance Abuse Subtle Screening Inventory (SASSI) ermittelt. Der größte gefundene Reliabilitätskoeffizient aus Primärquellen wurde von Capraro, M. M. et al. (2001) für die Mathematics Anxiety Rating Scale (MARS) berichtet und beträgt 1.00. Den kleinste, aus Primärquellen stammenden, Koeffizient berichteten Youngstrom und Green (2003) für die Subskala Interest der Differential Emotions Scale (DES). Er liegt bei -0.23. Dieser Wert wird durch Cronbach's Alpha Koeffizient der inneren Konsistenz beschrieben, welcher unter bestimmten Voraussetzungen auch einen negativen Wert annehmen kann. Einen ausführlichen Beweis hierfür bieten beispielsweise Henson (2001) oder Thompson (2003a).

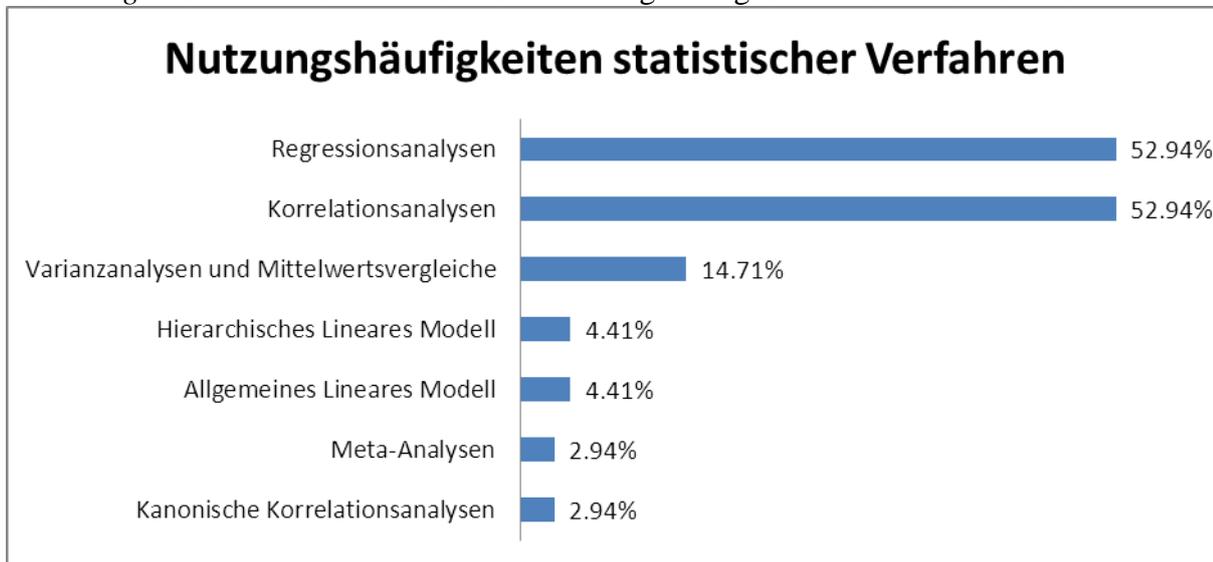
Tabelle 8: Häufigkeiten der ermittelten typischen Reliabilitäten ausgedrückt in Mittelwerten und Variabilitäten ausgedrückt in Standardabweichungen

<i>MW</i> Wertebereich	innere Konsistenz	zeitliche Stabilität	<i>SD</i> Wertebereich	innere Konsistenz	zeitliche Stabilität
über .90	39	1	über .25	1	0
.80 – .89	92	11	.20 – .24	1	6
.70 – .79	95	21	.15 – .19	8	6
.60 – .69	27	13	.10 – .14	53	13
.50 – .59	9	5	.05 – .09	105	13
.40 – .49	2	2	.00 – .04	48	6
.30 – .39	1				
.20 – .29	2				

3.4. Leitfrage D1: „Quellen der Variabilität der Reliabilität“

Abbildung 6 zeigt die prozentuale Verteilung der zur Anwendung gekommenen statistischen Verfahren zur Analyse potentieller Einflussgrößen auf die Variabilität der Reliabilität. Korrelations- sowie Regressionsanalysen wurden von je 36 Forschungsgruppen gewählt. Im Rahmen der hier vorliegenden Arbeit wurde zwischen einfachen und multiplen Regressionen nicht unterschieden. Mittelwertsvergleiche oder varianzanalytische Methoden kamen in zehn Forschungsprojekten zur Anwendung. Das Allgemeine Lineare Modell sowie das Hierarchische Lineare Modell wurden jeweils dreimal verwendet. In je zwei Studien wurden meta-analytische Verfahren beziehungsweise Kanonische Korrelationsanalysen gewählt. Zu beachten ist, dass pro Reliability Generalization-Studie mehr als ein hier genanntes Verfahren zur Anwendung kommen kann. Als Beispiel sei die Arbeit von Vassar et al. (2011) angeführt. Hier wurden in einem ersten Schritt bivariate Korrelationsanalysen angewandt um den Einfluss der einzelnen Variablen auf die Variabilität der Reliabilität zu untersuchen. Sodann wurde ein Regressionsmodell erstellt, welches als Prädiktoren jene Variablen enthielt, für welche ein signifikanter Zusammenhang mit der Höhe der Reliabilitätskoeffizienten festgestellt werden konnte.

Abbildung 6: Statistische Verfahren nach Nutzungshäufigkeit in Prozent der Studien



In sechs der 68 untersuchten Studien (8.82%) konnten keine entsprechenden Analysen durchgeführt werden, da keine beziehungsweise eine ungenügend große Anzahl an Daten aus

Primärquellen gefunden werden konnten (Capraro, R. M. & Capraro, M. M., 2002; Hellman et al., 2008; Henson & Hwang, 2002; Leach et al., 2006; Miller, C. S. et al., 2009; Mji & Alkhateeb, 2005; Wilder & Sudweeks, 2003).

3.5. Leitfrage D2: Meta-Analysen zu Quellen der Variabilität

Tabelle 9 zeigt die Ergebnisse der einzelnen Meta-Analysen unter der Annahme eines Fixed-Effect-Modells. Die einzelnen Effektstärken aus den Reliability Generalization-Studien, deren Konfidenzintervalle, Stichprobengrößen sowie die zugehörigen z - und p -Werte finden sich ebenso im Anhang wie die entsprechenden jeweiligen Gewichte für die Fixed-Effect- und Random-Effects-Modelle. Funnel plots finden sich ebenfalls im Anhang.

Tabelle 9: Ergebnisse der Meta-Analysen unter der Annahme eines Fixed-Effect-Modells

	N	Effektstärke [95% Konfidenzintervall]	Q^a	I^2	Fail-Safe- N
Alter	127	.17* [.14; .20]	363.77*	65.36	3086
Altershomogenität	59	.10* [.06; .14]	167.66*	65.41	308
Ethnie	113	.09* [.05; .13]	262.53*	57.34	367
Geschlecht	156	.13* [.10; .16]	523.58*	70.40	2173
Klinische Stichprobe	75	.17* [.13; .21]	220.36*	66.42	890
Standardabweichung des Score	49	.56* [.52; .60]	229.54*	79.09	4333
Sprache	93	.08* [.05; .11]	203.72*	54.84	557
Stichprobengröße	115	.08* [.05; .11]	204.02*	44.12	324

Anmerkung: * $p < .01$

^a df für $Q_i = N_i - 1$

Die Ergebnisse unter der Annahme eines Random-Effects-Modells lauten folgendermaßen: Für die Variable Alter (in Jahren) ergibt sich eine Effektstärke von $r = .18$ ($n = 127$, $p < .01$) und ein Konfidenzintervall von .12 bis .23. In Stichproben mit höherem Durchschnittsalter werden somit tendenziell höhere Reliabilitätskoeffizienten erzielt. Für die Variable Altershomogenität (homogene versus heterogene Gruppen) liegt das ermittelte r bei .11 ($n = 59$, $p < .01$) und das Konfidenzintervall reicht von .04 bis .18. Dieser Effekt deutet

darauf hin, dass in altersheterogenen Gruppen in der Regel höhere Reliabilitätswerte beobachtet werden können. Die Korrelation zwischen der Höhe der Reliabilität und der Variable Ethnie (Angehörige einer Minderheit versus Menschen europäischer Abstammung) beträgt $.09$ ($n = 113, p < .01$), das Konfidenzintervall dieser Effektstärke lautet $[.03; .16]$. Europäisch-stämmige Menschen erzielen also tendenziell höhere Reliabilitätswerte. Das Zusammenhangsmaß mit Geschlecht (Männer versus Frauen) liegt bei $.12$ und reicht von $.06$ bis $.17$ ($n = 165, p < .01$). Weibliche Stichproben erreichen somit in der Regel höhere Reliabilitätskoeffizienten. Für die Variable Klinische Stichprobe (gesunde Menschen versus Personen mit vorliegender klinisch-psychologischer Diagnose) ergibt sich eine Effektstärke von $r = .17$ ($n = 75, p < .01$) sowie ein Konfidenzintervall von $[.09; .25]$. In klinischen Stichproben werden also tendenziell höhere Reliabilitätskoeffizienten erreicht. Die Korrelation mit der Variable Sprache (Testvorgabe in anderer Sprache versus in englischer Sprache) beträgt $.08$ ($n = 93, p < .01$) und reicht von $.04$ bis $.13$. Wird ein Messinstrument in englischer Sprache vorgegeben, dann werden höhere Reliabilitätswerte erreicht, als wenn dasselbe Instrument in einer anderen Sprache eingesetzt wird. Der Zusammenhang zwischen der Höhe des Reliabilitätskoeffizienten und der Standardabweichung des Testscores beträgt $.57$ ($n = 49, p < .01$). Das entsprechende Konfidenzintervall lautet $[.47; .67]$. Eine höhere Standardabweichung des Scores weist auf Heterogenität des untersuchten Merkmals in der zu Grunde liegenden Stichprobe hin. Demnach sind in merkmalsheterogenen Gruppen in der Regel höhere Reliabilitätskoeffizienten zu erwarten. Für die Variable Stichprobengröße ergibt sich ein r von $.06$ ($n = 115, p < .01$) sowie ein Konfidenzintervall von $.01$ bis $.10$. Größere Stichproben erreichen also tendenziell höhere Reliabilitätswerte.

4. Diskussion

Ziel dieser Diplomarbeit war es, einen Überblick über bisher durchgeführte Reliability Generalization-Studien zu bieten und interessierten Forscherinnen und Forschern einen ersten Einblick in dieses Forschungsfeld zu geben. Testanwenderinnen und -anwender finden für die eigene Forschung relevante, reliabilitätsbezogene Hinweise sowie Verweise auf, für diese Stakeholder potentiell interessierende, Reliability Generalization-Studien. Es konnte gezeigt werden, dass diese, von Vacha-Haase (1998) als Vorschlag präsentierte Methode, in vielfältiger Weise umgesetzt und angewandt wurde. Nachfolgend werden die wichtigsten Erkenntnisse dieser Arbeit und deren Bedeutung für die wissenschaftliche Praxis entlang der vier formulierten Leitfragen diskutiert. Da es sich bei der vorliegenden Studie um eine erste Überblicksarbeit handelt, werden insbesondere Ideen für weiterführende Forschungsleistungen dargeboten.

4.1. Leitfrage A: „Das Forschungsfeld“

Die wichtigsten Autoren in diesem Feld sind Caruso, Shields, Vassar und Henson. Zu beachten ist allerdings, dass, von den insgesamt sieben Publikationen von Shields drei eigenständige Fachartikel aus einer einzigen Dissertation stammen. Die identifizierten Reliability Generalization-Studien erschienen in 27 verschiedenen Fachzeitschriften. Das Journal Educational and Psychological Measurement ist mit 35 Publikationen das bei weitem aktivste in diesem Forschungsfeld. In dieser Zeitschrift erschien auch die erste Reliability Generalization-Studie von Vacha-Haase (1998). Betrachtet man die Publikationstätigkeit im Zeitverlauf, so zeigt sich eine mehr oder minder gleichbleibende Anzahl an Veröffentlichungen pro Jahr. Dies spricht für ein anhaltendes Interesse an diesem Feld. Hinsichtlich der Motive für die Beschäftigung mit dieser Methode seien die Arbeiten von Miller, B. K. et al. (2009) und Bornmann et al. (2010) hervorgehoben. Sie erweiterten die Anwendungsmöglichkeiten dieser Methode auf zwei weitere Wissenschaftsdisziplinen.

Die innerhalb der Leitfrage A präsentierten Daten können aus Ausgangspunkt für zwei weiterführende Forschungsleistungen dienen. Für die Analyse der Zusammensetzung der Forschungsgruppen, also welche Personen gemeinsam an Projekten arbeiten, könnte das hier gelistete Verzeichnis der Autorinnen und Autoren genutzt werden. Als weitere

Forschungsfrage könnte die Entwicklung der hier berichteten Daten im Zeitverlauf formuliert werden.

4.2. Leitfrage B: „Literatursuche und Induktion“

Die Mehrheit der Forschungsgruppen wählte zur Identifikation relevanter Primärstudien eine oder zwei Literaturdatenbanken. Andere Strategien der Datenbeschaffung wurden im Abschnitt „Ergebnisse“ bereits genannt. Die Wahl des zu analysierenden Messinstruments beeinflusst vermutlich die Wahl der Datenbank. Criminal Justice Abstracts wurde beispielsweise nur einmal genutzt. Campbell et al. (2005) nutzten diese Datenbank in ihrer Arbeit zu einem Psychopathie-Fragebogen. Als Hauptproblem für die Durchführung einer Reliability Generalization-Studie konnte das Phänomen der Induktion (siehe Abschnitt „theoretischer Hintergrund“ und z.B. Shields & Caruso, 2004) identifiziert werden. Aus diesem Grund konnte in über drei von vier Studien weniger als die Hälfte der gefundenen Primärstudien in die Analysen einbezogen werden. Eine innovative Methode dieses Problem zu bearbeiten, präsentierten Henson et al. (2001). Bei Kenntnis des Mittelwerts, der Standardabweichung und der Anzahl der Items eines Tests ist die Schätzung der unteren Schranke der Reliabilität bei dichotomen Items mittels der Formel KR-21 möglich. Diese Möglichkeit wurde von Kieffer und Reese (2002) sowie von Lane et al. (2002) aufgegriffen. Diese Vorgehensweise löst das Problem der Induktion, zumindest teilweise, für Messinstrumente mit dichotomen Items, da nach Lane et al. (2002) Autorinnen und Autoren entsprechende deskriptiv-statistische Angaben häufiger präsentieren würden, als Reliabilitätskoeffizienten für deren eigene Daten. Die Anwendung der Formel KR-21 könnte demnach „dramatically increase the pool of usable studies“ (Lane et al., 2002, p. 688).

Rund um die in Leitfrage B besprochenen Themen könnten folgende weiterführende Forschungsarbeiten geleistet werden: Zur Bestimmung der Angemessenheit einer Induktion wurden bereits sogenannte Reliability Induction-Studien durchgeführt (beispielsweise Vacha-Haase et al., 2000; Vassar et al., 2008). Möglicherweise entwickelt sich hierzu gerade ein neues Forschungsfeld. Eine der vorliegenden Diplomarbeit ähnliche Studie könnte zu diesem Thema vorgenommen werden. Eine weitere Forschungsfrage könnte die Parameter des Prozesses der Literatursuche betreffen. Von Interesse könnte sein, wie die Wahl der Datenbanken die Ergebnisse der Suche beeinflussen und/oder welche Datenbanken in Kombination konsultiert werden.

4.3. Leitfrage C: „Ermittlung typischer Werte“

Die Mehrheit der Autorinnen und Autoren nutzte keine Form der Transformation. Auch eine Gewichtung der Koeffizienten wurde von der Mehrheit der Forschungsgruppen nicht durchgeführt. In parallelen Analysen mit transformierten und untransformierten, beziehungsweise mit gewichteten und ungewichteten Koeffizienten konnten keine relevanten Unterschiede festgestellt werden (Caruso, 2000; Caruso et al., 2001; Miller, C. S. et al., 2007; Shields & Caruso, 2003; 2004; Shields et al., 2007; 2008; Ryngala et al. 2005 Viswesvaran & Ones, 2000). Diese Befunde weisen darauf hin, dass beide Prozeduren nicht vorgenommen werden müssten. Andere Forscherinnen und Forscher sprechen sich allerdings deutlich für die Anwendung beider Prozesse aus (z.B. Botella, et al., 2010). Zu dieser Frage sind weiterführende Forschungsleistungen nötig. Die ermittelten typischen Reliabilitätskoeffizienten liegen mehrheitlich über .70. 15.36 Prozent der Koeffizienten aus inneren Konsistenzanalysen und 37.74 Prozent der mittels Retestmethode gewonnenen Kennzahlen unterschreiten diesen Wert. Reliabilitätswerte von über .70 werden für frühe Phasen der Testentwicklung und Werte ab .80 für die Anwendung zu Forschungszwecken empfohlen. Ein, für die Anwendung im Bereich der Individualdiagnostik geeignetes, Instrument sollte Reliabilitätskennzahlen ab .90 aufweisen (Henson, 2001; Huynh et al., 2009; Li & Bagger, 2007). Demnach wären lediglich 14.61 Prozent der hier berichteten Skalen hinsichtlich ihrer inneren Konsistenz für den Einsatz zur individuellen Diagnostik als geeignet einzustufen. Bezüglich der zeitlichen Stabilität trifft dies auf nur eine Skala zu. Die in Tabelle 8 dargestellten Häufigkeiten der ermittelten Variabilitäten (ausgedrückt in Standardabweichungen) verdeutlichen, dass Reliabilität keine feste Größe sein kann. Vacha-Haase (1998) formulierte als besonderes Ziel der von ihr vorgestellten Reliability Generalization die Bewusstseinsbildung unter Forscherinnen und Forschern bezüglich dieser Tatsache. Mit den hier zusammengefassten Befunden einzelner Studien wird deutlich auf diesen Umstand hingewiesen.

Inwiefern diese hier und in den einzelnen Reliability Generalization-Studien berichteten Ergebnisse von Verzerrungen betroffen sind, ist in weiterführenden Forschungsarbeiten zu klären. Das Phänomen der Induktion, insbesondere das gänzliche Unterlassen einer Diskussion der Daten hinsichtlich ihrer Zuverlässigkeit in Primärstudien, könnte diesbezüglich eine Rolle spielen.

4.4. Leitfrage D1: „Quellen der Variabilität der Reliabilität“

Verschiedene statistische Methoden zur Analyse möglicher Einflussgrößen auf die Höhe der Reliabilitätskoeffizienten konnten identifiziert werden. Die Wahl der Methode ist abhängig von den berichteten Daten aus Primärquellen. Reese et al. (2002) beispielsweise mussten auf Grund der gegebenen Datenlage univariate Analyseverfahren nutzen. Angestrebt wurde die Durchführung einer (multivariaten) Kanonischen Korrelationsanalyse. Die Art und die Anzahl der untersuchten potentiellen Einflussgrößen werden ebenfalls durch die berichteten Daten in Primärstudien, aber auch durch das zu analysierende Messinstrument bestimmt. Beispielsweise berücksichtigten Campbell et al. (2005), ob Informationen aus dem Strafregisterauszug der Studienteilnehmer genutzt wurden. Dieses Beispiel soll die Bandbreite möglicher zu untersuchender Faktoren illustrieren. Ähnliche Sachverhalte wurden in einzelnen Reliability Generalization-Studien teils sehr unterschiedlich kodiert. Dies sei am Beispiel der Variable Geschlecht verdeutlicht: Vacha-Haase (1998) bildete hierzu zwei Variablen. Einmal wurden rein weibliche Gruppen mit allen anderen verglichen, die andere Variable bestand aus einer Gruppe mit heterogenen Stichproben hinsichtlich der geschlechtlichen Zusammensetzung und einer zweiten Gruppe, welche rein weibliche oder rein männliche Stichproben enthielt. Beide Variablen waren Bestandteil eines einzigen multiplen Regressionsmodells. Diese Vorgehensweise wurde auf Grund der inhaltlichen Überschneidung der Gruppen von Sawilowsky (2000a) kritisiert. Die Bildung einer Variable, bestehend aus einer männlichen und einer weiblichen Gruppe, würde voraussetzen, dass das zu untersuchende Messinstrument in allen Primärstudien immer entweder ausschließlich Frauen oder ausschließlich Männern dargeboten wird. Da dies in den seltensten Fällen zutrifft, wurden in verschiedenen Reliability Generalization-Studien unterschiedlichste Gruppierungen versucht. Caruso und Edwards (2001) kodierten als erste Forschungsgruppe diese Variable als Prozentsatz männlicher Studienteilnehmer je Primärstudie. Diese Vorgehensweise wurde in den meisten nachfolgend durchgeführten Studien aufgegriffen. Ein ähnlicher Prozess zur Findung einer sich durchsetzenden Variante der Kodierung konnte bezüglich der Variable Ethnie beobachtet werden. Nilsson et al. (2002) drückten als erste Gruppe diese Variable als Prozentsatz europäisch-stämmiger Studienteilnehmerinnen und -teilnehmer aus. Umgekehrte Kodierungen, also ausgedrückt als Prozentsatz weiblicher Teilnehmerinnen beziehungsweise als Anteil einer Minderheit zugehöriger Personen, wurden in späteren Studien genutzt. Gründe hierfür wurden nicht genannt, es wird eine

gendergerechte respektive politisch korrekte Schreibweise vermutet. Die damit abgebildeten Sachverhalte ändern sich nicht.

Im Abschnitt „theoretischer Hintergrund“ wurde bereits beschrieben, dass unterschiedliche Arten von Reliabilitätskoeffizienten verschiedene Arten von Messfehlern beschreiben. Die gemeinsame Analyse verschiedener Koeffiziententypen ist daher nicht ratsam (z.B. Thompson, 2003b). In einigen Reliability Generalization-Studien wurden die Arten von Koeffizienten als Prädiktorvariablen in multiple Regressionsmodelle aufgenommen und zeigten innerhalb dieser Modelle bedeutsame Effekte (z.B. Vacha-Haase, 1998). Diese Vorgehensweise zeigt anschaulich, dass sie eben nicht sinnvoll ist. Die ersten durchgeführten Studien wollten unter anderem darauf hinweisen, dass unterschiedliche Koeffizienten auch unterschiedliche Sachverhalte abbilden. In späteren, nach 2002 durchgeführten, Studien wurde diese Vorgehensweise nicht mehr gewählt.

4.5. Leitfrage D2: Meta-Analysen zu Quellen der Variabilität

In acht meta-analytischen Untersuchungen wurde der Einfluss von Eigenschaften der Stichproben beziehungsweise der Messinstrumente auf die Höhe der Reliabilität ermittelt. Die Ergebnisse weisen darauf hin, dass höhere Reliabilitätswerte a) bei einem in der Regel höheren Durchschnittsalter der Stichprobe, b) bei tendenziell altersdiverseren Gruppen, c) in eher weiblichen Stichproben, d) in aus Menschen europäischer Abstammung zusammengesetzten Gruppen, e) bei Personen mit vorliegender klinisch-psychologischer Diagnose, f) bei Vorgabe des Messinstruments in englischer Sprache und g) in tendenziell größeren Stichproben zu erwarten sind. Die sieben eben genannten Faktoren weisen einen statistisch signifikanten Effekt auf die Höhe der Reliabilität auf. Allerdings zeigen alle Variablen eine nach Cohen (1992) als nicht hoch einzustufende Stärke des Effektes. Das bedeutet, dass die oben beschriebenen Einflüsse zwar vorhanden sind, allerdings einen wenig bedeutenden Beitrag zur Erklärung der Unterschiedlichkeit von Reliabilitätskoeffizienten aus einzelnen Messungen leisten können. Dieser Befund ist als erfreulich zu bewerten. Würde die Zuverlässigkeit einer Messung maßgeblich etwa vom ethnischen Hintergrund oder dem Geschlecht von Personen beeinflusst werden, wären Entscheidungen im Bereich der Individualdiagnostik oder Erkenntnisse aus der wissenschaftlichen Forschung unter ethischen und moralischen Gesichtspunkten zu hinterfragen. Wäre die Zuverlässigkeit einer Messung, etwa zur Identifikation klinischer Störungen, maßgeblich vom Vorliegen einer klinisch-

psychologischen Diagnose abhängig, würde die klinisch-psychologische Diagnostik ad absurdum geführt.

Einzig die Standardabweichung des Scores innerhalb einer Testvorgabe zeigt einen nach Cohen (1992) als hoch einzustufenden Effekt auf die Höhe der Reliabilität einer Messung. Gruppen, in welchen das zu untersuchende Merkmal in deutlich unterschiedlicher Weise ausgeprägt ist, erzielen somit Ergebnisse, welche als zuverlässiger zu bewerten sind, als jene aus merkmals-homogenen Gruppen. Dieser Befund steht im Einklang mit in verschiedenen Lehrbüchern beschriebenen Erkenntnissen bezüglich des Konzepts der Zuverlässigkeit (z.B. Bühner, 2004; Lienert & Raatz, 1994; Rost, 2004). Die Stichprobengröße weist, isoliert betrachtet, keinen besonders starken Effekt auf die Reliabilitätshöhe auf. Dies könnte den Schluss zulassen, dass eine größere Stichprobe nicht zwangsläufig auch eine größere Testwertstreuung bedeutet und somit auch nicht maßgeblich zur Steigerung der Zuverlässigkeit einer Messung beiträgt. Da allerdings alle hier präsentierten Effekte auf bivariaten Berechnungen beruhen, können keine Aussagen bezüglich eventuell vorliegender Wechselwirkungen zwischen einzelnen Variablen getätigt werden. Weiter Forschungsleistungen in diesem Bereich sind nötig.

4.6. *Einschränkungen dieser Arbeit*

Werden in einer Primärquelle Reliabilitätskoeffizienten für zwei oder mehr Skalen eines Messinstruments berichtet, so ist es denkbar, dass diese Werte aus nur einer Stichprobe stammen. Somit ist die Unabhängigkeit der Beobachtungen voneinander nicht mehr als gegeben anzusehen. Der Umgang mit diesem Umstand wurde bisher noch nicht behandelt. Statistisch wurde dieses Thema in nur drei untersuchten Reliability Generalization-Studien berücksichtigt. Beretvas et al. (2002 und 2008) verwendeten in ihren Analysen hierarchische lineare Modelle und Bornmann et al. (2010) nutzten die Methode der multilevel Meta-Analyse, um diesen Faktor in ihren Studien miteinzubeziehen. In einer Monte Carlo-Simulationsstudie konnten Romano und Kromrey (2008) allerdings zeigen, dass die Ergebnisse, welche mit fünf unterschiedlichen Ansätzen zum Umgang mit der Verletzung der Unabhängigkeit von Beobachtungen erzielt wurden, sich nicht nennenswert unterscheiden. Basierend auf diesem Befund wurden auch in den hier durchgeführten Meta-Analysen alle Beobachtungen als unabhängig behandelt.

Leitfrage D wurde ursprünglich anders formuliert und lautete „Gelingt es in Reliability Generalization-Studien die Schwankung der Reliabilitätskoeffizienten zu erklären und, falls ja, in welchem Ausmaß und wodurch?“. Die Annahme war, dass die meisten Autorinnen und Autoren multiple Regressionsmodelle, analog zur Vorgehensweise von Vacha-Haase (1998), nutzen würden. Eine Darstellung der Varianzerklärung der Modelle (mittels adj. R^2) sowie deren Zusammensetzung, also den jeweils enthaltenen Prädiktorvariablen, sollte diese Frage klären. Dieses Vorhaben wurde aus folgenden Gründen aufgegeben: In weniger als der Hälfte aller Studien wurden multiple Regressionsmodelle aufgestellt, wodurch die Anzahl der zu untersuchenden Studien drastisch eingeschränkt worden wäre. Die in den jeweiligen Modellen enthaltenen Prädiktorvariablen unterschieden sich oft deutlich und inhaltlich gleichbedeutende Prädiktoren wurden teilweise in erheblich unterschiedlicher Art und Weise berücksichtigt. Zusätzlich wurden in manchen Studien verschiedene Arten von Reliabilitäten gemeinsam untersucht. Die potentielle Aussagekraft einer derartigen Darstellung wurde als gering eingeschätzt und als Alternative wurde die Entscheidung getroffen, mehrere Meta-Analysen zu häufig betrachteten Einflussgrößen durchzuführen. Der Nachteil dieser Vorgehensweise liegt in der bivariaten Betrachtung der einzelnen Variablen, womit durch Faktorenkombinationen eventuell auftretende Effekte nicht untersucht werden konnten.

4.7. Kritik an der Methode der Reliability Generalization

Die Grundidee hinter der Methode ist die Betrachtung von Reliabilität als Eigenschaft von Daten und damit die (auch verbale) Abkehr von der Betrachtung derselben als Eigenschaft eines Tests (Thompson, 1994; Vacha-Haase, 1998). Diese Sicht wurde von Sawilowsky (2000a) in Frage gestellt. Er weist darauf hin, dass Reliabilität sehr wohl auch eine Eigenschaft eines Tests ist, denn falls dem nicht so wäre, „why should it be considered when choosing a standardized test?“ (p. 166). Zur Illustration der Annahme, Reliabilität könne keine testspezifische Eigenschaft sein, wird von Thompson und Vacha-Haase (2000) folgendes Beispiel genannt: Ein Verfahren zur Bestimmung des Mandarin-Wortschatzes würde, bei Vorgabe an eine Gruppe chinesischer und eine Gruppe US-amerikanischer Schüler, unterschiedliche Reliabilitätswerte produzieren. Dieses Beispiel wurde von Sawilowsky (2000b) scharf kritisiert, verletzt ein derartiges Vorgehen doch unterschiedliche andere testpsychologische Gütekriterien und ignoriert die Tatsache, dass Messinstrumente immer auch für einen bestimmten Zweck entwickelt werden und auch nur zu diesem sinnvoll

eingesetzt werden können. Ich halte auch die strenge verbale und konzeptionelle Abtrennung der Reliabilität von einem Messinstrument für unangebracht. In mehreren Lehrbüchern (z.B. Lienert & Raatz, 2004; Rost, 2004) wird darauf hingewiesen, dass Reliabilität keine feste Größe und unter anderem von den zu Grunde liegenden Daten abhängig ist. Jedoch werden diese Daten auch immer mit einem Instrument gewonnen, sind also von diesem Instrument abhängig. Wenn Reliabilität datenspezifisch ist und diese Daten testabhängig sind, dann muss meiner Ansicht nach Reliabilität *auch* testspezifischer Natur sein.

4.8. Fazit

Diese Diplomarbeit konnte zeigen, dass die Methode der Reliability Generalization einen vielfältigen Tätigkeitsbereich darstellt. Sie bietet ein geeignetes Mittel zur Bewusstseinsbildung bezüglich der Schwankung von Reliabilitätskoeffizienten eines Messinstruments über verschiedene Testvorgaben hinweg. Die Auseinandersetzung mit dieser Methode trägt dazu bei, das Konzept der Reliabilität neu zu begreifen.

In acht meta-analytischen Untersuchungen konnte als einziger bedeutender Einflussfaktor auf die Zuverlässigkeit einer Messung ein hoher Grad an Heterogenität der untersuchten Merkmalsausprägung festgestellt werden. Es konnte gezeigt werden, dass zwischen Reliabilität und den messungsspezifischen Eigenschaften Durchschnittsalter, Altershomogenität, Geschlecht und ethnischer Hintergrund der Teilnehmerinnen und Teilnehmer, dem Vorliegen einer klinisch-psychologischer Diagnose, der Vorgabe des Instruments in englischer Sprache sowie der Stichprobengröße statistisch signifikante Zusammenhänge bestehen. Diese beruhen allerdings auf eher niedrigen Effektstärken. Daraus wurde geschlossen, dass Reliabilität keinen systematischen Beeinflussungen durch die eben genannten Merkmale unterliegt. Entsprechende Effekte können allerdings durchaus spezifisch für einzelne Verfahren vorliegen (siehe Anhang). Die hier untersuchten Einflussgrößen sollten daher in künftigen Reliability Generalization-Studien weiterhin berücksichtigt werden.

Reliability Generalization-Studien bieten nützliche und praxisrelevante Informationen für Testanwenderinnen und -anwender. Für Wissenschaftlerinnen und Wissenschaftler stellen sie ein spannendes und abwechslungsreiches Forschungsfeld dar. Henson und Thompson (2002) halten passend fest, dass „conducting an RG study is seemingly limited by two things: (a) the creativity and insightfulness of the researcher and (b) the information reported in the prior studies examined in th RG meta-analysis“ (p. 114).

4.9. *Forschungsausblick*

In den einzelnen Abschnitten des Kapitels Diskussion, in welchen die formulierten Leitfragen behandelt wurden, wurde immer auch bereits ein Ausblick auf mögliche zusätzliche oder weiterführende Forschungsfragen gegeben. An dieser Stelle seien noch mal die beiden meiner Meinung nach wichtigsten und interessantesten zukünftigen Forschungsfelder angeführt. a) Das Phänomen der Induktion wurde in dieser Arbeit besprochen und die Erkenntnisse zur Leitfrage B: „Literatursuche und Induktion“ weisen darauf hin, dass Induktion ein häufig auftretendes Problem darstellt. Die in dieser Arbeit gewonnen Daten können als Grundlage für die Durchführung von Reliability Induction-Studien nach dem Vorbild von beispielsweise Vacha-Haase et al. (2000) und Vassar et al. (2008) dienen. In einem weiteren Schritt, wenn eine ausreichend hohe Anzahl an Reliability Induction-Studien vorliegt, wäre dann die Durchführung einer dieser Diplomarbeit ähnlichen Studie interessant. b) In den hier durchgeführten Analysen der Quellen der Variabilität der Reliabilität (Leitfrage D2) wurden alle untersuchten Einflussgrößen isoliert, also mittels bivariater Untersuchungsmethoden, betrachtet. Die Erforschung von Wechselwirkungen zwischen diesen und auch anderen, hier nicht berücksichtigten, Variablen stellt eine mögliche zukünftige Forschungsleistung dar.

5. Anhang

5.1. Daten auf Studienebene mit Legende

Tabelle A 1: Daten auf Studienebene

Autorinnen und Autoren	Motive	# Datenbanken	Datenbanken	Gefundene Studien	Induktion by omission %	Induktion by report %	Genutzte Studien %	Statistische Methoden	Publication bias	Transformation	Gewichtung	Konfidenzintervalle
Vacha-Haase (1998)	3a	1	1	628	413 66	133 21	57 9	1 2	1	0	0	0
Caruso (2000)	1	1	1	244	108 44	77 32	37 15	3 4	1	1	2	0
Viswesvaran & Ones (2000)	2	0						3	0	0	2	1
Yin & Fan (2000)	1	1	2	1200	961 80	149 12	90 8	3 5	0	0	1	0
Capraro, M. M. et al. (2001)	1	2	2 3	67			17 25	1 3	1	0	0	0
Caruso & Edwards (2001)	1	1	1	325	162 50	80 25	23 7	1	0	1	0	0
Caruso et al. (2001)	1	1	1	1540	767 50	384 25	44 3	1	0	2	0	0
Henson et al. (2001)	2	2	1 3				52	3	0	0	0	0
Vacha-Haase et al. (2001a)	1	1	1	1972	1704 86	115 6	47 2	1	0	0	0	0
Vacha-Haase et al. (2001b)	1	1	1	1972	1704 86	115 6	37 2	1	0	0	0	0
Barnes et al. (2002)	1	1	1	816	595 73	175 21	45 6	3	0	0	0	0
Beretvas et al. (2002)	1	4	1 3 4 5	1069	934 87	42 4	93 9	6	1	1	1	0
Capraro, R. M. & Capraro, M. M. (2002)	1	2	2 3	210	118 56	78 37	14 7		1	0	0	0
Deditius-Island & Caruso (2002)	1	1	1	244	214 88	9 4	21 9	3	0	1	0	0

Hanson et al. (2002)	2	2	1 3	62	37 60		25 40	3	0	0	0	0
Henson & Hwang (2002)	1	2	1 3	110	59 54	15 14	34 31		0	0	0	0
Kieffer & Reese (2002)	2	1	1	338	199 59	41 12	98 29	1	0	0	0	0
Lane et al. (2002)	1	2	1 3	277	152 55	92 33	33 12	1 4	0	0	0	0
Nilsson et al. (2002)	2	2	1 3	49	2 4	27 55	20 41	3 4	0	0	0	0
Reese et al. (2002)	2	1	1	154	98 64	3 2	53 34	3	0	0	0	0
Thompson & Cook (2002)	1	0						1	0	0	0	0
Wallace & Wheeler (2002)	2	1	1	157	94 60	15 10	30 19	3 4	1	1	0	0
Gronnerod (2003)	2	1	1				23	1 3	2	1	1	0
Shields & Caruso (2003)	1	2	1 6	104	70 67	11 11	23 22	1	1	2	0	0
Wilder & Sudweeks (2003)	2	3	2 3 18	106	7 7	90 85	9 8		0	0	0	0
Youngstrom & Green (2003).	2	5	1 3 7 8 9	49	49 100		0 0	1 3 4	0	0	1	0
Kieffer et al. (2004)	2	2	1 6	71	27 38	18 25	26 37	1 3	1	0	0	1
O'Rourke (2004)	1	3	1 9 10	155	49 32		106 68	1	1	1	0	0
Shields & Caruso (2004)	1	2	1 3 6	259	227 88	13 5	13 5	3	1	2	0	0
Campbell et al. (2005)	1	7	1 3 4 6 11 12 13	51	26 51		21 41	1 4	2	1	1	1
De Ayala et al. (2005)	1	3	1 14	145	82 57	14 10	47 32	1 3	0	1	0	1
Mji & Alkhateeb (2005)	2	2	1 15	5			5 100		0	1	1	1

Ross et al. (2005)	1	2	1 3	41	11 27		30 73	1 4	0	0	0	0
Ryngala et al. (2005)	1	0						1	0	2	0	0
Dunn et al. (2006)	1	10	1 3 4 12 16 17 18 19 20 21					3 4	0	0	0	0
Graham et al. (2006)	2	1	1	403	137 34	169 42	91 23	1	2	1	1	1
Hellman et al. (2006)	2	0		62	4 7	2 3	56 90	1 3	0	0	0	1
Leach et al. (2006)	1	2	1 3	112	49 44	7 6	56 50		1	0	0	0
Bachner & O'Rourke (2007)	2	4	1 4 6 10	258	149 58		102 40	1	1	0	0	0
Li & Bagger (2007)	2	3	1 3 7	236	86 36	40 17	110 47	1 2 3	1	0	0	0
Miller, C. S. et al. (2007)	1	2	1 6	221	180 81	20 9	21 10	3	0	0	2	1
Payne et al. (2007)	2	4	1 8 23 24					3	0	0	1	0
Rouse (2007)	1	1	1	17	7 41	5 29	5 29	4	1	0	0	1
Shields et al. (2007)	1	2	1 6	454			41 9	1 3	1	0	2	1
Beretvas et al. (2008)	1	1	1	445	369 84		76 16	6	0	1	1	1
Hellman et al. (2008)	1	2	1 25	197	93 47	15 7	89 45		0	0	0	1
Pearson (2008)	2	1	1	548	256 47	114 21	152 28	5	0	0	1	1
Rexrode et al. (2008)	1	5	1 6 21 25	661			92 14	4	1	0	0	0

			26									
Shields et al. (2008)	1	2	1 6	119	43 36	27 23	49 41	3	0	0	2	1
Vassar (2008a)	1	4	1 3 6 22	196	134 68		62 32	1 3	0	0	0	1
Vassar (2008b)	2	4	1 3 6 21	121	57 47	51 42	13 11	3	1	0	0	0
Vassar & Crosby (2008)	1	3	1 3 22	213	133 62		80 38	1 3	0	0	0	0
Victorson et al. (2008)	2	1	6	344	196 57	64 19	78 23	5	2	0	1	0
Graham & Christiansen (2009)	2	1	1	264	80 30	39 15	127 48	1 3	2	1	1	1
Huynh et al. (2009)	2	1	1	66	9 14	5 8	51 77	1 3	2	0	1	1
Lopez-Pina et al. (2009)	2	1	1	206	111 54		23 11	1 3	0	1	1	1
Miller, B. K. et al. (2009)	3b	2	7 23	75			75 100	1 3	0	1	1	0
Miller, C. S. Et al. (2009)	2	2	1 6	48	27 56	8 17	13 27		1	0	0	0
Nassen (2009)	2	3	1 3 6	2107	1622 77	423 20	40 2	3	0	1	1	1
Bornmann et al. (2010)	3c	7	1 3 6 14 24 27 28	84			48 57	6 7	2	1	1	1
Ha et al. (2010)	2	3	1 3 15	145	72 50	15 10	58 40	3	0	0	0	0
Mahapoonyanont et al. (2010)	2	1	26	25			14 56	1 3	0	1	1	1
Vassar & Bradley (2010)	2	5	1 3 6 21 22	332				1 3	1	1	1	0

Barlow & Zangaro (2010)	2	5	6 7 11 12 15	18			12 67	7	2	1	1	1
Graham et al. (2011)	1	2	1 7	1031	345 33	237 23	395 38	1 3	2	1	1	1
Kieffer & MacDonald (2011)	1	1	1	130	48 37	19 15	63 48	1	1	0	0	1
Vassar et al. (2011)	2	4	1 3 6 21	298			66 22	1 3	0	0	0	0
Wheeler et al. (2011)	2	3	1 3 21	221	79 36	58 26	84 38	1	0	1	1	1

Legende 1: Kodes zu Daten auf Studienebene

Motive	
Anwendung der Methode	1
Bedeutung des Instruments für die Praxis	2
Methode vorstellen	3a
Methode für Wirtschaftswissenschaften vorstellen	3b
Methode für Peer-Review Prozess vorstellen	3c
Datenbanken	
PsycInfo	1
PsycLit	2
ERIC	3
Sociological Abstracts	4
Social Sciences Abstracts	5
PubMed/Medline	6
Social Science Citation Index (SSCI)	7
Dissertation Abstracts International	8
Sociofile	9
AARP Ageline	10
Health and Psychosocial Instruments (HAPI)	11
Cumulative Index of Nursing and Allied Health Literature (CINAHL)	12
Criminal Justice Abstracts	13
Ingenta	14
EBSCO	15
Social Work Abstracts	16
Social Science Search	17

ProQuest	18
Family and Society Studies	19
PsychArticles	20
Academic Search Elite	21
PsycFirst	22
ABI Inform	23
Web of Science	24
Lexus Nexus	25
ThaiLIS	26
Scopus	27
Google	28
Statistische Methoden	
Regressionsanalysen	1
kanonische Korrelation	2
Korrelation	3
ANOVA/t-Test	4
Allgemeines Lineares Modell	5
Hierarchisches Lineares Modell	6
Meta-Analyse	7
Publication Bias	
nicht berücksichtigt	0
thematisiert	1
Berechnungen durchgeführt	2
Transformation	
nicht angewandt	0
angewandt	1
beide Vorgehensweisen	2
Gewichtung	
nicht angewandt	0
angewandt	1
beide Vorgehensweisen	2
Konfidenzintervalle	
nicht berechnet	0
berechnet	1

5.2. Daten auf Skalenebene mit Legende

Tabelle A 2: Daten auf Skalenebene

Autorinnen und Autoren	Skala	Reliabilitätstyp	Anzahl Koeffizienten	Mittelwert	Standardabweichung	Median	Range	Minimum	Maximum
Vacha-Haase (1998)	BSRI-M	3	87			.85			
Vacha-Haase (1998)	BSRI-F	3	87			.8			
Caruso (2000)	NEO-N	3	51	.88	.07	.88	.32		
Caruso (2000)	NEO-E	3	51	.83	.09	.83	.41		
Caruso (2000)	NEO-O	3	51	.79	.13	.79	.65		
Caruso (2000)	NEO-A	3	51	.75	.1	.77	.52		
Caruso (2000)	NEO-C	3	51	.83	.08	.84	.47		
Caruso (2000)	NEO-N	1	47	.88	.07				
Caruso (2000)	NEO-E	1	47	.83	.08				
Caruso (2000)	NEO-O	1	47	.79	.13				
Caruso (2000)	NEO-A	1	47	.76	.08				
Caruso (2000)	NEO-C	1	47	.84	.84				
Caruso (2000)	NEO-N	2	4	.82	.05				
Caruso (2000)	NEO-E	2	4	.81	.15				
Caruso (2000)	NEO-O	2	4	.78	.06				
Caruso (2000)	NEO-A	2	4	.58	.11				
Caruso (2000)	NEO-C	2	4	.76	.18				
Viswesvaran & Ones (2000)	Emotionale Stabilität	2	221	.75	.1				
Viswesvaran & Ones (2000)	Extraversion	2	176	.76	.12				
Viswesvaran & Ones (2000)	Offenheit	2	139	.71	.13				
Viswesvaran & Ones (2000)	Verträglichkeit	2	119	.69	.14				
Viswesvaran & Ones (2000)	Gewissenhaftigkeit	2	193	.72	.13				
Viswesvaran & Ones (2000)	Emotionale Stabilität	1	370	.78	.11				
Viswesvaran & Ones (2000)	Extraversion	1	307	.78	.09				

Viswesvaran & Ones (2000)	Offenheit	1	251	.73	.12				
Viswesvaran & Ones (2000)	Verträglichkeit	1	123	.75	.11				
Viswesvaran & Ones (2000)	Gewissenhaftigkeit	1	307	.78	.1				
Yin & Fan (2000)	BDI	3	165	.82	.01				
Yin & Fan (2000)	BDI	1	142	.84	.01				
Yin & Fan (2000)	BDI	2	23	.69	.01				
Capraro, M. M. et al. (2001)	MARS	3	35	.9	.09		.45	.55	1
Capraro, M. M. et al. (2001)	MARS	1	28	.92	.08		.45	.55	1
Capraro, M. M. et al. (2001)	MARS	2	7	.84	.07		.23	.72	.95
Caruso & Edwards (2001)	J-EPQ Psychoticism	1	44	.68	.09	.68	.36	.51	.87
Caruso & Edwards (2001)	J-EPQ Extraversion	1	44	.73	.07	.73	.31	.56	.87
Caruso & Edwards (2001)	J-EPQ Neuroticism	1	44	.78	.08	.8	.31	.55	.86
Caruso & Edwards (2001)	J-EPQ Lie	1	44	.77	.1	.79	.36	.55	.91
Caruso et al. (2001)	EPQ Psychoticism	1	69	.66	.13	.68	.55	.36	.91
Caruso et al. (2001)	EPQ Extraversion	1	69	.82	.05	.82	.25	.68	.93
Caruso et al. (2001)	EPQ Neuroticism	1	69	.83	.04	.83	.28	.69	.97
Caruso et al. (2001)	EPQ Lie	1	65	.77	.05	.78	.29	.59	.88
Henson et al. (2001)	RSA +	1	5	.76	.03		.07	.72	.79
Henson et al. (2001)	RSA -	1	5	.84	.04		.13	.75	.88
Henson et al. (2001)	TLC I+	5	7	.74	.02		.05	.71	.76
Henson et al. (2001)	TLC I-	5	7	.7	.13		.31	.51	.82
Henson et al. (2001)	STEBI PSTE	1	13	.89	.05		.18	.74	.92
Henson et al. (2001)	STEBI STOE	1	11	.76	.03		.07	.73	.8

Henson et al. (2001)	TES PTE	1	25	.78	.06		.26	.63	.89
Henson et al. (2001)	TES GTE	1	21	.7	.07		.27	.55	.82
Vacha-Haase et al. (2001a)	MMPI Skala 1	3	49	.72	.13		.52	.4	.92
Vacha-Haase et al. (2001a)	MMPI Skala 2	3	53	.7	.17		.93	0	.93
Vacha-Haase et al. (2001a)	MMPI Skala 3	3	48	.65	.16		.75	.16	.91
Vacha-Haase et al. (2001a)	MMPI Skala 4	3	54	.66	.16		.7	.18	.88
Vacha-Haase et al. (2001a)	MMPI Skala 5	3	44	.67	.2		.8	.13	.93
Vacha-Haase et al. (2001a)	MMPI Skala 6	3	47	.64	.15		.69	.18	.87
Vacha-Haase et al. (2001a)	MMPI Skala 7	3	48	.72	.18		.76	.17	.93
Vacha-Haase et al. (2001a)	MMPI Skala 8	3	51	.73	.18		.74	.22	.96
Vacha-Haase et al. (2001a)	MMPI Skala 9	3	49	.69	.14		.74	.17	.91
Vacha-Haase et al. (2001a)	MMPI Skala 0	3	46	.81	.14		.77	.18	.95
Vacha-Haase et al. (2001b)	MMPI Skala L	3	47	.68	.16	.71	.74	.19	.93
Vacha-Haase et al. (2001b)	MMPI Skala F	3	48	.68	.18	.72	.76	.16	.92
Vacha-Haase et al. (2001b)	MMPI Skala K	3	48	.73	.13	.76	.75	.16	.91
Barnes et al. (2002)	STAI Trait	1	51	.89	.05	.9	.24	.72	.96
Barnes et al. (2002)	STAI Trait	2	7	.88	.05	.88	.12	.82	.94
Barnes et al. (2002)	STAI State	1	52	.91	.05	.92	.31	.65	.96
Barnes et al. (2002)	STAI State	2	7	.7	.2	.68	.62	.34	.96
Beretvas et al. (2002)	MCSDS	1	161	.68					
Beretvas et al. (2002)	MCSDS	2	21				.48	.38	.86
Capraro, R. M. & Capraro, M. M. (2002)	MBTI	3	70	.82	.09		.49	.48	.97
Capraro, R. M.	MBTI	1	50	.82	.08		.42	.55	.97

& Capraro, M. M. (2002)									
Capraro, R. M. & Capraro, M. M. (2002)	MBTI	2	20	.81	.1		.43	.48	.91
Capraro, R. M. & Capraro, M. M. (2002)	MBTI Extravert-Introvert	3	17	.84	.05		.21	.74	.95
Capraro, R. M. & Capraro, M. M. (2002)	MBTI Sensing-Intuitive	3	17	.84	.05		.19	.78	.97
Capraro, R. M. & Capraro, M. M. (2002)	MBTI Thinking-Feeling	3	19	.76	.12		.5	.47	.97
Capraro, R. M. & Capraro, M. M. (2002)	MBTI Judgement-Perception	3	17	.82	.07		.34	.63	.97
Deditius-Island & Caruso (2002)	SSS-V Total	3	25	.76	.1	.78	.46	.39	.85
Deditius-Island & Caruso (2002)	SSS-V Thrill and Adventure	3	22	.75	.07	.75	.27	.56	.83
Deditius-Island & Caruso (2002)	SSS-V Experience	3	24	.69	.1	.66	.49	.4	.89
Deditius-Island & Caruso (2002)	SSS-V Disinhibition	3	21	.69	.08	.71	.31	.48	.79
Deditius-Island & Caruso (2002)	SSS-V Boredom Susceptibility	3	21	.62	.16	.61	.71	.17	.88
Hanson et al. (2002)	WAI Client	1	13	.93	.04		.14	.83	.97
Hanson et al. (2002)	WAI Client Goals	1	9	.87	.05		.15	.77	.92
Hanson et al. (2002)	WAI Client Tasks	1	8	.87	.03		.1	.82	.92
Hanson et al. (2002)	WAI Client Bond	1	8	.89	.03		.08	.84	.92
Hanson et al. (2002)	WAI-S Client	1	3	.95	.03		.06	.92	.98
Hanson et al. (2002)	WAI Therapist	1	5	.91	.05		.11	.84	.95
Hanson et al. (2002)	WAI Therapist Goals	1	6	.9	.02		.06	.87	.93
Hanson et al. (2002)	WAI Therapist Tasks	1	5	.87	.05		.1	.82	.92
Hanson et al. (2002)	WAI Therapist Bond	1	5	.84	.1		.24	.68	.92
Hanson et al. (2002)	WAI-S Therapist	1	2	.93	.04		.05	.9	.95

(2002)									
Hanson et al. (2002)	WAI Observer Total	1	3	.97	.02		.04	.94	.98
Hanson et al. (2002)	WAI Observer Total	4	6	.79	.12		.3	.62	.92
Henson & Hwang (2002)	LSI Concrete Experience	1	51			.75			
Henson & Hwang (2002)	LSI Reflective Observation	1	50			.79			
Henson & Hwang (2002)	LSI Active Experimentation	1	50			.81			
Henson & Hwang (2002)	LSI Abstract Conceptualization	1	51			.8			
Henson & Hwang (2002)	LSI Concrete Experience	2	38			.4			
Henson & Hwang (2002)	LSI Reflective Observation	2	38			.52			
Henson & Hwang (2002)	LSI Active Experimentation	2	38			.55			
Henson & Hwang (2002)	LSI Abstract Conceptualization	2	38			.56			
Kieffer & Reese (2002)	GDS	3	133	.85	.09		.58	.41	.99
Kieffer & Reese (2002)	GDS	5	134	.76	.14		.88	.11	.99
Lane et al. (2002)	CSEI	1	66	.73	.14				
Lane et al. (2002)	CSEI	2	69	.55	.17				
Lane et al. (2002)	CSEI	5	506	.67	.31				
Nilsson et al. (2002)	CDMSE	3	11	.95	.04		.14	.83	.97
Nilsson et al. (2002)	CDMSE Self-Appraisal	3	3	.86	.03		.05	.83	.88
Nilsson et al. (2002)	CDMSE Occupational Information	3	3	.88	.02		.04	.85	.89
Nilsson et al. (2002)	CDMSE Goal Selection	3	3	.85	.03		.04	.81	.85
Nilsson et al. (2002)	CDMSE Planning	3	3	.87	.03		.05	.84	.89
Nilsson et al. (2002)	CDMSE Problem Solving	3	3	.84	.03		.06	.8	.86
Nilsson et al. (2002)	CDMSE Short Form	3	11	.94	.01		.05	.92	.97
Nilsson et al.	CDMSE Short Form	3	5	.74	.04		.11	.71	.82

(2002)	Self-Appraisal								
Nilsson et al. (2002)	CDMSE Short Form Occupational Information	3	5	.78	.02		.04	.78	.82
Nilsson et al. (2002)	CDMSE Short Form Goal Selection	3	5	.83	0		0	.83	.83
Nilsson et al. (2002)	CDMSE Short Form Planning	3	5	.8	.03		.06	.77	.83
Nilsson et al. (2002)	CDMSE Short Form Problem Solving	3	5	.72	.03		.06	.69	.75
Reese et al. (2002)	AAS Depend	1	7	.78	.05				
Reese et al. (2002)	AAS Depend	2	1	.71					
Reese et al. (2002)	AAS Depend	3	8	.77	.06		.12	.71	.83
Reese et al. (2002)	AAS Close	1	7	.73	.06				
Reese et al. (2002)	AAS Close	2	1	.7					
Reese et al. (2002)	AAS Close	3	8	.72	.06		.2	.63	.83
Reese et al. (2002)	AAS Anxiety	1	7	.77	.09				
Reese et al. (2002)	AAS Anxiety	2	1	.64					
Reese et al. (2002)	AAS Anxiety	3	8	.75	.09		.22	.63	.85
Reese et al. (2002)	BORI Alienation	1	7	.84	.08		.2	.7	.9
Reese et al. (2002)	BORI Social Incompetence	1	7	.75	.07		.22	.68	.9
Reese et al. (2002)	BORI Insecure	1	7	.79	.07		.3	.6	.9
Reese et al. (2002)	BORI Egocentric	1	7	.75	.1		.2	.6	.8
Reese et al. (2002)	IPPA Trust	1	12	.96	.04		.12	.83	.95
Reese et al. (2002)	IPPA Communication	1	12	.98	.05		.19	.8	.99
Reese et al. (2002)	IPPA Alienation	1	12	.81	.15		.53	.42	.95
Reese et al. (2002)	PAQ Affective Quality	1	5	.93	.03				
Reese et al. (2002)	PAQ Affective Quality	2	3	.94	.02				

Reese et al. (2002)	PAQ Affective Quality	3	8	.93	.03		.08	.88	.96
Reese et al. (2002)	PAQ Parental Fostering	1	5	.89	.07				
Reese et al. (2002)	PAQ Parental Fostering	2	3	.89	.02				
Reese et al. (2002)	PAQ Parental Fostering	3	8	.89	.05		.17	.79	.96
Reese et al. (2002)	PAQ Parental Support	1	5	.84	.07				
Reese et al. (2002)	PAQ Parental Support	2	3	.85	.06				
Reese et al. (2002)	PAQ Parental Support	3	8	.85	.06		.18	.75	.93
Reese et al. (2002)	PBI Maternal Overprotection	1	22	.83	.08				
Reese et al. (2002)	PBI Maternal Overprotection	2	5	.7	.2				
Reese et al. (2002)	PBI Maternal Overprotection	3	27	.81	.12		.52	.42	.94
Reese et al. (2002)	PBI Maternal Care	1	22	.85	.06				
Reese et al. (2002)	PBI Maternal Care	2	5	.8	.13				
Reese et al. (2002)	PBI Maternal Care	3	27	.84	.08		.28	.67	.95
Reese et al. (2002)	PBI Paternal Overprotection	1	20	.84	.06				
Reese et al. (2002)	PBI Paternal Overprotection	2	5	.68	.23				
Reese et al. (2002)	PBI Paternal Overprotection	3	25	.8	.13		.6	.32	.92
Reese et al. (2002)	PBI Paternal Care	1	20	.85	.07				
Reese et al. (2002)	PBI Paternal Care	2	5	.75	.22				
Reese et al. (2002)	PBI Paternal Care	3	25	.83	.12		.52	.43	.95
Thompson & Cook (2002)	LibQUAL	1	43	.95	.01		.05	.93	.97
Thompson & Cook (2002)	LibQUAL Service Affect	1	43	.94	.01		.05	.92	.97
Thompson & Cook (2002)	LibQUAL Library as Place	1	43	.92	.01		.05	.89	.94
Thompson & Cook (2002)	LibQUAL Personal Control	1	43	.87	.02		.09	.82	.91

Thompson & Cook (2002)	LibQUAL Information Access	1	43	.75	.04		.19	.65	.84
Wallace & Wheeler (2002)	LSI	1	34	.79	.1	.79	.56	.42	.98
Gronnerod (2003)	Rohrschach	2	26	.68					
Gronnerod (2003)	Rohrschach CS	2	11	.77					
Shields & Caruso (2003)	AUDIT	1	24	.79	.1	.81	.35	.59	.94
Youngstrom & Green (2003)	DES Fear	1	30	.77	.09		.44	.44	.88
Youngstrom & Green (2003)	DES Self-hostility	1	30	.74	.15		.53	.37	.9
Youngstrom & Green (2003)	DES Shyness	1	30	.73	.11		.51	.41	.92
Youngstrom & Green (2003)	DES Sadness	1	30	.73	.1		.34	.52	.86
Youngstrom & Green (2003)	DES Enjoyment	1	30	.71	.13		.64	.25	.89
Youngstrom & Green (2003)	DES Anger	1	30	.71	.12		.64	.23	.87
Youngstrom & Green (2003)	DES Guilt	1	30	.63	.15		.66	.21	.87
Youngstrom & Green (2003)	DES Shame	1	30	.63	.13		.61	.19	.8
Youngstrom & Green (2003)	DES Contempt	1	30	.61	.11		.46	.31	.77
Youngstrom & Green (2003)	DES Disgust	1	30	.58	.15		.59	.18	.77
Youngstrom & Green (2003)	DES Surprise	1	30	.56	.19		.87	-.02	.85
Youngstrom & Green (2003)	DES Interest	1	30	.56	.19		1.05	-.23	.82
Youngstrom & Green (2003)	DES Negative affect	1	30	.92	.02		.07	.88	.95
Youngstrom & Green (2003)	DES Hostility	1	30	.77	.08		.45	.41	.86
Youngstrom & Green (2003)	DES Positive affect	1	30	.71	.12		.53	.3	.83
Kieffer et al. (2004)	AEQ	1	22	.82	.09				
Kieffer et al. (2004)	AEQ Global Positive Changes in Experience	1	14	.8	.09				
Kieffer et al. (2004)	AEQ Sexual Enhancement	1	14	.76	.14				

Kieffer et al. (2004)	AEQ Social and Physical Pleasure	1	13	.73	.09				
Kieffer et al. (2004)	AEQ Assertiveness	1	13	.79	.15				
Kieffer et al. (2004)	AEQ Relaxation/Tension Reduction	1	14	.7	.1				
Kieffer et al. (2004)	AEQ Arousal/Interpersonal Power	1	14	.59	.22				
Kieffer et al. (2004)	AEQ	2	4	.55	.19				
Kieffer et al. (2004)	AEQ Global Positive Changes in Experience	2	3	.76	.05				
Kieffer et al. (2004)	AEQ Sexual Enhancement	2	3	.7	.07				
Kieffer et al. (2004)	AEQ Social and Physical Pleasure	2	3	.67	.03				
Kieffer et al. (2004)	AEQ Assertiveness	2	3	.72	.07				
Kieffer et al. (2004)	AEQ Relaxation/Tension Reduction	2	3	.74	.08				
Kieffer et al. (2004)	AEQ Arousal/Interpersonal Power	2	3	.45	.07				
O'Rourke (2004)	CES-D	1	130	.88	.05	.89	.31	.63	.94
Shields & Caruso (2004)	CAGE	1	22	.74	.09	.74	.38	.52	.9
Campbell et al. (2005)	PCL:YV	1	18	.85					
Campbell et al. (2005)	PCL:YV	4	18	.91					
De Ayala et al. (2005)	BAI	1		.91	.03		.12	.83	.95
De Ayala et al. (2005)	BAI	2		.66	.22		.48	.35	.83
Mji & Alkhateeb (2005)	CMQ Fragmented	1	5	.89					
Mji & Alkhateeb (2005)	CMQ Cohesive	1	5	.91					
Ross et al. (2005)	PALS Extrinsic	1	15	.68	.07				
Ross et al. (2005)	PALS Task-Goal Orientation	1	41	.79	.05				
Ross et al.	PALS Performance-	1	37	.79	.07				

(2005)	Approach								
Ross et al. (2005)	PALS Performance-Avoid	1	10	.81	.04				
Ryngala et al. (2005)	RCMAS	1	48	.79	.8	.81	.45	.42	.87
Ryngala et al. (2005)	RCMAS Physiological	1	48	.59	.13	.61	.57	.18	.75
Ryngala et al. (2005)	RCMAS Worry & Oversensitivity	1	48	.76	.06	.77	.31	.61	.92
Ryngala et al. (2005)	RCMAS Concentration	1	48	.62	.11	.63	.62	.15	.77
Ryngala et al. (2005)	RCMAS Lie	1	48	.7	.1	.72	.51	.39	.9
Dunn et al. (2006)	MCI	1	26	.86	.04		.2	.75	.95
Dunn et al. (2006)	MCI Knowledge	1	34	.76	.11		.65	.23	.88
Dunn et al. (2006)	MCI Shills	1	34	.79	.05		.33	.62	.95
Dunn et al. (2006)	MCI Awareness	1	34	.78	.07		.4	.5	.9
Dunn et al. (2006)	MCI Relationship	1	34	.65	.06		.23	.55	.78
Dunn et al. (2006)	MAKSS	1	10	.89	.02		.06	.86	.92
Dunn et al. (2006)	MAKSS Knowledge	1	19	.78	.12		.58	.32	.9
Dunn et al. (2006)	MAKSS Skills	1	17	.91	.02		.09	.87	.96
Dunn et al. (2006)	MAKSS Awareness	1	19	.64	.09		.31	.47	.78
Dunn et al. (2006)	MCKAS	1	6	.88	.03		.08	.83	.91
Dunn et al. (2006)	MCKAS Knowledge	1	9	.89	.03		.08	.85	.93
Dunn et al. (2006)	MCKAS Awareness	1	9	.8	.07		.18	.71	.89
Dunn et al. (2006)	CCCI-R	1	17	.91	.05		.18	.79	.97
Dunn et al. (2006)	alle drei Instrumente	1	322	.8	.11		.74	.23	.97
Graham et al. (2006)	DAS	1	98	.92			.38	.58	.96
Graham et al. (2006)	DAS Dyadic Consensus	1	39	.87			.2	.73	.93
Graham et al.	DAS Dyadic	1	38	.85			.26	.7	.96

(2006)	Satisfaction								
Graham et al. (2006)	DAS Dyadic Cohesion	1	35	.79			.31	.58	.89
Graham et al. (2006)	DAS Affective Expression	1	34	.71			.3	.5	.8
Hellman et al. (2006)	SPOS	1	58	.88	.1	.9	.49	.49	.98
Bachner & O'Rourke (2007)	ZBI	1	127	.86	.06	.88	.33	.62	.95
Bachner & O'Rourke (2007)	ZBI	2	11	.59	.22		.65	.24	.89
Li & Bagger (2007)	BIDR	1	18	.8	.04	.8	.18	.68	.86
Li & Bagger (2007)	BIDR Impression Management	1	107	.74	.09	.76	.56	.32	.88
Li & Bagger (2007)	BIDR Self-Deception Enhancement	1	90	.68	.09	.69	.65	.27	.92
Miller, C. S. et al. (2007)	MAC	1	26	.47	.09	.47	.41	.24	.65
Miller, C. S. et al. (2007)	APS	1	2	.48		.48			
Miller, C. S. et al. (2007)	AAS	1	2	.66		.58			
Payne et al. (2007)	Learning Goal Orientation	2	20	.66					
Payne et al. (2007)	Prove Performance Goal Orientation	2	16	.7					
Payne et al. (2007)	Avoid Performance Goal Orientation	2	4	.73					
Rouse (2007)	MMPI-2 Aggressivness	1	63	.63	.08		.43	.36	.79
Rouse (2007)	MMPI-2 Psychoticism	1	63	.77	.09		.46	.44	.9
Rouse (2007)	MMPI-2 Disconstraint	1	63	.62	.11		.57	.3	.87
Rouse (2007)	MMPI-2 Negative Emotionality/Neuroticism	1	63	.83	.06		.37	.54	.91
Rouse (2007)	MMPI-2 Introversion/Low positive Emotionality	1	63	.75	.09		.41	.47	.88
Shields et al. (2007)	MAST	1	48	.83	.09	.85	.46	.47	.93
Beretvas et al. (2008)	I-E	1	116	.67	.13	.7	.74	.19	.93
Beretvas et al. (2008)	I-E	2	14	.66	.09	.64	.33	.53	.86
Beretvas et al.	NSIE	1	36	.66	.11	.68	.49	.37	.86

(2008)									
Beretvas et al. (2008)	NSIE	2	14	.71					
Hellman et al. (2008)	JWS	1	22	.64	.08	.64	.31	.48	.79
Hellman et al. (2008)	JWSR	1	47	.68	.09	.68	.46	.38	.84
Hellman et al. (2008)	GBJW	1	20	.81	.05	.82	.24	.65	.89
Pearson (2008)	EDI Total	1	6	.91	.06				
Pearson (2008)	EDI Body Dissatisfaction	1	55	.89	.04				
Pearson (2008)	EDI Drive for Thinness	1	49	.85	.05				
Pearson (2008)	EDI Bulimia	1	47	.75	.07				
Pearson (2008)	EDI Perfectionism	1	34	.68	.08				
Pearson (2008)	EDI Interoceptive Awareness	1	31	.75	.06				
Pearson (2008)	EDI Ineffectiveness	1	29	.8	.06				
Pearson (2008)	EDI Maturity Fears	1	31	.69	.09				
Pearson (2008)	EDI Interpersonal Distrust	1	29	.71	.08				
Pearson (2008)	EDI Asceticism	1	12	.52	.1				
Pearson (2008)	EDI Impuls Regulation	1	13	.76	.05				
Pearson (2008)	EDI Social Insecurity	1	13	.72	.04				
Pearson (2008)	EDI Total	2	2	.81	.06				
Pearson (2008)	EDI Body Dissatisfaction	2	3	.77	.14				
Pearson (2008)	EDI Drive for Thinness	2	4	.6	.15				
Pearson (2008)	EDI Bulimia	2	4	.64	.12				
Pearson (2008)	EDI Perfectionism	2	3	.63	.11				
Pearson (2008)	EDI Interoceptive Awareness	2	2	.6	.16				
Pearson (2008)	EDI Ineffectiveness	2	2	.71	.09				
Pearson (2008)	EDI Maturity Fears	2	3	.42	.11				
Pearson (2008)	EDI Interpersonal Distrust	2	3	.59	.1				
Pearson (2008)	EAT-40	1	15	.81	.09				
Pearson (2008)	EAT-26	1	54	.86	.05				
Pearson (2008)	EAT-26 Dieting	1	24	.8	.07				
Pearson (2008)	EAT-26 Bulimia and Food Preoccupation	1	23	.67	.11				
Pearson (2008)	EAT-26 Oral Control	1	18	.56	.1				
Pearson (2008)	EAT-26	2	4	.87	.02				

Pearson (2008)	BULIT-R	1	16	.93	.03				
Pearson (2008)	BULIT-R	2	2	.85	.01				
Rexrode et al. (2008)	WOCS Confrontive Coping	1	80			.63			
Rexrode et al. (2008)	WOCS Distancing	1	83			.65			
Rexrode et al. (2008)	WOCS Self-Controlling	1	83			.64			
Rexrode et al. (2008)	WOCS Seeking Social Support	1	106			.75			
Rexrode et al. (2008)	WOCS Accepting Responsibility	1	78			.64			
Rexrode et al. (2008)	WOCS Escape-Avoidance	1	86			.7			
Rexrode et al. (2008)	WOCS Planful Problem Solving	1	83			.72			
Rexrode et al. (2008)	WOCS Positive Reappraisal	1	81			.76			
Shields et al. (2008)	PESQ-PS	1	5	.92	.02	.92	.06	.9	.96
Shields et al. (2008)	RAPI	1	8	.91	.04	.92	.14	.81	.95
Shields et al. (2008)	DUSI-SUS	1	6	.9	.03	.89	.12	.87	.99
Shields et al. (2008)	POSIT-ADS	1	5	.86	.03	.85	.15	.77	.92
Shields et al. (2008)	YAAPST	1	4	.85	.05	.84	.19	.72	.91
Shields et al. (2008)	AAIS	1	11	.77	.12	.79	.43	.51	.94
Shields et al. (2008)	alle sechs Instrumente	1	40	.87	.07	.88	.48	.51	.99
Vassar (2008a)	SWLS	1	77	.78	.09	.79	.38	.57	.95
Vassar (2008b)	SOGS	1	15	.86	.07	.86	.28	.69	.97
Vassar & Crosby (2008)	UCLA	1	5			.89			
Vassar & Crosby (2008)	UCLA-R	1	62	.87	.06	.89	.42	.53	.95
Vassar & Crosby (2008)	UCLA-3	1	13	.92	.03	.92	.09	.86	.95
Victorson et al. (2008)	FACT-G	1	63	.88	.03		.16	.8	.96
Victorson et al. (2008)	FACT-G Physical Wellbeing	1	68	.81	.06				
Victorson et al. (2008)	FACT-G Social Wellbeing	1	66	.71	.11				

Victorson et al. (2008)	FACT-G Functional Wellbeing	1	65	.83	.05				
Victorson et al. (2008)	FACT-G Emotional Wellbeing	1	66	.73	.08				
Graham & Christiansen (2009)	Rubin Loving	1	57	.86					
Graham & Christiansen (2009)	Rubin Liking	1	36	.89					
Graham & Christiansen (2009)	LAS Eros	1	25	.72					
Graham & Christiansen (2009)	LAS Ludus	1	25	.69					
Graham & Christiansen (2009)	LAS Storge	1	24	.71					
Graham & Christiansen (2009)	LAS Pragma	1	26	.76					
Graham & Christiansen (2009)	LAS Mania	1	23	.72					
Graham & Christiansen (2009)	LAS Agape	1	24	.8					
Graham & Christiansen (2009)	LAS-SF Eros	1	12	.75					
Graham & Christiansen (2009)	LAS -SFLudus	1	13	.7					
Graham & Christiansen (2009)	LAS -SFStorge	1	13	.83					
Graham & Christiansen (2009)	LAS -SFPragma	1	14	.74					
Graham & Christiansen (2009)	LAS -SFMania	1	13	.68					
Graham & Christiansen (2009)	LAS -SFAgape	1	16	.82					
Graham &	PLS 30 item	1	7	.93					

Christiansen (2009)									
Graham & Christiansen (2009)	PLS 15 item	1	7	.88					
Graham & Christiansen (2009)	TLS Total	1	2	.98					
Graham & Christiansen (2009)	TLS Passion	1	11	.91					
Graham & Christiansen (2009)	TLS Intimacy	1	12	.9					
Graham & Christiansen (2009)	TLS Commitment	1	12	.92					
Huynh et al. (2009)	Nondominant Alle	1	42	.84	.09	.87	.34	.63	.97
Huynh et al. (2009)	VIA Nondominant	1	14	.83	.08	.85	.26	.66	.92
Huynh et al. (2009)	SMAS Nondominant	1	12	.82	.11	.85	.34	.63	.97
Huynh et al. (2009)	GEQ Nondominant	1	16	.86	.07	.88	.29	.64	.93
Huynh et al. (2009)	Dominant Alle	1	47	.83	.09	.85	.55	.39	.94
Huynh et al. (2009)	VIA Dominant	1	13	.83	.06	.84	.19	.7	.89
Huynh et al. (2009)	SMAS Dominant	1	11	.74	.14	.77	.51	.39	.9
Huynh et al. (2009)	GEQ Dominant	1	23	.87	.04	.88	.17	.77	.94
Lopez-Pina et al. (2009)	HAM-D	1	35	.79	.14		.48	.41	.89
Miller, B.K. et al. (2009)	POPS	1	95	.81	.08	.81	.4	.54	.94
Miller, C. S. et al. (2009)	SASSI Total	1	3	.87	.12	.93	.21	.74	.95
Miller, C. S. et al. (2009)	FVA	1	11	.88	.04	.89	.15	.78	.93
Miller, C. S. et al. (2009)	FVOD	1	11	.92	.05	.93	.2	.78	.98
Miller, C. S. et al. (2009)	SYM	1	6	.65	.26	.75	.67	.14	.81
Miller, C. S. et	OAT	1	11	.59	.14	.63	.52	.24	.76

al. (2009)									
Miller, C. S. et al. (2009)	SAT	1	11	.23	.16	.25	.48	.01	.49
Miller, C. S. et al. (2009)	DEF	1	11	.52	.13	.53	.37	.29	.66
Miller, C. S. et al. (2009)	SAM	1	11	.36	.11	.32	.35	.22	.57
Miller, C. S. et al. (2009)	FAM	1	10	.26	.08	.28	.21	.15	.36
Miller, C. S. et al. (2009)	COR	1	10	.63	.11	.65	.31	.47	.78
Nassen (2009)	CBCL Total	1	37	.92					
Nassen (2009)	External	1	38	.86					
Nassen (2009)	Internal	1	39	.83					
Nassen (2009)	Anxious	1	23	.83					
Nassen (2009)	Aggressive	1	23	.85					
Nassen (2009)	Attention	1	21	.77					
Nassen (2009)	Delinquent	1	23	.74					
Nassen (2009)	Social Problems	1	22	.74					
Nassen (2009)	Somatic	1	22	.75					
Nassen (2009)	Thought problems	1	19	.68					
Nassen (2009)	Withdrawn	1	23	.76					
Barlow & Zangaro (2010)	ATS	1	12	.89					
Bornmann et al. (2010)	Peer-review	4	44	.23					
Bornmann et al. (2010)	Peer-review	6	26	.17					
Ha et al. (2010)	F-MPS	1	50	.87	.05		.22	.72	.94
Ha et al. (2010)	F-MPS Concern over Mistakes	1	58	.84	.05		.21	.72	.93
Ha et al. (2010)	F-MPS Personal Standards	1	45	.77	.08		.39	.49	.88
Ha et al. (2010)	F-MPS Parental Expectation	1	32	.77	.07		.33	.56	.89
Ha et al. (2010)	F-MPS Parental Criticism	1	30	.75	.1		.35	.54	.89
Ha et al. (2010)	F-MPS Doubt about Actions	1	46	.71	.09		.36	.53	.89
Ha et al. (2010)	F-MPS Organization	1	35	.86	.07		.28	.68	.96
Mahapoonyanont et al. (2010)	critical thinking	1	14	.88					
Vassar & Bradley (2010)	LOT Composite	1	44	.73	.09		.38	.5	.88
Vassar & Bradley (2010)	LOT Optimism	1	14	.7	.04		.12	.63	.75

Vassar & Bradley (2010)	LOT Pessimism	1	14	.76	.08		.25	.61	.86
Vassar & Bradley (2010)	LOT-R Composite	1	100	.73	.11		.52	.35	.87
Vassar & Bradley (2010)	LOT-R Optimism	1	31	.72	.07		.28	.55	.83
Vassar & Bradley (2010)	LOT-R Pessimism	1	26	.73	.08		.26	.58	.84
Graham et al. (2011)	LWMAT	1	122	.79			.51	.42	.93
Graham et al. (2011)	KMS	1	105	.95			.27	.72	.99
Graham et al. (2011)	QMI	1	189	.94			.23	.75	.98
Graham et al. (2011)	RAS	1	196	.87			.3	.68	.98
Graham et al. (2011)	MOQ	1	19	.92			.13	.83	.96
Graham et al. (2011)	SMD	1	3	.98			.01	.97	.98
Graham et al. (2011)	CSI	1	5	.94			.08	.9	.98
Kieffer & MacDonald (2011)	WAYS	1	26	.84	.1		.41	.52	.93
Kieffer & MacDonald (2011)	WAYS Confrontive Coping	1	68	.69	.12		.52	.41	.93
Kieffer & MacDonald (2011)	WAYS Distancing	1	68	.7	.13		.67	.21	.88
Kieffer & MacDonald (2011)	WAYS Self-Controlling	1	68	.7	.1		.4	.48	.88
Kieffer & MacDonald (2011)	WAYS Seeking Social Support	1	69	.78	.08		.37	.5	.87
Kieffer & MacDonald (2011)	WAYS Accepting Responsibility	1	68	.7	.12		.53	.35	.88
Kieffer & MacDonald (2011)	WAYS Escape-Avoidance	1	66	.71	.1		.51	.4	.91
Kieffer & MacDonald (2011)	WAYS Planful Problem Solving	1	69	.73	.1		.54	.37	.91
Kieffer &	WAYS Positive	1	68	.74	.08		.37	.54	.91

MacDonald (2011)	Reappraisal								
Vassar et al. (2011)	IES	1	52	.89	.01		.29	.67	.96
Vassar et al. (2011)	IES Intrusion	1	52	.87	.01		.17	.78	.95
Vassar et al. (2011)	IES Avoidance	1	52	.84	.01		.22	.7	.92
Vassar et al. (2011)	IES-R	1	14	.91	.01		.1	.86	.96
Vassar et al. (2011)	IES-R Intrusion	1	14	.85	.02		.18	.75	.93
Vassar et al. (2011)	IES-R Avoidance	1	14	.83	.01		.14	.78	.92
Vassar et al. (2011)	IES-R Arousal	1	14	.81	.02		.25	.66	.91
Wheeler et al. (2011)	MBI Emotional Exhaustion	1	98	.87	.03				
Wheeler et al. (2011)	MBI Depersonalization	1	93	.74	.09				
Wheeler et al. (2011)	MBI Personal Accomplishment	1	90	.78	.05				

Legende 2: Kodes zu Daten auf Skalenebene

Reliabilitätstyp	
innere Konsistenz	1
zeitliche Stabilität	2
gemischte Gruppe	3
Interrater Agreement	4
KR-21	5
Cohen's Kappa	6

5.3. Daten zu Meta-Analysen mit Forest Plots

5.3.1. Alter

Tabelle A 3: Alter

Autorinnen und Autoren	Skala	Correlation	Lower limit	Upper limit	Z-Value	p-Value	Weight (Fixed)	Weight (Random)	Sample Size
Yin & Fan (2000)	BDI	-.23	-.39	-.06	-2.65	.01	3.18	1.26	131
Caruso & Edwards (2001)	J-EPQ Psychoticism	.17	-.13	.44	1.10	.27	1.02	1.01	44
Caruso & Edwards (2001)	J-EPQ Extraversion	.27	-.03	.52	1.77	.08	1.02	1.01	44
Caruso & Edwards (2001)	J-EPQ Neuroticism	.29	-.01	.54	1.91	.06	1.02	1.01	44
Caruso & Edwards (2001)	J-EPQ Lie	-.28	-.53	.02	-1.84	.07	1.02	1.01	44
Caruso et al. (2001)	EPQ Psychoticism	-.21	-.43	.03	-1.73	.08	1.64	1.14	69
Caruso et al. (2001)	EPQ Extraversion	-.19	-.41	.05	-1.56	.12	1.64	1.14	69
Caruso et al. (2001)	EPQ Neuroticism	.17	-.07	.39	1.39	.16	1.64	1.14	69
Caruso et al. (2001)	EPQ Lie	-.15	-.37	.09	-1.23	.22	1.64	1.14	69
Barnes et al. (2002)	STAI Trait	.21	-.07	.46	1.48	.14	1.19	1.06	51
Barnes et al. (2002)	STAI State	.31	.04	.54	2.27	.02	1.24	1.07	53
Lane et al. (2002)	CSEI	.13	-.13	.38	0.97	.33	1.37	1.09	58
Nilsson et al. (2002)	CDMSE	-.94	-.99	-.73	-4.26	.00	0.15	0.38	9
Nilsson et al. (2002)	CDMSE Short Form	-.28	-.89	.69	-0.50	.62	0.07	0.22	6
Wallace & Wheeler (2002)	LSI	.00	-.40	.40	0.00	1.00	0.52	0.79	24
Shields &	AUDIT	-.41	-.70	-.01	-2.00	.05	0.52	0.79	24

Caruso (2003)									
Youngstrom & Green (2003)	DES Enjoyment	.49	.16	.72	2.79	.01	0.67	0.88	30
Youngstrom & Green (2003)	DES Interest	.03	-.33	.39	0.16	.88	0.67	0.88	30
Youngstrom & Green (2003)	DES Surprise	-.06	-.41	.31	-0.31	.75	0.67	0.88	30
Youngstrom & Green (2003)	DES Anger	.21	-.16	.53	1.11	.27	0.67	0.88	30
Youngstrom & Green (2003)	DES Contempt	.34	-.02	.62	1.84	.07	0.67	0.88	30
Youngstrom & Green (2003)	DES Disgust	.11	-.26	.45	0.57	.57	0.67	0.88	30
Youngstrom & Green (2003)	DES Fear	.38	.02	.65	2.08	.04	0.67	0.88	30
Youngstrom & Green (2003)	DES Guilt	.09	-.28	.44	0.47	.64	0.67	0.88	30
Youngstrom & Green (2003)	DES Sadness	.29	-.08	.59	1.55	.12	0.67	0.88	30
Youngstrom & Green (2003)	DES Self-hostility	.20	-.17	.52	1.05	.29	0.67	0.88	30
Youngstrom & Green (2003)	DES Shame	.04	-.32	.39	0.21	.84	0.67	0.88	30
Youngstrom & Green (2003)	DES Shyness	.34	-.02	.62	1.84	.07	0.67	0.88	30
Youngstrom & Green (2003)	DES Negative affect	.25	-.12	.56	1.33	.18	0.67	0.88	30
Youngstrom & Green (2003)	DES Positive affect	.44	.09	.69	2.45	.01	0.67	0.88	30
Youngstrom & Green (2003)	DES Hostility	.25	-.12	.56	1.33	.18	0.67	0.88	30
Shields & Caruso (2004)	CAGE	.48	.05	.76	2.16	.03	0.42	0.72	20
Campbell et al. (2005)	PCL:YV	.07	-.48	.58	0.23	.82	0.27	0.57	14
Ryngala et al. (2005)	RCMAS	.30	.02	.54	2.07	.04	1.12	1.04	48
Ryngala et al. (2005)	RCMAS Physiological	-.13	-.40	.16	-0.90	.37	1.12	1.04	48
Ryngala et al. (2005)	RCMAS Worry & Oversensitivity	.28	-.01	.52	1.89	.06	1.12	1.04	48
Ryngala et al. (2005)	RCMAS Concentration	-.10	-.37	.19	-0.67	.51	1.12	1.04	48
Ryngala et al. (2005)	RCMAS Lie	.69	.51	.82	5.74	.00	1.12	1.04	48
Hellman et al.	SPOS	.57	.28	.76	3.55	.00	0.75	0.92	33

(2006)									
Leach et al. (2006)	SDQ I	.00	-.37	.37	0.00	1.00	0.65	0.87	29
Leach et al. (2006)	SDQ I General Self-Concept	.19	-.19	.52	0.98	.33	0.65	0.87	29
Leach et al. (2006)	SDQ I Math	.32	.06	.54	2.37	.02	1.27	1.07	54
Bachner & O'Rourke (2007)	ZBI	.07	-.10	.23	0.81	.42	3.36	1.27	138
Li & Bagger (2007) nt	BIDR Impression Manageme	.06	-.19	.30	0.47	.64	1.54	1.12	65
Li & Bagger (2007)	BIDR Self-Deception Enhancement	.03	-.25	.31	0.21	.84	1.17	1.05	50
Miller. C. S. et al. (2007)	MAC	-.14	-.52	.29	-0.63	.53	0.50	0.78	23
Shields et al. (2007)	MAST	.20	-.12	.48	1.23	.22	0.92	0.98	40
Pearson (2008)	EDI	-.58	-.95	.44	-1.15	.25	0.07	0.22	6
Pearson (2008)	EDI Body Dissatisfaction	.39	.14	.59	2.97	.00	1.29	1.08	55
Pearson (2008)	EDI Drive for Thinness	.19	-.10	.45	1.30	.19	1.14	1.05	49
Pearson (2008)	EDI Bulimia	.62	.40	.77	4.81	.00	1.09	1.03	47
Pearson (2008)	EDI Perfectionism	.65	.40	.81	4.32	.00	0.77	0.93	34
Pearson (2008)	EDI Interoceptive Awareness	.49	.16	.72	2.84	.00	0.70	0.89	31
Pearson (2008)	EDI Ineffectiveness	.53	.20	.75	3.01	.00	0.65	0.87	29
Pearson (2008)	EDI Maturity Fears	.59	.30	.78	3.59	.00	0.70	0.89	31
Pearson (2008)	EDI Interpersonal Distrust	.54	.22	.76	3.08	.00	0.65	0.87	29
Pearson (2008)	EDI Asceticism	.68	.17	.90	2.49	.01	0.22	0.50	12
Pearson (2008)	EDI Impuls Regulation	-.37	-.77	.23	-1.23	.22	0.25	0.53	13
Pearson (2008)	EDI Social Insecurity	.49	-.08	.82	1.70	.09	0.25	0.53	13
Pearson (2008)	EAT-40	.64	.19	.87	2.63	.01	0.30	0.60	15
Pearson (2008)	EAT-26	.10	-.18	.36	0.71	.48	1.24	1.07	53
Pearson (2008)	EAT-26 Dieting	.78	.55	.90	4.79	.00	0.52	0.79	24
Pearson (2008))	EAT-26 Bulimia and Food Preoccupation	.83	.64	.93	5.31	.00	0.50	0.78	23
Pearson (2008)	EAT-26 Oral	-.06	-.51	.42	-0.23	.82	0.37	0.67	18

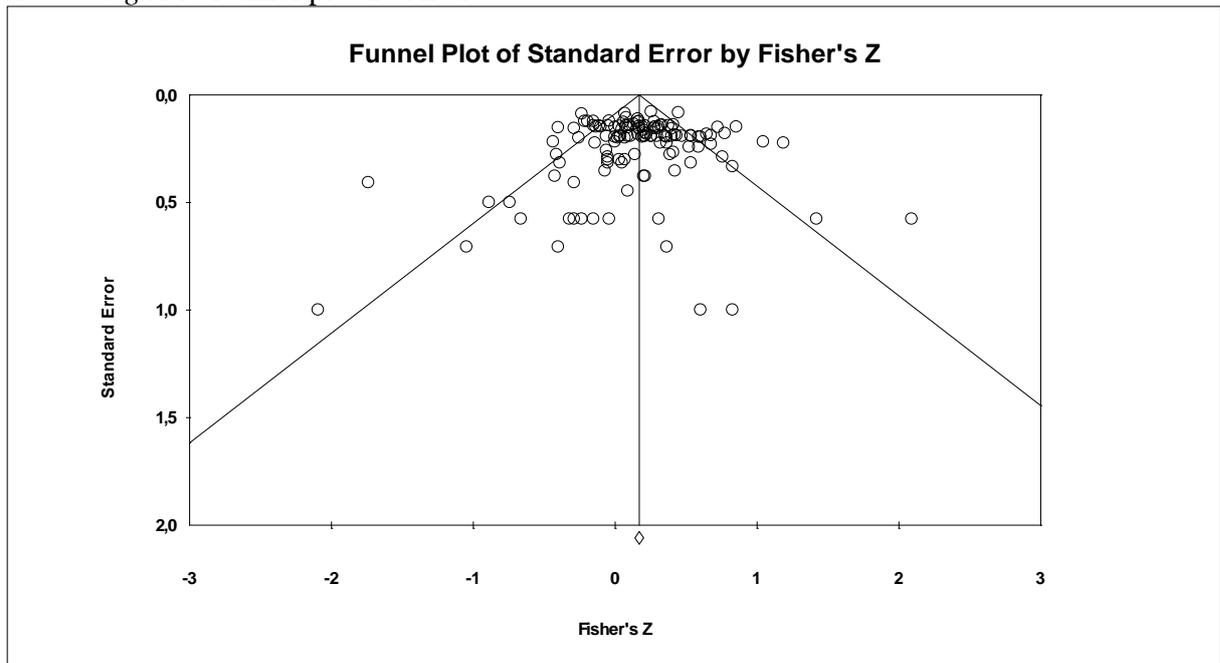
	Control								
Pearson (2008)	BULIT-R	.37	-.15	.73	1.40	.16	0.32	0.62	16
Shields et al. (2008)	PESQ-PS	-.38	-.95	.76	-0.57	.57	0.05	0.15	5
Shields et al. (2008)	RAPI	.68	-.81	.99	0.83	.41	0.02	0.08	4
Shields et al. (2008)	DUSI-SUS	-.15	-.86	.75	-0.26	.79	0.07	0.22	6
Shields et al. (2008)	POSIT-ADS	-.97	-1.00	-.13	-2.09	.04	0.02	0.08	4
Shields et al. (2008)	AAIS	.21	-.48	.74	0.56	.57	0.17	0.42	10
Shields et al. (2008)	alle sechs Instrumente	.16	-.19	.48	0.88	.38	0.75	0.92	33
Vassar (2008a)	SWLS	.22	-.13	.52	1.23	.22	0.75	0.92	33
Vassar (2008b)	SOGS	-.05	-.57	.49	-0.17	.87	0.27	0.57	14
Vassar & Crosby (2008)	UCLA-R	.16	-.09	.39	1.28	.20	1.57	1.13	66
Victorson et al. (2008)	FACT-G	.00	-.29	.29	0.00	1.00	1.09	1.03	47
Victorson et al. (2008)	FACT-G Physical Wellbeing	-.15	-.41	.13	-1.06	.29	1.22	1.06	52
Victorson et al. (2008)	FACT-G Social Wellbeing	.08	-.20	.35	0.56	.58	1.19	1.06	51
Victorson et al. (2008)	FACT-G Functional Wellbeing	.17	-.11	.43	1.18	.24	1.17	1.05	50
Victorson et al. (2008)	FACT-G Emotional Wellbeing	-.11	-.37	.17	-0.77	.44	1.19	1.06	51
Graham & Christiansen (2009)	Rubin Loving	-.38	-.60	-.10	-2.62	.01	1.07	1.03	46
Graham & Christiansen (2009)	Rubin Liking	-.25	-.57	.14	-1.28	.20	0.62	0.85	28
Graham & Christiansen (2009)	LAS Eros	.39	-.11	.73	1.54	.12	0.35	0.65	17
Graham & Christiansen (2009)	LAS Ludus	.14	-.38	.59	0.51	.61	0.32	0.62	16
Graham & Christiansen (2009)	LAS Storge	-.05	-.58	.52	-0.16	.87	0.25	0.53	13

Graham & Christiansen (2009)	LAS Pragma	.03	-.51	.55	0.10	.92	0.27	0.57	14
Graham & Christiansen (2009)	LAS Mania	-.05	-.55	.47	-0.17	.86	0.30	0.60	15
Graham & Christiansen (2009)	LAS Agape	.05	-.52	.58	0.16	.87	0.25	0.53	13
Graham & Christiansen (2009)	LAS-SF Eros	.35	-.77	.94	0.52	.61	0.05	0.15	5
Graham & Christiansen (2009)	LAS -SF Ludus	-.78	-.98	.33	-1.48	.14	0.05	0.15	5
Graham & Christiansen (2009)	LAS -SF Storge	-.31	-.90	.67	-0.56	.58	0.07	0.22	6
Graham & Christiansen (2009)	LAS -SF Pragma	-.23	-.88	.72	-0.41	.69	0.07	0.22	6
Graham & Christiansen (2009)	LAS -SF Mania	-.04	-.82	.80	-0.07	.94	0.07	0.22	6
Graham & Christiansen (2009)	LAS -SF Agape	-.71	-.95	.09	-1.77	.08	0.10	0.28	7
Graham & Christiansen (2009))	PLS 30 item	.30	-.68	.89	0.54	.59	0.07	0.22	6
Graham & Christiansen (2009)	PLS 15 item	-.63	-.94	.23	-1.48	.14	0.10	0.28	7
Graham & Christiansen (2009)	TLS Passion	-.28	-.80	.47	-0.70	.48	0.15	0.38	9
Graham & Christiansen (2009)	TLS Intimacy	.20	-.49	.74	0.54	.59	0.17	0.42	10
Graham & Christiansen (2009)	TLS Commitment	-.40	-.82	.31	-1.12	.26	0.17	0.42	10
Huynh et al. (2009)	Nondominant Alle	.05	-.26	.35	0.31	.75	0.97	1.00	42
Huynh et al. (2009)	Dominant Alle	.09	-.21	.37	0.59	.55	1.07	1.03	46
Lopez-Pina et al.	HAM-D	.35	-.02	.64	1.86	.06	0.65	0.87	29

(2009)									
Miller. B. K. et al. (2009)	POPS	.27	.03	.48	2.21	.03	1.59	1.13	67
Nassen (2009)	CBCL	.89	.28	.99	2.46	.01	0.07	0.22	6
Nassen (2009)	CBCL Internalizing	.09	-.66	.75	0.20	.84	0.12	0.33	8
Nassen (2009)	CBCL Externalizing	.97	.74	1.00	3.62	.00	0.07	0.22	6
Nassen (2009)	CBCL Withdrawn	.54	-.88	.99	0.60	.55	0.02	0.08	4
Ha et al. (2010)	F-MPS	.40	.06	.66	2.28	.02	0.72	0.90	32
Ha et al. (2010)	F-MPS Concern over Mistakes	.38	.09	.61	2.53	.01	1.00	1.00	43
Ha et al. (2010)	F-MPS Personal Standards	.41	.07	.67	2.31	.02	0.70	0.89	31
Ha et al. (2010)	F-MPS Parental Expectation	.59	.22	.81	2.95	.00	0.47	0.76	22
Ha et al. (2010)	F-MPS Parental Criticism	.53	.11	.79	2.43	.01	0.42	0.72	20
Ha et al. (2010)	F-MPS Doubt about Actions	.21	-.13	.51	1.21	.23	0.80	0.94	35
Ha et al. (2010)	F-MPS Organization	.31	-.12	.64	1.43	.15	0.50	0.78	23
Vassar & Bradley (2010)	LOT	.02	-.36	.39	0.10	.92	0.62	0.85	28
Vassar & Bradley (2010)	LOT Optimism	.40	-.26	.81	1.20	.23	0.20	0.46	11
Vassar & Bradley (2010)	LOT Pessimism	-.07	-.64	.55	-0.20	.84	0.20	0.46	11
Vassar & Bradley (2010)	LOT-R	-.04	-.27	.20	-0.33	.74	1.69	1.14	71
Vassar & Bradley (2010)	LOT-R Optimism	.07	-.31	.43	0.35	.73	0.62	0.85	28
Vassar & Bradley (2010)	LOT-R Pessimism	.35	-.07	.67	1.63	.10	0.50	0.78	23
Graham et al. (2011)	LWMAT	.08	-.13	.28	0.76	.45	2.21	1.20	92
Graham et al. (2011)	KMS	.16	-.06	.36	1.43	.15	1.97	1.18	82
Graham et al. (2011)	QMI	.25	.10	.39	3.24	.00	4.00	1.29	164
Graham et al. (2011)	RAS	.42	.28	.54	5.41	.00	3.63	1.28	149
Graham et al. (2011)	MOQ	-.39	-.74	.13	-1.48	.14	0.32	0.62	16
Vassar et al.	IES	.36	.10	.58	2.64	.01	1.22	1.06	52

(2011)									
Vassar et al. (2011)	IES Intrusion	.08	-.20	.35	0.56	.57	1.22	1.06	52
Vassar et al. (2011)	IES Avoidance	-.05	-.32	.23	-0.35	.73	1.22	1.06	52
Fixed-Effects		.17	.14	.20	1.76	.00			
Random-Effects		.18	.12	.23	6.15	.00			

Abbildung A 1: Funnel plot zu Alter



5.3.2. Altershomogenität

Tabelle A 4: Altershomogenität

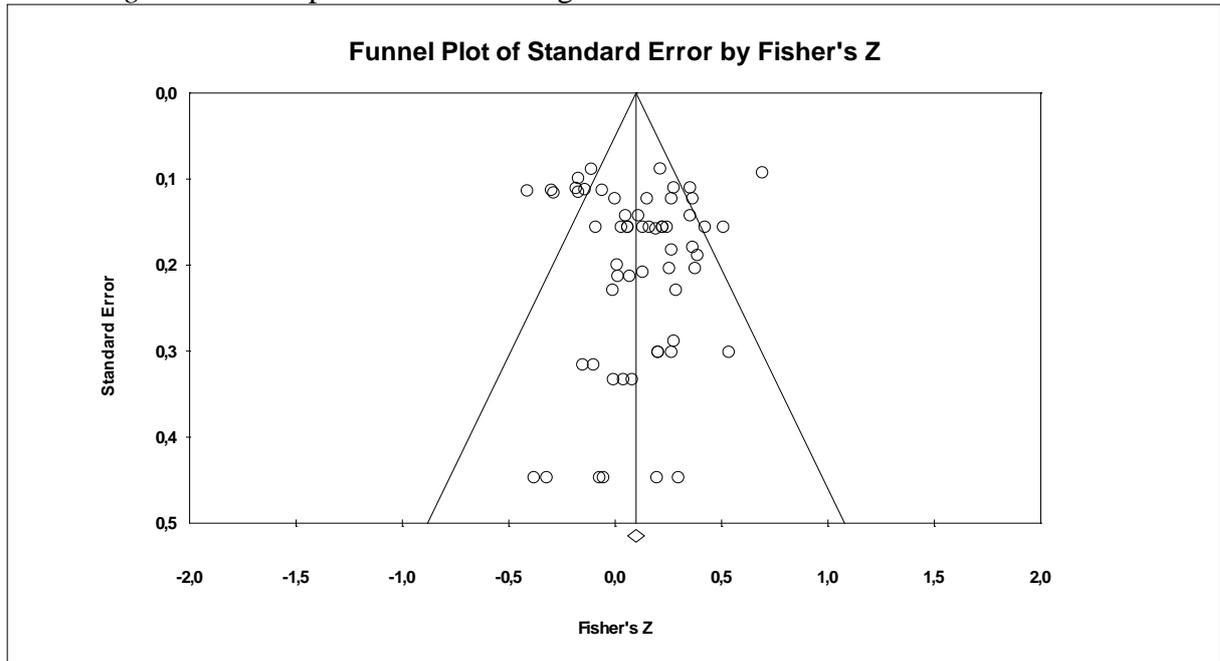
Autorinnen und Autoren	Skala	Correlation	Lower limit	Upper limit	Z-Value	p-Value	Weight (Fixed)	Weight (Random)	Sample Size
Yin & Fan (2000)	BDI	.21	.04	.37	2.41	.02	5.16	2.56	131
Capraro, M. M.	MARS	.01	-.36	.38	0.05	.96	1.01	1.60	28

et al. (2001)									
Caruso & Edwards (2001)	J-EPQ Psychoticism	.13	-.17	.41	0.84	.40	1.65	1.95	44
Caruso & Edwards (2001)	J-EPQ Extraversion	-.09	-.38	.21	-0.58	.56	1.65	1.95	44
Caruso & Edwards (2001)	J-EPQ Neuroticism	.47	.20	.67	3.27	.00	1.65	1.95	44
Caruso & Edwards (2001)	J-EPQ Lie	.40	.12	.62	2.71	.01	1.65	1.95	44
Caruso et al. (2001)	EPQ Psychoticism	.35	.12	.54	2.97	.00	2.66	2.25	69
Caruso et al. (2001)	EPQ Extraversion	.15	-.09	.37	1.23	.22	2.66	2.25	69
Caruso et al. (2001)	EPQ Neuroticism	.00	-.24	.24	0.00	1.00	2.66	2.25	69
Caruso et al. (2001)	EPQ Lie	.26	.02	.47	2.16	.03	2.66	2.25	69
Beretvas et al. (2002)	MCSDS	.60	.47	.70	7.47	.00	4.68	2.52	119
Reese et al. (2002)	AAS Depend	-.05	-.73	.68	-0.12	.90	0.20	0.56	8
Reese et al. (2002)	AAS Close	-.07	-.74	.67	-0.16	.87	0.20	0.56	8
Reese et al. (2002)	AAS Anxiety	-.36	-.85	.46	-0.85	.40	0.20	0.56	8
Reese et al. (2002)	IPPA Trust	.04	-.55	.60	0.12	.90	0.36	0.87	12
Reese et al. (2002)	IPPA Communication	-.01	-.58	.57	-0.02	.98	0.36	0.87	12
Reese et al. (2002)	IPPA Alienation	.08	-.52	.63	0.24	.81	0.36	0.87	12
Reese et al. (2002)	PAQ Affective Quality	-.31	-.83	.50	-0.72	.47	0.20	0.56	8
Reese et al. (2002)	PAQ Parental Fostering	.20	-.59	.79	0.44	.66	0.20	0.56	8
Reese et al. (2002)	PAQ Parental Support	.29	-.52	.83	0.67	.50	0.20	0.56	8
Reese et al. (2002)	PBI Maternal Overprotection	.25	-.14	.58	1.25	.21	0.97	1.57	27
Reese et al. (2002)	PBI Maternal Care	.36	-.02	.65	1.85	.06	0.97	1.57	27
Reese et al. (2002)	PBI Paternal Overprotection	.01	-.38	.41	0.07	.95	0.89	1.50	25
Reese et al. (2002)	PBI Paternal Care	.07	-.34	.45	0.32	.75	0.89	1.50	25
Kieffer et al.	AEQ	.28	-.16	.63	1.25	.21	0.77	1.39	22

(2004)									
Kieffer et al. (2004)	AEQ Global Positive Changes in Experience	.20	-.37	.66	0.67	.50	0.44	1.00	14
Kieffer et al. (2004)	AEQ Sexual Enhancement	.26	-.31	.69	0.88	.38	0.44	1.00	14
Kieffer et al. (2004)	AEQ Social and Physical Pleasure	-.10	-.62	.48	-0.32	.75	0.40	0.94	13
Kieffer et al. (2004)	AEQ Assertiveness	-.15	-.65	.44	-0.48	.63	0.40	0.94	13
Kieffer et al. (2004)	AEQ Relaxation/Tension Reduction	.20	-.37	.66	0.67	.50	0.44	1.00	14
Kieffer et al. (2004)	AEQ Arousal/Interpersonal Power	.49	-.05	.81	1.78	.08	0.44	1.00	14
O'Rourke (2004)	CES-D	-.11	-.28	.06	-1.24	.21	5.12	2.56	130
Rexrode et al. (2008)	WOCS Accepting Responsibility	-.28	-.47	-.06	-2.47	.01	2.98	2.32	77
Rexrode et al. (2008)	WOCS Confrontive Coping	-.17	-.38	.05	-1.49	.14	3.02	2.32	78
Rexrode et al. (2008)	WOCS Distancing	-.14	-.35	.08	-1.25	.21	3.19	2.35	82
Rexrode et al. (2008)	WOCS Escape-Avoidance	-.18	-.38	.04	-1.64	.10	3.27	2.36	84
Rexrode et al. (2008)	WOCS Planful Problem Solving	-.06	-.27	.16	-0.53	.60	3.15	2.34	81
Rexrode et al. (2008)	WOCS Positive Reappraisal	-.39	-.56	-.19	-3.61	.00	3.10	2.34	80
Rexrode et al. (2008)	WOCS Self-Controlling	-.29	-.48	-.08	-2.64	.01	3.15	2.34	81
Rexrode et al. (2008)	WOCS Seeking Social Support	-.17	-.35	.02	-1.73	.08	4.07	2.47	104
Vassar (2008a)	SWLS	.26	-.09	.55	1.46	.14	1.21	1.73	33
Vassar & Crosby (2008)	UCLA-R	.27	.06	.46	2.51	.01	3.31	2.37	85
Lopez-Pina et al. (2009)	HAM-D	.27	-.28	.69	0.96	.34	0.48	1.06	15
Vassar & Bradley (2010)	LOT	.35	.01	.62	2.03	.04	1.25	1.76	34
Vassar & Bradley (2010)	LOT-R	.34	.14	.52	3.21	.00	3.31	2.37	85
Vassar & Bradley (2010)	LOT-R Optimism	.37	.02	.64	2.06	.04	1.13	1.68	31
Vassar &	LOT-R	.13	-.27	.49	0.63	.53	0.93	1.53	26

Bradley (2010)	Pessimism								
Kieffer & MacDonald (2011)	WAYS	-.01	-.43	.41	-0.04	.97	0.77	1.39	22
Kieffer & MacDonald (2011)	WAYS Confrontive Coping	.03	-.27	.32	0.19	.85	1.65	1.95	44
Kieffer & MacDonald (2011)	WAYS Distancing	.06	-.24	.35	0.38	.70	1.65	1.95	44
Kieffer & MacDonald (2011)	WAYS Self-Controlling	.06	-.24	.35	0.38	.70	1.65	1.95	44
Kieffer & MacDonald (2011)	WAYS Seeking Social Support	.16	-.14	.44	1.03	.30	1.65	1.95	44
Kieffer & MacDonald (2011)	WAYS Accepting Responsibility	.22	-.08	.49	1.43	.15	1.65	1.95	44
Kieffer & MacDonald (2011)	WAYS Escape-Avoidance	.19	-.12	.46	1.22	.22	1.61	1.94	43
Kieffer & MacDonald (2011)	WAYS Planful Problem Solving	.22	-.08	.49	1.43	.15	1.65	1.95	44
Kieffer & MacDonald (2011)	WAYS Positive Reappraisal	.24	-.06	.50	1.57	.12	1.65	1.95	44
Vassar et al. (2011)	IES	.05	-.23	.32	0.35	.73	1.98	2.07	52
Vassar et al. (2011) I	ES Intrusion	.34	.07	.56	2.48	.01	1.98	2.07	52
Vassar et al. (2011)	IES Avoidance	.11	-.17	.37	0.77	.44	1.98	2.07	52
Fixed-Effects		.10	.06	.14	4.88	.00			
Random-Effects		.11	.04	.18	3.03	.00			

Abbildung A 2: Funnel plot zu Altershomogenität



5.3.3. Ethnie

Tabelle A 5: Ethnie

Autorinnen und Autoren	Skala	Correlation	Lower limit	Upper limit	Z-Value	p-Value	Weight (Fixed)	Weight (Random)	Sample Size
Lane et al. (2002)	CSEI	.53	.13	.78	2.50	.01	0.70	0.96	21
Nilsson et al. (2002)	CDMSE Short Form	-.81	-.98	.00	-1.95	.05	0.12	0.28	6
Reese et al. (2002)	AAS Depend	.28	-.53	.82	0.65	.51	0.19	0.42	8
Reese et al. (2002)	AAS Close	-.77	-.96	-.03	-2.02	.04	0.16	0.36	7
Reese et al. (2002)	AAS Anxiety	.50	-.41	.91	1.09	.28	0.16	0.36	7
Reese et al. (2002)	BORI Alienation	-.47	-.96	.71	-0.71	.48	0.08	0.20	5
Reese et al.	BORI Social	-.48	-.96	.70	-0.74	.46	0.08	0.20	5

(2002)	Incompetence								
Reese et al. (2002)	BORI Insecure	-.46	-.96	.71	-0.71	.48	0.08	0.20	5
Reese et al. (2002)	BORI Egocentric	-.49	-.96	.69	-0.76	.45	0.08	0.20	5
Reese et al. (2002)	IPPA Trust	-.45	-.90	.46	-0.97	.33	0.16	0.36	7
Reese et al. (2002)	IPPA Communication	.50	-.41	.91	1.09	.27	0.16	0.36	7
Reese et al. (2002)	IPPA Alienation	.58	-.31	.93	1.31	.19	0.16	0.36	7
Reese et al. (2002)	PAQ Affective Quality	-.28	-.93	.80	-0.41	.68	0.08	0.20	5
Reese et al. (2002)	PAQ Parental Fostering	.90	.10	.99	2.10	.04	0.08	0.20	5
Reese et al. (2002)	PAQ Parental Support	.82	-.22	.99	1.64	.10	0.08	0.20	5
Reese et al. (2002)	PBI Maternal Overprotection	.05	-.38	.46	0.21	.83	0.74	0.98	22
Reese et al. (2002)	PBI Maternal Care	.15	-.29	.54	0.67	.50	0.74	0.98	22
Reese et al. (2002)	PBI Paternal Overprotection	-.01	-.45	.43	-0.05	.96	0.66	0.93	20
Reese et al. (2002)	PBI Paternal Care	.12	-.34	.54	0.51	.61	0.66	0.93	20
Youngstrom & Green (2003)	DES Enjoyment	.31	-.06	.60	1.67	.10	1.05	1.14	30
Youngstrom & Green (2003)	DES Interest	-.11	-.45	.26	-0.57	.57	1.05	1.14	30
Youngstrom & Green (2003)	DES Surprise	-.02	-.38	.34	-0.10	.92	1.05	1.14	30
Youngstrom & Green (2003)	DES Anger	.15	-.22	.48	0.79	.43	1.05	1.14	30
Youngstrom & Green (2003)	DES Contempt	.03	-.33	.39	0.16	.88	1.05	1.14	30
Youngstrom & Green (2003)	DES Disgust	.05	-.32	.40	0.26	.79	1.05	1.14	30
Youngstrom & Green (2003)	DES Fear	.36	.00	.64	1.96	.05	1.05	1.14	30
Youngstrom & Green (2003)	DES Guilt	.28	-.09	.58	1.49	.13	1.05	1.14	30
Youngstrom & Green (2003)	DES Sadness	.23	-.14	.55	1.22	.22	1.05	1.14	30
Youngstrom & Green (2003)	DES Self-hostility	.38	.02	.65	2.08	.04	1.05	1.14	30
Youngstrom & Green (2003)	DES Shame	.12	-.25	.46	0.63	.53	1.05	1.14	30

Green (2003)									
Youngstrom & Green (2003)	DES Shyness	.53	.21	.75	3.07	.00	1.05	1.14	30
Youngstrom & Green (2003)	DES Negative affect	.07	-.30	.42	0.36	.72	1.05	1.14	30
Youngstrom & Green (2003)	DES Positive affect	.10	-.27	.44	0.52	.60	1.05	1.14	30
Youngstrom & Green (2003)	DES Hostility	-.04	-.39	.32	-0.21	.84	1.05	1.14	30
Kieffer et al. (2004)	AEQ	.38	-.05	.69	1.74	.08	0.74	0.98	22
Kieffer et al. (2004)	AEQ Global Positive Changes in Experience	-.59	-.85	-.09	-2.25	.02	0.43	0.73	14
Kieffer et al. (2004)	AEQ Sexual Enhancement	-.41	-.77	.15	-1.44	.15	0.43	0.73	14
Kieffer et al. (2004)	AEQ Social and Physical Pleasure	-.64	-.88	-.14	-2.40	.02	0.39	0.69	13
Kieffer et al. (2004)	AEQ Assertiveness	-.65	-.88	-.15	-2.45	.01	0.39	0.69	13
Kieffer et al. (2004)	AEQ Relaxation/Tension Reduction	-.61	-.86	-.12	-2.35	.02	0.43	0.73	14
Kieffer et al. (2004)	AEQ Arousal/Interpersonal Power	.25	-.32	.69	0.85	.40	0.43	0.73	14
Shields & Caruso (2004)	CAGE	-.61	-.90	.03	-1.88	.06	0.27	0.54	10
Campbell et al. (2005)	PCL:YV	.09	-.42	.56	0.33	.74	0.51	0.81	16
Ryngala et al. (2005)	RCMAS	.26	-.02	.51	1.79	.07	1.75	1.35	48
Ryngala et al. (2005)	RCMAS Physiological	.17	-.12	.43	1.14	.25	1.75	1.35	48
Ryngala et al. (2005)	RCMAS Worry & Oversensitivity	.21	-.07	.47	1.46	.14	1.75	1.35	48
Ryngala et al. (2005)	RCMAS Concentration	.53	.29	.71	3.94	.00	1.75	1.35	48
Ryngala et al. (2005)	RCMAS Lie	.16	-.14	.42	1.05	.29	1.75	1.35	48
Graham et al. (2006)	DAS	-.01	-.34	.32	-0.06	.95	1.28	1.23	36
Graham et al. (2006)	DAS Dyadic Consensus	-.18	-.59	.30	-0.73	.47	0.62	0.90	19
Graham et al. (2006)	DAS Dyadic Satisfaction	.15	-.31	.56	0.62	.53	0.66	0.93	20

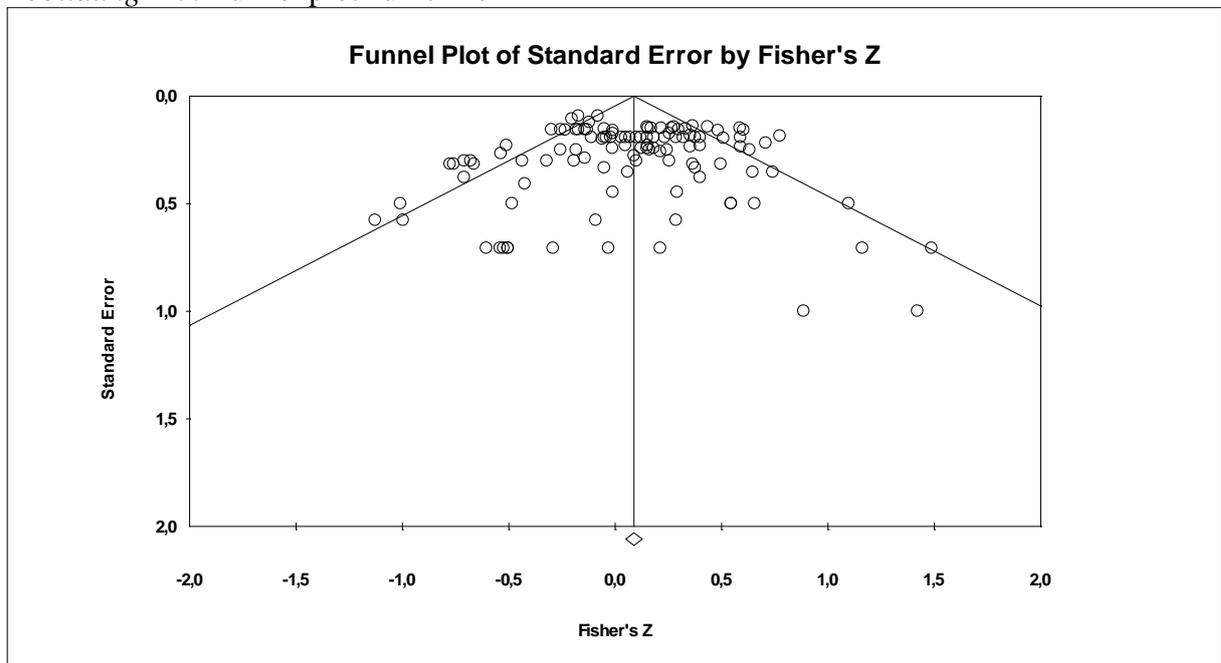
Graham et al. (2006)	DAS Dyadic Cohesion	.21	-.28	.62	0.83	.41	0.58	0.87	18
Graham et al. (2006)	DAS Affective Expression	-.49	-.79	-.01	-2.01	.04	0.54	0.84	17
Hellman et al. (2006)	SPOS	.34	.00	.61	1.94	.05	1.17	1.19	33
Miller, C. S. et al. (2007)	MAC	-.01	-.71	.70	-0.02	.98	0.19	0.42	8
Shields et al. (2007)	MAST	-.06	-.42	.32	-0.30	.76	0.97	1.11	28
Shields et al. (2008)	PESQ-PS	-.03	-.89	.88	-0.04	.97	0.08	0.20	5
Shields et al. (2008)	RAPI	-.76	-.97	.13	-1.73	.08	0.12	0.28	6
Shields et al. (2008)	DUSI-SUS	-.54	-.96	.65	-0.85	.39	0.08	0.20	5
Shields et al. (2008)	POSIT-ADS	.71	-.79	.99	0.89	.37	0.04	0.10	4
Shields et al. (2008)	AAIS	.80	.12	.97	2.20	.03	0.16	0.36	7
Shields et al. (2008)	alle sechs Instrumente	.18	-.19	.51	0.95	.34	1.05	1.14	30
Vassar (2008b)	SOGS	-.40	-.84	.36	-1.04	.30	0.23	0.49	9
Victorson et al. (2008)	FACT-G	.34	-.11	.67	1.50	.13	0.70	0.96	21
Victorson et al. (2008)	FACT-G Physical Wellbeing	.18	-.29	.58	0.75	.45	0.66	0.93	20
Victorson et al. (2008)	FACT-G Social Wellbeing	.16	-.32	.57	0.65	.52	0.62	0.90	19
Victorson et al. (2008)	FACT-G Functional Wellbeing	-.25	-.63	.23	-1.02	.31	0.62	0.90	19
Victorson et al. (2008)	FACT-G Emotional Wellbeing	.24	-.24	.63	0.98	.33	0.62	0.90	19
Graham & Christiansen (2009)	Rubin Loving	.61	.27	.81	3.25	.00	0.82	1.03	24
Graham & Christiansen (2009)	Rubin Liking	.56	.14	.81	2.53	.01	0.62	0.90	19
Graham & Christiansen (2009)	LAS Eros	-.05	-.61	.54	-0.15	.88	0.35	0.65	12
Graham & Christiansen (2009)	LAS Ludus	.46	-.12	.81	1.57	.12	0.39	0.69	13
Graham &	LAS Storge	.35	-.25	.76	1.16	.25	0.39	0.69	13

Christiansen (2009)									
Graham & Christiansen (2009)	LAS Pragma	-.14	-.61	.40	-0.49	.63	0.47	0.77	15
Graham & Christiansen (2009)	LAS Mania	-.58	-.86	-.04	-2.09	.04	0.39	0.69	13
Graham & Christiansen (2009)	LAS Agape	-.19	-.65	.38	-0.64	.52	0.43	0.73	14
Graham & Christiansen (2009)	LAS-SF Eros	.38	-.33	.81	1.06	.29	0.27	0.54	10
Graham & Christiansen (2009)	LAS -SF Ludus	.06	-.56	.64	0.17	.87	0.31	0.60	11
Graham & Christiansen (2009)	LAS -SF Storge	.63	.05	.89	2.10	.04	0.31	0.60	11
Graham & Christiansen (2009)	LAS -SF Pragma	.36	-.27	.77	1.13	.26	0.35	0.65	12
Graham & Christiansen (2009)	LAS -SF Mania	.57	-.05	.87	1.83	.07	0.31	0.60	11
Graham & Christiansen (2009)	LAS -SF Agape	.10	-.45	.60	0.33	.74	0.43	0.73	14
Graham & Christiansen (2009)	PLS 15 item	-.09	-.84	.78	-0.16	.88	0.12	0.28	6
Graham & Christiansen (2009)	TLS Passion	.89	-.49	1.00	1.42	.16	0.04	0.10	4
Graham & Christiansen (2009)	TLS Intimacy	.21	-.83	.92	0.30	.76	0.08	0.20	5
Graham & Christiansen (2009)	TLS Commitment	.28	-.69	.89	0.50	.62	0.12	0.28	6
Huynh et al. (2009)	Nondominant Alle	-.01	-.31	.29	-0.06	.95	1.52	1.29	42
Huynh et al. (2009)	Dominant Alle	-.05	-.34	.24	-0.33	.74	1.67	1.33	46
Miller, B. K. et al. (2009)	POPS	.25	-.09	.53	1.47	.14	1.28	1.23	36

Ha et al. (2010)	F-MPS	.32	.03	.56	2.17	.03	1.67	1.33	46
Ha et al. (2010)	F-MPS Concern over Mistakes	.35	.09	.56	2.64	.01	2.02	1.40	55
Ha et al. (2010)	F-MPS Personal Standards	.45	.17	.66	3.03	.00	1.52	1.29	42
Ha et al. (2010)	F-MPS Parental Expectation	-.05	-.40	.32	-0.26	.79	1.05	1.14	30
Ha et al. (2010)	F-MPS Parental Criticism	.47	.13	.71	2.60	.01	1.01	1.12	29
Ha et al. (2010)	F-MPS Doubt about Actions	.54	.29	.72	3.82	.00	1.56	1.30	43
Ha et al. (2010)	F-MPS Organization	.65	.39	.81	4.18	.00	1.13	1.17	32
Graham et al. (2011)	LWMAT	-.20	-.39	.01	-1.91	.06	3.46	1.56	92
Graham et al. (2011)	KMS	-.12	-.35	.12	-0.99	.32	2.60	1.48	70
Graham et al. (2011)	QMI	-.08	-.26	.10	-0.87	.39	4.55	1.62	120
Graham et al. (2011)	RAS	-.17	-.34	.01	-1.86	.06	4.55	1.62	120
Graham et al. (2011)	MOQ	-.31	-.72	.26	-1.06	.29	0.43	0.73	14
Kieffer & MacDonald (2011)	WAYS	-.47	-.74	-.06	-2.22	.03	0.74	0.98	22
Kieffer & MacDonald (2011)	WAYS Confrontive Coping	-.13	-.41	.17	-0.84	.40	1.59	1.31	44
Kieffer & MacDonald (2011)	WAYS Distancing	-.18	-.45	.12	-1.17	.24	1.59	1.31	44
Kieffer & MacDonald (2011)	WAYS Self-Controlling	-.17	-.44	.13	-1.10	.27	1.59	1.31	44
Kieffer & MacDonald (2011)	WAYS Seeking Social Support	-.29	-.54	.01	-1.91	.06	1.59	1.31	44
Kieffer & MacDonald (2011)	WAYS Accepting Responsibility	-.14	-.42	.16	-0.90	.37	1.59	1.31	44
Kieffer & MacDonald (2011)	WAYS Escape-Avoidance	-.23	-.50	.08	-1.48	.14	1.56	1.30	43
Kieffer & MacDonald	WAYS Planful Problem Solving	-.25	-.51	.05	-1.64	.10	1.59	1.31	44

(2011)									
Kieffer & MacDonald (2011)	WAYS Positive Reappraisal	.29	-.01	.54	1.91	.06	1.59	1.31	44
Vassar et al. (2011)	IES	.27	.00	.51	1.94	.05	1.91	1.38	52
Vassar et al. (2011)	IES Intrusion	.41	.15	.61	3.05	.00	1.91	1.38	52
Vassar et al. (2011)	IES Avoidance	.15	-.13	.41	1.06	.29	1.91	1.38	52
Fixed-Effects		.09	.05	.13	4.49	.00			
Random-Effects		.09	.03	.16	2.77	.01			

Abbildung A 3: Funnel plot zu Ethnie



5.3.4. Geschlecht

Tabelle A 6: Geschlecht

Autorinnen und Autoren	Skala	Correlation	Lower limit	Upper limit	Z-Value	p-Value	Weight (Fixed)	Weight (Random)	Sample size
Caruso & Edwards (2001)	J-EPQ Psychoticism	.39	.11	.62	2.64	.01	0.85	0.80	44
Caruso & Edwards (2001)	J-EPQ Extraversion	-.01	-.31	.29	-0.06	.95	0.85	0.80	44
Caruso & Edwards (2001)	J-EPQ Neuroticism	.30	.00	.55	1.98	.05	0.85	0.80	44
Caruso & Edwards (2001)	J-EPQ Lie	.07	-.23	.36	0.45	.65	0.85	0.80	44
Caruso et al. (2001)	EPQ Psychoticism	-.15	-.37	.09	-1.23	.22	1.37	0.88	69
Caruso et al. (2001)	EPQ Extraversion	.06	-.18	.29	0.49	.63	1.37	0.88	69
Caruso et al. (2001)	EPQ Neuroticism	-.03	-.26	.21	-0.24	.81	1.37	0.88	69
Caruso et al. (2001)	EPQ Lie	-.03	-.26	.21	-0.24	.81	1.37	0.88	69
Beretvas et al. (2002)	MCSDS	.89	.85	.92	15.31	.00	2.41	0.95	119
Nilsson et al. (2002)	CDMSE	-.24	-.78	.50	-0.60	.55	0.12	0.33	9
Nilsson et al. (2002)	CDMSE Short Form	-.34	-.90	.65	-0.61	.54	0.06	0.20	6
Reese et al. (2002)	AAS Depend	-.11	-.76	.64	-0.26	.80	0.10	0.29	8
Reese et al. (2002)	AAS Close	.02	-.69	.72	0.05	.96	0.10	0.29	8
Reese et al. (2002)	AAS Anxiety	-.23	-.81	.56	-0.53	.59	0.10	0.29	8
Reese et al. (2002)	BORI Alienation	-.63	-.99	.84	-0.75	.45	0.02	0.08	4
Reese et al. (2002)	BORI Social Incompetence	-.60	-.99	.85	-0.70	.49	0.02	0.08	4
Reese et al. (2002)	BORI Insecure	-.64	-.99	.84	-0.75	.45	0.02	0.08	4

Reese et al. (2002)	BORI Egocentric	-.54	-.99	.87	-0.61	.54	0.02	0.08	4
Reese et al. (2002)	IPPA Trust	-.47	-.82	.14	-1.54	.12	0.19	0.43	12
Reese et al. (2002)	IPPA Communication	.26	-.37	.72	0.79	.43	0.19	0.43	12
Reese et al. (2002)	IPPA Alienation	.36	-.27	.77	1.11	.27	0.19	0.43	12
Reese et al. (2002)	PAQ Affective Quality	.54	-.27	.90	1.34	.18	0.10	0.29	8
Reese et al. (2002)	PAQ Parental Fostering	.22	-.57	.80	0.50	.62	0.10	0.29	8
Reese et al. (2002)	PAQ Parental Support	-.04	-.72	.68	-0.09	.93	0.10	0.29	8
Reese et al. (2002)	PBI Maternal Overprotection	.38	-.06	.70	1.71	.09	0.37	0.61	21
Reese et al. (2002)	PBI Maternal Care	-.05	-.47	.39	-0.20	.84	0.37	0.61	21
Reese et al. (2002)	PBI Paternal Overprotection	.09	-.37	.51	0.36	.72	0.35	0.60	20
Reese et al. (2002)	PBI Paternal Care	-.01	-.45	.43	-0.04	.97	0.35	0.60	20
Thompson & Cook (2002)	LibQUAL Information Access	.01	-.29	.31	0.06	.95	0.83	0.80	43
Wallace & Wheeler (2002)	LSI	.00	-.35	.35	0.00	1.00	0.60	0.73	32
Shields & Caruso (2003)	AUDIT	-.13	-.51	.29	-0.60	.55	0.44	0.65	24
Youngstrom & Green (2003)	DES Enjoyment	.12	-.25	.46	0.63	.53	0.56	0.71	30
Youngstrom & Green (2003)	DES Interest	-.02	-.38	.34	-0.10	.92	0.56	0.71	30
Youngstrom & Green (2003)	DES Surprise	.30	-.07	.60	1.61	.11	0.56	0.71	30
Youngstrom & Green (2003)	DES Anger	.02	-.34	.38	0.10	.92	0.56	0.71	30
Youngstrom & Green (2003)	DES Contempt	-.24	-.55	.13	-1.27	.20	0.56	0.71	30
Youngstrom & Green (2003)	DES Disgust	.11	-.26	.45	0.57	.57	0.56	0.71	30
Youngstrom & Green (2003)	DES Fear	.27	-.10	.57	1.44	.15	0.56	0.71	30
Youngstrom & Green (2003)	DES Guilt	.18	-.19	.51	0.95	.34	0.56	0.71	30
Youngstrom & Green (2003)	DES Sadness	.16	-.21	.49	0.84	.40	0.56	0.71	30

Green (2003)									
Youngstrom & Green (2003)	DES Self-hostility	-.02	-.38	.34	-0.10	.92	0.56	0.71	30
Youngstrom & Green (2003)	DES Shame	.34	-.02	.62	1.84	.07	0.56	0.71	30
Youngstrom & Green (2003)	DES Shyness	-.10	-.44	.27	-0.52	.60	0.56	0.71	30
Youngstrom & Green (2003)	DES Negative affect	.10	-.27	.44	0.52	.60	0.56	0.71	30
Youngstrom & Green (2003)	DES Positive affect	.07	-.30	.42	0.36	.72	0.56	0.71	30
Youngstrom & Green (2003)	DES Hostility	-.07	-.42	.30	-0.36	.72	0.56	0.71	30
Kieffer et al. (2004)	AEQ	.25	-.19	.61	1.11	.27	0.40	0.63	22
Kieffer et al. (2004)	AEQ Global Positive Changes in Experience	.55	.03	.84	2.05	.04	0.23	0.49	14
Kieffer et al. (2004)	AEQ Sexual Enhancement	.77	.40	.92	3.38	.00	0.23	0.49	14
Kieffer et al. (2004)	AEQ Social and Physical Pleasure	.18	-.41	.67	0.58	.56	0.21	0.46	13
Kieffer et al. (2004)	AEQ Assertiveness	-.80	-.94	-.45	-3.47	.00	0.21	0.46	13
Kieffer et al. (2004)	AEQ Relaxation/Tension Reduction	.26	-.31	.69	0.88	.38	0.23	0.49	14
Kieffer et al. (2004)	AEQ Arousal/Interpersonal Power	.73	.33	.91	3.08	.00	0.23	0.49	14
Shields & Caruso (2004)	CAGE	.16	-.40	.64	0.54	.59	0.23	0.49	14
Ryngala et al. (2005)	RCMAS	.28	.00	.52	1.95	.05	0.94	0.82	48
Ryngala et al. (2005)	RCMAS Physiological	.23	-.06	.48	1.58	.11	0.94	0.82	48
Ryngala et al. (2005)	RCMAS Worry & Oversensitivity	-.08	-.36	.21	-0.54	.59	0.94	0.82	48
Ryngala et al. (2005)	RCMAS Concentration	.12	-.17	.39	0.79	.43	0.94	0.82	48
Ryngala et al. (2005)	RCMAS Lie	-.01	-.29	.28	-0.05	.96	0.94	0.82	48
Graham et al. (2006)	DAS	.00	-.33	.33	0.00	1.00	0.69	0.76	36
Graham et al. (2006)	DAS Dyadic Consensus	.31	-.17	.67	1.28	.20	0.33	0.58	19

Graham et al. (2006)	DAS Dyadic Satisfaction	.06	-.39	.49	0.25	.80	0.35	0.60	20
Graham et al. (2006)	DAS Dyadic Cohesion	.39	-.09	.72	1.59	.11	0.31	0.57	18
Graham et al. (2006)	DAS Affective Expression	.35	-.16	.71	1.37	.17	0.29	0.55	17
Hellman et al. (2006)	SPOS	.10	-.25	.43	0.55	.58	0.62	0.74	33
Li & Bagger (2007)	BIDR Impression Management	.09	-.11	.28	0.89	.37	2.04	0.93	101
Li & Bagger (2007)	BIDR Self-Deception Enhancement	-.03	-.24	.18	-0.28	.78	1.75	0.91	87
Miller, C. S. et al. (2007)	MAC	-.69	-.85	-.41	-4.07	.00	0.48	0.68	26
Rouse (2007)	MMPI-2 Aggressivness	.09	-.17	.34	0.68	.50	1.17	0.86	59
Rouse (2007)	MMPI-2 Psychoticism	-.04	-.29	.22	-0.30	.76	1.17	0.86	59
Rouse (2007)	MMPI-2 Disconstraint	-.01	-.27	.25	-0.07	.94	1.17	0.86	59
Rouse (2007)	MMPI-2 Negative Emotionality/Neuroticism	-.03	-.28	.23	-0.22	.82	1.17	0.86	59
Rouse (2007)	MMPI-2 Introversion/Low positive Emotionality	.04	-.22	.29	0.30	.76	1.17	0.86	59
Shields et al. (2007)	MAST	-.33	-.57	-.04	-2.22	.03	0.87	0.81	45
Beretvas et al. (2008)	I-E	-.35	-.58	-.06	-2.37	.02	0.87	0.81	45
Pearson (2008)	EDI Body Dissatisfaction	.83	.72	.90	8.06	.00	0.96	0.82	49
Pearson (2008)	EDI Drive for Thinness	.52	.25	.71	3.55	.00	0.79	0.79	41
Pearson (2008)	EDI Bulimia	.69	.48	.83	5.09	.00	0.75	0.78	39
Pearson (2008)	EDI Perfectionism	.34	-.01	.62	1.91	.06	0.60	0.73	32
Pearson (2008)	EDI Interoceptive Awareness	.59	.29	.79	3.46	.00	0.54	0.70	29
Pearson (2008)	EDI Ineffectiveness	.37	-.01	.66	1.90	.06	0.50	0.69	27
Pearson (2008)	EDI Maturity Fears	.00	-.37	.37	0.00	1.00	0.54	0.70	29

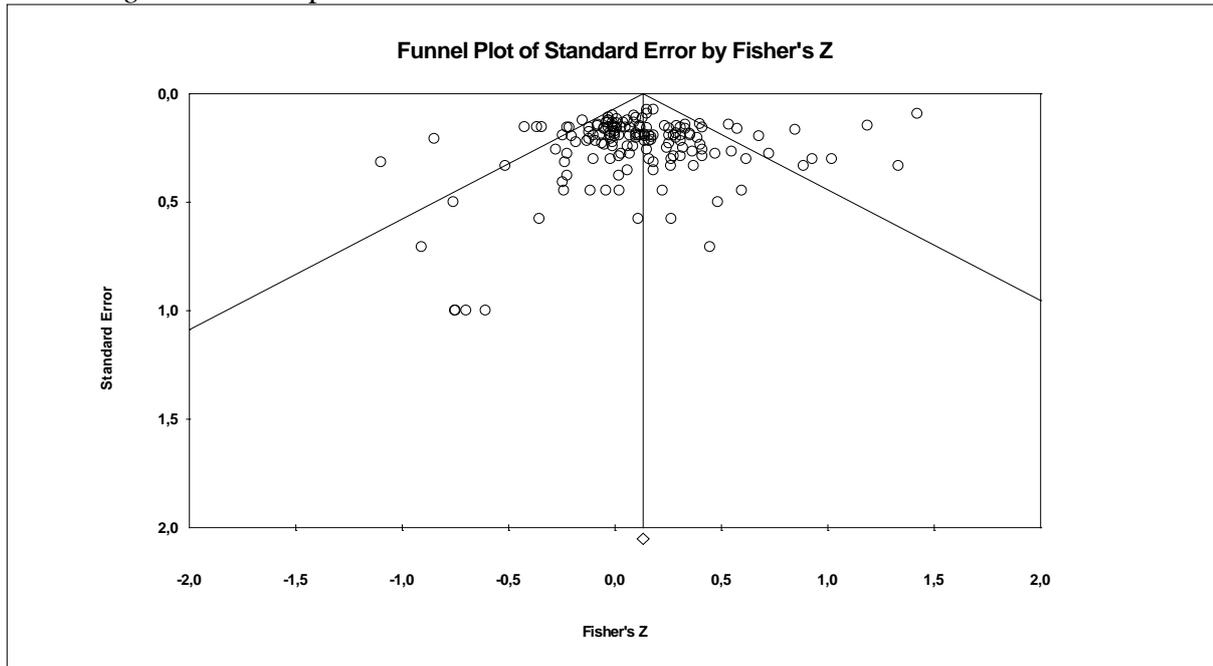
Pearson (2008)	EDI Interpersonal Distrust	.28	-.11	.60	1.41	.16	0.50	0.69	27
Pearson (2008)	EAT-40	.87	.59	.96	4.00	.00	0.19	0.43	12
Pearson (2008)	EAT-26	.00	-.29	.29	0.00	1.00	0.89	0.81	46
Pearson (2008)	EAT-26 Dieting	-.09	-.48	.33	-0.41	.68	0.44	0.65	24
Pearson (2008)	EAT-26 Bulimia and Food Preoccupation	-.18	-.55	.25	-0.81	.42	0.42	0.64	23
Pearson (2008)	EAT-26 Oral Control	-.27	-.65	.23	-1.07	.28	0.31	0.57	18
Pearson (2008)	BULIT-R	.62	.18	.85	2.61	.01	0.27	0.53	16
Shields et al.(2008)	PESQ-PS	.42	-.73	.95	0.63	.53	0.04	0.14	5
Shields et al. (2008)	RAPI	.26	-.70	.88	0.46	.64	0.06	0.20	6
Shields et al. (2008)	DUSI-SUS	.11	-.77	.85	0.19	.85	0.06	0.20	6
Shields et al. (2008)	POSIT-ADS	-.72	-.98	.44	-1.28	.20	0.04	0.14	5
Shields et al. (2008)	AAIS	-.22	-.75	.48	-0.59	.55	0.15	0.37	10
Shields et al.(2008)	alle sechs Instrumente	-.12	-.44	.22	-0.68	.50	0.67	0.75	35
Vassar (2008a)	SWLS	.28	-.07	.57	1.58	.12	0.62	0.74	33
Vassar (2008b)	SOGS	-.10	-.60	.45	-0.33	.74	0.23	0.49	14
Vassar & Crosby (2008)	UCLA-R	.13	-.09	.34	1.15	.25	1.62	0.90	81
Victorson et al. (2008)	FACT-G	.71	.23	.91	2.66	.01	0.19	0.43	12
Victorson et al. (2008)	FACT-G Physical Wellbeing	.44	-.07	.77	1.70	.09	0.27	0.53	16
Victorson et al. (2008)	FACT-G Social Wellbeing	.39	-.15	.75	1.43	.15	0.25	0.51	15
Victorson et al. (2008)	FACT-G Functional Wellbeing	.30	-.25	.70	1.07	.28	0.25	0.51	15
Victorson et al. (2008)	FACT-G Emotional Wellbeing	.27	-.28	.69	0.96	.34	0.25	0.51	15
Graham & Christiansen (2009)	Rubin Loving	.01	-.25	.27	0.07	.94	1.12	0.85	57
Graham & Christiansen (2009)	Rubin Liking	-.02	-.35	.32	-0.11	.91	0.67	0.75	35
Graham &	LAS Eros	.14	-.28	.51	0.65	.52	0.44	0.65	24

Christiansen (2009)									
Graham & Christiansen (2009)	LAS Ludus	.17	-.24	.53	0.81	.42	0.46	0.66	25
Graham & Christiansen (2009)	LAS Storge	.30	-.12	.63	1.42	.16	0.44	0.65	24
Graham & Christiansen (2009)	LAS Pragma	-.12	-.48	.28	-0.58	.56	0.48	0.68	26
Graham & Christiansen (2009)	LAS Mania	-.01	-.42	.40	-0.04	.96	0.42	0.64	23
Graham & Christiansen (2009)	LAS Agape	.16	-.26	.53	0.74	.46	0.44	0.65	24
Graham & Christiansen (2009)	LAS-SF Eros	.02	-.50	.53	0.07	.94	0.25	0.51	15
Graham & Christiansen (2009)	LAS -SF Ludus	-.22	-.65	.31	-0.81	.42	0.27	0.53	16
Graham & Christiansen (2009)	LAS -SF Storge	.03	-.47	.52	0.11	.91	0.27	0.53	16
Graham & Christiansen (2009)	LAS -SF Pragma	.50	.03	.79	2.06	.04	0.29	0.55	17
Graham & Christiansen (2009)	LAS -SF Mania	.07	-.44	.55	0.25	.80	0.27	0.53	16
Graham & Christiansen (2009)	LAS -SF Agape	.24	-.24	.63	0.98	.33	0.33	0.58	19
Graham & Christiansen (2009)	PLS 30 item	-.64	-.94	.22	-1.52	.13	0.08	0.25	7
Graham & Christiansen (2009)	PLS 15 item	.45	-.46	.90	0.97	.33	0.08	0.25	7
Graham & Christiansen (2009)	TLS Passion	.02	-.62	.64	0.05	.96	0.15	0.37	10
Graham & Christiansen (2009)	TLS Intimacy	.18	-.47	.70	0.51	.61	0.17	0.40	11

Graham & Christiansen (2009)	TLS Commitment	.06	-.56	.64	0.17	.87	0.17	0.40	11
Huynh et al. (2009)	Nondominant Alle	-.05	-.35	.26	-0.31	.76	0.79	0.79	41
Huynh et al. (2009)	Dominant Alle	-.40	-.62	-.12	-2.75	.01	0.87	0.81	45
Lopez-Pina et al. (2009)	HAM-D	-.20	-.53	.18	-1.03	.30	0.54	0.70	29
Miller, B. K. et al. (2009)	POPS	.01	-.22	.23	0.09	.93	1.52	0.90	76
Ha et al. (2010)	F-MPS	.15	-.16	.43	0.96	.34	0.83	0.80	43
Ha et al. (2010)	F-MPS Concern over Mistakes	.38	.12	.59	2.86	.00	1.06	0.84	54
Ha et al. (2010)	F-MPS Personal Standards	.25	-.06	.51	1.60	.11	0.81	0.79	42
Ha et al. (2010)	F-MPS Parental Expectation	.17	-.22	.51	0.86	.39	0.52	0.70	28
Ha et al. (2010)	F-MPS Parental Criticism	.10	-.29	.46	0.49	.62	0.50	0.69	27
Ha et al. (2010)	F-MPS Doubt about Actions	.32	.01	.57	2.04	.04	0.79	0.79	41
Ha et al. (2010)	F-MPS Organization	.25	-.12	.56	1.33	.18	0.56	0.71	30
Vassar & Bradley (2010)	LOT	.14	-.23	.47	0.75	.46	0.58	0.72	31
Vassar & Bradley (2010)	LOT Optimism	-.02	-.54	.52	-0.07	.95	0.23	0.49	14
Vassar & Bradley (2010)	LOT Pessimism	-.23	-.69	.37	-0.74	.46	0.21	0.46	13
Vassar & Bradley (2010)	LOT-R	.10	-.12	.31	0.90	.37	1.69	0.91	84
Vassar & Bradley (2010)	LOT-R Optimism	.14	-.24	.48	0.72	.47	0.54	0.70	29
Vassar & Bradley (2010)	LOT-R Pessimism	-.02	-.40	.37	-0.10	.92	0.48	0.68	26
Graham et al. (2011)	LWMAT	.15	-.03	.32	1.64	.10	2.46	0.95	121
Graham et al. (2011)	KMS	-.01	-.20	.18	-0.10	.92	2.08	0.93	103
Graham et al. (2011)	QMI	.15	.01	.29	2.03	.04	3.75	0.98	183
Graham et al. (2011)	RAS	.18	.04	.32	2.47	.01	3.83	0.98	187
Graham et al. (2011)	MOQ	.15	-.34	.58	0.59	.56	0.31	0.57	18

Kieffer & MacDonald (2011)	WAYS	-.06	-.47	.37	-0.26	.79	0.40	0.63	22
Kieffer & MacDonald (2011)	WAYS Confrontive Coping	-.12	-.40	.18	-0.77	.44	0.85	0.80	44
Kieffer & MacDonald (2011)	WAYS Distancing	.05	-.25	.34	0.32	.75	0.85	0.80	44
Kieffer & MacDonald (2011)	WAYS Self-Controlling	-.03	-.32	.27	-0.19	.85	0.85	0.80	44
Kieffer & MacDonald (2011)	WAYS Seeking Social Support	-.22	-.49	.08	-1.43	.15	0.85	0.80	44
Kieffer & MacDonald (2011)	WAYS Accepting Responsibility	.12	-.18	.40	0.77	.44	0.85	0.80	44
Kieffer & MacDonald (2011)	WAYS Escape-Avoidance	-.05	-.35	.25	-0.32	.75	0.83	0.80	43
Kieffer & MacDonald (2011)	WAYS Planful Problem Solving	-.21	-.48	.09	-1.36	.17	0.85	0.80	44
Kieffer & MacDonald (2011)	WAYS Positive Reappraisal	.03	-.27	.32	0.19	.85	0.85	0.80	44
Vassar et al. (2011)	IES	-.08	-.35	.20	-0.56	.57	1.02	0.83	52
Vassar et al. (2011)	IES Intrusion	.49	.25	.67	3.75	.00	1.02	0.83	52
Vassar et al. (2011)	IES Avoidance	.32	.05	.55	2.32	.02	1.02	0.83	52
Fixed-Effects		.13	.10	.16	9.16	.00			
Random-Effects		.12	.06	.17	4.14	.00			

Abbildung A 4: Funnel plot zu Geschlecht



5.3.5. Klinische Stichprobe

Tabelle A 7: Klinische Stichprobe

Autorinnen und Autoren	Skala	Correlation	Lower limit	Upper limit	Z-Value	p-Value	Weight (Fixed)	Weight (Random)	Sample Size
Yin & Fan (2000)	BDI	-.09	-.26	.08	-1.02	.31	6.00	2.09	131
Lane et al. (2002)	CSEI	.43	.21	.61	3.65	.00	2.95	1.89	66
Reese et al. (2002)	AAS Depend	-.70	-.94	.02	-1.92	.05	0.23	0.60	8
Reese et al. (2002)	AAS Close	.05	-.68	.73	0.12	.90	0.23	0.60	8
Reese et al. (2002)	AAS Anxiety	-.58	-.91	.21	-1.49	.14	0.23	0.60	8
Reese et al. (2002)	BORI Alienation	-.49	-.93	.53	-0.94	.35	0.14	0.40	6

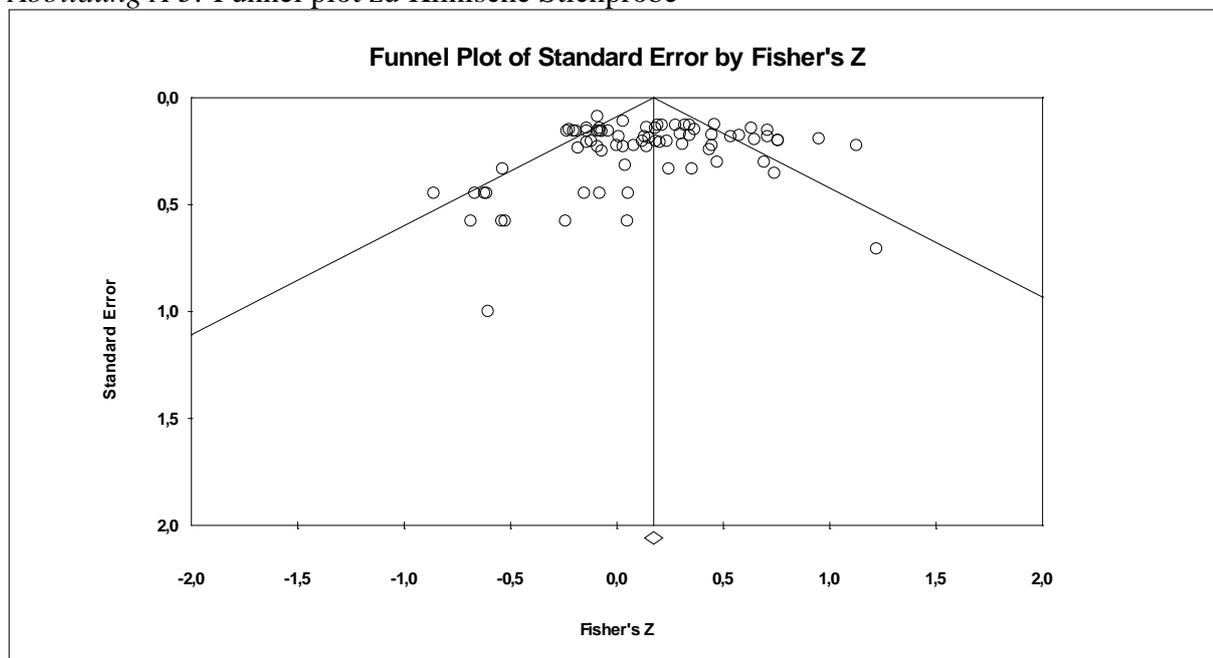
Reese et al. (2002)	BORI Social Incompetence	-.60	-.95	.42	-1.19	.24	0.14	0.40	6
Reese et al. (2002)	BORI Insecure	-.24	-.88	.71	-0.42	.68	0.14	0.40	6
Reese et al. (2002)	BORI Egocentric	-.48	-.93	.54	-0.91	.36	0.14	0.40	6
Reese et al. (2002)	PAQ Affective Quality	-.15	-.77	.62	-0.34	.73	0.23	0.60	8
Reese et al. (2002)	PAQ Parental Fostering	-.55	-.90	.25	-1.39	.16	0.23	0.60	8
Reese et al. (2002)	PAQ Parental Support	-.55	-.90	.26	-1.37	.17	0.23	0.60	8
Reese et al. (2002)	PBI Maternal Overprotection	.23	-.16	.56	1.16	.25	1.12	1.45	27
Reese et al. (2002)	PBI Maternal Care	-.12	-.48	.27	-0.59	.56	1.12	1.45	27
Reese et al. (2002)	PBI Paternal Overprotection	.18	-.21	.53	0.90	.37	1.12	1.45	27
Reese et al. (2002)	PBI Paternal Care	.12	-.27	.48	0.59	.55	1.12	1.45	27
Shields & Caruso (2003)	AUDIT	.30	-.12	.63	1.42	.16	0.98	1.38	24
Shields & Caruso (2004)	CAGE	.41	-.04	.72	1.80	.07	0.80	1.26	20
Miller, C. S. et al. (2007)	MAC	-.14	-.50	.26	-0.68	.50	1.08	1.43	26
Rouse (2007)	MMPI-2 Aggressivness	.21	-.04	.44	1.65	.10	2.81	1.87	63
Rouse (2007)	MMPI-2 Psychoticism	.31	.07	.52	2.48	.01	2.81	1.87	63
Rouse (2007)	MMPI-2 Disconstraint	.27	.02	.49	2.14	.03	2.81	1.87	63
Rouse (2007)	MMPI-2 Negative Emotionality/Neuroticism	.19	-.06	.42	1.49	.14	2.81	1.87	63
Rouse (2007)	MMPI-2 Introversion/Low positive Emotionality	.33	.09	.53	2.66	.01	2.81	1.87	63
Shields et al. (2007)	MAST	.35	.07	.58	2.45	.01	2.11	1.76	48
Pearson (2008)	EDI Total	.05	-.79	.83	0.09	.93	0.14	0.40	6
Pearson (2008)	EDI Body Dissatisfaction	.14	-.13	.39	1.01	.31	2.39	1.81	54
Pearson (2008)	EDI Drive for	-.22	-.47	.07	-1.50	.13	2.11	1.76	48

	Thinness								
Pearson (2008)	EDI Bulimia	.61	.39	.76	4.65	.00	2.01	1.74	46
Pearson (2008)	EDI Perfectionism	.61	.34	.79	3.88	.00	1.41	1.57	33
Pearson (2008)	EDI Interoceptive Awareness	.57	.26	.77	3.30	.00	1.22	1.50	29
Pearson (2008)	EDI Ineffectiveness	.64	.35	.82	3.79	.00	1.17	1.48	28
Pearson (2008)	EDI Maturity Fears	.74	.52	.87	4.94	.00	1.27	1.52	30
Pearson (2008)	EDI Interpersonal Distrust	.64	.35	.82	3.79	.00	1.17	1.48	28
Pearson (2008)	EDI Asceticism	.63	.05	.89	2.10	.04	0.37	0.83	11
Pearson (2008)	EDI Impuls Regulation	-.49	-.83	.12	-1.61	.11	0.42	0.90	12
Pearson (2008)	EDI Social Insecurity	.24	-.39	.72	0.73	.46	0.42	0.90	12
Pearson (2008)	EAT-40	.60	.10	.86	2.30	.02	0.52	1.01	14
Pearson (2008)	EAT-26	.56	.34	.72	4.43	.00	2.30	1.80	52
Shields et al. (2008)	PESQ-PS	.84	-.16	.99	1.73	.08	0.09	0.29	5
Shields et al. (2008)	RAPI	-.08	-.74	.66	-0.18	.86	0.23	0.60	8
Shields et al. (2008)	DUSI-SUS	-.54	-.99	.88	-0.60	.55	0.05	0.15	4
Shields et al. (2008)	AAIS	.34	-.29	.76	1.06	.29	0.42	0.90	12
Shields et al. (2008)	alle sechs Instrumente	.29	-.03	.56	1.77	.08	1.64	1.65	38
Vassar (2008a)	SWLS	.13	-.22	.45	0.72	.47	1.41	1.57	33
Lopez-Pina et al. (2009)	HAM-D	.33	.00	.60	1.94	.05	1.50	1.60	35
Nassen (2009)	CBCL	.49	.18	.71	2.94	.00	1.41	1.57	33
Nassen (2009)	CBCL Internalizing	.42	.11	.66	2.57	.01	1.55	1.62	36
Nassen (2009)	CBCL Externalizing	.52	.23	.73	3.26	.00	1.50	1.60	35
Nassen (2009)	CBCL Withdrawn	.00	-.41	.41	0.00	1.00	0.94	1.35	23
Nassen (2009)	CBCL Somatic	.14	-.30	.53	0.61	.54	0.89	1.32	22
Nassen (2009)	CBCL Anxious	.42	.01	.71	2.00	.05	0.94	1.35	23
Nassen (2009)	CBCL Social Problems	.03	-.40	.45	0.13	.90	0.89	1.32	22
Nassen (2009)	CBCL Thought problems	-.07	-.51	.40	-0.28	.78	0.75	1.23	19
Nassen (2009)	CBCL Attention	-.18	-.57	.27	-0.77	.44	0.84	1.29	21

Nassen (2009)	CBCL Delinquent	.08	-.34	.48	0.36	.72	0.94	1.35	23
Nassen (2009)	CBCL Aggression	.81	.60	.92	5.04	.00	0.94	1.35	23
Vassar & Bradley (2010)	LOT	.01	-.33	.35	0.05	.96	1.41	1.57	33
Vassar & Bradley (2010)	LOT Optimism	.44	-.12	.79	1.57	.12	0.52	1.01	14
Vassar & Bradley (2010)	LOT Pessimism	.04	-.52	.58	0.13	.90	0.47	0.96	13
Vassar & Bradley (2010)	LOT-R	.03	-.18	.24	0.27	.79	3.84	1.98	85
Vassar & Bradley (2010)	LOT-R Optimism	.15	-.22	.48	0.80	.42	1.31	1.54	31
Vassar & Bradley (2010)	LOT-R Pessimism	.20	-.20	.55	0.97	.33	1.08	1.43	26
Kieffer & MacDonald (2011)	WAYS	-.09	-.49	.34	-0.39	.69	0.89	1.32	22
Kieffer & MacDonald (2011)	WAYS Confrontive Coping	-.14	-.42	.16	-0.90	.37	1.92	1.72	44
Kieffer & MacDonald (2011)	WAYS Distancing	-.09	-.38	.21	-0.58	.56	1.92	1.72	44
Kieffer & MacDonald (2011)	WAYS Self- Controlling	-.08	-.37	.22	-0.51	.61	1.92	1.72	44
Kieffer & MacDonald (2011)	WAYS Seeking Social Support	-.20	-.47	.10	-1.30	.19	1.92	1.72	44
Kieffer & MacDonald (2011)	WAYS Accepting Responsibility	-.04	-.33	.26	-0.26	.80	1.92	1.72	44
Kieffer & MacDonald (2011)	WAYS Escape- Avoidance	-.07	-.36	.24	-0.44	.66	1.87	1.71	43
Kieffer & MacDonald (2011)	WAYS Planful Problem Solving	-.23	-.49	.07	-1.50	.13	1.92	1.72	44
Kieffer & MacDonald (2011)	WAYS Positive Reappraisal	-.19	-.46	.11	-1.23	.22	1.92	1.72	44
Vassar et al. (2011)	IES	-.08	-.35	.20	-0.56	.57	2.30	1.80	52
Vassar et al. (2011)	IES Intrusion	-.14	-.40	.14	-0.99	.32	2.30	1.80	52

Vassar et al. (2011)	IES Avoidance	.18	-.10	.43	1.27	.20	2.30	1.80	52
Fixed-Effects		.17	.13	.21	7.97	.00			
Random-Effects		.17	.09	.25	4.29	.00			

Abbildung A 5: Funnel plot zu Klinische Stichprobe



5.3.6. Standardabweichung des Score

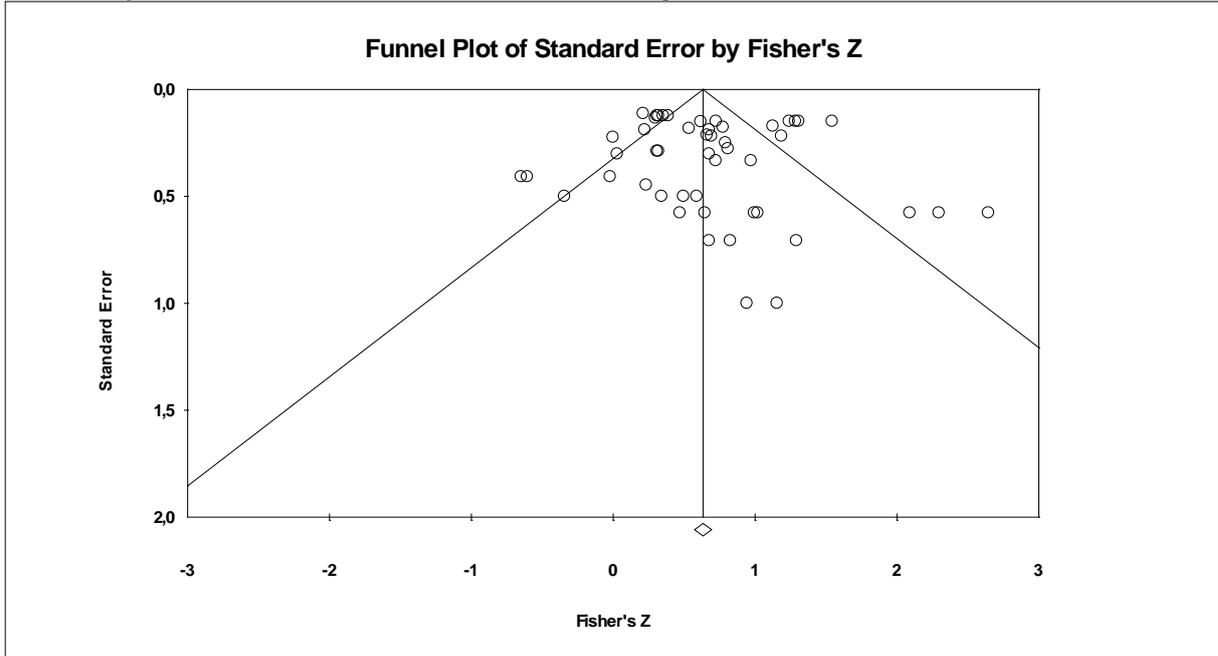
Tabelle A 8: Standardabweichung des Scores

Autorinnen und Autoren	Skala	Correlation	Lower limit	Upper limit	Z-Value	p-Value	Weight (Fixed)	Weight (Random)	Sample Size
Capraro, M. M. et al. (2001)	MARS	.31	-.24	.71	1.11	.27	1.13	2.21	15
Caruso et al. (2001)	EPQ Psychoticism	.37	.15	.56	3.16	.00	6.20	2.99	69
Caruso et al. (2001)	EPQ Extraversion	.30	.07	.50	2.51	.01	6.20	2.99	69
Caruso et al. (2001)	EPQ Neurotizismus	.31	.08	.51	2.60	.01	6.20	2.99	69

Caruso et al. (2001)	EPQ Lie	.34	.11	.53	2.88	.00	6.20	2.99	69
Henson et al. (2001)	STEBI PSTE	.68	-.51	.98	1.17	.24	0.19	0.85	5
Henson et al. (2001)	TES PTE	.86	-.09	.99	1.83	.07	0.19	0.85	5
Henson et al. (2001)	TES GTE	.74	-.77	.99	0.94	.35	0.09	0.49	4
Barnes et al. (2002)	STAI Trait	.59	.30	.78	3.59	.00	2.63	2.70	31
Barnes et al. (2002)	STAI State	.22	-.15	.53	1.18	.24	2.63	2.70	31
Nilsson et al. (2002)	CDMSE	-.02	-.68	.65	-0.05	.96	0.56	1.68	9
Nilsson et al. (2002)	CDMSE Short Form	.76	-.13	.97	1.73	.08	0.28	1.13	6
Wallace & Wheeler (2002)	LSI	.00	-.41	.41	0.00	1.00	1.88	2.54	23
Shields & Caruso (2003)	AUDIT	.83	.64	.92	5.44	.00	1.97	2.56	24
Shields & Caruso (2004)	CAGE	-.33	-.87	.56	-0.69	.49	0.38	1.35	7
Campbell et al. (2005)	PCL:YV	.59	.09	.85	2.25	.02	1.03	2.15	14
De Ayala et al. (2005)	BAI	.81	.66	.90	6.57	.00	3.19	2.79	37
Ryngala et al. (2005)	RCMAS	.85	.74	.91	8.33	.00	4.23	2.89	48
Ryngala et al. (2005)	RCMAS Physiological	.86	.77	.92	8.78	.00	4.23	2.89	48
Ryngala et al. (2005)	RCMAS Worry & Oversensitivity	.86	.76	.92	8.62	.00	4.23	2.89	48
Ryngala et al. (2005)	RCMAS Concentration	.91	.85	.95	10.37	.00	4.23	2.89	48
Ryngala et al. (2005)	RCMAS Lie	.62	.41	.77	4.87	.00	4.23	2.89	48
Leach et al. (2006)	SDQ I Parent Relations	.75	.31	.93	2.92	.00	0.85	2.00	12
Leach et al. (2006)	SDQ I Physical Ability	.62	.07	.88	2.18	.03	0.85	2.00	12
Leach et al. (2006)	SDQ II Verbal	-.57	-.90	.15	-1.59	.11	0.56	1.68	9
Leach et al. (2006)	SDQ II General Self-Concept	-.54	-.89	.19	-1.48	.14	0.56	1.68	9
Leach et al. (2006)	SDQ III General School	.59	-.61	.97	0.96	.34	0.19	0.85	5

Leach et al. (2006)	SDQ III Peer Relations/Same Sex	.67	.26	.88	2.92	.00	1.22	2.27	16
Miller, C. S. et al. (2007)	MAC	.46	-.45	.90	0.99	.32	0.38	1.35	7
Shields et al. (2007)	MAST	.58	.24	.79	3.11	.00	2.07	2.59	25
Shields et al. (2008)	POSIT-ADS	.82	-.67	1.00	1.16	.25	0.09	0.49	4
Shields et al. (2008)	AAIS	.33	-.56	.87	0.69	.49	0.38	1.35	7
Vassar (2008a)	SWLS	.49	.18	.71	2.94	.00	2.82	2.74	33
Vassar & Crosby (2008)	UCLA-R	.55	.31	.72	4.10	.00	4.13	2.88	47
Lopez-Pina et al. (2009)	HAM-D	.65	.40	.81	4.39	.00	3.00	2.76	35
Miller, B. K. et al. (2009)	POPS	.21	-.01	.41	1.89	.06	7.42	3.03	82
Ha et al. (2010)	F-MPS	.66	.29	.86	3.17	.00	1.50	2.40	19
Ha et al. (2010)	F-MPS Concern over Mistakes	.23	-.57	.80	0.52	.60	0.47	1.53	8
Ha et al. (2010)	F-MPS Personal Standards	.53	-.37	.92	1.18	.24	0.38	1.35	7
Ha et al. (2010)	F-MPS Parental Expectation	.44	-.58	.92	0.82	.41	0.28	1.13	6
Ha et al. (2010)	F-MPS Parental Criticism	.97	.74	1.00	3.62	.00	0.28	1.13	6
Ha et al. (2010)	F-MPS Doubt about Actions	.77	-.11	.97	1.77	.08	0.28	1.13	6
Ha et al. (2010)	F-MPS Organization	.98	.82	1.00	3.98	.00	0.28	1.13	6
Vassar & Bradley (2010)	LOT	.60	.26	.81	3.18	.00	1.97	2.56	24
Vassar & Bradley (2010)	LOT Optimism	.57	-.45	.94	1.12	.26	0.28	1.13	6
Vassar & Bradley (2010)	LOT Pessimism	.99	.91	1.00	4.58	.00	0.28	1.13	6
Vassar & Bradley (2010)	LOT-R	.29	.04	.51	2.23	.03	5.26	2.95	59
Vassar & Bradley (2010)	LOT-R Optimism	.30	-.25	.70	1.07	.28	1.13	2.21	15
Vassar & Bradley (2010)	LOT-R Pessimism	.03	-.51	.55	0.10	.92	1.03	2.15	14
Fixed-Effects		.56	.52	.60	20.68	.00			
Random-Effects		.57	.47	.67	8.61	.00			

Abbildung A 6: Funnel Plot zu Stabdardabweichung des Score



5.3.7. Sprache

Tabelle A 9: Sprache

Autorinnen und Autoren	Skala	Correlation	Lower limit	Upper limit	Z-Value	p-Value	Weight (Fixed)	Weight (Random)	Sample Size
Yin & Fan (2000)	BDI	.03	-.14	.20	0.34	.73	2.94	1.65	131
Caruso & Edwards (2001)	J-EPQ Psychoticism	.19	-.11	.46	1.23	.22	0.94	1.10	44
Caruso & Edwards (2001)	J-EPQ Extraversion	.23	-.07	.49	1.50	.13	0.94	1.10	44
Caruso & Edwards (2001)	J-EPQ Neuroticism	-.21	-.48	.09	-1.36	.17	0.94	1.10	44
Caruso & Edwards (2001)	J-EPQ Lie	-.36	-.59	-.07	-2.41	.02	0.94	1.10	44
Caruso et al. (2001)	EPQ Psychoticism	-.12	-.35	.12	-0.98	.33	1.52	1.35	69
Caruso et al.	EPQ Extraversion	-.08	-.31	.16	-0.65	.51	1.52	1.35	69

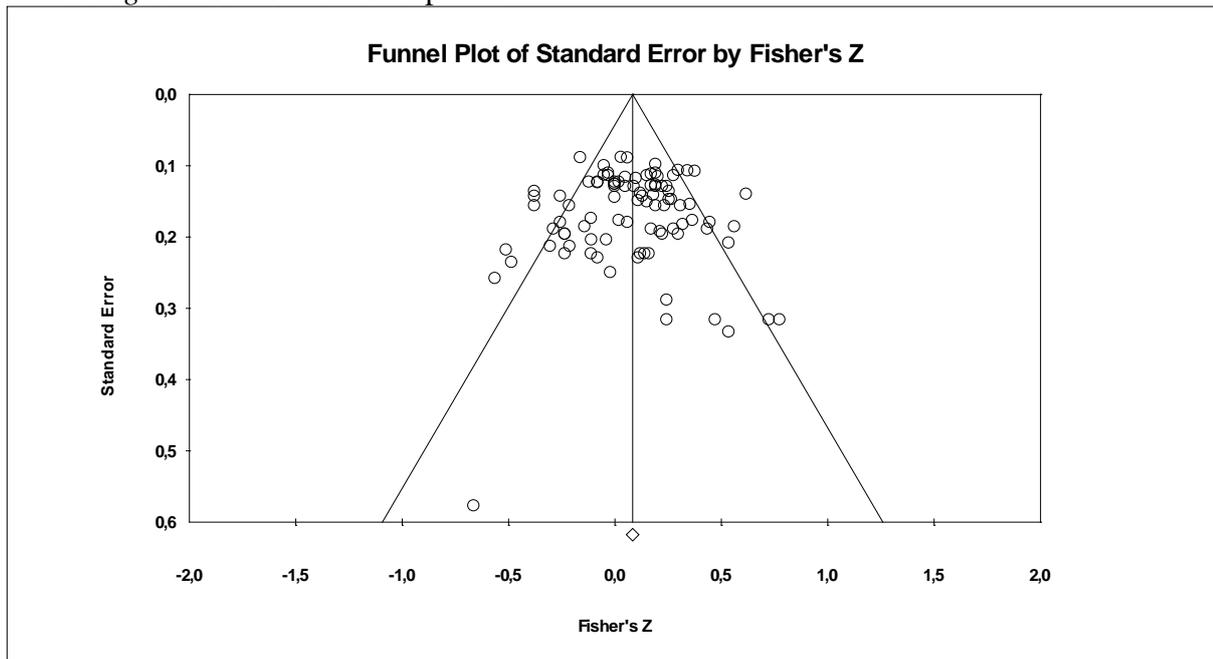
(2001)									
Caruso et al. (2001)	EPQ Neuroticism	.00	-.24	.24	0.00	1.00	1.52	1.35	69
Caruso et al. (2001)	EPQ Lie	.02	-.22	.26	0.16	.87	1.52	1.35	69
Barnes et al. (2002)	STAI Trait	.00	-.28	.28	0.00	1.00	1.10	1.19	51
Barnes et al. (2002)	STAI State	.18	-.09	.43	1.29	.20	1.15	1.21	53
Reese et al. (2002)	PBI Maternal Overprotection	-.11	-.47	.28	-0.54	.59	0.55	0.82	27
Reese et al. (2002)	PBI Maternal Care	-.04	-.41	.35	-0.19	.85	0.55	0.82	27
Reese et al. (2002)	PBI Paternal Overprotection	-.29	-.62	.11	-1.42	.16	0.51	0.78	25
Reese et al. (2002)	PBI Paternal Care	-.21	-.56	.20	-0.99	.32	0.51	0.78	25
Wallace & Wheeler (2002)	LSI	-.14	-.47	.22	-0.76	.45	0.67	0.92	32
O'Rourke (2004)	CES-D	-.16	-.32	.01	-1.82	.07	2.92	1.64	130
Leach et al. (2006)	SDQ I	.22	-.16	.54	1.14	.25	0.60	0.86	29
Leach et al. (2006)	SDQ I General Self-Concept	.29	-.09	.59	1.52	.13	0.60	0.86	29
Leach et al. (2006)	SDQ I Math	-.36	-.57	-.11	-2.77	.01	1.24	1.25	57
Bachner & O'Rourke (2007)	ZBI	.06	-.11	.23	0.67	.50	2.90	1.64	129
Li & Bagger (2007)	BIDR Impression Management	.19	.00	.37	1.96	.05	2.39	1.56	107
Li & Bagger (2007)	BIDR Self-Deception Enhancement	.33	.13	.50	3.20	.00	2.00	1.48	90
Rouse (2007)	MMPI-2 Aggressivness	.24	-.01	.46	1.90	.06	1.38	1.30	63
Rouse (2007)	MMPI-2 Psychoticism	.19	-.06	.42	1.49	.14	1.38	1.30	63
Rouse (2007)	MMPI-2 Disconstraint	.05	-.20	.29	0.39	.70	1.38	1.30	63
Rouse (2007)	MMPI-2 Negative Emotionality/Neuroticism	.09	-.16	.33	0.70	.48	1.38	1.30	63
Rouse (2007)	MMPI-2 Introversion/Low positive	.22	-.03	.44	1.73	.08	1.38	1.30	63

	Emotionality								
Shields et al. (2007)	MAST	.11	-.18	.38	0.74	.46	1.03	1.15	48
Beretvas et al. (2008)	I-E	.17	-.08	.40	1.34	.18	1.40	1.31	64
Pearson (2008)	EDI Total	-.58	-.95	.44	-1.15	.25	0.07	0.16	6
Pearson (2008)	EDI Body Dissatisfaction	.12	-.15	.37	0.87	.38	1.20	1.23	55
Pearson (2008)	EDI Drive for Thinness	.26	-.02	.50	1.80	.07	1.06	1.17	49
Pearson (2008)	EDI Bulimia	.15	-.14	.42	1.00	.32	1.01	1.14	47
Pearson (2008)	EDI Perfectionism	.06	-.28	.39	0.33	.74	0.71	0.96	34
Pearson (2008)	EDI Interoceptive Awareness	.41	.07	.67	2.31	.02	0.64	0.90	31
Pearson (2008)	EDI Ineffectiveness	-.23	-.55	.15	-1.19	.23	0.60	0.86	29
Pearson (2008)	EDI Maturity Fears	-.28	-.58	.08	-1.52	.13	0.64	0.90	31
Pearson (2008)	EDI Interpersonal Distrust	-.23	-.55	.15	-1.19	.23	0.60	0.86	29
Pearson (2008)	EDI Asceticism	.49	-.12	.83	1.61	.11	0.21	0.41	12
Pearson (2008)	EDI Impuls Regulation	.24	-.36	.70	0.77	.44	0.23	0.44	13
Pearson (2008)	EDI Social Insecurity	.65	.15	.88	2.45	.01	0.23	0.44	13
Pearson (2008)	EAT-40	.24	-.31	.67	0.85	.40	0.28	0.51	15
Pearson (2008)	EAT-26	.55	.33	.71	4.42	.00	1.17	1.22	54
Pearson (2008)	EAT-26 Dieting	-.47	-.73	-.08	-2.34	.02	0.48	0.76	24
Pearson (2008)	EAT-26 Bulimia and Food Preoccupation	-.23	-.59	.20	-1.05	.29	0.46	0.73	23
Pearson (2008)	EAT-26 Oral Control	-.51	-.79	-.06	-2.18	.03	0.34	0.60	18
Rexrode et al. (2008)	WOCS Accepting Responsibility	.10	-.13	.32	0.85	.39	1.66	1.40	75
Rexrode et al. (2008)	WOCS Confrontive Coping	.05	-.18	.27	0.43	.67	1.70	1.41	77
Rexrode et al. (2008)	WOCS Distancing	.15	-.07	.36	1.33	.18	1.77	1.43	80
Rexrode et al. (2008)	WOCS Escape-Avoidance	.17	-.05	.37	1.54	.12	1.84	1.45	83
Rexrode et al. (2008)	WOCS Planful Problem Solving	-.03	-.25	.19	-0.26	.79	1.77	1.43	80
Rexrode et al.	WOCS Positive	.20	-.02	.40	1.76	.08	1.72	1.41	78

(2008)	Reappraisal								
Rexrode et al. (2008)	WOCS Self-Controlling	.27	.05	.46	2.43	.02	1.77	1.43	80
Rexrode et al. (2008)	WOCS Seeking Social Support	-.05	-.24	.14	-0.50	.62	2.30	1.55	103
Vassar (2008a)	SWLS	.31	-.04	.59	1.76	.08	0.69	0.94	33
Vassar & Crosby (2008)	UCLA-R	-.05	-.27	.17	-0.44	.66	1.79	1.43	81
Victorson et al. (2008)	FACT-G	.00	-.25	.25	0.00	1.00	1.38	1.30	63
Victorson et al. (2008)	FACT-G Physical Wellbeing	-.08	-.31	.16	-0.65	.52	1.49	1.34	68
Victorson et al. (2008)	FACT-G Social Wellbeing	.00	-.24	.24	0.00	1.00	1.45	1.33	66
Victorson et al. (2008)	FACT-G Functional Wellbeing	.19	-.06	.41	1.51	.13	1.43	1.32	65
Victorson et al. (2008)	FACT-G Emotional Wellbeing	.00	-.24	.24	0.00	1.00	1.45	1.33	66
Lopez-Pina et al. (2009)	HAM-D	.35	.02	.61	2.07	.04	0.74	0.97	35
Nassen (2009)	CBCL	-.25	-.54	.10	-1.42	.16	0.71	0.96	34
Nassen (2009)	CBCL Internalizing	-.11	-.42	.23	-0.63	.53	0.76	0.99	36
Nassen (2009)	CBCL Externalizing	.02	-.32	.35	0.11	.91	0.74	0.97	35
Nassen (2009)	CBCL Withdrawn	-.11	-.50	.32	-0.49	.62	0.46	0.73	23
Nassen (2009)	CBCL Somatic	.11	-.33	.51	0.48	.63	0.44	0.71	22
Nassen (2009)	CBCL Anxious	.14	-.29	.52	0.63	.53	0.46	0.73	23
Nassen (2009)	CBCL Social Problems	-.08	-.49	.35	-0.35	.73	0.44	0.71	22
Nassen (2009)	CBCL Thought Problems	-.02	-.47	.44	-0.08	.94	0.37	0.63	19
Nassen (2009)	CBCL Attention	-.45	-.74	-.02	-2.06	.04	0.41	0.68	21
Nassen (2009)	CBCL Delinquent	.12	-.31	.51	0.54	.59	0.46	0.73	23
Nassen (2009)	CBCL Aggression	.16	-.27	.54	0.72	.47	0.46	0.73	23
Ha et al. (2010)	F-MPS	.25	-.03	.50	1.73	.08	1.06	1.17	49
Ha et al. (2010)	F-MPS Concern over Mistakes	.25	-.01	.48	1.88	.06	1.24	1.25	57
Ha et al. (2010)	F-MPS Personal Standards	.30	.00	.55	1.98	.05	0.94	1.10	44
Ha et al. (2010)	F-MPS Parental	.17	-.20	.49	0.91	.36	0.64	0.90	31

	Expectation								
Ha et al. (2010)	F-MPS Parental Criticism	.21	-.16	.53	1.11	.27	0.62	0.88	30
Ha et al. (2010)	F-MPS Doubt about Actions	.34	.05	.58	2.29	.02	0.97	1.12	45
Ha et al. (2010)	F-MPS Organization	.42	.10	.66	2.49	.01	0.71	0.96	34
Vassar & Bradley (2010)	LOT	.51	.20	.73	3.03	.00	0.67	0.92	32
Vassar & Bradley (2010)	LOT Optimism	.44	-.15	.80	1.49	.14	0.23	0.44	13
Vassar & Bradley (2010)	LOT Pessimism	.62	.10	.87	2.29	.02	0.23	0.44	13
Vassar & Bradley (2010)	LOT-R	.19	-.02	.39	1.74	.08	1.89	1.46	85
Vassar & Bradley (2010)	LOT-R Optimism	.27	-.09	.57	1.47	.14	0.64	0.90	31
Vassar & Bradley (2010)	LOT-R Pessimism	.49	.13	.74	2.57	.01	0.53	0.80	26
Vassar et al. (2011)	IES	.13	-.15	.39	0.92	.36	1.13	1.20	52
Vassar et al. (2011)	IES Intrusion	-.25	-.49	.02	-1.79	.07	1.13	1.20	52
Vassar et al. (2011)	IES Avoidance	-.36	-.58	-.10	-2.64	.01	1.13	1.20	52
Wheeler et al. (2011)	MBI Emotional Exhaustion	.29	.09	.47	2.80	.01	2.02	1.49	91
Wheeler et al. (2011)	MBI Depersonalization	.36	.16	.53	3.50	.00	1.98	1.48	89
Wheeler et al. (2011)	MBI Personal Accomplishment	-.03	-.24	.18	-0.27	.79	1.89	1.46	85
Fixed-Effects		.08	.05	.11	5.48	.00			
Random-Effects		.08	.04	.13	3.48	.00			

Abbildung A 7: Funnel Plot zu Sprache



5.3.8. Stichprobengröße

Tabelle A 10: Stichprobengröße

Autorinnen und Autoren	Skala	Correlation	Lower limit	Upper limit	Z-Value	p-Value	Weight (Fixed)	Weight (Random)	Sample Size
Viswesvaran & Ones (2000)	alle Skalen	.17	.11	.23	5.80	.00	23.39	2.44	1145
Capraro, M. M. et al. (2001)	MARS	.07	-.31	.43	0.35	.73	0.51	0.84	28
Caruso & Edwards (2001)	J-EPQ Psychoticism	.02	-.28	.32	0.13	.90	0.84	1.14	44
Caruso & Edwards (2001)	J-EPQ Extraversion	-.01	-.31	.29	-0.06	.95	0.84	1.14	44
Caruso & Edwards (2001)	J-EPQ Neurotizismus	.00	-.30	.30	0.00	1.00	0.84	1.14	44
Caruso & Edwards (2001)	J-EPQ Lie	-.10	-.39	.20	-0.64	.52	0.84	1.14	44
Henson et al.	RSA +	.62	-.58	.97	1.01	.31	0.04	0.10	5

(2001)									
Henson et al. (2001)	RSA -	.64	-.56	.97	1.06	.29	0.04	0.10	5
Henson et al. (2001)	STEBI PSTE	.28	-.32	.72	0.92	.36	0.20	0.42	13
Henson et al. (2001)	STEBI STOE	.93	.75	.98	4.69	.00	0.16	0.35	11
Henson et al. (2001)	TES PTE	-.12	-.49	.29	-0.55	.58	0.45	0.77	25
Henson et al. (2001)	TES GTE	-.50	-.77	-.09	-2.32	.02	0.37	0.67	21
Barnes et al. (2002)	STAI Trait	.10	-.18	.37	0.70	.49	0.98	1.24	51
Barnes et al. (2002)	STAI State	.09	-.18	.35	0.64	.52	1.02	1.26	53
Hanson et al. (2002)	WAI Client	.12	-.28	.48	0.58	.56	0.47	0.79	26
Hanson et al. (2002)	WAI Client Goals	.39	-.09	.72	1.59	.11	0.31	0.58	18
Hanson et al. (2002)	WAI Client Tasks	.00	-.50	.50	0.00	1.00	0.27	0.52	16
Hanson et al. (2002)	WAI Client Bond	-.31	-.70	.22	-1.16	.25	0.27	0.52	16
Hanson et al. (2002)	WAI-S Client	.25	-.70	.88	0.44	.66	0.06	0.14	6
Hanson et al. (2002)	WAI Therapist	-.11	-.69	.56	-0.29	.77	0.14	0.31	10
Hanson et al. (2002)	WAI Therapist Goals	.41	-.21	.80	1.31	.19	0.18	0.38	12
Hanson et al. (2002)	WAI Therapist Tasks	.36	-.35	.81	1.00	.32	0.14	0.31	10
Hanson et al. (2002)	WAI Therapist Bond	.58	-.08	.89	1.75	.08	0.14	0.31	10
Hanson et al. (2002)	WAI Observer	-.21	-.87	.73	-0.37	.71	0.06	0.14	6
Henson & Hwang (2002)	LSI Concrete Experience	-.29	-.52	-.02	-2.07	.04	0.98	1.24	51
Henson & Hwang (2002)	LSI Reflective Observation	-.40	-.61	-.14	-2.90	.00	0.96	1.22	50
Henson & Hwang (2002)	LSI Active Experimentation	-.47	-.66	-.22	-3.50	.00	0.96	1.22	50
Henson & Hwang (2002)	LSI Abstract Conceptualization	-.16	-.42	.12	-1.12	.26	0.98	1.24	51
Lane et al. (2002)	CSEI	.09	-.17	.34	0.66	.51	1.11	1.31	57

Nilsson et al. (2002)	CDMSE	.02	-.65	.68	0.05	.96	0.12	0.27	9
Nilsson et al. (2002)	CDMSE Short Form	.13	-.76	.85	0.23	.82	0.06	0.14	6
Reese et al. (2002)	AAS Depend	.19	-.60	.79	0.42	.68	0.10	0.23	8
Reese et al. (2002)	AAS Close	-.81	-.96	-.23	-2.49	.01	0.10	0.23	8
Reese et al. (2002)	AAS Anxiety	.46	-.36	.88	1.12	.26	0.10	0.23	8
Reese et al. (2002)	BORI Alienation	.31	-.67	.90	0.56	.58	0.06	0.14	6
Reese et al. (2002)	BORI Social Incompetence	.40	-.61	.92	0.74	.46	0.06	0.14	6
Reese et al. (2002)	BORI Insecure	.10	-.77	.84	0.18	.86	0.06	0.14	6
Reese et al. (2002)	BORI Egocentric	.31	-.67	.90	0.55	.58	0.06	0.14	6
Reese et al. (2002)	IPPA Trust	.38	-.25	.78	1.19	.23	0.18	0.38	12
Reese et al. (2002)	IPPA Communication	.37	-.26	.78	1.17	.24	0.18	0.38	12
Reese et al. (2002)	IPPA Alienation	.32	-.31	.75	0.99	.32	0.18	0.38	12
Reese et al. (2002)	PAQ Affective Quality	.43	-.40	.87	1.02	.31	0.10	0.23	8
Reese et al. (2002)	PAQ Parental Fostering	-.13	-.76	.63	-0.29	.77	0.10	0.23	8
Reese et al. (2002)	PAQ Parental Support	-.26	-.82	.54	-0.60	.55	0.10	0.23	8
Reese et al. (2002)	PBI Maternal Overprotection	.28	-.11	.60	1.42	.16	0.49	0.82	27
Reese et al. (2002)	PBI Maternal Care	.14	-.26	.49	0.68	.50	0.49	0.82	27
Reese et al. (2002)	PBI Paternal Overprotection	.33	-.08	.64	1.59	.11	0.45	0.77	25
Reese et al. (2002)	PBI Paternal Care	.24	-.17	.58	1.15	.25	0.45	0.77	25
Thompson & Cook (2002)	LibQUAL Information Access	-.02	-.32	.28	-0.13	.90	0.82	1.12	43
Wallace & Wheeler (2002)	LSI	.00	-.34	.34	0.00	1.00	0.63	0.97	34
Kieffer et al. (2004)	AEQ	.21	-.23	.58	0.93	.35	0.39	0.69	22
Kieffer et al.	AEQ Global	.05	-.49	.57	0.17	.87	0.23	0.45	14

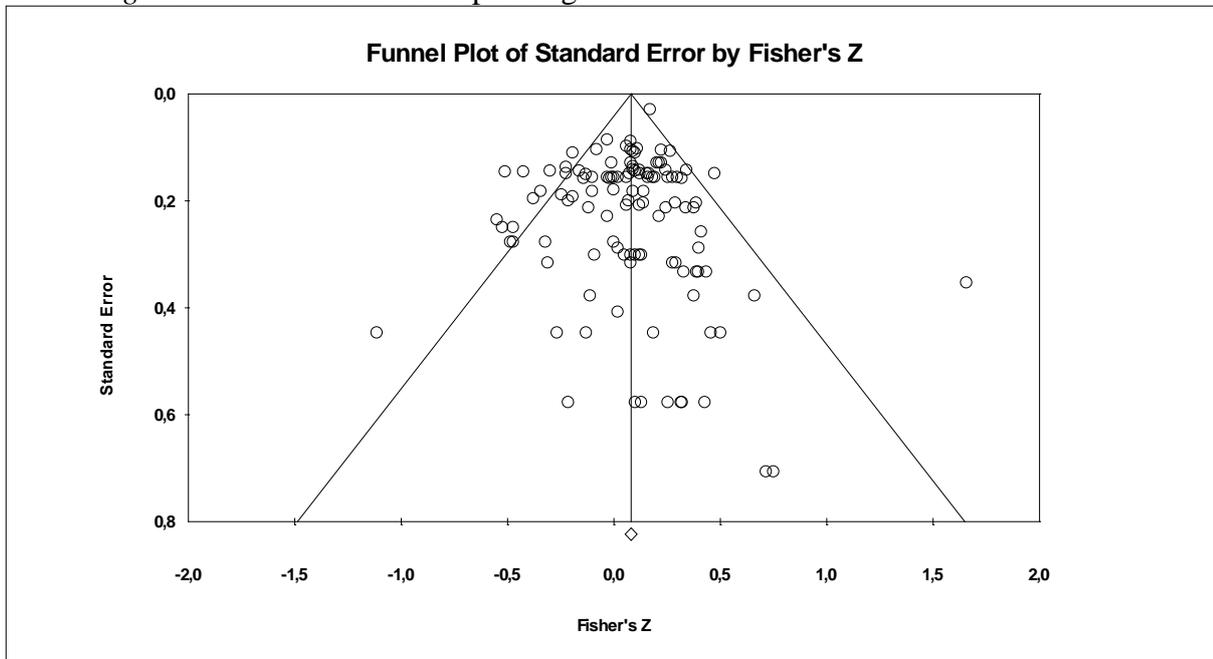
(2004)	Positive Changes in Experience								
Kieffer et al. (2004)	AEQ Sexual Enhancement	.13	-.43	.62	0.43	.66	0.23	0.45	14
Kieffer et al. (2004)	AEQ Social and Physical Pleasure	.27	-.33	.71	0.88	.38	0.20	0.42	13
Kieffer et al. (2004)	AEQ Assertiveness	.08	-.49	.60	0.25	.80	0.20	0.42	13
Kieffer et al. (2004)	AEQ Relaxation/Tension Reduction	.12	-.44	.61	0.40	.69	0.23	0.45	14
Kieffer et al. (2004) Power	AEQ Arousal/Interpersonal	-.09	-.59	.46	-0.30	.76	0.23	0.45	14
O'Rourke (2004)	CES-D	.08	-.09	.25	0.90	.37	2.60	1.82	130
De Ayala et al. (2005)	BAI	-.13	-.40	.16	-0.87	.39	0.90	1.18	47
Ross et al. (2005)	PALS	-.08	-.28	.12	-0.77	.44	1.88	1.64	95
Ryngala et al. (2005)	RCMAS	.16	-.13	.43	1.11	.27	0.92	1.20	48
Ryngala et al. (2005)	RCMAS Physiological	.15	-.14	.42	1.04	.30	0.92	1.20	48
Ryngala et al. (2005)	RCMAS Worry & Oversensitivity	.12	-.17	.39	0.82	.41	0.92	1.20	48
Ryngala et al. (2005)	RCMAS Concentration	.44	.18	.65	3.18	.00	0.92	1.20	48
Ryngala et al. (2005)	RCMAS Lie	.07	-.21	.35	0.50	.62	0.92	1.20	48
Hellman et al. (2006)	SPOS	-.10	-.43	.25	-0.55	.58	0.61	0.95	33
Leach et al. (2006)	SDQ I	-.36	-.64	.01	-1.92	.05	0.53	0.86	29
Leach et al. (2006)	SDQ I Academic	.37	-.01	.66	1.90	.06	0.49	0.82	27
Leach et al. (2006)	SDQ I Non-Academic	.36	-.04	.66	1.77	.08	0.45	0.77	25
Leach et al. (2006)	SDQ III Verbal	-.48	-.77	-.03	-2.09	.04	0.33	0.61	19
Leach et al. (2006)	SDQ III General School	-.44	-.75	.02	-1.89	.06	0.33	0.61	19
Leach et al. (2006)	SDQ III Parental Relations	-.45	-.77	.06	-1.75	.08	0.27	0.52	16
Leach et al. (2006)	SDQ III Physical Ability	.38	-.16	.75	1.39	.17	0.25	0.49	15
Leach et al.	SDQ III Problem	-.44	-.77	.07	-1.70	.09	0.27	0.52	16

(2006)	Solving								
Bachner & O'Rourke (2007)	ZBI	-.03	-.20	.14	-0.35	.73	2.77	1.85	138
Li & Bagger (2007)	BIDR Impression Management	.06	-.13	.25	0.61	.54	2.13	1.71	107
Li & Bagger (2007)	BIDR Self-Deception Enhancement	.09	-.12	.29	0.84	.40	1.78	1.61	90
Rouse (2007)	MMPI-2 Aggressivness	-.01	-.26	.24	-0.08	.94	1.23	1.38	63
Rouse (2007)	MMPI-2 Psychoticism	.20	-.05	.43	1.57	.12	1.23	1.38	63
Rouse (2007)	MMPI-2 Disconstraint	.08	-.17	.32	0.62	.53	1.23	1.38	63
Rouse (2007)	MMPI-2 Negative Emotionality/Neuroticism	.22	-.03	.44	1.73	.08	1.23	1.38	63
Rouse (2007)	MMPI-2 Introversion/Low positive Emotionality	.21	-.04	.44	1.65	.10	1.23	1.38	63
Vassar (2008a)	SWLS	.09	-.26	.42	0.49	.62	0.61	0.95	33
Vassar (2008b)	SOGS	.02	-.50	.53	0.07	.94	0.25	0.49	15
Vassar & Crosby (2008)	UCLA-R	-.19	-.39	.02	-1.74	.08	1.68	1.57	85
Miller, B. K. et al. (2009)	POPS	.08	-.12	.28	0.77	.44	1.88	1.64	95
Ha et al. (2010)	F-MPS	-.22	-.47	.07	-1.50	.13	0.92	1.20	48
Ha et al. (2010)	F-MPS Concern over Mistakes	-.22	-.46	.05	-1.63	.10	1.09	1.30	56
Ha et al. (2010)	F-MPS Personal Standards	-.14	-.42	.17	-0.89	.37	0.82	1.12	43
Ha et al. (2010)	F-MPS Parental Expectation	-.19	-.52	.18	-1.00	.32	0.55	0.88	30
Ha et al. (2010)	F-MPS Parental Criticism	-.21	-.54	.18	-1.07	.29	0.51	0.84	28
Ha et al. (2010)	F-MPS Doubt about Actions	-.03	-.32	.27	-0.19	.85	0.84	1.14	44
Ha et al. (2010)	F-MPS Organization	-.33	-.60	.02	-1.88	.06	0.61	0.95	33
Mahapoonyanont et al. (2010)	critical thinking	.08	-.47	.59	0.27	.79	0.23	0.45	14
Vassar & Bradley (2010)	LOT	.14	-.21	.46	0.77	.44	0.61	0.95	33
Vassar &	LOT Optimism	.10	-.45	.60	0.33	.74	0.23	0.45	14

Bradley (2010)									
Vassar & Bradley (2010)	LOT Pessimism	-.30	-.73	.30	-0.98	.33	0.20	0.42	13
Vassar & Bradley (2010)	LOT-R	.10	-.12	.31	0.91	.36	1.68	1.57	85
Vassar & Bradley (2010)	LOT-R Optimism	-.24	-.55	.12	-1.30	.20	0.57	0.91	31
Vassar & Bradley (2010)	LOT-R Pessimism	.06	-.34	.44	0.29	.77	0.47	0.79	26
Kieffer & MacDonald (2011)	WAYS	-.03	-.45	.40	-0.13	.90	0.39	0.69	22
Kieffer & MacDonald (2011)	WAYS Confrontive Coping	.27	-.03	.52	1.77	.08	0.84	1.14	44
Kieffer & MacDonald (2011)	WAYS Distancing	.29	-.01	.54	1.91	.06	0.84	1.14	44
Kieffer & MacDonald (2011)	WAYS Self-Controlling	.18	-.12	.45	1.17	.24	0.84	1.14	44
Kieffer & MacDonald (2011)	WAYS Seeking Social Support	.19	-.11	.46	1.23	.22	0.84	1.14	44
Kieffer & MacDonald (2011)	WAYS Accepting Responsibility	.25	-.05	.51	1.64	.10	0.84	1.14	44
Kieffer & MacDonald (2011)	WAYS Escape-Avoidance	.31	.01	.56	2.03	.04	0.82	1.12	43
Kieffer & MacDonald (2011)	WAYS Planful Problem Solving	.16	-.14	.44	1.03	.30	0.84	1.14	44
Kieffer & MacDonald (2011)	WAYS Positive Reappraisal	.06	-.24	.35	0.38	.70	0.84	1.14	44
Vassar et al. (2011)	IES	.33	.06	.55	2.40	.02	1.00	1.25	52
Vassar et al. (2011)	IES Intrusion	.24	-.04	.48	1.71	.09	1.00	1.25	52
Vassar et al. (2011)	IES Avoidance	.12	-.16	.38	0.84	.40	1.00	1.25	52
Wheeler et al. (2011)	MBI Emotional Exhaustion	.11	-.09	.30	1.08	.28	1.95	1.66	98
Wheeler et al. (2011)	MBI Depersonalization	.22	.02	.41	2.12	.03	1.84	1.63	93

Wheeler et al. (2011)	MBI Personal Accomplishment	.26	.06	.44	2.48	.01	1.78	1.61	90
Fixed		.08	.05	.11	5.65	.00			
Random		.06	.01	.10	2.51	.01			

Abbildung A 8: Funnel Plot zu Stichprobengröße



5.4. Autoren nach Anzahl der Publikationen

Tabelle A 11: Autorinnen und Autoren nach Anzahl der Publikationen

Caruso, J. C.	7	Campbell, J. S.	Meyers, J. L.
Shields, A. L.	7	Cella, D.	Miller, B. K.
Vassar, M.	6	Christiansen, K.	Mji, A.
Henson, R. K.	5	Cook, C.	Montoya, J. A.
Howell, R. T.	4	Cronin, C.	Muilenburg-Trevino, E. M.
Kieffer, K. M.	4	Crosby, J. W.	Murry, F.
Vacha-Haase, T.	4	Curry, K. T.	Mutz, R.
Graham, J. M.	3	Daniel, H.-D.	Nassen, L. M.
Kogan, L. R.	3	De Ayala, R. J.	Nilsson, J. E.
Miller, C. S.	3	Deditius-Island, H. K.	O'Toole, S.
Wallace, K. A.	3	Diebels, K. J.	Odom, L. R.
Worley, J. A.	3	Dunn, T. W.	Ones, D. S.
Barnes, L. L. B.	2	Durham, J. A.	Payne, S. C.
Beretvas, S. N.	2	Edwards, S.	Pearson, C. A.
Campfield, D. C.	2	Fan, X.	Petersen, S.
Capraro, M. M.	2	Fister, M. C.	Potter, J. S.
Capraro, R. M.	2	Forbes, S.	Puig, A.
Hellman, C. M.	2	Fuqua, D. R.	Pulos, S.
O'Rourke, N.	2	Gottlieb, J. D.	Rachasong, W.
Reese, R. J.	2	Green, K. W.	Rexrode, K. R.
Tani, C. R.	2	Gronnerod, C.	Rosa-Alcazar, A. I.
Thompson, B.	2	Ha, J. H.	Ross, M. E.
Weiss, R. D.	2	Hale, H.	Rouse, S. V.
Woodall, R. A.	2	Hale, W.	Rutherford, M. A.
Zangaro, G. A.	2	Hansen, A. M.	Ryngala, D. J.
		Hanson, W. E.	Sanchez-Meca, J.
		Harp, D.	Schmidt, C. K.
		Hogan, M.	Smith, T. B.
		Huynh, Q.-L.	Soeken, K. L.
		Hwang, D.-Y.	Song, J.
		Jeziorski, J. L.	Sudweeks, R. R.
		Jung, W. S.	Suizzo, M.-A.
		Kim, D.	Victorson, D.
		Knaup, K. G.	Viswesvaran, C.
		Kochakornjarupong, D.	Vonderharr-Carlson, D. J.
		Krahamwong, R.	Weins, R. D.
		Lane, G. G.	Wheeler, A. J.
		Leach, L. F.	Wheeler, D. L.
		Lee, S. M.	White, A. E.
		Leite, W. L.	Witkiewitz, K.
		Li, A.	Wilder, L. K.
		Liu, Y. J.	Woodson, J.
		Lopez-Pina, J. A.	Yarnell, L. M.
		MacDonald, G.	Yin, P.
		Mahapoonyanont, N.	Youngcourt, S. S.
		Meek, W. D.	Youngstrom, E. A.
			SUMME: 133
<u>Je eine Publikation von:</u>			
Alkhateeb, H. M.			
Bachner, Y. G.			
Bagger, J.			
Bandalos, D. L.			
Barlow, K.M.			
Barnow, Z. B.			
Barocas, J.			
Beaubien, J. M.			
Belcourt-Dittloff, A.			
Benet-Martinez, V.			
Blackburn, M.			
Bornmann, L.			
Bradley, G.			
Briggs, B. K.			
Byrne, Z. S.			
Cagle, L. S.			

5.5. *Fachzeitschriften nach Anzahl der Publikationen*

Tabelle A 12: Fachzeitschriften nach Anzahl der Publikationen

Journal	Anzahl	Prozent x/69
Educational and Psychological Measurement	35	50.72%
Journal of Personality Assessment	5	7.25%
Dissertation Abstracts International: Section B	2	2.90%
Psychological Reports	2	2.90%
South African Journal of Psychology	2	2.90%
Substance Use & Misuse	2	2.90%
Aging & Mental Health	1	1.45%
Assessment	1	1.45%
Education and Treatment of Children	1	1.45%
International Journal of Clinical and Health Psychology	1	1.45%
Journal of Applied Psychology	1	1.45%
Journal of Child & Adolescent Substance Abuse	1	1.45%
Journal of Counseling and Development	1	1.45%
Journal of Cross-Cultural Psychology	1	1.45%
Journal of Family Psychology	1	1.45%
Journal of Individual Differences	1	1.45%
Journal of Managerial Issues	1	1.45%
Journal of Marriage and Family	1	1.45%
Journal of Nursing Management	1	1.45%
Journal of Nursing Measurement	1	1.45%
Journal of Studies on Alcohol	1	1.45%
Personal Relationships	1	1.45%
Personality and Individual Differences	1	1.45%
PLoS ONE	1	1.45%
Procedia Social and Behavioral Sciences	1	1.45%
Qual Life Res	1	1.45%
Soc Indic Res	1	1.45%
SUMME: 27		

5.6. Datenbanken nach Nutzungshäufigkeit

Tabelle A 13: Datenbanken nach Nutzungshäufigkeit

Datenbank	#	Prozent x/68
PsycInfo	56	82.35%
ERIC	25	36.76%
PubMed/Medline	18	26.47%
Academic Search Elite	6	8.82%
Social Science Citation Index (SSCI)	5	7.35%
PsycLit	4	5.88%
Sociological Abstracts	4	5.88%
Cumulative Index of Nursing and Allied Health Literature (CINAHL)	3	4.41%
EBSCO	3	4.41%
PsycFirst	3	4.41%
AARP Ageline	2	2.94%
ABI Inform	2	2.94%
Dissertation Abstracts International	2	2.94%
Health and Psychosocial Instruments (HAPI)	2	2.94%
Ingenta	2	2.94%
Lexus Nexus	2	2.94%
ProQuest	2	2.94%
Sociofile	2	2.94%
ThaiLIS	2	2.94%
Web of Science	2	2.94%
Criminal Justice Abstracts	1	1.47%
Familiy and Society Studies	1	1.47%
Google	1	1.47%
PsychArticles	1	1.47%
Scopus	1	1.47%
Social Science Search	1	1.47%
Social Sciences Abstracts	1	1.47%
Social Work Abstracts	1	1.47%
SUMME: 28		

6. Literaturverzeichnis

Die mit einem Asterisk (*) markierten Artikel sind im Review inkludiert.

- Atteslander, P. (Ed.). (2010). *Methoden der empirischen Sozialforschung*. Berlin, Deutschland: Erich Schmidt Verlag.
- *Bachner, Y. G., & O'Rourke, N. (2007). Reliability generalization of responses by care providers to the Zarit Burden Interview. *Aging & Mental Health, 11*, 678-685.
- *Barlow, K. M., & Zangaro, G. A. (2010). Meta-analysis of the reliability and validity of the Anticipated Turnover Scale across studies of registered nurses in the United States. *Journal of Nursing Management, 18*, 862-873.
- Backhaus, K., Erichson, B., Plinke, W., & Weiber, R. (Eds.). (2011). *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung*. Heidelberg, Deutschland: Springer Verlag.
- *Barnes, L. L. B., Harp, D., & Jung, W. S. (2002). Reliability generalization of scores on the Spielberger State-Trait Anxiety Inventory. *Educational and Psychological Measurement, 62*, 603-618.
- Baugh, F. (2002). Correcting effect sizes for score reliability: A reminder that measurement and substantive issues are linked inextricably. *Educational and Psychological Measurement, 62*, 254-263.
- *Beretvas, S. N., Meyers, J. L., & Leite, W. L. (2002). A reliability generalization study of the Marlowe-Crowne Social Desirability scale. *Educational and Psychological Measurement, 62*, 570-589.
- Beretvas, S. N., & Pastor, D. A. (2003). Using mixed-effects models in reliability generalization studies. *Educational and Psychological Measurement, 63*, 75-95.
- *Beretvas, S. N., Suizzo, M.-A., Durham, J. A., & Yarnell, L. M. (2008). A reliability generalization study of scores on Rotter's and Nowicki-Strickland's locus of control scales. *Educational and Psychological Measurement, 68*, 97-119.
- Bonnett, D. G. (2010). Varying coefficient meta-analytic methods for alpha reliabilities. *Psychological Methods, 15*, 368-385.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (Eds.). (2009). *Introduction to meta-analysis*. Chichester, Großbritannien: John Wiley & Sons.

- *Bornmann, L., Mutz, R., & Daniel, H.-D. (2010). A reliability-generalization study of journal peer reviews: A multilevel meta-analysis of inter-rater reliability and its determinants. *PLoS ONE*, *5*, 1-10.
- Bortz, J., & Döring, N. (Eds.). (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Heidelberg, Deutschland: Springer Verlag.
- Botella, J., Suero, M., & Gambara, H. (2010). Psychometric inferences from a meta-analysis of reliability and internal consistency coefficients. *Psychological Methods*, *15*, 386-397.
- Bühner, M. (Ed.). (2004). *Einführung in die Test- und Fragebogenkonstruktion*. München, Deutschland: Pearson Studium.
- Campbell, J. S. (2003). Meta-analysis and reliability generalization of the psychopathy checklist applied to youth. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, *64*(3), 1544.
- *Campbell, J. S., Pulos, S., Hogan, M., & Murry, F. (2005). Reliability generalization of the Psychopathy Checklist applied in youthful samples. *Educational and Psychological Measurement*, *65*, 639-656.
- *Capraro, R. M., & Capraro, M. M. (2002). Myers-Briggs Type Indicator score reliability across studies: A meta-analytic reliability generalization study. *Educational and Psychological Measurement*, *62*, 590-602.
- *Capraro, M. M., Capraro, R. M., & Henson, R. K. (2001). Measurement error of scores on the Mathematics Anxiety Rating Scale across studies. *Educational and Psychological Measurement*, *61*, 373-386.
- *Caruso, J. C. (2000). Reliability Generalization of the NEO personality scales. *Educational and Psychological Measurement*, *60*, 236-254.
- *Caruso, J. C., & Edwards, S. (2001). Reliability generalization of the Junior Eysenck Personality Questionnaire. *Personality and Individual Difference*, *31*, 173-184.
- *Caruso, J. C., Witkiewitz, K., Belcourt-Dittloff, A., & Gottlieb, J. D. (2001). Reliability of scores from the Eysenck Personality Questionnaire: A reliability generalization study. *Educational and Psychological Measurement*, *61*, 675-689.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155-159.
- Comprehensive Meta-Analysis (Version 2) [Computer software]. Englewood, NJ, USA: Biostat.

- Cooper, H. M. (Ed.). (2010). *Research synthesis and meta-analysis: A step-by-step approach*. Thousand Oaks, CA, USA: Sage Publications.
- Courville, T., & Thompson, B. (2001). Use of structure coefficients in published multiple regression articles: β is not enough. *Educational and Psychological Measurement*, *61*, 229-248.
- *De Ayala, R. J., Vonderharr-Carlson, D. J., & Kim, D. (2005). Assessing the reliability of the Beck Anxiety Inventory scores. *Educational and Psychological Measurement*, *65*, 742-756.
- *Deditius-Island, H. K., Caruso, J. C. (2002). An examination of the reliability of scores from Zuckermann's Sensation Seeking Scales, Form V. *Educational and Psychological Measurement*, *62*, 728-734.
- *Dunn, T. W., Smith, T. B., & Montoya, J. A. (2006). Multicultural competency instrumentation: A review and analysis of reliability generalization. *Journal of Counseling and Development*, *84*, 471-482.
- Ellis, P. D. (2009). Effect size calculators [website]. Retrieved from <http://www.polyu.edu.hk/mm/effectsizafaqs/calculator/calculator.html>
- Fan, X., & Thompson, B. (2001). Confidence intervals about score reliability coefficients, please: An *EPM* guidelines editorial. *Educational and Psychological Measurement*, *61*, 517-531.
- Formann, A. K. (Ed.). (2004). *Skriptum zur Vorlesung Testtheorie und Testkonstruktion*. Wien, Österreich: Facultas Verlag.
- Goodwin, L. D. (2001). Interrater agreement and reliability. *Physical Education and Exercise Science*, *5*(1), 13-34.
- *Graham, J. M., & Christiansen, K. (2009). The reliability of romantic love: A reliability generalization meta-analysis. *Personal Relationships*, *16*, 49-66.
- *Graham, J. M., Diebels, K. J., & Barnow, Z. B. (2011). The reliability of relationship satisfaction: A reliability generalization meta-analysis. *Journal of Family Psychology*, *25*, 39-48.
- *Graham, J. M., Liu, Y. J., & Jeziorski, J. L. (2006). The Dyadic Adjustment Scale: A reliability generalization meta-analysis. *Journal of Marriage and Family*, *68*, 701-717.
- *Gronnerod, C. (2003). Temporal stability in the Rorschach method: A meta-analytic review. *Journal of Personality Assessment*, *80*, 272-293.

- *Ha, J. H., Lee, S. M., & Puig, A. (2010). A reliability generalization study of the Frost Multidimensional Perfectionism Scale (F-MPS). *Psychological Reports, 107*, 95-112.
- *Hanson, W. E., Curry, K. T., & Bandalos, D. L. (2002). Reliability generalization of Working Alliance Inventory Scale scores. *Educational and Psychological Measurement, 62*, 659-672.
- *Hellman, C. M., Fuqua, D. R., & Worley, J. A. (2006). A reliability generalization study on the Survey of Perceived Organizational Support: The effects of mean age and number of items on score reliability. *Educational and Psychological Measurement, 66*, 631-642.
- *Hellman, C. M., Muilenburg-Trevino, E. M., & Worley, J. A. (2008). The belief in a just world: An examination of reliability estimates across three measures. *Journal of Personality Assessment, 90*, 399-401.
- Henson, R. K. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement and Evaluation in Counseling and Development, 34*, 177-189.
- *Henson, R. K., & Hwang, D.-Y. (2002). Variability and prediction of measurement error in Kolb's Learning Style Inventory scores: A reliability generalization study. *Educational and Psychological Measurement, 62*, 712-727.
- *Henson, R. K., Kogan, L. R., & Vacha-Haase, T. (2001). A reliability generalization study of the Teacher Efficacy Scale and related instruments. *Educational and Psychological Measurement, 61*, 404-420.
- Henson, R. K., & Thompson, B. (2002). Characterizing measurement error in scores across studies: Some recommendations for conducting „reliability generalization“ studies. *Measurement and Evaluation in Counseling and Development, 35*, 113-126.
- Hogan, T. P., Benjamin, A., & Brezinski, K. L. (2000). Reliability methods: A note on the frequency of use of various types. *Educational and Psychological Measurement, 60*, 523-531.
- *Huynh, Q.-L., Howell, R. T., & Benet-Martinez, V. (2009). Reliability of bidimensional acculturation scores: A meta-analysis. *Journal of Cross-Cultural Psychology, 40*, 256-274.
- *Kieffer, K. M., Cronin, C., & Fister, M. C. (2004). Exploring variability and sources of measurement error in Alcohol Expectancy Questionnaire reliability coefficients: A

- meta-analytic reliability generalization study. *Journal of Studies on Alcohol*, 65, 663-671.
- *Kieffer, K. M., & MacDonald, G. (2011). Exploring factors that affect score reliability and variability in the Ways of Coping Questionnaire reliability coefficients: A meta-analytic reliability generalization study. *Journal of Individual Differences*, 32, 26-38.
- *Kieffer, K. M., & Reese, R. J. (2002). A reliability generalization study of the Geriatric Depression Scale. *Educational and Psychological Measurement*, 62, 969-994.
- Kieffer, K. M., Reese, R. J., & Vacha-Haase, T. (2010). Reliability generalization methods in the context of giftedness research. In B., Thompson, & R. F., Subotnik (Eds.), *Methodologies for conducting research on giftedness* (pp. 89-111). Washington, DC, USA: American Psychological Association
- Kubinger, K. D. (Ed.). (2006). *Psychologische Diagnostik: Theorie und Praxis psychologischen Diagnostizierens*. Göttingen, Deutschland: Hogrefe Verlag.
- *Lane, G. G., White, A. E., & Henson, R. K. (2002). Expanding reliability generalization methods with KR-21 estimates: An RG study of the Coopersmith Self-esteem Inventory. *Educational and Psychological Measurement*, 62, 685-711.
- *Leach, L. F., Henson, R. K., Odom, L. R., & Cagle, L. S. (2006). A reliability generalization study of the Self-Description Questionnaire. *Educational and Psychological Measurement*, 66, 285-304.
- *Li, A., & Bagger, J. (2007). The Balanced Inventory of Desirable Responding (BIDR): A reliability generalization study. *Educational and Psychological Measurement*, 67, 525-544.
- Lienert, G. A., & Raatz, U. (Eds.). (1994). *Testaufbau und Testanalyse*. Weinheim, Deutschland: Beltz Psychologie-Verlags-Union.
- Lipsey, M. W., Wilson, D. B. (Eds.). (2001). *Practical meta-analysis*. Thousand Oaks, CA, USA: Sage Publications.
- *Lopez-Pina, J. A., Sanchez-Meca, J., & Rosa-Alcazar, A. I. (2009). The Hamilton Rating Scale for Depression: A meta-analytic reliability generalization study. *International Journal of Clinical and Health Psychology*, 9, 143-159.
- *Mahapoonyanont, N., Krahamwong, R., Kochakornjarupong, D., & Rachasong, W. (2010). Critical thinking abilities assessment tools: Reliability generalization. *Procedia Social and Behavioral Sciences*, 2, 434-438.

- Meier, S. T., & Davis, S. R. (1990). Trends in reporting psychometric properties of scales used in counseling psychology research. *Journal of Counseling Psychology, 37*, 113-115.
- Meyer, H. (Ed.). (2004). *Theorie und Qualitätsbeurteilung psychometrischer Tests*. Stuttgart, Deutschland: Verlag W. Kohlhammer.
- *Miller, B. K., Byrne, Z. S., Rutherford, M. A., & Hansen, A. M. (2009). Perceptions of Organizational Politics: A demonstration of the reliability generalization technique. *Journal of Managerial Issues, 21*, 280-300.
- *Miller, C. S., Shields, A. L., Campfield, D., Wallace, K. A., & Weiss, R. D. (2007). Substance use scales of the Minnesota Multiphasic Personality Inventory: An exploration of score reliability via meta-analysis. *Educational and Psychological Measurement, 67*, 1052-1065.
- *Miller, C. S., Woodson, J., Howell, R. T., & Shields, A. L. (2009). Assessing the reliability of scores produced by the Substance Abuse Subtle Screening Inventory. *Substance Use & Misuse, 44*, 1090-1100.
- *Mji, A., & Alkhateeb, H. M. (2005). Combining reliability coefficients: Toward reliability generalization of the Conceptions of Mathematics Questionnaire. *Psychological Reports, 96*, 627-634.
- Moosbrugger, H. (2007). Klassische Testtheorie. In H. Moosbrugger, & A. Kelava (Eds.), *Testtheorie und Fragebogenkonstruktion* (pp. 99-112). Heidelberg, Deutschland: Springer Verlag.
- Moosbrugger, H., & Kelava, A. (2007). Qualitätsanforderungen an einen psychologischen Test (Testgütekriterien). In H. Moosbrugger, & A. Kelava (Eds.), *Testtheorie und Fragebogenkonstruktion* (pp. 7-25). Heidelberg, Deutschland: Springer Verlag.
- *Nassen, L. M. (2009). An internal consistency reliability generalization of the Child Behavior Checklist. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 69*(7), 4478.
- *Nilsson, J. E., Schmidt, C. K., & Meek, W. D. (2002). Reliability generalization: An examination of the Career Decision-Making Self-Efficacy Scale. *Educational and Psychological Measurement, 62*, 647-658.

- *O'Rourke, N. (2004). Reliability generalization of responses by care providers to the Center for Epidemiologic Studies-Depression Scale. *Educational and Psychological Measurement, 64*, 973-990.
- *Payne, S. C., Youngcourt, S. S., & Beaubien, J. M. (2007). A meta-analytic examination of the goal orientation nomological net. *Journal of Applied Psychology, 92*, 128-150.
- *Pearson, C. A. (2008). Measuring eating disorder attitudes and behaviors: A reliability generalization study. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 69*(1), 692.
- Petticrew, M., & Roberts, H. (Eds.). (2006). *Systematic reviews in the social sciences: A practical guide*. Malden, MA, USA: Blackwell Publishing.
- *Reese, R. J., Kieffer, K. M., & Briggs, B. K. (2002). A reliability generalization study of select measures of adult attachment style. *Educational and Psychological Measurement, 62*, 619-646.
- Rexrode, K. R. (2005). The Ways of Coping Questionnaire: A reliability generalization study. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 66*(6), 3424.
- *Rexrode, K. R., Petersen, S., & O'Toole, S. (2008). The Ways of Coping Scale: A reliability generalization study. *Educational and Psychological Measurement, 68*, 262-280.
- Romano, J. L., & Kromrey, J. D. (2009). What are the consequences if the assumption of independent observations is violated in reliability generalization meta-analysis studies? *Educational and Psychological Measurement, 69*, 404-428.
- Rosenberg, M. S. (2005). The file-drawer problem revisited: A general weighted method for calculating fail-safe numbers in meta-analysis. *Evolution, 59*, 464-468.
- Rost, J. (Ed.). (2004). *Lehrbuch Testtheorie – Testkonstruktion*. Bern, Schweiz: Verlag Hans Huber.
- *Ross, M. E., Blackburn, M., & Forbes, S. (2005). Reliability generalization of the Patterns of Adaptive Learning Survey goal orientation scales. *Educational and Psychological Measurement, 65*, 451-464.
- Rothrock, N. E., Barocas, J., & Victorson, D. (2004). A preliminary reliability generalization of the FACT-G. *Psycho-Oncology, 13*(S2), S170-S171.

- *Rouse, S. V. (2007). Using reliability generalization methods to explore measurement error: An illustration using the MMPI-2PSY-5 scales. *Journal of Personality Assessment*, 88, 264-275.
- *Ryngala, D. J., Shields, A. L., & Caruso, J. C. (2005). Reliability generalization of the Revised Children's Manifest Anxiety Scale. *Educational and Psychological Measurement*, 65, 259-271.
- Sawilowsky, S. S. (2000a). Psychometrics versus datametrics: Comment on Vacha-Haase's „reliability generalization“ method and some *EPM* editorial policies. *Educational and Psychological Measurement*, 60, 157-173.
- Sawilowsky, S. S. (2000b). Reliability: Rejoinder to Thompson and Vacha-Haase. *Educational and Psychological Measurement*, 60, 196-200.
- Schermelleh-Engel, K., & Werner, C. (2007). Methoden der Reliabilitätsbestimmung. In H. Moosbrugger, & Kelava, A. (Eds.), *Testtheorie und Fragebogenkonstruktion* (pp. 113-133). Heidelberg, Deutschland: Springer Verlag.
- Shields, A. L. (2003). Reliability generalizations of three alcohol screening measures: The Alcohol Use Disorders Identification Test, the CAGE Questionnaire, and the Michigan Alcoholism Screening Test. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, 64(5), 2406.
- *Shields, A. L., Campfield, D. C., Miller, C. S., Howell, R. T., Wallace, K. A., & Weiss, R. D. (2008). Score reliability of adolescent alcohol screening measures: A meta-analytic inquiry. *Journal of Child & Adolescent Substance Abuse*, 17, 75-97.
- *Shields, A. L., & Caruso, J. C. (2003). Reliability generalization of the Alcohol Use Disorders Identification Test. *Educational and Psychological Measurement*, 63, 404-413.
- *Shields, A. L., & Caruso, J. C. (2004). A reliability induction and reliability generalization study of the Cage Questionnaire. *Educational and Psychological Measurement*, 64, 254-270.
- *Shields, A. L., Howell, R. T., Potter, J. S., & Weiss, R. D. (2007). The Michigan Alcoholism Screening Test and its shortened form: A meta-analytic inquiry into score reliability. *Substance Use & Misuse*, 42, 1783-1800.
- Thompson, B. (1994). Guidelines for authors. *Educational and Psychological Measurement*, 54, 837-847.

- Thompson, B. (Ed.). (2002). *Score Reliability: Contemporary thinking on reliability issues*. Thousand Oaks, CA, USA: Sage Publications.
- Thompson, B. (2003a). Understanding reliability and coefficient alpha, really. In B. Thompson (Ed.), *Score reliability: Contemporary thinking on reliability issues* (pp. 3-30). Thousand Oaks, CA, USA: Sage Publications.
- Thompson, B. (2003b). A brief introduction to generalizability theory. In B. Thompson (Ed.), *Score reliability: Contemporary thinking on reliability issues* (pp. 43-58). Thousand Oaks, CA, USA: Sage Publications.
- *Thompson, B., & Cook, C. (2002). Stability of the reliability of LibQUAL+™ scores: A reliability generalization meta-analysis study. *Educational and Psychological Measurement, 62*, 735-743.
- Thompson, B., & Vacha-Haase, T. (2000). Psychometrics is datametrics: The test is not reliable. *Educational and Psychological Measurement, 60*, 174-195.
- *Vacha-Haase, T. (1998). Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measurement, 58*, 6-20.
- Vacha-Haase, T., Henson, R. K., & Caruso, J. C. (2002). Reliability Generalization: Moving toward improved understanding and use of score reliability. *Educational and Psychological Measurement, 62*, 562-569.
- *Vacha-Haase, T., Kogan, L. R., Tani, C. R., & Woodall, R. A. (2001a). Reliability generalization: Exploring variation of reliability coefficients of MMPI clinical scales scores. *Educational and Psychological Measurement, 61*, 45-59.
- Vacha-Haase, T., Kogan, L. R., & Thompson, B. (2000). Sample compositions and variabilities in published studies versus those in test manuals: Validity of score reliability inductions. *Educational and Psychological Measurement, 60*, 509-522.
- Vacha-Haase, T., Ness, C., Nilsson, I., & Reetz, D. (1999). Practices regarding reporting of reliability coefficients: A review of three journals. *Journal of Experimental Education, 67*, 335-341.
- *Vacha-Haase, T., Tani, C. R., Kogan, L. R., Woodall, R. A., & Thompson, B. (2001b). Reliability generalization: Exploring reliability variations on MMPI/MMPI-2 validity scale scores. *Assessment, 8*, 391-401.

- *Vassar, M. (2008a). A note on the score reliability for the Satisfaction With Life Scale: An RG study. *Soc Indic Res*, 86, 47-57.
- *Vassar, M. (2008b). Characterizing score reliability for the South Oaks Gambling Screen. *South African Journal of Psychology*, 38, 541-549.
- *Vassar, M., & Bradley, G. (2010). A reliability generalization study of coefficient alpha for the Life Orientation Test. *Journal of Personality Assessment*, 92, 362-370.
- *Vassar, M., & Crosby, J. W. (2008). A reliability generalization study of coefficient alpha for the UCLA Loneliness Scale. *Journal of Personality Assessment*, 90, 601-607.
- *Vassar, M., Knaup, K. G., Hale, W., & Hale, H. (2011). A meta-analysis of coefficient alpha for the Impact of Event Scales: a reliability generalization study. *South African Journal of Psychology*, 41, 6-16.
- Vassar, M., Ridge, J. W., & Hill, A. D. (2008). Inducing score reliability from previous reports: An examination of life satisfaction studies. *Soc Indic Res*, 87, 27-45.
- *Victorson, D., Barocas, J., Song, J., & Cella, D. (2008). Reliability across studies from the Functional Assessment of Cancer Therapy-General (FACT-G) and its subscales: A reliability generalization. *Qual Life Res*, 17, 1137-1146.
- *Viswesvaran, C., & Ones, D. S. (2000). Measurement error in „Big Five Factors“ personality assessment: Reliability Generalization across studies and measures. *Educational and Psychological Measurement*, 60, 224-235.
- *Wallace, K. A., & Wheeler, A. J. (2002). Reliability generalization of the Life Satisfaction Index. *Educational and Psychological Measurement*, 62, 674-684.
- *Wheeler, D. L., Vassar, M., Worley, J. A., & Barnes, L. L. B. (2011). A reliability generalization meta-analysis of coefficient alpha for the Maslach Burnout Inventory. *Educational and Psychological Measurement*, 71, 231-244.
- *Wilder, L. K., & Sudweeks, R. R. (2003). Reliability of ratings across studies of the BACS. *Education and Treatment of Children*, 26, 382-399.
- Willson, V. L. (1980). Research techniques in *AERJ* articles: 1969 to 1978. *Educational Researcher*, 9(6), 5-10.
- *Yin, P., & Fan, X. (2000). Assessing the reliability of Beck Depression Inventory Scores: Reliability generalization across studies. *Educational and Psychological Measurement*, 60, 201-223.

- *Youngstrom, E. A., & Green, K. W. (2003). Reliability generalization of self-report of emotions when using the Differential Emotions Scale. *Educational and Psychological Measurement, 63*, 279-295.
- Zangaro, G. A., & Soeken, K. L. (2005). Meta-Analysis of the reliability and validity of part B of the Index of Work Satisfaction across studies. *Journal of Nursing Measurement, 13*(1), 7-22.

7. Abstract

Deutsch

Die Methode der Reliability Generalization wurde 1998 von Vacha-Haase vorgestellt und ist als offener Vorschlag ohne festgelegte Durchführungsregeln formuliert. Ziel dieser Methode ist es, zur Bewusstseinsbildung bezüglich der Schwankung von Reliabilitätskoeffizienten beizutragen. Sowohl Wissenschaftlerinnen und Wissenschaftler als auch Praktikerinnen und Praktiker soll hiermit verdeutlicht werden, dass Reliabilität keine fixe, testspezifische Größe darstellt, sondern eine Eigenschaft von Daten ist. Die konkreten Zielsetzungen dieser Methode sind a) die Ermittlung eines für ein gegebenes Messinstrument typischen Reliabilitätskoeffizienten, b) die Bestimmung dessen Variabilität über verschiedene Vorgaben hinweg sowie c) die Identifikation potentieller Quellen dieser Variabilität.

Dieser Review stellt eine Zusammenschau bisher durchgeführter Reliability Generalization-Studien dar und folgt vier Leitfragen. Das Forschungsfeld (A) wird charakterisiert und Parameter des Literatursuchprozesse (B) werden berichtet. Methodische Fragen zur Ermittlung typischer Koeffizienten sowie deren Variabilitäten (C) werden besprochen. Statistische Ansätze zur Identifikation potentieller Einflussgrößen auf die Variabilität der Reliabilität (D1) werden berichtet. Zu acht häufig untersuchten potentiellen Quellen der Variabilität werden meta-analytische Untersuchungsergebnisse (D2) präsentiert.

Viele Autorinnen und Autoren, aber einige wenige Fachzeitschriften konnten als Akteure in diesem Forschungsfeld beobachtet werden. Ungenügendes Berichten von Koeffizienten in Primärstudien wurde als Hauptproblem dieser Methode erkannt. Vielfältige methodische Ansätze und Strategien zur Ermittlung typischer Reliabilitäten und deren Variabilitäten wurden gefunden. Acht meta-analytischen Untersuchungen zur Identifikation bedeutender Einflussgrößen auf die Höhe der Reliabilität wurden durchgeführt. Die Variablen Alter, Altershomogenität, Ethnie, Geschlecht, klinische Stichprobe, Sprache und Stichprobengröße zeigten keine nennenswerten Effekte. Für den meta-analytisch ermittelten Zusammenhang zwischen der Rohwertstreuung und der Höhe von Reliabilitätskoeffizienten konnte eine Effektstärke von $r = .56$ gefunden werden.

English

The Reliability Generalization method was presented by Vacha-Haase in 1998 and was formulated as an open preposition without any proscribed formal and/or methodological rules. This method's aim is to contribute to the consciousness within the scientific community that reliability coefficients show variations in height. To both scientists and practitioners shall be demonstrated that reliability is not an invariant, test-specific value but rather is to be seen as a characteristic of data. The explicitly formulated goals of this method are to a) characterize a typical reliability coefficient for a given measurement instrument, b) show its variability across different administrations and c) examine potential sources of this variability.

This review presents a summary of conducted Reliability Generalization-Studies and follows four main questions. The research field (A) is being characterized and parameters of the process of literature search (B) are being reported. Methodological issues concerning the computation of typical coefficients and their according variability (C) are being discussed. Statistical approaches for identifying potential sources of variability of reliability (D1) are being reported. Results of meta-analytic inquiries (D2) for eight most often analyzed potential sources of variability are being presented.

Many different authors but few different journals were found to be active in this research field. Insufficient reporting of reliability coefficients in primary research articles was identified as being the main problem in using this method. A wide variety of methodological approaches and strategies for characterizing typical reliability coefficients and their variability was found. Eight meta-analyses were conducted in order to identify important factors, which influence the height of reliability coefficients. The variables age, age homogeneity, ethnicity, gender, clinical population, language and sample size showed no noteworthy effects. The meta-analytically determined correlation between raw score heterogeneity and height of reliability coefficients showed an effect size of $r = .56$.

8. Eidesstattliche Erklärung

Hiermit bestätige ich, die vorliegende Diplomarbeit ohne Hilfe Dritter und nur mit den angegebenen Quellen und Hilfsmitteln angefertigt zu haben. Alle Inhalte, welche wörtlich oder sinngemäß übernommen wurden, sind mit der entsprechenden Quelle gekennzeichnet. Diese Arbeit hat in gleicher oder ähnlicher Weise noch keiner anderen Prüfungsbehörde vorgelegen.

Wien, Jänner 2012

Philipp Simbrunner

9. Curriculum Vitae

Name Philipp Simbrunner
Geburtsdatum/-ort 25. 02. 1985 in Linz

Ausbildung

2005- Universität Wien, Studium der Psychologie
2003-2005 Wirtschaftsuniversität Wien, Studium der BWL
1995-2003 Europagymnasium Auhof – Lycée Danube

Berufliche Erfahrungen

10. 2011 **FH Wien - Institut für Personal- und Wissensmanagement**
Hauptberuflich Lehrender
- Konzeption und Durchführung von Lehrveranstaltungen zum Thema wissenschaftliches Arbeiten im HR-Management
 - Mitarbeit an Forschungsprojekten
 - Mitwirkung an Publikationen
09. 2009 – 09. 2011 **FH Wien – Kompetenzzentrum für E-Learning**
Vortragstätigkeit in deutscher und englischer Sprache
- Durchführung und Konzeption von Schulungen für Studierende
01. - 04. 2011 **WU Wien – Institut für Wirtschaftsinformatik und Neue Medien**
Mitarbeiter am Projekt „Creative Business Process Redesign“
- Koordination und Durchführung von Gruppentestungen
 - Datenanalyse
10. 2010 – 01. 2011 **Universität Wien – Fakultät für Psychologie**
Mentor für Studienanfänger/-innen
- Betreuung und Begleitung einer Gruppe von Studierenden
 - Training von Schlüsselkompetenzen
- 05.-07. 2010 **Interface Consult GmbH**
Unternehmensberatung für Usability und Software-Ergonomie
Durchführung folgender Tätigkeiten als Praktikant:
- Usability Tests
 - Eye-Tracking Studien
 - User Interface Design

Sonstige Kenntnisse

EDV MS Office, OpenOffice, SPSS, AMOS, BMD, Navision
Sprachen Deutsch, Englisch, Französisch, Italienisch