

2. Methodik

2.1 Datenerhebung und Stichprobenbeschreibung

2.1.1 Untersuchungsmethode und technische Umsetzung

Als Untersuchungsmethode wurde die strukturierte und standardisierte Fragebogenerhebung mittels Onlineplattform gewählt, um möglichst ökonomisch und flächendeckend in sowohl Deutschland, Österreich und der Schweiz Psychologiestudierende zu akquirieren. Die Vorteile der computerisierten Fragebogenmethode liegen neben der hohen Auswertungsobjektivität, der großen Vergleichbarkeit der erhobenen Informationen und der Flexibilität und Anonymität der Erhebung (welche das Phänomen der sozialen Erwünschtheit reduziert) auch in der Möglichkeit des adaptiven Testens (s. Eid et al., 2013; Thielsch & Weltzin, 2009; Birnbaum, 2004), wovon in der vorliegenden Studie Gebrauch gemacht wurde. Da davon ausgegangen werden kann, dass sämtliche Psychologiestudierende Internetzugang haben (vgl. die aktuelle Studie der Arbeitsgemeinschaft Online Forschung e.V. 2013) und über E-Mailverteiler, Facebook- und andere Foren gut etablierte internetbasierte Kommunikationswege und Ansprechmöglichkeiten bestehen, scheinen die sonst häufig genannten Nachteile der mangelnden Erreichbarkeit von Randgruppen und potentiellen technischen Schwierigkeiten (s. Thielsch & Weltzin, 2009; Faas, 2006; Schoen, 2004) vernachlässigbar. Um letztere weiter abzuschwächen und eine hohe „Usability“ zu gewährleisten, wurden lediglich Fragen- bzw. Antwortformate verwendet, die nicht das Vorhandensein des Adobe Flash Players voraussetzten (Thielsch & Weltzin, 2009). Auch die Gefahr der Mehrfachteilnahme und der unklaren Identität der Befragten wurde zugunsten der Vorteile der Methode in Kauf genommen – nicht zuletzt, da die Mehrfachteilnahme in der Praxis laut Birnbaum (2004) eher selten vorzukommen scheint.

Realisiert wurde die Umfrage mittels der Onlinebefragungssoftware EFS-Survey 9.0 der QuestBack AG. Bei der Programmierung und Gestaltung wurde auf die Empfehlungen von Thielsch und Weltzin (2009) zurückgegriffen. So wurde auf eine angemessene Bildschirmbreite geachtet (Auflösung kleiner als 800 x 600 px), ein Fortschrittsbalken eingeblen-det und die Umfrage optisch ansprechend gestaltet. Zur Erhöhung der Effizienz wurden den Teilnehmerinnen mittels Filterfragen nur die für sie jeweils relevanten Fragen gestellt. Um Reihenfolgeeffekten entgegenzuwirken, wurden sowohl die Fragebogenseiten innerhalb eines Blocks randomisiert dargeboten als auch die einzelnen Fragen einer Seite, wo immer dies sinnvoll schien. Mittels „Zurück“-Buttons wurde den Probandinnen ermöglicht, vorherige Antworten zu ändern. Ebenso war eine vorübergehende Unterbrechung des Fragebogens

und spätere Wiederaufnahme möglich. Zur Verringerung der Dropout-Rate wurden gemäß den Empfehlungen von Thielsch und Weltzin (2009) bzw. Frick, Bächtiger und Reips (1999) die soziodemographischen Daten am Anfang der Umfrage erhoben (High-hurdle-technique). Um einen vollständigen Datensatz zu erwirken, wurde bei einem Großteil der Fragen ein Eingabezwang implementiert. Am Ende der Befragung bestand die Möglichkeit, Bemerkungen zur Studie und der Erhebungsmodalität abzugeben.

Vor dem Umfragestart wurde der Fragebogen an einer kleinen Stichprobe auf technische Funktionalität, inhaltliche Verständlichkeit und Bearbeitungsdauer getestet und – sofern nötig – überarbeitet.

2.1.2 Rekrutierung und Erhebung

Zielgruppe der Untersuchung waren Studierende in Deutschland, Österreich und der Schweiz, die in einem Präsenzstudiengang an einer öffentlichen oder privaten Hochschule das Hauptfach Psychologie studierten. Die Zielgruppe wurde auf Studierende in Präsenzstudiengängen eingegrenzt, weil beispielsweise Frensch (2013) zeigen konnte, dass sich Hochschulen mit Fernstudiengängen stark in den Lehrinhalten und der Zusammensetzung der Studierenden von den Präsenzstudiengängen unterscheiden. Für die an der vorliegenden Studie teilnehmenden Hochschulen galt wegen des klinischen Untersuchungsschwerpunkts zudem die Voraussetzung, dass sie das Fach Klinische Psychologie im Curriculum aufwiesen. Innerhalb der Schweiz wurde die Zielgruppe auf deutschsprachige Hochschulen beschränkt.

Um diese breite Zielgruppe zu erreichen, wurden aktive und passive Rekrutierungsstrategien kombiniert (s. Thielsch & Weltzin, 2009). Vorrangig wurden die lokalen Fachschaften bzw. Fachschaftsräte per E-Mail oder der Internetplattform Facebook kontaktiert mit der Bitte, die Studieneinladung (s. Anhang A) per E-Mail, Facebook- oder anderen Foren an die restlichen Kommilitoninnen weiterzuleiten. An 8 Hochschulen wurden auch private Kontakte zur Weiterverbreitung genutzt, um eine größere Resonanz zu erzielen (Snowball-Technique, Thielsch & Weltzin, 2009). Sofern die Mailverteiler oder Facebook-Seiten ganzer Studien- oder Jahrgänge öffentlich zugänglich waren, wurde die Einladung zur Untersuchung direkt über diese geschickt bzw. gepostet. Wenn keine Fachschaften oder Ähnliches zugänglich waren, wurde die Studieneinladung mit der Bitte um Weiterleitung an zuständige Sekretariate, Dozentinnen oder anderweitige Ansprechpartnerinnen geschickt. Um die Teilnahmemotivation zu erhöhen, wurden fünf Amazon-Gutscheine im Wert von 20€ verlost. Auf Grund der Anonymität der Umfrage sollten hierzu die daran interessierten Probandinnen eine leere E-Mail an die auf der Schlussseite des Fragebogens angegebene Adresse schicken. Zudem bestand hier die Möglichkeit, bei Interesse die Ergebnisse der Studie anzufordern.

Der Erhebungszeitraum begann mit dem Verschicken der E-Mails bzw. dem Posten der Einladung in den Facebook-Foren am 28. Mai 2013 und endete wie zuvor festgesetzt mit Ablauf des 30. Juni 2013. Am 24. Juni wurden Erinnerungsmails herausgeschickt mit dem Hinweis, dass das Internetportal noch eine Woche freigeschaltet sei.

Die durchschnittliche Bearbeitungsdauer betrug 19,7 Minuten, wobei bedingt durch die Filterführung deutliche Unterschiede in der Bearbeitungszeit vorlagen ($Std=16,4$). Die Hälfte der Probandinnen beendete den Fragebogen in weniger als 15,3 Minuten.

2.1.3 Aufbereitung der Daten

Von den insgesamt 1399 Teilnehmerinnen wurden 22 ausgeschlossen, da sie nicht das Fach Psychologie studierten. Von den restlichen 1377 beendeten 63,2% ($n=870$) den Fragebogen, während ihn 36,8% vorher abbrachen. Von diesen Abbrecherinnen klickten 60,5% ($n=307$) lediglich die Startseite an. Da diese Art des Abbruchs laut Thielsch und Weltzin (2009) unproblematisch, da unsystematisch ist, liegt die bereinigte Abbrecherquote bei nur 14,5%. Diese geringe Drop-out-Rate im Vergleich zu ähnlichen Onlinestudien (Schart 2011; Handerer 2011) mag neben der persönlichen Involviertheit der Probandinnen hinsichtlich des Themas auch an der Implementierung der High-hurdle-technique (Thielsch & Weltzin, 2009; Frick et al., 1999) und des Anreizes durch die Verlosung liegen.

Vor der Datenanalyse wurde mit der Software EFS-Survey eine Datenbereinigung durchgeführt. Hierbei wurden die Daten der Probandinnen bereinigt, die mittels „Zurück“-Button eine Filterfrage geändert hatten, so dass zuvor beantwortete Fragen nicht mehr relevant waren. Hiervon waren 24 Datensätze betroffen. Zudem wurde mittels EFS-Survey-Software eine Qualitätsanalyse durchgeführt. Hier sollten diejenigen Probandinnen identifiziert werden, die sich lediglich „durchgeklickt“ haben. Zu diesem Zweck wurde für jede Teilnehmerin ein Qualitätsindex berechnet, indem die individuelle Bearbeitungsdauer pro Seite zum Median der Bearbeitungszeit pro Seite über alle Probandinnen hinweg in Beziehung gesetzt wurde. Probandinnen, welche einen Qualitätsindex von kleiner 0,1 hatten (die pro Seite also nur ein Zehntel der durchschnittlichen Zeit brauchten) wurden ausgeschlossen. Dies traf auf insgesamt 8 Probandinnen zu. Diese wiesen auch für die gesamte Bearbeitungsdauer einen Mittelwert von lediglich 6,5 Minuten auf, so dass ein wenig gewissenhaftes Antworten nicht auszuschließen ist. Zuletzt wurde ein Proband von der Fernuniversität Hagen ausgeschlossen, da wie bereits erwähnt nur Präsenzstudiengänge untersucht werden sollten.

Die endgültige Stichprobe beinhaltet letztlich 861 Probandinnen. Auf Grund der Filterführung bekamen die Teilnehmerinnen je nach vorheriger Antwort verschiedene Fragenpfade vorgelegt, so dass die analysierte Anzahl der Probandinnen von Fragenblock zu Fragenblock variiert.

2.1.4 Stichprobenbeschreibung

Studienland und Hochschulart

Wie in Tabelle 5 abzulesen ist, studierte der mit Abstand größte Anteil der Studienteilnehmerinnen an deutschen Hochschulen (82% der Gesamtstichprobe). Davon waren innerhalb Deutschlands 94,9% an öffentlichen und 5,1% an privaten Hochschulen immatrikuliert. Aus Österreich nahmen 99 Studierende teil (11,5% der Gesamtstichprobe). Von ihnen studierten 77,8% an öffentlichen und 22,2% an privaten Hochschulen. Alle 56 Teilnehmerinnen aus der Schweiz (n=6,5% der Gesamtstichprobe) studierten an öffentlichen Hochschulen. Der Versuch der Kontaktaufnahme mit der nach Wissen der Autorin einzigen schweizerischen Privathochschule, die Psychologie als Präsenzstudiengang anbietet und das Fach Klinische Psychologie im Curriculum enthält (die „Zürcher Hochschule für Angewandte Wissenschaften“) blieb ohne Resonanz. 2 Teilnehmerinnen machten bei dem Item, das die Hochschulzugehörigkeit erfasst, keine Angabe und konnten daher keinem Land zugeordnet werden.

Über die Länder hinweg nahmen demnach 801 Studierende von öffentlichen (93,2%) und 58 von privaten Hochschulen teil (6,8%, s. Tabelle 5). Der vergleichsweise geringe Anteil an Privatstudierenden lässt sich sowohl auf Rekrutierungsschwierigkeiten zurückführen (primär, weil im Gegensatz zu den öffentlichen Hochschulen selten Fachschaftsadressen verfügbar waren und die alternativen Ansprechpartnerinnen die Anfrage seltener weiterleiteten) als auch auf die insgesamt geringere Anzahl von Studierenden an privaten Hochschulen im Vergleich zu den öffentlichen (s. Frensch, 2013).

Tabelle 5: Teilnehmerinnen der Studie unterteilt nach Studienländern und Hochschulart

Land		öffentliche Hochschule	private Hochschulen	Gesamt
Deutschland	Anzahl	668	36	704 (82,0%)
	% innerhalb Deutschlands	94,9%	5,1%	100,0%
Österreich	Anzahl	77	22	99 (11,5%)
	% innerhalb Österreichs	77,8%	22,2%	100,0%
Schweiz	Anzahl	56	0	56 (6,5%)
	% innerhalb der Schweiz	100,0%	0%	100,0%
Gesamt	Anzahl	801	58	859
	% insgesamt	93,2%	6,8%	100,0%

Hochschulen

In Deutschland bestand eine relative Ausgewogenheit der Hochschulen mit Teilnehmerinnen von insgesamt 46 Universitäten. Von den privaten Hochschulen nahmen lediglich 2 teil (die

„International Psychoanalytic University“ in Berlin und die „Fachhochschule für Gesundheit Gera“). Bei der österreichischen Stichprobe ist zu berücksichtigen, dass 89,6% der Studierenden an öffentlichen Hochschulen an der Alpen-Adria-Universität in Klagenfurt immatrikuliert waren ($n=69$) und damit nur 10,4% an anderen öffentlichen Hochschulen. Außerdem studierten 95,4% der Studierenden an privaten Hochschulen an der Sigmund-Freud-Universität (SFU) in Wien ($n=21$) und damit nur eine Person an der UMIT Hall. Dementsprechend beruhen die Werte der öffentlichen Hochschulen in Österreich primär auf den Werten der Universität Klagenfurt, während die der privaten vornehmlich die Meinung der Studierenden an der SFU abbilden. In der Schweiz bestand eine ausgeglichene Aufteilung mit 62,5% Teilnehmerinnen, die in Zürich studierten ($n=35$), 30,4% in Basel und 7,1% in Bern.

Geschlechter- und Altersverteilung

Insgesamt nahmen 708 weibliche Probandinnen teil (82,2%) und 153 männliche. Das durchschnittliche Alter lag bei 24,4 Jahren ($Std=4,6$), wobei die Spanne von 18 bis 59 Jahre reichte. Diese Zahlen entsprechen ziemlich genau denen anderer Studien (z.B. Mutz & Daniel, 2008; Hertwig & Stoltzke, 2001; Handerer 2011; Schart, 2011).

Studiengang und Fachsemester

59,5% der Probandinnen über die Länder hinweg ($n=512$) waren im Bachelorstudiengang immatrikuliert, 27,1% im Master- und 11,1% im Diplomstudiengang. 1,4% gaben an, das Studium mit einem Magister zu beenden und 0,8% sonstiges (hier wurden Lizentiat und Promotion genannt). Eine Person machte keine Angabe. Um sowohl den Bachelor- als auch den Master- und Diplomstudierenden gerecht zu werden, wurde nach dem „Fachsemester im derzeitigen Studiengang“ gefragt. Daher musste bei den Masterstudierenden noch die Fachsemesteranzahl des Bachelors addiert werden. Diese beträgt nach einer bundesweiten Absolventinnenbefragung aus dem Jahr 2011 (Wentura et al., 2013) 6,5 bis 6,7 Semester. Demnach wurden bei den Masterstudierenden 6,6 Semester zur angegebenen Semesterzahl addiert. Zudem muss erwähnt werden, dass die Fälle mit mindestens 17 Semestern in der Form „17+“ kodiert wurden und daher vereinfachend als „17“ in die Analyse eingingen. Davon waren 47 Fälle betroffen (s. Diskussionsteil 4.1.1 für eine kritische Erörterung). Es ergab sich ein Durchschnitt von 7,9 Fachsemestern ($Std=4,1$), der leicht über dem Mittelwert der Studie Handerers (2011) liegt, jedoch niedriger als in der Studie von Schiefele und Jacob-Ebbinghaus (2006) ausfällt. Der Median ist von der beschriebenen Abrundung unberücksichtigt und liegt bei 7 Fachsemestern.

Etwa die Hälfte der Probandinnen (51,6%; n=444) gab an, bereits einen inhaltlichen Schwerpunkt in ihrem Studiengang gewählt zu haben. Die andere Hälfte belegte (noch) keinen. Das Fach Klinische Psychologie bzw. Psychotherapie war mit Abstand am häufigsten vertreten: In 73,9% der Fälle wurde es als Schwerpunkt genannt. Darauf folgten Arbeits-/Betriebs- und Organisationspsychologie (24,8%), Gesundheitspsychologie (18,2%), Pädagogische Psychologie (11,9%) und eine Vielzahl weiterer Schwerpunktfächer, wie bspw. Gruppendynamik, Kognitionspsychologie, Psychoanalyse, Diagnostik bzw. Methoden, Verhaltensanalyse und -regulation oder Lehr-, Lern- & Trainingspsychologie.

Insgesamt 92,9% der Probandinnen bejahten die Frage, ob sie prinzipiell Interesse an Klinischer Psychologie haben (n=800). Den übrigen 7,1% der Teilnehmerinnen, die kein Interesse an der Klinischen Psychologie aufzeigten, wurde der Fragenblock zum Thema *Psychotherapiewissenschaft* nicht vorgelegt (s. u. in Abschnitt 2.3 zu den Messinstrumenten). 666 Teilnehmerinnen bejahten die Frage, ob sie derzeit in ihrem Studium das Fach Klinische Psychologie belegen oder es bereits früher an ihrer gegenwärtigen Hochschule belegt hatten (77,4%), die restlichen 22,6% verneinten diese Frage. Letzteren wurden die Fragen zum Fach Klinische Psychologie nicht vorgelegt (s. u. in Abschnitt 2.3 zu den Messinstrumenten).

2.2 Statistische Verfahren zur Datenanalyse

Im Folgenden werden die wichtigsten verwendeten statistischen Verfahren dargelegt. Aus Platzgründen wird hierbei lediglich bei den zentralen statistischen Verfahren (den Faktoren-, Varianz- und Regressionsanalysen sowie t-Tests) auf die Voraussetzungen und deren Überprüfungsmöglichkeiten eingegangen. Bei den anderen Verfahren sei auf Eid et al. (2013) verwiesen. Die Durchführung der statistischen Analysen erfolgte mit der Software SPSS Version 21.0. Das Signifikanzniveau wurde auf $\alpha = 5\%$ festgelegt.

2.2.1 Vorbemerkung zur Angemessenheit der statistischen Verfahren

Eine in den Sozialwissenschaften weit verbreitete Praxis ist die Anwendung von statistischen Verfahren, die mindestens Intervallskalenniveau voraussetzen (bspw. Varianz-, Faktoren- oder lineare Regressionsanalysen), obwohl strenggenommen lediglich Ordinalskalenniveau gegeben ist (Kempf, 2006; Eid et al., 2013; Westermann, 1985; Surber, 1984; Backhaus, Erichson, Plinke, & Weiber, 2000). Diese Praxis wird an anderer Stelle intensiv und kontrovers diskutiert (bspw. Westermann, 1985; Orth & Wegener, 1983; Jöreskog & Moustaki, 2001; Manisera, Dusseldorp, & van der Kooij, 2010). Während einige Autorinnen die Verwendung von metrischen Verfahren bei lediglich ordinalskalierten Variablen rigoros ablehnen und alternative Wege vorschlagen (bspw. Jöreskog & Moustaki, 2001; Linting, Meulman,

Groenen, & van der Kooij, 2007), geben andere gewisse Mindestvoraussetzungen an, die erfüllt sein müssen, um dennoch einigermaßen zuverlässige Ergebnissen zu erhalten (bspw. Urban & Mayerl, 2008). Andere prüfen empirisch, inwieweit die Anforderungen an intervallskalierte Daten auch von Daten erfüllt werden können, die gemeinhin als ordinalskaliert gelten (bspw. Westermann, 1985).

In der vorliegenden Untersuchung liegen die Antworten in der Regel in Form einer 6- bzw. 7-stufigen Likert-Skala vor. Diese Form erfüllt streng genommen lediglich die Voraussetzungen für ein ordinales Skalenniveau (s.o.). Theoretisch wäre daher wünschenswert gewesen, die Analysen mittels der „klassischen“ metrisch basierten Verfahren durchzuführen (nicht zuletzt um eine Vergleichsmöglichkeit mit anderen Studien herzustellen) und im Anschluss mit den äquivalenten Verfahren auf Ordinalskalenebene zu vergleichen. Leider hätte dies bei Anwendung auf alle Analysen den zeitlichen Rahmen und Umfang der Arbeit gesprengt. Daher werden im Folgenden überwiegend Analysemethoden verwendet, welche mindestens Intervallskalenniveau der Variablen voraussetzen. Um die Robustheit dieser Vorgehensweise zu überprüfen, werden in einigen ausgewählten Fällen zusätzliche Analysen durchgeführt, die dem ordinalen Skalenniveau besser Rechnung tragen. Die Ergebnisse dieser ergänzenden Analysen – insbesondere im Bereich der testtheoretischen Überprüfung des Fragebogens – werden im Folgenden lediglich knapp dargestellt, während die entsprechende Theorie aus Platzgründen im Anhang einsehbar ist.

2.2.2 Testtheoretische Überprüfung des Fragebogens

Zur Überprüfung des Fragebogens wurden explorative Faktorenanalysen durchgeführt, um die Eindimensionalität der Einzelskalen und die faktorielle Validität des Fragebogens zu überprüfen (s. Marcus & Bühner, 2009). Zusätzlich wurden als ordinales Äquivalent kategoriale Hauptkomponentenanalysen (PCA) mittels der in SPSS implementierten Methode „CATPCA“ durchgeführt. Die Ergebnisse dieser Variante sind jeweils kursiv in Klammern dargestellt, die Theorie dieses Verfahrens wird in Anhang C erläutert. Zur Itemselektion der so entstandenen Skalen wurden die korrigierten Trennschärfen der einzelnen Items berechnet sowie Cronbachs α , wenn das jeweilige Item entfernt wurde (Krohne & Hock, 2007). Diese Kennwerte bildeten die Grundlage für die Itemselektion der endgültig in die Analyse eingehenden Skalen, für die zuletzt jeweils die Gesamt-Reliabilität und die mittleren Inter-Item-Korrelationen berechnet wurden (Marcus & Bühner, 2009).

Faktorenanalysen

Die Faktoren der „klassischen“ metrischen Faktorenanalyse wurden mittels Maximum-Likelihood Schätzverfahren (Fisher, 1912; im Folgenden ML-Schätzverfahren) extrahiert (s. Eid et al., 2013 für die Vorteile dieses Verfahrens im Gegensatz zur Hauptkomponentenanalyse). Um die Voraussetzung der multivariaten Normalverteilung für das ML-Schätz-

verfahren zu überprüfen, wurde zunächst mittels Kolmogorov-Smirnov-Test (mittels Lilliefors-Korrektur der kritischen Werte) und Shapiro-Wilk-Test überprüft, ob eine univariate Normalverteilung vorliegt. Diese stellt eine notwendige, jedoch keineswegs hinreichende Bedingung für die multivariate Normalverteilung dar (DeCarlo, 1997). Da diese Testwerte bei großen Stichproben wie der vorliegenden selbst bei kleinen Abweichungen von der Normalverteilung signifikant werden (s. Eid et al., 2013), wurden zusätzlich die Histogramme, die Schiefe und Kurtosis untersucht. Als Grenzwerte für die Schiefe nennen West, Finch und Curran (1995) einen absoluten Schätzwert von < 2 , für die Kurtosis entsprechend < 7 . Allerdings ist die Parameterschätzung mithilfe der ML-Verfahren relativ robust gegenüber Verletzungen der multivariaten Normalverteilungsannahme bezüglich der Parameterschätzung, solange die univariaten Verteilungen nicht zu extrem sind und die Stichprobe nicht zu klein ist (Eid et al., 2013; Marcus & Bühner, 2009).

Vor Durchführung der Faktorenanalysen wurde mittels Bartlett Test auf Sphärizität überprüft, ob die Unabhängigkeit der Variablen vorliegt, was eine Faktorenanalyse obsolet machen würde (Eid et al., 2013). Allerdings hat sich dieser Test aus verschiedenen Gründen alternativen Prüfgrößen als unterlegen erwiesen (Marcus & Bühner, 2009). Daher wurde zusätzlich auf den *Kaiser-Meyer-Olkin-Koeffizienten* (im Folgenden KMO-Maß) zurückgegriffen, der als zusammenfassendes Maß für die Eignung der Stichprobe eine inhaltlich ähnliche Funktion erfüllt. Berechnet wird er anhand der Anti-Image-Korrelationsmatrix, welche die partiellen Korrelationen zwischen den Variablen als Basis hat. Zur Beurteilung wurde auf die Empfehlungen von Kaiser (1974) zurückgegriffen. Zudem wurden die Measure-of-Sampling-Adequacy-Werte betrachtet (MSA-Werte), mit deren Hilfe einzelne Variablen aus dem faktorenanalytischen Modell beurteilt und gegebenenfalls ausgeschlossen werden können (Brosius, 1998).

Zur Überprüfung der Modellgüte wurde der χ^2 -Test auf Anpassung verwendet. Da hier allerdings ebenfalls bei großen Stichproben wie der vorliegenden schon bei kleinsten Abweichungen die Nullhypothese der Passung verworfen wird (Eid et al., 2013), wurde eine Reihe weiterer Kriterien zur Auswahl der Anzahl der Faktoren und Items sowie zur Beurteilung der Güte des Faktorenmodells hinzugezogen. Zur Auswahl der Anzahl der Faktoren wurden sowohl das Kaiser-Guttman-Kriterium herangezogen als auch der Scree-Test (s. Eid et al., 2013). Des Weiteren wurden zur Überprüfung der Modellgüte die anhand der Faktorengleichungen reproduzierten Korrelationen zwischen den Variablen mit der beobachteten Korrelationsmatrix abgeglichen. Laut Krohne und Hock (2007) gilt für eine gute Reproduktion die Faustregel, dass alle korrelativen Diskrepanzen kleiner als 0,1 sein sollten. Außerdem wurde die Ladungshöhe der Faktoren selbst in Betracht gezogen, wobei nach Krohne und Hock (2007) Ladungen, deren Betrag kleiner als 0,30 ist, als niedrig bzw. vernachlässigbar gelten. Ab 0,40 kann man von einer moderaten und ab 0,60 von einer hohen Ladung sprechen. Auch die Kommunalitäten sollten laut Marcus und Bühner (2009) einen Wert von 0,20 nicht unterschreiten.

Nach der Bestimmung der Anzahl der Faktoren anhand der Anfangslösung wurde zur besseren Interpretierbarkeit eine Rotation durchgeführt. Hierbei wurde gemäß der Empfehlung von Eid et al. (2013) eine oblique Rotation durchgeführt, die Korrelationen zwischen

den Faktoren zulässt (was in der vorliegenden Studie bei den angenommenen Konstrukten hinter den Faktoren sinnvoll scheint). Eine Korrelation auf Faktorebene ist hierbei lediglich möglich, wird jedoch keineswegs aufgezwungen. Es wurde die direkte Quartimin-Rotation verwendet, welche ein Spezialfall der direkten Oblimin-Rotation darstellt, wenn δ auf 0 gesetzt wird (es wird demnach die schiefstmögliche Lösung erlaubt). Zur Interpretation wurde sowohl die Mustermatrix herangezogen, welche die Ladungen im rotierten Faktormodell enthält und in Abschnitt 2.3 zu den Messinstrumenten dargestellt wird, als auch die Strukturmatrix, welche die Korrelationen der Faktoren mit den manifesten Variablen darstellt.

Item- und Reliabilitätsanalysen

Um die Zuverlässigkeit der durch die Faktorenanalysen entstandenen Skalen zu beurteilen, wurde für jede Skala die interne Konsistenz Cronbachs α berechnet. Dieses Reliabilitätsmaß setzt konzeptionell voraus, dass die Testteile als Messung des gleichen Konstrukts angesehen werden können, also Homogenität bzw. Eindimensionalität der einzelnen Skalen besteht (Marcus & Bühner, 2009). Die Eindimensionalität kann am besten im Rahmen der probabilistischen Testtheorie oder mittels konfirmatorischer Faktorenanalyse untersucht werden (Marcus & Bühner, 2009), was beides in der vorliegenden Studie aus Gründen der fehlenden Software nicht möglich war. Alternativ wurde die Ladungsstruktur der explorativen Faktorenanalyse interpretiert (s.o.) und als notwendiges, wenn auch nicht hinreichendes Maß die mittlere Inter-Item-Korrelation (MIC) herangezogen (Marcus & Bühner, 2009). Dieses Maß sollte nach Bühner (2011) nicht unter 0,20 liegen sollten.

Während bei zumindest essentiell τ -äquivalenten Testitems Cronbachs α die Reliabilität präzise abbildet (s. Eid et al., 2013), stellt es bei kongenerischer Messung – bei der unterschiedliche Faktorladungen bestehen können – die Untergrenze der Reliabilität dar (s. Marcus & Bühner, 2009). Im Folgenden sollte daher von Cronbachs α als Untergrenze der Reliabilität ausgegangen werden, da eine lediglich kongenerische Messung nicht ausgeschlossen werden kann.

Klassisch bestehen gewisse Faustregeln, nach denen die Höhe der Reliabilitätskoeffizienten beurteilt werden kann.¹⁰ Allerdings lässt sich die Beurteilung laut Marcus und Bühner (2009) nicht durch derartige pauschale Faustregeln festlegen, da das erforderliche Niveau stark vom Anwendungsfall und -zweck abhängt. Auch die Testlänge muss meist im Abgleich mit der Praktikabilität in Betracht gezogen werden. In Abhängigkeit von der Testlänge empfiehlt bspw. Bagozzi (1980) unterschiedliche Untergrenzen, wonach ein Instrument mit nur zwei Indikatoren bereits bei einem $\alpha \geq 0,50$ als zuverlässig gelten kann. Ein niederländisches Ratingsystem für Testqualität, das von dem „Committee of Test Affairs of the Dutch Association of Psychologists“ veröffentlicht wurde (COTAN-Beurteilungssystem), spezifiziert drei Niveaus mit unterschiedlich strengen Anforderungen in Abhängigkeit vom

¹⁰ Nach George & Mallery (2003) kann bspw. ein α von $> 0,9$ als exzellent gewertet werden, $0,9 > \alpha > 0,8$ als gut, $0,8 < \alpha > 0,7$ als akzeptabel und $0,7 > \alpha > 0,6$ als fragwürdig. Werte darunter gelten als nicht akzeptabel.

Anwendungszweck: Für Niveau 1, das sich auf wichtige Einzelfallentscheidungen bezieht, wird eine Reliabilität von mindestens 0,80 gefordert. Für Niveau 2, welches weniger wichtige Einzelfalldiagnosen impliziert, mindestens 0,70 und für Niveau 3 bei Gruppenuntersuchungen eine Reliabilität von mindestens 0,60 (Evers, 2001). Da es sich bei der vorliegenden Studie um eine Gruppenuntersuchung mit zumindest keinen folgeschweren Konsequenzen handelt, wurde von einem Mindestwert von $\alpha \geq 0,60$ ausgegangen. Außerdem wäre eine Verlängerung des Fragebogens zwar aus Zuverlässigkeitszwecken wünschenswert gewesen, auf Grund der erhöhten Länge jedoch auf Kosten der Anzahl der Teilnehmerinnen gegangen.

Zur Analyse der Eignung der einzelnen Items wurde zunächst jeweils Cronbachs α auf Skalenebene ohne Berücksichtigung des entsprechenden Items berechnet. Sollte α bei Weglassen des Items steigen, ist es für die Skala ungeeignet (Krohne & Hock, 2007). Zudem wurde zur Beurteilung der Itemgüte die korrigierte Trennschärfe hinzugezogen. Nach Krohne & Hock (2007) gilt 0,30 als untere Grenze der Trennschärfe, während Werte um 0,50 und 0,60 Items kennzeichnen, die gut zwischen Personen mit hoher und niedriger Ausprägung trennen können.

2.2.3 Mittelwert-, Verteilungs- und Häufigkeitsvergleiche

Varianzanalysen

Zur Untersuchung der Mittelwertsunterschiede zwischen den Ländern (Deutschland, Österreich und Schweiz) und den Hochschularten (private und öffentliche Hochschulen) wurden bei den metrischen abhängigen Variablen zweifaktorielle Varianzanalysen (im Folgenden „Anova“ für „Analysis of Variance“) für unabhängige Stichproben mit festen Effekten durchgeführt (3x2-Design). Im Anschluss an diese Omnibustests wurden Post-hoc-Tests durchgeführt, wobei mittels Bonferroni-Korrektur die spezifischen Irrtumswahrscheinlichkeiten adjustiert wurden, um der Kumulierung des α -Fehlers bei den Paarvergleichen entgegenzuwirken (Eid et al., 2013; Janssen & Laatz, 2013). Die Annahme der Normalverteilung der Merkmalsvariablen – einer Voraussetzung des F-Tests der Anova – wurde mittels Kolmogorov-Smirnov-Test überprüft. Laut Eid et al. (2013) ist der F-Test bei hinreichend großen Stichproben von > 30 allerdings relativ robust gegen Verletzungen der Normalverteilungsannahme. Die Bedingung der Homoskedastizität wurde mittels Levene-Test überprüft. Da dieser jedoch sensibel auf die Stichprobengröße reagiert, wurde zusätzlich auf eine Daumenregel von Ziegler und Bühner (2009) zurückgegriffen, nach der ein Verhältnis der Varianzen von $\geq 10:1$ als kritisch gilt. Um die ungleichen Zellhäufigkeiten und damit Nonorthogonalität der Variablen zu berücksichtigen sowie die Tatsache, dass aus der Schweiz keine Studierenden von privaten Hochschulen teilnahmen, wurde Typ IV zur Berechnung der Quadratsummen gewählt (s. Janssen & Laatz, 2013, S.362 für detailliertere Informationen). Zudem wurde ein gesättigtes Modell gewählt, bei dem alle ausgewählten Faktoren und deren Wechselwirkungen in das Modell eingehen.

Die akademische Psychologie unter der Lupe
Dissonanz zwischen studentischen Vorstellungen und
wahrgenommener Studienrealität

Plischke, A.

2016, XI, 190 S. 16 Abb., Softcover

ISBN: 978-3-658-11177-9