



universität
wien

Diplomarbeit

Titel der Arbeit

Individualismus und Kollektivismus

Äquivalenzprüfung zweier Messinstrumente für Kinder
und Jugendliche mittels Rasch-Modell und
konfirmatorischer Mehrgruppen-Faktorenanalyse

Verfasser

Takuya Yanagida

Angestrebter akademischer Grad

Magister der Naturwissenschaften (Mag.rer.nat)

Wien, im Mai 2009

Studienkennzahl: 298

Studienrichtung: Psychologie

Betreuer: Dr. Dagmar Strohmeier

Danksagung

An dieser Stelle sei all jenen Personen gedankt, die einen Beitrag zur vorliegenden Arbeit geleistet haben bzw. mir während meiner Diplomarbeitsphase unterstützend zur Seite gestanden sind. Erst die Hilfe und Unterstützung dieser Personen insgesamt hat mir die Fertigstellung meiner Arbeit ermöglicht.

Allen voran gebührt der größte Dank meiner Betreuerin Univ.-Ass. Dr. Dagmar Strohmeier, die mich geduldig durch den gesamten Arbeitsprozess geführt hat und sich dabei stets Zeit für meine Anliegen genommen hat. Insbesondere danke ich ihr für das Vertrauen in meine Fähigkeiten, welches sie mir während des gesamten Arbeitsprozesses entgegen gebracht hat und mir das notwendige Selbstbewusstsein vermittelt hat.

Für die Organisation der Datenerhebung in Japan danke ich Ao.Univ.-Prof. Yuichi Toda, ebenso Mag. Anja Lampert und Mag. Eva-Maria Schiller für ihre Mitarbeit bei der Datenerhebung und -eigabe in Österreich.

Dr. Mag. Michael Weber und Univ.-Prof. Dr. Mag. Klaus D. Kubinger danke ich für die wegweisenden Anregungen und Hilfestellungen zur statistischen Auswertung. Darüber hinaus sei den beiden genannten Personen für das beharrliche Erinnern an die rasche Fertigstellung der Arbeit gedankt. In diesem Zusammenhang danke ich ebenfalls Mag. Lale Khorramdel für die motivierenden Worte, die sie durchweg für mich übrig hatte.

Weiters danke ich Mag. Clara Gomes und Mag. Carrie Kovacs für die Hilfestellung bei der englischen Übersetzung des Abstracts.

Abschließend gilt mein besonderer Dank meinen lieben Kolleginnen Mag. Michaela S. Fasching und Mag. Eva Schleicher, die stets ein offenes Ohr für meine Sorgen und Ängste hatten und mich unentwegt ermuntern und mich in meinem Tun bestärken konnten.

Abstract

Ziel der vorliegenden Untersuchung war (1) die psychometrische Äquivalenzprüfung mittels Rasch-Modell und konfirmatorischer Mehrgruppen-Faktorenanalyse und (2) die testtheoretische Analyse zweier Messinstrumente für Kinder- und Jugendliche.

277 Schüler¹in Österreich (46% Mädchen und 54% Buben) im Alter von 8 bis 15 Jahren und 532 Schüler in Japan (47% Mädchen und 53% Buben) im Alter von 9 bis 15 Jahren bearbeiteten dafür eine modifizierte Fassung der *Collectivism Scale* (Yamaguchi, Kuhlman & Sugimori, 1995) zur Erfassung von Kollektivismus und ein neu entwickeltes Erhebungsinstrument (*Individualism Difference Scale*) zur Erfassung von Individualismus. Zur Nachweis der Äquivalenz wurden beide Erhebungsinstrumente im ersten Schritt auf faktorielle Validität geprüft, um in weiterer Folge hinsichtlich (a) konfiguraler, (b) metrischer und (c) skalarer Invarianz zu prüfen.

(1) Die Äquivalenzprüfung zeigte, dass dem Rasch-Modell Analysen zufolge die faktorielle Validität der *Collectivism Scale* in beiden Kulturen nicht gegeben ist, während der *Individualism Difference Scale* Skalenäquivalenz attestiert werden konnte. Die konfirmatorische Mehrgruppen-Faktorenanalyse hingegen konnte die Messeinheitenäquivalenz der *Collectivism Scale* belegen, während die *Individualism Difference Scale* die Ebene der partiellen Messeinheitenäquivalenz erreichte. (2) Die testtheoretischen Analysen erbrachten bei ausreichend hoher Reliabilität keinen Nachweis für die Validität der *Collectivism Scale*. Die Analysen zur *Individualism Difference Scale* demgegenüber konnten eine ausreichend hohe Validität nachweisen. Allerdings weist diese Skala in der österreichischen Stichprobe eine signifikant geringere Reliabilität auf als in der japanischen Stichprobe.

Zusammenfassend konnte aufgrund der vorliegenden Untersuchung belegt werden, dass die *Individualism Difference Scale* einen vielversprechenden Ansatz zur Erfassung von Individualismus bei Kinder- und Jugendlichen darstellt, wohingegen sich die *Collectivism Scale* in der modifizierten Fassung nicht dafür zu eignen scheint.

¹ der einfachen Schreibweise wegen wird im folgenden bei allen Personenbeschreibungen die männliche Form verwendet

The aim of this empirical work was (1) to test statistical equivalence of two measurement tools for (early) adolescents using the *Rasch-Model* and a multiple group confirmatory factor analysis, and (2) to evaluate the psychometric properties of these instruments.

277 Austrian (46% girls and 54% boys) aged 8 to 15 years and 532 Japanese students (47% girls and 53% boys) aged 9 to 15 years completed an adapted version of the *Collectivism Scale* (Yamaguchi, Kuhlman & Sugimori, 1995), as well as a newly developed questionnaire to assess individualism (*Individualism Difference Scale*). In order to show equivalence, both instruments were first tested for factorial validity and subsequently for (a) configural, (b) metric and (c) scalar invariance.

(1) In regard to the psychometric equivalence analysis using the Rasch Model, no factorial validity across cultures could be found for the *Collectivism Scale*, in contrast to the *Individualism Difference Scale*, which did show scalar equivalence in both groups. In contrast, the confirmatory factor analysis showed measurement equivalence for Collectivism in both groups, while the *Individualism Difference Scale* was only partially measurement equivalent. (2) The test theoretical analysis showed acceptable reliability measures but no proof of validity for the *Collectivism Scale*. Validity of the *Individualism Difference Scale* was acceptable, however, the Austrian sample showed a significantly lower reliability in comparison to the Japanese sample.

In summary, the present results offer support for the Individualism Scale as a promising measurement instrument for assessing Individualism in (early) adolescence, whereas the adapted version of the *Collectivism Scale* could not be validated.

Inhaltsverzeichnis

Einleitung	1
THEORETISCHER TEIL	3
1 Kultur und Psychologie	5
1.1 Der Kulturbegriff in den Sozialwissenschaften	5
1.2 Kultur-, kulturvergleichende- und soziokulturelle Psychologie	6
1.3 Forschungsansätze der soziokulturellen Psychologie	8
1.4. Culture's Consequences	11
2 Individualismus – Kollektivismus (Ind-Col)	13
2.1 Begriffsbestimmung	13
2.2 Dimensionalität	14
2.3 Ansätze zur Erforschung von Ind-Col	15
2.4. Erhebungsmethoden und -instrumente	17
2.4.1 <i>Operationalisierung und Erhebungsmethoden</i>	17
2.4.2 <i>Inhaltliche Bereiche von Individualismus- und Kollektivismus-Skalen</i>	19
2.4.3 <i>Die drei gebräuchlichsten Erhebungsinstrumente für Ind-Col</i>	21
2.4 Kritische Betrachtung	23
3 Methodische Aspekte der kulturvergleichenden Forschung	25
3.1 Taxonomie kulturvergleichender Studien	26
3.2 Bias und Äquivalenz	27
3.2.1 <i>Arten von Bias</i>	27
3.2.2 <i>Umgang mit Bias</i>	28
3.2.3 <i>Ebenen der Äquivalenz</i>	29
3.3 Methoden zur psychometrischen Äquivalenzprüfung	30
3.3.1 <i>Das Rasch-Modell</i>	30
3.3.2 <i>Konfirmatorische (Mehrgruppen-)Faktorenanalyse</i>	34
EMPIRISCHER TEIL	37
4 Ziele der Untersuchung und Forschungsfragen	39
5 Methode	40
5.1 Untersuchungsdesign und Stichprobenbeschreibung	40

5.2.1 <i>Collectivism Scale</i>	42
5.2.2 <i>Individualism Difference Scale</i>	43
5.3 Auswertungsmethoden	44
5.3.1 <i>Äquivalenzprüfung nach dem Rasch-Modell</i>	45
5.3.2 <i>Äquivalenzprüfung mittels konfirmatorischer Mehrgruppen-Faktorenanalyse</i>	47
6 Ergebnisse	49
6.1 Rasch-Modell Analysen	49
6.1.1 <i>Collectivism Scale</i>	50
6.1.1.1 Österreichische Stichprobe	50
6.1.1.2 Japanische Stichprobe	51
6.1.1.3 Gesamtstichprobe	52
6.1.2 <i>Individualism Difference Scale</i>	52
6.1.2.1 Österreichische Stichprobe	52
6.1.2.2 Japanische Stichprobe	53
6.1.2.3 Gesamtstichprobe	54
6.2 Konfirmatorische (Mehrgruppen-)Faktorenanalyse	55
6.2.1 <i>Collectivism Scale</i>	55
6.2.1.1 Überprüfung des Messmodells	55
6.2.1.2 Invarianzprüfung in der Kalibrierungsstichprobe	57
6.2.1.3 Invarianzprüfung in der Prüfstichprobe	59
6.2.2 <i>Individualism Difference Scale</i>	60
6.2.2.1 Überprüfung des Messmodells	60
6.2.2.2 Invarianzprüfung in der Kalibrierungsstichprobe	62
6.2.2.3 Invarianzprüfung in der Prüfstichprobe	64
6.3 Testtheoretische Güte	65
6.3.1 <i>Collectivism Scale</i>	66
6.3.1.1. Reliabilität	67
6.3.1.2. Konkurrente Validität	67
6.3.2 <i>Individualism Difference Scale</i>	68
6.3.2.1. Reliabilität	68
6.3.2.1.1 Dichotome Individualism Difference Scale	68
6.3.2.1.2 Mehrkategoriale Individualism Difference Scale	69
6.3.2.2. Konkurrente Validität	69
6.3.2.2.1 Dichotome Individualism Difference Scale	69

6.3.2.2.2 Mehrkategorielle Individualism Difference Scale	70
Diskussion	71
Zusammenfassung	79
Literaturverzeichnis	81
Anhang	93
Anhang	93
Ad.) Kapitel 6.1.1.1 Österreichische Stichprobe	93
Ad.) Kapitel 6.1.1.2 Japanische Stichprobe	95
Ad.) Kapitel 6.1.2 Individualism Difference Scale	97
Ad.) Kapitel 6.1.2.1 Österreichische Stichprobe	99
Ad.) Kapitel 6.1.2.2 Japanische Stichprobe	99
Ad.) Kapitel 6.1.2.3 Gesamtstichprobe	100
Ad.) Kapitel 6.3.1.2 Konkurrente Validität	101
Ad.) Kapitel 6.3.2.2.1 Dichotome Individualism Difference Scale	101
Ad.) Kapitel 6.3.2.2.1 Mehrkategorielle Individualism Difference Scale	101

Tabellen-, Abbildungs- und Formelverzeichnis

Tabelle 1. Forschungsansätze der soziokulturellen Psychologie (Markus und Hamedami, 2007, S. 14)	8
Tabelle 2. Beispielvignette aus dem <i>Scenario Questionnaire of Cultural Orientations</i>	19
Tabelle 3. Inhaltliche Bereiche für Individualismus und Kollektivismus (Oyserman et al., 2002, S. 9)	20
Tabelle 4. Beispielitems aus der <i>Horizontal-Vertical Collectivism-Individualism scale</i>	22
Tabelle 5. Beispielitems aus der <i>Individualism-Collectivism (INDCOL) Scale</i>	22
Tabelle 6. Taxonomie kulturvergleichender Studien (Van de Vijver, 2007, S. 342) ..	26
Tabelle 7. Stichprobenbeschreibung	41
Tabelle 8. Itemliste der <i>Collectivism Scale</i> in deutscher Übersetzung.....	42
Tabelle 9. Itemliste der <i>Individualism Difference Scale</i> in deutscher Übersetzung..	44
Tabelle 10. Cut-off-Werte der verwendeten Fit-Indices und zur Beurteilung der Differenz in Abhängigkeit der Invarianzform.....	48
Tabelle 11. Likelihood-Ratio-Tests der <i>Collectivism Scale</i> in der österreichischen Stichprobe ($n_{\text{ö}} = 277$); Modelltest für die Teilungskriterien <i>Score</i> , <i>Geschlecht</i> und <i>Alter</i>	50
Tabelle 12. Likelihood-Ratio-Tests der <i>Collectivism Scale</i> in der japanischen Stichprobe ($n_j = 531$); Modelltest für die Teilungskriterien <i>Score</i> , <i>Geschlecht</i> und <i>Alter</i>	51
Tabelle 13. Likelihood-Ratio-Tests der <i>Individualism Difference Scale</i> in der österreichischen Stichprobe ($n_{\text{ö}} = 277$); Modelltest für die Teilungskriterien <i>Score</i> , <i>Geschlecht</i> und <i>Alter</i>	52
Tabelle 14. Likelihood-Ratio-Tests der <i>Individualism Difference Scale</i> in der japanischen Stichprobe ($n_j = 531$); Modelltest für die Teilungskriterien <i>Score</i> , <i>Geschlecht</i> und <i>Alter</i>	53
Tabelle 15. Likelihood-Ratio-Tests der <i>Individualism Difference Scale</i> für den reduzierten Itempool in der Kalibrierungstichprobe ($n_k = 567$); Modelltest für das Teilungskriterium <i>Kultur</i>	54

Tabelle 16. Likelihood-Ratio-Tests der <i>Individualism Difference Scale</i> für den reduzierten Itempool in der Prüfstichprobe ($n_p = 242$); Modelltest für das Teilungskriterium <i>Kultur</i>	54
Tabelle 17. Messmodells der <i>Collectivism Scale</i> in der österreichischen Stichprobe	56
Tabelle 18. Messmodell der <i>Collectivism Scale</i> in der japanischen Stichprobe	56
Tabelle 19. Geschätzte standardisierte Faktorladungen, Standardfehler und Kommunalitäten der reduzierten <i>Collectivism Scale</i> in der österreichischen und japanischen Stichprobe	57
Tabelle 20. Invarianzprüfung der <i>Collectivism Scale</i> in der Kalibrierungsstichprobe	59
Tabelle 21. Invarianzprüfung der <i>Collectivism Scale</i> in der Prüfstichprobe	60
Tabelle 22. Messmodells der <i>Individualism Difference Scale</i> in der österreichischen Stichprobe	61
Tabelle 23. Messmodell der <i>Individualism Difference Scale</i> in der japanischen Stichprobe	61
Tabelle 24. Geschätzte standardisierte Faktorladungen, Standardfehler und Kommunalitäten der <i>Individualism Difference Scale</i> in der österreichischen und japanischen Stichprobe	62
Tabelle 25. Invarianzprüfung der <i>Individualism Difference Scale</i> in der Kalibrierungsstichprobe	64
Tabelle 26. Invarianzprüfung der <i>Individualism Difference Scale</i> in der Prüfstichprobe	65
Tabelle 27. Reliabilitätskoeffizienten der <i>Collectivism Scale</i> für die Gesamt- und Teilstichproben	67
Tabelle 28. Klassifikationsanalyse der <i>Collectivism Scale</i>	68
Tabelle 29. Reliabilitätskoeffizienten der <i>Individualism Difference Scale</i> für die Gesamt- und Teilstichproben	68
Tabelle 30. Reliabilitätskoeffizienten der <i>Individualism Difference Scale</i> für die Gesamt- und Teilstichproben	69
Tabelle 31. Klassifikationsanalyse der <i>dichotomen Individualism Difference Scale</i>	70
Tabelle 32. Klassifikationsanalyse der <i>mehrkategorialen Individualism Difference Scale</i>	70

Tabelle 33. Z-Werte der Wald-type Tests und Differenzen der geschätzten Itemparameter für die <i>Collectivism Scale</i> in der österreichischen Stichprobe; Teilungskriterien Median, Geschlecht und Alter	93
Tabelle 34. Ausgeschiedene Items der <i>Collectivism Scale</i> in der österreichischen Stichprobe und Auffälligkeiten in den Teilungskriterien.....	94
Tabelle 35. Z-Werte der Wald-type Tests und Differenzen der geschätzten Itemparameter für <i>Collectivism Scale</i> in der japanische Stichprobe; Teilungskriterien Median, Geschlecht und Alter	95
Tabelle 36. Ausgeschiedene Items der <i>Collectivism Scale</i> in der japanischen Stichprobe und Auffälligkeiten in den Teilungskriterien.....	96
Tabelle 37. Verteilung der absoluten Differenzwerte, getrennt nach japanischer und österreichischer Teilstichprobe	97
Tabelle 38. Verteilung der Abweichung, getrennt nach japanischer und österreichischer Teilstichprobe	98
Tabelle 39. Z-Werte der Wald-type Tests und Differenzen der geschätzten Itemparameter für die <i>dichotome Individualism Difference Scale</i> in der österreichischen Stichprobe; Teilungskriterien Median, Geschlecht und Alter	99
Tabelle 40. Ausgeschiedene Items der <i>Individualism Difference Scale</i> in der österreichischen Stichprobe und Auffälligkeiten in den Teilungskriterien	99
Tabelle 41. Z-Werte der Wald-type Tests und Differenzen der geschätzten Itemparameter für die <i>dichotome Individualism Difference Scale</i> in der japanischen Stichprobe; Teilungskriterien Median, Geschlecht und Alter...	99
Tabelle 42. Z-Werte der Wald-type Tests und Differenzen der geschätzten Itemparameter für die <i>dichotome Individualism Difference Scale</i> in der Kalibrierungstichprobe; Teilungskriterium Kultur.....	100
Tabelle 43. Z-Werte der Wald-type Tests und Differenzen der geschätzten Itemparameter für die <i>dichotome Individualism Difference Scale</i> für die Kreuzvalidierungstichprobe; Teilungskriterium Kultur	100
Tabelle 44. Leave-One-Out Kreuzvalidierung für die <i>Collectivism Scale</i>	101
Tabelle 45. Leave-One-Out Kreuzvalidierung für die <i>dichtome Individualism Difference Scale</i>	101

Tabelle 46. Leave-One-Out Kreuzvalidierung für die <i>mehrkategorielle Individualism Difference Scale</i>	101
Abbildung 1. Itemcharakteristikkurven für drei <i>Rasch</i> -Modell konforme Items.....	32
Abbildung 2. Grafische Modelltests für die <i>Collectivism Scale</i> in der österreichischen Stichprobe ($n_{\delta} = 277$) nach Ausschluss von 5 Items, Teilungskriterien <i>Score</i> , <i>Geschlecht</i> und <i>Alter</i>	50
Abbildung 3. Grafische Modelltests für die <i>Collectivism Scale</i> in der japanischen Stichprobe ($n_j = 531$) nach Ausschluss von 5 Items, Teilungskriterien <i>Score</i> , <i>Geschlecht</i> und <i>Alter</i>	51
Abbildung 4. Grafische Modelltests für die <i>Individualism Difference Scale</i> in der österreichischen Stichprobe ($n_{\delta} = 277$) nach Ausschluss von 1 Item, Teilungskriterien <i>Score</i> , <i>Geschlecht</i> und <i>Alter</i>	53
Abbildung 5. Grafische Modelltests für die <i>Individualism Difference Scale</i> in der japanischen Stichprobe ($n_j = 531$), Teilungskriterien <i>Score</i> , <i>Geschlecht</i> und <i>Alter</i>	53
Abbildung 6. Grafische Modelltests für die <i>Individualism Difference Scale</i> in der Kalibrierungsstichprobe ($n_k = 567$), Teilungskriterium <i>Kultur</i>	54
Abbildung 7. Grafische Modelltests für die <i>Individualism Difference Scale</i> in der Prüfstichprobe ($n_p = 242$), Teilungskriterium <i>Kultur</i>	55
Abbildung 8. Messmodell der <i>Collectivism Scale</i> mit dem reduzierten Itempool...	57
Abbildung 9. Messmodell der <i>Individualism Difference Scale</i> mit dem reduzierten Itempool	62
Formel 1. Formel für das dichotome <i>Rasch</i> -Modell.....	31
Formel 2. Formel für das ordinalen <i>Rasch</i> -Modell	33
Formel 3. Regressionsgleichung der konfirmatorischen Faktorenanalyse	34
Formel 4. Formel zur Äquivalenzprüfung zweier Reliabilitätskoeffizienten.....	65

Einleitung

Der Beginn der modernen quantitativ orientierten Psychologie als Wissenschaft wird vielfach mit dem Namen Wilhelm Wundt in Verbindung gebracht. Dabei ist hinreichend bekannt, dass er mit der Begründung des ersten psychologischen Laboratorium der Welt in Leipzig im Jahr 1879 die experimentelle Forschung innerhalb der Psychologie einleitete (Becker-Carus, 2003). Weniger bekannt ist die Tatsache, dass Wundt schon damals die Ansicht vertrat, dass das Erleben und Verhalten von Menschen nur im Kontext ihrer Geschichte und ihrer Kultur verstanden werden kann (vgl. Wundt, 1900-1914). Aus heutiger Sicht kann man ihn aus diesem Grund auch als Kulturpsychologen bezeichnen (Nisbett, 2007). Mitte der 60er Jahre rückte man allerdings von einer Psychologie, die auch kulturellen Aspekten Aufmerksamkeit schenkt, ab. Infolgedessen wurde psychologische Forschung unter einem universalistischen Ansatz betrieben. Demnach glaubte man zur damaligen Zeit psychologische Prozesse als universelle Phänomene unabhängig vom Kontext erklären zu können (z. B. mit Hilfe der Lerntheorie). Erst die Forschungsarbeiten von Miller (1984) und insbesondere Markus und Kitayama (1991) erweckten die Aufmerksamkeit der Forscher. Die Autoren konnten nämlich zeigen, dass grundlegende psychologische Prozesse durchaus von den jeweiligen kulturellen Bedingungen abhängen. Im Zuge dieser Untersuchungen wurde nach und nach die universalistische Sichtweise aufgegeben (Nisbett, 2007; Triandis, 2007). Mittlerweile ist der Einfluss kultureller Gegebenheiten auf das Verhalten und Erleben von Menschen zur Genüge in der Forschung bekannt. Man kann sogar von einem wissenschaftlichen Trend sprechen, bei dem es darum geht interkulturelle Unterschiede aufzudecken oder Forschungsergebnisse in Hinblick auf interkulturelle Unterschiede zu erklären. Insbesondere stehen mit Individualismus und Kollektivismus (Hofstede, 1980, 2001) zwei Konstrukte zur Verfügung, die einen geeigneten theoretischen Rahmen dafür bieten. In diesem Zusammenhang besteht jedoch auch die Gefahr und sogar auch die Tendenz jegliche länderübergreifenden Unterschiede auf Individualismus und Kollektivismus zurückzuführen (Oyserman, Coon & Kemmelmeier, 2002). Dies ist besonders dann problematisch, wenn bestimmte Voraussetzungen für die Vergleichbarkeit der erhobenen Daten nicht erfüllt sind bzw. diese Voraussetzungen nicht geprüft werden. Schließlich dürfen Ergebnisse kulturvergleichender Untersuchungen nur dann interpretiert werden, wenn die statistischen Äquivalenz der erhobenen Daten nachgewiesen wurde (Chen &

West, 2008). Die vorliegende Arbeit nimmt sich dieser Problemstellung an und zeigt zwei Methoden zur psychometrischen Äquivalenzprüfung auf: Zum einen das *Rasch-Modell* und zum anderen die konfirmatorische Mehrgruppen-Faktorenanalyse. Exemplarisch werden zwei Messinstrumente mittels dieser Methoden auf statistische Äquivalenz in zwei Kulturen geprüft. Dabei handelt sich um zwei Erhebungsinstrumente für Kinder- und Jugendliche, die Kollektivismus bzw. Individualismus erfassen. Die Auswahl der untersuchten Erhebungsinstrumente ergab sich einerseits aufgrund der Popularität dieser Konstrukte innerhalb der kulturvergleichenden Forschung und andererseits aufgrund des aufkommenden Interesses an kulturvergleichenden Untersuchungen bei Kinder- und Jugendlichen (vgl. Alesi & Pepi, 2008; Lee, 2004; Nesdale & Naito, 2005).

Im theoretischen Teil dieser Arbeit werden zunächst die Rahmenbedingungen kulturvergleichender Forschung innerhalb der Psychologie umrissen. Danach folgt ein Kapitel zu den Konstrukten Individualismus und Kollektivismus, welches sich insbesondere den Erhebungsmethoden und -instrumenten widmet und die Konstrukte auch kritisch betrachtet. Der theoretische Teil schließt mit einem Kapitel zu methodischen Aspekten der kulturvergleichenden Forschung, welches das zentrale Thema dieser Arbeit beinhaltet, nämlich statistische Äquivalenz und psychometrische Äquivalenzprüfung. Im empirischen Teil dieser Arbeit werden die Erhebungsinstrumente der vorliegenden Arbeit vorgestellt. Hierauf erfolgt die Darstellung der psychometrischen Äquivalenzprüfung, die mit Hilfe zweier Methoden, die der notwendigen methodischen Stringenz entsprechen, durchgeführt werden. Abschließend werden beide Erhebungsinstrumente auch hinsichtlich der psychometrischen Güte evaluiert.

THEORETISCHER TEIL

1 Kultur und Psychologie

Das vorliegende Kapitel beschäftigt sich mit dem Verhältnis von Kultur und Psychologie. Ausgehend von zwei Definitionen zum Begriff Kultur werden verschiedene Konzepte innerhalb der Psychologie vorgestellt, die den Einfluss von Kultur auf das Verhalten und Erleben von Menschen bzw. die Wechselbeziehung von Kultur und Psychologie behandeln. Weiters werden fünf Forschungsansätze vorgestellt, die sich mit der Erforschung dieser Wechselbeziehung auseinandersetzen. Den Abschluss dieses Kapitels bildet die Darlegung der Forschungsarbeiten von Geert Hofstede, dessen Buch *Culture's Consequences* das Forschungsfeld nachhaltig beeinflusst haben.

1.1 Der Kulturbegriff in den Sozialwissenschaften

Der Gebrauch des Begriffs „Kultur“ scheint im alltäglichen Sprachegebrauch fest verankert zu sein und meint im Allgemeinen „Dinge und Werte der menschlichen Gesellschaft“ (Götz & Wellmann, 2009, S. 510). In den Sozialwissenschaften hingegen ist kein einheitlicher Gebrauch dieses Begriffs aufzufinden. Vielmehr herrscht sowohl innerhalb einer Fachdisziplin, wie etwa Anthropologie (vgl. Borowsky, Barth, Shweder, Rodseth & Stolzenberg, 2001), als auch zwischen den Fachdisziplinen Uneinigkeit über die Definition dieses Begriffs (vgl. Jahoda, 2007). Lonner und Adamopoulos (1997) schätzen, dass es in der sozialwissenschaftlichen Literatur mehr als 200 Definitionen von Kultur gibt. Dieser fehlende Konsens stellt nach Thomas (2004) auch das größten Manko innerhalb dieses Forschungsfeldes dar. Zusammenfassend weist Jahoda (2007) darauf hin, dass ein Konsens in der Debatte um die Begriffsdefinition möglicherweise prinzipiell unmöglich ist.

Vor dem Hintergrund dieser Sachlage werden im Folgenden lediglich die Kulturdefinition jener zweier Autoren vorgestellt, die eine Schlüsselrolle für die Entwicklung der kulturvergleichenden Forschung innerhalb der Psychologie eingenommen haben. Zum einen handelt es sich dabei um Harry C. Triandis, der mit seinem Buch über subjektive Kultur (1972) zur Begründung der kulturvergleichenden Psychologie beigetragen hat (Schimmack, Shigehiro & Diener, 2005). Zum anderen ist dies die Definition von Geert Hofstede, dessen Buch *Culture's Consequences* (1980) die kulturvergleichende Forschung nachhaltig geprägt hat.

Triandis (1995) erläutert das Konzept der Kultur anhand folgender Analogie: „Culture is to society what memory is to individuals“ (S. 4) und unterscheidet zwischen objektiver und subjektiver Kultur (Triandis, 1972). Ersteres beinhaltet beobachtbare Elemente, wie z. B. Artefakte oder soziale Strukturen, während Letzteres alle nicht beobachtbaren Elemente, wie z. B. Einstellungen, Normen oder Rollen, einschließt. Seiner Definition zufolge werden nur jene Elemente der Kultur durch die Gesellschaft über die Generationen hinweg weiter getragen, die sich in der Vergangenheit bewährt haben. Unterschiede zwischen den Kulturen werden dabei durch die Sprache, die Zeit (historische Periode) und den Ort (geografische Lage) festgelegt (Triandis, 2007).

Hofstede (2007) fasste Kultur hingegen als kollektives Phänomen auf und beschreibt sie als „kollektive Programmierung des Geistes, durch die sich die Mitglieder einer Gruppe oder Kategorie von Menschen von anderen unterscheiden“ (S. 387). Zwar kann sich der Begriff auf Organisationen oder Berufsgruppen beziehen, jedoch unterscheiden sich gesellschaftliche oder nationale Kulturen dadurch, dass sie als unbeewusste Werte im menschlichen Geist verwurzelt sind.

Trotz zahlreicher Definitionen des Begriffs „Kultur“ können drei Aspekte als Charakteristika einer Kultur extrahiert werden (Triandis, 2007). Erstens ergibt sich Kultur aus einer adaptiven Interaktion zwischen Mensch und Umwelt; zweitens besteht sie aus geteilten Elementen (z. B. Einstellungen, Werte und Normen); und drittens wird Kultur über die Zeit und Generationen hinweg weiter getragen.

1.2 Kultur-, kulturvergleichende- und soziokulturelle Psychologie

Die Kulturpsychologie einerseits und die kulturvergleichende Psychologie andererseits bezeichnen unterschiedliche Ansätze, die in ihren Forschungsbemühungen die Wechselbeziehung zwischen Kultur und Psyche einbeziehen, aber unter dem Begriff *culture inclusive psychology* zusammengefasst werden können (Boesch & Straub, 2007). Die Unterschiede beider Ausrichtungen sind nach Triandis (2007) zufolge nicht substantiell und betreffen lediglich die Schwerpunktsetzung. Auch Miller (2002) beschreibt die Kultur- und kulturvergleichende Psychologie als zwei Konzepte, die sich durch unterschiedliche Sichtweisen bezüglich der Rolle der Kultur innerhalb der psychologischen Theoriebildung und durch die differente Auffassung hinsichtlich der Zielsetzung einer kulturorientierten Forschung unterscheiden.

Die *Kulturpsychologie* schenkt auf der einen Seite dem Kontext mehr Beachtung als dem Inhalt und untersucht dementsprechend weniger stabile Eigenschaften einer Kultur (Triandis, 2007). Einer der Hauptziele kulturpsychologischer Untersuchungen ist das Generieren neuer Konzepte und Theorien, die zum Verständnis kulturspezifischer Aspekte beitragen (Miller, 2002). Häufig konzentriert sich die Forschungsarbeit darauf wenige Kulturen intensiv mittels ethnographische Methoden zu untersuchen, sodass die Kulturpsychologie zusammenfassend als kulturwissenschaftlicher, interpretativer bzw. hermeneutischer Ansatz gesehen werden kann (Boesch & Straub, 2007).

Die *kulturvergleichende Psychologie* beschäftigt sich auf der anderen Seite vermehrt mit dem Inhalt als mit dem Kontext und untersucht Attribute von Kulturen, die sich über die Zeit kaum verändern (Triandis, 2007). Hauptanliegen dieser Forschungsausrichtung ist die Untersuchung der Allgemeingültigkeit vorhandener psychologischer Theorien in unterschiedlichen kulturellen Kontexten und die Entwicklung universeller Modelle zur Vorhersage von Verhalten unabhängig der Kultur (Miller, 2002). In dieser Forschungstradition wird Kultur vorwiegend innerhalb groß angelegter Fragebogenuntersuchungen in vielen kulturellen Gruppen als weitere unabhängige Variable konzipiert. Darum wird die kulturvergleichende Psychologie als naturwissenschaftliche, nomothetischer bzw. nomologischer Ansatz gesehen (Boesch & Straub, 2007).

Eine zentrale Frage Kultur einbeziehender Psychologie stellt jene nach der Verortung der Kultur, nämlich außerhalb oder innerhalb der Person, dar. Während Kulturpsychologen die Kultur vorwiegend innerhalb der Person sehen, wird in der kulturvergleichenden Psychologie die Kultur weitestgehend außerhalb der Person verortet. Tatsächlich finden sich in jeder Kultur Bräuche, die bereits vor Existenz bestimmter Einzelpersonen bestehen (Kultur außerhalb der Person) und das Verhalten von Personen durch bestimmte internalisierte Interpretationsvorschriften beeinflussen (Kultur innerhalb der Person). Kurzum, Kultur ist sowohl innerhalb als auch außerhalb der Person zu sehen (Triandis, 2007).

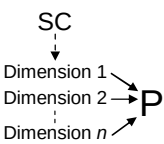
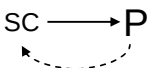
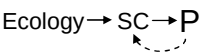
Die *soziokulturelle Psychologie* ist bemüht, die Interdependenz zwischen der Kultur innerhalb und außerhalb der Person hervorzuheben. Individuen sind stets in einem sozialen Kontext eingebettet; der soziale Kontext wiederum existiert jedoch nicht unabhängig von den Individuen. Die Rede ist von *mutual constitution*, einem dynamischen Prozess wechselseitiger Beeinflussung. Dementsprechend verfolgt die soziokulturelle Psychologie einerseits die Entdeckung systematischer Prinzipien

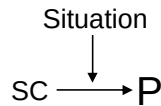
kulturell beeinflusster Gesellschaften und andererseits die Beschreibung des bestimmenden Prozesses menschlicher Sozialisation.

1.3 Forschungsansätze der soziokulturellen Psychologie

Laut Markus und Hamedami (2007) lassen sich Forschungsarbeiten, die sich mit der dynamischen Interdependenz des psychologischen und soziokulturellen Systems befassen, in fünf Forschungsansätzen zusammenfassen. Die Unterschiede zwischen diesen Ansätzen liegen in der theoretischen Konzeption und in der empirischen Vorgehensweise, psychologische und soziokulturelle Aspekte der Kultur und deren Wechselbeziehung zu untersuchen. Tabelle 1 zeigt eine schematische Auflistung der fünf Forschungsansätze. Jede schematische Darstellung enthält sowohl ein SC für das soziokulturelle System als auch ein P für das psychologische System. Deren relative Größe zueinander verdeutlicht die Gewichtung auf das Psychologische bzw. Soziokulturelle innerhalb des jeweiligen Forschungsansatzes. Drei der veranschaulichten Forschungsansätze enthalten zusätzliche Begriffe (*dimension*, *ecology* und *situation*), die weitere zentrale Konzepte innerhalb dieser Ansätze darstellen. Die Pfeile verdeutlichen die angenommene Interaktion zwischen dem soziokulturellen und psychologischen System; eine durchgezogene Linie steht dabei für eine genau vorgegebene, eine gestrichelte Linie für eine weniger genau vorgegebene Beziehung.

Tabelle 1. Forschungsansätze der soziokulturellen Psychologie (Markus und Hamedami, 2007, S. 14)

Forschungsansatz	Empirisches Ziel
<i>Dimensional</i> 	Beschreibt Dimensionen einer Kultur, die Unterschiede in Einstellungen, Werte und Verhalten erklären
<i>Models</i> 	Beschreibt Modelle, die die Beziehung zwischen dem soziokulturellen und dem psychologischen System erklären
<i>Cognitive Toolkit</i> 	Beschreibt wie Bedeutung und Bräuche einer Kultur grundlegende kognitive Prozesse beeinflussen
<i>Ecocultural</i> 	Beschreibt wie ökologische und sozialpolitische Faktoren die psychologische Adaptation an die Umwelt beeinflussen
<i>Dynamic Constructivist</i>	Beschreibt Situationsfaktoren und Rahmenbedingungen, die den kulturellen Einfluss regulieren



Anmerkung. SC: Soziokulturelle System; P : Psychologische System

(1) *Dimensional Approach*. Viele Arbeiten im Rahmen der kulturvergleichenden Forschung betrachten Kultur als komplexe, multidimensionale Struktur, welche anhand bestimmter Dimensionen beschrieben werden kann. Dies entspricht der Betrachtungsweise von Kulturen als *cultural syndromes* (Triandis, 1996), definiert durch eine Reihe geteilter Einstellungen, Überzeugungen, Normen, Rollen und Werte innerhalb einer Personengruppe, die eine bestimmte Sprache sprechen, während einer bestimmten historischen Periode und in einem umgrenzten geografischen Gebiet leben.

Bei diesem Forschungsansatz wird der Einfluss des soziokulturellen Systems auf das psychologische System durch Unterschiede in den Ausprägungen von verschiedenen Dimensionen, wie z. B. Individualismus-Kollektivismus, Machtdistanz, Unsicherheitsvermeidung (vgl. Hofstede, 1980), zwischen Mitgliedern unterschiedlicher Kulturen erfasst.

Der dimensionale Forschungsansatz bietet die Möglichkeit den Einfluss kultureller Elemente auf das Psychologische durch einfach strukturierte Dimensionen zu organisieren bzw. zu erfassen und diese über verschiedene kulturelle Kontexte hinweg zu vergleichen. Jedoch bietet er wenige Erklärungsansätze für den Interaktionsprozess zwischen dem soziokulturellen und dem psychologischen System.

(2) *Sociocultural Models Approach*. Dieser Forschungsansatz legt den Fokus auf den Interaktionsprozess zwischen dem soziokulturellen und dem psychologischen System. Demnach existieren soziokulturelle Modelle, die sowohl implizite als auch explizite kulturgefärbte Ideen und Bräuche enthalten und individuelle Erfahrungen, wie z. B. Wahrnehmung, Kognition, Emotion oder Motivation beeinflussen. Diese Modelle bestehen nicht nur in der Psyche von Individuen, sondern strukturieren in einem dynamischen Zusammenspiel auch die Welt, in der diese Individuen leben. In einem kontinuierlichen sozialen Austauschprozess wird so das soziokulturelle und psychologische System geformt.

ˆ Trotz theoretischer und empirischer Berücksichtigung der Interdependenz zwischen dem soziokulturellen und dem psychologischen Systems bleibt dieser

Ansatz einer Erklärung über den genauen Wirkungsprozess schuldig und verfügt darüber nur grobe Erklärungsansätze. Zudem bietet der soziokulturelle Modellansatz wenig empirische Erklärungsansätze wie die Kultur von Individuen geformt und über die Zeit verändert wird.

(3) *Tool Kit Approach*. Eine weitere Betrachtungsweise von Kultur stellt jene als Werkzeugkasten bzw. als ein Bündel von Hilfsmitteln dar, die Individuen bei der Wahrnehmung und Konstruktion von Bedeutung in der Welt unterstützen. Dabei wird angenommen, dass grundlegende psychologische Prozesse, wie beispielsweise Kognition und Wahrnehmung durch die Kultur beeinflusst werden. Demnach interagiert das Soziokulturelle mit dem Psychologischen über das kognitive System, weshalb das Hauptinteresse dieses Ansatzes an spezifischen psychologischen Funktionen liegt. Dies stellt auch den größten Kritikpunkt dar, nämlich die geringe Beachtung des Wirkungsprozesses selbst.

(4) *Ecocultural approach*. Im Rahmen dieses Ansatzes wird die Interaktion zwischen kulturellen und psychologischen Prozessen durch den kulturellen Kontext, bestimmt durch ökologische und sozialpolitische Faktoren, erklärt. Dieser Sichtweise liegt die Annahme einer adaptiven Anpassung an objektive Anforderungen des physischen und sozialen Lebensraums als Erklärung von kulturellen Unterschieden zugrunde. Die Kritik an diesem Ansatz bezieht sich auf den ökologischen Determinismus, der Kultur als statisches Gebilde auffasst. Jedoch wird von heutigen Vertretern des ökokulturellen Ansatzes auch die dynamische Interaktion mit der Umgebung berücksichtigt, die im Rahmen eines adaptiven Prozesses ebenfalls eine Veränderung durchläuft.

(5) *Dynamic constructivist Approach*. Vertreter dieses Forschungsansatzes heben Kultur als internalisiertes Konstrukt von losen Netzwerken von Wissensstrukturen, mentalen Abbildungen und Repräsentationen hervor. Dieses Konstrukt wird innerhalb eines bestimmten Kontexts geteilt und beeinflusst erst bei Aktivierung die Informationsverarbeitung. Das bedeutet, Kultur beeinflusst nur in bestimmten sozialen Situationen die Kognition, und das geschieht unter der Voraussetzung, dass relevante implizite Theorien oder geteilte Annahmen verfügbar, zugänglich und in der jeweiligen Situation anwendbar sind. Der Fokus dieses Ansatzes liegt somit vorwiegend auf den Bedingungen unter denen kulturelle Einflüsse vonstatten gehen und weniger auf dem Prozess, durch welchen sich Kultur und Psyche gegenseitig bedingen.

1.4. Culture's Consequences

Hofstede (1980, 2001) untersuchte zwischen 1973 und 1978 Angaben aus Mitarbeiterbefragungen, die zwischen 1967 und 1973 in lokalen Niederlassungen des US-amerikanischen IT- und Beratungsunternehmens *International Business Machines Corporation* (IBM) durchgeführt worden sind. Dabei wurden in über 116.000 Fragebögen in 20 verschiedenen Sprachen etwa 88.000 Personen aus 72 Ländern zu arbeitsbezogenen Themen befragt. Besonderes Interesse fanden 14 Items zu *work goal importance*, die in 40 Ländern vorgegeben wurden. Zusammen mit 18 weiteren Items führte Hofstede eine Faktorenanalyse durch, wobei als Untersuchungseinheit die Länder ($N = 40$) und nicht die einzelnen Individuen fungierten. Die Analysen ergaben drei Faktoren, die zusammen 49% der Varianz aufklärten. Davon wurde der erste Faktor in die beiden Dimensionen (1) „Individualismus vs. Kollektivismus“ und (2) „Machtdistanz“ aufgeteilt, weshalb diese Dimensionen eine Korrelation von -0.68 aufwiesen, obgleich sich dieser Zusammenhang bei Kontrolle des nationalen Wohlstandes (Bruttonationaleinkommen pro Kopf) auf $r = -0.32$ reduzierte. Die weiteren extrahierten Faktoren waren (3) „Maskulinität vs. Feminität“ und (4) „Unsicherheitsvermeidung“. Später wurden diese vier Dimensionen in Folge des *Chinese Value Survey project* (Hofstede & Bond, 1984; *The Chinese Culture Connection*, 1987) um die Dimension (5) „Langzeit- vs. Kurzzeitorientierung“ ergänzt. Demzufolge können kulturelle Unterschiede zwischen Nationen auf fünf Dimensionen abgebildet werden (siehe Hofstede, 2001):

(1) *Individualismus vs. Kollektivismus* als zwei Pole einer Dimension beschreibt die Beziehung zwischen dem Individuum und dem Kollektiv einer Gesellschaft und spiegelt die Gestaltung des Zusammenlebens wider. Die Beziehung zwischen Individuen kann entweder locker, d. h. jeder kümmert sich um sich selbst und seine unmittelbare Familie (Individualismus) oder stark, d. h. jeder ist von Geburt an in Wir-Gruppen integriert (Kollektivismus), sein.

(2) Die Dimension *Machtdistanz* beschreibt den Umgang von Mitgliedern einer Gruppe mit ungleicher Machtverteilung. Eine hohe Ausprägung dieser Dimension deutet insbesondere darauf hin, dass die niedriger Gestellten einer Gesellschaft diese Ungleichheit ebenso akzeptieren wie auch die höher Gestellten.

(3) Der Faktor *Maskulinität vs. Feminität* bezieht sich auf die Frage nach der emotionalen und sozialen Rollenverteilung zwischen den Geschlechtern. Während die Betonung auf dem einem Pol dieser Dimension auf der Durchsetzungsfähigkeit

und Wettbewerbsorientierung (Maskulinität) liegt, wird auf dem anderen Pol Bescheidenheit und Fürsorge hervorgehoben (Feminität).

(4) *Unsicherheitsvermeidung* ist klar abzugrenzen von dem Begriff Risikovermeidung und meint den Umgang einer Gesellschaft mit Unsicherheit. Eine hohe Unsicherheitsvermeidung ist durch strikte Verhaltensvorschriften und Regeln gekennzeichnet, die das Auftreten unstrukturierter Situationen verhindern sollen.

(5) *Langzeitorientierung* vs. *Kurzzeitorientierung* beinhalten Werte der Lehren von Konfuzius um 500 vor Christus und bezeichnen einerseits: Ausdauer und Sparsamkeit (Langzeitorientierung) und andererseits Respekt für Tradition (Kurzzeitorientierung).

Seit Erscheinen der ersten Auflage von *Culture's Consequence* im Jahr 1980 wurden Hofstede's Kulturdimensionen von unterschiedlichen fachlichen Disziplinen zur kulturvergleichenden Forschung aufgenommen. Innerhalb der Psychologie fand die Dimension Individualismus vs. Kollektivismus die größte Aufmerksamkeit, wie etwa die Veröffentlichung von Büchern zu diesem Thema (z.B. Kim, Triandis, Kagitcibasi, Choi & Yoon, 1994; Triandis, 1995) oder eines Buchkapitels in Handbüchern zu kulturvergleichenden Psychologie (z.B. Kagitcibasi, 1997) zeigen. In Abhängigkeit der akademischen Disziplin fanden die anderen Dimensionen mehr oder weniger Beachtung. So wurde in der vergleichenden Unternehmensführung und Organisationstheorie vermehrt die Dimensionen Machtdistanz und Unsicherheitsvermeidung untersucht, während in der Politikwissenschaft vorwiegend die Dimension Maskulinität vs. Feminität von Interesse war (Hofstede, 1994).

In Folge der Popularität der fünf Kulturdimensionen konnten zahlreiche Validierungsstudien durchgeführt werden (siehe Hofstede, 2001). Mehrheitlich wurden die Dimensionen mit den Ergebnissen anderer kulturvergleichender Studien korreliert. Dabei zeigte sich, dass die Länderunterschiede in den Dimensionen tatsächlich grundlegende und stabile Beschreibungen von Kulturunterschieden darstellen.

Abschließend sei angemerkt, dass sich neben Hofstede auch andere Wissenschaftler mit der Beschreibung von kulturellen Unterschieden beschäftigt haben: Schwartz (1994; Schwartz & Bardi, 2001) befragte Grundschullehrer und Studierende aus 50 Ländern zu 56 Werten und konnte sieben Dimensionen unterscheiden: Konservatismus, Beherrschung, Hierarchie, intellektuelle Autonomie, Egalitarismus, affektive Autonomie und Harmonie. Inglehart (1997)

hingegen fand im Rahmen der Weltwertestudie (Ingelhart, Basanez & Moreno, 1998) zwei Dimensionen: Wohlbefinden vs. Überleben und säkular-rational vs. traditionelle Autorität. Die Dimensionen beider Untersuchungen korrelieren signifikant mit den Dimensionen von Hofstede.

2 Individualismus – Kollektivismus (Ind-Col)

Seit Publikation von *Culture's Consequence* (Hofstede, 1980) haben die Konstrukte Individualismus und Kollektivismus große Beliebtheit in der kulturvergleichenden Forschung erfahren. Dies beschränkt sich nicht nur auf die Psychologie, sondern zieht sich durch alle Bereiche der Sozial-, Verhaltens- und Geisteswissenschaften (Kagitcibasi, 1997). Angefangen mit einer Begriffsbestimmung dieser Konstrukte, wird im Folgenden näher auf die Frage der Dimensionalität, der Ansätze zur Erforschung von Ind-Col und zu den Erhebungsmethoden und -instrumenten eingegangen. Zuletzt werden diese Konstrukte bzw. wird die Forschungsmethodologie im Zusammenhang mit diesen Konstrukte kritisch beleuchtet.

2.1 Begriffsbestimmung

Harry C. Triandis machte Ind-Col innerhalb der Psychologie bekannt (Kagitcibasi, 1997). Seiner Ansicht nach lässt sich Kultur als *cultural syndrom* betrachten (siehe Kapitel. 1.3). Auf dieser Betrachtungsweise von Kulturen baut die Begriffsbestimmung von Ind-Col auf, sodass sich diese Konstrukte nicht eindeutig bestimmen lassen und keine abgegrenzte Begriffsdefinition aufweisen. Vielmehr werden diese Begriffe mit Hilfe von Attribute beschrieben und stellen somit polythetische Konstrukte dar. Laut Triandis (1995) lässt sich Ind-Col anhand vier Hauptattribute definieren:

- (1) *The meaning of the self.* Während Individualisten über eine independente Selbstkonstruktion verfügen, weisen Kollektivisten eine interdependente Selbstkonstruktion auf.
- (2) *The structure of goals.* Individualisten gehen persönlichen Zielen nach, unabhängig davon, ob diese mit den Zielen ihrer Ingroup kompatibel sind oder nicht. Kollektivisten hingegen verfolgen nur Ziele, wenn sie auch mit jenen ihrer Ingroup vereinbar sind.

- (3) *Behaviour as function of norms and attitudes*. Das Verhalten von Individualisten wird über Situationen hinweg von ihrer eigenen Einstellung beeinflusst, wohingegen Kollektivistinnen sich in ihrem Verhalten eher nach der Norm richten.
- (4) *Focus on the needs of the ingroup or social exchanges*. Individualisten wägen Vorteile und Kosten einer Beziehung im Sinne der Austauschtheorie ab. Wenn die Kosten die Vorteile überwiegen, wird die Beziehung aufgegeben, während Kollektivistinnen den Fokus auf den Bedürfnissen der Mitglieder einer Ingroup setzen und die Beziehung trotz überwiegender Kosten aufrechterhalten.

Neben diesen vier Hauptattribute identifizierte Triandis (1995) 60 weitere Attribute, die in kollektivistischen oder individualistischen Kulturen zu finden sind. Das wichtigste Unterscheidungsmerkmal von Individualismus und Kollektivismus stellt dabei das Attribut horizontal – vertikal dar. Diese Dimension trägt dem Umstand Rechnung, dass manche Kulturen Gleichheit (z. B. Australien, Schweden oder Kibbuzim) betonen, wogegen andere Wert auf Hierarchie legen (z. B. Indien oder Amerika). Mithilfe dieser zusätzlichen Dimension lässt sich Individualismus bzw. Kollektivismus in Abhängigkeit der Gewichtung von horizontalen oder vertikalen Aspekten sozialer Beziehungen in vier kulturelle Orientierungen unterteilen (Triandis, 1995; Triandis & Gelfand, 1998): (1) horizontaler Individualismus (Personen wollen einzigartig und unabhängig von der Gruppe sein), (2) vertikaler Individualismus (Personen wollen herausstechen und hohen Status erlangen), (3) horizontaler Kollektivismus (Personen sehen sich als Teil der Gruppe und verfolgen gemeinsame Ziele) und (4) vertikaler Kollektivismus (Personen ordnen sich Autoritäten einer Gruppe unter und sind bereit sich für die Gruppe aufzuopfern).

2.2 Dimensionalität

Im Zuge einer Untersuchung in 40 verschiedenen Ländern aller Welt beschrieb Hofstede (1980) Individualismus als eindimensionales Konstrukt, dessen Gegenpol er als Kollektivismus bezeichnete. Obwohl sich diese Beschreibung auf die Ebene der Länder bezieht, wurde in nachfolgende Forschungsarbeiten auf Individualebene ebenfalls die Betrachtungsweise von Individualismus und Kollektivismus als zwei Pole ein und derselben Skala vertreten. Dieser ökologischer Fehlschluss (Bond, 2002) stellte lange Zeit das größte Missverständnis im Rahmen der Forschung zu Ind-Col dar (Hoppe, 2004). Mittlerweile wird Individualismus und Kollektivismus auf Individualebene als zwei unabhängige Dimensionen betrachtet

(vgl. Bond, 2002; Fiske, 2002; Li & Aksoy, 2006; Oyserman et al., 2002). Um die Individualebene der Analyse hervorzuheben, wurden hierauf neue Konzepte eingeführt, wie z. B. die Theorie vom *independent* und *interdependent self-construal* (Markus & Kitayama, 1991), die selbstbezogene Aspekte von Individualismus und Kollektivismus betont. Triandis, Leung, Villareal und Clack (1985) schlugen weiters vor, auf Individualebene von *idiocentrism* und *allocentrism* zu sprechen und Individualismus und Kollektivismus auf die Kulturebene zu beziehen. Somit wäre es möglich, sowohl von idiozentrischen Personen (denken, fühlen und handeln wie Personen in individualistischen Kulturen) als auch von allozentrischen Personen (wie Personen in kollektivistischen Kulturen) zu sprechen, die entweder innerhalb individualistischen oder kollektivistischen Kultur leben. Laut Triandis (Triandis, 2004) leben schätzungsweise 35 bis 100% idiozentrische und 0 bis 35% allozentrische Personen in individualistischen Kulturen, wohingegen 0 bis 35% idiozentrische und 30 bis 100% allozentrische Personen in kollektivistischen Kulturen leben.

Zusammenfassend kann nach Oyserman (2006) die Betrachtungsweise der Konstrukte Individualismus bzw. Kollektivismus in drei Phasen unterteilt werden. Die anfängliche Forschung verfolgte ein einfaches Konzept vom Individualismus als eindimensionales Konstrukt. Eine zweite Welle von Forschungsarbeiten beschäftigte sich mit Individualismus und Kollektivismus als zwei unabhängige Konstrukte. Die aktuellen Forschungsbestrebungen weisen die Tendenz auf Individualismus und Kollektivismus als dynamische Merkmale einer Kultur zu sehen. Die mit den Konstrukten assoziierten Verhaltensweisen sind dabei jeder Person universell verfügbar, werden jedoch mit unterschiedlicher Wahrscheinlichkeit ins Arbeitsgedächtnis gerufen.

2.3 Ansätze zur Erforschung von Ind-Col

Oyserman et al. (2002) unterscheiden drei allgemeine Forschungsansätze, die zum Studium von Individualismus bzw. Kollektivismus in den heutigen Forschungsarbeiten verfolgt werden:

(1) *Applying Hofstede*. Dieser Forschungsansatz beruht auf den Arbeiten von Hofstede (1980), dessen Rangliste des nationalen Ausprägungsgrades von Individualismus die Grundlage dieses Zugangs darstellt. Hierbei werden entweder die Einschätzungen bzgl. der Ausprägung von Individualismus von Hofstede (1980) übernommen oder Unterschiede in anderen untersuchten Variablen zwischen

zwei oder mehr Ländern auf die ungleiche Ausprägung von Individualismus gemäß Hofstede zurückgeführt. Beide Herangehensweisen gehen davon aus, dass die von Hofstede erforschten nationalen Unterschiede im Ausprägungsgrad von Individualismus (a) über verschiedene Lebensbereiche, also auch über den Arbeitskontext hinaus, generalisierbar sind, (b) stabil über die Zeit sind und (c) relevant auch für die individuelle Ebene sind. Der fehlende empirische Nachweis für diese Annahmen stellt die größte Kritik für dieses Vorgehen dar. Nichtsdestotrotz stellt *Applying Hofstede* den am häufigsten gebrauchte Forschungsansatz dar, welcher zu wichtigen Erkenntnissen im Bereich der kulturvergleichen Forschung beigetragen (Oyserman & Uskul, 2008).

(2) *Direct Assessment*. Eine weitere Möglichkeit Individualismus bzw. Kollektivismus zu untersuchen stellt jene der direkten Erhebung dieser beiden Konstrukte auf individueller Ebene dar. Für gewöhnlich werden dann im Rahmen einer interkulturellen Untersuchung neben interessierenden Variable auch Individualismus bzw. Kollektivismus erhoben, um diese Variablen miteinander zu korrelieren. Unterschiede in der Ausprägung einer Variable können so in Zusammenhang mit unterschiedlicher Ausprägung von Individualismus bzw. Kollektivismus gesetzt werden. Voraussetzung für dieses Vorgehen ist allerdings die psychometrische Äquivalenz der Erhebungsinstrumente in den jeweiligen Kulturen. Eine derartige Äquivalenzprüfung von Erhebungsinstrumenten von Individualismus und Kollektivismus stellen nach wie vor mehr die Ausnahme als die Regel dar (Oyserman et al., 2002).

(3) *Priming Cultural Frame*. Die Forschungsarbeiten zu *social cognition* zeigen, dass einerseits das Verhalten durch das aktivierte soziale Wissen beeinflusst wird und andererseits diese Beeinflussung auch in Laborbedingung auftreten. Auf Grundlage dieser Erkenntnisse wird bei diesem Forschungsansatz versucht, Ind-Col bzw. Werte, die mit diesen Dimensionen assoziiert sind, zu aktivieren (*priming*), um anschließend die Auswirkungen dieses aktivierten Wissens auf die Kognition, die Emotion und das Verhalten zu untersuchen.

Die Vorteile dieses Vorgehens liegen in der Möglichkeit, den Einfluss von Kultur als aktiven Prozess einer Situation (*situated cognition*) zu untersuchen und bestimmte Aspekte von Ind-Col hinsichtlich Kausalität zu überprüfen (siehe Oyserman & Lee, 2007). Die geringere ökologische Validität aufgrund des Fokus auf einzelne Aspekte der beiden Konstrukte stellt wiederum ein Nachteil dieses Ansatzes dar (Oyserman & Uskul, 2008).

2.4. Erhebungsmethoden und -instrumente

Dieses Kapitel beschäftigt sich mit verschiedenen Möglichkeiten zur Erfassung von Ind-Col, einschließlich den inhaltlichen Bereichen von Ind-Col. Zur Veranschaulichung sollen noch drei der am häufigsten genutzten Erhebungsinstrumente näher dargelegt werden.

2.4.1 Operationalisierung und Erhebungsmethoden

Triandis, McCusker und Hui (1990) empfehlen für eine valide Erfassung kultureller Syndrome, wie etwa Individualismus und Kollektivismus, einen *multimethod* Ansatz und beschreiben fünf Zugänge zur Erfassung von Ind-Col:

(1) *Social content of the self*. Bei der „I am ...“-Methode werden Personen aufgefordert Sätze mit „Ich bin ...“ zu vervollständigen. Eine anschließende Inhaltsanalyse kategorisiert die Antworten danach, ob es sich um eine soziale (S) Antwort, beispielsweise „...ein Sohn“ (Familie) oder „...römisch-katholisch“ (Religion), handelt oder nicht. Der prozentuale Anteil an S-Antworten steht dabei für die jeweilige Ausprägung von Kollektivismus einer Person.

(2) *Perceived homogeneity of ingroups and outgroups*. Diese Methode baut auf der theoretischen Überlegung auf, dass Kollektivisten ihre Ingroup homogener beurteilen als Individualisten. Dementsprechend werden Personen gebeten im ersten Schritt verschiedene ethnische Gruppe hinsichtlich Homogenität zu beurteilen, z. B. „Wie sehr sind sich die Mitglieder dieser Gruppe in Bezug auf ihre Ziele einig“. Im zweiten Schritt wird die psychologische Distanz zur beurteilten ethnischen Gruppen erfasst, indem jede Person die Ähnlichkeit des eigenen Selbstbildes zur jeweiligen Gruppe einschätzt. Kollektivisten werden dabei ihre Ingroups homogener betrachten als ihre Outgroups.

(3) *Responses to attitude items*. Bei dieser Methode werden auf einer sechs- oder neustufigen Likert-Skala verschiedene Fragen zu Einstellungen erfasst, die in Zusammenhang mit Individualismus bzw. Kollektivismus stehen. Beispielsweise wird dem Item „Betagte Eltern sollten zusammen mit ihren Kindern leben“ von Kollektivisten eher zugestimmt, während es von Individualisten eher abgelehnt wird (Triandis, Chan, Bhawuk, Iwao & Sinha, 1995).

(4) *Responses to value items*. Schwartz und Bisky (1987) entwickelten 56 Items, die nach der Wichtigkeit von 30 verschiedenen Werten, beispielsweise „*Equality (equal opportunity for all)*“ und 26 verschiedenen Attributen, beispielsweise „*Independent (self-reliant, self-sufficient)*“, fragen. Personen beurteilen diese

Werte und Attribute von *not important*“ bis *of supreme importance*. Während Kollektivisten Werten wie *„honoring parents and elders“* größere Bedeutung beimessen, schätzen Individualisten Werte wie *„an exciting life“* als wichtig ein.

(5) *Perceptions of social behavior as a function of social distance*. Die Theorie, dass im Vergleich zu Individualisten Kollektivisten sich gegenüber ingroup und outgroup differenzierter verhalten, stellt die Grundlage dieser Methode dar. Personen beurteilen die psychologische Distanz zu 20 sozialen Stimuli (z. B. Vater). Diese sozialen Stimuli werden hierauf in Bezug zu sechs Qualitäten sozialen Verhalten gesetzt: *association* (z. B. helfen), *dissociation* (z. B. vermeiden), *superordination* (z. B. kritisieren), *subordination* (z. B. nach Hilfe fragen), *intimacy* (z. B. küssen) und *formality* (z. B. schriftliche Einladung schicken). Daraus ergeben sich Fragen wie z. B.: „Würdest du deinen Vater küssen?“, die mit „Ja“ bzw. „Nein“ beantwortet werden können. Die Skala der psychologischen Distanz und das soziale Verhalten werden mittels direkter Skalierung sensu Stevens (1966) ermittelt. Pro soziale Kategorie wird die Anzahl der bejahten Antworten gegen die soziale Distanz grafisch als Funktion dargestellt. Kollektivisten zeigen dabei mehr positive (*associative, intimate* und *subordinate*) Verhaltensweisen gegenüber Mitgliedern ihrer ingroup und mehr negative (*dissociative, formal* und *subordinate*) gegenüber Mitgliedern ihrer outgroup als Individualisten.

Weitere Ansätze zur Erhebung von Ind-Col wurden von Triandis (1995) folgendermaßen zusammengefasst:

(6) *Content Analyses*. Inhaltsanalysen von verbalen oder schriftlichen Material können Aufschluss über das Ausmaß von Ind-Col geben. Beispielsweise haben Inhaltsanalysen von Autobiografien oder Kinderbüchern gezeigt, dass in individualistischen Länder vermehrt auch individualistische Inhalte in diesen Büchern zu finden sind (vgl. Morsbach, 1980; Shannon, 1986). Weiters konnte Semin und Rubini (1990) zeigen, dass Beleidigungen in individualistischen Ländern häufiger direkt auf die Person gerichtet sind (z. B. *you are a pig, stupid, cretin ugly, uncivil, sucker*), während in kollektivistischen Ländern die nahestehenden Personen beleidigt werden (z. B. *f--- your mother and 36 other relatives*).

(7) *Observation*. Beobachtungsverfahren stellen eine weitere Möglichkeit zur Erhebung von Ind-Col dar. Triandis (1980) beobachtete zum Beispiel systematisch Personen beim Gehen und Autofahren dahingehend, ob sie alleine oder mit anderen Personen unterwegs waren. Dabei zeigte sich, dass beim Verhältnis von

alleine/zusammen in Bezug auf das Gehen keine kulturellen Unterschiede nachweisbar waren, im Bezug auf das Autofahren jedoch schon.

Cirkov, Lnych und Niwa (2005) stellen eine weitere Methode zur Erfassung von Ind-Col vor:

(8) *Scenario Format*. Im *Scenario Questionnaire of Cultural Orientations* (Cirkov et al., 2005) werden Vignetten (kurze Fallbeispiele oder Szenarien), mit vier Erklärungsvorschlägen vorgegeben. Tabelle 2 zeigt eine Beispiel-Vignette.

Tabelle 2. Beispielvignette aus dem *Scenario Questionnaire of Cultural Orientations*

A student is going over Graduate Programs brochures in order to decide which program to attend. What is the most important factor in making this decision?

- 1. It is a very competitive program and one of the best. (VI)*
- 2. The student's professor/supervisor approves of it. (VC)*
- 3. It is the program most of the student's classmates are going to. (HC)*
- 4. It is the program which fits the student's interest and needs the most. (HI)*

Jede der vier Erklärungsvorschläge wird auf einer fünfstufigen Likert-Skala von *strongly disagree* bis *strongly agree* beurteilt. Das Ausmaß der Zustimmung zum ersten Erklärungsvorschlag spiegelt vertikalen Individualismus (VI), zum zweiten vertikalen Kollektivismus (VC), zum dritten horizontalen Kollektivismus (HK) und zum vierten horizontalen Individualismus (HI) wider.

2.4.2 Inhaltliche Bereiche von Individualismus- und Kollektivismus-Skalen

Oyserman et al. (2002) betonen, dass trotz Empfehlung von *multimethod* Ansätzen Ind-Col nach wie vor fast ausschließlich über Fragen zu konstruktrelevanten Einstellungen und Werten erfasst werden. Die Autoren fanden in ihrer Meta-Analyse hierzu 27 verschiedene Skalen zur Erhebung von Individualismus bzw. Kollektivismus. Dabei heben sie hervor, dass viele Skalen Modifikation vorhandener Skalen darstellen, die lediglich durch einzelne neue Items ergänzt wurden. Elf dieser Skalen erfassen Individualismus und Kollektivismus als bipolares Konstrukt, während 16 der Skalen Individualismus (oder *independence* oder *idiocentrism*) und Kollektivismus (oder *interdependence* oder *allocentrism*) als unabhängige Konstrukte erheben.

Angesichts der mehrdeutigen Definition von Individualismus und Kollektivismus ist es nicht verwunderlich, dass auch die verwendeten Erhebungsinstrumente in

Bezug auf die inhaltlich erfassten Bereiche eine starke Heterogenität aufweisen. Aufgrund dessen führten Oyserman et al. (Oyserman et al., 2002) eine Inhaltsanalyse aller Items durch, um die tatsächlich mit den Konstrukten assoziierten Inhalte zu ergründen. Hierbei fanden sie sechs inhaltliche Bereiche für Individualismus und acht Inhaltsbereiche für Kollektivismus (siehe Tabelle 3).

Tabelle 3. Inhaltliche Bereiche für Individualismus und Kollektivismus (Oyserman et al., 2002, S. 9)

Inhaltsbereich	Beschreibung
Individualismus	
<i>Unabhängigkeit</i>	Freiheit, Selbstgenügsamkeit, Kontrolle über das eigene Leben
<i>Ziele</i>	Streben nach den eigenen Zielen, Wünschen und Erfolgen
<i>Wettbewerb</i>	Persönlicher Wettkampf und Gewinn
<i>Einzigartigkeit</i>	Fokus auf die eigenen einzigartigen und idiosynkratischen Qualitäten
<i>Privatheit Selbstwissen</i>	Privatheit eigener Handlungen und Gedanken; Starke Identität
<i>Direkte Kommunikation</i>	Klarer Artikulation der eigenen Wünsche und Bedürfnisse
Kollektivismus	
<i>Verbundenheit</i>	Betrachtung von Nahestehenden als Bestandteil der eigenen Identität
<i>Zugehörigkeit</i>	Wunsch zu einer Gruppe zu gehören
<i>Pflicht</i>	Pflichten und Opfer, die mit einer Gruppenzugehörigkeit einhergehen
<i>Harmonie</i>	Gruppenharmonie als persönliches Anliegen
<i>Ratschlag</i>	Ratschlag von Nahestehenden in die eigene Entscheidung einbeziehen
<i>Kontext</i>	Anpassung an den Kontext oder die Situation
<i>Hierarchie</i>	Fokus auf Hierarchie und Status
<i>Gruppe</i>	Bevorzugung von Gruppenarbeit

Zu den von Oyserman et al. (2002) identifizierten inhaltlichen Bereichen lassen sich 88% der Items aller untersuchten Erhebungsinstrumente zuordnen und diese spiegeln die von verschiedenen Forschern mehr oder weniger einheitlich verwendete Operationalisierung von Ind-Col wider. Dabei beinhalten in 83% der Skalen zu Individualismus mindestens ein Item zu „Wertschätzung persönlicher Unabhängigkeit“ und 85% der Skalen zu Kollektivismus mindestens ein Item zu „Pflichtgefühl gegenüber der Gruppe“. Demnach können die inhaltlichen Bereiche „Unabhängigkeit“ und „Pflicht“ als Hauptkomponenten von Ind-Col betrachtet werden.

2.4.3 Die drei gebräuchlichsten Erhebungsinstrumente für Ind-Col

Innerhalb der Meta-Analyse von Oyserman et al. (2002) bei der 83 Studien zu Individualismus und Kollektivismus berücksichtigt wurden, wurden folgende Erhebungsinstrumente am häufigsten eingesetzt (Shulruf, Hattie & Dixon, 2003): 19 Untersuchungen verwendeten die (1) *Independent-Interdependent (SCS) Scale* von Singelis (1994), 16 Untersuchungen die (2) *Horizontal-Vertical Collectivism-Individualism scale* von Singelis, Triandis, Bhawuk & Gelfand (1995) und 10 Untersuchungen die (3) *Individualism-Collectivism (INDCOL) Scale* von Hui (1988). Aufgrund der relativen Wichtigkeit werden diese Erhebungsinstrumente im Folgenden näher vorgestellt:

(1) *Self-Construal Scale (SCS)*. Die SCS (Singelis, 1994) besteht aus insgesamt 24 Items, die Überarbeitungen bereits bestehender Erhebungsinstrumente (vgl. Bhawuk & Brislin, 1992; Cross & Markus, 1991; Hui, 1988; Yamaguchi, 1994) darstellen. Jeweils 12 Items erfassen *independent self* (z. B. „*I enjoy being unique and different from others in many respects*“) und *interdependent self* (z. B. „*It is important to me to respect decisions made by the group*“) als zwei unabhängige Konstrukte. Als Antwortformat dient eine siebenstufige Likert-Skala von *strongly disagree* bis *strongly agree*. Die Reliabilität kann mit $\alpha = .69$ (*independent items*) und $\alpha = .73$ (*interdependent items*) als akzeptabel beurteilt werden. Konstruktvalidität wurde mittels Vergleich von Amerikanern asiatischer und kaukasischer Herkunft hinreichend nachgewiesen.

(2) *Horizontal-Vertical Collectivism-Individualism scale*. Singelis et al. (1995) konstruierten ein Erhebungsinstrument für die Erfassung von Ind-Col in Zusammenhang mit den Attributen vertikal bzw. horizontal. Diese Kombination ergibt eine Typologie von vier distinkte Muster kultureller Orientierung: Horizontaler Individualismus (HI), vertikaler Individualismus (VI), horizontaler Kollektivismus (HC) und vertikaler Kollektivismus (VC). Für die Erfassung dieser Typologie analysierten die Autoren einen Itempool bestehend aus neu entwickelten und bereits vorhandenen Items (vgl. Triandis et al., 1995). Nach faktorenanalytischen und testtheoretischen Gütekriterien (Kommunalitäten, Trennschärfe etc.) selektierten sie insgesamt 32 Items, von denen jeweils acht Items eines der vier kulturellen Orientierungen zugeordnet wurden. Tabelle 4 listet pro Dimension ein Beispielitem auf. Als Antwortformat diente eine neustufige Skala mit den Polen *never* oder *definitely* und *always* oder *definitely yes*.

Tabelle 4. Beispielitems aus der *Horizontal-Vertical Collectivism-Individualism scale*

Dimension	Beispielitem
Horizontaler Individualismus	<i>I enjoy being unique and different from others in many ways.</i>
Vertikaler Individualismus	<i>It annoys me when other people perform better than I do.</i>
Horizontaler Kollektivismus	<i>The well-being of my co-workers is important to me.</i>
Vertikaler Kollektivismus	<i>I would do what would please my family, even if I detested that activity.</i>

Die Reliabilität der Skalen kann mit $\alpha = .67$ (horizontaler Individualismus), $\alpha = .74$ (vertikaler Individualismus), $\alpha = .74$ (horizontaler Kollektivismus) und $\alpha = .68$ (vertikaler Kollektivismus) durchweg als akzeptabel beurteilt werden. Anhand Korrelationen mit der *Self-Construal Scale* (Singelis, 1994) und der *Sinha scale* (Sinha & Verma, 1994) wurde die konvergente Validität untersucht. Diese zeigte zwar keinen konsistenten jedoch hinreichenden Nachweis für Konstruktvalidität.

(3) *Individualism-Collectivism (INDCOL) Scale*. Die *INDCOL Scale* (Hui, 1988) erfasst Ind-Col als eindimensionales Konstrukt und besteht aus Items, die auf Basis der Untersuchung von Hui und Triandis (1986) entwickelt wurden. Zusätzlich ergänzten modifizierte Items einer bereits bestehenden Skala von Muzynski (1984) den Itempool. Den Empfehlungen von Hui und Triandis (1986) entsprechend, wurden sechs Referenzgruppen bestimmt: Ehepartner, Eltern, Familie, Nachbar, Freunde und Arbeitskollegen. Als Antwortformat wurde eine sechsstufige Skala mit den Polen *strongly disagree* oder *false* und *strongly agree* oder *true* verwendet. Die Anzahl der Items pro Referenzgruppe inklusive eines Beispielitems sind in Tabelle 5 aufgelistet.

Tabelle 5. Beispielitems aus der *Individualism-Collectivism (INDCOL) Scale*

Referenzgruppe	Itemzahl	Beispielitem
Ehepartner	8	<i>The decision of where one is to work should be jointly made with one's spouse, if one is married.</i>
Eltern	16	<i>It is reasonable for a son to continue his father's business</i>
Familie	8	<i>I count on my relatives for help if I find myself in any kind of trouble.</i>
Nachbar	10	<i>My neighbours always tell me interesting stories that have happened around them.</i>
Freunde	9	<i>My good friends and I agree on the best places to shop.</i>
Arbeitskollegen	11	<i>Classmates' assistance is indispensable to getting a good grade at school.</i>

Die Reliabilität der einzelnen Subskalen belaufen sich zwischen $\alpha = .46$ (Ehepartner) bis $\alpha = .76$ (Eltern) und können als gering bis akzeptabel beurteilt werden. Konstruktvalidität wurde in zahlreichen Studien zur konvergenten, divergenten und prädiktive Validität hinreichend belegt (vgl. Hui, 1988).

2.4 Kritische Betrachtung

Ind-Col stellen zwei Konstrukte dar, die in der Lage sind interkulturelle Unterschiede im Erleben und Verhalten zu erklären und sich besonders durch ihre theoretische Sparsamkeit auszeichnen (Oyserman et al., 2002). Trotzdem darf nicht übersehen werden, dass auch vermehrt Kritik an den beiden Konstrukten geübt wird, die im Folgenden skizziert werden. Konkret können diese Kritikpunkte in folgende Punkte zusammengefasst werden: (1) Definition, (2) Stichprobe und (3) Erhebungsmethode.

(1) *Definition.* Einer der größten Kritikpunkte von Ind-Col betrifft die uneinheitliche Definition dieser Konstrukte (vgl. Kapitel 2.1). Dieses Problem ergibt sich aus der Tatsache, dass Ind-Col ursprünglich soziologische und nicht psychologische Dimensionen darstellten (Hofstede, 1994). Dementsprechend gibt es per se keinen psychologischen Individualismus oder Kollektivismus, sondern lediglich individualistische oder kollektivistische Verhaltensweisen (Vornov & Singer, 2002). Aus diesem Grund werden Ind-Col als polythetische Konstrukte betrachtet, welche anhand einer Reihe von Attributen definiert werden. Folglich werden normalerweise Items aus ausgewählten inhaltlichen Bereichen formuliert, die eine Skala zur Erfassung Individualismus bzw. Kollektivismus bilden sollen (vgl. Kapitel 2.4.2). Daraus ergeben sich drei Probleme: Die Behandlung eines polythetischen Konstruktes als eindimensionales Konstrukt führt (a) zur Verletzung der Inhaltsvalidität, denn wichtige Facetten des Konstruktes könnten unterrepräsentiert bzw. unwichtige Facetten überrepräsentiert sein. Weiters könnten konstruktirrelevante Facetten erfasst sein oder bestimmte Facetten eine unterschiedliche Gewichtung in den untersuchten Gruppe haben (Chen & West, 2008). Darüber hinaus weisen Erhebungsinstrumente, die Ind-Col als breites und diffuses Konstrukt mittels verschiedenen inhaltlichen Bereichen erfassen, (b) eine geringe Reliabilität (zumeist unter $\alpha = .5$) auf (Kagitcibasi, 1997). Dieses Problem ist auch unter der Bezeichnung *bandwidth vs. fidelity dilemma* (Cronbach, 1990) bekannt. Zuletzt kann die Operationalisierung von Ind-Col anhand der mit den Konstrukten assoziierten Verhaltensweisen (c) zu einer Tautologie führen

(Oyserman et al., 2002). Nicht selten wird schließlich der Zusammenhang zwischen bestimmten Inhalten und Ind-Col untersucht, ohne zu berücksichtigen, dass diese Inhalte bereits Teil der Operationalisierung von Ind-Col waren (Miller, 2002).

(2) *Stichprobe*. Ein allgemein verbreitetes Problem innerhalb der Psychologie stellt die mangelnde Repräsentativität der untersuchten Stichproben dar. Während dies bei manchen Untersuchungen ein vernachlässigbarer, zumindest begründbarer Umstand (z. B. innerhalb der Biologischen Psychologie) darstellt, schränkt dies innerhalb der kulturvergleichenden Forschung die Generalisierbarkeit von Forschungsergebnissen erheblich ein. Bereits die Untersuchung von Hofstede (1980) bezog sich ausschließlich auf Mitarbeiter eines internationalen Elektronik Konzerns, die aus hoch ausgebildeten Managern, Technikern und anderen Angestellten bestanden (Green, Deschamps & Paez, 2005; Vornov & Singer, 2002). Auch nachfolgende Forschungen beschränkten sich traditionell hauptsächlich auf die Population der Studierenden. Diese stellen jedoch keine verlässliche Repräsentation von Kulturen dar. Vielmehr kann angenommen werden, dass Studierende aufgrund ihrer höheren Bildung und ihres höheren sozioökonomischen Status andere Werte und Einstellungen vertreten als die restlichen Mitglieder einer Kultur (Fiske, 2002; Miller, 2002).

Problematisch ist ebenfalls die Operationalisierung von kultureller Orientierung anhand des nationalen Hintergrundes im Sinne von *Applying Hofstede* (vgl. Kapitel 2.3), denn dies setzt die Homogenität dieser Länder hinsichtlich aller geografischen Gebiete und in Bezug auf alle Lebensbereiche (Familie, Schule, Arbeit usw.) voraus. Besonders in multikulturellen Ländern ist diese Annahme jedoch höchst unwahrscheinlich (Cirkov et al., 2005).

(3) *Erhebungsmethode*. Neben der mangelnden Repräsentativität der untersuchten Stichproben stellt eine weitere Tradition der psychologischen Forschung die Verwendung von Fragebögen als bevorzugte Methode der quantitativen Datenerhebung dar. Auch die Erhebung von Ind-Col erfolgt mittels Fragen zu Einstellungen und Werten verschiedener inhaltlicher Bereiche. Innerhalb der kulturvergleichenden Forschung können dabei drei Problembereiche identifiziert werden: Fragebögen durchlaufen für den Einsatz in verschiedenen Kulturen einen Prozess der Übersetzung und anschließender Rückübersetzung (siehe Rippl & Seipl, 2008). Dies stellt jedoch (a) keine Garantie für die Äquivalenz der verschiedenen Sprachversionen dar (Li & Aksoy, 2006). Vielmehr setzt die Beantwortung von Fragen stets die Interpretation derselben durch die befragten

Personen voraus (Schwartz & Oyserman, 2001). Da diese Interpretation von verschiedenen kulturbedingten Aspekten abhängig ist, kann nicht davon ausgegangen werden, dass in den jeweiligen untersuchten Kulturen dasselbe erfasst wird (siehe Uskul & Oyserman, 2006). Weiters können (b) Antworttendenzen in Abhängigkeit der untersuchten Kultur unterschiedlich stark zum Tragen kommen (Schimmack et al., 2005). Dadurch werden neben der untersuchten Eigenschaft einer Person Unterschiede in der kulturellen Ausprägung von bestimmten Antworttendenzen, wie etwa Tendenz zur Mitte bzw. zum extremen Urteil (siehe Jankisz & Moosbrugger, 2007), erfasst. Zuletzt stellt sich (c) die Frage, inwieweit kulturelle Werte bewusst zugänglich sind, sodass Personen innerhalb einer Fragebogenerhebungen davon berichten können (Bond, 2002). Fiske (2002) betont, dass Kultur kein deklaratives, semantisches oder episodisches Wissen darstellt, sondern aus diversen, lose verbundenen und nicht artikulierbaren Bestandteilen besteht. Zudem werden kulturelle Aspekte von Kognition, Emotion oder Motivation spontan im Zuge der Interaktion mit der Umwelt aktiviert, sodass Kitayama (2002) meint „Culture is much more than what people indicate in response to attitudinal survey questions“ (S. 95).

Zusammenfassend kann gesagt werden, dass der zukünftige, wissenschaftliche Wert von Ind-Col für die Zukunft abhängig ist von: (1) einer engeren und einheitlichen Definition, welche ausschließlich Kernelementen dieser Konstrukte beinhaltet, (2) einer Stichprobenziehung, die die Kriterien der Repräsentativität erfüllen und (3) bestimmten Standards zur Sicherstellung der Äquivalenz (v. a. die Durchführung einer stringenten psychometrischen Äquivalenzprüfung) von Erhebungsinstrumenten, die den Einsatz in kulturvergleichenden Untersuchungen erst rechtfertigen.

3 Methodische Aspekte der kulturvergleichenden Forschung

Die nachfolgenden Ausführungen beziehen sich in erster Linie auf Van de Vijver, der mit *Methods and Data Analysis for Cross-Cultural Research* (Van de Vijver & Leung, 1997) eine systematische Darstellung der Methodologie kulturvergleichender Studien vorlegte. An dieser Stelle wird auch auf das Buchkapitel von Cohen (2007) verwiesen, der sich mit Fragen der Kausalität, der Operationalisierung, der Stichprobenziehung und der Interpretation innerhalb kulturvergleichender Studien auseinandersetzt.

3.1 Taxonomie kulturvergleichender Studien

Kulturvergleichende Studien können anhand dreier Merkmalen eingeteilt werden: (1) Berücksichtigung von kontextuellen Faktoren, (2) Betonung auf Exploration oder Hypothesentestung und (2) Level- oder Strukturorientierung (siehe Tabelle 6).

Tabelle 6. Taxonomie kulturvergleichender Studien (Van de Vijver, 2007, S. 342)

Berücksichtigung kontextueller Faktoren	Betonung liegt auf...	
	<i>Exploration</i>	<i>Hypothesentestung</i>
<i>Struktur-orientiert</i>		
<i>Nein</i>	Struktur-orientierte Studie zu psychologischen Unterschieden	Struktur-orientierte Generalisierbarkeitsstudie
<i>Ja</i>	Struktur-orientierte ökologische Linkage-Studie	Struktur-orientierte kontexttheoretische Studie
<i>Level-orientiert</i>		
<i>Nein</i>	Level-orientierte Studie zu psychologischen Unterschieden	Level-orientierte Generalisierbarkeitsstudie
<i>Ja</i>	Level-orientierte ökologische Linkage-Studie	Level-orientierte kontexttheoretische Studie

(1) Kontextuelle Faktoren, die im Design einer Studie Berücksichtigung finden können, beziehen sich auf jegliche Variablen, die interkulturelle Unterschiede beeinflussen können. Damit sind sowohl subjektgebundene Merkmale (z. B. Bildungsstand) sowie kulturelle Merkmale (z. B. das Bruttosozialprodukt eines Landes) angesprochen. Als einer der populärsten Kontextvariablen sei die Unterscheidung zwischen individualistischen und kollektivistischen Ländern genannt.

(2) Explorative Studien stehen am Anfang von kulturvergleichenden Forschungsprozessen, wenn noch keine Theorien über die zu erwartenden Kulturunterschiede oder die Anwendbarkeit von Erhebungsinstrumenten vorhanden sind. Nach der Entwicklung von Theorien bzw. Modellen können diese in hypothesentestenden Studien auf ihre Richtigkeit geprüft werden.

(3) Forschungen der kulturvergleichenden Psychologie können zum einen strukturelle Unterschiede von Konzepten thematisieren und zum anderen Unterschiede in den Ausprägungen von Variablen untersuchen.

Aus der Kombination dieser drei Merkmale können acht unterschiedliche Arten von Studien abgeleitet werden, die die Forschungsfragen und das Design kulturvergleichender Studien bestimmen.

3.2 Bias und Äquivalenz

Bias und Äquivalenz stellen zwei wesentliche Konzepte dar, die sich auf unterschiedliche Aspekte kulturvergleichender Forschung beziehen. Dabei bezeichnet der Begriff Bias alle systematischen Fehlerquellen, die innerhalb des Untersuchungsprozesses auftreten können. Der Begriff Äquivalenz bezieht sich hingegen auf die Auswirkungen von Bias auf die Messung und spricht daher deren Messniveau an. Beide Konzepte kommen bei der Anwendung eines Messinstrumentes in verschiedenen kulturellen Gruppen zum Tragen, wodurch sie die Vergleichbarkeit von (Mess-)Ergebnissen über Kulturen hinweg grundlegend bestimmen.

3.2.1 Arten von Bias

In der kulturvergleichenden Forschung werden drei Arten von Bias unterschieden: (1) Konstrukt-Bias, (2) Methoden-Bias und (3) Item-Bias.

(1) Der *Konstrukt-Bias* tritt dann auf, wenn das gemessene Konstrukt in den verschiedenen Kulturen nicht identisch ist. Dies kann dann der Fall sein, wenn die Definition eines Konstrukts zwischen den kulturellen Gruppen nicht übereinstimmt, folglich unterschiedliche Verhaltensweisen mit dem Konstrukt assoziiert sind. Beispielsweise spielen in nicht westlichen Ländern die sozialen Aspekte der Intelligenz eine größere Rolle in der Alltagsdefinition von Intelligenz als in westlichen Ländern, wo Intelligenz in erster Linie mit logischem Denkvermögen und einem großen Wissen in Verbindung gebracht wird. Als eine weitere Ursache für einen Konstrukt-Bias kann die inadäquate Auswahl, das heißt eine unzureichende bzw. unvollständige Erfassung von konstruktrelevanten Verhaltensweisen, eines zu erfassenden breiten Konstruktes gesehen werden.

(2) Der *Methoden-Bias* bezieht sich auf alle methodischen Aspekte einer Untersuchung, die die Validität der Ergebnisse einschränken können. Dazu zählt (1) die Nichtvergleichbarkeit der Stichproben aus verschiedenen Kulturen (*Sample-Bias*), die beispielsweise aufgrund des differenten Bildungsstandes oder der unterschiedlichen Motivation bei der Bearbeitung eines Leistungstests

verursacht werden. Außerdem umfasst der Methoden-Bias (2) bestimmte Merkmale des Erhebungsinstruments (Instrumenten-Bias), die unterschiedlicher Antworttendenzen hervorrufen können oder für bestimmte Gruppen ungleich vertraut sein können (z. B. Antwortformat oder Stimulusmaterial) und (3) Probleme bei der Durchführung der Studie (Durchführungs-Bias), wie etwa Kommunikationsprobleme zwischen Interviewer und Interviewten.

(3) Der *Item-Bias* (differentielle Itemfunktion, DIF) kommt dann zum Tragen, wenn Personen aus unterschiedlichen kulturellen Gruppen mit gleicher Ausprägung auf einem latenten Konstrukt nicht den selben Durchschnittswert auf einem Item aufweisen. Derartige Verzerrungen von Items resultieren für gewöhnlich aus inadäquaten Übersetzungen oder Formulierungen von Items, die angesichts abweichender Konnotation von Wortlauten zustande kommen können.

3.2.2 Umgang mit Bias

Eine häufig angewandte, jedoch keinesfalls vertretbare Methode im Umgang mit Bias ist das Ignorieren desselben; die entgegengesetzte Strategie wäre die Unterlassung des Kulturvergleichs bei der Feststellung eines Bias. Von beiden Extrempositionen ist abzusehen, um sowohl bei der Planung die Reduktion von Bias im Vorfeld zu bedenken als auch das Auftreten von Bias im Rahmen der statistischen Analysen zu berücksichtigen. Darüber hinaus kann das Auftreten von Bias als Hinweis für kulturelle Unterschiede angesehen werden, wonach Bias für kulturspezifische Aspekte (*emic*) steht, während Messinstrumente bzw. einzelne Items ohne Bias als universelle (*etic*) Aspekte gesehen werden können.

Zwei Methoden zur Reduktion von Konstrukt-Bias sind das kulturelle Dezentrieren und der Konvergenzansatz. Ersteres meint die simultane Entwicklung desselben Erhebungsinstruments in mehreren Kulturen zur nachfolgenden Analyse von kulturgebundenen Wörtern und Konzepten, während letzteres die unabhängige Entwicklung von Instrumenten in mehrere Kulturen zur nachfolgenden Vorgabe aller (übersetzten) Instrumente in allen kulturellen Gruppen vorsieht. Weitere Strategien, die sich mit der Reduktion von Konstrukt-Bias beschäftigen sind eng mit der Berücksichtigung des Methoden-Bias geknüpft. Dazu zählt das Hinzuziehen von lokalen Informanten mit Kenntnissen über die jeweilige Kultur bzw. der Sprache, zum Beispiel zur Beurteilung des Erhebungsinstruments, oder das Einbeziehen von zweisprachigen Personen als Studienteilnehmer. Ferner kann die

Durchführung nicht-standardisierter lokaler Pilotstudien Aufschluss über die Angemessenheit von Instrumenten geben, beispielsweise wenn man Probanden sowohl Instruktion als auch Stimuli beurteilen lässt oder versucht, mittels der Methode des lauten Denkens einen Einblick in die kognitiven Prozesse zu gewinnen. Eine weitere Möglichkeit zur Reduktion von Konstrukt- und Methoden-Bias bietet die kulturvergleichende Analyse von nomologischen Netzwerken (Monotrait-Multimethod-Analyse oder Analyse der konvergenten und diskriminanten Validität).

Zur Berücksichtigung des Methoden-Bias ist eine umfassende Einschulung von Testleitern mit Spezifizierung von Testhandbüchern zur Standardisierung der Durchführungsbedingung unerlässlich. Durchführungsprotokolle ergänzen den Prozess der Datenerhebung, in denen alle relevanten Zusatzinformationen (z. B. Verhalten der Teilnehmer während der Erhebung) dokumentiert werden. Auch auf die Erhebung potentiell relevanter Subjekt- und Kontextvariablen (z. B. soziale Erwünschtheit oder Bildungsstand) ist zu achten. Diese können bei der statistischen Analyse, z. B. mit Hilfe der Kovarianzanalyse, Berücksichtigung finden. Zuletzt bietet die Durchführung von Test-Retest-Studien, Trainingsstudien bzw. Interventionsstudien Aufschluss über das Vorhandensein eines Methoden-Bias. Ein Bias auf Item-Ebene wird für gewöhnlich mittels beurteilenden und psychometrischen Methoden untersucht. Bei beurteilenden Methoden erfolgt eine linguistische und/oder psychologische Expertenbeurteilung des Stimulusmaterials, wohingegen sich psychometrische Methoden mannigfaltiger statistischer Verfahren zur Aufdeckung von differenziellen Itemfunktionen bedienen.

Der Bias in kulturvergleichenden Studien wird vorwiegend auf der Item-Ebene behandelt ohne den Konstrukt- und Methoden-Bias als weitere Quellen für Verzerrungen zu berücksichtigen. Dies führt zwangsläufig zur Herabsetzung der Äquivalenzebene und letztlich zu eingeschränkter Validität von Untersuchungsergebnissen. Um dies zu verhindern müssen bei der Planung, Durchführung, Auswertung und Interpretation kulturvergleichender Untersuchungen alle drei Arten von Bias berücksichtigt werden.

3.2.3 Ebenen der Äquivalenz

Eine zentrale Frage innerhalb der kulturvergleichenden Forschung stellt jene der Äquivalenz dar, welche die interkulturelle Vergleichbarkeit der Daten anspricht. Dabei werden drei Ebenen der Äquivalenz unterschieden: (1) Strukturelle

Äquivalenz, (2) Messeinheitsäquivalenz und (3) Skalenäquivalenz. Diese Ebenen bauen hierarchisch aufeinander auf, sodass die Skalenäquivalenz als höchste Äquivalenzebene nur erreicht werden kann, wenn die Messeinheitsäquivalenz gegeben ist, welche wiederum strukturelle Äquivalenz voraussetzt.

(1) *Strukturelle Äquivalenz* ist gegeben, wenn in verschiedenen kulturellen Gruppen dasselbe Konstrukt erfasst wird.

(2) *Messeinheitsäquivalenz* wird unter der Voraussetzung der strukturellen Äquivalenz erfüllt, wenn dieselben Messeinheiten eines Erhebungsinstrumentes in verschiedenen Kulturen vorliegen. Dabei dürfen sie einen unterschiedlichen Ursprung (Nullpunkt) der Skala besitzen.

(3) *Skalenäquivalenz* liegt unter der Voraussetzung der *Messeinheitsäquivalenz* vor, wenn das Erhebungsinstrument in jeder kulturellen Gruppe auf derselben Rationalskala liegt, d. h. die Skala besitzt denselben Ursprung (Nullpunkt).

3.3 Methoden zur psychometrischen Äquivalenzprüfung

Im folgenden Kapitel werden zwei methodische Ansätze zur Prüfung der Äquivalenz dargestellt, die auf der probabilistischen und der klassischen Testtheorie basieren. Aus einer Vielzahl in der Literatur vorgeschlagenen Verfahren (siehe Van de Vijver & Leung, 1997) werden einerseits das *Rasch-Modell* aufgrund der methodischen Stringenz und andererseits die konfirmatorische Mehrgruppen-Faktorenanalyse als *state of the art* Vorgehensweise für die statistische Analysen zur Äquivalenzprüfung in der vorliegenden Arbeit ausgewählt.

3.3.1 Das Rasch-Modell

Die probabilistische Testtheorie basiert auf der Annahme einer nicht direkt beobachtbaren latenten Eigenschaftsdimension, die mit der Berücksichtigung einer stochastischen Beziehung zwischen dieser latenten Variable und des manifesten Antwortverhaltens verbunden ist (Fischer, 1974).

Das dichotom-logistische Testmodell von Rasch (1960, kurz „dichotome Rasch-Modell“) stellt das zentrale Modell der probabilistischen Testtheorie dar. Aus diesem Modell lassen sich eine Reihe von weiteren Testmodelle entweder als Spezialfälle oder als Verallgemeinerungen ableiten (Kubinger, 1999). Diese Testmodelle werden unter dem Stichwort *Item-Response-Theory* (IRT) zusammengefasst (Bortz & Döring, 2006). Einen Überblick über die verschiedenen

Modelle gibt Rost (2004). Die folgenden Ausführungen beschränken sich auf das (dichotome) *Rasch-Modell* (Rasch, 1960) und auf das ordinale *Rasch-Modell*, das auch unter der Bezeichnung *Partial Credit Model* bekannt ist (Masters, 1982).

Das *Rasch-Modell* formuliert eine logistische Wahrscheinlichkeitsfunktion, in Abhängigkeit einer latenten Eigenschaftsdimension ξ (Personenparameter) und der Schwierigkeit² eines Items σ (Itemparameter):

$$P(+ | \xi_v, \sigma_i) = \frac{e^{\xi_v - \sigma_i}}{1 + e^{\xi_v - \sigma_i}}$$

Formel 1. Formel für das dichotome *Rasch-Modell*

Dieser Formel lässt sich entnehmen, dass das Lösen („+“) eines Items i von der Person v neben einer wahrscheinlichkeitstheoretisch modellierten Zufallskomponente lediglich von der Eigenschaftsausprägung ξ_v ($-\infty \leq \xi_v \leq \infty$) und der Itemschwierigkeit σ_i ($-\infty \leq \sigma_i \leq \infty$) abhängt (siehe Kubinger, 2004).

Der Zusammenhang zwischen der Eigenschaftsausprägung ξ_v , der Itemschwierigkeit σ_i und der Lösungswahrscheinlichkeit p lässt sich mittels „Item-Charakteristikkurve“ (ICC) grafisch veranschaulichen. Die Itemschwierigkeit σ_i ist definiert als jene Eigenschaftsausprägung ξ_v , bei der die Lösungswahrscheinlichkeit für dieses Item genau $\frac{1}{2}$ ($p = 0.5$) beträgt. Dies erlaubt sowohl Personenparameter ξ_v (Ausprägung der Eigenschaft der Person v) als auch Itemparameter σ_i (Schwierigkeitsausprägung des Items i) auf derselben Skala aufzutragen (*joint scale*) und der Lösungswahrscheinlichkeit p gegenüberzustellen. Abbildung 1 zeigt IC-Kurven für drei *Rasch-Modell* konforme Items mit unterschiedlichen Itemparametern.

² Im Kontext der Einstellungsmessung entspricht die „Schwierigkeit“ eines Items dem Grad der „Herausforderung“ ein bestimmtes Verhalten zu zeigen und beeinflusst die Zustimmungswahrscheinlichkeit (p) auf einen Iteminhalt.

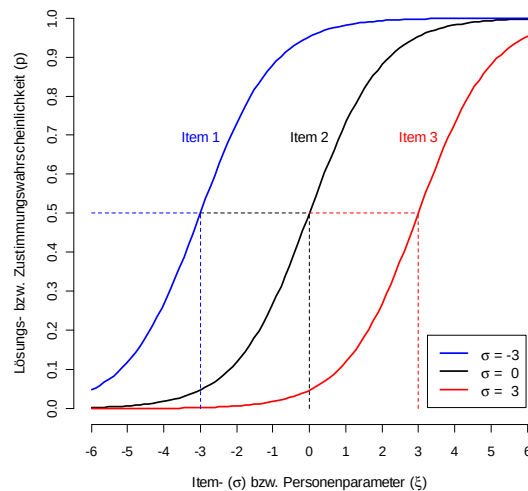


Abbildung 1. Itemcharakteristikkurven für drei Rasch-Modell konforme Items

Der Abbildung lässt sich entnehmen, dass alle Items den gleichen Verlauf der Lösungswahrscheinlichkeiten zeigen. Dies spiegelt eine weitere Modelleigenschaft wider, nämlich jene der gleichen Trennschärfen aller Items.

Eine wesentliche Voraussetzung des *Rasch-Modells* stellt jene der „lokalen stochastischen Unabhängigkeit“ dar. Damit ist gemeint, dass das Antwortverhalten einer Person ausschließlich von der latenten Eigenschaftsdimension abhängt. Unter Erfüllung dieser Voraussetzung stellt die Anzahl der gelösten Aufgaben eine „erschöpfende Statistik“ dar. Das bedeutet, die Anzahl der gelösten Items beinhaltet die gesamte Information hinsichtlich der fraglichen Eigenschaftsdimension einer Person, sodass kein Interesse daran besteht zu berücksichtigen, welche Items die Person gelöst bzw. nicht gelöst hat (Fischer, 1974).

Die Gültigkeit des *Rasch-Modells* impliziert eine eindimensionale Messung einer Eigenschaft, wobei bei Eindimensionalität im Sinne des *Rasch-Modells* durchaus verschiedene psychologische Prozesse involviert sein können, solange bei der Bearbeitung eines Items immer dieselben Prozesse in gleicher Gewichtung beteiligt sind (Bühner, 2006).

Eine besondere Eigenschaft des *Rasch-Modells* stellt jenes der „spezifischen Objektivität“ dar. Diese besagt einerseits, dass der Vergleich zweier Personen hinsichtlich der interessierenden Eigenschaft von den bearbeiteten Items eines Itempools unabhängig ist (Itemhomogenität). Andererseits ist der Vergleich zweier Items bezüglich deren Schwierigkeiten von der verwendeten Stichprobe unabhängig (Personenhomogenität). In diesem Zusammenhang wird auch von der Separierbarkeit der Parameter (Item- und Personenparameter) gesprochen

(Kubinger, 1999). Die Eigenschaft spezifisch objektive Vergleiche zu ermöglichen wird oft mit der Bezeichnung „Stichprobenunabhängigkeit“ in Verbindung gebracht (Fischer, 1999) und dient als Grundlage zur Prüfung auf Rasch-Modell-Konformität in einer Population. Anhand unterschiedlicher Modelltests kann einerseits die Personenhomogenität (z. B. mittels Likelihood-Ratio-Test nach Andersen, 1973) andererseits die Itemhomogenität (z. B. mittels Martin-Löf-Test nach Martin-Loef, 1973) geprüft werden. Aufgrund des „Falsifikationsprinzips“ in der Erkenntnislogik kann jedoch die Geltung des Rasch-Modells für einen Itempool nicht bewiesen, sondern nur hinsichtlich der „Grad der Bewährung“ sensu Popper (vgl. z. B. Popper, 1976) geprüft werden. Erst nach ausreichendem Grad an Bewährung wird die Geltung des Rasch-Modells für einen bestimmten Itempool als gegeben erachtet (Kubinger & Draxler, 2007).

Das ordinale Rasch-Modell (Masters, 1982) stellt eine Verallgemeinerung des dichotomen Rasch-Modells für ordinale Daten dar und ist das allgemeinste Modell für Ratingskalen (siehe Wright & Masters, 1982). In diesem Modell wird die Wahrscheinlichkeit, dass eine Person v mit der Eigenschaftsausprägung ξ_v bei Item i mit der Schwierigkeit σ_i die Antwortkategorie h erreicht durch einen kombinierten Itemkategorienparameter $\sigma_i^{(h)}$, $h = 1, 2, \dots, (m_i-1)$ bestimmt (siehe Kubinger, 1999):

$$P(h | \xi_v, (\sigma_i^{(1)}, \sigma_i^{(2)}, \dots, \sigma_i^{(m_i-1)})) = \frac{e^{\sum_{j=1}^{h-1} (\xi_v - \sigma_i^{(j)})}}{\sum_{g=1}^{m_i} e^{\sum_{j=1}^{g-1} (\xi_v - \sigma_i^{(j)})}}$$

Formel 2. Formel für das ordinalen Rasch-Modell

Folgt ein Itempool dem ordinalen Rasch-Modell, dann weist die summierte Anzahl erfolgter Antwortkategorienstritte ebenfalls die Modelleigenschaft der erschöpfenden Statistik sowie der spezifischen Objektivität auf.

Die Schätzung der Parameter erfolgt unter der Bedingung der spezifischen Objektivität mittels *Conditional Maximum Likelihood* (CML). Hierzu stehen Programme wie z. B. LPCM-WIN (Fischer & Ponocny-Seliger, 1998), WINMIRA (von Davier, 2001) oder das im „R“ (R Developmental Core Team, 2008) implementierte Programmpaket eRm (Mair & Hatzinger, 2007) zur Verfügung. In diesem Zusammenhang sei angemerkt, dass die Parameter in anderen Programmen, beispielsweise in ConQuest (Wu, Adams, Wilson & Haldane, 2007) anhand der *Marginal Maximum Likelihood* (MML) geschätzt werden, wodurch jedoch die

Modelleigenschaft der spezifischen Objektivität und damit die Prüfbarkeit des Modells verloren geht.

3.3.2 Konfirmatorische (Mehrgruppen-)Faktorenanalyse

Die konfirmatorische Faktorenanalyse (CFA, Jöreskog, 1969) wurde als Erweiterung der Maximum-Likelihood-Faktorenanalyse (Jöreskog, 1963) vorgestellt (Rietz, Rudinger & Andres, 1996) und dient der Beurteilung der Konsistenz eines theoretisch fundierten Messmodells mit den empirischen Daten (Homburg, Klarmann & Pfesser, 2008). Der allgemeine Ablauf einer konfirmatorischen Faktorenanalyse lässt sich nach Moosbrugger und Schermelleh-Engel (2007) in folgende Schritte unterteilen: (1) Modellspezifikation (2) Parameterschätzung und (3) Modellevaluation.

(1) *Modellspezifikation.* Im ersten Schritt wird theoriegeleitet ein Messmodell festgelegt, das sowohl die Anzahl latenter Variablen (Faktoren) als auch die konkrete Zuordnung manifester Variablen (Items) als Indikatoren latenter Variablen spezifiziert (Homburg et al., 2008). Das Messmodell kann unter der Voraussetzung zentrierter bzw. z-standardisierter Werte formal als folgende Regressionsgleichung³ dargestellt werden:

$$x = \Lambda \cdot \xi + \delta$$

Formel 3. Regressionsgleichung der konfirmatorischen Faktorenanalyse

Die vorliegende Gleichung stellt ein reflektives Messmodell als Regelfall dar, welches annimmt, dass die Veränderungen in den Ausprägungen der manifesten Variablen (x) durch die latente Variable (ξ), gewichtet mit den Faktorladungen (Λ), kausal verursacht werden und einen Messfehler (δ) aufweisen (Backhaus, Erichson, Plinke & Weiber, 2008). In einem Pfaddiagramm kann die Struktur grafisch veranschaulicht werden, wobei latente Variablen sowie Fehlervariablen durch Kreise und manifeste Variablen durch Rechtecke symbolisiert werden. Zudem stellen einfache Pfeile eine gerichtete Beziehung (partielle Regressionsgewichte) und doppelte Pfeile einen ungerichteten Zusammenhang (Kovarianz bzw. Korrelation) dar.

(2) *Parameterschätzung.* Die eindeutige Schätzung der Modellparameter erfordert die Identifizierbarkeit des Modells (Kline, 2005). Dies setzt einerseits voraus, dass mindestens genau so viele empirisch bekannte wie zu schätzende Parameter

³ Die Notation erfolgt nach LISREL (Jöreskog & Sörbom, 1996)

(Faktorladungen, Faktorvarianzen und Fehlervarianzen) vorhanden sind und andererseits, dass die latenten Variablen sowie die Fehlervarianzen eine Metrik aufweisen (Moosbrugger & Schermelleh-Engel, 2007). Um Probleme bei der Identifikation zu vermeiden empfiehlt Bühner (2006) mindestens drei Indikatoren pro latenter Variable zu spezifizieren. Die Metrik einer latenten Variable kann entweder dadurch festgelegt werden, indem die Ladung einer manifesten Variable auf den Wert Eins fixiert wird, um für jede latente Variable eine Referenzvariable zu bestimmen, oder indem die Varianz der latenten Variablen auf den Wert Eins fixiert wird.

Als Grundlage der Parameterschätzung wird im Allgemeinen die empirische Kovarianzmatrix herangezogen. Zur Schätzung der Modellparameter stehen unterschiedliche Schätzmethoden zur Verfügung von denen die Maximum-Likelihood-Methode am häufigsten verwendet wird (Bühner, 2006). Diese setzt voraus, dass die manifesten Variablen intervallskaliert und multivariat normalverteilt sind. Alle Schätzmethoden haben das Ziel, die unbekannten Parameter des Modells so zu schätzen, dass die vom Modell theoretisch implizierte Kovarianz (implizierte Kovarianzmatrix) mit den empirisch beobachteten Kovarianzen (empirische Kovarianzmatrix) möglichst gut übereinstimmen (vgl. Bollen, 1989).

(3) *Modellevaluation*. Zur Beurteilung der Schätzergebnisse bzw. Anpassungsgüte eines Modells (*Model Fit*) stehen einerseits globale Anpassungsmaße und lokale Anpassungsmaße zur Verfügung (Homburg et al., 2008).

Die globale Beurteilung eines Modells erfolgt inferenzstatistisch mit Hilfe eines χ^2 -Modelltests. Da der χ^2 -Test in Abhängigkeit der Stichprobengröße und dem Ausmaß der Verletzung der Voraussetzung der multivariaten Normalverteilung häufig zur Ablehnung eines passenden Modells führt (Bühner, 2006), müssen stets auch deskriptive Kriterien (sog. Fit-Indizes) Berücksichtigung finden (La Du & Tanaka, 1989; Marsh, Balla & McDonald, 1988). Die Auswahl deskriptive Kriterien und die Grenzen der Modellakzeptanz bei der Beurteilung dieser Kriterien basieren im Wesentlichen auf Erfahrungswerten und sind somit arbiträr (Cheung & Rensvold, 2002; Hu & Bentler, 1998, 1999; Schermelleh-Engel, Moosbrugger & Müller, 2003). Nach Beauducel und Wittmann (2005) empfiehlt es sich neben dem χ^2 -Wert samt dazugehörigen p-Wert folgende Fit-Indizes zur Beurteilung eines Modells heranzuziehen: *Comparative-Fit-Index* (CFI; Bentler, 1990), *Root-Mean-Square-Error*

of Approximation (RMSEA; Steiger, 1990) und *Standardized-Root-Mean-Residual* (SRMR; Bentler, 1995).

Lokale Anpassungsmaße beziehen sich auf einzelne manifeste oder latente Variablen des Messmodells. Dazu gehören die Indikatorreliabilität, die Faktorreliabilität und die durchschnittlich erfasste Varianz eines Faktors (Homburg et al., 2008).

Die konfirmatorische Mehrgruppen-Faktorenanalyse (MCFA, Jöreskog, 1971) bietet als Weiterentwicklung der CFA die Möglichkeit der simultanen Analyse von modellspezifischen Parametern in mehreren voneinander unabhängigen Gruppen. Anhand stichprobenübergreifender Beschränkungen bei der Parameterschätzung können Gruppenunterschiede hinsichtlich bestimmter Parametern untersucht werden. Die Parameterrestriktion führt dazu, dass die Schätzung dieser Parameter nicht mehr an die einzelnen Teilstichproben angepasst, sondern über die Gruppen hinweg gleichgesetzt werden. Dadurch verschlechtert sich zwangsläufig die Modellanpassung (*Model Fit*) an die empirischen Daten. Das Ausmaß dieser Verschlechterung gibt Auskunft darüber, inwieweit die einzelnen Parameter über die Stichproben hinweg als gleich angenommen werden dürfen. Wenn sich das Modell mit Beschränkungen im Vergleich zum Modell ohne Beschränkungen nicht signifikant verschlechtern, so ist das ein Hinweis dafür, dass sich die beschränkten Parameter in den einzelnen Teilstichproben nicht voneinander unterscheiden.

Zur Durchführung einer konfirmatorischen (Mehrgruppen-)Faktorenanalyse stehen verschiedene Programme zur Verfügung, wie z. B. LISREL (Jöreskog & Sörbom, 1996), EQS (Bentler, 1995), Mplus (Muthén & Muthén, 2004) oder AMOS (Arbuckle, 2007).

EMPIRISCHER TEIL

4 Ziele der Untersuchung und Forschungsfragen

Ziel der vorliegenden Arbeit ist die Evaluation zweier Ansätze zur Erhebung von Individualismus und Kollektivismus bei Kinder- und Jugendlichen hinsichtlich ihrer Anwendbarkeit bei kulturvergleichenden Studien. Zum einen wird in einer modifizierten Fassung der *Collectivism Scale* von Yamaguchi (Yamaguchi et al., 1995) nach der Selbsteinschätzung der Kinder- und Jugendlichen in Bezug auf konstruktrelevante Einstellungen und Verhaltensweisen gefragt, während in der *Individualism Difference Scale*, einem neu entwickelten Erhebungsinstrument von Strohmeier und Lampert, die Differenz zwischen der eigenen Meinung und der wahrgenommener Norm als Indikator für Individualismus angesehen wird. Beide Erhebungsinstrumente sollen im ersten Schritt auf faktorielle Validität überprüft werden, um diese in einem weiteren Schritt einer psychometrischen Äquivalenzprüfung zwischen zwei Kulturen zu unterziehen. Abschließend soll die testtheoretische Güte beider Erhebungsinstrumente bestimmt werden. Aus den dargelegten Zielen lassen sich folgende Forschungsfragen und Hypothesen ableiten:

- (1) Kann die angenommene Dimensionsstruktur der *Collectivism Scale* inferenzstatistisch bestätigt werden?
 - Hypothese 1.1: Die faktorielle Validität der *Collectivism Scale* kann empirisch belegt werden.
- (2) Kann die angenommene Dimensionsstruktur der *Individualism Difference Scale* inferenzstatistisch bestätigt werden?
 - Hypothese 1.1: Die faktorielle Validität der *Individualism Difference Scale* kann empirisch belegt werden.
- (3) Erfolgt die Messung des Konstrukts Kollektivismus mittels *Collectivism Scale* statistisch äquivalent über zwei Kulturen hinweg?
 - Hypothese 3.1: Die strukturelle Äquivalenz der *Collectivism Scale* kann über beide Kulturen hinweg empirisch belegt werden.
 - Hypothese 3.2: Die Messeinheitsäquivalenz der *Collectivism Scale* kann über beide Kulturen hinweg empirisch belegt werden.
 - Hypothese 3.3: Die Skalenäquivalenz der *Collectivism Scale* kann über beide Kulturen hinweg empirisch belegt werden.

(4) Erfolgt die Messung des Konstrukts Individualismus mittels *Individualism Difference Scale* statistisch äquivalent über zwei Kulturen hinweg?

- Hypothese 4.1: Die strukturelle Äquivalenz der *Individualism Difference Scale* kann über beide Kulturen hinweg empirisch belegt werden.
- Hypothese 4.2: Die Messeinheitsäquivalenz der *Individualism Difference Scale* kann über beide Kulturen hinweg empirisch belegt werden.
- Hypothese 4.3: Die Skalenäquivalenz der *Individualism Difference Scale* kann über beide Kulturen hinweg empirisch belegt werden.

(5) Wie ist die testtheoretische Güte des *Collectivism Scale* bei Kinder- und Jugendlichen einzuschätzen?

- Hypothese 1.1: Die *Collectivism Scale* weist eine zufrieden stellende Reliabilität auf.
- Hypothese 1.2: Die Kriteriumsvalidität der *Collectivism Scale* konnte empirisch belegt werden.

(6) Wie ist die testtheoretische Güte des *Individualism Difference Scale* bei Kinder- und Jugendlichen einzuschätzen?

- Hypothese 1.1: Die *Individualism Difference Scale* weist eine zufrieden stellende Reliabilität auf.
- Hypothese 1.2: Die Kriteriumsvalidität der *Individualism Difference Scale* konnte empirisch belegt werden.

5 Methode

In diesem Kapitel wird das Untersuchungsdesign dargelegt und die Stichprobe beschrieben. Danach folgen die Beschreibung der Erhebungsinstrumente und die Erläuterung der statistischen Auswertungsverfahren, die im Rahmen dieser Arbeit zur Anwendung kommen.

5.1 Untersuchungsdesign und Stichprobenbeschreibung

Das Untersuchungsdesign sah die theoriegeleitete Auswahl von zwei Kulturen bzw. Ländern vor, die einerseits in Bezug auf die Konstrukte Individualismus und Kollektivismus möglichst unterschiedlich, andererseits jedoch in Bezug auf die Dimension horizontal-vertikal möglichst ähnlich sind. Diesem Kriterium entsprechend, wurden Österreich (vertikaler Individualismus) und Japan (vertikaler

Kollektivismus) als Untersuchungsländer ausgewählt, in welchen als Querschnittsuntersuchung eine Fragebogenerhebung stattfand.

Die Datenerhebung in Österreich fand im Dezember 2006 an zwei Schulklassen einer Volksschule und an neun Schulklassen eines Gymnasiums statt. Die Erhebung wurde von Strohmeier und Lampert während der regulären Unterrichtszeit durchgeführt und dauerte etwa eine Schulstunde (50 Minuten). Diesem ging die Einverständniserklärung des Stadtschulrates, des Direktors der jeweiligen Schule und der Eltern voraus. Etwa 90% aller Schüler der jeweiligen Schulklassen nahmen an dieser Befragung teil.

Die Datenerhebung in Japan wurde im November 2006 an sieben Schulklassen einer Grundschule⁴ und an neun Schulklassen einer Mittelschule von den jeweiligen Klassenlehrern während der Unterrichtszeit durchgeführt und dauert etwa eine Schulstunde (50 Minuten). Die notwendige Einverständniserklärung seitens des Direktors und der betreffenden Klassenlehrer wurde vorab eingeholt. In der Summe konnte etwa 96% aller Schüler der betreffenden Schulklassen befragt werden.

Insgesamt nahmen 809 Kinder- und Jugendliche an der Untersuchung teil, davon 277 (125 Mädchen und 149 Jungen, bei 3 Personen fehlte die Geschlechtsangabe) im Alter zwischen 8 und 15 Jahren ($M = 11.96$, $SD = 1.81$) aus Österreich und 532 (250 Mädchen und 282 Jungen) im Alter zwischen 9 und 15 Jahren ($M = 12.33$, $SD = 1.78$) aus Japan (siehe Tabelle 7).

Tabelle 7. Stichprobenbeschreibung

	Österreich			Japan		
Stichprobenumfang	$N = 277$			$n = 532$		
Geschlecht	Mädchen	Jungen		Mädchen	Jungen	
	125	149		250	282	
	45.6%	54.4%		47%	53%	
Alter	M	SD	Range	M	SD	Range
	11.96	1.81	8-15	12.33	1.78	9-15

Anmerkungen. Angaben bei Geschlecht in absoluten und relativen Zahlen. Aufgrund fehlender Angaben summieren sich die absoluten Zahlen nicht auf den Stichprobenumfang.

⁴ Das japanische Schulsystem sieht die Einschulung mit der Vollendung des 6. Lebensjahres in die Grundschule (Dauer: 6 Jahre) und den anschließenden Besuch einer Mittelschule (Dauer: 3 Jahre) vor.

5.2 Erhebungsinstrumente

Dieses Kapitel widmet sich der Beschreibung der verwendeten Erhebungsinstrumente, nämlich einer modifizierten Fassung der *Collectivism Scale* von Yamaguchi (1995) und der *Individualism Difference Scale*, einem neu entwickelten Erhebungsinstrument von Strohmeier und Lampert. Beide Erhebungsinstrumente wurden für die Vorgabe in der jeweiligen Landessprache aus der englischen Version in die deutsche bzw. japanische Sprache übersetzt und zur Kontrolle rückübersetzt.

5.2.1 *Collectivism Scale*

Die *Collectivism Scale* (Yamaguchi et al., 1995) baut auf dem Konzept des Selbst (vgl. Markus & Kitayama, 1991) auf, welches Unterschiede in der individuellen Ausprägung von Kollektivismus erklärt. Ursprünglich enthielt die *Collectivism Scale* zehn Items und wurde für den Einsatz in kollektivistischen Ländern entwickelt. Erst nachträglich wurde die Skala um vier „individualistische Items“ erweitert, die den Einsatz dieses Erhebungsinstruments auch in individualistischen Ländern ermöglichen sollten (Yamaguchi et al., 1995). Tabelle 8 listet alle Items der *Collectivism Scale* in deutscher Übersetzung auf.

Tabelle 8. Itemliste der *Collectivism Scale* in deutscher Übersetzung

Item	Inhalt
1	Ich verzichte oft auf Dinge, die ich gerne tun würde. Stattdessen mache ich etwas, das meine Mitschüler gerne tun wollen.
2	Ich mache immer das, was meine Mitschüler machen wollen.
3	Ich halte immer zu meinen Mitschülern, egal was passiert.
4	Ich versuche immer dafür zu sorgen, dass sich in der Klasse alle gut verstehen.
5	Ich ändere meine Meinung, wenn die Mehrheit meiner Mitschüler anders denkt.
6	Ich halte immer zu meinen Mitschülern, auch dann, wenn sie etwas tun, was ich nicht mag.
7	Ich unterstütze meine Mitschüler immer, egal ob sie Recht haben oder nicht.
8	Ich respektiere die Entscheidungen meiner Mitschüler immer.
9	Ich sage meinen Mitschülern nicht, wenn ich eine andere Meinung habe.
10	Ich helfe immer mit, Unstimmigkeiten in meiner Klasse zu verhindern.
11	Ich mache meinen Mitschülern keine Vorwürfe, wenn ich denke, dass sie etwas falsch gemacht haben.
12	Ich bleibe auch dann bei meinen Mitschülern, wenn sie etwas tun wollen, was ich nicht machen mag.

- 13 Ich mache Dinge gern so, wie ich es will, egal, was meine Mitschüler von mir erwarten.
- 14 Ich bleibe bei meiner Meinung, egal, ob meine Mitschüler mich dabei unterstützen oder nicht.
-

In jedem dieser 14 Items wird nach dem Ausmaß gefragt in der Personen bereit sind, ihre eigenen Ziele hinter die der Gruppe zu stellen. Dabei ist die *Collectivism Scale* nicht gruppenspezifisch, sodass die Bezugsgruppe je nach Forschungsfrage variiert werden kann. Für die Anwendung bei Kinder- und Jugendlichen wurde für die vorliegende Untersuchung als Referenzgruppe die Gruppe der Mitschüler festgesetzt. Zusätzlich wurde die Formulierung der Items dem sprachlichen Niveau der Kinder- und Jugendlichen entsprechend vereinfacht. Auf einer fünfstufigen Rating-Skala mit den Polen „Stimmt genau“ und „Stimmt nicht“ wird bei unterschiedlichen Einstellungsfragen wie z. B. „Ich mache immer das, was meine Mitschüler machen wollen.“ nach einer Entscheidung für das Individuum oder der Gruppe der Mitschüler gefragt – je öfters bzw. ausgeprägter die Entscheidung für die Gruppe ist, desto kollektivistischer werden die befragten Kinder- und Jugendlichen angesehen.

5.2.2 Individualism Difference Scale

Die *Individualism Difference Scale* baut ebenfalls auf dem theoretischen Konzept des Selbst auf, nämlich auf dem Konzept der unterschiedlichen Sichtweise des Selbst, die Markus und Kitayama (1991) *interdependent view of the self* und *independent view of the self* bezeichnen. Während im ersteren Fall der soziale Kontext mitsamt der Bezugsgruppe als Teil der eigenen Identität wahrgenommen wird (Kollektivismus), spielen Bezugsgruppen im letzteren Fall eine untergeordnete Rolle für die Selbstdefinition (Individualismus). Folglich unterscheiden sich beide Sichtweisen auch in Hinblick auf die Wichtigkeit von Normen (Gruppe) und Einstellungen (Selbst). So sind in individualistischen Ländern Einstellungen wichtiger als Normen, während es sich in kollektivistischen Ländern umgekehrt verhält (Triandis, 2004; Triandis & Gelfand, 1998). Desweiteren können in kollektivistischen Ländern die Ziele einer Bezugsgruppe soweit ins Bewusstsein rücken, dass sie nicht mehr von den eigenen unterschieden werden können (Markus & Kitayama, 1991).

Aufbauend auf diesen theoretischen Überlegungen wurde die *Individualism Difference Scale* konstruiert, die sowohl die persönlichen Einstellungen als auch die

wahrgenommene Norm einer Bezugsgruppe erheben und die Differenz beider Angaben als Indikator für Individualismus ansieht. Im vorliegenden Fall bei Kinder- und Jugendlichen wurde als Bezugsgruppe die Schulklasse festgesetzt und sieben Items mit kontextrelevanten Inhalten formuliert: (1) andere um Rat fragen, (2) Aufgaben in der Gruppe lösen, (3) andere Meinung als die Lehrer haben (4) sich nicht an Gruppenaufgaben beteiligen, (5) die eigene Meinung durchsetzen (6) eine schwierige Aufgabe allein lösen, (7) die Meinung nicht ändern wollen, obwohl alle anderen anders denken. Kinder- und Jugendliche gaben auf einer fünfstufigen Rating-Skala mit den Polen „Sehr gut“ und „Sehr schlecht“ jeweils Auskunft über ihre Meinung zu diesen Verhaltensweisen (Einstellung) und der antizipierten Meinung ihrer Klasse (wahrgenommene Norm). Tabelle 9 listet alle Items der *Individualism Difference Scale* in deutscher Übersetzung auf.

Tabelle 9. Itemliste der *Individualism Difference Scale* in deutscher Übersetzung

Item	Inhalt
1	Wie denkt du/deine Klasse über Mitschüler, die andere um Rat fragen, wenn sie ein Problem haben?
2	Wie denkt du/deine Klasse über Aufgaben, die man in der Gruppe lösen soll?
3	Wie denkt du/deine Klasse über Mitschüler, die eine andere Meinung haben als die Lehrer?
4	Wie denkt du/deine Klasse über Mitschüler, die sich nicht an Gruppenaufgaben beteiligen wollen?
5	Wie denkt du/deine Klasse über Mitschüler, die ihre eigene Meinung durchsetzen wollen?
6	Wie denkt du/deine Klasse über Mitschüler, die eine schwierige Aufgabe ganz allein lösen?
7	Wie denkt du/deine Klasse über Mitschüler, die ihre Meinung nicht ändern wollen, auch wenn alle anderen anders denken?

Die Besonderheit der *Individualism Difference Scale* liegt im Ansatz zur nachfolgenden Quantifizierung von Individualismus, bei der der Absolutbetrag der Differenz zwischen eigener Meinung und wahrgenommener Klassennorm gebildet wird und den eigentlichen Kennwert eines Items darstellt. Der Theorie entsprechend werden Kinder- und Jugendliche, die einen hohen Differenzwert aufweisen als individualistisch angesehen.

5.3 Auswertungsmethoden

Dieses Kapitel behandelt das konkrete Vorgehen bei der statistischen Auswertung zur Prüfung auf faktorielle Validität im ersten Schritt und der

anschließenden Prüfung auf statistische Äquivalenz im zweiten Schritt. Die Analysen erfolgen sowohl auf der Basis der probabilistischen Testtheorie (*Rasch-Modell*) als auch der klassischen Testtheorie (konfirmatorische Mehrgruppen-Faktorenanalyse). An dieser Stelle sei angemerkt, dass die Modellgeltung nach dem *Rasch-Modell* in den einzelnen Teilstichproben ein hinreichender Nachweis für faktorielle Validität und eine darauf aufbauende Modellgeltung für die Gesamtstichprobe auch den Nachweis der Äquivalenz darstellen, sodass die zusätzliche Analysen mittels konfirmatorischer Mehrgruppen-Faktorenanalyse redundant sind. Da die Analysen zur Prüfung der Äquivalenz in der Literatur für gewöhnlich faktorenanalytisch erfolgen, sollen in dieser Arbeit trotzdem beide Verfahren zur Anwendung kommen.

5.3.1 Äquivalenzprüfung nach dem *Rasch-Modell*

Der Nachweis der Datenäquivalenz zwischen verschiedenen Kulturen auf Grundlage des *Rasch-Modells* erfordert sowohl den Nachweis der Modellgeltung innerhalb der Stichproben einzelner Kulturen, als auch innerhalb der Gesamtstichprobe über die Kulturen hinweg (Wagner-Menghin, 2003). Die Gültigkeit des *Rasch-Modells* in den Teilstichproben belegt in diesem Zusammenhang die strukturelle Äquivalenz und die darauf aufbauende interkulturelle Gültigkeit des *Rasch-Modells* die Skalenäquivalenz. Die Messeinheitsäquivalenz stellt aufgrund der Voraussetzung der gleichen Trennschärfe aller Items (siehe Kapitel 3.3.1) bereits ein grundlegendes Merkmal des *Rasch-Modells* dar (Salzberger, Sinkovics & Schlegelmilch, 2001).

Die *Rasch-Modell* Analysen erfolgen dem beschriebenen Vorgehen entsprechend getrennt für die japanische und österreichische Stichprobe, um nach erfolgreichem Nachweis der Modellgültigkeit in den beiden Teilstichproben die Analysen zur Prüfung der *Rasch-Modell*-Konformität für die Gesamtstichprobe durchzuführen.

Vor Beginn der Analysen werden jeweils für die entsprechende Skala jene Personen mit maximal einem bearbeiteten Item ausgeschlossen, da die Parameter für diese Personen nach der *Conditional Maximum Likelihood Schätzung* nicht bestimmt werden können.

Folgende Teilungskriterien werden für die Modelltests festgelegt: „hoher vs. niedriger Score“, *Geschlecht* und „Alter 8-12 Jahre vs. 13-15 Jahre“. Die Analysen zur Überprüfung der interkulturellen Gültigkeit des *Rasch-Modells* erfolgt mittels

Teilungskriterium „Kultur Österreich vs. Japan“. Beim Teilungskriterium *Geschlecht* werden Schüler mit diesbezüglich fehlender Information ($N = 3$) zufällig auf die beiden Gruppen aufgeteilt.

Das Signifikanzniveau pro durchgeführten Modelltest beläuft sich auf $\alpha = 0.01$. Dies soll sicherstellen, dass das globale Risiko 1. Art von 5% eingehalten wird (siehe Kubinger, 2005a).

Nach Kubinger (2005a) empfiehlt es sich den Datensatz in eine Kalibrierungs- und Prüfstichprobe zu teilen, um die α -posteriori erreichte Modellanpassung anhand einer unabhängigen Stichprobe zu validieren (eine Art „Kreuzvalidierung“). Für die Analysen in der Gesamtstichprobe wird daher der Datensatz, bestehend aus 809 Schülern per Zufall im Verhältnis 70:30 in eine Kalibrierungs- und einer Prüfstichprobe aufgeteilt. Sollten die in der Kalibrierungsstichprobe als modellkonform attestierten Items in der Prüfstichprobe ebenfalls einer Modellprüfung standhalten, so würde dies für eine hinreichende Bewährung dieses Itempools sprechen.

Am Anfang der Prüfung auf Rasch-Modell-Konformität steht der Likelihood-Ratio-Test von Andersen (1973), der als Globaltest den Itempool insgesamt pro Teilungskriterium auf Modellverträglichkeit prüft. Erst bei Signifikanz in mindestens einem Teilungskriterium wird anschaulich mit Hilfe von Grafischen Modelltests und unter Berücksichtigung der z -Werten der *Wald-typ tests* (Fischer & Scheiblechner, 1970; Glas & Verhelst, 1995) versucht, nicht Rasch-Modell konforme Items zu identifizieren. Dabei wird in der Regel das auffälligste Item aus dem Itempool ausgeschieden, um den reduzierten Itempool anschließend einer erneuten Modelltestung zu unterziehen. Dieser Prozess wird iterativ so lange fortgesetzt bis eine α -posteriori-Modellanpassung erreicht wird (Kalibrierung nach dem Rasch-Modell). Zu Berücksichtigen ist hierbei stets die inhaltliche Relevanz der geschätzten Parameterdifferenzen zwischen zwei Teilstichproben (Effektstärke). Denn bei großen Stichproben führen die Modelltests trotz irrelevanten Abweichungen zur Ablehnung der Modellgültigkeit (siehe Kubinger, 2009). Nach einer Faustregel von Goethals (1994) sollen darum lediglich Abweichungen, die mehr als 1/10 aller Schwierigkeitsparameter betragen, als inhaltlich relevant erachtet werden. Angemerkt sei an dieser Stelle, dass neben diesen „technischen“ Selektionskriterien ein „inhaltlich“ begründeter Selektionsvorgang auf Grundlage psychologisch-inhaltlicher Überlegungen in Vordergrund zu stehen hat (Gittler, 1986). Denn laut Gittler „ist es nicht wichtig zu wissen, dass ein Item

nicht passt; es ist wichtig zu wissen warum es nicht passt“ (Kubinger, persönl. Mitteilung, 13.04.2009).

Sollte nach Ausscheiden von 30% des ursprünglichen Itempools (optimaler Weise jedoch höchstens 10%; siehe Kubinger, 2007) nach wie vor kein Rasch-Modell konformer Itempool vorliegen, so wird dieser Itempool als nicht skalierbar nach dem Rasch-Modell angesehen (Kubinger, persönl. Mitteilung, 07.11.2008).

Die Analysen zur Prüfung der Rasch-Modell-Konformität wird mit dem in „R“ (R Developmental Core Team, 2008) implementierten Programmpaket eRm (Mair & Hatzinger, 2007) durchgeführt.

5.3.2 Äquivalenzprüfung mittels konfirmatorischer Mehrgruppen-Faktorenanalyse

Die Äquivalenzprüfung mit Hilfe der konfirmatorischen Mehrgruppen-Faktorenanalyse erfolgt hierarchisch durch den Vergleich von Messmodellen dessen Parameterschätzungen über die Teilstichproben hinweg sukzessive beschränkt (*constrained*) werden (vgl. Steenkamp & Baumgartner, 1998) .

Als Ausgangspunkt der Analysen dient ein Messmodell, dessen Modellgüte in beiden Kulturen unabhängig voneinander hinreichend nachgewiesen wurde. Aus diesem Grund wird im ersten Schritt jeweils eine konfirmatorische Faktorenanalyse sowohl in der österreichischen wie auch in der japanischen Stichprobe durchgeführt und in Hinblick auf die Modellanpassung modifiziert. Erst nach Sicherstellung der Eindimensionalität des Messinstruments werden die Modellparameter einer stichprobenübergreifenden simultanen Analyse unterzogen.

Die Äquivalenz der Erhebungsdaten wird durch den Nachweis der Invarianz bestimmter Parameter belegt und kann den Äquivalenzebenen entsprechend (siehe Kapitel 3.2.3) ebenfalls in drei Ebenen unterteilt werden (vgl. Horn & McArdle, 1992; Steenkamp & Baumgartner, 1998):

Bei der (1) *konfiguralen Invarianz* wird überprüft, ob die gleiche Faktorenstruktur in allen untersuchten Gruppen gegeben ist. Das bedeutet, dass alle Items auf die selben latenten Variablen laden und damit dieselben signifikanten Ladungen bzw. Null-Ladungen aufweisen. Unter der Voraussetzung der konfiguralen Invarianz bedingt die (2) *metrische Invarianz* dass die Höhe der Ladungen über die Gruppen hinweg gleich ist, d. h. die Ladungsparameter können zusätzlich zum Ladungsmuster beschränkt werden. Nach erfolgreichem Nachweis der konfiguralen und metrischen Invarianz kann die (3) *skalare Invarianz* untersucht

werden, die besagt, dass die Itemursprünge (Item-Intercepts) über die untersuchten Gruppen hinweg als gleich angenommen bzw. beschränkt werden können.

Die Restriktion auf jeder Ebene der Invarianzprüfung führt zwangsläufig zur Verschlechterung der Modellanpassung (Model-Fit). Ob diese Verschlechterung statistisch bedeutsam ist, wird mit einem χ^2 -Differenztest ($\Delta\chi^2$) überprüft. Chen (2007) empfiehlt als zusätzliche Kriterien folgende Fit-Indizes heranzuziehen: CFI, RMSEA und SRMR.

Die Analysen in der vorliegenden Arbeit orientieren sich an den Cut-off-Werten (Schwellenwerten) von Chen (2007) und jenen Schwellenwerten die der gängigen Konvention zur Beurteilung von Fit-Indizes entsprechen (vgl. Browne & Cudeck, 1993; Hu & Bentler, 1995, 1998, 1999; Rodgers, 1999; Schermelleh-Engel et al., 2003). Tabelle 10 fasst die Cut-Off-Werte für einen guten bzw. akzeptablen Fit und die Cut-off-Werte für die Beurteilung der Differenz in Abhängigkeit der Invarianzform zusammen

Tabelle 10. Cut-off-Werte der verwendeten Fit-Indices und zur Beurteilung der Differenz in Abhängigkeit der Invarianzform

Fit Indizes	Guter Fit	Akzeptabler Fit	Differenz: Fit Indizes	Metrische Invarianz	Skalarere Invarianz
CFI	$95 \leq \text{CFI} \leq 1.00$	$.90 \leq \text{CFI} < 0.95$	$\Delta \text{ CFI}$	$\geq -.01$	$\geq -.01$
RMSEA	$0 \leq \text{RMSEA} \leq .05$	$.05 < \text{RMSEA} \leq .10$	$\Delta \text{ RMSEA}$	$\geq .015$	$\geq .015$
SRMR	$0 \leq \text{SRMR} \leq .05$	$.05 < \text{SRMR} \leq .10$	$\Delta \text{ SRMR}$	$\geq .030$	$\geq .010$

Anmerkung. CFI: Comparative-Fit-Index; RMSEA: Root-Mean-Square-error-of-Approximation; SRMR: Standardized-Root-Mean-Square-Residual.

Die Invarianzprüfung erfolgt in der Regel für das gesamte Messmodells, d. h. die Parameterbeschränkungen auf jeder Invarianzebene beziehen sich auf alle Items. In der Praxis jedoch stellt dies einen hohen Anspruch an das Messinstrument dar, wodurch selten volle metrische Invarianz oder gar volle skalare Invarianz erreicht wird. Das Konzept der partiellen Invarianz bietet die Möglichkeit einigen Parametern in jeder Gruppe spezifische Schätzungen zuzuweisen. Neben der Referenzvariable ist hierbei bereits ein metrisch bzw. skalar invariantes Item für die Vergleichbarkeit der Daten ausreichend (siehe Byrne, Shavelson & Muthén, 1989; Singh, 1995; Steenkamp & Baumgartner, 1998).

Als allgemeine Vorgehensweise wird nach Erreichen der konfiguralen Invarianz zuerst auf vollständige metrische Invarianz geprüft; sollte diese nicht gegeben sein, werden einzelne Parameterrestriktionen aufgehoben, um auf partielle metrische Invarianz zu prüfen. Für die nachfolgende Prüfung auf partielle skalare Invarianz werden hierauf nur jene Items einbezogen, die die metrische Invarianzbedingung erfüllen konnten. Die Identifikation jener Parameter, dessen Restriktion aufgehoben werden, erfolgt mit Hilfe so genannter Modifikationsindizes (Sörbom, 1989). Diese Kennwerte zeigen an wie stark sich der χ^2 -Wert bei entsprechender Modifikation reduziert. Dementsprechend werden die Restriktionen jener Parameter, die den höchsten Modifikationsindex aufweisen, aufgehoben (Bühner, 2006). Die Berechnung dieser Modifikationsindizes erfordert einen vollständigen Datensatz, weshalb vor Beginn der Analysen ein listenweise Ausschluss fehlender Werte erfolgt.

Die Anpassung eines Modells an einen Datensatz erfordert stets einer Validierung an einer unabhängigen Stichprobe, um eine artifizielle Modellanpassung ausschließen zu können (Steiger, 1990). Aus diesem Grund wird für die Invarianzprüfung die Gesamtstichprobe per Zufall im Verhältnis 50:50 in eine Kalibrierungs- und Prüfstichprobe geteilt, um die in der Kalibrierungsstichprobe erreichte (partielle) Invarianzebene anhand der Prüfstichprobe zu validieren.

Die Analysen mittels konfirmatorischen (Mehrgruppen-)Faktorenanalyse erfolgten mit AMOS 16.0 (Arbuckle, 2007).

6 Ergebnisse

Nachfolgend werden die Ergebnisse der Rasch-Modell Analysen und der konfirmatorischen Faktorenanalysen für die *Collectivism Scale* und der *Individualism Difference Scale* dargestellt. Dem folgen die Bestimmung der Reliabilität beider Erhebungsinstrumente mitsamt Prüfung auf Äquivalenz der Reliabilitätskoeffizienten in beiden Kulturen, sowie die Bestimmung der konkurrenten Validität mittels sog. „Extremgruppenvalidierung“ (siehe Kubinger, 2006).

6.1 Rasch-Modell Analysen

6.1.1 Collectivism Scale

6.1.1.1 Österreichische Stichprobe

Die Analysen ergaben signifikante Likelihood-Ratio-Tests in den Teilungskriterien *Score*, *Geschlecht* und *Alter*. Nach Ausscheiden von fünf Items (siehe Anhang Tabelle 34), das sind 35.7% des ursprünglichen Itempools, waren die Likelihood-Ratio-Tests in den Teilungskriterien *Geschlecht* und *Alter* nach wie vor signifikant (siehe Tabelle 11 und Abbildung 2). Zudem zeigten die Differenzen der in beiden Teilstichproben geschätzten Itemparameter insbesondere im Teilungskriterium *Median* bedeutsame Abweichungen (siehe Anhang Tabelle 33). Da nach Ausscheiden von über 30% des ursprünglichen Itempools die Kalibrierung nach dem *Rasch-Modell* nicht gelang, wird die *Collectivism Scale* in der österreichischen Stichprobe als nicht skalierbar nach dem *Rasch-Modell* angenommen.

Tabelle 11. Likelihood-Ratio-Tests der *Collectivism Scale* in der österreichischen Stichprobe ($n_s = 277$); Modelltest für die Teilungskriterien *Score*, *Geschlecht* und *Alter*

Teilungskriterium		χ^2 (LRT)	df	χ^2 ($\alpha = .01$)
<i>Score</i>	vor Ausschluss von Items	103.0676	47	72.44331
	nach Ausschluss von 5 Items	57.31123	31	52.1914
<i>Geschlecht</i>	vor Ausschluss von Items	67.70077	55	82.29212
	nach Ausschluss von 5 Items	85.3451	35	57.34207
<i>Alter</i>	vor Ausschluss von Items	118.6985	51	77.38596
	nach Ausschluss von 5 Items	65.15726	31	52.1914

Anmerkung. Item 2 erwies sich nach Ausschluss von 5 Items als „not well-conditioned“ im Teilungskriterium *Score* und *Alter*; weiters Item 3 als „not well-conditioned“ im Teilungskriterium *Score*.

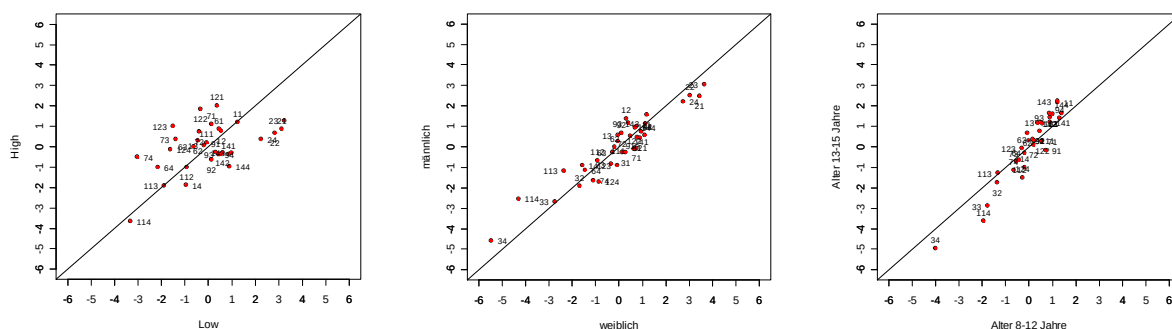


Abbildung 2. Grafische Modelltests für die *Collectivism Scale* in der österreichischen Stichprobe ($n_s = 277$) nach Ausschluss von 5 Items, Teilungskriterien *Score*, *Geschlecht* und *Alter*

6.1.1.2 Japanische Stichprobe

Die Analysen ergaben signifikante Likelihood-Ratio-Tests in den Teilungskriterien *Geschlecht* und *Alter*. Nach Ausschluss von fünf Items (siehe Anhang Tabelle 36), das sind 35,7% des ursprünglichen Itempools, war der Likelihood-Ratio-Test im Teilungskriterium *Alter* nach wie vor signifikant (siehe Tabelle 12 und Abbildung 3). Dementsprechend weisen die Differenzen der in beiden Teilstichproben geschätzten Itemparameter für einzelne Items bedeutsame Abweichungen auf (siehe Anhang Tabelle 35). Da nach Ausscheiden von über 30% des ursprünglichen Itempools die Kalibrierung nach dem *Rasch-Modell* nicht gelang, wird die *Collectivism Scale* in der japanischen Stichprobe als nicht skalierbar nach dem *Rasch-Modell* angenommen.

Tabelle 12. Likelihood-Ratio-Tests der *Collectivism Scale* in der japanischen Stichprobe ($n_j = 531$); Modelltest für die Teilungskriterien *Score*, *Geschlecht* und *Alter*

Teilungskriterium		χ^2 (LRT)	df	χ^2 ($\alpha = .01$)
Score	vor Ausschluss von Items	57.929	55	82.292
	Nach Ausschluss von 5 Items	56.329	35	57.342
Geschlecht	vor Ausschluss von Items	113.061	55	82.292
	Nach Ausschluss von 5 Items	49.791	35	57.342
Alter	vor Ausschluss von Items	146.802	55	82.292
	Nach Ausschluss von 5 Items	84.054	35	57.342

Anmerkung. Item 7 erwies sich nach Ausschluss von 5 Items als „not well-conditioned“ im Teilungskriterium *Score*

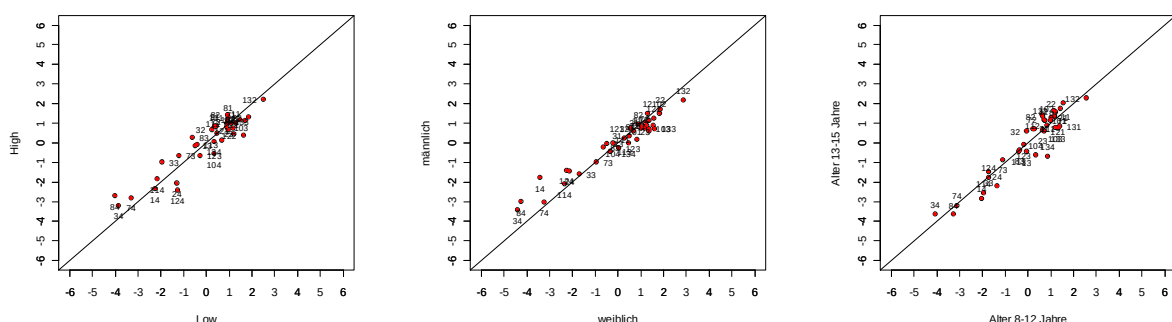


Abbildung 3. Grafische Modelltests für die *Collectivism Scale* in der japanischen Stichprobe ($n_j = 531$) nach Ausschluss von 5 Items, Teilungskriterien *Score*, *Geschlecht* und *Alter*

6.1.1.3 Gesamtstichprobe

Da der *Collectivism Scale* weder in der japanischen noch in der österreichischen Stichprobe die Rasch-Modell-Konformität attestiert werden konnte, erübligten sich die weiteren Analysen in der Gesamtstichprobe.

6.1.2 Individualism Difference Scale

Die Rasch-Modell Analysen der *Individualism Difference Scale* nach dem ordinalen Rasch-Modell zeigten, dass das Teilungskriterium *Score* in beiden Teilstichproben aufgrund der geringen Zellbesetzungen in den oberen Kategorien (siehe Anhang Tabelle 37) nicht berechnet werden konnte. Hierauf wurde entschieden, die Differenzwerte in zwei Kategorien zusammenzufassen, nämlich *keine Abweichung* zwischen eigener Meinung und wahrgenommener Klassennorm (bzw. absoluter Differenzwert: 0) vs. *Abweichung* (bzw. absoluter Differenzwert: 1 bis 4) (siehe im Anhang Tabelle 38). Infolgedessen erfolgte die Auswertung mittels dichotomen Rasch-Modell.

6.1.2.1 Österreichische Stichprobe

Die Analysen ergaben einen signifikanten Likelihood-Ratio-Test im Teilungskriterium *Geschlecht*. Nach Ausschluss von Item 5 (siehe Anhang Tabelle 40) konnte Modellkonformität erzielt werden (siehe Tabelle 13 und Abbildung 4). Angemerkt sei jedoch, dass die Differenzen der in beiden Teilstichproben geschätzten Itemparameter durchweg in allen Teilungskriterien bedeutsame Abweichungen aufwiesen (siehe Anhang Tabelle 39).

Tabelle 13. Likelihood-Ratio-Tests der *Individualism Difference Scale* in der österreichischen Stichprobe ($n_0 = 277$); Modelltest für die Teilungskriterien *Score*, *Geschlecht* und *Alter*

Teilungskriterium		χ^2 (LRT)	<i>df</i>	χ^2 ($\alpha = .01$)
<i>Score</i>	vor Ausschluss von Items	2.561	6	16.812
	Nach Ausschluss von 1 Items	6.301	5	15.086
<i>Geschlecht</i>	vor Ausschluss von Items	20.219	6	16.812
	Nach Ausschluss von 1 Items	10.861	5	15.086
<i>Alter</i>	vor Ausschluss von Items	0.788	6	16.812
	Nach Ausschluss von 1 Items	0.777	5	15.086

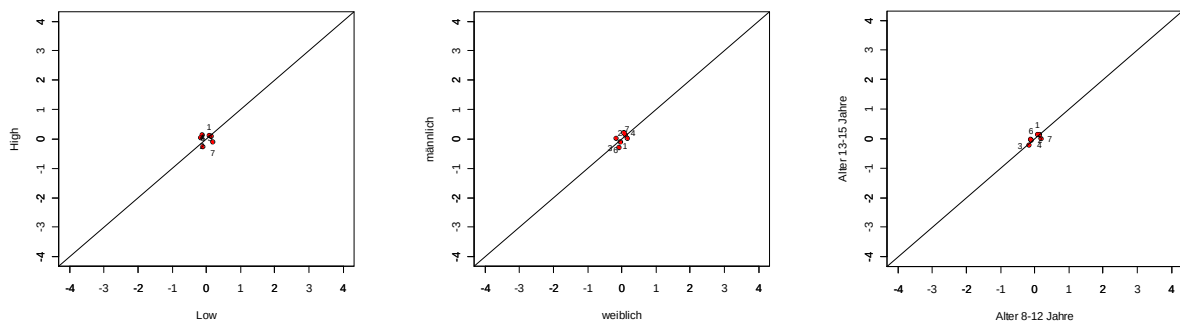


Abbildung 4. Grafische Modelltests für die *Individualism Difference Scale* in der österreichischen Stichprobe ($n_o = 277$) nach Ausschluss von 1 Item, Teilungskriterien Score, Geschlecht und Alter

6.1.2.2 Japanische Stichprobe

In den Analysen waren die Likelihood-Ratio-Tests in allen Teilungskriterien nicht signifikant, d. h. der gesamte Itempool konnte sich in der entsprechenden Population als Rasch-Modell konform bewähren (siehe Tabelle 14 und Abbildung 5). Angemerkt sei jedoch, dass die Differenzen der in beiden Teilstichproben geschätzten Itemparameter durchweg in allen Teilungskriterien bedeutsame Abweichungen aufwiesen (siehe Anhang Tabelle 41).

Tabelle 14. Likelihood-Ratio-Tests der *Individualism Difference Scale* in der japanischen Stichprobe ($n_j = 531$); Modelltest für die Teilungskriterien Score, Geschlecht und Alter

Teilungskriterium		χ^2 (LRT)	df	χ^2 ($\alpha = .01$)
Score	Ohne Ausschluss von Items	10.025	6	16.812
Geschlecht	Ohne Ausschluss von Items	3.608	6	16.812
Alter	Ohne Ausschluss von Items	5.144	6	16.812

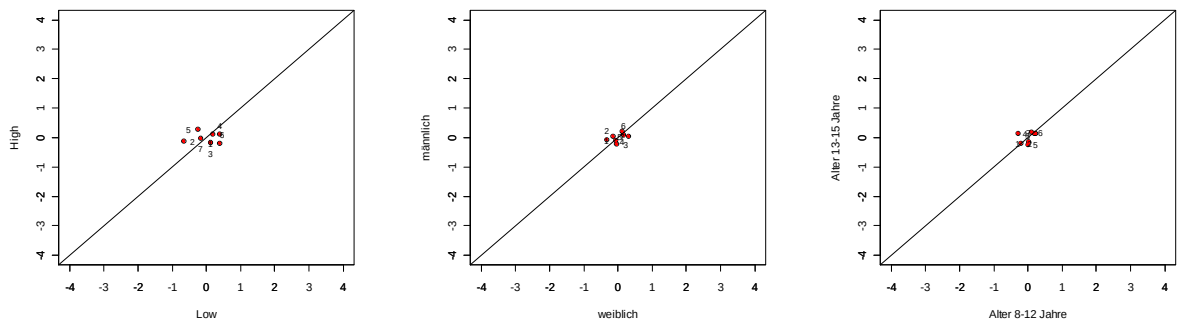


Abbildung 5. Grafische Modelltests für die *Individualism Difference Scale* in der japanischen Stichprobe ($n_j = 531$), Teilungskriterien Score, Geschlecht und Alter

6.1.2.3 Gesamtstichprobe

Die Rasch-Modell Analysen in der Gesamtstichprobe erfolgten mit dem reduzierten Itempool, dem bereits Rasch-Modell-Konformität in beiden Teilstichproben attestiert werden konnte. In der Kalibrierungsstichprobe war der Likelihood-Ratio-Test im Teilungskriterium Kultur nicht signifikant, was für die Gültigkeit des Rasch-Modells sprach (siehe Tabelle 15 und Abbildung 6). Angemerkt sei jedoch, dass die Differenzen der in beiden Teilstichproben geschätzten Itemparameter durchweg bedeutsame Abweichungen aufwiesen (siehe Anhang Tabelle 42).

Tabelle 15. Likelihood-Ratio-Tests der *Individualism Difference Scale* für den reduzierten Itempool in der Kalibrierungsstichprobe ($n_k = 567$); Modelltest für das Teilungskriterium Kultur

Teilungskriterium		χ^2 (LRT)	df	χ^2 ($\alpha = .01$)
Kultur	Ohne Ausschluss von Items	6.791	5	15.086

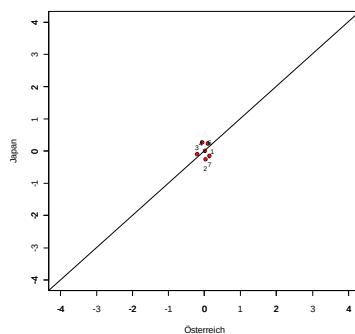


Abbildung 6. Grafische Modelltests für die *Individualism Difference Scale* in der Kalibrierungsstichprobe ($n_k = 567$), Teilungskriterium Kultur

In der Prüfstichprobe war der Likelihood-Ratio-Test im Teilungskriterium Kultur nicht signifikant (siehe Tabelle 16 und Abbildung 7.-1-3), d. h. der reduzierte Itempool konnte sich in der Prüfstichprobe ebenfalls hinsichtlich der Gültigkeit des Rasch-Modells bewähren. Angemerkt sei jedoch, dass die Differenzen der in beiden Teilstichproben geschätzten Itemparameter durchweg bedeutsame Abweichungen aufwiesen (siehe Anhang Tabelle 43).

Tabelle 16. Likelihood-Ratio-Tests der *Individualism Difference Scale* für den reduzierten Itempool in der Prüfstichprobe ($n_p = 242$); Modelltest für das Teilungskriterium Kultur

Teilungskriterium		χ^2 (LRT)	df	χ^2 ($\alpha = .01$)
Kultur	Ohne Ausschluss von Items	4.632	5	15.086

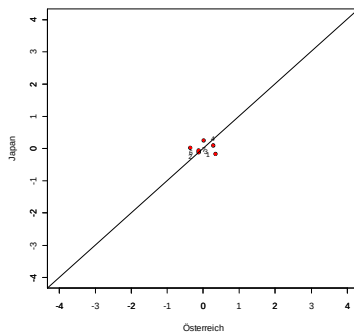


Abbildung 7. Grafische Modelltests für die *Individualism Difference Scale* in der Prüfstichprobe ($n_p = 242$), Teilungskriterium Kultur

Nachdem sich der reduzierte Itempool sowohl in der Kalibrierungs- als auch in der Prüfstichprobe hinreichend bewährt hat, kann in Folge die interkulturelle Gültigkeit des Rasch-Modells für diesen Itempool angenommen werden.

6.2 Konfirmatorische (Mehrgruppen-)Faktorenanalyse

6.2.1 Collectivism Scale

6.2.1.1 Überprüfung des Messmodells

Österreichische Stichprobe. Die Analyse der *Collectivism Scale* in der österreichischen Stichprobe mit Hilfe der konfirmatorischer Faktorenanalyse ergab einen signifikanten χ^2 -Wert ($\chi^2_{(77)} = 250.80, p < .001$); während RMSEA mit .092 und SRMR mit .085 auf eine akzeptablen Modellanpassung hindeuteten, wies der CFI mit .697 hingegen auf einen schlechten Modell-Fit hin. Hierauf wurden fünf Items (Item 1, 8, 10, 13 und 14) aus den Analysen ausgeschlossen, die sowohl in den Modifikationsindizes auffällig waren, als auch geringe Faktorladungen bzw. Kommunalitäten aufwiesen. Zwar war nach Ausschluss der Items der χ^2 -Wert auf $\alpha = .01$ nach wie vor signifikant ($\chi^2_{(27)} = 53.37$), jedoch deuteten alle Fit-Indizes auf eine akzeptable Modellanpassung hin (CFI = .92, RMSEA = .061, SRMR = .054). Aufgrund dessen wurde dem reduzierten Itempool eine hinreichende Modellanpassung im Sinne der Eindimensionalität dieser Skala in der österreichischen Stichprobe attestiert (siehe Tabelle 17).

Tabelle 17. Messmodells der *Collectivism Scale* in der österreichischen Stichprobe

	χ^2	<i>df</i>	<i>p</i>	CFI	RMSEA 90% CI	SRMR
vor Ausschluss von Items	250.80	77	< .001	.697	.092 .080-.105	.085
nach Ausschluss von 5 Items	53.37	27	> .01	.92	.061 .036-.085	.054

Anmerkung. CFI: Comperative Fit Index; RMSEA: Root Mean Square Error of Approximation; SRMR: Standardized Root Mean Square Residual.

Japanische Stichprobe. Die Analyse der *Collectivism Scale* in der japanische Stichprobe mittels konfirmatorischer Faktorenanalyse ergab einen signifikanten Chi²-Wert ($\chi^2_{(77)} = 431.15$, $p < .001$); RMSEA mit einem Wert von .102 und CFI mit einem Wert von .663 deuteten auf eine schlechte Modellanpassung hin, während SRMR mit .088 im akzeptablen Wertebereich lag. Nach Ausschluss von vier Items (Item 1, 9, 10 und 14) wiesen die Fit-Indizes trotz signifikanten Chi²-Wert ($\chi^2_{(77)} = 78.018$, $p < .001$) auf eine gute bis akzeptable Modellanpassung hin: CFI = .924, RMSEA = .053 und SRMR = .047. In Folge dieses Befundes wurde dem reduzierten Itempool eine hinreichende Modellanpassung im Sinne der Eindimensionalität dieser Skala in der japanischen Stichprobe attestiert (siehe Tabelle 18).

Tabelle 18. Messmodell der *Collectivism Scale* in der japanischen Stichprobe

	χ^2	<i>df</i>	<i>p</i>	CFI	RMSEA 90% CI	SRMR
vor Ausschluss von Items	431.15	77	< .001	.663	.102 .093-.111	.0876
nach Ausschluss von 4 Items	78.02	35	< .001	.924	.053 .037-.068	.0467

Anmerkung. CFI: Comperative Fit Index; RMSEA: Root Mean Square Error of Approximation; SRMR: Standardized Root Mean Square Residual.

Im Hinblick auf die Invarianzprüfung verblieben acht Items, die sich in den Analysen sowohl in der österreichischen, als auch in der japanischen Stichprobe bewährt haben. Das Messmodell mit dem reduzierten Itempool ist in Abbildung 8 veranschaulicht.

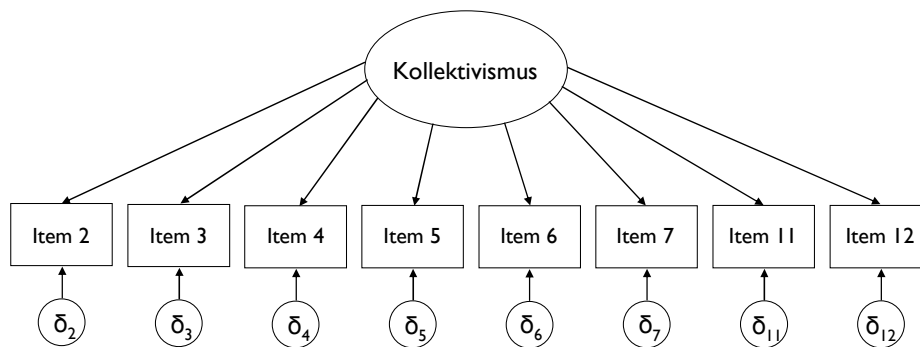


Abbildung 8. Messmodell der *Collectivism Scale* mit dem reduzierten Itempool

Tabelle 19 zeigt die geschätzten standardisierten Faktorladungen, Standardfehler und erklärten Varianzanteile (Kommunalitäten) der latenten Variable „Kollektivismus“ getrennt für die österreichische und japanische Stichprobe.

Tabelle 19. Geschätzte standardisierte Faktorladungen, Standardfehler und Kommunalitäten der reduzierten *Collectivism Scale* in der österreichischen und japanischen Stichprobe

Item	Österreich			Japan		
	λ	SE	h^2	λ	SE	h^2
2	.494	.082	.244	.321	.112	.103
3	.487	.077	.237	.543	.139	.295
4	.286	.085	.082	.264	.111	.070
5	.447	.114	.200	.319	.143	.102
6	.658	.105	.433	.615	.144	.379
7	.724*	-	.524	.555*	-	.308
11	.278	.086	.077	.445	.139	.198
12	.583	.099	.340	.520	.138	.271

Anmerkung. λ = standardisierte Faktorladung; h^2 = Kommunalität; alle Ladungen waren signifikant;

* Ladung wurde auf 1 fixiert und diente als Referenzvariable.

6.2.1.2 Invarianzprüfung in der Kalibrierungstichprobe

Die Überprüfung der drei Ebenen der Invarianz erfolgte mittels konfirmatorischer Mehrgruppen-Faktorenanalyse bei sukzessiver Beschränkung der Modellparameter.

Konfigurale Invarianz. Die Analyse des reduzierten Itempools der *Collectivism Scale* ergab bei Beschränkung der Faktorenstruktur einen signifikanten χ^2 -Wert ($\chi^2_{(40)} = 250.80$, $p < .001$), wobei RMSEA mit .055 und SRMR mit .070 auf eine akzeptable Modellanpassung hindeuteten. Allerdings unterschritt der CFI mit .866 deutlich den Schwellenwert für eine akzeptable Modellanpassung (Tabelle 20., Modell A1).

Hierauf wurde Item 5 aus dem Itempool herausgenommen, welches zu einer deutlichen Verbesserung des Modells mit einem χ^2 -Wert von 51.565 mit 28 Freiheitsgraden ($p = .004$) führte; alle Werte der Fit-Indizes lagen im akzeptablen Bereich (CFI = .907, RMSEA = .050, SRMR = .051; Tabelle 20., Modell A2). Aufgrund dieses Befundes wurde den sieben Items des Modells A2 konfigurale Invarianz über die beiden Teilstichproben hinweg attestiert.

Vollständige metrische Invarianz. Die Prüfung auf vollständige metrische Invarianz erfolgte durch den Vergleich des Modells A2 (Konfigurale Invarianz) mit einem Modell bei der zusätzlich alle Ladungen der Items über die Gruppen hinweg beschränkt wurden. Der Modellvergleich zeigte keine signifikante Verschlechterung der beiden Modelle ($\Delta\chi^2_{(6)} = 8.009, p = .237$). Die Differenz der Fit Indizes-Werte lagen ebenfalls im akzeptablen Bereich ($\Delta\text{CFI} = .008, \Delta\text{RMSEA} = -.003, \Delta\text{SRMR} = .011$) und sprachen insgesamt für die metrische Invarianz der 7 Items der *Collectivism Scale* (Tabelle 20., Modell B).

Vollständige Skalare Invarianz. Zur Prüfung der skalaren Invarianz wurde das Modells B (Vollständige metrische Invarianz) mit einem weiteren Modell dessen Itemursprünge (*Intercept*) zusätzlich zur Faktorenstruktur und Ladungen beschränkt wurden, verglichen. Der Modellvergleich ergab einen signifikanten χ^2 -Wert ($\Delta\chi^2_{(7)} = 91.868, p < .001$); die Differenz der Fit Indizes-Werte wiesen mit $\Delta\text{CFI} = .336, \Delta\text{RMSEA} = .042$ und $\Delta\text{SRMR} = .012$ ebenfalls eine deutliche Verschlechterung des Modells auf (Tabelle 20, Modell C1). Da die vollständige skalare Invarianz nicht bestätigt werden konnte, erfolgte die Prüfung auf partielle skalare Invarianz.

Partielle skalare Invarianz. Für die Prüfung auf partielle skalare Invarianz wurden die Beschränkungen der Itemursprünge für Item 2, 3, 4, 6, und 11 aufgehoben, sodass neben der Referenzvariable (Item 7) lediglich der Itemursprung von Item 12 über die Gruppen hinweg beschränkt blieb. Der Vergleich von Modell B (Vollständige metrische Invarianz) mit dem Modell partieller Invarianz (Modell C2) zeigte eine nicht signifikante Verschlechterung des Modells mit einem χ^2 -Differenz-Wert von $\Delta\chi^2 = 5.550$ mit 2 Freiheitsgraden ($p = .062$) auf. Die Verschlechterung der Werte der Fit-Indizes lagen beim $\Delta\text{CFI} = .012$ außerhalb des akzeptablen Bereichs; $\Delta\text{RMSEA} = .002$ und $\Delta\text{SRMR} = .0003$ hingegen deuten auf keine Modellverschlechterung hin (Tabelle 20., Modell C1). Aufgrund der signifikanten Verschlechterung des CFI konnte dem reduzierten Itempool keine partielle skalare Invarianz attestiert werden.

Tabelle 20. Invarianzprüfung der *Collectivism Scale* in der Kalibrierungsstichprobe

Modell	χ^2	df	p	CFI	RMSEA 90% CI	SRMR	Modellvergleich		
							$\Delta\chi^2$	Δdf	p
A1 Konfigurale Invarianz	80.452	40	< .001	.866	.055 .037-.072	.0702	-	-	-
A2 Konfigurale Invarianz	51.565	28	.004	.907	.050 .028-.071	.0510	-	-	-
B Vollständige Metrische Invarianz	59.573	34	.004	.899	.047 .026-.066	.0620	8.009	6	.237
C1 Vollständige Skalare Invarianz	151.442	41	< .001	.563	.089 .074-.104	.0740	91.868 ¹	7	< .001
C2 Partielle Skalare Invarianz	51.565	28	.004	.885	.049 .029-.067	.0623	5.550 ¹	2	.062

Anmerkung. CFI: Comparative Fit Index; RMSEA: Root Mean Square Error of Approximation; SRMR: Standardized Root Mean Square Residual. ¹ Vergleich zum Modell B Vollständige Metrische Invarianz.

6.2.1.3 Invarianzprüfung in der Prüfstichprobe

In der Prüfstichprobe konnte sich der reduzierte Itempool ebenfalls im Sinne der konfiguralen Invarianz bewähren, wobei diese Invarianzbedingung auch ohne Ausschluss von Item 5 erfüllt war ($\chi^2_{(40)} = 60.262$, $p = .021$, CFI = .951, RMSEA = .037, SRMR = .053) und der Ausschluss dieses Items zu keiner bedeutsamen Verbesserung des Modells führte (Tabelle 21, Modell A1, A2).

Die Prüfung auf vollständige metrische Invarianz ergab eine signifikante Verschlechterung des Modells ($\Delta\chi^2_{(6)} = 22.747$, $p = < .001$, $\Delta CFI = .045$) im Vergleich zum Modell konfiguraler Invarianz (Tabelle 21, Modell B). Die Fit-Indizes für sich genommen, sprachen allerdings für eine akzeptable Modellanpassung (CFI = .908, RMSEA = .052, SRMR = .075), sodass die vollständige metrische Invarianz als gegeben erachtet werden konnte.

Auch in der Prüfstichprobe konnte wie in der Kalibrierungsstichprobe keine vollständige skalare Invarianz erreicht werden ($\Delta\chi^2_{(7)} = 118.80$, $p < .001$, $\Delta CFI = .3$, $\Delta RMSEA = .0047$, $\Delta SRMR = .009$) (Tabelle 21, Modell C1). Die nachfolgende Prüfung auf partielle skalare Invarianz ergab einen nicht signifikanten Chi²-Differenz-Wert ($\Delta\chi^2_{(2)} = 6.737$, $p = .034$) für den Vergleich mit dem Modell vollständiger metrischer Invarianz (Modell B); abgesehen vom ΔCFI mit .013 sprachen die Differenzen der Fit-Indizes-Werte ebenfalls gegen eine Modellverschlechterung ($\Delta RMSEA = .002$, $\Delta SRMR = .0001$) (Tabelle 21, Modell C2).

Tabelle 21. Invarianzprüfung der *Collectivism Scale* in der Prüfstichprobe

Modell	χ^2	df	p	CFI	RMSEA 90% CI	SRMR	Modellvergleich		
							$\Delta\chi^2$	Δdf	p
A1 Konfigurale Invarianz	60.262	40	.021	.951	.037 .015-.055	.053	-	-	-
A2 Konfigurale Invarianz	45.470	28	.020	.953	.041 .017-.062	.055	-	-	-
B Vollständige metrische Invarianz	62.218	34	< .001	.908	.052 .034-.070	.075	22.747	6	< .001
C1 Vollständige skalare Invarianz	187.021	41	< .001	.608	.099 .084-.113	.084	118.80 ¹	7	< .001
C2 Partielle skalare Invarianz	74.954	36	< .001	.895	.054 .037-.072	.076	6.737 ¹	2	.034

Anmerkung. CFI: Comparative Fit Index; RMSEA: Root Mean Square Error of Approximation; SRMR: Standardized Root Mean Square Residual. ¹ Vergleich zum Modell B Vollständige Metrische Invarianz.

Insgesamt sprachen die Ergebnisse in der Prüfstichprobe für die Konsistenz der in der Kalibrierungss Stichprobe erhaltenen Befunde. In beiden Teilstichproben konnte in den Analysen nicht die Ebene partieller skalarer Invarianz erreicht werden. Somit kann dem reduzierten Itempool der *Collectivism Scale* lediglich vollständige metrische Invarianz attestiert werden.

6.2.2 Individualism Difference Scale

Die Analysen der *Individualism Difference Scale* mittels konfirmatorischer (Mehrgruppen-)Faktorenanalyse erfolgte im Gegensatz zu den Rasch-Modell Analysen nicht mit den dichotomisierten, sondern mit den fünfkategorialen Daten.

6.2.2.1 Überprüfung des Messmodells

Österreichische Stichprobe. Die Analyse der *Individualism Difference Scale* in der österreichischen Stichprobe mit Hilfe der konfirmatorischer Faktorenanalyse ergab einen nicht signifikanten χ^2 -Wert ($\chi^2_{(14)} = 16.975$, $p = .258$); übereinstimmend mit diesem Ergebnis wiesen alle Fit-Indizes auf eine gute Modellanpassung hin: CFI = .975, RMSEA = .029 und SRMR = .041. Dieses Ergebnis sprach für die Eindimensionalität der *Individualism Difference Scale* in der österreichischen Stichprobe (siehe Tabelle 22).

Tabelle 22. Messmodells der *Individualism Difference Scale* in der österreichischen Stichprobe

	χ^2	<i>df</i>	<i>p</i>	CFI	RMSEA 90% CI	SRMR
ohne Ausschluss von Items	16.975	14	.258	.975	.029 .000-.070	.0409

Anmerkung. CFI: Comperative Fit Index; RMSEA: Root Mean Square Error of Approximation; SRMR: Standardized Root Mean Square Residual.

Japanische Stichprobe. Die Analyse der *Individualism Difference Scale* in der japanische Stichprobe mittels konfirmatorischer Faktorenanalyse ergab einen signifikanten χ^2 -Wert ($\chi^2_{(14)} = 61.948$, $p < 0.001$). Zwar deuten RMSEA mit .084, CFI mit .910 und SRMR mit .0480 auf eine akzeptable Modellanpassung hin, trotzdem wurde Item 2 ausgeschlossen, da es deutliche Auffälligkeiten in den Modifikationsindizes zeigte. Die Analyse des reduzierten Itempools ergab einen χ^2 -Wert von 19.195 mit 9 Freiheitsgraden ($p = .024$); alle Fit Indizes deuteten auf eine gute Modellanpassung hin: CFI = .974, RMSEA = .048 und SRMR = .031. Diesem reduzierten Itempool wurde somit eine hinreichende Modellanpassung im Sinne der Eindimensionalität dieser Skala in der japanischen Stichprobe attestiert (siehe Tabelle 23).

Tabelle 23. Messmodell der *Individualism Difference Scale* in der japanischen Stichprobe

	χ^2	<i>df</i>	<i>p</i>	CFI	RMSEA 90% CI	SRMR
vor Ausschluss von Items	61.948	14	< .001	.910	.084 .063-.106	.0480
nach Ausschluss von 1 Item	19.195	9	.024	.974	.048 .017-.079	.0309

Anmerkung. CFI: Comperative Fit Index; RMSEA: Root Mean Square Error of Approximation; SRMR: Standardized Root Mean Square Residual.

Für die Invarianzprüfung verblieben dechs Items, die sich in den Analysen sowohl in der österreichischen, als auch in der japanischen Stichprobe bewährt haben. Das Messmodell mit dem reduzierten Itempool ist in Abbildung 9 veranschaulicht.

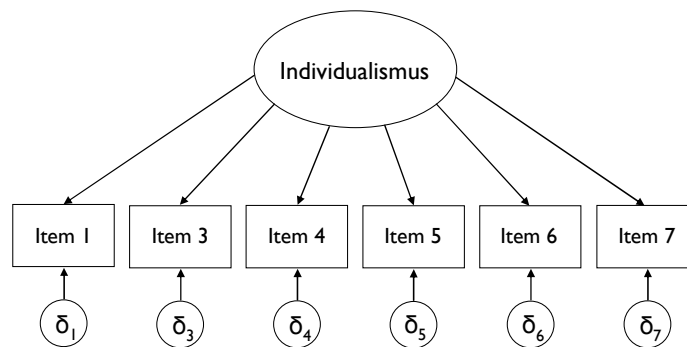


Abbildung 9. Messmodell der *Individualism Difference Scale* mit dem reduzierten Itempool

Tabelle 24 zeigt die geschätzten standardisierten Faktorladungen, Standardfehler und erklärten Varianzanteile (Kommunalitäten) der latenten Variable „Individualismus“ getrennt für die österreichische und japanische Stichprobe.

Tabelle 24. Geschätzte standardisierte Faktorladungen, Standardfehler und Kommunalitäten der *Individualism Difference Scale* in der österreichischen und japanischen Stichprobe

Item	Österreich			Japan		
	λ	SE	h^2	λ	SE	h^2
Item 1	.456	.187	.208	.420	.155	.177
Item 3	.538*	-	.289	.450*	-	.202
Item 4	.429	.201	.184	.546	.197	.298
Item 5	.444	.182	.197	.626	.200	.391
Item 6	.372	.156	.139	.595	.220	.354
Item 7	.203	.159	.041	.529	.184	.279

Anmerkung. λ = standardisierte Faktorladung; h^2 = Kommunalität; alle Ladungen waren signifikant;

* Ladung wurde auf 1 fixiert und diente als Referenzvariable.

6.2.2.2 Invarianzprüfung in der Kalibrierungstichprobe

Die Prüfung des reduzierten Itempools der *Individualism Difference Scale* auf konfigurale, metrische und skalare Invarianz erfolgte mit Hilfe der konfirmatorischen Mehrgruppen-Faktorenanalyse, indem Modellparameter auf jeder Ebene der Invarianzprüfung sukzessive beschränkt wurden.

Konfigurale Invarianz. Die Analyse des reduzierten Itempools der *Individualism Difference Scale* ergab bei Beschränkung der Faktorenstruktur einen χ^2 -Wert von 17 mit 18 Freiheitsgraden ($p = .523$). Diesem nicht signifikanten χ^2 -Wert entsprechend, deuteten alle Fit Indizes auf eine gute Modellanpassung hin (CFI = 1.0, RMSEA = 0.0, SRMR = .036; Tabelle 25, Modell A1). Aufgrund dieses Befundes

wurde den sechs Items der *Individualism Difference Scale* konfigurale Invarianz über die Teilstichproben hinweg attestiert.

Vollständige metrische Invarianz. Zur Prüfung der vollständigen metrischen Invarianz wurde das Modell konfiguraler Invarianz (Tabelle 25, Modell A1) mit einem weiteren Modell mit zusätzlicher Beschränkung aller Itemladungen verglichen. Der Modellvergleich ergab einen signifikanten χ^2 -Wert ($\Delta\chi^2_{(5)} = 20.252, p = .031$), alle Differenzen der Fit Indizes sprachen ebenfalls für eine bedeutsame Verschlechterung des Modells ($\Delta CFI = .071, \Delta RMSEA = .041$ und $\Delta SRMR = .059$; Tabelle 25, Modell B1). Da die vollständige metrische Invarianz für die sechs Items nicht erreicht werden konnte, erfolgte die Prüfung auf partielle metrische Invarianz.

Partielle metrische Invarianz. Für die Prüfung auf partielle metrische Invarianz wurden die Beschränkungen der Itemladungen für die Items 6 und 7 aufgehoben. Der Vergleich von Modell A1 (Konfigurale Invarianz) mit dem Modell B2 (Partielle Metrische Invarianz) ergab einen nicht signifikanten χ^2 -Differenzwert von 6.55 mit 3 Freiheitsgraden ($p = .088$). Die Differenz der Fit Indizes ergaben folgende Werte: $\Delta CFI = .013, \Delta RMSEA = .018$ und $\Delta SRMR = .002$ (Tabelle 25, Modell B2). Abgesehen von der Differenz der SRMR-Werte deuteten ΔCFI und $\Delta RMSEA$ zwar auf eine Modellverschlechterung hin, aufgrund des nicht signifikanten χ^2 -Wertes wird jedoch partielle metrische Invarianz für die verbleibenden 3 Items⁵ (Item 1, 4 und 5) angenommen.

Partielle skalare Invarianz. Die Prüfung auf partielle skalare Invarianz erfolgt mit jenen Items, denen im vorangegangenen Schritt metrische Invarianz zugesprochen werden konnte. Dementsprechend werden die Itemursprünge der Items 1, 3, 4 und 5 über die Teilstichproben hinweg beschränkt. Die Analyse ergab eine signifikante Verschlechterung des Modells ($\Delta\chi^2_{(4)} = 74.461, p < .001$; $\Delta CFI = .353, \Delta RMSEA = .072, \Delta SRMR = .020$) im Vergleich zum Modell partieller metrischer Invarianz (Tabelle 25, Modell C1). Da den drei metrisch invarianten Items im Modell C1 keine skalare Invarianz attestiert werden konnte, wurde ein weiteres Modell aufgestellt, das neben der Referenzvariable (Item 3) lediglich für Item 4 eine Beschränkung der Itemursprünge vorsieht. Der Vergleich dieses Modells mit Modell B2 (partielle metrische Invarianz) zeigte, dass sich das Modell bei Beschränkung der Itemursprünge nach wie vor signifikant verschlechtert ($\Delta\chi^2_{(2)} = 43.348, p < .001$). Abgesehen vom SRMR deuteten ΔCFI mit .207 und $\Delta RMSEA$ mit .054 auf eine

⁵ Item 3 bleibt unberücksichtigt, da dieses die Referenzvariable darstellt.

deutliche Verschlechterung des Modells hin (Tabelle 25, Modell C2). Diesem Ergebnis folgend konnte der *Individualism Difference Scale* keine partielle skalare Invarianz in der Kalibrierungsstichprobe attestiert werden.

Tabelle 25. Invarianzprüfung der *Individualism Difference Scale* in der Kalibrierungsstichprobe

Modell	χ^2	df	p	CFI	RMSEA 90% CI	SRMR	Modellvergleich		
							$\Delta\chi^2$	Δdf	p
A Konfigurale Invarianz	17.004	18	.523	1.0	.0 .0-.044	.036	-	-	-
B1 Vollständige metrische Invarianz	37.256	23	.031	.929	.041 .013-.065	.094	20.252	5	.001
B2 Partielle metrische Invarianz	23.554	21	.315	.987	.018 .0-.049	.054	6.550 ¹	3	.088
C1 Partielle skalare Invarianz	98.015	25	< .001	.634	.090 .071-.109	.0744	74.461 ²	4	< .001
C2 Partielle skalare Invarianz	66.902	23	< .001	.780	.072 .053-.093	.052	43.348 ²	2	< .001

Anmerkung. CFI: Comparative Fit Index; RMSEA: Root Mean Square Error of Approximation; SRMR: Standardized Root Mean Square Residual. ¹ Vergleich zum Modell A1 Konfigurale Invarianz. ² Vergleich zum Modell B2 Partielle Metrische Invarianz.

6.2.2.3 Invarianzprüfung in der Prüfstichprobe

In der Prüfstichprobe konnte ebenfalls die Gültigkeit der konfigurale Invarianz für den reduzierten Itempool der *Individualism Difference Scale* nachgewiesen werden ($\chi^2_{(18)} = 32.402$, $p = .02$). Alle Fit Indizes deuteten dabei auf eine akzeptable Modellanpassung hin (CFI = .948, RMSEA = .046, SRMR = .052; Tabelle 26, Modell A).

Der nachfolgende Vergleich eines Modells mit Beschränkung der Itemladungen mit dem Modell A (Konfigurale Invarianz) zeigte entgegen den Ergebnissen in der Kalibrierungsstichprobe keine signifikante Verschlechterung dieses Modells ($\Delta \chi^2_{(40)} = 6.245$, $p = .283$, $\Delta CFI = .005$, $\Delta RMSEA = .004$, $\Delta SRMR = .017$). In der Kalibrierungsstichprobe folgte die Prüfung auf partielle metrische Invarianz, welches in der Prüfstichprobe eine geringfügig schlechtere Modellanpassung aufwies als das Modell vollständiger metrischer Invarianz ($\chi^2_{(21)} = 37.401$, $p = .015$; CFI = .941, RMSEA = .0464, SRMR = .064; Tabelle 26, Modell B2).

Die Prüfung auf partielle skalare Invarianz mit Beschränkung der Itemursprünge der Items 1, 3, 4 und 5 zeigte, dass diese Items auch in der Prüfstichprobe keine skalare Invarianz aufweisen. Dabei deuten sowohl $\Delta\chi^2$ -Wert mit 35.294 ($p < .001$)

als auch die Differenzwerte der Fit Indizes ($\Delta CFI = .114$, $\Delta RMSEA = .025$, $\Delta RMR = .014$) auf eine signifikante Verschlechterung des Modells im Vergleich zum Modell B2 (Partielle metrische Invarianz) hin (Tabelle 26, Modell C1). Der weitere Vergleich des Modells C2 (Partielle skalare Invarianz) mit Beschränkung des Itemursprungs des Items 4 neben der Referenzvariable mit dem Modell B2 (Partielle metrische Invarianz) ergab nach wie vor eine signifikante Verschlechterung des Modells ($\chi^2_{(2)} = 19.353$, $p = .015$). Mit Ausnahme von SRMR ($\Delta SRMR = .001$) deuteten alle Fit Indizes ebenfalls auf eine signifikante Modellverschlechterung hin ($\Delta CFI = .114$, $\Delta RMSEA = .025$; Tabelle 26, Modell C2).

Tabelle 26. Invarianzprüfung der *Individualism Difference Scale* in der Prüfstichprobe

Modell	χ^2	df	p	CFI	RMSEA 90% CI	SRMR	Modellvergleich		
							$\Delta\chi^2$	Δdf	p
A Konfigurale Invarianz	32.402	18	.020	.948	.046 .018-.071	.052	-	-	-
B1 Vollständige Metrische Invarianz	38.647	23	.022	.943	.042 .016-.065	.069	6.245	5	.283
B2 Partielle Metrische Invarianz	37.401	21	.015	.941	.046 .020-.069	.064	4.99	3	.172
C1 Partielle Skalare Invarianz	72.695	25	< .001	.827	.071 .052-.090	.078	35.294	4	< .001
C2 Partielle Skalare Invarianz	56.754	23	< .001	.878	.062 .042-.083	.065	19.353	2	< .001

6.3 Testtheoretische Güte

Das vorliegende Kapitel beschäftigt sich zum einen mit der Bestimmung der Reliabilität beider Erhebungsinstrumente unter Berücksichtigung der Frage nach der Äquivalenz der Reliabilitätskoeffizienten in beiden Kulturen. Nach Van de Vijver und Leung (1997) wird die Äquivalenz mittels folgender Formel berechnet:

$$F_{emp} = \frac{1 - \alpha_1}{1 - \alpha_2}$$

Formel 4. Formel zur Äquivalenzprüfung zweier Reliabilitätskoeffizienten

Dabei steht α_1 und α_2 für die Reliabilitäten, die mit demselben Messinstrument in zwei verschiedenen kulturellen Gruppen ermittelt wurden. Für große Stichproben folgt diese Teststatistik eine F-Verteilung mit $N_1 - 1$ und $N_2 - 1$ Freiheitsgraden (N_1 und N_2 stehen für die Stichprobengröße).

Zum anderen soll die Kriteriumsvalidität mit Hilfe einer Extremgruppenvalidierung (siehe Kubinger, 2005b) bestimmt werden, was der Überprüfung der Übereinstimmungsvalidität (konkurrente Validität) entspricht.

Die Extremgruppenvalidierung stellt zwei Gruppen gegenüber, die sich hinsichtlich der Ausprägung der interessierenden Eigenschaft möglichst voneinander unterscheiden. Im Falle dieser Untersuchung stellen die österreichischen und die japanische Schülern zwei Extremgruppen hinsichtlich der Ausprägung von Individualismus bzw. Kollektivismus dar. Denn es kann davon ausgegangen werden, dass die österreichischen Schüler über eine hohe Ausprägung von Individualismus bzw. geringe Ausprägung von Kollektivismus verfügen, während es sich bei den japanischen Schülern umgekehrt verhält. Sollten die Erhebungsinstrumente diese Merkmale valide erfassen, müsste sich (1) der Testkennwert in den Gruppen signifikant voneinander unterscheiden und (2) jede Person aufgrund des Testkennwertes korrekt in die richtige Gruppe klassifizieren lassen. Ersteres wird mit einem Welch-Test (Welch, 1974), auch bekannt als *t*-Test für heterogene Varianzen (zur Verwendung des Welch-Tests siehe Kubinger, Rasch & Moder, 2009) und letzteres mittels Diskriminanzanalyse (Klassifikationsanalyse) mit einer Leave-one-out-Kreuzvalidierung bzw. mit der Jack-knife-Methode (siehe Rodgers, 1999) überprüft.

Zu beachten ist, dass für die Bestimmung der Kriteriumsvalidität jeweils Mittelwerte über die Items der jeweiligen Skalen gebildet werden, um einen Kennwert für die Ausprägung einer Person auf dieser Skala zu haben. Für die interkulturelle Vergleichbarkeit dieses Kennwerts müsste allerdings skalare Äquivalenz der Skala gegeben sein, was für die *Collectivism Scale* und die *mehrkategorielle Individualism Difference Scale* nicht der Fall ist. In der Interpretation bzw. Diskussion der Ergebnisse soll dieser Umstand berücksichtigt werden.

6.3.1 *Collectivism Scale*

Die Überprüfung der testtheoretischen Güte erfolgte für den reduzierten Itempool der *Collectivism Scale* (siehe Kapitel 6.1.1.3). Die reduzierte Skala umfasste folgende Items: Item 2, 3, 4, 5, 6, 7, 11 und 12.

6.3.1.1. Reliabilität

Die Reliabilitätsanalyse der *Collectivism Scale* in der Gesamtstichprobe ergab einen Reliabilitätskoeffizienten von $\alpha = .70$. Die Analysen getrennt nach den Teilstichproben ergaben ein $\alpha = .72$ für die österreichische und ein $\alpha = .69$ für die japanische Stichprobe (siehe Tabelle 27). Dieser Unterschied war statistisch nicht signifikant, $F(261, 484) = 0.92$, $p = .21$, d. h. die Reliabilitätskoeffizienten unterscheiden sich nicht signifikant in den beiden Teilstichproben.

Die Beurteilung der Höhe der Reliabilität erfolgt in der Regel in Abhängigkeit des Kontexts. Beachtet man, dass die *Collectivism Scale* für Forschungszwecke entwickelt wurde und Merkmale im Persönlichkeits- bzw. Einstellungsbereich erfasst, kann die Reliabilität als ausreichend hoch angesehen werden (Moosbrugger & Kelava, 2007).

Tabelle 27. Reliabilitätskoeffizienten der *Collectivism Scale* für die Gesamt- und Teilstichproben

	Itemzahl	Itemtrennschärfe	Cronbach's Alpha
Gesamtstichprobe	8	.267-.518	.695
Österreichische Stichprobe	8	.232-.588	.720
Japanische Stichprobe	8	.253-.503	.694

6.3.1.2. Konkurrente Validität

Die Analyse mittels Welch-Test zeigte keine signifikanten Unterschiede ($t(461.56) = 1.29$, $p = .232$, $r = 0.06$) zwischen Schülern in Österreich ($M = 2.84$, $SD = 0.73$) und Japan ($M = 2.90$, $SD = 0.58$) hinsichtlich der Mittelwerte der *Collectivism Scale*.

Die Klassifikationsanalyse mittels Diskriminanzanalyse im zweiten Schritt zeigte bei einer a-priori Zuordnungswahrscheinlichkeit von 0.5, dass 52.8% der Schüler (48.4% der österreichischen Schüler und 55.4% der japanischen Schüler) durch die gefundene Diskriminanzfunktion richtig zugeordnet werden konnten (siehe Tabelle 28). Das bedeutet, dass aufgrund der Information über die Ausprägung auf der *Collectivism Scale* eine 5.4% höhere Trefferquote bei den japanischen Schülern erzielt werden konnte als bei zufälliger Einteilung in beide Gruppen. Bei den österreichischen Schülern hingegen führte die Klassifikation auf Grundlage der *Collectivism Scale* zu einer 1.6% geringeren Trefferquote als bei Zuteilung nach dem Zufall. Die Kreuzvalidierung nach der Leave-one-out Methode lieferte exakt dasselbe Ergebnis (siehe Anhang Tabelle 44).

Tabelle 28. Klassifikationsanalyse der *Collectivism Scale*

		Vorhergesagte Gruppenzugehörigkeit		Gesamt (N = 808)
		Österreich	Japan	
Tatsächliche Gruppenzugehörigkeit				
	Österreich	48.4% (134)	51.6% (143)	34.2% (277)
	Japan	44.6% (237)	55.4% (294)	65.7% (531)

Anmerkung. A-priori Wahrscheinlichkeit: 0.5

6.3.2 Individualism Difference Scale

Für die *Individualism Difference Scale* ergeben sich aufgrund der Rasch-Modell Analysen (siehe Kapitel 6.1.2) und den Analysen der konfirmatorischen (Mehrgruppen-)Faktorenanalyse (siehe Kapitel 6.2.2) zwei unterschiedliche Skalen. Einerseits die dichotomisierte Skala mit den Items 1, 2, 3, 4, 6 und 7; andererseits die fünfkategorielle Skala mit den Items 1, 3, 4, 5, 6 und 7. Die nachfolgenden testtheoretische Analysen erfolgen getrennt für beide Skalen.

6.3.2.1. Reliabilität

6.3.2.1.1 Dichotome Individualism Difference Scale

Die Bestimmung der Reliabilität bei dichotomen Items erfolgte mit Hilfe der Kuder-Richardson-Formel (Lienert & Raatz, 1998). Die Reliabilitätsanalysen ergaben einen Reliabilitätskoeffizienten von $\alpha = .69$ in der Gesamtstichprobe, einen $\alpha = .48$ in der österreichische Stichprobe und einen $\alpha = .70$ in der japanischen Stichprobe (siehe Tabelle 29). Die Unterschiede der Reliabilitätskoeffizienten zwischen der österreichischen und der japanischen Stichprobe waren statistisch signifikant, $F(261, 484) = 0.92, p = < .001$.

Die Reliabilität der *dichotomen Individualism Difference Scale* kann in der Gesamtstichprobe sowie in der japanischen Stichprobe als zufrieden stellend beurteilt werden, wohingegen die Reliabilität in der österreichischen Stichprobe – trotz geringer Itemanzahl – als gering zu beurteilen ist.

Tabelle 29. Reliabilitätskoeffizienten der *Individualism Difference Scale* für die Gesamt- und Teilstichproben

	Itemzahl	Itemtrennschärfe	Cronbach's Alpha
Gesamtstichprobe	6	.394-.439	.688
Österreichische Stichprobe	6	.190-.311	.480

6.3.2.1.2 Mehrkategorielle *Individualism Difference Scale*

Die Reliabilitätsanalyse der mehrkategorialen *Individualism Difference Scale* ergab einen Reliabilitätskoeffizienten von $\alpha = .69$ in der Gesamtstichprobe. In der österreichischen Stichprobe wurde ein $\alpha = .54$, sowie in der japanischen Stichprobe ein $\alpha = .70$ ermittelt (siehe Tabelle 30). Dieser Unterschied war mit $F(262, 486) = 0.92$ ($p = < .001$) statistisch signifikant.

Die Reliabilität der mehrkategorialen *Individualism Difference Scale* kann in der Gesamtstichprobe sowie in der japanischen Stichprobe als ausreichend hoch beurteilt werden. In der österreichischen Stichprobe weist die Skala trotz geringer Itemanzahl eine geringe Reliabilität auf.

Tabelle 30. Reliabilitätskoeffizienten der *Individualism Difference Scale* für die Gesamt- und Teilstichproben

	Itemzahl	Itemtrennschärfe	Cronbach's Alpha
Gesamtstichprobe	6	.375-.467	.691
Österreichische Stichprobe	6	.157-.357	.540
Japanische Stichprobe	6	.352-.499	.700

6.3.2.2. Konkurrente Validität

6.3.2.2.1 Dichotome *Individualism Difference Scale*

Der Analyse mittels Welch-Test ergab einen signifikanten Gruppenunterschied mittlerer bis großer Effektstärke ($t(574,53) = 12,56$, $p < .001$, $r = 0.46$) zwischen österreichischen ($M = 0,49$, $SD = 0.26$) und japanischen Schülern ($M = 0.25$, $SD = 0.27$) bezüglich den Gruppenmittelwerten der dichotomen *Individualism Difference Scale*.

Die Klassifikationsanalyse der Diskriminanzanalyse zeigte, dass 68.9% der Schüler (61.7% der österreichischen Schüler und 72.7% der japanischen Schüler) aufgrund der ermittelten Diskriminanzfunktion bei einer a-priori Zuordnungswahrscheinlichkeit von 0.5 richtig in die Gruppen zugeordnet werden konnten (siehe Tabelle 31). Dies entspricht einem Zugewinn von 11.7% bei der österreichischen und 22.7% bei der japanischen Stichprobe in Bezug auf die Trefferquote im Vergleich zu einer Zufallszuordnung. Die Kreuzvalidierung nach der

Leave-one-out Methode sprach für die Konsistenz dieses Ergebnisses (siehe Anhang 45).

Tabelle 31. Klassifikationsanalyse der *dichotomen Individualism Difference Scale*

		Vorhergesagte Gruppenzugehörigkeit		Gesamt (N = 794)
		Österreich	Japan	
Tatsächliche Gruppenzugehörigkeit				
	Österreich	61.7% (169)	38.3% (105)	34.5% (274)
	Japan	27.3% (142)	72.7% (378)	65.5% (520)

Anmerkung. A-priori Wahrscheinlichkeit: 0.5

6.3.2.2.2 Mehrkategorielle *Individualism Difference Scale*

Die Bestimmung der konkurrenten Validität ergab einen signifikanten Unterschied mittlerer bis großer Effektstärke in Bezug auf die Gruppenmittelwerte in der *mehrkategorialen Individualism Difference Scale* zwischen Schülern in Österreich ($M = 0.73$, $SD = 0.48$) und jenen in Japan ($M = 0.34$, $SD = 0.43$), $t(504,42) = 11,29$, $p < .001$, $r = 0.45$.

Die ermittelte Diskriminanzfunktion der Klassifikationsanalyse konnte bei einer a-priori Zuordnungswahrscheinlichkeit von 0.5 insgesamt 69.6% der Schüler (58.4% der österreichischen Schüler und 75.4% der japanischen Schüler) in die richtige Gruppe einteilen (siehe Tabelle 32). Das bedeutet, dass sich die Trefferquote bei den Schülern in Österreich um 8.4% und bei denen in Japanern um 25.4% im Vergleich zur Klassifikation nach dem Zufall erhöhte. Die Kreuzvalidierung nach der Leave-one-out Methode ergab dasselbe Ergebnis (siehe Anhang 46).

Tabelle 32. Klassifikationsanalyse der *mehrkategorialen Individualism Difference Scale*

		Vorhergesagte Gruppenzugehörigkeit		Gesamt (N = 795)
		Österreich	Japan	
Tatsächliche Gruppenzugehörigkeit				
	Österreich	58.4% (160)	41.6% (114)	34.5% (274)
	Japan	24.6% (128)	75.4% (393)	65.5% (521)

Anmerkung. A-priori Wahrscheinlichkeit: 0.5

Diskussion

Die Ergebnisse der vorliegenden Arbeit können Bezug nehmend auf die Fragestellungen und den dazugehörigen Hypothesen folgendermaßen zusammengefasst werden:

Collectivism Scale. Den Rasch-Modell Analysen zufolge kann die faktorielle Validität der *Collectivism Scale* sowohl für die österreichische als auch die japanische Stichprobe nicht angenommen werden. Infolgedessen wurde keine Äquivalenzprüfung nach dem Rasch-Modell durchgeführt. Die Analysen mit Hilfe der konfirmatorischen (Mehrgruppen-)Faktorenanalyse ergaben demgegenüber ein konträres Bild. Nach dieser Analyseverfahren zufolge kann die faktorielle Validität der *Collectivism Scale* nach Ausschluss von fünf bzw. vier Items in beiden Teilstichproben angenommen werden. Die anschließende Äquivalenzprüfung konnte die Messeinheitsäquivalenz und somit auch die strukturelle Äquivalenz der *Collectivism Scale* empirisch belegen. Allerdings erreichte die *Collectivism Scale* nicht die Ebene der Skalenäquivalenz, das bedeutet die Messung mit Hilfe dieser Skala erfolgt nicht skalar invariant über die Kulturen hinweg. In Bezug auf die testtheoretische Güte kann der *Collectivism Scale* eine ausreichend hohe Reliabilität attestiert werden. Zudem konnte die Äquivalenz der Reliabilitätskoeffizienten über beide Kulturen hinweg nachgewiesen werden. In Hinblick auf die Kriteriumsvalidität zeigten die Analysen, dass sich Schüler in Österreich hinsichtlich des Gesamtscores der *Collectivism Scale* im Mittel nicht von Schülern in Japan unterscheiden. Dementsprechend war die *Collectivism Scale* nicht in der Lage Schüler aus Österreich bzw. Japan richtig zuzuordnen. Nach diesen Ergebnissen zufolge ist die Kriteriumsvalidität der *Collectivism Scale* als praktisch nicht vorhanden einzuschätzen.

Individualism Difference Scale. Zur Durchführbarkeit der Rasch-Modell Analysen wurden die Differenzwerte der *Individualism Difference Scale* dichotomisiert (Abweichung: nicht vorhanden/vorhanden). Die darauf aufbauenden Analysen der *dichotomen Individualism Difference Scale* zeigten, dass die faktorielle Validität nach Ausschluss eines Items in der österreichischen Stichprobe angenommen werden kann. In der japanischen Stichprobe hingegen konnte die faktorielle Validität ohne Ausschluss eines Items nachgewiesen werden. Dies stellt darüber hinaus den Beleg der strukturellen Äquivalenz der *dichotomen Individualism Difference Scale* über beide Kulturen hinweg dar. In der nachfolgenden Äquivalenzprüfung konnte zudem die Skalenäquivalenz und

damit auch die Messeinheitsäquivalenz der *dichotomen Individualism Difference Scale* empirisch belegt werden. Die Analysen mittels konfirmatorischer (Mehrgruppen-) Faktorenanalyse erfolgten mittels *mehrkategorieller Individualism Difference Scale*. Hierbei konnte die faktorielle Validität in der österreichischen Stichprobe ohne Ausschluss eines Items nachgewiesen werden, während diese in der japanischen Stichprobe nach Ausschluss eines Items belegt werden konnte. Im Rahmen der Äquivalenzprüfung der *mehrkategorieller Individualism Difference Scale* konnte die partielle Messeinheitsäquivalenz, nicht jedoch die Skalenäquivalenz nachgewiesen werden. Die testtheoretischen Analysen erfolgten getrennt für die *dichotome* und die *mehrkategorielle Individualism Difference Scale*. Die Reliabilität sowohl der *dichotomen* als auch der *mehrkategoriellen Individualism Difference Scale* kann in der Gesamtstichprobe als ausreichend hoch beurteilt werden. Allerdings unterschieden sich die Reliabilitätskoeffizienten in Abhängigkeit der untersuchten Stichprobe dahingehend, dass diese in der österreichischen Stichprobe geringer ist als in der japanischen Stichprobe. Dabei zeigte sich, dass die Reliabilität in der österreichischen Stichprobe als gering zu beurteilen ist. Dies trifft sowohl für die *dichotome* als auch die *mehrkategorielle Individualism Difference Scale* zu, wobei die *mehrkategorielle* Skala eine geringfügig höhere Reliabilität aufweist als die *dichotome* Skala. In Bezug auf die Kriteriumsvalidität zeigte sich, dass Schüler in Österreich hinsichtlich des Gesamtscores sowohl der *dichotomen* als auch der *mehrkategoriellen Individualism Difference Scale* im Mittel von Schülern in Japan unterscheiden. So waren beide Skalen in der Lage einen Beitrag zur richtigen Zuordnung von Schülern aus Österreich bzw. Japan zu leisten, wobei die Trefferquote der *mehrkategoriellen Individualism Difference Scale* höher war als bei der *dichotomen* Skala. Weiters zeigte sich, dass die Trefferquote beider Skalen bei Schülern aus Japan wesentlich höher ist als bei Schülern aus Österreich. Zusammenfassend kann der *Individualism Difference Scale*, sowohl als *dichotome* als auch *mehrkategorielle* Skala eine zufrieden stellende Kriteriumsvalidität attestiert werden.

Vor dem Hintergrund dieser Ergebnisse werden folgende Punkte diskutiert: (1) *Collectivism Scale* als Erhebungsinstrument für Kollektivismus bei Kinder- und Jugendlichen, (2) *Individualism Difference Scale* als Erhebungsinstrument für Individualismus bei Kinder- und Jugendlichen und (3) *Rasch-Modell* und

konfirmatorische Mehrgruppen-Faktorenanalyse als Methoden zur psychometrischen Äquivalenzprüfung.

(1) *Collectivism Scale* als Erhebungsinstrument für Kollektivismus bei Kinder- und Jugendlichen. Den Ergebnissen der vorliegenden Arbeit zufolge, muss die psychometrische Güte der *Collectivism Scale* als mangelhaft beurteilt werden. Diese Einschätzung ergibt sich zum einen daraus, dass die Skala keine eindimensionale Messung im Sinne des Rasch-Modells aufweist und zum anderen aus der fehlenden Validität dieser Skala. In Anbetracht der Tatsache, dass es sich hierbei um ein bereits publiziertes Erhebungsinstrument handelt, das bereits in mehreren kulturvergleichenden Studien eingesetzt wurde, mag dieser Befund überraschen. Allerdings darf nicht vergessen werden, dass die *Collectivism Scale* ursprünglich nicht für den Einsatz bei Kinder- und Jugendlichen entwickelt wurde, weshalb es sich bei der vorliegenden Untersuchung um eine für diesen Einsatzbereich modifizierte Fassung handelt. So kann festgehalten werden, dass sich die *Collectivism Scale* zumindest in der vorliegenden Fassung nicht für den Einsatz bei Kinder- und Jugendlichen eignet. Dabei zeigten die Rasch-Modell Analysen, dass die Itemschwierigkeiten in beiden Kulturen insbesondere in Abhängigkeit des Alters variieren. In der österreichischen Stichprobe unterscheiden sich darüber hinaus gehäuft die Schwierigkeiten der Items zwischen Personen mit hoher und niedriger Ausprägung auf der Skala. Eine mögliche Erklärung für den Einfluss des Alters stellt die kognitive Anforderung des eingesetzten Verfahrens dar. Schüler müssen bei der Bearbeitung der *Collectivism Scale* einschätzen, inwieweit sie bestimmte Verhaltensweisen im Schulalltag zeigen. Dies erfordert allerdings ein bestimmtes Maß an Selbstreflexion, das sich erst im Laufe der kognitiven Entwicklung ausprägt. Dementsprechend könnten jüngere Schüler mit dem Bearbeiten der *Collectivism Scale* überfordert, zumindest nicht in der Lage gewesen sein die Items adäquat zu beantworten. Die Abhängigkeit der Itemschwierigkeiten von der Ausprägung auf der Skala könnte durch Antworttendenzen bedingt sein. Dies wäre dann der Fall, wenn Kollektivismus mit bestimmten Antworttendenzen einhergeht, sodass kollektivistische Schüler die Items beispielsweise besonders sozial erwünscht beantworten. Dass sich dieses Phänomen vorwiegend nur in der österreichischen Stichprobe zeigt, kann damit erklärt werden, dass in der kollektivistisch orientierten japanischen Stichprobe alle annähernd die gleiche Ausprägung von sozialer Erwünschtheit zeigen. In Bezug auf die Validität der *Collectivism Scale* ist zunächst festzuhalten, dass diese Skala

nach den vorliegenden Befunden zufolge scheinbar nicht in der Lage ist Kollektivismus zu erfassen. Allerdings müssen die Ergebnisse differenzierter betrachtet werden, denn für den mangelnden Nachweis der Validität lassen sich drei Erklärungen finden: (a) das Kriterium zur Überprüfung der Validität war ungeeignet, (b) aufgrund der fehlenden Skalenäquivalenz konnte die Validität nicht nachgewiesen werden oder (c) die *Collectivism Scale* weist tatsächlich keine Validität auf. Im Sinne der Extremgruppenvalidierung müsste Kollektivismus in der japanische Stichprobe hoch ausgeprägt sein, während die österreichische Stichprobe über eine geringe Ausprägung verfügt. Denkbar wäre es, dass (a) sich beide Stichproben nicht in der Ausprägung von Kollektivismus unterscheiden. Schließlich gibt es in der Literatur keine Untersuchungen bei Kinder- und Jugendlichen zur Ausprägung von Individualismus oder Kollektivismus. Unter diesen Umständen wäre aufgrund des ungeeigneten Kriteriums die Extremgruppenvalidierung nicht in der Lage die Validität der *Collectivism Scale* nachzuweisen. Für den Validierungsprozess wurden Mittelwerte über alle Items der *Collectivism Scale* gebildet. Zu bedenken ist jedoch, dass (b) der nachfolgende interkulturelle Vergleich wegen der fehlenden Skalenäquivalenz methodisch gesehen nicht zulässig ist. Aufgrund der fehlenden skalaren Invarianz haben sich möglicherweise dadurch tatsächlich vorhandene Unterschiede aufgehoben, wodurch diese in den anschließenden Analysen nicht mehr nachweisbar waren. Zuletzt muss auch in Betracht gezogen werden, dass die *Collectivism Scale* (c) tatsächlich keine Validität aufweist. In diesem Fall würde sich die Frage stellen, was diese Skala bei einer zufrieden stellenden Reliabilität dann tatsächlich erfasst. Diese und die Frage nach der Ursache des fehlenden Validitätsnachweises lassen sich in Rahmen dieser Arbeit nicht klären.

(2) *Individualism Difference Scale* als Erhebungsinstrument für Individualismus bei Kinder- und Jugendlichen. Die *Individualism Difference Scale* verfolgt einen innovativen Ansatz zur Erhebung von Individualismus bei der der Absolutbetrag der Differenz zwischen der eigenen Meinung und der wahrgenommener Klassennorm für die Ausprägung von Individualismus steht. Nach dem Validitätsbefund dieser Arbeit zufolge, kann mit Hilfe dieses Ansatzes tatsächlich ausreichend valide Individualismus erfasst werden. Zudem konnte die *Individualism Difference Scale* in Hinblick auf die Äquivalenz als dichotome Skala die höchste Äquivalenzebene und als mehrkategorielle Skala zumindest partielle Messeinheitsäquivalenz erreichen. Diese Befunde sprechen für den neuartigen

Ansatz der *Individualism Difference Scale*. Andererseits ist die geringe Reliabilität der *Individualism Difference Scale* in der österreichischen Stichprobe kritisch zu betrachten. Denn dies bedeutet nicht nur, dass in dieser Teilstichprobe nicht hinreichend genau gemessen wird, sondern auch dass der Messfehler in Zusammenhang mit der untersuchten Stichprobe steht. Die möglichen Ursachen für die unterschiedliche Höhe der Reliabilität, wie z. B. Probleme bei der Vorgabe der Erhebungsinstrumente, unterschiedliche Antworttendenzen etc. (siehe Van de Vijver & Leung, 1997) sind demnach noch zu klären. Eine plausible Erklärung ergibt sich aus der Tatsache, dass die Itemschwierigkeiten eine äußerst geringe Streuung aufweisen. So kann angenommen werden, dass die *Individualism Difference Scale* zu wenige Items enthält, die im Ausprägungsbereich der österreichischen Stichprobe diskriminieren können. Weiters steht in diesem Zusammenhang auch die bekannten Probleme der klassischen Veränderungsmessung, allen voran die geringe Reliabilität von Differenz-Scores. Das linear-logistische Testmodell (LLTM) von Fischer (1974) zur Veränderungsmessung würde eine Möglichkeit darstellen, diese Probleme zu überwinden. Insofern könnte die Anwendung des LLTM statt Bildung der Differenzwerte ein möglicher Ansatz sein, um eine höhere Reliabilität der *Individualism Difference Scale* in beiden Teilstichproben zu erlangen.

(3) *Individualism Difference Scale* als Erhebungsinstrument für Individualismus bei Kinder- und Jugendlichen. Die Äquivalenzprüfung der *Collectivism Scale* und der *Individualism Difference Scale* erfolgten sowohl auf Basis der probabilistischen Testtheorie (Rasch-Modell) als auch der klassischen Testtheorie (konfirmatorische Mehrgruppen-Faktorenanalyse). Vergleicht man die Ergebnisse beider Analysemethoden, so fällt auf, dass diese sowohl der *Collectivism Scale* als auch der *Individualism Difference Scale* unterschiedliche Ebenen der Äquivalenz attestieren. Bei der *Individualism Difference Scale* mag diese Inkonsistenz nicht überraschen, da die Analysen auf unterschiedlicher Datenbasis erfolgten, nämlich einmal als dichotome Skala und ein andermal als mehrkategoriale Skala. Bei der *Collectivism Scale* hingegen konnte mittels Rasch-Modell nicht die Ebene der Strukturäquivalenz erreicht werden, während mittels konfirmatorischer Mehrgruppen-Faktorenanalyse die Ebene der Messeinheitsäquivalenz erreicht wurde. Dies spricht dafür, dass die Rasch-Modell Analysen wesentlich strengere Anforderungen an das Messinstrument stellen als die konfirmatorische Mehrgruppen-Faktorenanalyse. Fraglich ist in diesem Zusammenhang, ob diese Anforderungen im Rahmen der Einstellungsmessung noch realistisch sind.

Schließlich weist Kubinger (2005b) darauf hin, dass die Konstruktion von Persönlichkeitsfragebogen nach dem *Rasch*-Modell problematisch ist, da die interessierende Eigenschaft stets von weiteren Eigenschaften (z. B. Antworttendenzen) überlagert ist. Eine wesentliche Stärke der *Rasch*-Modell Analysen stellt jedoch die Prüfbarkeit bestimmter Modellannahmen dar, wodurch sich einzelne nicht modellkonforme Items identifizieren lassen. Darüber hinaus besteht die Möglichkeit aufgrund der verwendeten Teilungskriterien Hypothesen zu den Ursachen der Modellabweichung aufzustellen. Im Gegensatz dazu erfolgen die Analysen bei der konfirmatorischen (Mehrgruppen-) Faktorenanalyse auf Grundlage der Modellanpassung (*Model-Fit*) bzw. anhand von Modellvergleichen. Daraus resultieren zumindest zwei Probleme: (a) Aufgrund der arbiträren Cut-Off-Werten kommt es nicht selten vor, dass beispielsweise der RMSEA auf eine akzeptable Modellanpassung hindeutet, während der CFI zur Ablehnung des Modells rät. So bleibt die Beurteilung eines Modells oftmals der Willkür des Forschers überlassen. Ein ähnliches Problem ergibt sich auch (b) bei Modellvergleichen, die Grundlage der Äquivalenzprüfung darstellen. Nach dem strengen Vorgehen der Äquivalenzprüfung dürften sich die Modelle bei Beschränkung der Ladungen bzw. der Intercepts nicht signifikant verschlechtern. Bei diesem Vorgehen besteht einerseits das Problem, dass ein Modell eher die Möglichkeit hat sich zu verschlechtern je besser das Ausgangsmodell passt (siehe z. B. Kapitel 6.2.2.2). Andererseits führt dieses Vorgehen dazu, Modelle bei signifikanter Verschlechterung trotz akzeptabler Modellanpassung zu verwerfen. Dabei stellt sich die Frage, ob ein Modell bei signifikanter Verschlechterung tatsächlich nicht über die jeweilige Äquivalenzebene verfügt, wenn die Modellanpassung für sich genommen im akzeptablen Bereich liegt (siehe z. B. Kapitel 6.2.2.3). Aus diesen Gründen kann festgehalten werden, dass das Vorgehen bei der Äquivalenzprüfung mittels Modellvergleichen womöglich nicht die optimale Strategie darstellt. Strohmeier (persönl. Mitteilung, 30.03.09) schlägt in diesem Zusammenhang eine elegante Alternative vor, nämlich die Äquivalenzprüfung mittels *Latent mean structures* (Byrne, 2001) bei gleichzeitiger Validitätsprüfung. Dabei wird ein Modell beurteilt, dessen Ladungen und Intercepts über die Gruppen hinweg beschränkt werden und lediglich die Fehlerterme und die latenten Mittelwerte der Gruppen variieren lassen. Weist dieses Modell eine akzeptablen *Model-Fit* auf, so stellt dies den Nachweis der skalaren Äquivalenz auf. Darüber hinaus kann im Sinne einer

Extremgruppenvalidierung beurteilt werden, ob das Messinstrumente in der Lage ist Mittelwertsunterschiede zwischen Gruppen aufzudecken.

Zusammenfassend können aus den Ergebnissen der vorliegenden Arbeit folgende Erkenntnisse und Empfehlungen abgeleitet werden. Wie die Analysen zur *Collectivism Scale* gezeigt haben, ist die traditionelle Erhebungsmethode mittels Likert-Skalen nur eingeschränkt bei Kinder- und Jugendlichen zu empfehlen. Die neuartige Erhebungsmethode, die der *Individualism Difference Scale* zugrunde liegt erscheint in diesem Zusammenhang eine verfolgenswerte mögliche Alternative. Als Weiterentwicklung bedarf es jedoch an Items, die die Bandbreite an Itemschwierigkeiten ausfüllen, wodurch sich möglicherweise das Problem der geringen Reliabilität erübricht. In Hinblick auf die Äquivalenzprüfung bedarf es weiterer Untersuchungen, inwiefern Rasch-Modell Analysen trotz methodischer Stringenz für die Entwicklung von Erhebungsinstrumenten für Individualismus bzw. Kollektivismus geeignet sind. Dies gilt auch auf die Äquivalenzprüfung mittels konfirmatorischer Mehrgruppen-Faktorenanalyse bei der das Vorgehen anhand Modellvergleiche kritisch zu betrachten ist.

Zusammenfassung

Die vorliegende Arbeit befasst sich mit der psychometrischen Äquivalenzprüfung zweier Messinstrumente für Kinder- und Jugendliche. Zum einen handelt es sich dabei um eine modifizierte Fassung der *Collectivism Scale* von Yamaguchi (1995) zur Erfassung von Kollektivismus. Zum anderen um die *Individualism Difference Scale*, einer neu entwickelten Skala zur Erfassung von Individualismus. Beide Erhebungsinstrumente wurden bei 277 Schülern in Österreich im Alter von 8 und 15 Jahren und 532 Schülern in Japan im Alter zwischen 9 und 15 Jahren in der jeweiligen Übersetzung vorgegeben. Dadurch ergab sich die Notwendigkeit der Prüfung auf Äquivalenz, die in drei Ebenen eingeteilt werden kann: strukturelle Äquivalenz, Messeinheitsäquivalenz und Skalenäquivalenz. Die Äquivalenzprüfung erfolgte einerseits mit Hilfe des Rasch-Modells und andererseits mittels konfirmatorischer Mehrgruppen-Faktorenanalyse. Die Analysen auf Basis des Rasch-Modells zeigten, dass die faktorielle Validität der *Collectivism Scale* sowohl in der österreichischen als auch in der japanischen Stichprobe nicht gegeben war. Infolgedessen erüblichte sich die Äquivalenzprüfung nach dieser Auswertungsmethode. Die *Individualism Difference Scale* als dichotome Skala erreichte hingegen die Ebene der Skalenäquivalenz, was die interkulturelle Vergleichbarkeit der erhobenen Daten belegt. Die Analysen anhand der konfirmatorischer (Mehrgruppen-)Faktorenanalyse jedoch belgten nicht nur die faktorielle Validität der *Collectivism Scale* in den beiden Teilstichproben, sondern konnten auch die Messeinheitsäquivalenz dieser Skala nachweisen. Die *Individualism Difference Scale* als mehrkategorielle Skala hingegen erreichte die partielle Messeinheitsäquivalenz. Die anschließenden testtheoretischen Analysen ergaben eine ausreichend hohe Realibilität für die *Collectivism Scale*. Allerdings konnte in der nachfolgenden Extremgruppenvalidierung kein Nachweis für die Validität dieser Skala erbracht werden, sodass diese Skala nach den vorliegenden Ergebnissen zufolge praktisch über keine Validität verfügt. Die Analysen der *Individualism Difference Scale* zeigten, dass zwar für die Gesamtstichprobe eine zufrieden stellend hohe Reliabilität erreicht wird, der Reliabilitätskoeffizient jedoch in der österreichischen Stichprobe signifikant geringer ausfällt als in der japanischen Stichprobe. Die anschließenden Analysen zur Validität konnten eine ausreichend hohe Validität der *Individualism Difference Scale* nachweisen.

Für nachfolgende Untersuchungen wird empfohlen weitere Items für die *Individualism Difference Scale* mit geringen Itemschwierigkeiten zu entwickeln.

Weiters bedarf es einer kritischen Reflexion in Bezug auf die Methoden der Äquivalenzprüfung. Zum einen in Hinblick auf die Durchführbarkeit der Konstruktion von Erhebungsinstrumenten für Individualismus bzw. Kollektivismus nach dem Rasch-Modell. Zum anderen in Bezug auf die Äquivalenzprüfung mittels konfirmatorischen Mehrgruppen-Faktorenanalyse, dessen Vorgehen mit Hilfe von Modellvergleichen in Frage gestellt wird. Hierbei sollen alternative Vorgehensweisen wie etwa mit Hilfe von *Latent mean structures* in Betracht gezogen werden.

Literaturverzeichnis

- ALESI, M. & PEPI, A. (2008). Individualism-collectivism and school achievement in Italian students. *International Journal of Psychology*, 43, 397.
- ANDERSEN, E. B. (1973). A goodness of fit test for the Rasch model. *Psychometrika*, 38, 123-140.
- ARBUCKLE, J. L. (2007). *AMOS 16.0 User's Guide*. Chicago: Smallwaters Corporation.
- BACKHAUS, K., ERICHSON, B., PLINKE, W. & WEIBER, R. (2008). *Multivariate Analysemethoden*. Heidelberg: Springer.
- BEAUDUCEL, A. & WITTMANN, W. W. (2005). Simulation study on fit indices in confirmatory factor analysis based on data with slightly distorted simple structure. *Structural Equation Modeling*, 12, 41-75.
- BECKER-CARUS, C. (2003). *Allgemeine Psychologie: Eine Einführung*. Heidelberg: Spektrum Akademischer Verlag.
- BENTLER, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107, 238-246.
- BENTLER, P. M. (1995). *EQS structural equation with latent variables*. Newbury Park, CA: Sage.
- BHAWUK, D. P. S. & BRISLIN, R. W. (1992). The measurement of intercultural sensitivity using the concepts of individualism and collectivism *International Journal of Intercultural Relations*, 16, 413-436.
- BOESCH, E. E. & STRAUB, J. (2007). Kulturpsychologie - Prinzipien, Orientierungen, Konzeptionen. In G. Trommsdorff (Hrsg.), *Theorien und Methoden der kulturvergleichenden Psychologie* (S. 25-96). Göttingen: Hogrefe.
- BOLLEN, K. A. (1989). *Structural equations with latent variables*. Oxford: John Wiley and Sons.
- BOND, M. H. (2002). Reclaiming the individual from Hofstede's ecological analysis - a 20-year odyssey: Comment on Oyserman et al. (2002). *Psychological Bulletin*, 128, 73-77.
- BOROWSKY, R., BARTH, F., SHWEDER, R. A., RODSETH, L. & STOLZENBERG, N. M. (2001). When: A conversation about culture. *American Anthropologist*, 103, 432-446.

- BORTZ, J. & DÖRING, N. (2006). *Forschungsmethoden und Evaluation: für Human- und Sozialwissenschaftler*. Berlin: Springer.
- BROWNE, M. W. & CUDECK, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Hrsg.), *Testing structural equation models* (S. 136-162). Newbury Park, CA: Sage.
- BÜHNER, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion*. München: Pearson Studium.
- BYRNE, B. M. (2001). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. Mahwah, NJ: Lawrence Erlbaum Associates.
- BYRNE, B. M., SHAVELSON, R. J. & MUTHÉN, B. O. (1989). Testing for the equivalence of factor covariance and mean structures: The issues of partial measurement invariance. *Psychological Bulletin*, 105, 456-466.
- CHEN, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14, 464-504.
- CHEN, F. F. & WEST, S. G. (2008). Measuring individualism and collectivism: The importance of considering differential components, reference groups, and measurement invariance. *Journal of Research in Personality*, 42, 259-294.
- CHEUNG, G. W. & RENSVOOLD, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9, 233-255.
- CIRKOV, V. I., LYNCH, M. & NIWA, S. (2005). Application of the scenario questionnaire of horizontal and vertical individualism and collectivism to the assessment of cultural distance and cultural fit. *International Journal of Intercultural Relations*, 29, 469-490.
- COHEN, D. (2007). Methods in cultural psychology. In S. Kitayama (Hrsg.), *Handbook of cultural psychology*. New York: Guilford Press.
- CRONBACH, L. J. (1990). *Essentials of psychological testing*. New York: HarperCollins.
- CROSS, S. E. & MARKUS, H. R. (1991). Cultural adaptation and the self: Self-construal, coping, and stress *Annual meeting of the American Psychological Association, San Francisco*.
- FISCHER, G. H. (1974). *Einführung in die Theorie psychologischer Tests*. Bern: Huber.

- FISCHER, G. H. (1999). Spezifische Objektivität: Eine wissenschaftstheoretische Grundlage. In K. D. Kubinger (Hrsg.), *Moderne Testtheorie - Ein Abriss samt neusten Beiträgen* (S. 87-111). München: Beltz.
- FISCHER, G. H. & PONOCNY-SELIGER, E. (1998). *Structural Rasch modeling. Handbook of the usage of LPCM-WIN 1.0*. Groningen: ProGAMMA.
- FISCHER, G. H. & SCHEIBLECHNER, H. (1970). Algorithmen und Programme für das probabilistische Testmodell von Rasch. *Psychologische Beiträge*, 12, 23-51.
- FISKE, A. P. (2002). Using individualism and collectivism to compare cultures - A critique of the validity and measurement of the constructs: Comment on Oyserman et al. (2002). *Psychological Bulletin*, 128, 78-88.
- GLAS, C. A. W. & VERHELST, N. D. (1995). Testing the Rasch model. In G. H. Fischer & I. W. Molenaar (Hrsg.), *Rasch models - Foundations, recent developments, and applications* (S. 69-95). New York: Springer.
- GOETHALS, R. (1994). *Die praktische Erprobung von Alternativen zur multiple-choice-Vorgabe bei Computertests*: Unveröffentlichte Dissertation, Universität Wien.
- GÖTZ, D. & WELLMANN, H. (2009). *Langenscheidt Power Wörterbuch Deutsch*. München: Langenscheidt.
- GREEN, E. G. T., DESCHAMPS, J.-C. & PAEZ, D. (2005). Variation of individualism and collectivism within and between 20 countries: A typological analysis. *Journal of Cross-Cultural Psychology*, 36, 321-359.
- HOFSTEDE, G. (1980). *Culture's Consequences*. Newbury Park, CA: Sage
- HOFSTEDE, G. (1994). Foreword. In U. Kim, H. C. Triandis, C. Kagitcibasi, S. C. Choi & G. Yoon (Hrsg.), *Individualism and collectivism: Theory, method, and application*.
- HOFSTEDE, G. (2001). *Culture's Consequences*. Newbury Park, CA: Sage.
- HOFSTEDE, G. (2007). Der kulturelle Kontext psychologischer Prozesse. In G. Trommsdorff (Hrsg.), *Theorien und Methoden der kulturvergleichenden Psychologie*. Göttingen: Hogrefe.
- HOFSTEDE, G. & BOND, M. H. (1984). Hofstede's culture dimensions: An independent validation using Rokeach's value survey. *Journal of Cross-Cultural Psychology*, 15, 417-433.

- HOMBURG, C., KLARMANN, M. & PFESSER, C. (2008). Konfirmatorische Faktorenanalyse. In A. Herrmann, C. Homburg & M. Klarmann (Hrsg.), *Handbuch Marktforschung: Methoden - Anwendungen - Praxisbeispiele* (S. 271-303). Wiesbaden: Gabler.
- HOPPE, M. H. (2004). An interview with Geert Hofstede. *Academy of Management Executive*, 18, 75-79.
- HORN, J. L. & MCARDLE, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, 18, 117-144.
- HU, L. & BENTLER, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Hrsg.), *Structural equation modeling. Concepts, issues, and applications* (S. 76-99). London: Sage.
- HU, L. & BENTLER, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3, 424-453.
- HU, L. & BENTLER, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.
- HUI, C. H. (1988). Measurement of individualism-collectivism. *Journal of Research in Personality*, 22, 17-36.
- HUI, C. H. & TRIANDIS, H. C. (1986). Individualism-collectivism: A study of cross-cultural researchers. *Journal of Cross-Cultural Psychology*, 17, 225-248.
- INGELHART, R. (1997). *Modernization and postmodernization: Cultural, economic, and political change in 43 societies*. Princeton, NJ: Princeton University Press.
- INGELHART, R., BASANEZ, M. & MORENO, A. (1998). *Human values and beliefs: A cross-cultural sourcebook: Political, religious, sexual, and economic norms in 43 societies. Findings from the 1990-1993 World Values Survey*. Ann Arbor, MI: The University of Michigan Press.
- JAHODA, G. (2007). Kulturkonzepte im Wandel. In G. Trommsdorff (Hrsg.), *Theorien und Methoden der kulturvergleichenden Psychologie* (S. 3-24). Göttingen: Hogrefe.

- JANKISZ, E. & MOOSBRUGGER, H. (2007). Planung und Entwicklung von psychologischen Tests und Fragebogen. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 27-72). Heidelberg: Springer.
- JÖRESKOG, K. G. (1963). *Statistical estimation in factor analysis*. Stockholm: Almqvist & Wiksell.
- JÖRESKOG, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34, 183-202.
- JÖRESKOG, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, 36, 409-426.
- JÖRESKOG, K. G. & SÖRBOM, D. (1996). *LISREL 8: User's Reference Guide*. Lincolnwood, IL: Scientific Software international, Inc.
- KAGITCIBASI, C. (1997). Individualism and collectivism. In J. W. Berry (Hrsg.), *Handbook of cross-cultural psychology* (S. 1-49). Boston: Allyn & Bacon.
- KIM, U., TRIANDIS, H. C., KAGITCIBASI, C., CHOI, S. C. & YOON, G. (1994). *Individualism and collectivism: Theory, method, and application*. Thousand Oaks, CA: Sage.
- KITAYAMA, S. (2002). Culture and basic psychological processes-toward a system view of culture: Comment on Oyserman et al. (2002). *Psychological Bulletin*, 128, 89-96.
- KLINE, R. B. (2005). *Principles and Practices of Structural Equation modeling*. New York: The Guilford Press.
- KUBINGER, K. D. (1999). Aktueller Stand und kritische Würdigung der Probabilistischen Testtheorie. In K. D. Kubinger (Hrsg.), *Moderne Testtheorie. Ein Abriss samt neuesten Beiträgen* (S. 19-83). Weinheim: Beltz.
- KUBINGER, K. D. (2004). Probabilistische Testtheorie. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 415-423). Weinheim: Beltz.
- KUBINGER, K. D. (2005a). Psychological test calibration using the Rasch model - Some critical suggestions on traditional approaches. *International Journal of Testing*, 5, 377-394.
- KUBINGER, K. D. (2005b). *Psychologische Diagnostik: Theorie und Praxis psychologischen Diagnostizierens*. Göttingen: Hogrefe.

- KUBINGER, K. D. (2009). *Adaptives Intelligenz Diagnostikum - Version 2.2 (AID2) samt AID 2-Türkisch* Göttingen: Beltz.
- KUBINGER, K. D. & DRAXLER, C. (2007). Probleme bei der Testkonstruktion nach dem Rasch-Modell. *Diagnostica*, 53, 131-143.
- KUBINGER, K. D., RASCH, D. & MODER, K. (2009). Zur Legende der Voraussetzungen des t-tests für unabhängige Stichproben. *Psychologische Rundschau*, 60, 26-27.
- LA DU, T. & TANAKA, J. S. (1989). Influence of sample size, estimation method, and model specification on goodness-of-fit assessments in structural equation models. *Journal of Applied Psychology*, 74, 625-635.
- LEE, H. (2004). The effects of individualism-collectivism cultural tendency on perception of school bullying victims. *International Journal of Psychology*, 39, 587-587.
- LI, F. & AKSOY, L. (2006). Dimensionality of individualism-collectivism and measurement equivalence of Triandis and Gelfand's scale. *Journal of Business and Psychology*, 21, 313-329.
- LIENERT, G. A. & RAATZ, U. (1998). *Testaufbau und Testanalyse*. Weinheim: Beltz.
- LONNER, W. J. & ADAMOPOULOS, J. (1997). Culture as antecedent to behavior. In J. W. Berry, Y. H. Poortinga & J. Pandey (Hrsg.), *Handbook of cross-cultural Psychology. Second Edition, Vol. 1: Theory and Method* (S. 43-83). Boston: Allyn and Bacon.
- MAIR, P. & HATZINGER, R. (2007). Extended Rasch modeling: The eRm package for the application of IRT models in R. *Journal of Statistical Software*, 20, 1-20.
- MARKUS, H. R. & HAMEDANI, M. G. (2007). Sociocultural psychology: The dynamic interdependence among self system and social system. In S. Kitayama (Hrsg.), *Handbook of cultural psychology* (S. 3-39). New York: Guilford Press.
- MARKUS, H. R. & KITAYAMA, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, 98, 224-253.
- MARSH, H. W., BALLA, J. R. & MCDONALD, R. P. (1988). Goodness-of-Fit Indexes in Confirmatory Factor Analysis: The effect of sample size. *Psychological Bulletin*, 103, 391-410.
- MARTIN-LOEF, P. (1973). *Statistika modeller* (Statistical models. Notes from seminars 1969-70 by Rolf Sunberg. 2nd ed.). Stockholm.

- MASTERS, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-174.
- MILLER, J. G. (1984). Culture and the development of everyday social explanation. *Journal of Personality and Social Psychology*, 81, 961-978.
- MILLER, J. G. (2002). Bringing culture to basic psychological theory - beyond individualism and collectivism: Comment on Oyserman et al. (2002). *Psychological Bulletin*, 128, 97-109.
- MOOSBRUGGER, H. & KELAVA, A. (2007). *Testtheorie und Fragebogenkonstruktion*. Heidelberg: Springer.
- MOOSBRUGGER, H. & SCHERMELLEH-ENGEL, K. (2007). Exploratorische (EFA) und Konfirmatorische Faktorenanalyse (CFA). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 308-324). Heidelberg: Springer.
- MORSBACH, H. (1980). Major psychological factors influencing Japanese interpersonal relations. In N. Warren (Hrsg.), *Studies in cross-cultural psychology* (S. 33-50). London: Academic Press.
- MUTHÉN, L. & MUTHÉN, B. O. (2004). *Mplus user's guide*. Los Angeles: Muthén & Muthén.
- NESDALE, D. & NAITO, M. (2005). Individualism-Collectivism and the attitudes to school bullying of Japanese and Australian students. *Journal of Cross-Cultural Psychology*, 36, 537-556.
- NISBETT, R. E. (2007). A psychological perspective: Cultural psychology-past, present, and future. In S. Kitayama (Hrsg.), *Handbook of cultural psychology* (S. 837-843). New York: Guilford Press.
- OYSERMAN, D. (2006). High power, low power, and equality: Culture beyond individualism and collectivism. *Journal of Consumer Psychology*, 16, 352-357.
- OYSERMAN, D., COON, H. M. & KEMMELMEIER, M. (2002). Rethinking individualism and collectivism: Evaluation of theoretical assumptions and meta-analyses. *Psychological Bulletin*, 128, 3-72.

- OYSERMAN, D. & LEE, S. (2007). Priming "culture": Cultural accommodation: Hybridity and the framing of social obligation. *Journal of Personality and Social Psychology*, 74, 1606-1618.
- OYSERMAN, D. & USKUL, A. K. (2008). Individualism and collectivism: Societal-level processes with implications for individual-level and society-level outcomes. In F. J. R. van de Vijver, D. A. van Hemer & Y. H. Poortinga (Hrsg.), *Multilevel analysis of individuals and cultures* (S. 145-173). New York: Lawrence Erlbaum Associates.
- POPPER, K. R. (1976). *Logik der Forschung*. Tübingen: Mohr.
- R DEVELOPMENTAL CORE TEAM. (2008). R: A language and environment for statistical computing (Version 2.7.2) [Computer software]. R Foundation for Statistical Computing, Vienna, Austria.
- RASCH, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press.
- RIETZ, C., RUDINGER, G. & ANDRES, J. (1996). Lineare Strukturgleichungsmodelle. In E. Erdfelder, R. Mausfeld & T. Meiser (Hrsg.), *Handbuch Quantitative Methoden* (S. 253-268). Weinheim: Beltz.
- RIPPL, S. & SEIPL, C. (2008). Fragebogenentwicklung. In S. Rippl & C. Seipl (Hrsg.), *Methoden kulturvergleichender Sozialforschung* (S. 97-114). Wiesbaden: VS Verlag für Sozialwissenschaften.
- RODGERS, J. L. (1999). The Bootstrap, the Jackknife, and the Randomization Test: A sampling Taxonomy. *Multivariate Behavioral Research*, 34, 441-456.
- ROST, J. (2004). *Lehrbuch Testtheorie - Testkonstruktion*. Bern: Huber.
- SALZBERGER, T., SINKOVICS, R. R. & SCHLEGELMILCH, B. B. (2001). Die Bedeutung der Datenäquivalenz in der internationalen Marketing- und Konsumentenforschung. *Jahrbuch der Absatz- und Verbrauchsforschung*, 2, 190-209.
- SCHERMELLEH-ENGEL, K., MOOSBRUGGER, H. & MÜLLER, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research-Online*, 8, 23-74.

- SCHIMMACK, U., SHIGEHIO, O. & DIENER, E. (2005). Individualism: A valid and important dimension of cultural differences between nations. *Personality and Social Psychology Review*, 9, 17-31.
- SCHWARTZ, S. H. (1994). Beyond individualism/collectivism: New cultural dimensions of values. In U. Kim, C. Triandis, C. Kagitcibasi, S. C. Choi & G. Yoon (Hrsg.), *Individualism and collectivism: Theory, method and applications* (S. 85-119). Thousand Oaks, CA: Sage.
- SCHWARTZ, S. H. & BARDI, A. (2001). Value hierarchies across culture: Taking a similarities perspective. *Journal of Cross-Cultural Psychology*, 32, 268-290.
- SCHWARTZ, S. H. & BILSKY, W. (1987). Toward a universal psychological structure of human values. *Journal of Personality and Social Psychology*, 53, 550-562.
- SCHWARTZ, S. H. & OYSERMAN, D. (2001). Asking questions about behavior: Cognition, communication, and questionnaire construction. *American Journal of Evaluation*, 22, 127-160.
- SEMIN, G. R. & RUBINI, M. (1990). Unfolding the concepts of personal by verbal abuse. *European Journal of Social Psychology*, 20, 463-474.
- SHANNON, P. (1986). Hidden within the pages: A study of social perspectives in young children's favourite books. *Reading Teachers*, 39, 656-663.
- SHULRUF, B., HATTIE, J. & DIXON, R. (2003). Development of a new measurement tool for individualism and collectivism NZARE/AARE Joint Conference 2003, Auckland, New Zealand.
- SINGELIS, T. M. (1994). The measurement of independent and interdependent self-construals. *Personality and Social Psychology Bulletin*, 20, 580-591.
- SINGELIS, T. M., TRIANDIS, H. C., BHAWUK, D. P. S. & GELFAND, M. J. (1995). Horizontal and vertical dimensions of individualism and collectivism: A theoretical and measurement refinement. *Cross-Cultural Research*, 29, 240-275.
- SINGH, J. (1995). Measurement issues in cross-national research. *Journal of International Business Studies*, 26, 597-619.
- SINHA, J. B. P. & VERMA, J. (1994). Social support as a moderator of the relationship between allocentrism and psychological well-being. In U. Kim, H. C. Triandis, C. Kagitcibasi, S. C. Choi & G. Yoon (Hrsg.), *Individualism and collectivism: Theory, method, and applications* (S. 267-275). Thousand Oaks, CA: Sage.

- SÖRBOM, D. (1989). Model modification. *Psychometrika*, 54, 371-384.
- STEENKAMP, H.-B. E. M. & BAUMGARTNER, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, 25, 78-90.
- STEIGER, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, 25, 51-62.
- STEVENS, S. S. (1966). Matching functions between loudness and ten other continua. *Perception and Psychophysics*, 1, 5-8.
- THE CHINESE CULTURE CONNECTION. (1987). Chinese values and the search for culture-free dimensions of culture. *Journal of Cross-Cultural Psychology*, 18, 143-164.
- THOMAS, A. (2004). *Kulturvergleichende Psychologie*. Göttingen: Hogrefe.
- TRIANDIS, H. C. (1972). *The analysis of subjective culture*. New York: Wiley.
- TRIANDIS, H. C. (1980). Values, attitudes, and interpersonal behaviour. In H. E. Howe & M. M. Page (Hrsg.), *Nebraska symposium on motivation*, 1979. Lincoln: University of Nebraska Press.
- TRIANDIS, H. C. (1995). *Individualism & Collectivism*. Boulder, CO: Westview Press.
- TRIANDIS, H. C. (1996). The psychological measurement of cultural syndromes. *American Psychologist*, 51, 407-415.
- TRIANDIS, H. C. (2004). The many dimensions of culture. *Academy of Management Executive*, 18, 88-93.
- TRIANDIS, H. C. (2007). Culture and Psychology: A history of the study of their relationship. In S. Kitayama (Hrsg.), *Handbook of cultural psychology* (S. 59-76). New York: Guilford Press.
- TRIANDIS, H. C., CHAN, D. K.-S., BHAWUK, D. P. S., IWAO, S. & SINHA, J. B. P. (1995). Multi-method probes of allocentrism and idiocentrism. *Journal of Psychology*, 30, 461-480.
- TRIANDIS, H. C. & GELFAND, M. J. (1998). Converging measurement of horizontal and vertical individualism and collectivism. *Journal of Personality and Social Psychology*, 74, 118-128.

- TRIANDIS, H. C., LEUNG, K., VILLAREAL, M. & CLACK, F. L. (1985). Allocentric vs. idiocentric tendencies: Convergent and discriminant validation. *Journal of Research in Personality*, 19, 395-415.
- TRIANDIS, H. C., MCCUSKER, C. & HUI, C. H. (1990). Multimethod probes of individualism and collectivism. *Journal of Personality and Social Psychology*, 59, 1006-1020.
- USKUL, A. K. & OYSEMAN, D. (2006). Question comprehension and response: Implications of individualism and collectivism. In B. Mannix, M. Neale & Y. Chen (Hrsg.), *Research on managing groups and teams: National culture & groups* (S. 173-201). Oxford: Elsevier Science Press.
- VAN DE VIJVER, F. J. & LEUNG, K. (1997). *Methods and data analysis for cross-cultural research*. Thousand Oaks, CA: Sage.
- VON DAVIER, M. (2001). WINMIRA (Version 2001) [Computer Software]. University Ave, St. Paul: Assessment Systems Corporation.
- VORNOV, M. & SINGER, J. A. (2002). The myth of individualism-collectivism: A critical review. *The Journal of Social Psychology*, 142, 461-480.
- WAGNER-MENGHIN, M. (2003). Äquivalenzprüfung. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 32-34). Weinheim: Beltz.
- WELCH, B. L. (1974). The generalisation of 'Student's' problem when several different population variances are involved. *Biometrika*, 34, 28-35.
- WRIGHT, B. D. & MASTERS, G. N. (1982). *Rating scale analysis: Rasch measurement*. Chicago: Mesa Press.
- WU, M., ADAMS, R. J., WILSON, M. & HALDANE, S. (2007). *ACER ConQuest, Version 2.0, Generalised Item Response Modeling Software*. Camberwell, Victoria: ACER Press, Australian Council for Educational Research Ltd.
- WUNDT, W. (1900-1914). *Völkerpsychologie*. Leipzig: Engelmann.
- YAMAGUCHI, S. (1994). Collectivism among the Japanese: A perspective from the self. In U. Kim, C. Triandis, C. Kagitcibasi, S. C. Choi & G. Yoon (Hrsg.), *Individualism and collectivism: Theory, method, and application* (S. 175-188). Thousand Oaks, CA.: Sage.

YAMAGUCHI, S., KUHLMAN, D. M. & SUGIMORI, S. (1995). Personality correlates of individualist and collectivist cultures. *Journal of Cross-Cultural Psychology*, 26, 658-672.

Anhang

Ad.) Kapitel 6.1.1.1 Österreichische Stichprobe

Tabelle 33. Z-Werte der Wald-type Tests und Differenzen der geschätzten Itemparameter für die *Collectivism Scale* in der österreichischen Stichprobe; Teilungskriterien Median, Geschlecht und Alter

ItemNr	Median		Geschlecht		Alter	
	z-Wert	Diff	z-Wert	Diff	z-Wert	Diff
1.c1	0,095	0,061	-0,806	-0,384	-1,652	-0,948
1.c2	-0,334	-0,214	-2,102	-1,052	-1,021	-0,584
1.c3	-0,381	-0,255	-1,227	-0,643	-1,240	-0,715
1.c4	1,220	0,927	-1,011	-0,638	1,194	0,799
2.c1	1,826	2,276	0,786	0,958		
2.c2	1,852	2,164	0,440	0,519		
2.c3	1,789	1,959	0,535	0,597		
2.c4	1,744	1,874	0,509	0,550		
3.c1			1,842	0,532	-1,116	-0,320
3.c2			0,526	0,235	0,798	0,352
3.c3			-0,164	-0,088	2,124	1,119
3.c4			-0,880	-0,865	0,997	0,956
4.c1						
4.c2						
4.c3						
4.c4						
5.c1						
5.c2						
5.c3						
5.c4						
6.c1	-0,658	-0,420	0,796	0,316	-1,256	-0,550
6.c2	-1,171	-0,753	-0,373	-0,168	-1,616	-0,758
6.c3	-0,996	-0,646	0,013	0,006	-0,463	-0,215
6.c4	-1,472	-1,129	-0,620	-0,334	0,299	0,156
7.c1	-1,272	-0,956	2,088	0,827	0,713	0,283
7.c2	-1,574	-1,148	0,986	0,420	0,207	0,085
7.c3	-2,331	-1,748	1,115	0,518	0,223	0,098
7.c4	-2,792	-2,547	0,998	0,526	0,944	0,465
8.c1						
8.c2						
8.c3						
8.c4						
9.c1	0,734	0,547	-0,075	-0,041	1,312	0,873
9.c2	1,083	0,759	-1,041	-0,588	-0,545	-0,313
9.c3	1,302	0,855	-1,488	-0,791	-1,011	-0,532
9.c4	1,191	0,805	-0,467	-0,250	-0,170	-0,088
10.c1						
10.c2						
10.c3						
10.c4						
11.c1	-0,821	-0,301	-0,968	-0,313	-0,260	-0,085
11.c2	0,225	0,098	-0,701	-0,274	0,486	0,188
11.c3	0,027	0,014	-2,474	-1,191	-0,156	-0,071
11.c4	0,405	0,323	-2,526	-1,747	2,627	1,683
12.c1	-1,606	-1,611	1,129	0,466	-1,899	-0,788
12.c2	-2,290	-2,203	1,685	0,736	0,627	0,273
12.c3	-2,598	-2,511	1,691	0,825	-0,460	-0,217
12.c4	-1,539	-1,480	1,560	0,856	2,326	1,234
13.c1						
13.c2						
13.c3						
13.c4						
14.c1	1,776	1,225	0,377	0,216	-0,419	-0,298
14.c2	1,419	0,898	0,939	0,537	-1,172	-0,784
14.c3	2,080	1,285	-0,050	-0,027	-1,662	-1,039
14.c4	2,827	1,848	-0,431	-0,244	-0,425	-0,262

Anmerkung. **Rot:** Item ausgeschlossen; **gelb:** „not well-conditioned“; **blau:** auffälliger z-Wert; **rosa:** bedeutsame Abweichung (Effektstärke); krit. Differenz Median: 0,61; krit. Differenz Geschlecht: 0,84; krit. Differenz Alter: 0,63.

Tabelle 34. Ausgeschiedene Items der *Collectivism Scale* in der österreichischen Stichprobe und Auffälligkeiten in den Teilungskriterien

Ausgeschieden	Auffälligkeiten
Item 4	Auffällig im Teilungskriterium <i>Alter</i> (relativ schwieriger für die höhere Altersgruppe)
Item 8	Auffällig im Teilungskriterium <i>Alter</i> (relativ schwieriger für die höhere Altersgruppe)
Item 10	Auffällig im Teilungskriterium <i>Alter</i> (relativ schwieriger für die höhere Altersgruppe)
Item 11	Auffällig im Teilungskriterium <i>Geschlecht</i> (relativ schwieriger für Burschen)
Item 13	Auffällig in den Teilungskriterien <i>Score</i> , <i>Geschlecht</i> und <i>Alter</i> (relativ schwieriger für die hohe Scoregruppe, Burschen und für die hohe Altersgruppe)

Ad.) Kapitel 6.1.1.2 Japanische Stichprobe

Tabelle 35. Z-Werte der Wald-type Tests und Differenzen der geschätzten Itemparameter für Collectivism Scale in der japanische Stichprobe; Teilungskriterien Median, Geschlecht und Alter

ItemNr	Median		Geschlecht		Alter	
	z-Wert	Diff	z-Wert	Diff	z-Wert	Diff
1.c1	0,496	0,273	0,584	0,206	-0,567	-0,205
1.c2	0,560	0,294	-0,616	-0,211	-1,102	-0,387
1.c3	-0,598	-0,342	-1,172	-0,466	0,937	0,378
1.c4	0,114	0,087	-2,695	-1,659	1,414	0,787
2.c1	1,097	0,719	0,386	0,155	-0,635	-0,266
2.c2	0,921	0,546	0,305	0,115	-1,255	-0,490
2.c3	-0,081	-0,049	0,141	0,056	-0,121	-0,050
2.c4	1,047	0,754	-1,255	-0,692	1,528	0,843
3.c1	-1,170	-0,496	0,440	0,117	-1,875	-0,516
3.c2	-0,970	-0,407	0,104	0,028	-2,300	-0,645
3.c3	-1,862	-0,971	-0,353	-0,131	-0,639	-0,240
3.c4	-0,804	-0,693	-1,624	-1,003	-0,784	-0,467
4.c1						
4.c2						
4.c3						
4.c4						
5.c1						
5.c2						
5.c3						
5.c4						
6.c1						
6.c2						
6.c3						
6.c4						
7.c1	0,210	0,103	1,134	0,357	-0,439	-0,140
7.c2	-0,282	-0,135	1,321	0,417	-1,471	-0,470
7.c3	-1,010	-0,563	0,009	0,004	-0,668	-0,268
7.c4	-0,577	-0,510	-0,410	-0,243	0,124	0,073
8.c1	-0,642	-0,390	0,646	0,204	-2,394	-0,807
8.c2	-0,912	-0,531	-0,183	-0,057	-2,175	-0,725
8.c3	-1,503	-0,916	-1,168	-0,425	-0,015	-0,006
8.c4	-1,142	-1,326	-1,714	-1,257	0,524	0,351
9.c1						
9.c2						
9.c3						
9.c4						
10.c1	0,318	0,247	1,726	0,748	-0,914	-0,395
10.c2	0,753	0,532	0,752	0,302	-0,865	-0,346
10.c3	0,087	0,061	1,579	0,650	1,077	0,441
10.c4	1,231	0,873	-0,455	-0,202	2,171	0,959
11.c1	-0,786	-0,432	2,009	0,653	-1,355	-0,441
11.c2	-0,166	-0,086	0,917	0,281	-0,186	-0,057
11.c3	-0,628	-0,344	0,203	0,072	0,097	0,034
11.c4	-0,500	-0,336	-0,575	-0,269	1,337	0,619
12.c1	1,143	0,537	-0,463	-0,169	0,260	0,097
12.c2	2,807	1,217	0,926	0,307	-1,444	-0,489
12.c3	0,787	0,384	0,776	0,296	-0,272	-0,105
12.c4	1,932	1,130	-1,766	-0,860	0,054	0,025
13.c1	0,387	0,416	1,139	0,636	0,988	0,524
13.c2	0,285	0,273	1,359	0,680	0,603	0,285
13.c3	-0,186	-0,175	1,703	0,882	1,118	0,552
13.c4	0,280	0,259	0,903	0,479	3,025	1,547
14.c1						
14.c2						
14.c3						
14.c4						

Anmerkung. **Rot:** Item ausgeschlossen; **blau:** auffälliger z-Wert; **rosa:** bedeutsame Abweichung (Effektstärke); krit. Differenz Median: 0,60; krit. Differenz Geschlecht: 0,65; krit Differenz Alter: 0,63.

Tabelle 36. Ausgeschiedene Items der *Collectivism Scale* in der japanischen Stichprobe und Auffälligkeiten in den Teilungskriterien

Ausgeschieden	Auffälligkeiten
Item 4	Auffällig in den Teilungskriterien <i>Geschlecht</i> und <i>Alter</i> (relativ schwieriger für Burschen und für die hohe Altersgruppe)
Item 5	Auffällig im Teilungskriterium <i>Alter</i> (relativ schwieriger die hohe Altersgruppe)
Item 6	Auffällig im Teilungskriterium <i>Geschlecht</i> (relativ schwieriger für Mädchen)
Item 9	Auffällig in den Teilungskriterien <i>Geschlecht</i> und <i>Alter</i> (relativ schwieriger für Burschen und für die hohe Altersgruppe)
Item 14	Auffällig in den Teilungskriterien <i>Score</i> , <i>Geschlecht</i> und <i>Alter</i> (relativ schwieriger für die hohe Scoregruppe, Burschen und für die hohe Altersgruppe)

Ad.) Kapitel 6.1.2 Individualism Difference Scale

Tabelle 37. Verteilung der absoluten Differenzwerte, getrennt nach japanischer und österreichischer Teilstichprobe

	Item	Absoluter Differenzwert					n
		0	1	2	3	4	
Japan	1	393	94	27	1	1	516
	2	398	87	23	1	3	512
	3	392	86	33	0	1	512
	4	371	102	27	4	3	507
	5	366	108	24	5	1	504
	6	372	94	32	7	3	508
	7	388	95	22	5	2	512
Österreich	1	131	85	50	5	1	272
	2	143	99	23	6	2	273
	3	147	83	26	8	7	271
	4	129	100	27	9	7	272
	5	121	102	38	6	2	269
	6	143	91	34	5	0	273
	7	131	84	43	11	2	271

Anmerkung. Aufgrund fehlender Angaben summiert sich n nicht auf die jeweilige Stichprobengröße.

Tabelle 38. Verteilung der Abweichung, getrennt nach japanischer und österreichischer Teilstichprobe

	Item	Abweichung		<i>n</i>
		Nein	Ja	
Japan	1	393	123	516
	2	398	114	512
	3	392	120	512
	4	371	136	507
	5	366	138	504
	6	372	136	508
	7	388	124	512
Österreich	1	131	141	272
	2	143	130	273
	3	147	124	271
	4	129	143	272
	5	121	148	269
	6	143	130	273
	7	131	140	271

Anmerkung. Aufgrund fehlender Angaben summiert sich *n* nicht auf die jeweilige Stichprobengröße.

Ad.) Kapitel 6.1.2.1 Österreichische Stichprobe

Tabelle 39. Z-Werte der Wald-type Tests und Differenzen der geschätzten Itemparameter für die *dichotome Individualism Difference Scale* in der österreichischen Stichprobe; Teilungskriterien Median, Geschlecht und Alter

ItemNr	Median		Geschlecht		Alter	
	z-Wert	Diff	z-Wert	Diff	z-Wert	Diff
1	-0,036	-0,009	-0,185	-0,044	-0,315	-0,076
2	-1,315	-0,338	0,196	0,047	-0,392	-0,094
3	-1,483	-0,381	0,906	0,218	0,175	0,042
4	0,417	0,104	-0,684	-0,164	-0,003	-0,001
5						
6	1,306	0,310	-0,798	-0,190	-0,237	-0,057
7	1,295	0,314	0,556	0,134	0,770	0,186

Anmerkung. **Rot:** Item ausgeschlossen; **blau:** auffälliger z-Wert; **rosa:** bedeutsame Abweichung (Effektstärke); krit. Differenz Median: 0,05; krit. Differenz Geschlecht: 0,04; krit. Differenz Alter: 0,03.

Tabelle 40. Ausgeschiedene Items der *Individualism Difference Scale* in der österreichischen Stichprobe und Auffälligkeiten in den Teilungskriterien

Ausgeschieden	Auffälligkeit
Item 5	Auffällig im Teilungskriterium <i>Geschlecht</i> (relativ schwieriger für Burschen)

Ad.) Kapitel 6.1.2.2 Japanische Stichprobe

Tabelle 41. Z-Werte der Wald-type Tests und Differenzen der geschätzten Itemparameter für die *dichotome Individualism Difference Scale* in der japanischen Stichprobe; Teilungskriterien Median, Geschlecht und Alter

ItemNr	Median		Geschlecht		Alter	
	z-Wert	Diff	z-Wert	Diff	z-Wert	Diff
1	2,136	0,569	0,087	0,020	-1,971	-0,447
2	-1,533	-0,549	-1,157	-0,266	-0,097	-0,022
3	1,056	0,300	0,755	0,172	1,065	0,243
4	0,967	0,269	-0,376	-0,085	0,225	0,051
5	-1,554	-0,525	1,190	0,269	0,420	0,095
6	0,216	0,062	0,316	0,071	-0,403	-0,091
7	-0,407	-0,127	-0,797	-0,181	0,749	0,171

Anmerkung. **blau:** auffälliger z-Wert; **rosa:** bedeutsame Abweichung (Effektstärke); krit. Differenz Median: 0,04; krit. Differenz Geschlecht: 0,03; krit. Differenz Alter: 0,03.

Ad.) Kapitel 6.1.2.3 Gesamtstichprobe

Tabelle 42. Z-Werte der Wald-type Tests und Differenzen der geschätzten Itemparameter für die *dichotome Individualism Difference Scale* in der Kalibrierungsstichprobe; Teilungskriterium Kultur

ItemNr	Kultur	
	z-Wert	Diff
1	-0,056	-0,011
2	-1,462	-0,285
3	0,531	0,103
4	0,686	0,132
5		
6	1,781	0,342
7	-1,438	-0,281

Anmerkung. **Rot:** Item ausgeschlossen; **rosa:** bedeutsame Abweichung (Effektstärke); krit. Differenz Kultur: 0,04.

Tabelle 43. Z-Werte der Wald-type Tests und Differenzen der geschätzten Itemparameter für die *dichotome Individualism Difference Scale* für die Kreuzvalidierungsstichprobe; Teilungskriterium Kultur

ItemNr	Kultur	
	z-Wert	Diff
1	-1,677	-0,505
2	1,281	0,387
3	0,171	0,051
4	-0,63	-0,190
5		
6	0,053	0,016
7	0,804	0,241

Anmerkung. **Rot:** Item ausgeschlossen; **rosa:** bedeutsame Abweichung (Effektstärke); krit. Differenz Kultur: 0,03.

Ad.) Kapitel 6.3.1.2 Konkurrente Validität

Tabelle 44. Leave-One-Out Kreuzvalidierung für die *Collectivism Scale*

	Vorhergesagte Gruppenzugehörigkeit		Gesamt (N = 808)
	Österreich	Japan	
Tatsächliche Gruppenzugehörigkeit			
Österreich	48.4% (134)	51.6% (143)	34.2% (277)
Japan	44.6% (237)	55.4% (294)	65.7% (531)

Ad.) Kapitel 6.3.2.2.1 Dichotome Individualism Difference Scale

Tabelle 45. Leave-One-Out Kreuzvalidierung für die *dichotome Individualism Difference Scale*

	Vorhergesagte Gruppenzugehörigkeit		Gesamt (N = 794)
	Österreich	Japan	
Tatsächliche Gruppenzugehörigkeit			
Österreich	61.7% (169)	38.3% (105)	34.5% (274)
Japan	27.3% (142)	72.7% (378)	65.5% (520)

Ad.) Kapitel 6.3.2.2.1 Mehrkategoriale Individualism Difference Scale

Tabelle 46. Leave-One-Out Kreuzvalidierung für die *mehrkategoriale Individualism Difference Scale*

	Vorhergesagte Gruppenzugehörigkeit		Gesamt (N = 795)
	Österreich	Japan	
Tatsächliche Gruppenzugehörigkeit			
Österreich	58.4% (160)	41.6% (114)	34.5% (274)
Japan	24.6% (128)	75.4% (393)	65.5% (521)

LEBENS LAUF

Angaben zur Person

Vorname:	Takuya	Geschlecht:	männlich
Zuname:	Yanagida	Geburtsdatum:	7. Jänner 1980
Adresse:	Süßenbrunnerstr. 29 2201 Gerasdorf	Geburtsort:	Wien
E-Mail:	takuya.yanagida@univie.ac.at	Familienstand:	ledig
		Staatsangeh.:	Japan

Berufserfahrung

02 / 2009-	Mitarbeiter am Projekt „Adaptives Intelligenz Diagnostikum-Englisch (AID2-Englisch)“, Leitung: Klaus D. Kubinger (Fakultät für Psychologie, Universität Wien)
01 / 2009- 02 / 2009	Freier Mitarbeiter am Projektantrag „Formation of relationships in immigrant youth: the environments, networks, dyads, and students“, Leitung: Christiane Spiel (Fakultät für Psychologie, Universität Wien)
11 / 2008-	IT-Trainer für SPSS am Zentralen Informatikdienst (ZID) der Universität Wien
03 / 2008- 05 / 2008	Mitarbeiter am Projekt „Anerkennung non-formalen und informellen Lernens an Universitäten“, Leitung: Margarete Lützenberger & Christiane Spiel (Fakultät für Psychologie, Universität Wien)
03 / 2008- 08 / 2008	Mitarbeiter am Projekt „Evaluation: EU-Lexikon“ im Auftrag des Instituts für Staatswissenschaften, Universität Wien
02 / 2008- 01 / 2009	Mitarbeiter am Projekt „Entwicklung von Kompetenz-Tests im Rahmen der Testung von nationalen Bildungsstandards in Österreich“, Test- und Beratungsstelle des Arbeitsbereichs Psychologische Diagnostik, Universität Wien
11 / 2007- 01 / 2008	Mitarbeiter am Projekt „Adaptives Intelligenz Diagnostikum (AID 2) Version 2.2“, Leitung: Klaus D. Kubinger (Fakultät für Psychologie, Universität Wien)
10 / 2007- 06 / 2008 03 / 2009-	Studienassistent am Institut für Wirtschaftspsychologie, Bildungspsychologie und Evaluation an der Fakultät für Psychologie der Universität Wien
07 / 2007- 08 / 2007	Praktikum an der Stabstelle für Evaluation und Qualitätsmanagement an der medizinischen Universität Wien
03 / 2007-	Freier Mitarbeiter am Center for Teaching and Learning (CTL) der Universität Wien
02 / 2007- 06 / 2007	Fachpraktikum in der Abteilung für Klinische Psychodiagnostik an der Universitätsklinik für Psychiatrie und Psychotherapie des AKH Wien
11 / 2006- 03 / 2007	Praktikum (Volontariat) am Institut für Wirtschaftspsychologie, Bildungspsychologie und Evaluation an der Fakultät für Psychologie der Universität Wien

Ausbildung

seit 2008- Bakkalaureatsstudium Statistik
seit 2001- Diplomstudium Psychologie an der Universität Wien

EDV-Kenntnisse

ECDL Core (MS-Office) AMOS, LISREL, R, SPSS	EndNote
--	---------

Sprachkenntnisse

Muttersprache: Japanisch
1. Fremdsprache: Deutsch
2. Fremdsprache: Englisch