

**Entwicklung eines Tests zur Erfassung
interkultureller Handlungskompetenz (TIHK)**

Julia Loboda

Diplomarbeit als Teil
der Diplomprüfung für Psychologen

Universität Regensburg

03.02.2003

Inhaltsverzeichnis

1	EINLEITUNG	1
2	ANALYSE BESTEHENDER TESTVERFAHREN	3
2.1	Darstellung ausgewählter Messinstrumente	3
2.1.1	Persönlichkeitstests	4
2.1.2	Selbsteinschätzungstests	5
2.1.3	Attributionstests	6
2.2	Zum Kriterienproblem interkultureller Diagnostik	8
2.3	Zum Methodenproblem interkultureller Diagnostik	9
2.4	Zusammenfassung der Analyse	10
3	THEORETISCHE GRUNDLAGEN FÜR DIE ENTWICKLUNG DES TIHK	11
3.1	Grundlagen interkultureller Psychologie	11
3.1.1	Kultur und Kulturstandards	11
3.1.2	Kulturelle Überschneidungssituation	12
3.1.3	Interkulturelles Lernen	14
3.2	Interkulturelle Handlungskompetenz	15
3.2.1	Begriffsabgrenzung	15
3.2.2	Die Forschungsproblematik bzgl. interkultureller Handlungskompetenz	16
3.2.3	Zum Forschungsparadigma interkultureller Handlungskompetenz	17
3.2.4	Definition interkultureller Handlungskompetenz	21
3.2.5	Einteilung interkultureller Handlungskompetenz in 3 Ebenen	21
3.3	Perspektivenwechsel als Indikator für interkulturelle Handlungskompetenz	23
3.3.1	Das Konzept des Perspektivenwechsels	23
3.3.2	Isomorphe Attribution (TRIANDIS, 1975, 1977)	25
4	KRITERIEN DER TESTENTWICKLUNG	27
4.1	Ableitungen aus der Analyse der Testverfahren	27
4.2	Ableitungen aus den theoretischen Grundlagen	27
4.2.1	Zur Erfassung des Perspektivenwechsels	28
4.2.2	Die Methode der kritischen Interaktionssituationen zur Erfassung des Perspektivenwechsels	29

5	DIE METHODE ZUR ENTWICKLUNG DES TIHK	31
5.1	Das Culture Assimilator Material	31
5.2	Die Methode der Paralleltestkonstruktion	33
5.2.1	Die Aufgabenanalyse.....	35
5.2.1.1	<i>Schwierigkeitsanalyse</i>	35
5.2.1.2	<i>Trennschärfenanalyse</i>	36
5.2.2	Die Auswertung der Daten.....	38
5.2.2.1	<i>Ermittlung der Rohwerte</i>	38
5.2.2.2	<i>Prüfung der Verteilung auf Normalität</i>	41
5.2.3	Die Bestimmung der Reliabilität.....	41
5.2.4	Der Paralleltestaufbau anhand der Analysedaten.....	42
5.2.5	Die Aufgabenpaarung.....	43
5.2.6	Die Prüfung der Äquivalenzkriterien von Paralleltests.....	44
5.2.7	Die Bestimmung der Validität.....	46
6	DIE PRAKTISCHE ENTWICKLUNGSARBEIT	48
6.1	Die Konstruktion der Testform A	48
6.1.1	Das Testmaterial.....	48
6.1.1.1	<i>Die Testaufgaben</i>	48
6.1.1.2	<i>Demographische Daten</i>	50
6.1.2	Durchführung der Datenerhebung.....	50
6.1.3	Analyse der Stichprobe Test A.....	50
6.1.4	Auswertung der Daten Test A.....	51
6.1.4.1	<i>Ergebnisse Test A</i>	51
6.1.4.2	<i>Aufgabenanalyse Test A</i>	51
6.1.4.3	<i>Überprüfung der Verteilung auf Normalität</i>	53
6.1.4.4	<i>Bestimmung der Reliabilität (Test A)</i>	54
6.1.5	Die Endform Test A.....	55
6.2	Die Konstruktion der Testform B	56
6.2.1	Das Testmaterial.....	56
6.2.2	Analyse der Stichprobe Test B.....	57
6.2.3	Auswertung der Daten Test B.....	58
6.2.3.1	<i>Ergebnisse Test B</i>	58
6.2.3.2	<i>Kriteriumsorientierte Messung</i>	59
6.2.3.3	<i>Der c-Koeffizient</i>	60
6.2.3.4	<i>Aufgabenanalyse Test B</i>	61
6.2.3.5	<i>Bestimmung der Reliabilität (Gesamt)</i>	63

6.2.4	Die Aufgabenpaarung	64
6.2.5	Die Prüfung der Äquivalenzkriterien	68
6.3	Darstellung des Gesamtergebnisses des TIHK.....	71
6.4	Validierung des TIHK	72
6.5	Interpretation der Testergebnisse	74
7	DISKUSSION.....	78
7.1	Die Bewertung der Methode aus dem Blickwinkel der theoretischen Modelle.....	78
7.2	Die Kritische Reflexion des Entwicklungsdesigns	79
7.2.1	Das Datenmaterial	79
7.2.2	Die Skalierung der Aufgaben.....	80
7.2.3	Die Objektivität des Tests	81
7.2.4	Die Parallelität der Testformen	82
7.2.5	Die Testkonstruktion	83
7.2.6	Die Rohwertverteilung	84
7.2.6.1	<i>Das Culture Assimilator Format</i>	<i>84</i>
7.2.6.2	<i>Die Aufgabenanordnung</i>	<i>85</i>
7.2.6.3	<i>Die Stichprobe</i>	<i>86</i>
7.2.6.4	<i>Das zu messende Konstrukt</i>	<i>88</i>
7.3	Die Praktikabilität des Tests.....	90
7.4	Ausblick auf weiterführende Forschung.....	90
7.5	Anwendungsmöglichkeiten in der Praxis	92
8	ZUSAMMENFASSUNG	94
9	LITERATUR.....	95
10	ANHANG.....	103

Tabellenverzeichnis

Tabelle 1:	Ausgewählte kritische Interaktionssituationen Test A (Vorform).....	49
Tabelle 2:	Ergebnis der Schwierigkeitsanalyse.....	51
Tabelle 3:	Ergebnis der Trennschärfenanalyse	52
Tabelle 4:	Items der Testform A.....	55
Tabellen 5-7:	Testaufgaben (Skalen A, B, C) zur Konstruktion von Testform B.....	56-57
Tabelle 8:	Umkodierung der Aufgaben und Zuordnung zu Kulturstandards.....	57
Tabelle 9:	Deskriptive Daten von Test A inklusive der Subskalen A, B, C	58
Tabelle 10:	Schwierigkeitsindices geordnet nach Kulturstandards.....	62
Tabelle 11:	Trennschärfe geordnet nach Kulturstandards.....	63
Tabelle 12:	Auswahl der Items für den Paralleltest aufgrund der Aufgabenpaarung	64
Tabelle 13:	Aufgabenpaarungen ohne die Items A3, A6, A12.....	65
Tabelle 14:	Aufgabenpaarungen ohne die Items A1, A3, A6, A12	65
Tabelle 15:	Aufgabenpaarungen ohne die Items A1, A3, A5, A6, A12.....	66
Tabelle 16:	Mittelwert (M), Standardabweichung (SD) und innere Konsistenz (c) der Parallelskalen A und B	67
Tabelle 17:	Schwierigkeitsindex (P) und Trennschärfe (c) der Skala A.....	67
Tabelle 18:	Schwierigkeitsindex (P) und Trennschärfe (c) der Skala B.....	67
Tabelle 19:	Der TIHK: Gesamtergebnis der Paralleltestkonstruktion	71
Tabelle 20:	Übereinstimmung der Test- und Rollenspielergebnisse im Rahmen der Validierung.....	73
Tabelle 21:	Häufigkeiten der Punkteverteilung in der Analysestichprobe des TIHK.....	75
Tabelle 22:	Zensierungsmodell nach HERBIG (1974) zur Interpretation der Testergebnisse des TIHK.....	76
Tabelle 23:	Häufigkeiten der Notenverteilung in der Analysestichprobe des TIHK	77
Tabelle 24:	Deskriptive Statistik der Gruppe der Psychologen und „Nichtpsychologen“ der Analysestichprobe von Test B.....	87

Abbildungsverzeichnis

Abbildung 1:	Ablaufschema der Paralleltestkonstruktion des TIHK.....	34
Abbildung 2:	Rohwertverteilung Test A.....	54
Abbildung 3:	Rohwertverteilung der Parallelskala A	69
Abbildung 4:	Rohwertverteilung der Parallelskala B	69

1 Einleitung

Im Zuge zunehmender internationaler Zusammenarbeit von Politik, Wirtschaft und Wissenschaft treffen Menschen unterschiedlicher Kulturen und Organisationen aufeinander. Die sich durch Internationalisierung der Märkte und der Entwicklung der EU verändernde gesellschaftliche Situation führt zu neuen Herausforderungen und Aufgaben für das Zusammenleben und Zusammenarbeiten über kulturelle Grenzen hinweg. Menschen sehen sich verstärkt mit Fremdheitserlebnissen konfrontiert, die bei unzulänglichem Umgang unabsehbare Folgen nach sich ziehen können. Daraus resultiert die Notwendigkeit einer interkulturellen Verständigung oder, weitergefasst, einer interkulturellen Orientierung. Damit ist gemeint, dass der Mensch fähig ist, effektiv mit einem fremdkulturellen Gegenüber zu interagieren (LANGE, 1994). Diese Fähigkeit wird insbesondere für Fach- und Führungskräfte, die im Ausland oder mit fremdkulturellen Arbeitskräften in Deutschland arbeiten zu einer unabdingbaren Schlüsselqualifikation. Da diese aber nicht von alleine entsteht, sondern der Unterstützung bedarf, ist es notwendig geeignete Lehr- und Lernprozesse zu organisieren.

An diesem Punkt setzt die Universität Regensburg in Kooperation mit der Fachhochschule Regensburg an und bietet das Zusatzstudium „Internationale Handlungskompetenz“ für Studierende aller Fachrichtungen im zweiten Studienabschnitt an. Ziel des Studiums ist es zum einen, Wissen über das eigene und über fremde Kultursysteme zu erwerben und sensibel für Unterschiede zu werden. Zum anderen sollen darauf aufbauend Fähigkeiten entwickelt werden, produktiv und sozialverträglich mit interkulturellen Problemen umzugehen.

Im Rahmen der Evaluation des Zusatzstudiums soll der Lernzuwachs der Studierenden und damit die Güte des Studienganges anhand eines interkulturellen Tests bewertet werden. Gegenstand der vorliegenden Arbeit ist die Entwicklung dieses Tests zur Erfassung interkultureller Handlungskompetenz (TIHK) mit einer parallelen Testform, die beide den testtheoretischen Gütekriterien von Reliabilität und Validität gerecht werden.

Interkulturelle Handlungskompetenz ist ein Sammelbegriff für spezifische Kenntnisse, Fähigkeiten und Fertigkeiten, die als Prädiktoren interkulturellen Handlungserfolgs fungieren (MÜLLER, 1992). Um welche Kenntnisse, Fähigkeiten und Fertigkeiten es sich jedoch hierbei im Konkreten handelt, darüber herrscht Uneinigkeit auf dem Gebiet der interkulturellen Forschung. Das Problem der Kriterien zur Vorhersage von Handlungserfolg im Ausland spiegelt sich auch auf der Ebene interkultureller Messinstrumente wider. Deshalb befasst sich diese Arbeit zunächst mit der Analyse bereits bestehender Testverfahren zur Erfassung interkultureller Handlungskompetenz. Diese werden auf Stärken und Schwächen untersucht und hinsichtlich ihrer Eignung als Messinstrument im Zusatzstudium „Internationale Handlungskompetenz“ diskutiert. Als

theoretische Grundlagen für die Entwicklung des Tests zur Erfassung interkultureller Handlungskompetenz (TIHK) werden die bisherigen Erkenntnisse der interkulturellen Psychologie herangezogen. Dazu werden theoretische Definitionsversuche des zu erfassenden Konstrukts interkultureller Handlungskompetenz sowie empirisch unterstützte Konzeptionen berücksichtigt.

Im Anschluss daran wird sukzessive die Entwicklung des TIHK in den darauf folgenden Kapiteln nachgezeichnet. Zur Konstruktion werden kritische Interaktionssituationen der Culture Assimilators USA (MÜLLER & THOMAS, 1995), England (SCHMID, 1999) und Indonesien (MARTIN, 2000) als Vertreter des westlichen, europäischen und östlichen Kulturraumes ausgewählt und nach testtheoretischen Gütekriterien an einer Analysestichprobe von Studenten ausgewertet. In Anlehnung an LIENERT UND RAATZ (1994) wird zu einer Testform A in einem zweiten Durchlauf eine nach statistischen Gesichtspunkten parallele Testform B konstruiert. Die Validierung des TIHK wird von GÖTZ (2003) vorgenommen. Hierauf wird zum Schluss dieser Arbeit Bezug genommen.

2 Analyse bestehender Testverfahren

Ein vor allem von der Wirtschaftsseite oft geäußelter Wunsch ist die Identifizierung von Prädiktoren, die internationalen Handlungserfolg von Mitarbeitern im Ausland vorhersagen. Die Forderung nach einem internationalen Auswahlverfahren oder zumindest nach einem Instrument zur Messung interkultureller Handlungskompetenz wird immer dringlicher. Bisher wurden jedoch Faktoren, die interkulturelle Effektivität voraussagen, nur mit mäßigem Erfolg identifiziert (DETWEILER, BRISLIN, MCCOMACK, 1983). Das liegt zum großen Teil daran, dass weder auf der Begriffsebene noch auf der Messebene Klarheit darüber herrscht, was erfolgreiches und interkulturell kompetentes Handeln ausmacht (KÜHLMANN, 1995). Die Forschungslage zu Auswahlverfahren für die Entsendung ins Ausland (Verfahren zur Erfassung interkultureller Handlungskompetenz) ist ebenso komplex und undurchsichtig wie das Konstrukt selber.

Um zu überprüfen ob bereits bestehende interkulturelle Testverfahren für die Evaluation des Studienganges „Internationale Handlungskompetenz“ an der Universität Regensburg in Betracht gezogen werden können, werden diese im Folgenden vorgestellt. Dabei werden ihre Stärken und Schwächen hinsichtlich eines Einsatzes in besagtem Zusatzstudiengang diskutiert.

Es bleibt vorab zu bemerken, dass überwiegend aus dem angloamerikanischen Raum Veröffentlichungen existieren, die wiederum unterschiedliche Aspekte interkultureller Handlungskompetenz erfassen. Die Forschungen, die auf diesem Gebiet im deutschsprachigen Raum vorliegen, werden ebenfalls berücksichtigt. Hierbei beschränken sich die Messinstrumente jedoch auf einzelne Übungen, die in interkulturellen Assessment Centern Verwendung finden (KÜHLMANN & Stahl, 1998) oder auf Tests, die zur Evaluation von interkulturellen Trainings eingesetzt werden (LAYES, 1995; HENTSCHEL, 2000).

2.1 Darstellung ausgewählter Messinstrumente

Bei der Sichtung der relevanten Literatur stößt man auf folgende Schwierigkeiten: Der größte Teil der interkulturellen Testverfahren ist kulturspezifisch, d.h., sie sind an einer bestimmten Stichprobe erhoben oder für eine spezifische Zielgruppe entwickelt worden. Das wiederum bedeutet, dass diese nur bei Mitgliedern einer speziellen Kultur, die in ein bestimmtes Gastland ausreisen, angewandt werden können. Die Instrumente interkultureller Handlungskompetenz sind größtenteils in den USA entwickelt worden und nur schlecht oder gar nicht auf

andere Länder übertragbar (vgl. LOO & SHIOMI, 1999; PRUEGER & ROGERS, 1993). Hier spielt vor allem auch die Problematik der adäquaten Übersetzung eine Rolle (vgl. GEISINGER, 1994; SPERBER, DEVELLIS & BOELECKE, 1994).

Weiterhin ist zu bemerken, dass unterschiedliche Tests auch unterschiedliche Aspekte interkultureller Handlungskompetenz (z.B. interkulturelle Anpassungsfähigkeit, interkulturelle Kommunikationskompetenz, Sensibilität bezüglich kultureller Unterschiede) erfassen und deswegen nur bedingt miteinander vergleichbar sind. Darüber hinaus verfolgen unterschiedliche Tests unterschiedliche Ziele, wie z.B. Evaluation von Trainings, Vorhersage von Auslandserfolg oder Bewertung des Wissensstandes und des Niveaus interkultureller Handlungskompetenz.

Aus den genannten Gründen wird in dieser Arbeit auf eine inhaltliche Einteilung der vorhandenen Testverfahren zugunsten einer methodischen Einteilung verzichtet. Grundsätzlich kann man aufgrund der gesichteten Literatur folgende Klassifizierung interkultureller Testverfahren vornehmen:

1. Persönlichkeitstests
2. Selbsteinschätzungstests
3. Attributionstests

Was unter diesen Tests zu verstehen ist, wird im Folgenden dargelegt und ein Instrument jeder Kategorie wird exemplarisch vorgestellt.

2.1.1 Persönlichkeitstests

Den größten Teil der Instrumente zur Erfassung interkultureller Handlungskompetenz bilden **Persönlichkeitstests**. Mit diesen werden Persönlichkeitseigenschaften getestet, die z.T. aus anerkannten Persönlichkeitsinventaren vor allem aus dem klinischen Kontext (wie z.B. Big Five Personality Inventory (BFI), Eysenck Personality Inventory (EPI) etc.) entliehen sind, oder die aus dem impliziten Verständnis interkultureller Handlungskompetenz der Autoren abgeleitet werden (Aufzählung von Persönlichkeitseigenschaften, wie z.B. Empathie, Offenheit, etc.). Mit Hilfe dieser Persönlichkeitseigenschaften werden Prognosen über das Niveau interkultureller Handlungskompetenz von Personen bzw. über die Erfolgswahrscheinlichkeit eines Auslandsaufenthaltes aufgestellt. An dieser Stelle wird nun beispielhaft für die Kategorie der Persönlichkeitstests das „Intercultural Sensitivity Inventory“ (ICSI) von BHAWUK UND BRISLIN (1992) vorgestellt.

Das „**Intercultural Sensitivity Inventory**“ (**ICSI**) von BHAWUK UND BRISLIN (1992) wurde mit dem Ziel entwickelt, die Fähigkeit von Personen zur Verhaltensmodifikation zu messen. Die Autoren gehen bei der Entwicklung des ICSI davon aus, dass Menschen im Ausland dann erfolgreich sind, wenn sie sich in die Lage von Menschen anderer Kulturen hineinversetzen können und bereit sind ihr Verhalten zu ändern. Besonders heben sie die Fähigkeit von Menschen hervor, eigenes Verhalten an kulturelle Gegebenheiten anpassen zu können. Dies werten sie als Zeichen interkultureller Sensitivität, die wiederum als Indikator für interkulturellen Erfolg gesehen wird. Als theoretische Konzepte ziehen sie die Dimensionen Individualismus und Kollektivismus von HOFSTEDE (1980) sowie „open-mindedness“ und „flexibility“ nach KEALEY UND RUBEN (1983) heran. Das Instrument besteht aus 46 Items zur Feststellung des Verhaltens in fremdkulturellen Situationen und misst a) *das Verständnis für unterschiedliche Verhaltensweisen abhängig von der Interaktion mit einer individualistischen oder kollektivistischen Kultur*, b) *die Offenheit hinsichtlich Unterschieden in der Begegnung mit anderen Kulturen* und c) *die Flexibilität, sich in ungewohnter und nicht vertrauter Umgebung anders zu verhalten*. Gute Reliabilitäts- und Validitätswerte sprechen für den Einsatz der Dimensionen Kollektivismus und Individualismus und eine Verbesserung des Tests durch die Kombination mit den Skalen „open-mindedness“ und „flexibility“. Der ICSI kann darüber hinaus effektiv hoch interkulturelle von durchschnittlich interkulturellen Personen unterscheiden.

In Bezug auf das Zusatzstudium „Internationale Handlungskompetenz“ ist hier zu konstatieren, dass das ICSI methodischen Anforderungen durchaus gerecht wird. Inhaltlich gesehen basiert dieser Test allerdings auf den theoretischen Annahmen von HOFSTEDE (1980), so dass sich das Testdesign nicht ohne weiteres auf Verfahren mit anderen theoretischen Hintergründen übertragen lässt. Da der Fokus im Zusatzstudium „Internationale Handlungskompetenz“ nicht explizit auf einer Differenzierung zwischen individualistischen und kollektivistischen Kulturen im Hofstede'schen Sinne beruht, erscheint es hier wenig sinnvoll einen Test zu verwenden, der sich auf diesen Ansatz stützt.

2.1.2 Selbsteinschätzungstests

Innerhalb der Literatur interkultureller Forschung findet man weiterhin Tests in Form von **Selbsteinschätzungen**. Diese Tests haben sich in der Praxis der Personalauswahl jedoch nicht bewährt: „[...] the way people see themselves and assess their own competencies is often very different from the way they are seen and assessed by others“ (KEALEY, 1996, S. 95). Sie sind eher für einen Einsatz innerhalb einer Personalentwicklungsmaßnahme geeignet, da sie helfen, die persönlichen Stärken und Schwächen hinsichtlich interkultureller

Handlungskompetenz zu identifizieren. Häufig werden diese auch als Instrument in interkulturellen Trainings angewandt. An dieser Stelle ist zu erwähnen, dass Instrumente, die mit Selbsteinschätzungen zur Erfassung interkultureller Handlungskompetenz arbeiten, die Gefahr einer gewissen Verzerrung durch soziale Erwünschtheit bergen (TRIANDIS, CHEN & CHAN, 1998). Damit ist gemeint, dass Personen nicht unbedingt ihre eigene Meinung wiedergeben, sondern Antworttendenzen in Richtung der von ihnen antizipierten sozial erwünschten Antwort zeigen.

Das „**Cross-Cultural Adaptability Inventory**“ (CCAI) von KELLY UND MEYERS (1992) ist ein Beispiel für einen Test zur Selbsteinschätzung der persönlichen interkulturellen Anpassungsfähigkeit, der primär zu Trainingszwecken eingesetzt werden soll. Er zielt auf die Erfassung der persönlichen Stärken und Schwächen hinsichtlich interkultureller Handlungskompetenz, um darauf aufbauend individuelle Fähigkeiten weiterzuentwickeln und Wissen zu erwerben. Der CCAI macht keine Vorhersagen über zukünftigen Auslandserfolg, sondern ist als Instrument für interkulturelles Lernen gedacht, das der persönlichen Entscheidungshilfe dient, im Ausland tätig zu werden oder nicht.

Das CCAI umfasst 50 Items, die auf den Subskalen Emotionale Stabilität, Flexibilität/Offenheit, Persönliche Selbständigkeit und Wahrnehmungsschärfe von den Probanden beantwortet werden. Die Selbsteinschätzung wird dabei auf einer 6er-Rating Skala von (1) „absolut nicht zutreffend auf meine Person“ bis (6) „absolut zutreffend auf meine Person“ vorgenommen. Das Instrument besitzt eine Reliabilität von 0,89 und wird von den Autoren als valide erachtet, ohne dass sie jedoch die empirische Beweisführung erbringen.

Da der CCAI ein Test zur Selbsteinschätzung ist, verfolgt er ein anderes Ziel als ein Test, der im Zusatzstudiengang an der Universität Regensburg eingesetzt und im Rahmen der Evaluation die Leistungen der Studenten objektiv erfassen soll. Zudem sind methodische Zweifel angebracht, die sich zum einen auf die erwähnte Problematik der sozialen Erwünschtheit und zum anderen auf die fehlende Validierung des Instrumentes beziehen. Aus diesen Gründen wird auch der CCAI von KELLY UND MEYERS (1992) für einen Einsatz im Zusatzstudium als ungeeignet erachtet.

2.1.3 Attributionstests

Die dritte Kategorie interkultureller Testverfahren umfasst so genannte **Attributionstests**. Attributionstests bestehen aus kritischen Interaktionssituationen (vgl. Kap. 3.1.2) zwischen 2 Kulturen, die von den Probanden gelöst werden müssen. Sie sind dadurch kulturspezifisch

konzipiert und nicht auf andere Kulturen übertragbar. In der Mehrzahl der Fälle werden diese Testverfahren zur Evaluation von interkulturellen Trainings eingesetzt (vgl. WELDON ET AL., 1975; CUSHNER, 1989; LAYES, 1995; HENTSCHEL, 2001). Vereinzelt werden kritische Interaktionssituationen auch verwendet, um die Fähigkeit kulturisomorpher Attributionen in interkulturellen Begegnungen zu messen (BHAWUK, 1998). Indem der Testperson offene Antwortmöglichkeiten oder die Auswahl aus mehreren Antwortalternativen angeboten werden, verringert sich hier das Problem der sozialen Erwünschtheit ganz entscheidend (TRIANDIS ET AL., 1998).

Der „**Test of Intercultural Sensitivity**“ (TICS) von WELDON ET AL. (1975) kann hier als Beispiel eines Attributionstests angeführt werden. Er behandelt 25 kritische Interaktionssituationen zwischen schwarzen und weißen Amerikanern zum Abbau von Vorurteilen bei weißen Amerikanern der Mittelschicht gegenüber schwarzen Amerikanern der Unterschicht. Die Trainingsteilnehmer müssen sich hier, genau wie im Culture Assimilator, für eine Antwortalternative, die das Verhalten in der Situation am besten erklärt, entscheiden. Sie erhalten jedoch weder eine Rückmeldung noch erfahren sie Lösungsstrategien. Für die gewählte Antwortalternative bekommen die Trainingsteilnehmer die Punktzahl, die dem Prozentsatz der Schwarzen entspricht, die in einem Vortest ebenfalls diese Erklärungsalternative gewählt haben.

Bei der Evaluation von Culture Assimilator Trainings ist es durchaus üblich, kritische Interaktionssituationen als Aufgaben im Sinne eines Leistungstests von den Trainingsteilnehmern lösen zu lassen (vgl. ALBERT, 1983; LANGE, 1994). In diesem Sinn ist der „Test of Intercultural Sensitivity“ von WELDON ET AL. (1975) eine sinnvolle Methode interkulturelle Kompetenzen zu erfassen. Allerdings handelt es sich hier um kritische Interaktionssituationen zwischen schwarzen und weißen Amerikanern, so dass diese nicht auf einen Test für deutsche Studenten übertragen werden können. Ferner ist es in diesem Zusammenhang erforderlich, für einen Studiengang, der interkulturelle Handlungskompetenz allgemein fördert, auch einen kulturallgemeinen Test einzusetzen und nicht wie im Fall des TICS, die Sensibilität für eine Kultur zu erfassen. Auch hier sind methodische Aspekte der Reliabilität und Validität zu bedenken, über die von den Autoren keine Aussage gemacht wird, die jedoch für einen Test zur Erfassung interkultureller Handlungskompetenz unerlässlich sind.

Die Klassifizierung interkultureller Testverfahren sowie die exemplarische Darstellung einiger Instrumente legen 2 Probleme offen, mit der die interkulturelle Diagnostik konfrontiert ist. Damit sind in Bezug auf das Konstrukt Schwierigkeiten auf der Kriterienebene einerseits so-

wie Schwierigkeiten auf methodischer Ebene andererseits angesprochen, die im Folgenden skizziert werden.

2.2 Zum Kriterienproblem interkultureller Diagnostik

Aus der Analyse der Testverfahren wird deutlich, wie unterschiedlich sich die Herangehensweise an die Erfassung interkultureller Handlungskompetenz in der interkulturellen Forschung gestaltet. BHAWUK und BRISLIN (1992) zufolge liegt die Schwierigkeit bei der Erfassung interkultureller Handlungskompetenz im Auffinden geeigneter Indikatoren. Das wiederum liegt zum großen Teil im Konstrukt selbst begründet. Die mehrdimensionale Auffassung von interkultureller Handlungskompetenz bietet zwar ein grobes Einteilungsraster, jedoch wird nichts darüber ausgesagt, anhand welcher Kriterien, Prädiktoren oder Indikatoren die einzelnen Dimensionen beschrieben und gemessen werden können. Trotz reger Forschungsbemühungen wird oft von einem „Kriterienproblem“ gesprochen (BENSON, 1978; TUCKER, 1974), zu dessen Lösung jedoch bis zum heutigen Zeitpunkt wenige Konzepte angeboten werden. Dieses Kriterienproblem äußert sich bereits in der Unterscheidung zwischen einer allgemeinen Fähigkeit, in interkulturellen Überschneidungssituationen kompetent zu agieren (interkulturelle Handlungskompetenz) und einer speziellen Fähigkeit, im Ausland erfolgreich im Auftrag eines Unternehmens zu arbeiten und zu handeln (interkultureller Handlungserfolg).

In diesem Zusammenhang differenzieren CUI UND AWA (1992) in ihrer Studie zur Messung interkultureller Effektivität zwischen „job performance“ und „cross-cultural adjustment“. Die Resultate ihrer Untersuchung zeigen, dass interkulturelle Anpassung und Leistungsfähigkeit nach der Selbsteinschätzung der Befragten signifikant korrelieren, so dass nach Meinung der Autoren, Anpassung an die Kultur sowie effektive Arbeitsleistung als zueinander in Beziehung stehende Aspekte zu verstehen sind. BENSON (1978) dagegen weist daraufhin, Arbeitsleistung nicht mit kultureller Anpassung zu verwechseln. Es handelt sich hier um 2 eigenständige Konzepte, auch wenn beide charakteristisch für interkulturelle Erfahrungen während Auslandsaufenthalten sind (HAMMER, 1989). Der Aspekt, im Ausland erfolgreich für eine Firma tätig zu sein, umfasst im Gegensatz zu der allgemeinen interkulturellen Handlungskompetenz auch andere Gesichtspunkte. So hängt der erfolgreiche Einsatz des Mitarbeiters im Ausland u.a. entscheidend von der Fachkompetenz, Vertrautheit mit der Unternehmenskultur oder Anpassung des (Ehe)-Partners ab (STAHL, 1998).

Da auf Kriterienebene unklar ist, was genau unter interkultureller Handlungskompetenz zu verstehen ist, bzw. wodurch sich eine interkulturell kompetente Person wirklich auszeichnet, stellt die Entwicklung valider interkultureller Testverfahren eine große Herausforderung dar. Infolgedessen kann ein bestehendes Problem auf Kriterienebene unweigerlich zu Schwierigkeiten auf Methodenebene führen. Denn wenn keine Klarheit über das zu messende Konstrukt herrscht oder identifizierte Kriterien nur unzureichend operationalisiert sind, ist es schwierig geeignete Messmethoden zu generieren.

2.3 Zum Methodenproblem interkultureller Diagnostik

Wie aus der Analyse ersichtlich wird, ist das Methodenspektrum zur Erfassung interkultureller Handlungskompetenz breit gefächert. Zu bemängeln ist hier die empirische Überprüfung der aufgestellten Prädiktoren und die z.T. fehlende Überprüfung der Gültigkeit der Verfahren.

Nach DELLER (1996) wird Auslandserfolg häufig über Anpassungsfähigkeit und Leistung im Bereich der Forschung operationalisiert. Dabei finden Selbst- und Fremdeinschätzungen Anwendung. Geprüfte Kriterien, mit denen Auslandserfolg objektiv oder „pseudo-objektiv“ gemessen oder auch nur subjektiv fremdbeurteilt werden können, finden sich in der Literatur nicht. Die eingesetzten Methoden der interkulturellen Personalauswahl sind vielfältig, jedoch wird keine bestimmte Methode empfohlen (DELLER, 1996). KEALEY UND RUBEN (1983) weisen darauf hin, dass interkulturelle Eignungsdiagnostik die Aspekte Identifizierung der notwendigen Kenntnisse, Definition von Erfolg (Anpassung und/oder Effektivität), ein Arbeitsplatz- und Kulturprofil (z.B. bei kurzfristiger versus langfristiger Entsendung) und die Auswahl angemessener Datensammlungs- und Beurteilungstechniken beinhalten muss.

Die Verfahren, die als Prädiktor für Anpassungsfähigkeit und Leistung im Ausland eingesetzt werden, variieren von Persönlichkeitstests, Interviews, Verhaltensbeobachtungen über Selbst- und Fremdeinschätzungen bis hin zu biographisch orientierten Auswahlmethoden. Weiterführende Ideen fordern eine Hinwendung zu arbeitsbezogenen Tests, wie z.B. Rollenspielen (HARRIS, 1975) oder Fallstudien und kleinen Gruppenexperimenten (CHURCH, 1982). Welche Methode sich jedoch zur Messung der Kriterien am besten eignet, hängt vom Charakter der zu beurteilenden Eigenschaft (z.B. Anpassungsfähigkeit im Gegensatz zu Kommunikationsverhalten) ab (DELLER, 2000).

2.4 Zusammenfassung der Analyse

Die Literatur hinsichtlich der Instrumente interkultureller Diagnostik ist durch das weitgehende Fehlen testdiagnostischer Erhebungen, die Vernachlässigung empirischer Forschung und unzureichende Forschungsmethoden gekennzeichnet. Viele Studien entbehren einer empirischen Überprüfung der postulierten Prädiktoren und repräsentieren vielmehr das implizite Verständnis interkultureller Eignung der Autoren. Dazu kommt, dass häufig Prädiktoren gesucht werden, die mangelhaft oder gar nicht operationalisiert sind und denen keine Kriterien explizit zugrunde gelegt werden (STAHL, 1995).

Das Problem, das sich auf Seiten der Messebene ergibt, lässt sich auf das Problem der Kriterien zurückführen. Auch wenn die Bedeutung von interkultureller Handlungskompetenz in den verschiedensten Studien interkultureller Psychologie einheitlich hervorgehoben wird, so herrscht wenig Einigkeit darüber, um welche Eigenschaften (Kriterien) es sich genau handelt. Einerseits werden hier Prädiktoren für interkulturellen Handlungserfolg identifiziert, andererseits geht es um hintergründige Dimensionen interkultureller Handlungskompetenz, wie z.B. Persönlichkeitsmerkmale, Kommunikationskompetenzen oder -verhalten etc.

Aus diesem Grund sind auch den beschriebenen Testverfahren jeweils unterschiedliche theoretische Annahmen zugrunde gelegt worden. Die dargestellten Messinstrumente sind zur Entwicklung des TIHK im Rahmen der Evaluation des Studienganges „Internationale Handlungskompetenz“ nicht zu übernehmen, da einerseits die verwendeten Methoden und Skalen sehr spezifisch auf den jeweiligen Test zugeschnitten sind. Andererseits sind die Kriterien der Testverfahren, die zum einen Persönlichkeitseigenschaften (vgl. ICSI von BHAWUK & BRISLIN, 1992; CCAI von KELLY & MEYERS, 1992) und zum anderen kulturspezifisches Wissen (vgl. TICS von WELDON ET AL. 1975) messen, für die Erfassung des Lernerfolges im Zusatzstudium nicht relevant.

Da keine Einigkeit darüber herrscht, welche Prädiktoren interkulturelle Handlungskompetenz vorhersagen, wird im folgenden Kapitel dieses Konstrukt kritisch beleuchtet. Für die Entwicklung eines Tests zur Erfassung interkultureller Handlungskompetenz (TIHK) werden Kriterien aus den theoretischen Befunden der interkulturellen Psychologie bzw. der kulturvergleichenden Psychologie abgeleitet, die hinreichend operationalisiert werden müssen, um darauf aufbauend eine geeignete Methode zu generieren.

3 Theoretische Grundlagen für die Entwicklung des TIHK

Die vorliegende Arbeit beschäftigt sich mit der Entwicklung eines Tests zur Erfassung interkultureller Handlungskompetenz (TIHK). Das impliziert, dass sie sich mit Kultur im weiteren Sinne auseinandersetzt. Im Folgenden soll deswegen zuerst deutlich gemacht werden, welches Verständnis von „Kultur“ hier zugrunde gelegt wird, wie darauf aufbauend "Kulturstandards" als Bestimmungsstücke ausgemacht werden können und welche Implikationen dies für die Begegnung in kulturellen Überschneidungssituationen beinhaltet.

Im zweiten Teil dieses Kapitels wird geklärt, wie sich das Konstrukt interkultureller Handlungskompetenz fassen und beschreiben lässt. Dazu wird die aktuelle Literatur interkultureller Forschung analysiert und vor dem Hintergrund der Aufgabenstellung die Fähigkeit zum Perspektivenwechsel als Indikator zur Erfassung interkultureller Handlungskompetenz herausgearbeitet.

3.1 Grundlagen interkultureller Psychologie

3.1.1 Kultur und Kulturstandards

Betrachtet man die aktuelle Forschungslage zum Konstrukt „**Kultur**“, so ergibt sich ein sehr vielseitiges und heterogenes Bild. Es existiert eine Vielzahl von Kulturdefinitionen (KROEBER & KLUCKHOHN, 1952), die je nach Forschungsdisziplin andere Schwerpunkte setzen. Zur Erfassung interkultureller Handlungskompetenz erweist sich die Kulturdefinition von THOMAS (1993) als geeignet, da sie auf handlungstheoretischen Konzepten beruht und „der kognitiv orientierten Sozial- und Persönlichkeitspsychologie verpflichtet“ ist (THOMAS, 1993, S. 158).

“Kultur ist ein universelles, für eine Gesellschaft, Organisation und Gruppe aber sehr typisches Orientierungssystem. Dieses Orientierungssystem wird aus spezifischen Symbolen gebildet und in der jeweiligen Gesellschaft, Organisation und Gruppe tradiert. Es beeinflusst Wahrnehmen, Denken, Werten und Handeln aller ihrer Mitglieder und definiert somit deren Zugehörigkeit zur Gesellschaft. Kultur als Orientierungssystem strukturiert ein für die sich der Gesellschaft zugehörig fühlenden Individuen spezifisches Handlungsfeld und schafft damit die Voraussetzungen zur Entwicklung eigenständiger Formen der Umweltbewältigung“ (THOMAS, 1993, S. 380).

In der Definition von THOMAS wird deutlich, dass Kultur etwas vom Menschen konstruiertes ist. TRIANDIS (1989) nennt diesen Aspekt, „den vom Menschen gemachten Teil der Umwelt“ (S.306). Zum anderen bestimmt Kultur die Umwelt der Menschen: „...sie legt für den einzelnen fest, wie er seine soziale und materielle Umwelt strukturiert“ (TRIANDIS, 1989, S.306).

Kultur stellt damit nicht einfach nur die vom Menschen gestaltete Umwelt dar, sondern ist zugleich Handlungsfeld, das Handlungsbedingungen und -grenzen festlegt und somit einen für alle Mitglieder der kulturellen Gruppe verbindlichen Handlungsspielraum determiniert (vgl. BOESCH, 1980, S. 17). Kultur ist also Bedingung und Ergebnis von Handlung (KROEBER & KLUCKHOHN, 1982). TRIANDIS (1972) unterscheidet zwischen „objektiver“ und „subjektiver“ Kultur. Objektiv bezieht sich auf Elemente einer Kultur, die offen sichtbar sind, wie z.B. Kleidung, Häuser, Werkzeug, Straßen usw., wogegen subjektiv die spezifische Art und Weise impliziert, wie eine kulturelle Gruppe ihre Umwelt wahrnimmt (z.B. Einstellungen, Überzeugungen, Normen und Werte). Die angesprochene subjektive Kultur bei TRIANDIS (1972) findet sich auch bei THOMAS (1993) wieder, der in diesem Zusammenhang die zentralen Merkmale eines kulturspezifischen Orientierungssystems als „Kulturstandards“ bezeichnet.

Unter **Kulturstandards** werden nach THOMAS (1993) „alle Arten des Wahrnehmens, Denkens, Wertens und Handelns verstanden, die von der Mehrzahl der Mitglieder einer bestimmten Kultur für sich und für andere als normal, selbstverständlich, typisch und verbindlich angesehen werden“ (S. 381). Analog zur subjektiven Kultur bei TRIANDIS (1972) äußern sich Kulturstandards auf verschiedenen Abstraktionsniveaus und können in allgemeinen Werten bis hin zu konkreten Verhaltensvorschriften einer Kultur zum Ausdruck kommen. Dabei gibt es individuelle und gruppenspezifische Ausprägungen von Kulturstandards, die sich jedoch innerhalb eines bestimmten Toleranzbereiches bewegen. Verhaltensweisen, die außerhalb der Toleranzgrenzen liegen, werden abgelehnt und sanktioniert. Da Kulturstandards im Laufe der Sozialisation gelernt und verinnerlicht werden, sind sie den Angehörigen einer Kultur nicht ständig bewusst, sondern werden erst im Kontakt mit fremdkulturell sozialisierten Partnern bemerkt.

Kulturstandards fungieren demnach als „Maßstäbe, Gradmesser und Bezugssysteme“ (THOMAS & SCHENK, 1996, S. 24) und sind nach THOMAS (1993) besonders bei der Regulation sozialen Verhaltens in einer Kultur, also in der Interaktion von Menschen bedeutsam. Eine mögliche Interaktion, in der Kulturstandards relevant werden, stellen kulturelle Überschneidungssituationen dar, die im Folgenden genauer erklärt werden.

3.1.2 Kulturelle Überschneidungssituation

Eine **kulturelle Überschneidungssituation** ist nach KINAST (1998) ein Handlungsfeld, in dem Kulturstandards unterschiedlicher kulturspezifischer Orientierungssysteme aufeinander treffen. Sie entsteht dadurch, dass ein Individuum in eine fremde kulturelle Umwelt oder in Kontakt mit einem fremdkulturellen Individuum tritt (DADDER, 1987). Da alles menschliche Tun

sich in einem kulturspezifischen Kontext vollzieht (THOMAS & SCHENK, 1996), wird das Wahrnehmen, Denken, Werten und Handeln innerhalb einer solchen Begegnung an unterschiedlichen oder unterschiedlich gewichteten Kulturstandards ausgerichtet. Dieser Tatsache werden sich die betroffenen Personen in aller Regel vorerst nicht bewusst sein, da Kulturstandards während der Sozialisation internalisiert werden und dadurch in der Folge nicht mehr bewusstseinspflichtig sind, wenngleich sie potentiell bewusstseinsfähig bleiben (THOMAS, 1993) (vgl. Kap. 3.1.1). Beide Handlungspartner werden daher ihr Verhalten zunächst an den jeweils eigenen, gewohnten Kulturstandards ausrichten und die Handlungen des jeweils anderen daran beurteilen. Dies kann im harmlosesten Fall zu kleineren Missverständnissen führen. Da nach THOMAS (1995) Menschen aber dazu neigen, „das, was verschieden ist, auch für „falsch““ (S. 89) zu halten, werden diese anderen Arten der Umweltbewältigung oft negativ bewertet und abgelehnt. Dies gilt insbesondere dann, wenn der gewisse Toleranzbereich, der jeden Kulturstandard umgibt, überschritten ist (vgl. Kap. 3.1.1). Es kann dann zu schweren Interaktionskonflikten bis hin zum Abbruch von Beziehungen kommen. Aber auch dann, wenn diese negative Bewertung ausbleibt, kann das Bewusstwerden der verschiedenen kulturellen Orientierungen ohne adäquate Handlungsstrategie zur Handlungsunfähigkeit führen. Wiederholen sich derartige Erlebnisse, kann die Person schlimmstenfalls einen Kulturschock erleiden (OBERG, 1960).

Um solche Auswirkungen unterschiedlicher kultureller Orientierungssysteme in der interkulturellen Begegnung zu vermeiden, dienen in der interkulturellen Forschung kulturelle Überschneidungssituationen zur Beschreibung und Analyse fremdkulturellen Verhaltens. Aufbauend auf der Analyse kultureller Überschneidungssituationen wird versucht, zu einem besseren gegenseitigen, also interkulturellen Verständnis zu gelangen. Dazu werden kulturelle Überschneidungssituationen in so genannten „**kritischen Interaktionssituationen**“ festgehalten. Das deutsche Wort „kritische Interaktionssituation“ (im Folgenden auch als KI bezeichnet) geht auf die „critical incident“ Methode von FLANAGAN (1954) zurück und hat sich als geeignetes Konzept zur Identifizierung zentraler Kulturstandards durchgesetzt. Ein „critical incident“ ist demnach eine Situation, die nach FIEDLER ET AL. (1971) folgende 3 Kriterien erfüllen muss:

1. Eine Person aus der Anwenderkultur muss in dieser Situation mit einer Person der Zielkultur interagieren.
2. Das Verhalten, das der Zielkulturelle in dieser Situation zeigt, muss von der Person aus der Anwenderkultur als verwirrend oder unverständlich wahrgenommen werden.
3. Dieses Verhalten muss für eine Person, die mit der Zielkultur vertraut ist, eindeutig interpretierbar sein.

Die dargestellte KI sollte sich also dadurch auszeichnen, dass sie von den Beteiligten als konflikthaft oder unverständlich erlebt wird und bei ausreichenden Kenntnissen über die Zielkultur eindeutig interpretierbar ist, d.h., der dargestellte Konflikt begründet sich in kulturellen Unterschieden. Das Konzept der kritischen Interaktionssituationen hat sich im Bereich interkultureller Culture Assimilator Trainings durchgesetzt und findet weite Verbreitung im internationalen Trainingsgeschäft. Zentral ist hierbei der Aufbau der Fähigkeit, das Verhalten eines Mitgliedes einer fremden Kultur „isomorph attribuieren“ (TRIANDIS, 1975, 1977; vgl. Kap. 3.3.2) zu können. Auf die Methode des Culture Assimilators wird zu einem späteren Zeitpunkt dieser Arbeit (vgl. Kap. 5.1) näher eingegangen.

Damit kulturelle Überschneidungssituationen nicht zu schwerwiegenden Konflikten eskalieren oder um allgemein mit dem fremdkulturellen Gegenüber effektiv interagieren zu können, muss sich das Individuum in der interkulturellen Begegnungssituation interkulturell orientieren, also in sich selbst zu einem Vermittler beider Kulturen werden (LANGE, 1994). Die Entwicklung dieser individuellen Fähigkeiten wird durch interkulturelles Lernen ermöglicht, das sich wiederum implizit in allen kulturellen Überschneidungssituationen ereignen kann (LANGE, 1994).

3.1.3 Interkulturelles Lernen

Die Fähigkeit zu kulturadäquatem Handeln ist nicht einfach vorhanden oder nicht vorhanden, sondern entwickelt sich nach THOMAS (in Druck) in und aus dem Verlauf eines hochgradig lernsensitiven interkulturellen Begegnungs- und Erfahrungsprozesses. **Interkulturelles Lernen** unterscheidet sich von Lernen unter eigenkulturellen Bedingungen lediglich hinsichtlich des Lerninhalts, des Lernkontextes und der Lernreize (BREITENBACH, 1975) und wird von verschiedenen Forschern als Entwicklungsprozess beschrieben (HOOPES, 1981; BENNETT, 1986; WINTER, 1988; BHAWUK, 1998). Wie zuvor erwähnt, können kulturelle Überschneidungssituationen einen Weg zum interkulturellen Lernen darstellen. Sie bieten dem Individuum die Möglichkeit, das fremdkulturelle Orientierungssystem kennen zu lernen und in das eigene Orientierungssystem zu integrieren und sich somit die fremde Kultur als Handlungsfeld zu erschließen. „Interkulturelles Lernen findet statt, wenn eine Person bestrebt ist, im Umgang mit Menschen einer anderen Kultur deren spezifisches Orientierungssystem der Wahrnehmung, des Denkens, Wertens und Handelns zu verstehen, in das eigene Orientierungssystem zu integrieren und auf ihr Denken und Handeln im fremdkulturellen Handlungsfeld anzuwenden. Interkulturelles Lernen bedingt neben dem Verstehen fremdkultureller Orientierungssysteme eine Reflexion des eigenkulturellen Orientierungssystems“ (THOMAS,

1993, S. 382). Die Definition von THOMAS (1993) impliziert, dass der Lernende eine monokulturelle Sozialisation an relevanten Werten, Normen und Verhaltensregeln durchlaufen hat, die einen festen Bestandteil seines alltäglichen Lebens (Orientierungssystem) darstellen. Der Lernende geht grundsätzlich davon aus, dass das, was er tut, wie er denkt und empfindet richtig ist und für alle Menschen gilt. Begegnet er nun in einer kulturellen Überschneidungssituation abweichenden und für ihn ungewöhnlichen Verhaltensweisen, reagiert er mit Verunsicherung und Ablehnung. Um nun aus einer solchen Situation zu lernen, muss er in der Lage sein, seinen eigenkulturellen Orientierungsrahmen zu reflektieren, um so die Andersartigkeit eines fremdkulturellen Orientierungssystems zu begreifen. Es handelt sich dabei um einen interkulturellen Lernprozess, dessen oberstes Ziel interkulturell kompetentes Handeln in kulturellen Überschneidungssituationen ist (KAMMHUBER, 2000). Der Endzustand interkulturellen Lernens ist nach SANDHAAS (1988) ein Individuum, das in dem fremden wie eigenen Orientierungssystem handlungsfähig ist, die Handlungen des fremdkulturellen Interaktionspartners in der gleichen Weise interpretieren und antizipieren kann wie in der Heimatkultur sowie das Handeln in der Fremdkultur als befriedigend und angenehm erlebt. Kurz gefasst ist der Endzustand interkulturellen Lernens demnach eine Person, die zu interkulturell kompetentem Handeln befähigt ist. Was genau unter interkulturell kompetentem Handeln oder anders formuliert unter interkultureller Handlungskompetenz zu verstehen ist, bzw. welche Komponenten kulturadäquates Handeln determinieren, wird im folgenden Abschnitt diskutiert.

3.2 Interkulturelle Handlungskompetenz

3.2.1 Begriffsabgrenzung

Im Zentrum des Interesses steht in diesem Abschnitt das Konstrukt **interkultureller Handlungskompetenz** als Voraussetzung für interkulturellen Handlungserfolg.

Im Rahmen interkultureller Forschungen ist eine Fülle von Begriffen entstanden, die die Vermittlung eines homogenen Bildes erschwert, gleichzeitig aber die Komplexität des Konstrukts interkultureller Handlungskompetenz repräsentiert. Aus diesem Grund wird in der vorliegenden Arbeit einheitlich der Begriff der interkulturellen Handlungskompetenz verwendet. Darunter werden auch Begriffe aus dem angloamerikanischen Raum wie „intercultural competence“, „intercultural sensitivity“ oder „intercultural effectiveness“ gefasst, die teilweise als dazu synonym anzusehen sind, da sie oft mit ähnlichen oder gleichen Variablen umschrieben werden.

3.2.2 Die Forschungsproblematik bzgl. interkultureller Handlungskompetenz

Darüber, dass interkulturelle Handlungskompetenz eine Schlüsselqualifikation für international tätige Personen ist und zunehmend an Bedeutung gewinnt, sind sich Politiker, Wissenschaftler und Manager einig. Um erfolgreich im Ausland oder mit fremdkulturellen Partnern agieren zu können, bedarf es bestimmter interkultureller Kompetenzen, die es zu entwickeln bzw. fördern gilt. Auf die Frage, was interkulturelle Handlungskompetenz genau ausmacht, werden in der interkulturellen Forschung unterschiedliche Ansichten vertreten.

Grundsätzlich herrscht Konsens darüber, interkulturelle Handlungskompetenz als multidimensionales Konstrukt zu beschreiben (KEALEY & RUBEN, 1983). Dabei ist das übergeordnete Ziel vieler wissenschaftlicher Arbeiten, Variablen zu identifizieren, die als Prädiktoren für interkulturellen Handlungserfolg dienen können. Bei der Definition von interkultureller Handlungskompetenz handelt es sich häufig um eine Aufzählung von umfangreichen Eigenschaftslisten, die auf unterschiedlichstem Abstraktionsniveau einzelne Komponenten des Konstrukts erfassen. Diese Listen von Fähigkeiten und Fertigkeiten, die erfüllt werden müssen, um im Ausland erfolgreich interagieren zu können, lesen sich nach THOMAS (in Druck) „wie das Persönlichkeitsprofil des modernen Menschen, mit stark idealisierten, von allen angestrebten, aber von niemandem erreichten Leistungsmerkmalen.“

Bei näherer Betrachtung der einzelnen Definitionen wird deutlich, dass von den Autoren häufig der Schwerpunkt auf die kommunikativen Fähigkeiten der beteiligten Personen gelegt wird. Dabei wird immer wieder die Fremdsprachenkompetenz als Prädiktor für interkulturellen Erfolg betont, die sicher einen wichtigen Aspekt darstellt, aber nicht das ganze Themenspektrum umfasst.

Ein weiteres Problem, das in diesem Zusammenhang offensichtlich wird, ist das Fehlen hinreichender Definitionen des Konstrukts bzw. dessen mangelnde empirische Überprüfungen. Die zahlreichen Definitionsversuche in der Literatur sind nach STAHL (1998) entweder aus Plausibilitätsüberlegungen abgeleitet oder sind, sofern sie überhaupt empirisch gewonnen wurden, mit schwerwiegenden methodischen Fehlern behaftet (z.B. schlechte Operationalisierungen, Erhebung impliziter Eignungstheorien der Befragten). „Die resultierenden Merkmalslisten muten deshalb oft wie die Karikatur eines erfolgreichen Auslandsaufenthaltes an“ (S. 60).

3.2.3 Zum Forschungsparadigma interkultureller Handlungskompetenz

Die Anfänge verstärkter Forschungsaktivitäten auf dem Gebiet der interkulturellen Psychologie sind auf Personalauswahlverfahren für Freiwillige der Peace Corps in den USA in den 60er Jahren zurückzuführen. Zunehmende interkontinentale wirtschaftliche Verflechtungen haben zudem Investitionen in die Erforschung interkultureller Kommunikation notwendig gemacht. „Project failures“, „botched negotiations“, „early return of workers“ und „lost time and money“ (RUBEN, 1989) waren Gründe dafür, Maßnahmen zur Mitarbeiterqualifizierung und Personalauswahl einzuleiten. In Deutschland sind in diesem Bereich erst seit den 90er Jahren verstärkte Aktivitäten zu verzeichnen, die im Zusammenhang mit der zunehmenden Internationalisierung und EU-Erweiterung gesehen werden können.

In den vergangenen 50 Jahren hat die Forschung zur interkulturellen Kompetenz ähnlich wie viele andere psychologische Konstrukte einen Paradigmenwechsel vollzogen, den STAHL (1998) wie folgt beschreibt: „ein Wandel ausgehend vom personalistischen über den situationistischen hin zum interaktionistischen Ansatz“ (S. 53). Die verschiedenen Forschungstraditionen des angloamerikanischen und deutschsprachigen Raumes werden in dieser Arbeit gleichermaßen berücksichtigt, obwohl die angloamerikanische Forschung auf eine längere Tradition und zahlreichere Veröffentlichungen zurückblicken kann.

Im **personalistischen Ansatz** wird die Auffassung vertreten, dass bestimmte Persönlichkeitseigenschaften interkulturelle Handlungskompetenz determinieren. Vertreter dieser Forschungsrichtung sind u.a. GARDNER (1962), der von einem „universal communicator“ ausgeht, dessen Charaktereigenschaften ihn dazu befähigen, erfolgreich in interkulturellen Situationen zu agieren. Im Rahmen des personalistischen Forschungsparadigmas identifizieren Wissenschaftler einzelne Komponenten interkultureller Handlungskompetenz. HINZ-ROMMEL (1994, S. 60) hat die so entstandenen wenig trennscharfen Kategorien aus der Literatur zusammengetragen:

- „Abilities“ bei DEAN/POPP (1990), HAMMER, GUDYKUNST/WISEMAN, (1978), ABE/WISEMAN (1983)
- „Attitudes“ bei RUBEN (1989) und HANNIGAN (1990)
- „Behaviours“ bei RUBEN (1989)
- „Traits“ bei HANNIGAN (1990)
- „Motivation“ bei TODD IMAHORI/LANIGAN (1989)
- „Skills“ bei HANNIGAN (1990), TODD IMAHORI/LANIGAN (1989)
- „Knowledge“ bei RUBEN (1989), TODD IMAHORI/LANIGAN (1989)

Eine im Rahmen personalistischer Ansätze viel zitierte Studie ist die von HAMMER, GUDYKUNST UND WISEMAN (1978), die mittels Faktorenanalyse 24 Persönlichkeitseigenschaften auf 3 allgemeine Verhaltenskomponenten zurückführen konnte. Dieser Analyse zufolge sind die Dimensionen „*Fähigkeit zur psychologischen Stressbewältigung*“, „*Fähigkeit zur effektiven Kommunikation*“ und „*Fähigkeit zum Aufbau interpersonaler Beziehungen*“ entscheidend für interkulturelle Effektivität, was von HAMMER (1987) repliziert werden konnte.

Als weitere empirisch gut überprüfte Persönlichkeitseigenschaften sind die Konzepte „*category width*“ (DETWEILER, 1978) und „*open-mindedness*“ (KEALEY & RUBEN, 1979) erwähnenswert. Beide Konzepte beschreiben, zu welchem Grad eine Person für neue und abweichende Ideen und Reize offen ist, wozu auf empirischer Ebene ein Zusammenhang mit erfolgreichem interkulturellem Handeln hergestellt werden konnte. Besonders hervorzuheben ist in diesem Zusammenhang die einschlägige Studie von KEALEY UND RUBEN (1979). Die Autoren konnten neben „*open-mindedness*“ 6 weitere Fähigkeiten für interkulturellen Erfolg generieren: „*display of respect*“, „*interaction posture*“, „*empathy*“, „*role behaviour*“, „*interaction management*“ und „*tolerance for ambiguity*“.

An dieser Stelle wird noch auf die in Deutschland erschienene Auflistung von DELKESKAMP (1991) hingewiesen. Sie nennt „*Interaktionsfreudigkeit*“, „*Flexibilität*“, „*Selbstsicherheit*“, „*eigenkulturelle Bewusstheit*“, „*Stresstoleranz*“, „*Fähigkeit Widersprüchlichkeiten zu ertragen*“ und „*Empathie*“ als relevante Faktoren und kommt damit den amerikanischen Veröffentlichungen sehr nahe.

Bei Betrachtung des personalistischen Ansatzes wird trotz einiger Gemeinsamkeiten deutlich, welche unterschiedlichen Auffassungen interkultureller Handlungskompetenz innerhalb der Forschergemeinde existieren. Neben den in dieser Arbeit erwähnten mehrfach identifizierten Eigenschaften Offenheit, Empathie und Ambiguitätstoleranz (Fähigkeit Widersprüchlichkeiten bzw. Unsicherheiten zu ertragen) wird an anderer Stelle vor allem eine Fähigkeit als zentrale Voraussetzung für kompetentes interkulturelles Handeln von verschiedenen Forschern hervorgehoben: die „**Fähigkeit zum Perspektivenwechsel**“ (vgl. KIM, 1991; KIM & RUBEN, 1988; TAYLOR, 1994; BRISLIN, LANDIS & BRANDT, 1983 etc.). Darunter wird allgemein die Fähigkeit einer Person verstanden, multiple Perspektiven einzunehmen und für ihr Handeln berücksichtigen zu können.

So wird z.B. auch bei THOMAS ET AL. (1997) neben dem kulturspezifischen Wissen, die generelle Fähigkeit zur Übernahme der Sichtweise einer anderen Person in einer kulturellen Überschneidungssituation als unbedingt erforderlich erachtet. TAYLOR (1994) dagegen beschreibt interkulturelle Handlungskompetenz als transformatives und adaptives Konstrukt,

welches das Ergebnis eines Lernprozesses darstellt. Bei Entwicklung interkultureller Handlungskompetenz werden vor allem durch kommunikatives Lernen über Wertvorstellungen, Moralurteile, Ideale etc. anderer Personen Änderungen in den Deutungsmustern und Perspektiven erreicht. BRISLIN, LANDIS & BRANDT (1983) verweisen mit dem Begriff „Lability of Self Perception“ auf die Fähigkeit einer Person, sein Selbstbild und seine Weltsicht in Frage stellen zu können, was einer Person ermöglicht, aus einer ethnozentristischen Sichtweise herauszutreten. KIM UND RUBEN (1988) und KIM (1991) betonen in ihrem Konzept der „intercultural transformation“ ebenfalls die Rolle von Perspektivenwechsel als Bestandteil interkultureller Kompetenz. KIM (1991) geht davon aus, dass ein Individuum dann interkulturell kompetent ist, wenn es fähig ist, seine Perspektive zu verändern, um fremdkulturellen Anforderungen gerecht zu werden. Durch die Veränderung der eigenen Sichtweise wird laut STAHL (1998) versucht, „zu einem größeren Verständnis für das Handeln der anderen Person zu gelangen, indem man sich bemüht, deren Beweggründe nachzuvollziehen, sich in deren Lage zu versetzen oder die Legitimität ihres Verhaltens anzuerkennen“ (S.189).

Anhand der Fülle der Veröffentlichungen zur Thematik des Perspektivenwechsels wird seine Bedeutung für das Konstrukt interkulturelle Handlungskompetenz offensichtlich und aus diesem Grund an dieser Stelle besonders hervorgehoben. Die Relevanz des Perspektivenwechsels für diese Arbeit wird an späterer Stelle konkretisiert (vgl. Kap. 3.3).

Im **situativen Ansatz** wird die Ansicht vertreten, dass die Arbeits- und Lebensbedingungen und deren Bewältigung interkulturellen Erfolg bestimmen: „Personality concepts have very little predictive value, rather situations predict behaviour“ (DETWEILER, BRISLIN & MC CORMACK, 1983, S.102). In diesem Zusammenhang sind wenig aussagekräftige Listen angeblich relevanter Situationsfaktoren entstanden, wie z.B. die von BRISLIN (1981), der u.a. „*klimatische Bedingungen*“, „*persönliche und unpersönliche Interaktion*“, „*Vertrautheit vs. Anonymität*“, „*Status und Benehmen des Gegenübers*“, „*zeitliche Rahmenbedingungen*“, „*Führungskraft-Mitarbeiter-Beziehung*“, „*Bekanntheitsgrad*“ etc. als Faktoren für interkulturellen Handlungserfolg identifizieren konnte. Verschiedene Forscher vertreten die Annahme, dass in unbekanntem kulturellen Situationen situative Variablen erfolgskritischer sind als Persönlichkeitsvariablen. STAHL (1998) hält fest, dass nicht die Umweltgegebenheiten als solches, sondern die Art und Weise, wie diese wahrgenommen, bewertet und verarbeitet werden ausschlaggebend für die interkulturelle Kooperation ist. Für ihn sind die Anzahl und Güte der Beziehungen zu Landsleuten und die kulturelle Distanz zwischen Heimat- und Gastland wichtige Determinanten des Anpassungserfolgs.

Der **interaktionistische Ansatz** kombiniert personale und situationale Aspekte interkultureller Handlungskompetenz. Interkulturelle Handlungskompetenz wird hier als das Resultat der Interaktion von Individuum und Situation betrachtet. Anpassungsvorgänge, so die Meinung verschiedener Autoren, vollziehen sich in einer ständigen Wechselwirkung zwischen der Person und der Umwelt. Die Definition von FURNHAM UND BOCHNER (1986) betont, dass interkulturelle Handlungskompetenz sich in der Auseinandersetzung der Person mit ihrem Gegenüber innerhalb der gemeinsam erlebten und geschaffenen Situation zeigt: „not simply a personal trait or a learned skill, but a social phenomenon“ (S. 217). KINAST (1998) spricht von interkultureller Handlungskompetenz als „Kenntnisse, Fähigkeiten und Fertigkeiten, die Bedingung und zugleich Ergebnis der Bewältigung einer kulturellen Überschneidungssituation sind“ (S. 9) und verdeutlicht auf diese Weise ebenfalls die Bedeutung von Person und Situation. Auch die Ausführungen von THOMAS (1996) können in diese Kategorie eingereiht werden: „die Fähigkeit zur Person- und Situationswahrnehmung, zur Verhaltensbeurteilung und zum Situationserleben im Kontext des fremdkulturellen Orientierungssystems ist die Voraussetzung zu kulturadäquatem Handeln“ (S. 115).

Umstritten bleibt an dieser Stelle jedoch die unterschiedliche Bedeutung und Gewichtung personaler und situativer Faktoren zur Bestimmung interkultureller Handlungskompetenz. Viele aktuelle Studien thematisieren Interaktionseffekte zwischen Personen- und Situationsvariablen, wobei die Autoren die Faktoren je nach Überzeugung unterschiedlich gewichten, so dass sowohl Studien mit Fokus auf Situationsaspekten (z.B. DINGES & LIEBERMANN, 1989) als auch Studien mit Fokus auf Personenfaktoren (z.B. KEALEY, 1989) existieren. Die meisten der angeführten Veröffentlichungen gehen implizit davon aus, dass interkulturelle Handlungskompetenz ein nicht kulturspezifisches, sondern weitgehend übertragbares Kompetenzprofil ist (HINZ-ROMMEL, 1994, DELLER, 2000). Als Fazit sei hier die Aussage von DINGES UND BALDWIN (1996) erwähnt, die in ihrem Review zu interkultureller Forschung zu dem Schluss kommen, dass sich die meisten Forscher heute als „interactionists“ bezeichnen. Persönlichkeitsaspekte werden jedoch wohl auch deswegen weiterhin als zentral angesehen, da man zur Vorbereitung auf den Auslandseinsatz die Menschen und nicht die gegebenen Situationsfaktoren vor Ort trainieren kann.

Da in der aktuellen interkulturellen Forschung die Interdependenz von Personen- und Situationsfaktoren innerhalb interkultureller Begegnungen allgemein anerkannt ist, soll auch in dieser Arbeit die interaktionistische Sichtweise bei der Entwicklung des Tests zur Erfassung interkultureller Handlungskompetenz berücksichtigt werden. Im Anschluss an die verschiedenen Forschungsansätze der interkulturellen Psychologie, wird nun eine Definition interkultureller Handlungskompetenz gegeben, die das dieser Arbeit zugrunde liegende Vorverständnis des Konstrukts präzisiert.

3.2.4 Definition interkultureller Handlungskompetenz

Eine Definition des Konstrukts interkultureller Handlungskompetenz, die dem interaktionistischen Ansatz gerecht wird und sich in der interkulturellen Psychologie durchgesetzt hat, ist die von THOMAS (in Druck):

„Interkulturelle Handlungskompetenz zeigt sich in der Fähigkeit, kulturelle Bedingungen und Einflussfaktoren im Wahrnehmen, Urteilen, Empfinden und Handeln bei sich selbst und bei fremden Personen zu erfassen, zu würdigen und produktiv zu nutzen im Sinne einer wechselseitigen Anpassung, von Toleranz gegenüber Inkompatibilitäten und einer Entwicklung hin zu synergieträchtigen Formen der Zusammenarbeit, des Zusammenlebens und handlungswirksamer Orientierungsmuster in Bezug auf Weltinterpretation und Weltgestaltung“ (Thomas, in Druck).

THOMAS liefert hier, analog zu seiner Kulturdefinition, einerseits eine sehr umfassende, allgemein gehaltene Definition interkultureller Handlungskompetenz. Er betont andererseits das Zusammenspiel von personalen Grundbedingungen, situativen interkulturellen Erfahrungen, Prozessen des interkulturellen Lernens und Verstehens sowie der Interdependenzen im Begegnungs- und Handlungsvollzug und ist deshalb auch von besonderem Wert für diese Arbeit. Darüber hinaus integriert er kognitive, affektive und behaviourale Aspekte interkultureller Handlungskompetenz, die in interkulturellen Begegnungssituationen auf beiden Seiten (in der eigenkulturellen und in der fremdkulturellen Sichtweise) handlungswirksam werden. Seiner Einschätzung zufolge reicht es zur Bewältigung einer interkulturellen Begegnung nicht aus, fremdkulturelle Einflüsse zu erfassen (Kognition), man muss diese auch wertschätzen (Emotion) und produktiv nutzen (in Handlung umsetzen) können. Diese Einteilung interkultureller Handlungskompetenz in 3 Komponenten wird noch in weiteren Ansätzen vorgenommen (vgl. BRISLIN, 1989; DADDER, 1987; GUDYKUNST, GUZLEY & HAMMER (1996); LANDIS & BAGHAT, 1996) und wird im folgenden Abschnitt näher beleuchtet.

3.2.5 Einteilung interkultureller Handlungskompetenz in 3 Ebenen

Ein Versuch, sich dem Konstrukt interkultureller Handlungskompetenz zu nähern, kann über die Zerlegung in 3 interdependente Dimensionen geschehen: Einer kognitiven, einer affektiven und einer verhaltensbezogenen Dimension. Diese klassische Aufteilung der Psychologie ist häufig in der interkulturellen Forschung anzutreffen, so z.B. bei der Klassifikation der Lernziele interkultureller Trainings (Lernzieltaxonomie von BLOOM, 1972 und KRATHWOHL ET AL. 1975) oder der Aufschlüsselung der Evaluation interkultureller Trainings nach kognitiven, affektiven und behaviouralen Effekten (vgl. KINAST, 1998). Dieses verbreitete Schema der

Zerlegung in Komponenten scheint hier geeignet, um das Konstrukt interkulturelle Handlungskompetenz greifbarer zu machen. Im Folgenden ist die Einteilung interkultureller Handlungskompetenz in Anlehnung an HERBRAND (2000) dargestellt, da dessen Definition sich mit der von THOMAS (in Druck) deckt.

1. Kognitive Dimension

Unter die kognitive Dimension interkultureller Handlungskompetenz fällt hier das Wissen um Gemeinsamkeiten und Unterschiede zwischen Kulturen, was dazu beiträgt, die Komplexität einer Kultur zu verstehen. Dazu gehören das Erkennen, Einschätzen und Strukturieren einer kulturellen Überschneidungssituation (DADDER, 1987). Analog hierzu führt THOMAS (in Druck) in seiner Definition die Aspekte des Erfassens und Urteilens in der interkulturellen Überschneidungssituation an.

2. Affektive Dimension

Neben der persönlichen Einstellung zu interkulturellen Kontakten sind Eigenschaften wie Empathiebereitschaft, Offenheit, Toleranz und Geduld von zentraler Bedeutung. Diese Eigenschaften sind auch in den zuvor dargestellten Eigenschaftslisten zu finden (vgl. Kap. 3.2.3) und beziehen sich auf eine positive Einstellung zum fremdkulturellen Interaktionspartner. Die Sensibilität für die Sichtweise und Emotionen des fremdkulturellen Partners werden ebenfalls zu dieser Dimension gerechnet. Die Parallele kann auch hier wieder zur Definition von THOMAS (in Druck) gezogen werden, der von der Würdigung fremdkultureller Einflüsse spricht.

3. Verhaltensbezogene Dimensionen

Interkulturelles Wissen und interkulturelle Sensibilität bedürfen einer Ergänzung durch Fähigkeiten und Fertigkeiten auf der Verhaltensebene, um effektiv mit Menschen einer anderen Kultur interagieren und die aus kultureller Andersartigkeit resultierenden Probleme und Konflikte bewältigen oder von vornherein vermeiden zu können. An dieser Stelle spricht THOMAS (in Druck) von der Fähigkeit kulturelle Bedingungen und Einflussfaktoren produktiv zu nutzen, um synergieträchtige Formen der Zusammenarbeit, des Zusammenlebens und handlungswirksamer Orientierungsmuster zu entwickeln.

Die dargestellte Einteilung in kognitive, affektive und behaviourale Dimensionen betont die verschiedenen Ebenen interkultureller Handlungskompetenz und verdeutlicht damit die Komplexität der sich abspielenden Prozesse innerhalb interkulturellen Begegnungen. Bei der Entwicklung eines Tests zur Erfassung interkultureller Handlungskompetenz sollte daher dieser Aspekt nicht außer Acht gelassen werden.

Im Sinne einer Begriffsklärung des zu erfassenden Konstrukts interkultureller Handlungskompetenz wird an dieser Stelle festgehalten, dass interkulturelle Handlungskompetenz ein mehrdimensionales dynamisches Konstrukt ist, das sowohl von Persönlichkeitseigenschaften als auch Situationsvariablen beeinflusst wird. Dabei wird die Fähigkeit zum Perspektivenwechsel zur Bewältigung einer kulturellen Überschneidungssituation in verschiedenen Studien besonders hervorgehoben. An diesem Punkt wird nun angesetzt und im Folgenden die Relevanz des Perspektivenwechsels für einen Test zur Erfassung interkultureller Handlungskompetenz herausgearbeitet.

3.3 Perspektivenwechsel als Indikator für interkulturelle Handlungskompetenz

3.3.1 Das Konzept des Perspektivenwechsels

Wie unter Kapitel 3.2.3 gezeigt werden konnte, wird der Fähigkeit zum Perspektivenwechsel als Indikator für interkulturelle Handlungskompetenz in der interkulturellen Forschung ein hoher Stellenwert zugesprochen. Eine Arbeit, die in diesem Zusammenhang besonders hervorgehoben werden soll, ist die von HATZER (2000), die die Fähigkeit zum Perspektivenwechsel als Indikator für interkulturelle Handlungskompetenz empirisch bestätigen konnte. HATZER hat dazu die Bedeutung des Perspektivenwechsels als Prädiktor für interkulturelle Handlungskompetenz aufgegriffen. Sie geht davon aus, dass eine grundsätzliche Wahrnehmungssensitivität einer Person überhaupt erst dazu führt, in der interkulturellen Begegnung eine Handlungsproblematik zu erkennen und darauf aufbauend kulturadäquat interagieren zu können. Die Fähigkeit in der kulturellen Überschneidungssituation andere Perspektiven im Zuge einer umfassenden Situationsanalyse einnehmen zu können und bei der Situationsinterpretation zu berücksichtigen, erweitert so die soziale Wahrnehmung dieser Person, was wiederum zu einer besseren Bewältigung der interkulturellen Problemsituation führt. HATZER (2000) konnte diese Annahmen in einer regressionsanalytischen Auswertung beweisen und damit die Fähigkeit zum Perspektivenwechsel als einen wichtigen Indikator für interkulturelle Handlungskompetenz bestätigen.

Die Bestätigung des Perspektivenwechsels als Indikator für interkulturelle Handlungskompetenz auf empirischer Ebene durch HATZER (2000) unterstützt die theoretischen Ausführungen von Kap. 3.2.3. und legt eine genauere Betrachtung des Konstrukts des Perspektivenwechsels nahe. In diesem Zusammenhang sei auf ein Zitat von RUBEN (1976) hingewiesen. Sein Ausdruck „to put oneself in another's shoes“ verdeutlicht sehr anschaulich den Aspekt des Perspektivenwechsels. Er unterstreicht die Fähigkeit, sich zum einen in die Lage eines

fremdkulturellen Gegenübers hineinzuversetzen und zum anderen dessen Gefühle, Gedanken und Erfahrungen angemessen zu reflektieren.

Als **Arbeitsdefinition** des Konzeptes des Perspektivenwechsels dient in Anlehnung an die Definition interkultureller Handlungskompetenz von THOMAS (in Druck) die folgende:

Unter Perspektivenwechsel wird die Fähigkeit verstanden, **andere Perspektiven** unter Berücksichtigung des fremdkulturellen Orientierungssystems **einzunehmen** und diese bei der **Situationsinterpretation zu bedenken**. Dies impliziert die **kritische Reflexion des eigenkulturellen Orientierungssystems** und **das Verlassen gewohnter Denkmuster**, gegebenenfalls auch deren Aufgabe, wenn alternative Denkmuster sich konstruktiver, adäquater oder vorteilhafter erweisen.

Diese umfassende Definition integriert mehrere in der Literatur beschriebene Aspekte des Perspektivenwechsels und ist aus diesem Grund für die vorliegende Arbeit besonders geeignet. Zum einen wird die Übernahme der Sichtweise einer anderen Person, (vgl. THOMAS ET AL., 1997; KIM, 1998 etc.) und zum anderen das Verlassen gewohnter Denkmuster angesprochen (vgl. BRISLIN, LANDIS & BRANDT, 1983) (vgl. Kap. 3.2.3). Die Berücksichtigung der kritischen Reflexion des eigenkulturellen Orientierungssystems, kommt auch in der zuvor gegebenen Definition interkultureller Handlungskompetenz von THOMAS (in Druck) (vgl. Kap. 3.2.4) zum Ausdruck. Er sieht in der kritischen Reflexion des eigenen kulturellen Hintergrundes und der daraus resultierenden Fähigkeit, fremdkulturelle Blickwinkel einzunehmen eine entscheidende Voraussetzung für die weitere Ausprägung interkultureller Handlungskompetenz.

Wie zuvor gezeigt werden konnte (vgl. 3.2.2), verlangen viele Modelle interkultureller Handlungskompetenz bezüglich des Perspektivenwechsels ein Heraustreten aus dem eigenkulturellen Referenzrahmen. Diese Fähigkeit muss vorhanden sein bzw. gelernt werden (BENNET, 1986), um ein interkulturelles Verständnis zu entwickeln. In diesem Zusammenhang wird eine weitere Fähigkeit relevant, die einen engen Bezug zum Perspektivenwechsel aufweist: die Fähigkeit zu isomorphen Attributionen (TRIANDIS, 1975, 1977). Darunter versteht man die Fähigkeit, Verhalten oder Ereignissen eine der Kultur adäquate Erklärung zuzuschreiben. Beide Konzepte hängen nun insofern zusammen, als dass Perspektivenwechsel eine Voraussetzung für isomorphe Attributionen ist, die wiederum eine Voraussetzung für interkulturelle Handlungskompetenz darstellen. Das bedeutet, kulturisomorphe Attributionen werden erst durch die Fähigkeit zum Perspektivenwechsel möglich, oder anders formuliert die Fähig-

keit zum Perspektivenwechsel äußert sich in kulturadäquaten Attributionen. Der Ansatz der isomorphen Attributionen geht auf TRIANDIS (1975, 1977) zurück. Dieser gründet neben attributionstheoretischen Gedanken auf der Notwendigkeit des Perspektivenwechsels, was im nächsten Abschnitt dargelegt wird.

3.3.2 Isomorphe Attribution (TRIANDIS 1975, 1977)

Im Rahmen dieser Arbeit steht die Fähigkeit des Perspektivenwechsels als Indikator für interkulturelle Handlungskompetenz im Vordergrund (vgl. Kap. 3.3.1). Interpersonales Verstehen kann nur erfolgreich sein, wenn es gelingt, sich in die andere Person mit ihrem kulturellen Hintergrund hineinzuversetzen und ihre Absichten und ihr Verhalten nachzuvollziehen (GUDYKUNST & HAMMER., 1983). In diesem Zusammenhang spricht man auch von **isomorpher Attribution** (TRIANDIS, 1975, 1977).

Grundsätzlich gibt es keine richtige oder falsche Erklärung von Verhalten (JONES & NISBETT, 1971), da Wahrnehmung und Interpretation keine Frage von richtig oder falsch sind. In der zwischenmenschlichen Begegnung gibt es immer mehrere Erklärungen und Gründe für gezeigte Reaktions- und Verhaltensweisen eines Interaktionspartners. Eine Voraussetzung für gegenseitiges Verständnis stellen gleiche oder zumindest ähnliche Attributionen dar. In der interkulturellen Psychologie wird davon ausgegangen, dass fehlerhafte Attributionsprozesse über Handlungsursachen und -ziele Missverständnisse in kulturellen Überschneidungssituationen hervorrufen. Diese Missverständnisse entstehen dadurch, dass die Handelnden in einer Überschneidungssituation implizit davon ausgehen, dass sich das Handeln des fremdkulturellen Interaktionspartners auf der Basis des eigenen kulturellen Hintergrunds begründen und auf seine Ziele hin einschätzen lässt. Abweichendes, ungewöhnliches Verhalten des fremdkulturellen Interaktionspartners wird tendenziell dann dessen Persönlichkeit zugeschrieben, weil es auf der Basis des eigenen Orientierungssystems keinen Sinn ergibt (MÜLLER & THOMAS, 1991, S. 9).

TRIANDIS (1975) sieht deswegen die Fähigkeit zu „isomorphen Attributionen“, also das Handeln des fremdkulturellen Gegenübers auf der Grundlage seines Orientierungssystems zu interpretieren, als zentral für eine effektive interkulturelle Orientierung. Er spricht dann von isomorphen Attributionen, wenn Personen Verhalten in der gleichen oder ähnlichen Weise wie der fremdkulturelle Interaktionspartner attribuieren. Dabei muss die Attribution beider Interaktionspartner nicht notwendigerweise identisch sein, aber zumindest so ähnlich, dass durch die Interpretation der Situation keine Konflikte entstehen (TRIANDIS, 1994).

Für TRIANDIS liegt die Ursache interkultureller Probleme nicht darin, dass eine Person eine bestimmte Handlung ausführt, sondern darin, dass ihr fremdkultureller Interaktionspartner diese Handlung anders erwartet, wahrnimmt, interpretiert und wertet. Um Kommunikations- und Interaktionsstörungen zu vermeiden, ist es demnach erforderlich, dass Menschen in interkulturellen Begegnungen „richtig“, also isomorph attribuieren können und somit über die Kenntnis der „subjektiven Kultur“ (TRIANDIS, 1972) des fremdkulturellen Partners verfügen. Während des Attributionsprozesses reicht es nicht aus, Kulturwissen bloß abzurufen, sondern gewohnte Denkmuster zu verlassen und alternative Überlegungen mit einzubeziehen, d.h., es wird die Fähigkeit, multiple Perspektiven einzunehmen gefordert.

Die Fähigkeit zu isomorphen Attributionen und damit zum Perspektivenwechsel unterliegt einem Entwicklungsprozess und benötigt Lernerfahrung, auf die bei zukünftigen kulturellen Überschneidungssituationen zurückgegriffen werden kann. Isomorphe Attributionen sind aus diesem Grund häufig der Gegenstand interkultureller Trainings (Culture Assimilator).

Mit der Betonung des Perspektivenwechsels als Indikator für interkulturelle Handlungskompetenz sowohl auf theoretischer als auch auf empirischer Ebene wurde der Ausgangspunkt zur Entwicklung eines Tests zur Erfassung interkultureller Handlungskompetenz geschaffen. Das nächste Kapitel widmet sich nun den Anforderungen, denen ein solcher Test gerecht werden muss. Dazu werden zum einen aus der Analyse der Testverfahren und zum anderen aus den dargestellten theoretischen Ausführungen Kriterien für die Testentwicklung abgeleitet.

4 Kriterien der Testentwicklung

4.1 Ableitungen aus der Analyse der Testverfahren

Eine Anforderung abgeleitet aus den Analyseverfahren betrifft die empirische Überprüfung des TIHK auf testtheoretische Gütekriterien. Dazu gehört die Überprüfung der Validität (Gültigkeit) sowie die Überprüfung der Reliabilität (Genauigkeit), d.h., ob und wie genau dieser Test die Fähigkeit des Perspektivenwechsels bei Personen zu erfassen vermag.

Eine Forderung bezüglich interkultureller Testverfahren betrifft die Festlegung auf bestimmte Indikatoren und deren hinreichender Operationalisierung. Der TIHK hat sich dementsprechend zum Ziel gesetzt, die empirisch und theoretisch abgesicherte Fähigkeit zum Perspektivenwechsel als Indikator für interkulturelle Handlungskompetenz zu erfassen. Klassischer Weise kann die Fähigkeit zum Perspektivenwechsel anhand isomorpher Attributionen in KIs erfasst werden. Dies ist deswegen zulässig, da die Fähigkeit zum Perspektivenwechsel eine Voraussetzung für kulturadäquate Ursachenzuschreibungen darstellt. Zu beachten ist im Zusammenhang mit dem TIHK, dass dieser im Rahmen des Zusatzstudiums „Internationale Handlungskompetenz“ eingesetzt werden kann, um die Fortschritte der Studenten innerhalb des Studiums zu erfassen.

Hinsichtlich der Methodik hat der TIHK nicht zum Ziel, Persönlichkeitseigenschaften anhand bereits vorhandener Skalen zu messen (vgl. Kap. 2.1.1). Auch ein Verfahren in Form von Selbsteinschätzungen erscheint im Rahmen der Evaluation des Zusatzstudiums wenig sinnvoll. Zum einen birgt sich hier die Tendenz zur sozialen Erwünschtheit im Antwortverhalten der Studierenden und zum anderen die Gefahr von Verzerrungen zwischen Selbsteinschätzung und tatsächlicher Leistung (vgl. Kap. 2.1.2)

4.2 Ableitungen aus den theoretischen Grundlagen

Perspektivenwechsel ist eine entscheidende kognitive Fähigkeit interkulturellen Lernens (THOMAS ET AL., 1997, HELFRICH, 1996, BHAWUK & BRISLIN, 1992) und ein Prädiktor für interkulturelle Handlungskompetenz. Interkulturelle Forschung beschäftigt sich häufig mit der Förderung des Perspektivenwechsels mittels interkultureller Trainings. Die Forschergruppe um THOMAS in Deutschland hat sich mit der Entwicklung des Perspektivenwechsels und dessen Eignung als Prädiktor für interkulturelle Handlungskompetenz auseinandergesetzt. Die Erfassung von interkultureller Handlungskompetenz anhand der Fähigkeit zum Perspektivenwechsel mittels geeigneter Messinstrumente wurde bis dato in der Literatur noch nicht thematisiert. Aus diesem Grund hat sich die vorliegende Arbeit zum Ziel gesetzt, ein Instru-

ment zu entwickeln, welches die Fähigkeit zum Perspektivenwechsel bei Personen messen kann. Mit Hilfe des zu entwickelnden Tests sollen Aussagen darüber getroffen werden, ob Personen Perspektivenwechsel in einer interkulturellen Begegnung vollziehen können oder nicht. Diese Fähigkeit wird dann als eine Voraussetzung für interkulturelle Handlungskompetenz gewertet.

4.2.1 Zur Erfassung des Perspektivenwechsels

Bei der Fähigkeit zum Perspektivenwechsel handelt es sich, wie bereits erwähnt, um eine kognitive Fähigkeit. Da interkulturelle Handlungskompetenz jedoch auf mehreren Dimensionen zum Ausdruck kommt (vg. Kap. 3.2.5), ist es erstrebenswert in Bezug auf ihre Erfassung auch affektive oder verhaltensbezogene Aspekte zu berücksichtigen. Aus diesem Grund wird bei der Erfassung des Perspektivenwechsels darauf geachtet einen Handlungsbezug herzustellen. Eine Möglichkeit, diesen Handlungsbezug bei der Erfassung des Perspektivenwechsels zu erreichen, ist die Berücksichtigung der Situation, in der Personen verschiedener Kulturen interagieren. Auf die Bedeutung der Situation in einer interkulturellen Begegnung wird in der Diskussion um die verschiedensten Facetten von interkulturellem Lernen, Verstehen oder Handeln allgemein hingewiesen. In diesem Zusammenhang betont THOMAS (in Druck), dass Erfahrungen in als kritisch erlebten interkulturellen Interaktionssituationen und den darauf folgenden Reaktionen der entscheidende Anstoß für das Entstehen dieses Lernprozesses sind. KRUMING (2000) dagegen hebt die Bedeutung der Situation im Speziellen für die Fähigkeit zum Perspektivenwechsel hervor: „Eine der entscheidenden Voraussetzungen, damit ein Perspektivenwechsel aktiviert wird, ist die Situation, in der die normalen erlernten Verhaltensmuster und Wertmaßstäbe nicht mehr anwendbar erscheinen“ (S. 43).

Für die Berücksichtigung der Situation bei der Erfassung des Perspektivenwechsels spricht zudem auch die dieser Arbeit zugrunde liegende interaktionistische Sichtweise interkultureller Handlungskompetenz: Da personale und situationale Aspekte interkulturelle Handlungskompetenz bedingen, ist die Erfassung einer (Personen)Fähigkeit in der Situation notwendig.

Eine Möglichkeit, die Fähigkeit zum Perspektivenwechsel unter Berücksichtigung der dargestellten Kriterien (interaktionistische Sichtweise eines mehrdimensionalen Konstrukts) zu erfassen, bietet die Methode der kritischen Interaktionssituationen (FLANAGAN, 1954), wie an den folgenden Ausführungen verdeutlicht wird.

4.2.2 Die Methode der kritischen Interaktionssituationen zur Erfassung des Perspektivenwechsels

Für eine Verwendung von kritischen Interaktionssituationen zur Erfassung des Perspektivenwechsels als Indikator für interkulturelle Handlungskompetenz sprechen verschiedene Erkenntnisse in der interkulturellen Forschung, die im Folgenden dargestellt werden.

Da eine kritische Interaktionssituation Handlungen in einer interkulturellen Überschneidungssituation thematisiert, stellt sie eine geeignete Methode dar, interkulturelle Handlungskompetenz zu erfassen. Mit ihr wird der Forderung von KEALEY (1989) nachgekommen, dass die Instrumente zur Erfassung interkultureller Handlungskompetenz verhaltensverankert, also in einer konkreten Situation, messen sollen.

Ebenso unterstreicht CUSHNER (1989) im Zusammenhang mit der Culture Assimilator Methode die Bedeutung kritischer Interaktionssituationen für den Prozess interkulturellen Lernens, der wiederum Voraussetzung für interkulturelle Handlungskompetenz ist. Er vertritt die Meinung, dass das Lösen von kritischen Interaktionssituationen der Art und Weise entspricht, in der Personen Kulturwissen und kulturelle Unterschiede erlernen. Lerntheoretiker haben die Bedeutung von Vorwissen und Vorerfahrungen auf den Prozess der Aneignung von neuem Wissen identifiziert. Da nicht jedes Individuum über interkulturelle Erfahrungen und Wissen verfügt, bietet der Culture Assimilator durch das Lösen von KIs eine gute Ausgangslage, um neues interkulturelles Wissen aufzubauen. Durch die Darstellung von Alltagserfahrungen in KIs kann sich der Lernende besser in die Situationen hineinversetzen und diese auch besser erinnern als z.B. Listen von Do's und Don'ts, wie sie in Vorbereitungsseminaren präsentiert werden: „The Critical incident format presents real-life experiences in the form of short scripts that are easy to recall, therefore making it easy for trainees to integrate the experience as their own. Further learning can then develop from this starting point“ (CUSHNER, 1989, S. 139). Wenn kritische Interaktionssituationen gut geeignet sind, interkulturelles Wissen aufzubauen, liegt der Umkehrschluss nahe, dass es möglich ist, dieses Wissen auch in kritischen Interaktionssituationen abrufen zu können. Diese Annahme findet ebenfalls in der Studie von CUSHNER (1989) Bestätigung, der zu dem Schluss kommt, dass sich die Methode der kritischen Interaktionssituationen in Form des Culture Assimilators für die Erfassung von Wissenszuwachs eignet: „Critical incidents are able to provide an effective dependent variable measure of knowledge gain“ (S. 139).

Auch die Ergebnisse von Evaluationsstudien sprechen für die Verwendung kritischer Interaktionssituationen zur Erfassung der Fähigkeit zum Perspektivenwechsel bzw. zu isomorphen

Attributionen. Laut den Studienergebnissen von BIRD ET AL., 1993; CUSHNER, 1989; RANDOLPH ET AL., 1977 (zitiert nach KINAST, 1998) steht die Fähigkeit der isomorphen Attribution als eine wichtige Komponente interkultureller Handlungskompetenz in einem positiven Zusammenhang mit der Teilnahme an einem Culture Assimilator Training, welches aus kritischen Interaktionssituationen besteht. Zum Lösen der KIs im Culture Assimilator wird die Fähigkeit zu isomorphen Attributionen gefordert, die wiederum nur möglich sind, wenn Personen in der Lage sind, die Perspektive des fremdkulturellen Handlungspartners zu übernehmen.

Die in diesem Abschnitt angeführten Gründe sprechen für einen Einsatz kritischer Interaktionssituationen zur Erfassung des Perspektivenwechsels als Indikator für interkulturelle Handlungskompetenz. Die Methode, die die Grundlage für die Entwicklung eines Tests zur Erfassung interkultureller Handlungskompetenz (TIHK) bildet, wird im folgenden Kapitel dargestellt.

5 Die Methode zur Entwicklung des TIHK

Die Aufgabe bei der Entwicklung des Tests zur Erfassung interkultureller Handlungskompetenz (TIHK) besteht in der Konstruktion zweier paralleler Testformen, deren Testaufgaben das Konstrukt des Perspektivenwechsels als Indikator für interkulturelle Handlungskompetenz zu erfassen vermögen. Es gilt also adäquate Aufgaben auszuwählen, die zum einen den testtheoretischen Gütekriterien entsprechen und zum anderen der Prüfung auf Äquivalenzkriterien standhalten.

Die Entwicklung eines Paralleltests zur Erfassung interkultureller Handlungskompetenz (TIHK) umfasst mehrere Schritte, die im Einzelnen in den folgenden Abschnitten dargelegt werden. Zunächst wird auf das Culture Assimilator Material eingegangen, das aufgrund der theoretischen Überlegungen dieser Arbeit als methodisches Konzept ausgewählt wurde und somit das Grundgerüst für die Konstruktion des Tests darstellt. In einem nächsten Schritt wird ein Überblick über die Methode der Paralleltestkonstruktion allgemein gegeben und ein Ablaufschema der praktischen Entwicklungsarbeit zur Orientierung für den Leser dargestellt. Die sukzessive Entwicklung und die praktische Umsetzung der Testform A sowie ihrer Parallelförm B bilden den Schwerpunkt dieses Kapitels. Dabei wird auf die Bestimmung der testtheoretischen Gütekriterien und der Äquivalenzkriterien für Paralleltests eingegangen.

5.1 Das Culture Assimilator Material

Das Testmaterial zur Entwicklung des TIHK wird anhand des Konzeptes der kritischen Interaktionssituationen konstruiert, die wiederum aus einem spezifischen Trainingsformat zur Vorbereitung auf den Auslandsaufenthalt, dem Culture Assimilator entnommen wurden. Im Folgenden wird nun zum besseren Verständnis für den Leser die Methode des Culture Assimilators vorgestellt.

Der Culture Assimilator ist das kulturelle Trainingsformat, dem in der interkulturellen Forschung am meisten Interesse zu Teil wurde (BRISLIN & PEDERSON, 1976). Der Culture Assimilator findet seit seiner Entwicklung international weite Verbreitung und Anwendung. Das Konzept wurde 1962 an der University of Illinois von den Psychologen FIEDLER, OSGOOD, TRIANDIS UND STOLUROW (FIEDLER ET AL., 1971) als kulturspezifisches Training entwickelt und in den 80er Jahren von BRISLIN ET AL. (1986) um das Konzept kulturallgemeiner Culture Assimilator zur Sensibilisierung für fremdkulturelle Orientierungssysteme ergänzt. Der Culture Assimilator soll primär kognitive Komponenten interkultureller Handlungskompetenz

aufbauen und zählt nach THOMAS (1995) zu den verstehensorientierten interkulturellen Trainings.

Inhaltlich betrachtet besteht der Culture Assimilator aus verschiedenen Trainingsabschnitten, die sich jeweils auf einen zentralen Kulturstandard (THOMAS, 1993) beziehen. Er beinhaltet aufbereitete Fallbeispiele aus der interkulturellen Begegnungspraxis, die kritische Interaktionssituationen zwischen Mitgliedern der eigenen Kultur und Mitgliedern der fremden Kultur beschreiben (vgl. Anhang A und B). Aufgabe des Trainingsteilnehmers ist es dabei, für das gezeigte Verhalten des fremdkulturell geprägten Handlungspartners eine kulturadäquate Erklärungsalternative zu finden. Die dargestellte Situation im Culture Assimilator zeichnet sich dadurch aus, dass sie von den Beteiligten als konflikthaft oder unverständlich erlebt wird und dass sich der dargestellte Konflikt in kulturellen Unterschieden begründet (vgl. Kap. 3.1.2). Im Anschluss an die Situationsschilderung folgen mehrere Erklärungsalternativen für das Verhalten der beteiligten Personen, aus denen diejenige Interpretation ausgewählt werden soll, die aus der Sicht des fremdkulturellen Interaktionspartners am ehesten zutrifft. Dabei entspricht eine Erklärungsalternative genau der Sichtweise der fremdkulturellen Person. Die anderen dagegen beinhalten Fehlinterpretationen, die typisch für die eigene Kultur sind und auf Unkenntnis kultureller Einflussfaktoren, auf ethnozentrischem Denken oder Vorurteilen beruhen. Nach der Entscheidung für eine Erklärungsalternative bekommt der Trainingsteilnehmer ein Feedback bezüglich seiner Wahl (ob diese richtig oder falsch war) und zusätzlich Informationen über wesentliche Merkmale des jeweils zutreffenden zentralen Kulturstandards. Auf diese Weise lernt der Trainingsteilnehmer, das gezeigte Verhalten des fremdkulturellen Gegenübers aus dessen Perspektive zu betrachten, er lernt also kulturisomorph zu attribuieren. Der TIHK nutzt nun diese Methode, um festzustellen, ob und inwieweit die Testperson in der Lage ist die fremdkulturelle Perspektive einzunehmen und somit fähig zu kulturisomorphen Attributionen ist. Dazu wird das Antwortverhalten der Probanden bei jeder KI analysiert und überprüft, ob die kulturadäquate Interpretation gewählt wurde. Die genaue Auswertung des Antwortverhaltens der Probanden im TIHK folgt in Kapitel 5.2.2.1.

Die häufige Kritik bezüglich der Vereinfachung und Generalisierung komplexer interkultureller Konflikte in Culture Assimilator Trainings wie bspw. durch das Element des Lösungsweges kritischer Interaktionssituationen in 3 „falsche“ (kulturinadäquate) und nur eine „richtige“ (kulturadäquate) Antwortalternative, hat zu neueren Ansätzen der Culture Assimilator Methode geführt.

So sind z.B. an der Universität Regensburg in der Abteilung für Sozial- und Organisationspsychologie im Institut für Psychologie die Culture Assimilator abweichend von der U.S.-amerikanischen Standardversion inzwischen so aufgebaut, dass nach jeder kritischen Inter-

aktionssituation nicht einfach nur 3 „falsche“ und eine „richtige“ Antwortalternative zur Auswahl vorgegeben werden. Vielmehr wird der Trainingsteilnehmer gebeten, zu jeder Antwortalternative auf einer 4-stufigen Skala von „am meisten zutreffend“ über „eher zutreffend“, „eher nicht zutreffend“ bis „am wenigsten zutreffend“ anzukreuzen, inwieweit er die angebotene Antwortalternative als kulturisomorph ansieht. Zudem werden die Kulturstandards mit mehreren Sätzen genauer beschrieben, was im Einzelnen darunter zu verstehen ist und auf welche kulturhistorischen Entwicklungsprozesse und Verankerungen sie zurückgeführt werden können. Hinzu kommen Schilderungen über kulturadäquate Lösungsmöglichkeiten der als problematisch geschilderten kritischen Interaktionssituation.

Da es fraglich ist, ob ein dichotomisiertes „Richtig-oder-Falsch“-Lösungsschema der interkulturellen Realität tatsächlich entspricht, bezieht sich der TIHK auf die neueren Ansätze der Culture Assimilator Methode. Er nutzt kritische Interaktionssituationen verschiedener Culture Assimilators, um Perspektivenwechsel als Indikator für interkulturelle Handlungskompetenz bei Personen zu ermitteln. Es werden, wie oben theoretisch erläutert, mehrere Situationen eines kritischen Zusammentreffens Deutscher mit Menschen anderer Kulturen beschrieben. An jede dieser Situationen schließen sich 4 Antwortalternativen an, die mögliche Gründe für das jeweilige Verhalten nennen. Der Proband hat dann die Aufgabe, die Antwortalternativen bezüglich ihrer Güte in eine Reihenfolge zu bringen. Dabei muss er entscheiden, welche der 4 Antworten aus Sicht des fremdkulturellen Interaktionspartners am meisten zutrifft, eher zutrifft, eher nicht zutrifft und am wenigsten zutrifft. Auf diese Weise möchte der TIHK eruieren, wie gut der Proband sich in die Situation hineinversetzen kann und in der Lage ist, diese aus der Sicht des fremdkulturellen Partners zu erklären. Dabei erweist sich die Berücksichtigung der Rangreihe als vorteilhaft, da mehr Informationen über den Probanden gewonnen werden, als bei einem „Richtig-oder-Falsch“-Lösungsschema. Der TIHK gibt Aufschluss darüber, ob der Proband die KI in seiner Tendenz richtig erkannt hat oder nicht und ermöglicht damit ein differenzierteres Bild des zu messenden Konstrukts des Perspektivenwechsels. Die Auswahl der Testaufgaben (KIs) und die Zusammenstellung von Test A zur Vorlage der Analysestichprobe werden im Kapitel „Die Praktische Entwicklungsarbeit“ (vgl. Kap. 6.1) beschrieben.

5.2 Die Methode der Paralleltestkonstruktion

Es gibt grundsätzlich 2 Möglichkeiten Paralleltests aufzubauen (LIENERT & RAATZ, 1994). Der Simultanaufbau eines Tests berücksichtigt bereits in der Planung die Konstruktion zweier paralleler Testformen, indem im Rahmen der Aufgabenanalyse eine möglichst große Anzahl

von Aufgaben analysiert wird. Die qualifizierten Aufgaben werden im Anschluss zu Paaren gleicher Schwierigkeit und gleicher Trennschärfe geordnet. Reliabilität und Validität für beide Formen werden in einem Verfahren ermittelt. Mit Hilfe des Sukzessivaufbaus hingegen werden die Paralleltestformen nacheinander entwickelt. Man geht hier nach LIENERT UND RAATZ (1994) im Allgemeinen so vor, dass man zu jeder einzelnen Aufgabe eine zweite konstruiert, „die der ersten nach bestem Ermessen inhaltlich und formal ähnelt (...)“ (S.307). In einem zweiten Schritt werden möglichst alle Aufgaben beider Parallelformen einer erneuten Aufgabenanalyse unterzogen. Danach wird entschieden, ob man die Aufgaben in den Parallelformen belässt oder sie nach Zufall neu aufteilt. Bezüglich deren Validität vertraut man auf die gleiche Validität wie in der ersten Testform. Die Reliabilität der neuen Testform kann über das Halbierungsverfahren, die Konsistenzanalyse oder die Paralleltestmethode ermittelt werden. Die Entwicklung des Tests zur Erfassung interkultureller Handlungskompetenz (TIHK) erfolgt nach dem Prinzip des Sukzessivaufbaus. Das folgende Ablaufschema veranschaulicht den Prozess der Paralleltestkonstruktion des TIHK:

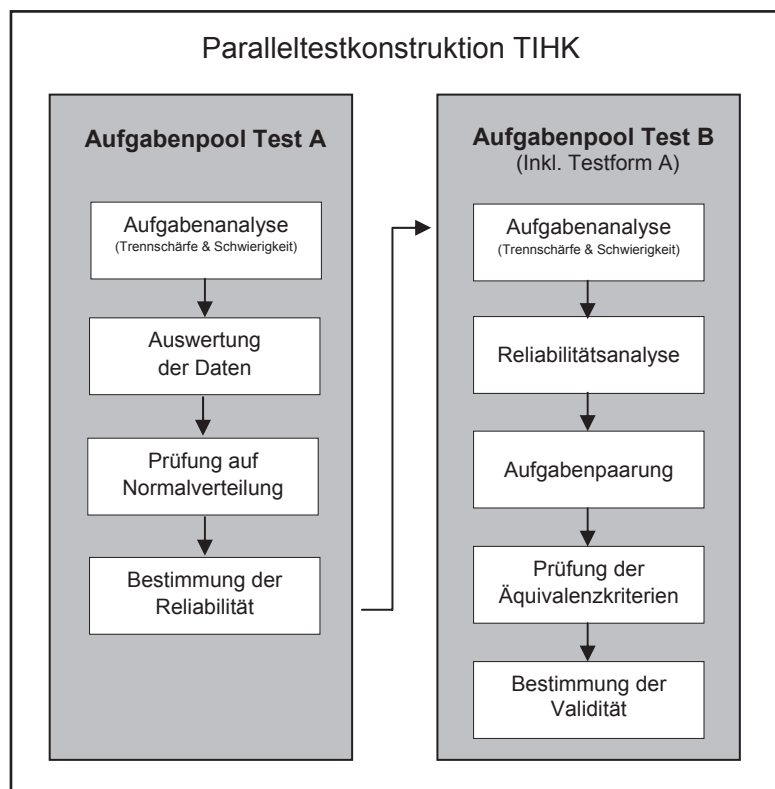


Abb. 1: Ablaufschema der Paralleltestkonstruktion des TIHK

5.2.1 Die Aufgabenanalyse

Unter der Aufgabenanalyse versteht man die Darbietung der Testvorform an einer Analysestichprobe. Sie gibt Aufschluss über die Güte der verwendeten Testaufgaben und dient der Selektion von Aufgaben für die Testendform. Die Aufgabenanalyse verfolgt nach LIENERT UND RAATZ (1994) den Zweck der Revision nicht geeigneter Testaufgaben hinsichtlich einer höheren Reliabilität und Validität sowie die Überprüfung der Testpunktverteilung und Durchführungsmodalitäten des Tests. Die Testpunktverteilung sollte für den Fall einer späteren Eichung des Tests eine hinreichende Streuung, Symmetrie und Eingipfligkeit wie bei einer Normalverteilung aufweisen. Für die Aufgabenanalyse sollte die Wahl der Analysestichprobe in höchstmöglichem Maße für das später zu testende Kollektiv repräsentativ sein. Weiterhin sollte bei reinen Niveautests nach LIENERT UND RAATZ (1994) eine Gesamtdarbietung der Testaufgaben erfolgen, d.h., die Testvorform wird bei konstanter Reihenfolge der Aufgaben vollständig der Analysestichprobe zur Beantwortung dargeboten, wobei jeder Proband sich für jede Aufgabe theoretisch so viel Zeit nehmen kann, wie er dazu benötigt. Im Rahmen der Aufgabenanalyse werden nun Schwierigkeit und Trennschärfe der Testaufgaben ermittelt, um die geeigneten Aufgaben für die Testendform auszuwählen.

5.2.1.1 Schwierigkeitsanalyse

Nach LIENERT UND RAATZ (1994) ist der Schwierigkeitsindex einer Aufgabe gleich dem prozentualen Anteil P der auf diese Aufgabe entfallenden richtigen Antworten in einer Analysestichprobe von der Größe N ; der Schwierigkeitsindex liegt also bei schwierigen Aufgaben niedrig, bei leichten hoch. Die Schwierigkeit einer Aufgabe wird hier aufgrund wissenschaftlicher Sparsamkeit anhand dichotomer Variablen bestimmt, da es letztendlich um die Frage geht, ob der Proband das Item richtig oder falsch beantwortet hat. Eine Aufgabe gilt hierbei als „richtig“ beantwortet, wenn der Proband die „richtige“ Erklärungsalternative mit „am meisten zutreffend“ beantwortet hat. Aus diesem Grund wird hier das komplette Antwortverhalten, nämlich die Rangreihenbildung des Probanden nicht berücksichtigt. Die Berechnung des Schwierigkeitsindex bei Niveautestaufgaben (Aufgaben, die entweder überhaupt keine Testzeit vorschreiben oder eine solche, innerhalb der alle Probanden alle Aufgaben bearbeiten können) wird nach LIENERT UND RAATZ (1994) nach folgender Definition berechnet:

$$P = 100 \times \frac{N_R}{N}$$

Das bedeutet, man berechnet den Prozentsatz der Aufgaben, die richtig beantwortet werden und setzt diesen ins Verhältnis zur Gesamtzahl der Probanden. Der Schwierigkeitsindex einer Aufgabe kann sich demnach zwischen den Werten 0 (von keinem Teilnehmer gelöst, extrem schwierig) und 100 (von allen Teilnehmern gelöst, extrem leicht) bewegen. Normalerweise wird bei Testkonstruktionen angestrebt, dass die Schwierigkeitsindizes über den gesamten Bereich streuen und in etwa normalverteilt sind, d.h., einige leichte und schwere Aufgaben sowie viele mittelschwere Aufgaben im Test enthalten sind. Eliminiert werden sollen nach LIENERT UND RAATZ (1994) die Aufgaben, die außerhalb des Schwierigkeitsbereiches von $95 > P > 5$ liegen, da sie mit hoher Wahrscheinlichkeit nicht trennscharf sind. Bei Aufgaben außerhalb des Schwierigkeitsgrads $80 > P > 20$ ist zumindest Vorsicht geboten.

Im Rahmen der Schwierigkeitsanalyse sollen demnach die Aufgaben mit einem mittleren Schwierigkeitsgrad herausgefiltert werden. Dies ist deswegen erforderlich, da zwischen Schwierigkeit und Trennschärfe eine paraboloiden Abhängigkeit besteht. Damit ist gemeint, dass bei geringer Schwierigkeit einer Aufgabe auch ihre Trennschärfe gering ist, die aber mit ansteigender Schwierigkeit wachsen kann, bis sie bei einer mittleren (50 %igen) Schwierigkeit ihr Maximum erreicht und dann bei weiter ansteigender Schwierigkeit wieder abnimmt bis zu einem Minimum bei höchster Schwierigkeit. Aufgaben von mittlerer Schwierigkeit bieten demnach die besten Voraussetzungen für eine hohe Trennschärfe.

5.2.1.2 Trennschärfeanalyse

Für die Güte einer Aufgabe verwenden LIENERT UND RAATZ (1994) den Begriff der Trennschärfe, die ein Maß dafür ist, wie gut eine Aufgabe „gute“ Probanden mit hohem Gesamtpunktwert von „schlechten“ Probanden mit niedrigem Punktwert unterscheidet.

Dazu sind in einem ersten Schritt nach LIENERT UND RAATZ (1994) solche Aufgaben zu eliminieren, die von „guten“ und „schlechten“ Probanden gleich häufig richtig beantwortet werden (dies entspricht einem Trennschärfe-Koeffizienten von 0) oder die von „schlechten“ Probanden häufiger richtig beantwortet werden als von „guten“ Probanden (dies entspricht einem negativen Trennschärfe-Koeffizienten). Statistisch gesehen ist „die Trennschärfe eines Items (...) durch die Korrelation der Itemlösungen mit den Gesamtpunktwerten [, die gewöhnlich aus der Summe aller Itemwerte berechnet werden,] der Probanden definiert“ (ROGGE, 1995, S. 94). Die Korrelation des Einzelitems mit dem Summenwert gibt an, ob das Item inhaltlich wirklich zu der Skala passt (SCHMID, 1992).

Nach LIENERT UND RAATZ (1994) wird der Trennschärfe-Koeffizient bei einem quantitativen Analyseverfahren und vollständiger Aufgabendarbietung mit der Methode der so genannten punktbiserialen Korrelation berechnet. Erreichen die Variablenwerte jedoch Intervallniveau,

und ist die Beziehung zwischen Variablenwert und Analysekriterium linear, kann die Trennschärfe auch als Produkt-Moment-Korrelation berechnet werden (FISSENI, 1997). Im Falle der vorliegenden Arbeit handelt es sich hinsichtlich der Variablen als auch hinsichtlich des Analysekriteriums um intervallskalierte Daten, so dass als Trennschärfe-Koeffizient die Produkt-Moment-Korrelation berechnet wird. Mit Hilfe der Berechnung der Trennschärfe der einzelnen Aufgaben soll geklärt werden, welche der kritischen Interaktionssituationen im Rahmen der Untersuchung sich für die Erfassung interkultureller Handlungskompetenz am besten eignen.

MUMMENDEY (1995) berichtet, dass Fragebogenkonstrukteure in der Praxis häufig die Signifikanz des Korrelationskoeffizienten als Auslesekriterium wählen, obwohl sie seiner Meinung nach eigentlich keine sinnvolle Größenangabe darstellt. Es sei „im wesentlichen dem Geschick des Forschers und seinem Anspruchsniveau überlassen, welches Trennschärfe-Kriterium er festlegt“ (MUMMENDEY, 1995, S. 74). Bei BORTZ UND DÖRING (1995) finden sich dagegen Angaben, dass Werte zwischen 0,30 und 0,50 auf eine mittlere Trennschärfe, Werte ab 0,50 auf eine hohe Trennschärfe hinweisen. Diese Kriterien werden auch für die Analyse im Rahmen der vorliegenden Arbeit angesetzt. Abschließend ist zu beachten, dass die Trennschärfe-Koeffizienten so hoch wie möglich ausfallen sollten, denn: „Je höher die Trennschärfe-Koeffizienten liegen, desto höher ist die Reliabilität; und unter sonst gleichen Bedingungen ist auch die Validitätserwartung höher“ (LIENERT & RAATZ, 1994, S. 59).

Zusammenfassend lässt sich bezüglich der Aufgabenanalyse Folgendes festhalten: Für Niveautests ist nach LIENERT UND RAATZ (1994) im Allgemeinen die Trennschärfe von größerer Bedeutung als die Schwierigkeit. Die Trennschärfe soll möglichst hoch sein, wobei Aufgaben mit fehlender oder sogar negativer Trennschärfe eliminiert werden müssen. Die Schwierigkeit dagegen soll entweder etwa bei 50 % liegen oder über einen möglichst großen Bereich um 50 % streuen, so dass Aufgaben von minimaler wie solcher von maximaler Schwierigkeit im Test vorkommen. Wenn es sich um ziemlich homogene Testaufgaben mittlerer Trennschärfe ($r \approx 0,6$) handelt, soll die Verteilung der Schwierigkeitsindices eher breitgipflig sein; bei heterogenen Aufgaben mit einer eher geringeren Trennschärfe ($r \approx 0,3$) soll sie erfahrungsgemäß eher schmalgipflig sein. Da die simultane Selektion von Testaufgaben nach Trennschärfe und Schwierigkeit nicht einfach ist und nicht selten mit vielerlei Kompromissen endet (LIENERT & RAATZ 1994), wird von der Autorin im Zweifelsfall auch immer die psychologische Validität der einzelnen Aufgaben berücksichtigt. D.h., die Aufgaben werden zugunsten der inhaltlichen Ausgeglichenheit bevorzugt, so dass aus jeder Kultur die gleiche Anzahl kritischer Interaktionssituationen bearbeitet werden müssen.

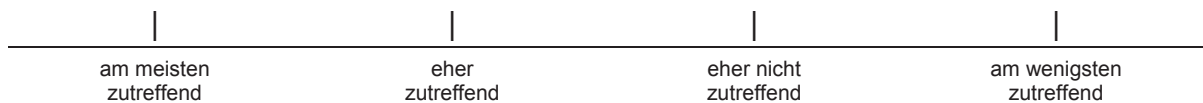
5.2.2 Die Auswertung der Daten

5.2.2.1 Ermittlung der Rohwerte

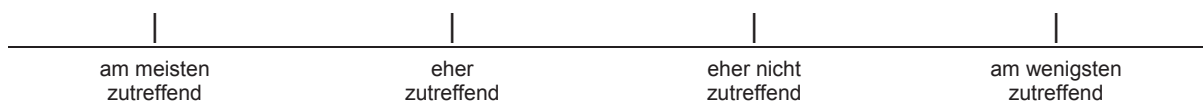
Wie zu Beginn dieses Kapitels näher erläutert, gehen neuere Konzepte der Culture Assimilator Methode davon aus, dass sich die 4 Erklärungsalternativen zur Lösung einer kritischen Interaktionssituation im Hinblick auf ihre Güte in eine Rangreihe bringen lassen. Die Aufgabe der Probanden ist es, diese Rangreihe der Antwortalternativen zu erkennen, d.h., sie müssen pro Erklärungsalternative auf einer 4-stufigen Skala entscheiden, welche der 4 Antworten am meisten zutrifft, eher zutrifft, eher nicht zutrifft und am wenigsten zutrifft. Zur Veranschaulichung der Antwortskala der Aufgaben im TIHK dient das folgende Beispiel:

Wie erklärst Du Dir das Verhalten der Engländer?

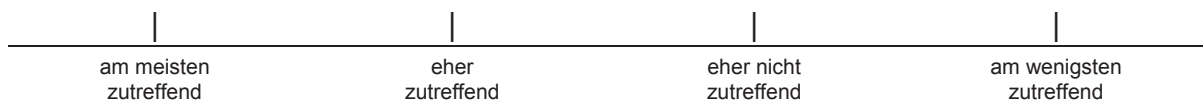
- a) Jemanden auf einen Fehler hinzuweisen ist eine Einmischung in des anderen Angelegenheiten; sogar wenn es einen selbst betrifft, hält man sich wenn möglich zurück.



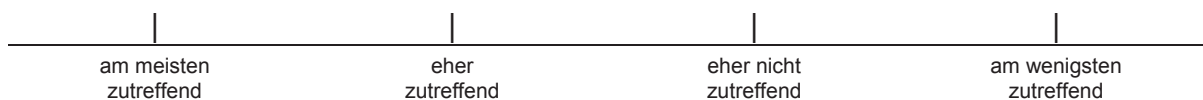
- b) Den Engländern ist es einfach nicht so wichtig, wer in welcher Reihenfolge an die Reihe kommt.



- c) Haltung zu bewahren und keine Szene zu machen ist ein wichtiges Ideal in England.



- d) Die Engländer hätten den Kassierer darauf hingewiesen, dass die Deutschen gedrängelt haben. Sie hätten sich nicht direkt mit den Drängelnden auseinandergesetzt.



Zur Auswertung von Culture Assimilators oder entsprechenden Attributionstests gibt es grundsätzlich verschiedene Möglichkeiten. Geht man davon aus, dass bei jeder kritischen Interaktionssituation genau eine Antwort die Situation richtig und vollständig erklärt, bekommt der Proband dann einen Punkt, wenn er diese „richtige“ Antwort auf der vorgegebenen Skala höher als alle anderen Erklärungsalternativen einstuft.

Bei Culture Assimilators, die nach einem „verzweigten Beantwortungsformat“ (MALPASS & SALANCIK, 1977) bearbeitet werden, bekommt der Proband Punkte in Abhängigkeit davon, beim wievielten Ankreuzen einer Erklärungsalternative er die „richtige“ Antwort trifft. Beim „verzweigten Beantwortungsformat“ (siehe Studie von LANDIS ET AL., 1976) muss der Proband sich zunächst für eine Antwort entscheiden und liest den dazugehörigen Kommentar auf der nächsten Seite. War die Entscheidung richtig, ist die Bearbeitung dieser kritischen Interaktionssituation beendet. War die Entscheidung falsch, so muss er sich für eine andere Erklärungsalternative entscheiden. Dies geht so lange, bis er die richtige Erklärung trifft. Bezüglich der Auswertung fällt die zu erreichende Punktzahl umso niedriger aus, je eher der Proband die richtige Antwort findet.

Eine weitere Art und Weise der Auswertung findet sich bei LAYES (1995), der davon ausgeht, dass sich die 4 Erklärungsalternativen hinsichtlich der Güte der Antworten in eine Rangreihe bringen lassen. Dass dies ein sinnvoller Ansatz ist, wird von LAYES (1995) durch verschiedene Angaben in der Literatur belegt. Zum einen ist fraglich, ob ein dichotomisiertes „Richtig-oder-Falsch“-Lösungsschema der interkulturellen Realität tatsächlich entspricht (vgl. Kap. 5.1), zum anderen birgt es nach WINTER (1994) die Gefahr einer zu engen Typisierung fremdkulturellen Verhaltens. Darüber hinaus zeigen andere Culture Assimilators bzw. Attributionstests (z.B. der „Test of Intercultural Sensitivity“ (TICS) von WELDON ET AL., 1975), dass auch unter den Angehörigen der Zielkultur keineswegs völlige Einigkeit darüber herrscht, wie bestimmte kulturspezifische Verhaltensweisen erklärt werden können.

Beim Auswertungsmodus nach WELDON ET AL. (1975) erhält jeder Teilnehmer für die von ihm gewählte Bestantwort die Punktzahl, die dem Rangplatz dieser Antwort in der „wahren“ Rangreihe entspricht. Dort ist nach LAYES (1995) ein weiterer Schwachpunkt dieses Verfahrens zu finden: „Dadurch, dass nur die Antwort bewertet wird, die der Teilnehmer auf den ersten Rangplatz gesetzt hat, geht viel Information verloren, da nicht berücksichtigt wird, auf welche Rangplätze der Teilnehmer die 3 anderen Antwortalternativen gesetzt hat. (...) Zur Illustrierung dieses Sachverhalts kann man sich folgenden Extremfall vorstellen: Angenommen, ein Teilnehmer würde immer die richtige Bestantwort treffen, gleichzeitig aber immer die schlechteste Antwort auf den zweiten Platz setzen und die zweitbeste Antwort auf den letzten Platz. Dieser Teilnehmer bekäme nach dem Auswerteverfahren von WELDON ET AL. (1975) die maximale Gesamtpunktzahl, womit seine Kompetenz aber offensichtlich überschätzt würde“ (S. 41).

Vor dem Hintergrund dieser Überlegungen wird bei LAYES (1995) zur Auswertung die komplette Rangreihe berücksichtigt, mit der ein Teilnehmer eine Aufgabe beantwortet. Dazu wird für jede Aufgabe die vom Teilnehmer gewählte Rangreihe der Antwortalternativen mit der

zuvor ermittelten „wahren“ Rangreihe verglichen. Anschließend werden die absoluten Differenzen für jede Aufgabe berechnet und aufsummiert, so dass sich für jeden Teilnehmer ein Differenzmaß zwischen 0 und 96 Punkten ergibt. Um die Interpretation der Testergebnisse zu erleichtern, wird die Gesamtfehlerzahl noch in eine „positive“ Punktzahl transformiert, indem die Gesamtfehlerzahl vom Wert 100 subtrahiert wird. Anhand dieses Fehlerpunktwertes kann gesehen werden, wie gut oder schlecht ein Proband die kritischen Interaktionssituationen lösen konnte: Je höher die Punktzahl, desto besser stimmt die Rangreihe des Teilnehmers mit der „wahren“ Rangreihe überein.

Die Auswertung des TIHK in dieser Arbeit lehnt sich an den oben beschriebenen Auswertungsmodus von LAYES (1995) an, wird aber modifiziert verwendet. Eine Schwierigkeit, die sich ergab, um das Verfahren von LAYES anzuwenden, liegt darin begründet, dass bei den im TIHK eingesetzten Culture Assimilators England (SCHMID, 1999), USA (MÜLLER & THOMAS, 1995) und Indonesien (MARTIN, 2000) keine Rangreihe hinsichtlich der Güte der Antwortalternativen vorgegeben ist. Um die Auswertung des Tests überhaupt vornehmen zu können, wurde nun versucht, eine Art „Musterlösung“ der verwendeten kritischen Interaktionssituationen zu generieren. Dazu wurden die relevanten kritischen Interaktionssituationen den Autoren der verschiedenen Culture Assimilators sowie Wissenschaftlern und Praktikern auf dem Gebiet der interkulturellen Forschung mit der Bitte vorgelegt, die Antwortalternativen in eine Rangreihe von 1 (trifft am meisten zu) bis 4 (trifft am wenigsten zu) zu bringen. Da die Ergebnisse jedoch ein ähnliches Bild ergaben wie die Antworten der Zielkulturangehörigen in besagten TICS von WELDON ET AL. (1975), nämlich keine eindeutige Übereinstimmung einer Rangreihe der Antwortalternativen, kann der oben beschriebene Auswertungsmodus in dieser Arbeit nur in modifizierter Form verwendet werden.

Die Auswertung der Rohdaten des TIHK ist nun folgendermaßen vorgenommen worden: Ausgangspunkt sind 4 Antwortalternativen, die jeweils auf einer 4-stufigen Skala zu beantworten sind. Ausgewertet wird nur das Antwortverhalten der Versuchspersonen auf dem kulturadäquaten Item, während die anderen Items lediglich als Distraktoren dienen. Dieses Vorgehen ist deswegen gerechtfertigt, da jeder Wert der Skala nur einmal vergeben werden kann und somit die Information der anderen Skalen in der ausgewählten Antwortalternative implizit enthalten ist. Die interkulturelle Handlungskompetenz der Personen wird somit auf einer 4-fach gestuften Intervallskala erfasst. Je nachdem, welche Ausprägung der 4-stufigen Skala der Proband gewählt hat, werden Punktwerte vergeben. Hat der Proband die richtige Antwort mit „am meisten zutreffend“ bewertet, bekommt er für diese Aufgabe den Punktwert 4. Hat er auf der Skala „eher zutreffend“ angekreuzt bekommt er 3 Punkte, für „eher nicht zutreffend“ 2 und für „am wenigsten zutreffend“ 1 Punkt. Für den Gesamtestwert

werden die erreichten Punkte der einzelnen Aufgaben addiert und im Anschluss daran gemittelt. Demzufolge kann das Gesamtergebnis der Probanden im Test Werte zwischen 1 und 4 annehmen, wobei 4 die höchste zu erreichende Punktzahl ist.

5.2.2.2 Prüfung der Verteilung auf Normalität

An dieser Stelle schließt sich die Betrachtung der Rohwertverteilung an. Es wird geprüft, ob es sich bei der vorliegenden Verteilung um eine Normalverteilung handelt. Hinreichende Streuung, Symmetrie und Eingipfligkeit der Rohwertverteilung sind zwar keine notwendigen Bedingungen eines guten Tests, jedoch für die spätere Eichung sehr wünschenswert (LIENERT & RAATZ, 1994). Zur Überprüfung der Normalverteilung kann das Häufigkeitspolygon der Ergebnisse betrachtet werden. Sind dort Abweichungen zur Normalverteilung auffällig, wird anhand des Kolmogorov-Smirnov-Anpassungstests (KSA) die Prüfung auf Normalverteilung statistisch vorgenommen. Zusätzlich können Schiefe und Exzess der Verteilung zur Überprüfung auf Normalität betrachtet werden.

5.2.3 Die Bestimmung der Reliabilität

Die Reliabilität oder die Zuverlässigkeit eines Tests bezeichnet die Genauigkeit, mit der ein Test ein Merkmal misst. Der Test ist dann reliabel, wenn die Testergebnisse die Testperson fehlerfrei beschreiben bzw. auf der Testskala lokalisieren. Die Reliabilität ist ein Maß für die Reproduzierbarkeit eines Testergebnisses und wird auf verschiedene Arten methodisch erfasst: die „Paralleltest-Reliabilität“, die „Retest-Reliabilität“ und die „innere Konsistenz eines Tests“, (LIENERT & RAATZ, 1994, S. 9). Die **Paralleltest-Reliabilität** entspricht dem Korrelationskoeffizienten, der sich ergibt, wenn man einer Stichprobe sowohl den in Frage stehenden Test als auch einen parallelen, d.h. absolut gleichartigen Test vorlegt und die beiden Ergebnisreihen korreliert. Zur Berechnung der **Retest-Reliabilität** legt man einer Stichprobe den in Frage stehenden Test mit einem nicht zu kleinen zeitlichen Abstand noch ein zweites Mal vor und korreliert die beiden Ergebnisreihen. Die **innere Konsistenz** eines Tests kann entweder nach der Methode der Testhalbierung oder der Konsistenzanalyse berechnet werden.

Um eine zweckmäßige Verwendung bzw. Berechnung der Reliabilität zu gewährleisten, gilt es, die unterschiedliche Bedeutung der 3 oben dargestellten Reliabilitätskennwerte zu berücksichtigen. Misst ein Test ein Merkmal, dessen Bedeutung sich mit der Testwiederholung ändert, erweist sich demnach eine hohe Paralleltest-Reliabilität als günstig. Besonderer Wert

auf die Retest-Reliabilität sollte dagegen dann gelegt werden, wenn der Test zeitlich überdauernde Merkmalsausprägungen zu messen beansprucht. Ungeachtet dessen ist, so BORTZ UND DÖRING (1995), eine hohe innere Konsistenz von jedem Test zu fordern. LIENERT UND RAATZ (1994) bezeichnen sie auch als „instrumentale Reliabilität“ (S. 201), die die Leistungsfähigkeit des Tests als Messinstrument kennzeichnet und dem Wesen der Reliabilität am nächsten kommt. Dies ist ein Grund, warum in der vorliegenden Arbeit die Reliabilität des TIHK anhand der inneren Konsistenz eines Tests überprüft wird. Da eine Überprüfung der Reliabilität so früh wie möglich im Verlauf der Testentwicklung stattfinden soll, schlagen LIENERT UND RAATZ (1994) die Berechnung der inneren Konsistenz als eine geeignete Methode zur Vorschätzung der Reliabilität vor. Dies ist der zweite Grund, warum im Anschluss an die Aufgabenanalyse die Reliabilität anhand der inneren Konsistenz bestimmt wird.

Die **innere Konsistenz** gibt Auskunft darüber, wie zuverlässig ein Test ein Konstrukt zu erfassen vermag. Die durchschnittliche Korrelation jedes Items mit jedem ergibt dann ein hohes Maß für den inneren Zusammenhang eines Tests, wenn einzelne Items eines Tests alle hoch miteinander korrelieren. Sie messen dann alle die gleiche Information, d. h. sie messen alle das gleiche Konstrukt. Ein Maß für die innere Konsistenz ist der CRONBACHS α -Koeffizient (1951), der die mittlere Korrelation zwischen den einzelnen Items untereinander angibt bzw. die mittlere Testhalbierungs-Reliabilität eines Tests für alle möglichen Testhalbierungen. Er wurde unter mehreren Methoden aus dem Grund ausgewählt, da er sich nach BORTZ (1984) hierfür am stärksten durchgesetzt hat. Der Alphakoeffizient sollte Werte von $\geq 0,80$ erreichen (BORTZ, 1984; LIENERT & RAATZ, 1994), um von einer reliablen Testform sprechen zu können.

Fällt die Ermittlung der Analysedaten befriedigend aus, sollten noch andere Aspekte der Reliabilität zu einem späteren Zeitpunkt der Testentwicklung untersucht werden. Da das Ziel der vorliegenden Arbeit die Entwicklung eines Paralleltests ist, kann anhand der Korrelation der 2 konstruierten Testformen die Paralleltest-Reliabilität überprüft werden, die als das beste Verfahren zur Beurteilung der Testreliabilität gilt (LIENERT & RAATZ, 1994).

5.2.4 Der Paralleltestaufbau anhand der Analysedaten

Soweit beim Aufbau eines Tests eine Parallelförmigkeit vorgesehen ist, sollte das Aufgabenmaterial im Rahmen des Sukzessivaufbaus nach LIENERT UND RAATZ (1994) gleichzeitig für beide Formen an einer Analysestichprobe dargeboten und analysiert werden. Zur Konstruktion der zu Test A parallelen Testform B müssen allerdings zunächst weitere Aufgaben ausgesucht

werden. Dazu werden erneut kritische Interaktionssituationen aus den Culture Assimilators USA, England und Indonesien ausgewählt, die den gleichen Kulturstandards wie die kritischen Interaktionssituationen von Test A entsprechen. Die kritischen Interaktionssituationen werden nach dem Gesichtspunkt inhaltlicher Gleichartigkeit durch einen Experten auf dem Gebiet der interkulturellen Psychologie bestimmt. Es werden pro Testform mehr kritische Interaktionssituationen mitaufgenommen, um das Zusammenstellen der äquivalenten Testform B nach der Aufgabenanalyse zu erleichtern. Die Aufgaben des Testmaterials setzen sich demnach aus Test A und jeweils 2 weiteren kritischen Interaktionssituationen der dort verwendeten Kulturstandards zusammen. Die dadurch entstehenden zusätzlichen Skalen werden mit B und C bezeichnet.

Ziel ist es nun, aus den Skalen B und C eine nach testtheoretischen Gütekriterien hinreichende Paralleltestform zu entwickeln. Dazu werden alle 3 Testformen wiederum einer Analysestichprobe vorgelegt. Um Reihenfolgeeffekte zu vermeiden, werden die Skalen A, B und C zufällig und in unterschiedlicher Reihenfolge nach einem Überkreuzungsplan (cross-over-design) den Probanden im zweiten Versuchsdurchlauf präsentiert. Dies sieht wie folgt aus: Ein Drittel der Probanden bekommt den Fragebogen in der Reihenfolge ABC (12 Items aus Test A, 12 Items für die Paralleltestskala B, sowie 12 Items für die Paralleltestskala C), ein Drittel der Probanden bekommt den Fragebogen in der Reihenfolge BCA und ein Drittel bekommt ihn in der Reihenfolge CAB.

5.2.5 Die Aufgabenpaarung

Strukturhomogene (inhaltlich einheitliche) Tests werden nach LIENERT UND RAATZ (1994) meist ohne vorherige Paarbildung zur Aufgabenanalyse dargeboten und erst aufgrund der Analysedaten in Aufgabenpaare geteilt. Zum Aufbau des Paralleltests werden zunächst, analog zu einem Test ohne Parallelform, die unzureichenden Aufgaben eliminiert. Der Rest der verbleibenden Aufgaben wird so in 2 gleiche Hälften geteilt, dass jeder Aufgabe ein Paarling von gleicher Schwierigkeit und gleicher Trennschärfe zugeordnet wird. Die Aufgabenpaarung kann nun unter 2 Gesichtspunkten vorgenommen werden: Zum einen können Aufgabenpaarungen bevorzugt werden, die eher gleiche Schwierigkeit gewährleisten, zum anderen können Aufgabenpaarungen aufgrund gleicher Trennschärfe begünstigt werden.

Im Fall des vorliegenden TIHK muss man aufgrund des Testmaterials folgendes berücksichtigen. Da jede Aufgabe im TIHK einen Kulturstandard der jeweiligen Kultur repräsentiert, werden für den Paralleltest nur die kritischen Interaktionssituationen der gleichen Kulturstandards aus Test A ausgewählt. Diese Tatsache schränkt die Aufgabenpaarung des Parallel-

tests insofern ein, als für jedes Item aus Test A nur 2 Items für den Paralleltest zur Auswahl zur Verfügung stehen. Zur Aufgabenpaarung wird demnach in der vorliegenden Arbeit lediglich zwischen 2 Items entschieden, die dem jeweiligen Item aus Test A bezüglich Schwierigkeit und Trennschärfe am ehesten entsprechen.

Da bei der Konstruktion von Paralleltests bestimmte Äquivalenzkriterien (vgl. Kap. 5.2.6) erfüllt sein müssen, findet an dieser Stelle ein weiterer Aspekt bei der Aufgabenpaarung Berücksichtigung. Um tatsächlich von einer Parallelität der beiden Testformen des TIHK sprechen zu können, müssen die Skalen A und B gleiche Mittelwerte, Varianzen und Reliabilität besitzen (vgl. Kap. 5.2.6). Aus diesem Grund werden bei der Parallelskala B verschiedene Itemkombinationen (Aufgaben aus den Skalen B und C werden unterschiedlich zu einer Testform B zusammengesetzt) hinsichtlich dieser Äquivalenzkriterien untersucht. Auf diese Weise soll die Skala generiert werden, die hinsichtlich der Verteilungskennwerte (Mittelwerte, Varianzen) und der Reliabilität von Test A den höchsten Grad an Äquivalenz aufweist. Dazu werden für jede dieser Itemkombinationen Mittelwert, Standardabweichung und innere Konsistenz berechnet und die Skala mit den „besten“ Werten ausgewählt. Unter „beste“ Werte werden die zu Test A gleich hohen Mittelwerte, Varianzen und Reliabilitätswerte verstanden. Das letzte ausschlaggebende Kriterium zur Überprüfung der Äquivalenz der Testformen ist die Paralleltest-Reliabilität, die Auskunft über die Übereinstimmung der beiden Skalen gibt.

5.2.6 Die Prüfung der Äquivalenzkriterien von Paralleltests

Bei Paralleltests bzw. Alternativ- oder Äquivalenztests, wie sie auch genannt werden, handelt es sich nach LIENERT UND RAATZ (1994) um 2 oder mehr gleichwertige Formen ein und desselben Tests, die wechselseitig austauschbar sind. Mit der Parallelförmigkeit kann ein und derselbe Test an ein und demselben Probanden wiederholt werden, was zu Evaluationszwecken für den Studiengang „Internationale Handlungskompetenz“ notwendige Voraussetzung ist. Um davon ausgehen zu können, dass 2 oder mehrere Testformen gleichwertig sind, müssen so genannte Äquivalenzkriterien bestimmt werden. Äquivalenzkriterien sind nach LIENERT UND RAATZ (1994) Kennwerte dafür, dass 2 Paralleltests mit Recht als äquivalent bezeichnet werden dürfen. Es müssen folgende Punkte überprüft werden:

– Äquivalenz der Verteilungskennwerte

Die Überprüfung von Mittelwerten und Varianzen der beiden Testformen ist nach LIENERT UND RAATZ (1994) der Schritt, der bei der Bestimmung der Äquivalenzkriterien als erstes vorzunehmen ist. Dazu werden die beiden Mittelwerte in üblicher Weise

berechnet und ihrem Absolutbetrag nach verglichen. Zeigt sich kein auffälliger Unterschied, dann wird mit dem t-Test geprüft. In gleicher Weise wird mit den Varianzen verfahren, deren F-Quotient keinen signifikanten Wert ergeben darf.

– **Äquivalenz der Reliabilität**

Die Reliabilität eines Paralleltests ist nach LIENERT UND RAATZ (1994) durch die Korrelation der einzelnen Testformen definiert, so dass sich die primäre Forderung bei Paralleltests auf eine genügend hohe Paralleltest-Reliabilität richtet. Die Paralleltest-Reliabilität wird in der Weise bestimmt, dass einer Stichprobe von Probanden mehrere miteinander streng vergleichbare Tests (Paralleltests) vorgelegt und deren Ergebnisse korreliert werden. Da es sich laut LIENERT UND RAATZ (1994) empfiehlt, nach Möglichkeit mehrere Reliabilitätskoeffizienten für die Parallelformen zu ermitteln, wird in der vorliegenden Arbeit auf die Berechnung der inneren Konsistenz, auf die in Kapitel 5.2.3 genau eingegangen wurde, zurückgegriffen.

– **Äquivalenz der Häufigkeitsverteilung**

Um von parallelen Testformen sprechen zu können, müssen die Häufigkeitsverteilungen der Paralleltestformen übereinstimmen. Die Häufigkeitsverteilung der Testformen kann allerdings trotz gleicher Mittelwerte und Varianzen unterschiedlich sein, was nach LIENERT UND RAATZ (1994) meistens auf eine ungleichwertige Anordnung der Aufgaben bezüglich der Schwierigkeit zurückzuführen ist.

– **Äquivalenz der Validität**

Nach LIENERT UND RAATZ (1994) ist die Gleichartigkeit des Testinhalts zweier Parallelformen eine hinreichende Bedingung für gleiche inhaltliche Validität. Die Autoren erläutern weiter, dass gleiche Validität nur empirisch ermittelt und bewiesen werden kann, indem beide Formen an demselben Validitätskriterium validiert werden und ihre Validitätskoeffizienten sich nicht statistisch signifikant unterscheiden.

Sind nicht alle Äquivalenzkriterien hinreichend gegeben, muss der Paralleltest nicht direkt als unbrauchbar verworfen werden. Man kann, um nicht-äquivalente Parallelformen anzugleichen, verschiedene Maßnahmen durchführen (vgl. LIENERT & RAATZ, 1994, Kap. 13).

5.2.7 Die Bestimmung der Validität

Unter der Validität eines Tests versteht man allgemein, wie gut der Test in der Lage ist, genau das zu messen, was er zu messen vorgibt (BORTZ & DÖRING, 1995). Hierbei unterscheidet man 3 Hauptarten der Validität: Inhaltsvalidität, Kriteriumsvalidität und Konstruktvalidität.

Die **Inhaltsvalidität** ist nach BORTZ UND DÖRING (1995) dann gegeben, wenn der Inhalt der Testitems das zu messende Konstrukt in seinen wichtigsten Aspekten erschöpfend erfasst. Dabei kann die Höhe der Inhaltsvalidität eines Tests nicht numerisch bestimmt werden, da sie allein auf subjektiven Einschätzungen beruht. BORTZ UND DÖRING (1995) sprechen an dieser Stelle davon, dass es sich bei der Inhaltsvalidität nicht um ein Testgütekriterium, sondern um eine Zielvorgabe handelt, die bei der Testkonstruktion bedacht werden sollte.

Die **Konstruktvalidität** zielt darauf ab, Beschreibungsmerkmale des zu erfassenden Konstrukts, die nicht unmittelbar operational erfassbar sind, zu analysieren. Zu ihrer Bestimmung formuliert man ein Netz aus Hypothesen über das zu messende Konstrukt, die es anhand der Testwerte zu bestätigen gilt. Je mehr Hypothesen ihre Überprüfung bestehen, umso überzeugender ist die Konstruktvalidierung. Im Rahmen der vorliegenden Arbeit wird aufgrund der Komplexität des Verfahrens und aus ökonomischen Gründen auf die Überprüfung der Konstruktvalidität verzichtet. LIENERT UND RAATZ (1994) sprechen davon, dass dieses Verfahren wegen der fehlenden operationalen Fassbarkeit „ein Vorgang von erheblicher Schwierigkeit und mit großem technisch-ökonomischem Aufwand verbunden“ ist (S. 226). An dieser Stelle wird auf die Kriteriumsvalidität verwiesen, die nach MESSICK (1988) auch als eine Form der Konstruktvalidität betrachtet werden kann. Diese Aussage leuchtet ein, bedenkt man, dass das von außen zu bestimmende Kriterium zur Validierung ebenfalls vom Konstrukt interkulturelle Handlungskompetenz abgeleitet werden muss.

Die **Kriteriumsvalidität** eines Tests ist in der Regel die einzige Art der Validität, die sich durch eine Maßzahl quantitativ eindeutig bestimmen lässt. Sie wird daher von einigen Autoren auch als die „wichtigste Validitätsart“ (BORTZ, 1984, S.138) bezeichnet. Sie liegt dann vor, wenn das Ergebnis eines Tests zur Messung eines latenten Merkmals oder Konstrukts mit Messungen eines korrespondierenden manifesten Merkmals bzw. Kriteriums übereinstimmt (BORTZ & DÖRING, 1995). Dazu korreliert man die Testergebnisse einer Stichprobe mit einem so genannten Außenkriterium. Da dem Verfahren der kriterienbezogenen Validierung in der Literatur eine hohe Bedeutung zugesprochen wird (LIENERT & RAATZ (1994): „Historisch und praktisch gesehen ist die kriterienbezogene Validität der bedeutsamste

Aspekt.“), dessen Überprüfung aber gleichzeitig mit einem hohen ökonomischen Aufwand verbunden ist, wird an dieser Stelle auf die Arbeit von GÖTZ (2003) (vgl. Kap. 6.4) verwiesen, die sich mit der Validierung des TIHK anhand von Rollenspielen auseinandersetzt. In ihrer Studie wurden die Probanden dazu aufgefordert, als naive Versuchspersonen an instruierten Rollenspielen teilzunehmen, die kritische Interaktionssituationen des TIHK beinhalteten. Das gezeigte Verhalten in den Rollenspielen wurde von unabhängigen Beobachtern als interkulturell adäquat bzw. inadäquat eingeschätzt und mit den Ergebnissen des TIHK verglichen. Die Ergebnisse der Validierung des TIHK werden unter Kapitel 6.4 ausführlich dargestellt.

6 Die Praktische Entwicklungsarbeit

Die Entwicklungsarbeit des Tests zur Erfassung interkultureller Handlungskompetenz (TIHK) wird anhand der einzelnen Entwicklungsschritte der zuvor vorgestellten Methode der Paralleltestkonstruktion nachgezeichnet.

6.1 Die Konstruktion der Testform A

6.1.1 Das Testmaterial

6.1.1.1 Die Testaufgaben

Die Testform A des TIHK liegt in Form eines Fragebogens vor und wurde anhand kritischer Interaktionssituationen aus Culture Assimilator Material entwickelt. Um das Konstrukt interkulturelle Handlungskompetenz, das in dieser Arbeit - wie im theoretischen Teil näher erläutert - als kulturallgemeine Fähigkeit verstanden wird, adäquat zu erfassen, sollen stellvertretend für den westlichen, europäischen und östlichen Kulturraum Aufgaben verwendet werden. Von den theoretischen Ausführungen ausgehend, sollen empirisch gewonnene kritische Interaktionssituationen, die von Deutschen im Ausland erlebt wurden, als Testaufgaben im TIHK fungieren. Dazu werden aus dem Culture Assimilator England (SCHMID, 1999) und dem Culture Assimilator Indonesien (MARTIN, 2000) jeweils 7 kritische Interaktionssituationen sowie aus dem Culture Assimilator USA (MÜLLER & THOMAS, 1995) 9 kritische Interaktionssituationen aus unterschiedlichen Kulturstandards gewählt. Die Anzahl der kritischen Interaktionssituationen wurde nach praktischen Kriterien bestimmt. Unter Berücksichtigung der Tatsache, dass nach der Aufgabenanalyse ausreichend kritische Interaktionssituationen in die Testendform mitaufgenommen werden können und um die Probanden bei der Lösung des Tests nicht zu überfordern, wurde eine Anzahl von insgesamt 23 kritischen Interaktionssituationen für die Testvorform festgelegt. Die 2 zusätzlichen kritischen Interaktionssituationen des Culture Assimilators USA wurden zu anderweitigen Forschungszwecken miterhoben. TRIANDIS ET AL. (1998) schlagen für einen Test in Form von kritischen Interaktionssituationen eine Anzahl von mindestens 16 KIs vor, so dass die verwendete Anzahl der ausgewählten KIs angemessen ist. Die kritischen Interaktionssituationen werden nach folgenden Kriterien ausgewählt:

- Die kritischen Interaktionssituationen sollen aus deutscher Perspektive nicht zu schwer und nicht zu leicht zu beantworten sein (vgl. Kap. 5.2.1.2 zur Trennschärfe).
- Die Situationen sollten für die Zielgruppe, in diesem Fall Studenten, nachvollziehbar sein, d.h., es sollen nicht zu viele geschäftliche Details enthalten sein.
- Die kritischen Interaktionssituationen sollen ein möglichst breites Spektrum fremdkulturellen Handelns und Verhaltens (unterschiedliche Bereiche des Alltags) abdecken.
- Es wird aus dem jeweiligen Kulturstandard die kritische Interaktionssituation ausgewählt, die nach Einschätzung eines Experten den jeweiligen Kulturstandard am besten beschreibt.

	Kultur	KI	Kulturstandard
1.	England	„Royal Opera“	Selbstdisziplin
2.	USA	Situation 4	Patriotismus
3.	Indonesien	„Diskussionen“	Harmoniestreben
4.	USA	Situation 7	Gleichheitsdenken
5.	Indonesien	„Gruppenzwang“	Gruppenorientierung
6.	England	„Wohin mit den Gästen?“	Indirektheit interpersoneller Kommunikation
7.	USA	Situation 10	Gelassenheit
8.	England	„Der Kneipenjob“	Ritualisierung
9.	Indonesien	„Die Hausangestellte“	Hierarchie
10.	Indonesien	„Die Geduldsprobe“	Ganzheitlichkeit
11.	USA	Situation 11	Handlungsorientierung
12.	England	„Der Streit“	Pragmatismus
13.	Indonesien	„Zusagen“	Paternalismus
14.	USA	Situation 16	Leistungsorientierung
15.	England	„Die Bürodekoration“	Ritualisierte Regelverletzung
16.	USA	Situation 22	Individualismus
17.	USA	Situation 26	Streben nach sozialer Anerkennung
18.	England	„How are you?“	Indirektheit interpersoneller Kommunikation
19.	Indonesien	„Der Sprachkurs“	Gesicht
20.	USA	Situation 30	Interpersonale Distanzreduzierung
21.	Indonesien	„Das Betriebsfest“	Geschehen lassen
22.	USA	Situation 37	Zwischengeschlechtliche Beziehungsmuster
23.	England	„Heil Hitler“	Deutschlandstereotyp

Tabelle 1: Ausgewählte kritische Interaktionssituationen Test A (Vorform)

Die Reihenfolge der kritischen Interaktionssituationen wird zufällig bestimmt, da nach LIENERT UND RAATZ (1994) bei zeitlich nicht begrenzten Tests keine bestimmte Reihenfolge eingehalten werden braucht.

6.1.1.2 Demographische Daten

Zusätzlich zu den Testaufgaben wurden die demographischen Daten Alter, Geschlecht und Studienrichtung der Probanden erfasst, um sicher zu stellen, dass diese auch zur intendierten Zielgruppe gehören.

6.1.2 Durchführung der Datenerhebung

Der Test liegt in Form eines Paper-and-Pencil-Fragebogens vor und umfasst die demographischen Fragen, eine Testinstruktion sowie die 23 Testaufgaben (kritische Interaktionssituationen). Die Fragebögen wurden den Probanden zum Ausfüllen mit nach Hause gegeben. Die Teilnahme basierte auf freiwilliger Basis und ohne zeitliche Vorgabe.

6.1.3 Analyse der Stichprobe Test A

Für die Datenerhebung zur Konstruktion der Testform A konnten 60 Studenten der Universität und Fachhochschule Regensburg gewonnen werden. Die Studenten gehören den Fachrichtungen Psychologie, Pädagogik, Betriebswirtschaftslehre, Wirtschaftsinformatik, Lehramt, Maschinenbau, Physiotherapie, Jura, Elektrotechnik, Volkswirtschaftslehre, Biologie und Biochemie an. Die Studenten der Psychologie stellen mit $N = 32$ (53 %) den größten der Teil der Stichprobe dar. Teilgenommen haben 17 Männer (28 %) und 43 (72 %) Frauen im Alter zwischen 19 und 33 Jahren ($M = 23,18$). In die Berechnungen zur Testkonstruktion konnten nur 57 Testbögen berücksichtigt werden, da 3 Probanden den Test nur unvollständig ausgefüllt hatten und somit eine korrekte Auswertung nicht möglich war. Die zukünftige Anwendung des TIHK im Zusatzstudium „Internationale Handlungskompetenz“ an der Universität bzw. Fachhochschule Regensburg setzt eine ähnliche Verteilung der teilnehmenden Studierenden voraus. Da das Zusatzstudium vom Lehrstuhl für ABO- und Sozialpsychologie initiiert wurde, ist davon auszugehen, dass ein Großteil der Studenten Geisteswissenschaften, im besonderen Psychologie studieren. Besteht der Großteil der Zusatzstudienteilnehmer aus Geisteswissenschaftlern, kann man weiterhin davon ausgehen, dass ein Großteil der Probanden weiblich ist (Verhältnis weiblich/männlich ungefähr 80 zu 20 Prozent). Das Durchschnittsalter von 23 Jahren entspricht gewöhnlich dem Alter von Studenten im Hauptstudium, das Voraussetzung für die Teilnahme am Zusatzstudium darstellt. Die Verteilung von Alter, Geschlecht und Studienrichtung entspricht somit der Forderung, dass die Analysestichprobe im höchstmöglichen Maße für das später zu testende Kollektiv repräsentativ ist (LIENERT & RAATZ, 1994).

6.1.4 Auswertung der Daten Test A

6.1.4.1 Ergebnisse Test A

Anhand der Rohwertermittlung wie sie in Kapitel 5.2.2.1 beschrieben ist, ergeben sich für die Testvorform A die folgenden Ergebnisse: Die Spannweite der Testergebnisse der analysierten Stichprobe liegt zwischen 2,43 und 3,83 Testpunkten mit einem Mittelwert von $M = 3,28$ und einer Standardabweichung von $SD = 0,29$. Die Interpretation der Testergebnisse hat zunächst keine Bedeutung für die Konstruktion des TIHK, könnte aber ein Hinweis darauf sein, dass viele Probanden im TIHK gut abschneiden. Die Interpretation der Ergebnisse wird in Kapitel 6.5 dieser Arbeit beschrieben.

6.1.4.2 Aufgabenanalyse Test A

Im Rahmen der Aufgabenanalyse werden der Schwierigkeitsindex und der Trennschärfe-koeffizient der 23 Aufgaben der Testvorform A bestimmt. Die Ergebnisse der entsprechenden Analysen werden im Folgenden vorgestellt.

Schwierigkeitsanalyse

Wie in Kapitel 5.2.1.1 beschrieben, wird mit Hilfe der Formel $P = 100 \times N_R/N$ der Schwierigkeitsindex der Testaufgaben anhand Analysestichprobe ($N = 57$) ermittelt. Die Ergebnisse sind tabellarisch dargestellt.

	KI_1	KI_2	KI_3	KI_4	KI_5	KI_6	KI_7	KI_8	KI_9	KI_10	KI_11	KI_12
P	47,4	72,9	77,2	35,1	38,6	37,9	45,6	75,4	68,4	30,0	30,5	15,5
	KI_13	KI_14	KI_15	KI_16	KI_17	KI_18	KI_19	KI_20	KI_21	KI_22	KI_23	
P	63,8	10,3	41,4	34,5	60,3	82,8	72,4	94,8	32,8	88,3	43,1	

P = Schwierigkeit

Tabelle 2: Ergebnis der Schwierigkeitsanalyse

Man erkennt, dass die Aufgaben über einen weiten Schwierigkeitsbereich von $P = 10,3$ bis $P = 94,8$ streuen. Es existieren 4 schwere Aufgaben mit $P < 30$, 6 leichte Aufgaben mit $P > 70$ sowie 13 mittelschwere Aufgaben von $30 > P < 70$. Aus der Analyse wird ersichtlich, dass die Items KI_12, KI_14, KI_18, KI_20 und KI_22 außerhalb des als Auswahlkriterium benannten Schwierigkeitsbereiches von $95 > P > 5$ liegen und deswegen mit hoher Wahrscheinlichkeit nicht trennscharf sind. Diese Items müssen aufgrund des Ergebnisses der

Schwierigkeitsanalyse eliminiert werden. Um nicht zu viele Items aufgrund der Schwierigkeitsanalyse zu „verlieren“, wird jedoch auf den von LIENERT UND RAATZ (1994) weniger konservativen Geltungsbereich zwischen $80 > P > 20$ zurückgegriffen, woraufhin lediglich die Items 14 und 22 eliminiert werden.

Trennschärfeanalyse

Bei der Bestimmung der Trennschärfe-Koeffizienten kann von der Richtig-Falsch-Klassifikation der Aufgaben zur Berechnung der Schwierigkeitsindizes wieder abgerückt und mit den exakten Aufgabenfehlerzahlen gerechnet werden. Wieder werden die 23 Aufgaben der Analyse der oben beschriebenen Stichprobe ($N = 57$) unterzogen. Die Resultate zeigt unten stehende Tabelle.

	KI_1	KI_2	KI_3	KI_4	KI_5	KI_6	KI_7	KI_8	KI_9	KI_10	KI_11	KI_12
r	0,44	0,01	0,29	0,03	0,33	0,26	0,47	0,51	0,09	0,15	0,38	0,45
	KI_13	KI_14	KI_15	KI_16	KI_17	KI_18	KI_19	KI_20	KI_21	KI_22	KI_23	
r	0,46	0,10	0,53	0,05	0,56	0,60	0,51	0,38	0,29	0,10	0,35	

r = Trennschärfe

Tabelle 3: Ergebnis der Trennschärfeanalyse

Die Trennschärfeanalyse ergibt Werte von $r = 0,01$ bis $r = 0,60$, woraus ersichtlich wird, dass einige Aufgaben der Testvorform A den Kriterien einer mittleren Trennschärfe mit Werten von $0,30$ bis $0,50$ von BORTZ UND DÖRING (1995) nicht gerecht werden. Aus diesem Grund müssen die Items KI_2, KI_3, KI_4, KI_6, KI_9, KI_10, KI_14, KI_16, KI_21, KI_22 aus Test A herausgenommen werden. Die den Items 20 (USA) und 18 (England) zugrunde liegenden Kulturstandards behandeln ähnliche Verhaltensweisen des sozialen Miteinanders in den USA bzw. in England, die zu den gleichen Schwierigkeiten auf deutscher Seite führen. Da die Testaufgaben jedoch ein möglichst breites Spektrum fremdkulturellen Verhaltens und Handelns abdecken sollen (vgl. Kap. 6.1.1.1), wird an dieser Stelle das Item 20 ebenfalls eliminiert. Es besitzt gegenüber Item 18 eine geringere Trennschärfe und wird aus diesem Grund aus der Testform herausgenommen. Test A besteht demnach an diesem Punkt lediglich aus 12 Aufgaben.

6.1.4.3 Überprüfung der Verteilung auf Normalität

An dieser Stelle soll die Überprüfung der Rohwertverteilung auf Normalität vorgenommen werden, die für die Testkonstruktion von praktischer Bedeutung ist. Nach LIENERT UND RAATZ (1994) ist die Normalverteilung im Allgemeinen ein günstiges Symptom sowohl für die Wahl der Analysestichprobe wie auch für die Konstruktion des Tests. Außerdem sind normalverteilte Rohwerte einer interindividuellen Differenzierung förderlich und stellen die beste Voraussetzung für eine Standardnormierung dar.

Um nun sicher zu gehen, dass es sich im vorliegenden Fall um eine Normalverteilung handelt, wird der Kolmogorov-Smirnov-Anpassungstest (KSA) durchgeführt. Der KSA prüft die Nullhypothese, dass die untersuchten Daten normalverteilt sind. Testet man diese Nullhypothese im vorliegenden Fall über die Rohwerte aller Probanden, so wird die Nullhypothese nicht verworfen ($Z = ,811$; $p = ,527$). Das bedeutet, dass die Rohwerte von Testvorform A als normalverteilt angesehen werden können. Betrachtet man jedoch die graphische Darstellung des Häufigkeitspolygons, wird ersichtlich, dass die Verteilung nicht der Normalverteilung entsprechen kann, sondern es sich um eine rechtsschiefe Verteilung handelt. Die Diskrepanz zwischen Häufigkeitspolygon auf der einen Seite und Kolmogorov-Smirnov-Anpassungstest auf der anderen Seite ist darauf zurückzuführen, dass der KSA lediglich die ermittelten Werte der Testergebnisse (von 2,43 bis 3,83) und nicht alle möglichen Testerwerte (von 1 bis 4) in die Berechnung miteinbezieht. Dadurch entsteht eine Verzerrung, so dass an dieser Stelle zur Überprüfung der Verteilung auf Normalität auf statistischer Ebene die Schiefe und der Exzess der Rohwertverteilung berechnet werden. Die Schiefe der Rohwertverteilung wird mittels des z-Tests für den Schiefekennwert S berechnet und weist mit $S = -2,47$ auf eine rechtsgipflige Verteilung hin. Der Exzess der Rohwertverteilung wird mittels des Exzessivitätskennwertes E berechnet und weist mit $E = 1,17$ auf eine hyperexzessive Girlandenverteilung (schmalgipflige Verteilung) hin. Hier wird demnach der Eindruck der graphischen Darstellung des Häufigkeitspolygons bestätigt. Die anomale Verteilung der Rohwerte kann an dieser Stelle als Hinweis gewertet werden, dass die Verfahren der klassischen Testtheorie zur Testkonstruktion nicht angewandt werden können und auf andere Methoden zurückgegriffen werden muss (vgl. Kap. 6.2.3.2). Das Häufigkeitspolygon der Rohwertverteilung von Test A zeigt untenstehende Abbildung.

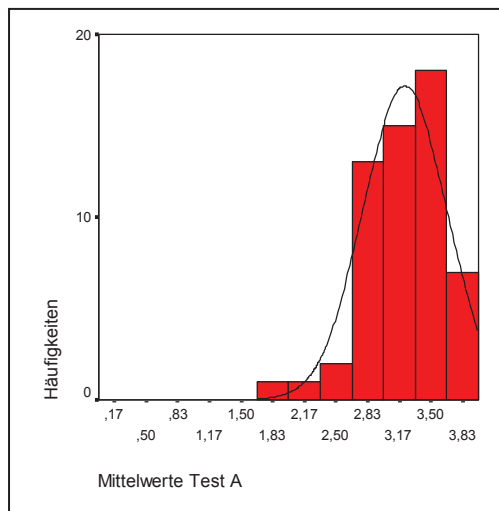


Abb. 2: Rohwertverteilung Test A

6.1.4.4 Bestimmung der Reliabilität (Test A)

Zur Kontrolle der Reliabilität des TIHK wird an dieser Stelle die Reliabilität anhand der Methode der inneren Konsistenz ermittelt (vgl. Kap. 5.2.3). Als Maß für die innere Konsistenz von TIHK ergibt sich nach der Elimination der ungeeigneten Aufgaben ein CRONBACHS α -Koeffizient von $r = 0,72$. Dieser Wert reicht erwartungsgemäß nicht an die Reliabilitätskoeffizienten von standardisierten Tests heran, für die eine Reliabilität von $r \geq 0,80$ gefordert wird (BORTZ, 1984; LIENERT & RAATZ, 1994). Hier macht sich wohl vor allem die kleine Zahl von nur 12 Aufgaben negativ bemerkbar, da die Reliabilität auch von der Testlänge abhängt: Je länger ein Test ist, desto höher ist in der Regel auch seine Reliabilität (LIENERT & RAATZ, 1994). Aufgrund dieses Zusammenhangs kann mit Hilfe der „Spearman-Brown-Formel“ (LIENERT & RAATZ, 1994, S.185) geschätzt werden, wie hoch die Reliabilität würde, verlängerte man den Test um das n -fache. LORD UND NOVICK (1968) haben mit Hilfe dieser Spearman-Brown-Formel die Reliabilität von verlängerten Tests für verschiedene Ausgangswerte extrapoliert. Dort kann man sehen, dass ein Test mit einer Anfangsreliabilität von $r = 0,72$ nach einer Verdopplung der Testlänge eine Reliabilität von $r = 0,80$ erreichen kann. Für die vorliegende Untersuchung bedeutet das: Würde der TIHK nicht aus 12, sondern aus etwa 24 Aufgaben dieser Art bestehen (was durchaus denkbar wäre), so wäre wohl auch mit diesem Test eine Reliabilität von $r \geq 0,80$ erreichbar.

6.1.5 Die Endform Test A

Nach Berechnung von Aufgabenanalyse und Reliabilität sowie der Erörterung der Rohwertverteilung ergibt sich für die Testendform A des TIHK folgendes Datenmaterial:

	Kultur	KI	Schwierigkeitsindex (P)	Trennschärfe (r)
1.	England	„Royal Opera“ (1)	47,4	0,44
2.	Indonesien	„Gruppenzwang“ (5)	38,6	0,33
3.	USA	Situation 10 (7)	45,6	0,47
4.	England	„Der Kneipenjob“ (8)	75,4	0,51
5.	USA	Situation 11 (11)	30,5	0,38
6.	England	„Der Streit“ (12)	15,5	0,45
7.	Indonesien	„Zusagen“ (13)	63,8	0,46
8.	England	„Die Bürodekoration“ (15)	41,4	0,53
9.	USA	Situation 26 (17)	60,3	0,56
10.	England	„How are you?“ (18)	82,8	0,60
11.	Indonesien	„Der Sprachkurs“ (19)	72,4	0,50
12.	England	„Heil Hitler“ (23)	43,1	0,35

Tabelle 4: Items der Testform A

Die Testform A des TIHK besteht demnach aus 12 kritischen Interaktionssituationen der Länder England (stellvertretend für den europäischen Kulturraum), USA (stellvertretend für den westlichen Kulturraum) und Indonesien (stellvertretend für den östlichen Kulturraum). Dabei sind 6 Aufgaben des Culture Assimilators England (SCHMID, 1999) sowie jeweils 3 Aufgaben des Culture Assimilators USA (MÜLLER & THOMAS, 1995) und des Culture Assimilators Indonesien (MARTIN, 2000) in die Version A aufgenommen worden. Der Forderung, dass bei Testkonstruktionen die Schwierigkeitsindizes über den gesamten Bereich streuen, also einige leichte und schwere Aufgaben sowie viele mittelschwere Aufgaben verwendet werden sollen, wird das Datenmaterial hiermit gerecht. Die Ergebnisse der Trennschärfenberechnungen sind mit Werten zwischen $r = 0,33$ und $r = 0,60$ ebenfalls zufrieden stellend, wenngleich mehr Werte über 0,60 für die Güte der Trennschärfe wünschenswert wären. Zur Reliabilität des Tests ist anzumerken, dass die innere Konsistenz mit einem CRONBACHS α -Koeffizienten von $r = 0,72$ ausreicht, wenn man die geringe Anzahl an Testitems berücksichtigt.

Die Rohwertverteilung der Testendform A ergibt die folgenden Ergebnisse: Der Range der dargestellten Stichprobe liegt zwischen 1,83 und 3,92 Testpunkten bei einem Mittelwert von $M = 3,22$ und einer Standardabweichung von $SD = 0,44$. Ausgehend von der oben dargestellten Testendform A kann nun die zu A parallele Testform B konstruiert werden, deren Entwicklung im nächsten Abschnitt beschrieben wird.

6.2 Die Konstruktion der Testform B

6.2.1 Das Testmaterial

Die Konstruktion der Testform B des TIHK erfolgt im Wesentlichen nach dem gleichen Prinzip wie die Konstruktion von Test A. Da Paralleltests für beide Formen an einer Analysestichprobe dargeboten und analysiert werden müssen (LIENERT & RAATZ, 1994), werden die 12 Items aus Testform A zusammen mit 2 mal 12 neuen Items an einer neuen Stichprobe geprüft. Wie in Kapitel 5.2.4 beschrieben, besteht demnach die Testvorform B zur Vorlage der Analysestichprobe aus 36 Aufgaben. Die Testvorform B liegt wiederum in Form eines Paper-and-Pencil-Fragebogens vor und umfasst neben den demographischen Fragen die Testinstruktion sowie die 36 Testaufgaben. Die Fragebögen wurden den Probanden auch dieses Mal zum Ausfüllen ohne zeitliche Vorgabe mit nach Hause gegeben. Das für die Konstruktion von Test B verwendete Datenmaterial zeigen folgende Tabellen:

Skala A

	Kultur	KI	Kulturstandard
A1	England	„Royal Opera“	Selbstdisziplin
A2	Indonesien	„Gruppenzwang“	Gruppenorientierung
A3	USA	„Situation 10“	Gelassenheit
A4	England	„Der Kneipenjob“	Ritualisierung
A5	USA	„Situation 11“	Handlungsorientierung
A6	England	„Der Streit“	Pragmatismus
A7	Indonesien	„Zusagen“	Paternalismus
A8	England	„Die Bürodekoration“	Ritualisierte Regelverletzung
A9	USA	„Situation 26“	Streben nach sozialer Anerkennung
A10	England	„How are you?“	Indirektheit interspers. Kommunikation
A11	Indonesien	„Der Sprachkurs“	Gesicht
A12	England	„Heil Hitler“	Deutschlandstereotyp

Skala B

	Kultur	KI	Kulturstandard
B1	England	„Feueralarm“	Selbstdisziplin
B2	Indonesien	„Vitamin B“	Gruppenorientierung
B3	England	„Cheers!“	Ritualisierung
B4	USA	„Situation 8“	Gelassenheit
B5	England	„Die Standpauke“	Pragmatismus
B6	Indonesien	„Verantwortung“	Paternalismus
B7	England	„Ein Geburtstagsgeschenk“	Ritualisierte Regelverletzung
B8	USA	„Situation 12“	Handlungsorientierung
B9	England	„Happy Birthday“	Indirektheit interspers. Kommunikation
B10	Indonesien	„Der Handwerker“	Gesicht
B11	England	„Die Brandanschläge“	Deutschlandstereotyp
B12	USA	„Situation 28“	Streben nach sozialer Anerkennung

Skala C

	Kultur	KI	Kulturstandard
C1	England	„Die Diskussion“	Selbstdisziplin
C2	Indonesien	„Vitamin B“	Gruppenorientierung
C3	England	„Die Verlobte“	Ritualisierung
C4	USA	„Situation 9“	Gelassenheit
C5	England	„Die Putzfrau“	Pragmatismus
C6	Indonesien	„Entscheidungen“	Paternalismus
C7	England	„Let's talk about Sex“	Ritualisierte Regelverletzung
C8	USA	„Situation 13“	Handlungsorientierung
C9	England	„Eine dt.-engl. Freundschaft“	Indirektheit interspers. Kommunikation
C10	Indonesien	„Prüfungen“	Gesicht
C11	England	„Die Kündigung“	Deutschlandstereotyp
C12	USA	„Situation 27“	Streben nach sozialer Anerkennung

Tabellen 5-7: Testaufgaben (Skalen A, B, C) zur Konstruktion von Testform B

Bei der Eingabe der Daten in das Statistikprogramm SPSS 11.0 wurden die Aufgaben der Testformen folgendermaßen umkodiert:

	Kultur	Kulturstandards	Test A	Test B	Test C
1.	England	Selbstdisziplin	A1	B1	C1
2.	Indonesien	Gruppenorientierung	A2	B2	C2
3.	England	Ritualisierung	A4	B3	C3
4.	USA	Gelassenheit	A3	B4	C4
5.	England	Pragmatismus	A6	B5	C5
6.	Indonesien	Paternalismus	A7	B6	C6
7.	England	Ritualisierte Regelverletzung	A8	B7	C7
8.	USA	Handlungsorientierung	A5	B8	C8
9.	England	Indirektheit interspers. Kommunikation	A10	B9	C9
10.	Indonesien	Gesicht	A11	B10	C10
11.	England	Deutschlandstereotyp	A12	B11	C11
12.	USA	Streben nach sozialer Anerkennung	A9	B12	C12

Tabelle 8: Umkodierung der Aufgaben und Zuordnung zu Kulturstandards

6.2.2 Analyse der Stichprobe Test B

Für die Datenerhebung zur Konstruktion der Testform B konnten 72 Studenten der Universität und der Fachhochschule Regensburg gewonnen werden. Die Studenten gehören den Fachrichtungen Psychologie, Pädagogik, Sozialpädagogik, Betriebswirtschaftslehre, Verwaltungswissenschaft, Lehramt, Wirtschaftsingenieurwesen, Informatik, Jura, Architektur, Kommunikationswissenschaften, Zahnmedizin, Biologie, Maschinenbau, Geografie und Theologie an. Die Studenten der Psychologie stellen mit $N = 45$ (62,5 %) den größten Teil der Stichprobe dar. Teilgenommen haben 25 Männer (35 %) und 47 Frauen (65 %) im Alter zwischen 18 und 29 Jahren ($M = 22,81$). Hier zeichnet sich eine ähnliche Verteilung wie

die der Stichprobe von Test A ab, so dass auch hier der Forderung der Repräsentativität der Analysestichprobe bezüglich des später zu testenden Kollektivs nachgekommen wird.

6.2.3 Auswertung der Daten Test B

6.2.3.1 Ergebnisse Test B

Die Ermittlung der Rohwerte des Tests wird mit der gleichen Methode wie zuvor zur Konstruktion von Testform A vorgenommen. Die Gesamtpunktzahl (Rohwertverteilung) in Test B über alle 36 Items der Analysestichprobe ergibt einen Range zwischen 2,40 und 3,72 Testpunkten, einen Mittelwert von $M = 3,26$ und eine Standardabweichung von $SD = 0,21$. Die Ergebnisse der Testergebnisse einschließlich der Subskalen A, B und C sind in Tabelle 9 dargestellt:

	N	Min	Max	M	SD
Gesamttest	72	2,40	3,72	3,27	0,21
Skala A	72	2,50	3,83	3,26	0,33
Skala B	72	1,92	4,00	3,29	0,32
Skala C	72	2,42	3,67	3,22	0,26

Tabelle 9: Deskriptive Daten von Test A inklusive der Subskalen A, B, C

An dieser Stelle folgt die Berechnung von Aufgabenschwierigkeit, Trennschärfe und Reliabilität. Dazu wurde im Rahmen der Konstruktion von Test A auf das Verfahren der Korrelation zurückgegriffen. Bei Betrachtung des Häufigkeitspolygons der Testergebnisse ergibt sich allerdings folgendes Bild: In beiden Testdurchläufen sind die erreichten Rohwerte der Probanden rechtsschief verteilt, d.h., viele Testaufgaben wurden von vielen Probanden gut gelöst. Der zuvor verwendete Produkt-Moment-Korrelationskoeffizient lässt sich jedoch nur unter bestimmten Voraussetzungen anwenden. Die wichtigsten Voraussetzungen sind die bivariate Normalverteilung der Variablen und deren Erfassung auf Intervallskalenniveau. Sind die Daten nicht intervallskaliert, kann auf spezielle Korrelationstechniken zurückgegriffen werden, wie die punktbiserale Korrelation, die Rangkorrelation oder den Phi-Koeffizienten (vgl. BORTZ, 1984; FAHRMEIER ET AL, 1997). Diese Techniken setzen jedoch alle voraus, dass die Merkmale normalverteilt sind. Zwar ist der Korrelationskoeffizient gegenüber Verletzungen der Normalverteilungsannahme relativ robust (vgl. BORTZ & DÖRING, 1995), mit der rechtsschiefen Verteilung geht aber zugleich auch eine geringe Varianz der Daten einher, so dass dieser nicht mehr anwendbar ist (vgl. GUILFORD, 1954; POPHAM &

HUSEK, 1969; HERBIG, 1972). Für diesen Fall wurden verschiedene Verfahren entwickelt, die als so genannte kriteriumsorientierte Leistungsmessung bekannt wurden.

6.2.3.2 Kriteriumsorientierte Messung

Unter einer kriteriumsorientierten Messung versteht man nach FRICKE (1974) ein wissenschaftliches Routineverfahren zur Untersuchung der Frage, ob und wie gut ein bestimmtes Lehrziel erreicht ist. Die dabei verwendeten Testaufgaben repräsentieren das Lehrziel und dienen dazu, den individuellen Fähigkeitsgrad eines Schülers mit einem gewünschten Fähigkeitsgrad zu vergleichen.

Kriteriumsorientierte Tests interpretieren die Testergebnisse auf der Grundlage eines Kriteriums und unterscheiden sich damit von normorientierten Tests, die Testergebnisse anhand des Mittelwerts einer Eichstichprobe interpretieren. Bei normorientierten Tests bedient sich die Interpretation von Testergebnissen der klassischen Testtheorie, die nach FRICKE (1974) in erster Linie formal-statistisch orientiert ist und wenig Hilfen bei der inhaltlichen Interpretation von Testwerten liefert. Die Methoden der klassischen Testtheorie zielen darauf ab, möglichst gut zwischen Personen zu diskriminieren und versagen, wenn alle Personen eine Idealnorm erreichen. Auf statistischer Ebene bedeutet das, wie POPHAM UND HUSEK (1969) nachgewiesen haben, dass viele Formeln der klassischen Testtheorie deswegen versagen, da in fast allen von ihnen der Korrelationskoeffizient enthalten ist, der im Fall, dass alle Personen das Ziel erreicht haben, nicht definiert ist. Grund dafür ist, dass bei den klassischen Tests die Varianz in den Testergebnissen relevant ist: Alle Gütekriterien beziehen sich auf Varianz- bzw. Kovarianzverhältnisse bzw. korrelative Beziehungen zwischen „wahren“ Testwerten und beobachteten Testwerten. Haben alle Personen das Ziel erreicht, ist die wahre Varianz Null, d.h., es gibt keine Unterschiede in den „wahren“ Testwerten der Personen. Damit werden die klassischen Methoden der Trennschärfe-, der Zuverlässigkeits- und Validitätsbestimmung nicht mehr anwendbar, da sich in einem solchen Fall durchwegs Null-Korrelationen ergeben (HERBIG, 1972).

Daraus ergibt sich für den TIHK die folgende Schlussfolgerung: Bei einer derart schiefen Verteilung wie sie im Fall von TIHK vorliegt und der damit einhergehenden geringen Varianz der Testwerte, können die meisten Formeln der klassischen Testtheorie nicht mehr angewandt werden. Für diesen Fall können neue mathematisch-statistische Verfahren der kriteriumsorientierten Messung verwendet werden. Die geringe Varianz der 4-stufigen Antwortskala des TIHK ist häufig bei Fragebögen anzutreffen, die auf Rating-Skalen mit nur wenigen Abstufungen beruhen (wie z.B. die am häufigsten verwandte Likert-Skala mit nur 5 Abstu-

fungen von „stimme zu“ bis „stimme gar nicht zu“) (SIMON ET AL., in Vorbereitung). Ein Grund, warum diese Verfahren zur Konstruktion von Test A keine Anwendung fanden, ist zum einen in der Unbekanntheit der Verfahren kriteriumsorientierter Messung in der „Main-Stream-Psychologie“ zu sehen bzw. darauf zurückzuführen, dass diese Verfahren größtenteils noch recht unausgereift sind (SIMON ET AL., in Vorbereitung). Darüber hinaus sind die Ergebnisse der Testkonstruktion A trotz rechtsschiefer Verteilung so gut ausgefallen, dass kein Anlass gegeben war, die Verfahren der klassischen Testtheorie in Frage zu stellen. Die oben geschilderte Problematik hat sich erst bei der Auswertung zur Konstruktion von Test B herausgestellt, was als Beweis dafür gelten kann, wie schnell es in der klassischen Testtheorie bei Verwendung der Korrelation zu Artefakten kommen kann.

Aufgrund der geringen Varianz und der rechtsschiefer Verteilung im TIHK, kann die Trennschärfe der Items sowie die Reliabilität und Validität des Fragebogens mit dem Korrelationskoeffizienten nicht adäquat bestimmt werden. In diesem Fall stellt der c-Koeffizient nach SIMON ET AL. (in Vorbereitung) ein optimales Verfahren dar. Der c-Koeffizient wird im zweiten Testdurchlauf zur Konstruktion der Paralleltestform B eingesetzt und im nächsten Abschnitt ausführlich dargestellt.

6.2.3.3 *Der c-Koeffizient*

Der c-Koeffizient von SIMON ET AL. (in Vorbereitung) kann zur Objektivitäts-, Reliabilitäts- und Validitätsbestimmung von Tests oder Fragebögen angewandt werden, wenn die Daten nicht normalverteilt sind und/oder nur eine geringe Varianz aufweisen.

Der c-Koeffizient geht in Anlehnung an den Ü-Koeffizient von FRICKE (1972) davon aus, dass das Verhältnis der Nichtübereinstimmung zwischen den Beurteilern zur maximal möglichen Nichtübereinstimmung Auskunft über die vorhandene Nichtübereinstimmung zwischen den Beurteilern gibt. Der Zahlenwert 1 als absolut mögliche Übereinstimmung minus der vorhandenen Nichtübereinstimmung liefert dann ein Maß für die Übereinstimmung zwischen den Beurteilern. Der c-Koeffizient ist nach SIMON ET AL. (in Vorbereitung) gegenüber dem Ü-Koeffizienten deswegen geeigneter, da beim c-Koeffizienten die Nichtübereinstimmung nicht anhand der Varianz zwischen den Beurteilern bestimmt wird, sondern anhand der Differenzen zwischen den Werten. Der c-Koeffizient ist dementsprechend wie folgt definiert:

$$c = 1 - \frac{d^2}{d^2_{\max}}$$

Zur Bestimmung der Differenzen zwischen den Beurteilungen wird auf die Quadrierte Euklidische Distanz zurückgegriffen, die wie folgt definiert ist:

$$d_{ii'}^2 = \sum_{j=1}^n (x_{ij} - x_{i'j})^2$$

Die maximal mögliche Differenz zwischen den Beurteilungen - und damit die maximal mögliche Nichtübereinstimmung d_{\max}^2 zwischen den Beurteilern - wird nach unten folgender Formel bestimmt. Hierbei ist jedoch zu unterscheiden, ob die Anzahl der Beurteiler k gerade oder ungerade ist.

Formel zur Berechnung von d_{\max}^2 :

$$d_{\max}^2 = N \cdot (b - a)^2 \cdot \frac{k^2}{4} \quad \text{für } k \text{ gerade}$$

$$d_{\max}^2 = N \cdot (b - a)^2 \cdot \frac{k^2 - 1}{4} \quad \text{für } k \text{ ungerade}$$

k = Anzahl der Beurteiler

N = Anzahl der Versuchspersonen

a = obere Skalengrenze

b = untere Skalengrenze

Zur Bestimmung der Signifikanz von c muss das 95 %- und 99 %-Quantil der zufälligen Verteilung von c unter der Nullhypothese ermittelt werden. Die Verwendung des c -Koeffizienten zur Konstruktion des TIHK hat Auswirkungen auf die Berechnung von Trennschärfe, innerer Konsistenz und Paralleltest-Reliabilität, die in den folgenden Kapiteln dargestellt werden.

6.2.3.4 Aufgabenanalyse Test B

Schwierigkeitsanalyse

Für die Konstruktion der Paralleltestformen erfolgt nun erneut die Bestimmung von Schwierigkeitsindex und Trennschärfe im Rahmen der Aufgabenanalyse. Die Bestimmung der Schwierigkeit wird in gleicher Weise vorgenommen wie in Kapitel 5.2.1.1 der vorliegenden Arbeit, da die Formel des Schwierigkeitsindex keinen Einsatz des c -Koeffizienten erforderlich macht. Die Ergebnisse sind in der folgenden Tabelle dargestellt:

Kultur-standards	1	2	3	4	5	6	7	8	9	10	11	12
P (A)	60,56	48,61	64,79	48,61	19,44	61,11	44,44	35,21	80,56	79,17	37,5	56,94
P (B)	38,57	40,85	54,93	63,38	58,57	76,06	54,93	28,17	49,30	53,52	77,14	80,28
P (C)	31,43	25,35	56,34	69,01	23,94	74,65	64,79	53,52	63,38	80,00	36,62	28,57

P = Schwierigkeit

Tabelle 10: Schwierigkeitsindices geordnet nach Kulturstandards

Es wird ersichtlich, dass alle 3 Skalen über einen breiten Schwierigkeitsbereich streuen und sich hinsichtlich ihrer Gesamtschwierigkeit ähneln. Skala C ist mit 2 schweren Aufgaben ($P < 30$), 8 mittelschweren Aufgaben ($30 > P < 70$) und 2 leichten Aufgaben ($P > 70$) gegenüber Skala A (1 schwere, 9 mittelschwere und 2 leichte Aufgaben) und Skala B (1 schwere, 8 mittelschwere und 3 leichte Aufgaben) die „schwierigste“ Testform. Bezüglich der im Methodenteil dargestellten Schwierigkeitskriterien verfehlen lediglich die Aufgaben A9, C10 und B12 nur sehr knapp den erforderlichen Bereich von $20 > P < 80$, so dass nach Ermessen der Autorin aufgrund der Schwierigkeitsanalyse keine Aufgaben eliminiert werden müssen.

Trennschärfeanalyse

Wie unter Kapitel 6.2.3.3 geschildert, bietet sich der c-Koeffizient von SIMON ET AL. (in Vorbereitung) als geeignetes Maß zur Trennschärfeanalyse an. Die Bestimmung der Trennschärfe bei der Konstruktion von Skala B des TIHK anhand des c-Koeffizienten gibt Auskunft darüber, inwieweit ein Item zur Unterscheidung zwischen interkulturell handlungskompetenten und nicht kompetenten Personen beitragen kann. Dies ist dann gegeben, wenn eine Person, die mittels des Tests als interkulturell handlungskompetent identifiziert wurde, auch auf einem einzelnen Item zur Erfassung interkultureller Handlungskompetenz einen hohen Wert erreicht und umgekehrt. In diesem Zusammenhang geht es also um die Frage der Übereinstimmung des Itemwerts mit dem Gesamtwert des Tests.

SIMON ET AL. (in Vorbereitung) weist daraufhin, dass bei der Verwendung des c-Koeffizienten zur Trennschärfenermittlung - im Gegensatz zur Korrelation - auf die Signifikanz des c-Wertes zu achten ist. Da sich nur sehr hohe Übereinstimmungswerte als signifikant erweisen, kann der c-Koeffizient keineswegs wie der Korrelationskoeffizient interpretiert werden, bei dem bereits Korrelationen ab 0,30 auf eine ausreichende Trennschärfe hinweisen (BORTZ & DÖRING, 1995). Beim c-Koeffizient ist deswegen weniger als beim Korrelationskoeffizient auf die Größe der Übereinstimmung zu achten, als vielmehr auf die Signifikanz der Übereinstimmung.

stimmung, da sich bereits per Zufall hohe Übereinstimmungen ergeben können und somit nur sehr hohe Übereinstimmungen auf eine substantielle Übereinstimmung hinweisen. Bei Berechnung der Trennschärfe anhand des c-Koeffizienten mit $N = 72$ und $k = 2$ ergeben sich die folgenden Werte:

Kulturstandards	1	2	3	4	5	6	7	8	9	10	11	12
c (A)	0,91	0,93	0,92	0,90	0,89	0,93	0,92	0,91	0,94	0,94	0,81	0,93
c (B)	0,82	0,93	0,92	0,91	0,88	0,93	0,92	0,89	0,94	0,89	0,93	0,91
c (C)	0,86	0,87	0,96	0,93	0,80	0,92	0,91	0,96	0,92	0,92	0,88	0,88

c = Trennschärfe

Tabelle 11: Trennschärfe geordnet nach Kulturstandards

Laut Prüfstatistik sind alle in der Tabelle dargestellten Aufgaben mit Werten $\geq 0,88$ auf dem 1 %-Niveau hochsignifikant, besitzen also eine hohe Trennschärfe. Für den TIHK als ungeeignet gelten demnach die Items B1, C1, C2, C5 und A11. Sie müssen aus der Testform herausgenommen werden.

6.2.3.5 Bestimmung der Reliabilität (Gesamt)

Als Maß für die **innere Konsistenz** wird analog zur Entwicklung von Test A der CRONBACHS α -Koeffizient berechnet. Die Berechnung von CRONBACHS α -Koeffizient basiert allerdings ebenfalls, wie die Berechnung der Trennschärfe, auf der Methode der Korrelation, die sich bei nicht normalverteilten Daten als nicht geeignet erwiesen hat. Aus diesem Grund wird wie bei der Bestimmung der Trennschärfe auf den c-Koeffizienten von SIMON ET AL. (in Vorbereitung) zurückgegriffen. Mit seiner Hilfe wird zur Bestimmung der inneren Konsistenz von Testform B des TIHK die Interkorrelation der Items getrennt für die Skalen A und B berechnet.

Die **Paralleltest-Reliabilität** wird normalerweise über die Paralleltest-Korrelation bestimmt. Dies ist abermals im vorliegenden Test nicht zulässig, da auch hier der Fall der nicht vorhandenen Varianz mit in Betracht gezogen werden muss. Bei Bestimmung der Paralleltest-Reliabilität geht es um die Frage der Reproduzierbarkeit der Daten. Reproduzierbarkeit meint die Übereinstimmung der Testwerte bei wiederholter Messung, so dass der c-Koeffizient als Maß der Übereinstimmung ein adäquates Verfahren zur Bestimmung der Reliabilität eines Tests darstellt. Im TIHK wird demnach die Übereinstimmung der Testergebnisse von Skala A jedes Probanden mit den Testergebnissen der Skala B jedes Probanden berechnet.

6.2.4 Die Aufgabenpaarung

Zur Aufgabenpaarung wird in der vorliegenden Arbeit zwischen 2 Items der Skalen B und C entschieden (vgl. Kap. 5.2.5), die hinsichtlich des Kulturstandards dem jeweiligen Item aus Test A bezüglich Schwierigkeit und Trennschärfe am ehesten entsprechen.

Aufgrund der geringen Trennschärfe werden zunächst die Aufgaben der amerikanischen Kulturstandards 4, 5 und 11 ($c = 0,90$, $c = 0,89$, $c = 0,81$) aus Test A eliminiert. Es verbleiben 9 Items, für die es im nächsten Schritt Aufgabenpaarlinge zu finden gilt. Bei der Aufgabenpaarung wird darauf geachtet, dass die Schwierigkeitsindices und die Trennschärfenwerte der resultierenden Skalen in etwa gleich sind. Nach LIENERT UND RAATZ (1994) ist die Begünstigung der Trennschärfe bei heterogenen Aufgaben sinnvoller. Die Tatsache, dass es sich beim vorliegenden TIHK jedoch um strukturhomogene Aufgaben handelt (es gibt nur kritische Interaktionssituationen in diesem Test), würde für eine Bevorzugung der Schwierigkeit der Aufgaben sprechen. Hohe Trennschärfenwerte dagegen haben generell einen positiven Einfluss auf die innere Konsistenz von Testskalen (LIENERT & RAATZ, 1994), was wiederum auf eine Begünstigung der Trennschärfe hindeuten würde. Die Entscheidung bleibt demnach bei der Autorin, die sich an dieser Stelle auf eine Bevorzugung der Trennschärfe zugunsten einer hohen inneren Konsistenz festlegt. Das Ergebnis der Aufgabenpaarung unter Berücksichtigung von Trennschärfe und Schwierigkeitsindex sieht wie folgt aus:

Test A	P	c	Test B	P	c
A1	60,56	0,91	B1	38,57	0,82
A2	48,61	0,94	B2	40,85	0,93
A4	64,79	0,92	C3	56,34	0,96
A5	35,21	0,91	B8	28,17	0,89
A7	61,11	0,94	B6	76,06	0,93
A8	44,44	0,92	B7	54,93	0,92
A9	56,94	0,93	B12	80,28	0,91
A10	80,56	0,96	B9	49,30	0,94
A11	79,17	0,95	C10	80,00	0,92

P = Schwierigkeit, c = Trennschärfe

Tabelle 12: Auswahl der Items für den Paralleltest aufgrund der Aufgabenpaarung

Da neben Trennschärfe und Schwierigkeit, den im Methodenteil dieser Arbeit vorgestellten Äquivalenzkriterien eine entscheidende Rolle bei der Konstruktion von Paralleltests zukommt, werden zur Aufgabenpaarung die Verteilungskennwerte (Mittelwert, Standardabweichung) und die innere Konsistenz ebenfalls berücksichtigt. Zur Auswahl der Aufgaben aus

Skala B und C sind demnach fünf Maße von der Autorin gleichermaßen berücksichtigt worden. Die Ergebnisse der verschiedenen Aufgabenpaarungen zeigen die nun folgenden Tabellen.

		M	SD	c
Test A	a1, a2, a4, a5, a7, a8, a9, a10, a11	3,39	0,34	0,73
Test B	1. b1, b2, b3, b6, b7, c8, b9, c10, b12	3,36	0,33	0,69
	2. c1, b2, c3, b6, b7, c8, b9, c10, b12	3,39	0,31	0,73
	3. b1, b2, c3, b6, b7, c8, b9, c10, b12	3,34	0,32	0,71
	4. c1, b2, b3, b6, b7, b8, b9, c10, b12	3,29	0,35	0,68
	5. b1, b2, b3, b6, b7, b8, b9, c10, b12	3,32	0,34	0,66
	6. c1, b2, c3, b6, b7, c8, b9, c10, b12	3,41	0,30	0,73

Tabelle 13: Aufgabenpaarungen ohne die Items A3, A6, A12 (M = Mittelwert, SD = Standardabweichung, c = innere Konsistenz)

Hier wird ersichtlich, dass Mittelwerte und Standardabweichungen der ersten, zweiten und sechsten Itemkombination von B der Testform A am nächsten kommen. Die innere Konsistenz der zweiten und sechsten Itemkombination ist gleich der inneren Konsistenz von Testform A. Mit Werten von $c = 0,73$ entspricht dieses Ergebnis laut Prüfstatistik nicht einer hinreichenden inneren Konsistenz. Um die innere Konsistenz der Skala zu erhöhen, wird zunächst aufgrund der geringsten Trennschärfe das Item B1 ($c = 0,82$) aus Test B sowie das entsprechende Item A1 aus Test A eliminiert. Daraus ergeben sich weitere Aufgabenkombinationen mit neuen Ergebnissen hinsichtlich der Äquivalenzkriterien.

		M	SD	c
Test A	a2, a4, a5, a7, a8, a9, a10, a11	3,39	0,36	0,75
Test B	1. b2, b3, b6, b7, b8, b9, c10, b12	3,37	0,37	0,74
	2. b2, c3, b6, b7, c8, b9, c10, b12	3,47	0,33	0,78
	3. b2, c3, b6, b7, b8, b9, c10, b12	3,40	0,38	0,74
	4. b2, b3, b6, b7, c8, b9, c10, b12	3,45	0,35	0,76

Tabelle 14: Aufgabenpaarungen ohne die Items A1, A3, A6, A12 (M = Mittelwert, SD = Standardabweichung, c = innere Konsistenz)

Die Tabelle zeigt, dass sich die innere Konsistenz der Testform A erwartungsgemäß erhöht hat ($c = 0,75$) und somit auf dem 1 %-Niveau hochsignifikant wird. Die erste und dritte Itemkombination kommt hinsichtlich Mittelwerten und Standardabweichungen der Testform A am nächsten. Die Werte der inneren Konsistenz kommen mit $c = 0,74$ annähernd an die innere Konsistenz von Test A heran. Die zweite und vierte Itemkombination wird von der Autorin

aufgrund zu A unterschiedlicher Verteilungskennwerte (Mittelwert, Standardabweichung) als weniger geeignet erachtet, obwohl beide Skalen eine höhere innere Konsistenz aufweisen.

Um die innere Konsistenz noch weiter zu erhöhen, werden das Item A5 und die dazugehörigen Items B8 bzw. C8 aufgrund zu geringer Trennschärfe aus der Skala herausgenommen. Es ergeben sich folgende Werte:

		M	SD	c
Test A	a2, a4 a7, a8, a9, a10, a11	3,45	0,39	0,77
Test B	1. b2, c3, b6, b7, b9, c10, b12	3,47	0,35	0,77
	2. b2, b3, b6, b7, b9, c10, b12	3,44	0,37	0,74

Tabelle 15: Aufgabenpaarungen ohne die Items A1, A3, A5, A6, A12 (M = Mittelwert, SD = Standardabweichung, c = innere Konsistenz)

Wie die Tabelle zeigt, erhöht sich die innere Konsistenz erneut auf $c = 0,77$. Dieser Wert wird von der ersten Itemkombination von Test B ebenfalls erreicht, wogegen sich Mittelwert ($M = 3,47$) und Standardabweichung ($SD = 0,35$) von Test A unterscheiden. Aufgrund des Mittelwertes ($M = 3,44$) und der Standardabweichung ($SD = 0,37$) scheint die zweite der Itemkombinationen von B geeigneter.

Ergebnis Aufgabenpaarung:

Berücksichtigt man alle Äquivalenzkriterien gleichermaßen, schneiden in der Reihe der 8er Skalen Nummer 1 ($M = 3,37$, $SD = 0,37$ und $c = 0,74$) und Nummer 3 ($M = 3,40$, $SD = 0,38$ und $c = 0,74$) am besten ab. Sie sind hinsichtlich Mittelwerten, Varianzen und Innerer Konsistenz nahezu äquivalent zu Test A ($M = 3,39$, $SD = 0,36$ und $c = 0,75$). Die 7er Skala von A weist wie zu erwarten eine höhere innere Konsistenz auf ($M = 3,45$, $SD = 0,35$ und $c = 0,77$). Die Itemkombinationen 1 ($M = 3,47$, $SD = 0,35$ und $c = 0,77$) und 2 ($M = 3,44$, $SD = 0,37$ und $c = 0,74$) von Test B schneiden aber hinsichtlich gleicher Mittelwerte, Varianzen und Reliabilität schlechter ab, als dies bei den 8er Skalen der Fall ist. Eine 8er Skala wird von der Autorin auch vor dem Hintergrund der psychologischen Validität bevorzugt: Die ohnehin geringe Anzahl an Aufgaben wird so nicht weiter verringert und die Aufgaben verteilen sich gleichmäßiger über die verschiedenen Kulturräume England, USA und Indonesien. Somit wird hier sowohl dem theoretischen Hintergrund zur Erfassung interkultureller Handlungskompetenz als auch den praktischen Überlegungen zur Konstruktion des TIHK in dieser Arbeit Rechnung getragen. Diese Gesichtspunkte sind zudem hinsichtlich der inhaltlichen Gleichartigkeit für die psychologische Validität von Relevanz. Dadurch wird die Akzeptanz des Tests gesteigert, was wiederum einen ernsthaften Umgang mit dem Testmaterial begünstigt.

Da bei homogenen Niveautests eine maximale Trennschärfe nach LIENERT UND RAATZ (1994) das primäre Kriterium ist, nach dem die Auswahl der Aufgaben erfolgen soll, entscheidet sich die Autorin an dieser Stelle für die Skala 3. Die Skala 3 hat das Item C3 mitaufgenommen, das mit einer Trennschärfe von $c = 0,96$ eine höhere Trennschärfe aufweist als das Item B3 der Skala 1, das eine Trennschärfe von $c = 0,92$ besitzt. Demnach bestehen die Skalen A und B aus jeweils 8 Items, wobei jeweils 3 kritische Interaktionssituationen der Culture Assimilators England und Indonesien sowie 2 kritische Interaktionssituationen des Culture Assimilator USA angehören. Für Skala A ergibt sich ein Mittelwert von $M = 3,39$ und eine Standardabweichung von $SD = 0,38$, für Skala B $M = 3,40$ und $SD = 0,36$. Die Ergebnisse der Aufgabenpaarung hinsichtlich der zu berücksichtigenden 5 Kriterien Schwierigkeitsindex, Trennschärfenindex, Mittelwert, Varianz und innere Konsistenz sind in der folgenden Übersicht tabellarisch dargestellt:

		M	SD	c
Test A	a2, a4, a5, a7, a8, a9, a10, a11	3,39	0,36	0,75
Test B	b2, c3, b6, b7, b8, b9, c10, b12	3,40	0,38	0,74

Tabelle 16: Mittelwert (M), Standardabweichung (SD) und innere Konsistenz (c) der Parallelskalen A und B

Test A	A2	A4	A5	A7	A8	A9	A10	A11
P	48,61	64,79	35,21	61,11	44,44	56,94	80,56	79,17
c	0,95	0,93	0,91	0,94	0,93	0,93	0,96	0,95

Tabelle 17: Schwierigkeitsindex (P) und Trennschärfe (c) der Skala A

Test B	B2	C3	B8	B6	B7	B12	B9	C10
P	40,85	56,34	28,17	76,06	54,93	80,28	49,30	80,00
c	0,94	0,96	0,89	0,94	0,93	0,93	0,95	0,94

Tabelle 18: Schwierigkeitsindex (P) und Trennschärfe (c) der Skala B

Die Tabelle zeigt, dass beide Skalen bezüglich ihrer Schwierigkeitsindices ähnliche Werte aufweisen. Insgesamt betrachtet, handelt es sich bei Werten von $28,17 > P < 80,56$ eher um einen „leichten“ Test (5 Werte $P > 70$), was sich auch in der Verteilung der Rohwerte widerspiegelt.

Hinsichtlich der Trennschärfe fällt das Item B8 auf, das mit 0,89 die geringste Trennschärfe besitzt. Dieses Item zu eliminieren, wurde aus den oben genannten Gründen der psychologischen Validität abgelehnt. Die Verteilungskennwerte (Mittelwert und Varianz) und die innere Konsistenz der Skalen A und B sind augenscheinlich sehr ähnlich. Ob dies auch in einem

statistischen Sinn zutrifft, muss anhand der Überprüfung der Äquivalenzkriterien im nächsten Abschnitt vorgenommen werden.

6.2.5 Die Prüfung der Äquivalenzkriterien

Die Äquivalenz der Verteilungskennwerte

Vergleicht man nun **Mittelwerte** und **Varianzen** beider Testformen, erkennt man, dass die Verteilungskennwerte beider Skalen augenscheinlich etwa gleich sind. Um sicher zu gehen, dass sich die Testformen tatsächlich nicht voneinander unterscheiden, werden Mittelwerte und Varianzen statistisch geprüft. LIENERT UND RAATZ (1994) geben zur Überprüfung der Mittelwerte den t-Test an, der in dieser Arbeit aufgrund nicht normal verteilter Rohwerte jedoch nicht verwendet werden darf. Stattdessen wird der Wilcoxon-Test für nicht-parametrische Daten und abhängige Stichproben berechnet. Er testet die Nullhypothese, ob die Mittelwerte homogen sind. Mit einem z-Wert von -0,253 bei $P = 0,801$ wird die Nullhypothese nicht verworfen, so dass von der Äquivalenz der Mittelwerte ausgegangen werden kann.

Zur Überprüfung der Äquivalenz der Varianzen wird der Levene-Test verwendet. Der Levene-Test testet die Nullhypothese, ob die Varianzen homogen sind. Im vorliegenden Fall (Levene-Statistik: 0,618 und $P = 0,433$) kann diese Nullhypothese auch auf dem 20 %-Niveau nicht verworfen werden, so dass von Varianzhomogenität ausgegangen werden kann.

Die Äquivalenz der Reliabilität

Bezüglich der Äquivalenz der Reliabilität des Paralleltests sind 2 Maße berechnet worden: die innere Konsistenz jeder Skala sowie die Paralleltest-Reliabilität. Die **innere Konsistenz** kennzeichnet die Leistungsfähigkeit als Messinstrument für jede Paralleltestform gesondert. Sie wurde ebenfalls mit dem c-Koeffizient von SIMON ET AL. (in Vorbereitung) berechnet und beträgt bei Test A 0,75 und bei Test B 0,74. Beide Werte werden laut Prüfstatistik auf dem 1 %-Niveau statistisch signifikant. Damit sind die inneren Konsistenzen der Parallelformen augenscheinlich fast identisch, so dass man hier von Äquivalenz der Reliabilität ausgehen kann.

Die letzte notwendige Bedingung, die bei der Konstruktion von Paralleltests erfüllt sein muss, ist die Überprüfung der **Paralleltest-Reliabilität**. Dazu wird wiederum die Übereinstimmung der Testskalen A und B anhand des c-Koeffizienten von SIMON ET AL. (in Vorbereitung) be-

rechnet. Mit einem c von 0,98 ist die Übereinstimmung auf dem 1 %-Niveau statistisch hochsignifikant, so dass man hier von einer hohen Paralleltest-Reliabilität sprechen kann.

Die Äquivalenz der Häufigkeitsverteilung

Zur Überprüfung dieses Äquivalenzkriteriums werden die Häufigkeitspolygone der Rohwertverteilungen beider Testformen analysiert (vgl. Abbildungen 3 und 4). Die Überprüfung auf Normalverteilung spielt an dieser Stelle eine untergeordnete Rolle.

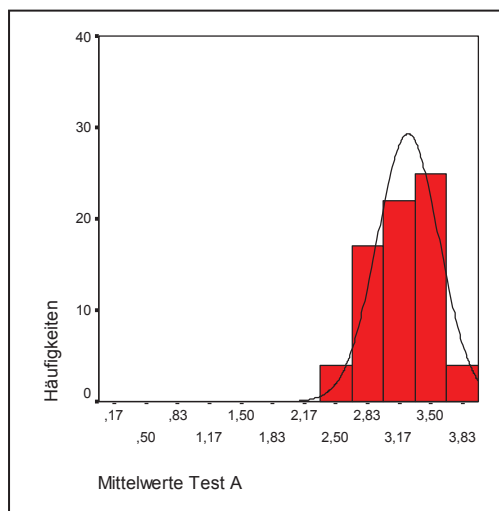


Abb. 3: Rohwertverteilung der Parallelskala A

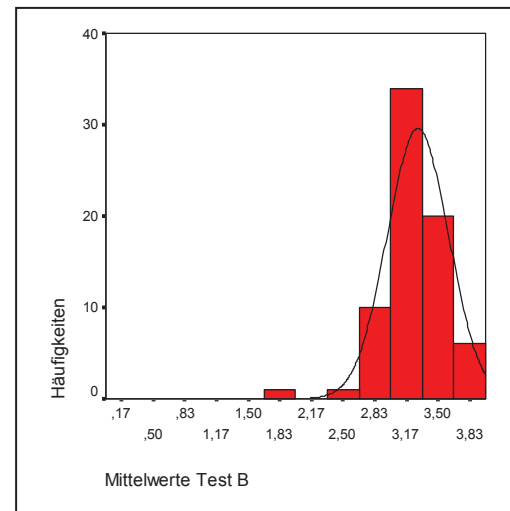


Abb. 4: Rohwertverteilung der Parallelskala B

Betrachtet man die Häufigkeitspolygone der beiden Testformen A und B, erkennt man Unterschiede in den Rohwertverteilungen. Bei beiden Testformen handelt es sich zwar um rechtschiefe Verteilungen (Test A: $S = -3,01$; Test B: $S = -3,36$), Test A ($E = 0,87$) unterscheidet sich jedoch hinsichtlich des Exzesses deutlich von Test B ($E = 3,81$). Das Häufigkeitspolygon von A ist breitgipflig im Gegensatz zu dem von Test B, der bezüglich des Exzesses annähernd normalverteilt ist. Dieser Umstand lässt sich auf die Anordnung der Aufgaben von unterschiedlicher Schwierigkeit zurückführen (LIENERT & RAATZ, 1994). Welche Rolle der Vergleichsvergleich in der Praxis bei nichtnormalen Verteilungen spielt, wird von LIENERT UND RAATZ (1994) zwar angesprochen, jedoch nicht weiter behandelt. Die unterschiedlichen Rohwertverteilungen können aber als Hinweis dafür gewertet werden, dass die Parallelität der Testformen nicht perfekt ist.

Die Äquivalenz der Validität

Wie in Kapitel 5.2.7 erläutert, wird aufgrund des enormen ökonomischen Aufwandes von Validierungsstudien die Validität des TIHK in der Arbeit von GÖTZ (2003) überprüft. Diese Arbeit wird im nächsten Abschnitt genauer dargestellt, so dass an dieser Stelle lediglich auf die **inhaltliche Validität** des Paralleltests eingegangen wird.

Im Fall des TIHK soll die im Theorieteil hervorgehobene Fähigkeit des Perspektivenwechsels als Kernkompetenz für interkulturell kompetentes Handeln durch kulturadäquates Attribuieren messbar gemacht werden. Die Fähigkeit des kulturadäquaten Attribuierens wird in den Aufgaben von Test A unmittelbar durch die Aufgaben in Form von kritischen Interaktionssituationen getestet (vgl. Kap. 5.1). Durch das vorgegebene Antwortschema, die Antwortalternativen hinsichtlich ihrer Güte in eine Rangreihe zu bringen, ist es erforderlich, dass die Probanden die Perspektive des fremdkulturellen Interaktionspartners einnehmen, um zu beurteilen, welche Antwort die Situation am treffendsten erklärt. Aus diesem Grund kann man hier von gegebener inhaltlicher Validität sprechen.

Hinsichtlich Testform B ist festzuhalten, dass diese ebenfalls aus kritischen Interaktionssituationen besteht, die aus dem gleichen Culture Assimilator Material wie Testform A zusammengestellt wurde. Demzufolge können beide Testformen als inhaltlich gleichartig betrachtet werden, was wiederum nach LIENERT UND RAATZ (1994) eine hinreichende Bedingung für gleiche inhaltliche Validität ist.

Was die Übereinstimmungs- und Vorhersagevalidität betrifft, wird an späterer Stelle ausführlich auf die Arbeit von GÖTZ (2003) hingewiesen, die den Validitätskoeffizienten für die gesamte Testform an einem Außenkriterium bestimmt. Die Berücksichtigung lediglich der Gesamttestform (und nicht jeder Testform getrennt) ist gerechtfertigt, da man bei der Überprüfung der Gesamtvalidität nach LIENERT UND RAATZ (1994) davon absieht, dass 2 Parallelformen verwendet worden sind.

6.3 Darstellung des Gesamtergebnisses des TIHK

Der TIHK besteht aus 2 parallelen Testformen zu je 8 Aufgaben (kritische Interaktionssituationen) der Culture Assimilators England (3 KIs), USA (2 KIs) und Indonesien (3 KIs). Folgende KIs wurden in den TIHK aufgenommen:

	Kultur	Skala A			Skala B		
		(M = 3,39; SD = 0,36; innere Konsistenz: c = 0,75)	(P)	(c)	(M = 3,40; SD = 0,38; innere Konsistenz: c = 0,74, Paralleltest-Reliabilität: c = 0,98)	(P)	(c)
1.	Indonesien	„Gruppenzwang“ (Gruppenorientierung)	48,61	0,95	„Vitamin B“ (Gruppenorientierung)	40,85	0,94
2.	England	„Der Kneipenjob“ (Ritualisierung)	64,79	0,93	„Die Verlobte“ (Ritualisierung)	56,34	0,94
3.	USA	„Situation 11“ (Handlungsorientierung)	35,21	0,91	„Situation 12“ (Handlungsorientierung)	28,17	0,93
4.	Indonesien	„Zusagen“ (Paternalismus)	61,11	0,94	„Verantwortung“ (Paternalismus)	76,06	0,89
5.	England	„Die Bürodekoration“ (Ritualisierte Regelverletzung)	44,44	0,93	„Ein Geburtstagsgeschenk“ (Ritualisierte Regelverletzung)	54,93	0,95
6.	USA	„Situation 26“ (Soziale Anerkennung)	56,94	0,93	„Situation 28“ (Soziale Anerkennung)	80,28	0,93
7.	Indonesien	„Der Sprachkurs“ (Gesicht)	80,56	0,96	„Prüfungen“ (Gesicht)	49,30	0,96
8.	England	„How are you?“ (Indirektheit interspers. Komm.)	79,17	0,95	„Happy Birthday“ (Indirektheit interspers. Komm.)	80,00	0,94

P= Schwierigkeit, c = Trennschärfe

Tabelle 19: Der TIHK: Gesamtergebnis der Paralleltestkonstruktion

Aus der Darstellung wird ersichtlich, dass die Aufgaben des TIHK eine mittlere bis leichte Schwierigkeit besitzen (Test A: P = 35,21 bis P = 80,56) und durchwegs eine hohe Trennschärfe aufweisen (Test A: c = 0,91 bis c = 0,96; Test B: c = 0,89 bis c = 0,96).

Weiterhin ist festzuhalten, dass es sich beim TIHK um einen Paralleltest handelt, der hinreichend reliabel ist und den Äquivalenzkriterien von Paralleltests gerecht wird. Die Verteilungskennwerte der Testform A sind mit M = 3,39 und SD = 0,36 äquivalent zu Testform B mit M = 3,40 und SD = 0,38. Die innere Konsistenz der Testform A ist mit c = 0,75 äquivalent zu Testform B mit c = 0,74. Beide werden auf dem 1 %-Niveau statistisch hochsignifikant. Die Paralleltest-Reliabilität beträgt c = 0,98 und wird ebenfalls auf dem 1 %-Niveau statistisch hochsignifikant.

Hinsichtlich der Validierung des TIHK sowie der Interpretation der Testergebnisse wird auf die folgenden Kapitel hingewiesen.

6.4 Validierung des TIHK

Der TIHK wird mit dem Anspruch konzipiert, das Konstrukt des Perspektivenwechsels als Indikator für interkulturelle Handlungskompetenz zu messen. Dabei wird nicht nur überprüft, wie gut sich der Proband in einer fremden Kultur auskennt, sondern auch, wie sicher er im Umgang mit verschiedenen kulturellen Verhaltensweisen und Situationen ist. Um aufgrund von Testpunktwerten Rückschlüsse auf das Verhalten ziehen zu können, muss der Test Kriteriumsvalidität besitzen. Als Kriterium diente in der Validitätsstudie das Verhalten der Probanden in einem standardisierten Rollenspiel, innerhalb dessen sie mit einer kritischen Interaktionssituation konfrontiert wurden. Inwieweit sich die Probanden in dieser Situation kulturadäquat verhielten, wurde auf einer 4-stufigen Skala gemessen, deren Abstufungen inhaltlich mit der 4-stufigen Skala des TIHK übereinstimmten. Die Fragestellung, ob eine signifikante Übereinstimmung zwischen Test- und Rollenspieldaten zu verzeichnen ist, wurde von GÖTZ (2003) untersucht.

Als Grundlage für die Konzeption der Rollenspiele wurden die kritischen Interaktionssituationen 9 („Skifahrt“), 11 („Abendessen“) und 26 („Kirche“) des Culture Assimilators USA entnommen. Bei den Testpersonen handelte es sich um 30 Studenten der Universität Regensburg, die von 2 geschulten, unabhängigen Beobachtern hinsichtlich ihrer interkulturellen Handlungskompetenz im Rollenspiel eingeschätzt wurden. Die Testpersonen wurden im Allgemeinen eher positiv bewertet – 68 Ratings von 90 Ratings (3 Rollenspiele pro 30 Versuchspersonen) entfallen auf die positiven Skalenwerte 3 und 4, 22 auf die negativen Skalenwerte 1 und 2. Damit ergibt sich ähnlich wie im TIHK eine rechtsschiefe Verteilung der Rohwerte des Rollenspiels.

Um die Testpersonen hinsichtlich ihrer interkulturellen Handlungskompetenz in den Rollenspielen einschätzen zu können, wurden die Beobachterwerte der 3 Rollenspiele für jede Person gemittelt. Somit ergab sich für jede Person ein Wert als Indikator für den Grad an gezeigter interkultureller Handlungskompetenz in den Rollenspielen. Aufgrund dieses Wertes, der zwischen 1 und 4 liegt, konnten die Testpersonen unter Berücksichtigung des positiven (3 und 4) und negativen Bereichs (1 und 2) des Kodierungsmanuals als interkulturell handlungs(in)kompetent eingestuft werden.

Die in den Rollenspielen und im TIHK gewonnenen Daten werden mit Hilfe des c-Koeffizienten (SIMON ET AL., in Vorbereitung) auf ihre Übereinstimmung geprüft. Es wurden verschiedene Übereinstimmungen berechnet: Zum einen wurden die Rollenspielmittelwerte mit dem Gesamttestergebnis aus Test A und B, zum anderen mit den Ergebnissen der beiden Parallel-

tests A und B getrennt auf Übereinstimmung untersucht. Bei vorliegender Datenlage – 30 Probanden und jeweils 2 auf Übereinstimmung zu testende Werte, nämlich die Rollenspiel- und Testwerte - sind die Punktwerte der Rollenspiele und der Testergebnisse hochsignifikant homogen ab Werten von 0,91. In Tabelle 20 sind die Ergebnisse der c-Koeffizienten-Berechnungen einzusehen.

	Übereinstimmung
Mittelwert über die 3 Rollenspiele mit Gesamtergebnis Test A und B	0,95
Mittelwert über die 3 Rollenspiele mit Ergebnis Test A	0,94
Mittelwert über die 3 Rollenspiele mit Ergebnis Test B	0,95

Tabelle 20: Übereinstimmung der Test- und Rollenspielergebnisse im Rahmen der Validierung

Testet man die Übereinstimmung der Rollenspiele mit dem Gesamtergebnis der beiden Paralleltests A und B zusammen auf Übereinstimmung, ergibt sich ein hochsignifikanter c-Koeffizient ($c = 0,95$). Auch die Berechnung der gemittelten Rollenspielwerte mit den Einzelergebnissen der Paralleltests A und B ergibt sowohl mit dem Paralleltest A ($c = 0,94$) als auch mit dem Paralleltest B ($c = 0,95$) hochsignifikante c-Koeffizienten.

Aufgrund dieser hochsignifikanten Übereinstimmungen zwischen den Ergebnissen der Paralleltests A und B und den Rollenspieldaten sowie dem Gesamtergebnis der Tests A und B und den Rollenspieldaten konnte gezeigt werden, dass Probanden mit guten bzw. schlechten Leistungen im Test auch gute bzw. schlechte Leistungen in den Rollenspielen erbrachten. Somit erfüllen die Paralleltests des TIHK den Anspruch der Kriteriumsvalidität. Da Test A ($c = 0,94$) und Test B ($c = 0,95$) augenscheinlich fast identische Werte in der Übereinstimmung des Rollenspielergebnisses mit dem TIHK aufweisen, wird hier die Parallelität der beiden Skalen als bestätigt angesehen. Weiterhin wird dadurch ebenfalls das Äquivalenzkriterium der Validität erfüllt.

6.5 Interpretation der Testergebnisse

Der Paralleltest TIHK besteht pro Skala aus 8 Aufgaben. Zur Interpretation der Testergebnisse werden die erreichten Punkte in den Einzelaufgaben der Probanden addiert: die Probanden können maximal 32 Punkte erlangen (pro richtig gelöster Aufgabe (Skalenausprägung „am meisten zutreffend“) werden 4 Punkte vergeben) und minimal 8 Punkte (bei Ankreuzen der Skalenausprägung „am wenigsten zutreffend“ wird 1 Punkt vergeben). Die Interpretation der Testergebnisse wird anhand der kriteriumsorientierten Testauswertung vorgenommen, d.h., ein Testrohwert wird in Bezug auf das Kriterium (in diesem Fall interkulturelle Handlungskompetenz) interpretiert. Im Gegensatz zur klassischen Testtheorie kann die Normierung bei kriteriumsorientierten Messungen in mehrfacher Hinsicht erfolgen (vgl. FRICKE, 1974):

1. Man kann die Differenz zwischen geschätztem und gewünschtem „universe score“ (der Wert für die Fähigkeit der Person) bilden.
2. Man kann diese Differenz in die zweistufige Entscheidung „Kriterium erreicht bzw. nicht erreicht“ transformieren.
3. Man kann Differenzmaße in mehrstufige Entscheidungen transformieren, z.B. bei der Vergabe von Noten.

Im Zusammenhang mit dem TIHK interessiert uns, ob der Proband das Kriterium interkulturelle Handlungskompetenz erfüllt hat oder nicht. Es ist also ein kritischer Punktwert (cut-off-point) festzulegen, mit dem jeder individuelle Testwert verglichen werden kann. Erhält der Proband einen Punktwert, der gleich oder größer einem kritischen Punktwert ist, besteht er den Test, andernfalls besteht er ihn nicht. Zur Festlegung eines kritischen Punktwertes für den Testrohwert wird in dieser Arbeit zum einen das Binomialmodell nach KLAUER (1972) und zum anderen das Zensierungsmodell von HERBIG (1974) verwendet.

In Anlehnung an das Modell von KLAUER (1972) wird der individuelle „universe score“ p_i aus der relativen Lösungshäufigkeit x_i/n der betreffenden Person geschätzt. Das 95%- oder 99%- Konfidenzintervall für diesen geschätzten Wert p_i kann aus Tabellen zur Binomialverteilung abgelesen werden (z.B. Documenta Geigy). Wie hoch der kritische Punktwert (wie viel Prozent des Tests der Proband lösen muss) des Tests liegen soll, wird vom Testentwickler festgelegt. Das Konfidenzintervall der Person gibt dann Auskunft darüber, ob diese den erwünschten kritischen Punktwert erreicht hat oder nicht.

Ein weiteres Modell der kriteriumsorientierten Auswertung stellt das Zensierungsmodell nach dem einfachen Binomialmodell von HERBIG (1974) dar. HERBIG quantifiziert den Grad der Erreichung des Kriteriums anhand von Noten. Die Note „4“ impliziert hier ein ausreichendes Ergebnis und damit eine Mindestleistung. Dieser Mindestleistung wird nun ein bestimmter

erwünschter „universe score“ p zugeordnet und die verbleibende Strecke $1-p$ in 4 äquidistante Teilstrecken aufgeteilt. HERBIG schlägt 4 Zensierungsmodelle vor, die von Mindest- p -Werten $p = 0,6$; $p = 0,7$; $p = 0,8$ und $p = 0,9$ ausgehen und dann zu 4 gewünschten p -Werten für die einzelnen Notenstufen führen. Hat man sich auf ein bestimmtes Zensierungsmodell festgelegt, resultieren für jede Testlänge 4 kritische Punktwerte, die sich unter Zuhilfenahme der Documenta-Geigy-Tabellen berechnen lassen. Das Zensierungsmodell ist abhängig von der Anzahl der Testaufgaben und muss bei unterschiedlichen Testlängen jedes Mal von neuem aus den Tabellen abgelesen werden.

Wendet man das oben dargestellte Modell nach KLAUER (1972) auf die Daten der Konstruktion des TIHK an, hätten laut Documenta-Geigy-Tabelle bei einem „universe score“ von $100 \times p = 90$ alle Probanden mit 25 Testpunkten das Kriterium erfüllt. Das impliziert, dass zum Bestehen des TIHK z.B. 7 Aufgaben mit 3 Punkten und eine Aufgabe mit 4 Punkten beantwortet werden müssen. Damit wird hier die Mindestanforderung pro Aufgabe bei 3 Punkten angesetzt. Da der Punktwert 3 der Skalenausprägung „eher zutreffend“ entspricht, deutet ein solches Antwortverhalten darauf hin, dass der Proband die dargestellte kritische Interaktionssituation in ihrer Tendenz richtig erkannt hat und annähernd kulturisomorph attribuieren kann. Dieses Vorgehen ist daher auch vor dem Hintergrund inhaltlicher Überlegungen gerechtfertigt. Bei Test A beläuft sich das Ergebnis auf 60 von 72 Testpersonen (83 %), die das Kriterium erreicht haben, also fähig zum Perspektivenwechsel sind. Für Test B ist ein Ergebnis von 59 von 72 Testpersonen (82 %) zu verzeichnen. Die Verteilung der Punktwerte der 72 Probanden der Analysestichprobe des TIHK zeigt die folgende Tabelle:

Rohwerte Testergebnis	Probanden Test A	Probanden Test B
16		1
19	2	
20	2	
21		1
22	1	1
23	3	4
24	4	6
25	5	3
26	12	9
27	5	12
28	12	14
29	8	4
30	10	8
31	8	8
32		1
Σ :	60	59

 = Kriterium erfüllt

Tabelle 21: Häufigkeiten der Punkteverteilung in der Analysestichprobe des TIHK (N = 72)

Hier wird erneut ersichtlich (vgl. Kap. 5.2.6), dass die Rohwerte von Test A verschieden zu Test B verteilt sind. Obwohl fast die gleiche Anzahl der Probanden im Test A ($N = 60$) und im Test B ($N = 59$) das Kriterium erreicht haben, stimmen einige Einzelergebnisse der Probanden in Test A nicht mit den Einzelergebnissen in Test B überein. Das bedeutet, dass im TIHK nicht zwingend die gleichen Punktwerte in Test A und B erreicht werden. Damit bestätigt sich analog den Häufigkeitspolygonen die nicht perfekte Parallelität.

Möchte man im TIHK die Ergebnisse der Probanden genauer differenzieren, könnte man zusätzlich das Zensierungsmodell von HERBIG (1974) anwenden. Setzt man dazu den „universe score“ für die ausreichende Leistung ebenfalls auf $p_4 = 0,90$ fest, lesen wir bei einem α -Fehler von 2,5 % aus den Documenta-Geigy-Tabellen die kritischen Punktwerte für $p_4 = 0,90$; $p_3 = 0,925$; $p_2 = 0,95$ und $p_1 = 0,975$ ab und erhalten $k_4 = 25$; $k_3 = 26$; $k_2 = 28$ und $k_1 = 29$. Die Beziehung zwischen Testpunktwert und Note zeigt unten stehende Tabelle:

X %	Testwert x	Note
78 %	25	4
81 %	26	3
84 %	27	3
88 %	28	2
91 %	29	1
94 %	30	
97 %	31	
100 %	32	

X % = Prozentsatz der richtig gelösten Aufgaben im Test

Tabelle 22: Zensierungsmodell nach HERBIG (1974) zur Interpretation der Testergebnisse des TIHK

Hier wird ersichtlich, dass ab 25 Punkten die Probanden die Mindestleistung (Note „4“) erreicht haben, was dem „universe score“ bei KLAUER (1972) entspricht. Die Note „3“ wird nach diesem Modell mit 26 und 27 Punkten vergeben, die Note „2“ mit 28 Punkten und die Note „1“ ab 29 bis 32 Punkten. Nach dem Modell von HERBIG (1994) konnten in Test A des TIHK 26 Probanden die Note „1“, 12 die Note „2“, 17 die Note „3“ und 5 die Note „4“ erreichen. Bei Test B ergibt sich ein ähnliches Bild mit 21 Probanden mit der Note „1“, 14 Probanden mit der Note „2“, 19 Probanden mit der Note „3“ und 3 Probanden mit der Note 4. Jeweils 12 (Test A) und 13 (Test B) Probanden haben die Mindestleistung nicht erreicht. Eine weitere Differenzierung der Notenstufen „5“ und „6“ ist an dieser Stelle nach Einschätzung der Autorin nicht weiter bedeutsam, da lediglich interessiert, ob die Probanden das Kriterium erfüllen konnten oder nicht. Das gute Abschneiden der Probanden (Test A: 38 Probanden erreichen die Noten „1“ oder „2“, Test B: 35 Probanden erreichen die Noten „1“ oder „2“) ist auf die Kürze des Tests zurückzuführen, da kürzere Tests größere Messfehler haben, die sich zu-

gunsten der Personen auswirken: Die gleiche prozentuale Leistung führt bei kürzeren Tests zu besseren Noten (FRICKE, 1974). Die Verteilung der Noten der 72 Probanden der Analysestichprobe des TIHK zeigt die folgende Tabelle im Überblick:

Note	Testpunkte	Probanden Test A	Probanden Test B
4	25	5	3
3	26	12	9
3	27	5	12
2	28	12	14
1	29	8	4
	30	10	8
	31	8	8
	32		1

Tabelle 23: Häufigkeiten der Notenverteilung in der Analysestichprobe des TIHK (N = 72)

Mit den Modellen von KLAUER (1972) und HERBIG (1974) sind 2 Interpretationswege des TIHK aufgezeigt worden. Je nach Zielsetzung und Einsatz des TIHK kann zwischen den Modellen entschieden werden bzw. können beide auch kombiniert angewandt werden. Die Entscheidung, wie hoch der kritische Wert des TIHK angesetzt wird, liegt beim Testanwender. Aufgrund der theoretischen Überlegungen ist ein kritischer Punktwert von 24 sinnvoll, der aber je nach Gesamtleistung der Stichprobe im Test höher oder niedriger angesetzt werden kann.

7 Diskussion

Die Diskussion umfasst die Reflexion der an den Test zur Erfassung interkultureller Handlungskompetenz gestellten Anforderungen. Dies erfolgt vor dem Hintergrund der ausgearbeiteten theoretischen Überlegungen. Weiterhin wird das gewählte Entwicklungsdesign kritisch beleuchtet und Anregungen zur Modifizierung des bestehenden Messinstrumentes gegeben, um dessen Anwendbarkeit zu verbessern. Zum Schluss dieses Kapitels wird ein Ausblick auf weiterführende Forschung sowie auf Anwendungsmöglichkeiten in der Praxis gewährt.

7.1 Die Bewertung der Methode aus dem Blickwinkel der theoretischen Modelle

Ein Aspekt, der an bestehenden interkulturellen Testverfahren immer wieder bemängelt wurde, ist die fehlende Anbindung an theoretische Konzepte bzw. hinreichende Operationalisierungen der zu erfassenden Konstrukte (vgl. Kap. 2).

Der vorliegende TIHK basiert auf der Definition interkultureller Handlungskompetenz, welche die Fähigkeit zum Perspektivenwechsel als Indikator für interkulturellen Handlungserfolg hervorhebt. Die Bedeutung der Fähigkeit zum Perspektivenwechsel ist in der interkulturellen Forschung allgemein anerkannt (vgl. KIM & RUBEN, 1988; TAYLOR, 1994; BRISLIN, LANDIS & BRANDT, 1983 etc.) und konnte empirisch untermauert werden (HATZER, 2000). Ein bewährtes und verbreitetes Konzept, das auf der Fähigkeit zum Perspektivenwechsel aufbaut, ist das der isomorphen Attributionen von TRIANDIS (1975, 1977). Eine Methode, die Fähigkeit zu kulturisomorphen Attributionen zu entwickeln, stellt der Culture Assimilator dar. Daher wurde der TIHK anhand empirisch gewonnenen Culture Assimilator Materials, das diese Fähigkeit in Form von kritischen Interaktionssituationen fördert, konstruiert. Somit wurde interkulturelle Handlungskompetenz im vorliegenden TIHK anhand der Fähigkeit des Perspektivenwechsels durch isomorphes Attribuieren messbar gemacht.

Der TIHK wird damit der Forderung gerecht, das zu messende Konstrukt gut operationalisiert und eine empirisch abgesicherte Methode verwendet zu haben. Die Ergebnisse der Validierung von GÖTZ (2003), wonach das Antwortverhalten im Test mit dem im Rollenspiel gezeigten Verhalten der Probanden übereinstimmt, sprechen für die Gültigkeit des Instruments. Das Resultat des dargestellten Entwicklungsprozesses ist somit eine hinreichend reliable und valide Testform zur Erfassung interkultureller Handlungskompetenz.

7.2 Kritische Reflexion des Entwicklungsdesigns

7.2.1 Das Datenmaterial

Der vorliegende Test zur Erfassung interkultureller Handlungskompetenz (TIHK) besteht aus 2 Parallelformen zu jeweils 8 Aufgaben (kritischen Interaktionssituationen). Die kritischen Interaktionssituationen des TIHK sind den Culture Assimilators England, USA und Indonesien entnommen. Bei der Wahl des Datenmaterials sind 2 Aspekte kritisch zu beleuchten. Zum einen stellt sich die Frage der Repräsentativität der ausgewählten Kulturen für die kulturelle Vielfalt unserer Welt. Ob der Engländer den typischen Europäer verkörpert oder der Amerikaner aufgrund der Kulturhistorie eher dem englischen Volk ähnelt bzw. ob der indonesische Landsmann sich symptomatisch für die asiatische Bevölkerung verhält, wird allgemein schwer zu beantworten sein und bei Vertretern der jeweiligen Kultur auf wenig Zustimmung treffen. Zum anderen ist die generelle Anzahl der verwendeten Länder in einem Test interkultureller Handlungskompetenz zu kritisieren.

Dem ist entgegenzuhalten, dass die Wahl gerade dieser 3 Kulturen vor allem pragmatische Gründe (vgl. Kap. 6.1.1) hatte. Um die existierende kulturelle Vielfalt in einem Attributionstest wiedergeben zu können, müssten sicherlich mehrere Kulturen berücksichtigt werden. Welche Kulturen dann auf jeden Fall mitaufgenommen werden sollten, ist schwer zu beurteilen, bedenkt man die Einzigartigkeit jeder Kultur. Trotz allem wurde stellvertretend für den europäischen, westlichen und östlichen Kulturraum Culture Assmilator Material gewählt, um zu überprüfen, ob der Proband sich prinzipiell in die verschiedenen Kulturräume hineinversetzen kann. Darüber hinaus erschien es wichtiger, dass die verwendeten kritischen Interaktionssituationen für die Zielgruppe, also für Studenten, nachvollziehbar sind. Vor diesem Hintergrund ist das beschriebene Vorgehen durchaus gerechtfertigt.

Die Anzahl von 8 Aufgaben je Testform fällt aufgrund der Paralleltestkonstruktion geringer aus als ursprünglich geplant. Obwohl dieser Gedanke bei der Generierung des Itempools berücksichtigt wurde, mussten viele Aufgaben auf statistischer Ebene zugunsten einer zufrieden stellenden Reliabilität aufgegeben werden. Eine weitere Schwierigkeit, die sich bei der Paralleltestkonstruktion ergab, war die Beachtung des Kulturstandardkonzepts innerhalb der verwendeten KIs. Eine Forderung an Paralleltests betrifft die inhaltliche Äquivalenz der Aufgaben, so dass die Aufgaben für Testform B aus den gleichen Kulturstandards wie aus Testform A entnommen werden mussten und damit die Auswahlmöglichkeiten eingeschränkt waren. Zudem hätte eine noch größere Anzahl von Aufgaben in den Testvorformen die Probanden bei der Lösung überfordert. Bei 36 Testaufgaben müssen zur Bearbeitung des Tests

eineinhalb bis 2 Stunden veranschlagt werden. Es ist leicht vorstellbar, dass die Motivation mit steigender Anzahl der Aufgaben nachlassen kann und deswegen zu befürchten ist, dass der Test weniger sorgfältig bearbeitet wird. Um aber valide Aussagen über die Fähigkeit zum Perspektivenwechsel treffen zu können, ist es erforderlich, dass der Proband die einzelnen Situationen sorgfältig liest und versucht, sich in sie hineinzusetzen, bevor er eine Aussage über die Antwortalternativen treffen kann.

7.2.2 Die Skalierung der Aufgaben

Die Aufgaben des TIHK bestehen aus der Schilderung der kritischen Interaktionssituationen sowie jeweils 4 Antwortalternativen, die mögliche Gründe für das jeweilige Verhalten des fremdkulturellen Partners in der dargestellten Situation nennen. Die Antwortalternativen sind durch eine 4-stufige Skala mit den Ausprägungen „am meisten zutreffend“, „eher zutreffend“, „eher nicht zutreffend“ und am „wenigsten zutreffend“ gekennzeichnet und sollen von den Probanden im Hinblick auf ihre Güte in eine Rangreihe gebracht werden.

Eine Schwierigkeit, die sich bei der Konstruktion von Skalen häufig ergibt, ist deren Abstufung: „Die Zahl der Stufen – üblicherweise zwischen 3 und 11 – muss einerseits auf das Differenzierungsvermögen des Urteilers und andererseits den Differenzierungsbedarf des Forschers abgestimmt sein“ (ROHRMANN, 1978, S. 223). Das bedeutet im Fall des TIHK, dass eine Antwortalternative auf einer 5-, 6- oder 7-stufigen Skala für den Testanwender zugunsten eines besseren Differenzierungsvermögens wünschenswert wäre. Damit lässt sich auch die geringe Varianz in den Testergebnissen der Analysestichprobe erklären, da mehrstufige Skalen zu einer größeren Varianz im Antwortverhalten führen (MUMMENDEY, 1995). Ein weiteres Argument für eine Verbesserung der Differenzierungsfähigkeit der Skala im TIHK bezieht sich auf die Erfassung des Lernzuwachses anhand des TIHK. Sind die Testwerte von Probanden im TIHK von vornherein sehr hoch, ist es schwierig, eine deutliche Verbesserung der Perspektivenwechselfähigkeit messbar zu machen. Dies könnte auch bedeuten, dass Veränderungen der Personen mit hohen Werten sich nicht auf der Skala bemerkbar machen.

Dagegen spricht die von ROHRMANN (1978) erwähnte Differenzierungsfähigkeit der Probanden, die bei einer 5-, 6- oder 7-stufigen Skala nicht mehr in der Lage sind, inhaltlich zwischen den Stufen zu unterscheiden. Aus diesem Grund erscheint eine 4-stufige Skala, mit der Einschränkung des Informationsverlusts für den Testanwender, durchaus als vertretbar, wenn auch Forschungen zur Weiterentwicklung einer Skalierung wünschenswert wären.

7.2.3 Die Objektivität des Tests

Die Objektivität eines Test ist neben Reliabilität und Validität ein weiteres testtheoretisches Gütekriterium. Sie gibt an, inwiefern die Ergebnisse eines Tests unabhängig vom Untersucher sind. Man unterscheidet hier „Durchführungsobjektivität“, „Auswertungsobjektivität“ und „Interpretationsobjektivität“ (LIENERT & RAATZ, 1994, S.8).

Die **Durchführungsobjektivität** betrifft die Frage, inwieweit das Verhalten des Untersuchers während der Testdurchführung das Ergebnis beeinflussen haben könnte. Bei der vorliegenden Untersuchung wurde das Testmaterial den Probanden mit nach Hause gegeben, so dass die Autorin während der Testbearbeitung nie persönlich anwesend war. Die Testanweisung war schriftlich abgefasst und für alle Probanden identisch. Die Durchführungsobjektivität hinsichtlich der Testinstruktion kann demnach als „hoch“ bezeichnet werden. Da jedoch die Bedingungen unter denen die Probanden den Test bearbeiteten, der Autorin nicht bekannt sind, ist die Durchführungsobjektivität trotzdem als kritisch zu beurteilen. Aufgrund der großen Anzahl an Probanden ist es nicht möglich, das Bearbeiten des Tests bei allen Probanden zu kontrollieren. Die hohe Paralleltest-Reliabilität spricht allerdings dafür, dass die Probanden unter vergleichbaren Bedingungen gearbeitet haben.

Die **Auswertungsobjektivität** gibt an, inwieweit die Vergabe von Testpunkten für bestimmte Testantworten von der Person des Auswerters unabhängig ist. Im Fall des TIHK sind keine freien Beantwortungen der Fragen möglich, sondern der Proband hat lediglich die Möglichkeit, zwischen mehreren fest vorgegebenen Kategorien zu wählen. Aus diesem Grund kann festgehalten werden, dass die Auswertungsobjektivität bei dieser Art der Auswertung nahezu vollkommen verwirklicht ist. Sie könnte nur durch Fehler beim Eintragen der Lösungen des Teilnehmers in das Statistikprogramm SPSS oder durch einen Fehler dieses Auswerteprogramms geschmälert worden sein. Beides kann aufgrund mehrfacher Überprüfung ausgeschlossen werden.

Die **Interpretationsobjektivität** besagt, dass individuelle Deutungen nicht in die Interpretation eines Testwertes miteinfließen dürfen, sie beschäftigt sich also mit der Frage, inwieweit die inhaltliche Interpretation des Gesamttestergebnisses unabhängig vom Untersucher ist. Die Interpretationsobjektivität ist dann maximal, wenn jeder numerische Wert eindeutig einer inhaltlichen Aussage zugeordnet wird. Im Fall des TIHK ergibt sich pro Proband ein numerischer Gesamtwert, dem eine inhaltliche Aussage eindeutig zugeordnet wird (vgl. Kap. 6.5), so dass man auch in diesem Fall von einer ausreichenden Interpretationsobjektivität sprechen kann.

Zusammenfassend kann also festgehalten werden, dass der vorliegende TIHK dem testtheoretischen Gütekriterium der Objektivität gerecht wird.

7.2.4 Die Parallelität der Testformen

Die sehr guten Ergebnisse der Berechnung der inneren Konsistenz und der Paralleltestmethode zeigen, dass die Parallelität des TIHK gegeben ist. Aufgrund der unterschiedlichen Rohwertverteilungen (vgl. Kap. 6.2.5) und der abweichenden Testergebnisse (vgl. Kap. 6.5) der Skalen A und B muss allerdings eingeräumt werden, dass keine perfekte Parallelität erreicht werden konnte.

Das Ergebnis der abweichenden Rohwertverteilungen kann methodisch erklärt werden. Die Reihung der Aufgaben im Test kann zu unterschiedlichen Rohwertverteilungen führen (LIENERT UND RAATZ, 1994). Um eine regelmäßige Rohwertverteilung zu begünstigen, sollten die Aufgaben bei Niveautests hinsichtlich ihrer Schwierigkeit monoton ansteigend geordnet werden. Da die Schwierigkeitsindices der Testformen A und B nicht identisch sind und die Aufgaben hinsichtlich der Schwierigkeitsindices ungeordnet der Analysestichprobe präsentiert wurden, sind unterschiedliche Rohwertverteilungen und abweichende Testergebnisse der Testformen A und B zu erwarten.

Neben den methodischen Mängeln ist allerdings auch zu überlegen, wie realistisch eine perfekte Parallelität des TIHK auf theoretischer Ebene ist. Es ist doch eher unwahrscheinlich, dass Probanden zu 2 verschiedenen Messzeitpunkten zwingend genau die gleiche Punktzahl erreichen, berücksichtigt man Lerneffekte, Tagesform und Kontextvariablen beim Lösen des Tests. Damit ist zum einen gemeint, dass Lerneffekte zu höheren Testergebnissen in Test B führen können. Zum anderen ist die Aufmerksamkeits- und Leistungsfähigkeit abhängig von der Tagesform des Probanden, so dass auch in diesem Fall abweichende Punktwerte in den Testformen zu erwarten sind. Ebenso können sich unterschiedliche Rahmenbedingungen, unter denen die Testformen bearbeitet werden, auf die Testergebnisse auswirken.

Vor diesem Hintergrund kann die Erzielung der Ergebnisse der Paralleltest-Reliabilität als zufrieden stellend beurteilt werden. Es liegt somit ein Paralleltest vor, der sinnvoll anwendbar ist und von dem gute Ergebnisse zu erwarten sind.

7.2.5 Die Testkonstruktion

Zur Testkonstruktion des TIHK wurden die Mittelwerte der Probanden über den gesamten Test berechnet. Da durch die Berechnung von Mittelwerten Informationen verloren gehen, wäre es im Sinne einer besseren Differenzierungsfähigkeit des Tests im Nachhinein sinnvoller gewesen, den TIHK anhand einer anderen Maßzahl zu entwickeln. Hier würden sich die Summenwerte über alle Aufgaben pro Proband anbieten, die auch bei der Interpretation der Testergebnisse verwendet worden sind (vgl. Kap. 6.5). Auf diese Weise ist eine höhere Varianz der Testergebnisse zu erwarten, die sich positiv auf die Differenzierungsfähigkeit des Tests auswirkt. In der vorliegenden Arbeit wurde jedoch auf die Methode der Mittelwerte zurückgegriffen, um die Probanden auf der verwendeten 4er Skala lokalisieren zu können. Unabhängig von der inhaltlichen Interpretation wird vorgeschlagen, die Testkonstruktion zukünftig grundsätzlich an den Gesamtrohwerten vorzunehmen. Dort ist der statistische Mehrgewinn bedeutsamer als bei der Interpretation der Testergebnisse.

Als weitere methodische Kritik im Rahmen der Testkonstruktion des TIHK sind 2 statistische Verfahren anzuführen. Zum einen betrifft dies die Vorgehensweise zur Bestimmung der Schwierigkeitsindices (vgl. Kap. 5.2.1.1). Der gewählten Vorgehensweise zur Bestimmung der Aufgabenschwierigkeit wurde ein „Richtig-Falsch“-Antwortschema zugrunde gelegt. Pro Aufgabe gab es eine Antwortalternative mit der richtigen Lösung, die 3 anderen Antworten wurden mit falsch bewertet. Somit wurde die geforderte Rangreihenbildung der Antworten (vgl. Kap. 5.2.2.1) wieder außer Kraft gesetzt. Dadurch gingen Informationen über die Güte des Antwortverhaltens des Probanden verloren, nämlich wie nah der Proband an der richtigen Lösung der Aufgabe war. Eine Möglichkeit, dem entgegenzuwirken, stellt der plausibilitätskorrigierte Schwierigkeitsindex von HORST (1971) dar, der die meistfrequentierteste falsche Antwortalternative zur Berechnung des Schwierigkeitsindex berücksichtigt. Da die Formel von HORST (1971) jedoch wenig erprobt und damit in der Anwendung zweifelhaft ist, wurde auf die bewährte Methode zur Berechnung des Schwierigkeitsindex zurückgegriffen.

Das zweite zu kritisierende Vorgehen bei der Konstruktion des TIHK betrifft die Verwendung der c-Koeffizienten von SIMON ET AL. (in Vorbereitung) (vgl. Kap. 6.2.3.3). Dieser hätte sinnvollerweise bereits zur Konstruktion von Testform A eingesetzt werden müssen, um einem einheitlichen methodischen Vorgehen gerecht zu werden und aussagekräftige Ergebnisse zu gewährleisten. Da die Ergebnisse der Testkonstruktion A allerdings keinen Anlass gaben, die Methoden der klassischen Testtheorie anzuzweifeln, wurde erst an späterer Stelle (bei der Konstruktion von Test B) offensichtlich, dass die Verfahren der klassischen Testtheorie im vorliegenden Fall unzureichend waren (vgl. Kap. 6.2.3.2). Infolgedessen wurde von der Auto-

rin im Rahmen der kriteriumsorientierten Messung der c-Koeffizient als adäquate Methode identifiziert und eingesetzt. Die Ergebnisse der Konstruktion von Test A sind somit ein generelles Beispiel dafür, wie schnell Artefakte in der Psychologie entstehen können. Es ist anzunehmen, dass der ein oder andere in der Literatur publizierte Test (speziell jene, die auf Likert-Skalen-Niveau beruhen) demzufolge vielleicht auch andere Reliabilitäts- und Validitätswerte aufweist als angegeben. Dies gilt insbesondere für Tests mit geringer Varianz und nicht normalverteilten Daten.

Trotz kleiner methodischer Mängel ist zu konstatieren, dass mit der angewandten Methode ein durchaus adäquater Test zur Erfassung interkultureller Handlungskompetenz mit einer Parallelförmigkeit konstruiert werden konnte. Dafür sprechen die hohen Werte der inneren Konsistenz und der Paralleltest-Reliabilität. Die erkannten Mängel führen zu der Überlegung, dass mit anderen KIs ein noch besseres Ergebnis erreicht werden könnte. Wenn man auf der anderen Seite jedoch davon ausgeht, dass alle Aufgaben aus dem gleichen Culture Assimilator Material stammen, ist anzunehmen, dass selbst die Auswechslung von Aufgaben keine großen Veränderungen mit sich bringen würde.

7.2.6 Die Rohwertverteilung

Ein auffälliges Phänomen, das bei der Konstruktion des TIHK aufgetreten ist und von der klassischen Testtheorie zur Hinwendung zur kriteriumsorientierten Leistungsmessung führte, sind die guten Ergebnisse der Probanden und die damit verbundene rechtsschiefe Rohwertverteilung. Die vorliegende anomale Verteilungsform der Rohwerte ist kein Grund, den TIHK zu verwerfen, legt aber eine Ergründung der Ursachen nahe. Das Auftreten hoher Ergebnisse einer großen Zahl von Probanden sowohl bei der Konstruktion von Test A als auch bei der Konstruktion von Test B (dies entspricht einer Gesamtstichprobe von 119 Versuchspersonen) legt mehrere Schlussfolgerungen nahe, die im Folgenden dargestellt werden.

7.2.6.1 Das Culture Assimilator Format

Eine Erklärung für die rechtsschiefe Verteilung der Rohwerte könnte die Formulierung der Antwortalternativen des Culture Assimilator Materials sein. Da viele Aufgaben von bis zu 80 Prozent der Probanden gelöst werden konnten, liegt die Überlegung nahe, dass es Antwortalternativen gibt, deren Inhalte weniger glaubwürdig erscheinen oder die weniger gut

formuliert sind, so dass diese per se nicht als mögliche richtige Interpretationsmöglichkeit in Betracht gezogen werden. Diese Erklärung leuchtet ein, wenn man das Procedere der Gewinnung des Culture Assimilator Materials berücksichtigt: In der Regel werden die kritischen Interaktionssituationen des Culture Assimilators aufgrund von Interviewmaterial generiert. Dabei wird die kulturadäquate Antwort der dargestellten Situation von Experten der relevanten Kulturen gegeben. Die kulturinadäquaten Antwortalternativen werden durch Befragung von Laien (im Fall des TIHK durch Deutsche) gewonnen. Da das Culture Assimilator Training ein bestimmtes Format, nämlich 4 Antwortalternativen mit einer kulturadäquaten und 3 kulturinadäquaten Interpretationen vorgibt, ist es durchaus denkbar, dass bei diesem Vorgehen Antwortalternativen z.T. aus der Not entstehen, das vorgegebene Antwortmuster zu erfüllen. Diese sind dann unter Umständen anderen Antwortalternativen qualitativ unterlegen und werden infolgedessen mit geringerer Wahrscheinlichkeit vom Probanden als relevant erachtet. Die kulturadäquate Antwort könnte deswegen bevorzugt ausgewählt werden, da die anderen Antwortalternativen zu unglaubwürdig erscheinen.

7.2.6.2 Die Aufgabenanordnung

Eine weitere Ursache dafür, dass viele Aufgaben von vielen Probanden richtig gelöst wurden, könnte in der Anordnung der Aufgaben begründet liegen. Statistisch gesehen sollte der ideale Schwierigkeitsverlauf nach LIENERT UND RAATZ (1994) wie bereits erwähnt (vgl. 7.2.4) annähernd geradlinig sein, nämlich monoton ansteigend bei Aufgaben, die - wie beim Niveautest üblich - nach aufsteigender Schwierigkeit geordnet sind. „Unter diesen Bedingungen kann man - homogene Stichprobe und normalverteiltes Persönlichkeitsmerkmal vorausgesetzt - eine symmetrische Verteilung der Testwerte erwarten“ (LIENERT UND RAATZ, 1994, S.156). Da dies beim TIHK nicht der Fall ist, könnte darin die Ursache der rechtsschiefen Verteilung gesehen werden. Auf der anderen Seite ist es bei einem so eindeutigen Ergebnis wie im Fall des TIHK jedoch höchst unwahrscheinlich, dass die Rohwertverteilung selbst bei monoton ansteigender Aufgabenschwierigkeit ein anderes Bild ergeben würde. Vielmehr scheint es einleuchtend, die Gründe bei der Auswahl der Stichprobe oder dem zu messenden Konstrukt zu suchen.

Im Zusammenhang mit den Testaufgaben ist darüber hinaus ein weiterer Aspekt zu beachten. Um ausschließlich hohe Testergebnisse zu vermeiden und damit die Bandbreite der möglichen Testergebnisse im TIHK zu erweitern, ist es notwendig, den Test „schwerer“ zu machen. Konträr zu den Empfehlungen der klassischen Testkonstruktion müssen dazu schwierigere Aufgaben (also mit niedrigeren Schwierigkeitsindices) bevorzugt verwandt wer-

den. Schwierigere Aufgaben führen zu einer größeren Varianz der Testergebnisse und zu einer größeren Differenzierungsfähigkeit zwischen den Probanden. Dieser Gedanke findet sich auch bei CUSHNER (1989) wieder, der ebenfalls schwierige kritische Interaktionssituationen für einen Attributionstest empfiehlt.

7.2.6.3 Die Stichprobe

Was die Auswahl der Stichprobe betrifft, so ist anzunehmen, dass das Bildungsniveau und die intellektuellen Fähigkeiten der Probanden nicht den Durchschnitt der Gesamtbevölkerung widerspiegeln. Die untersuchten Personen könnten durch ihr Wissen und ihre Fähigkeit zur Reflexion einen höheren Grad an interkultureller Handlungskompetenz mitbringen als andere Bevölkerungsgruppen und deswegen im TIHK besser abschneiden. Diese Hypothese erscheint auf den ersten Blick zwar plausibel, muss aber durch weitere Untersuchungen erst noch überprüft werden.

Ein weiterer Grund für unregelmäßige Rohwertverteilungen können heterogene Stichproben sein. Diese entstehen dann, „wenn sich 2 Gruppen nach ihren Mittelwerten und nach ihrer Größe wesentlich voneinander unterscheiden; die Varianz kann dabei gleich oder auch unterschiedlich sein“ (LIENERT UND RAATZ, 1994, S. 154). Da die Mehrzahl der untersuchten Probanden Psychologiestudenten waren, wird an der zweiten Analytestichprobe zur Konstruktion des Paralleltests nun überprüft, ob sich diese Gruppe in ihrer Größe und in ihrem Mittelwert von der Gruppe der Studenten aller anderen Studiengängen unterscheidet und deswegen auf eine rechtsschiefe Rohwertverteilung geschlossen werden kann. Die Unterteilung der Stichprobe in Psychologen und „Nichtpsychologen“ beruht auf der Annahme, dass Psychologiestudenten möglicherweise aufgrund ihres Studienfaches über ein höheres Reflexionsvermögen als Studenten anderer Studiengänge verfügen und deshalb im TIHK besser abschneiden. Eine weitere Überlegung bezieht sich auf die Mutmaßung, dass die Psychologiestudenten aufgrund des Lehrangebots im Psychologiestudium an der Universität Regensburg bereits die Gelegenheit nutzen konnten, sich mit der interkulturellen Thematik auseinanderzusetzen und deswegen im TIHK höhere Ergebnisse erzielen. Diese Hypothese ließ sich allerdings nicht bestätigen. Die Gruppe der Psychologen unterscheidet sich zwar in ihrer Größe ($N = 45$) deutlich von der Gruppe der anderen Studienfächer ($N = 27$), nicht jedoch in ihren Mittelwerten (Psychologen: $M = 3,41$ in beiden Testformen, „Nichtpsychologen“: Test A: $M = 3,35$, Test B: $M = 3,38$). Die deskriptive Statistik der beiden Gruppen zeigt die folgende Tabelle:

	"Psychologen"		"Nichtpsychologen"	
	Test A	Test B	Test A	Test B
N	45	45	27	27
Mittelwert	3,41	3,41	3,35	3,38
Median	3,50	3,38	3,25	3,50
Modus	3,50	3,38	3,25	3,50
Standardabweichung	0,35	0,30	0,41	0,44
Minimum	2,38	2,88	2,38	2,00
Maximum	3,88	4,00	3,88	3,88

Tabelle 24: Deskriptive Statistik der Gruppe der Psychologen und „Nichtpsychologen“ der Analysestichprobe zur Konstruktion von Test B

Da aufgrund der dargestellten deskriptiven Statistik die Hypothese heterogener Untergruppen von Psychologiestudenten und Studenten anderer Fachrichtungen nicht bestätigt werden kann, liegt die Ursache möglicherweise in einer anderen Variable begründet.

Im Zusammenhang heterogener Stichproben könnte weiter angenommen werden, dass die Stichprobe sich aufgrund anderer Variablen in 2 Gruppen einteilen lässt. Eine denkbare Hypothese zielt darauf ab, dass Probanden mit großer interkultureller Vorerfahrung besser im TIHK abschneiden als solche, die keine interkulturellen Vorerfahrungen vorzuweisen haben. Dadurch wird das Ergebnis in Richtung einer rechtsschiefen Verteilung verzerrt. Um diese Hypothese zu überprüfen, könnte in Anlehnung an BHAWUK UND BRLISLIN (1992) die Einteilung der Analysestichprobe in interkulturell Vorerfahrene und interkulturell nicht erfahrene Probanden anhand des Kriteriums „Auslandserfahrung in Jahren“ vorgenommen werden. Im Anschluss daran würde man die Testergebnisse der beiden Gruppen wie im Abschnitt zuvor anhand ihrer Größe und ihrer Mittelwerte vergleichen. Da zur Überprüfung der Hypothese die relevanten Daten (Auslandserfahrung) jedoch nicht erhoben wurden, kann an dieser Stelle lediglich angeregt werden, diesen Aspekt bei weiteren Forschungsvorhaben zu berücksichtigen.

Da die Frage nach den Ursachen für die sehr guten Ergebnisse im TIHK nicht aufgrund des verwendeten Testmaterials, der Aufgabenanordnung oder der Konstitution der Stichprobe zufrieden stellend beantwortet werden kann, wird das Augenmerk im Folgenden auf den Aspekt des zu erfassenden Konstrukts gelenkt.

7.2.6.4 Das zu messende Konstrukt

Zur Beschreibung des Konstrukts interkultureller Handlungskompetenz setzen einige Forschungsansätze bei der Betrachtung des Konstrukts der sozialen Kompetenz an (vgl. FURNHAM & BOCHNER, 1986; EDER, 1996; BOLTEN, 2001). BOLTEN (2001) geht davon aus, dass es sich bei interkultureller Handlungskompetenz (Fremdsprachenkenntnisse und kulturelles Wissen ausgeschlossen) um Eigenschaften handelt, die allgemein das Handeln von Menschen prägen. Er versteht interkulturelle Handlungskompetenz nicht als eigenständigen Kompetenzbereich, sondern als die Fähigkeit, individuelle, soziale, fachliche und strategische Teilkompetenzen in ihrer bestmöglichen Verknüpfung auf interkulturelle Handlungskontexte beziehen zu können. Menschen, die im eigenkulturellen Orientierungsrahmen sozial kompetent handeln, haben demnach auch in der interkulturellen Überschneidungssituation das richtige Gespür, sich in einer fremden Kultur adäquat zu verhalten. Die Besonderheit im Gegensatz zu einer intrakulturellen Begegnung stellt hier, so BOLTEN (2001), lediglich der Transfer der sozialen Kompetenz auf die interkulturelle Situation dar.

Bezogen auf den TIHK hätte das zur Konsequenz, dass Probanden den Test aufgrund ihrer allgemeinen sozialen Kompetenzen lösen können. Die Testergebnisse fallen so hoch aus, da fast alle Menschen sozial kompetent sind, wie die Arbeit von SCHMICK (2001) zeigt, die nachweisen konnte, dass der Aspekt der Teamfähigkeit (als eine Komponente sozialer Kompetenz) bei fast allen Personen stark ausgeprägt ist. Soziale Kompetenz stellt zwar sicherlich einen relevanten Aspekt in der kulturellen Überschneidungssituation dar, kann aber die Fähigkeit interkultureller Handlungskompetenz nicht hinreichend erklären, bedenkt man die Häufigkeit mit der kritische Interaktionssituationen in der interkulturellen Begegnung entstehen. An dieser Stelle sei auf THOMAS (1993) hingewiesen, der davon ausgeht, dass in der kulturellen Überschneidungssituation Schwierigkeiten deswegen entstehen, weil Wahrnehmen, Denken, Werten und Handeln an unterschiedlichen Kulturstandards ausgerichtet sind, die während der monokulturellen Sozialisation internalisiert werden und dadurch nicht bewusstseinspflichtig sind (vgl. Kap. 3.1.2).

Diese Ausführungen sowie die Bestätigung der hohen Testergebnisse in der Validierung von GÖTZ (2003) führen zu der Überlegung, das Konstrukt der interkulturellen Handlungskompetenz aus einem anderen Blickwinkel zu betrachten. So wird hier angenommen, dass Menschen, die für eine bestimmte Zeit im Ausland leben und arbeiten, sich zu Beginn eines Auslandsaufenthaltes bewusst sind, auf fremdartige Menschen und Lebensweisen zu stoßen und demnach auf fremde Situationen eingestellt sind. Aus der intrakulturellen Erfahrung wissen sie, wie man in sozialen Situationen adäquat oder auch sozialverträglich handelt.

Dadurch, dass sie für die Kultur sensibel sind, gehen sie zunächst offen auf fremdkulturelle Partner zu und sind besonders bemüht, sich in der fremden Situation zurechtzufinden. Erst im Verlauf der Zeit eines Auslandsaufenthalts ergeben sich Schwierigkeiten in der kulturellen Überschneidungssituation. Die betroffenen Personen machen sich die Immanenz der fremden Kultur nicht ständig bewusst, was zur Folge hat, nicht immer sensibel auf die jeweilige Situation reagieren zu können. Daraus folgt, dass Probleme innerhalb interkultureller Begegnungen erst im Verlauf der Zeit entstehen. Mit fortlaufender Dauer des Auslandsaufenthaltes kommt zusätzlich die emotionale Belastung der Personen hinzu. Nach einer meist kurzen „honeymoon stage“ (APFELTHALER, 1999, S. 16), in der die Fremdheit der anderen Kultur als an- oder aufregend empfunden wird, werden Unsicherheiten in und Zweideutigkeit von Situationen evident, was wiederum zu Enttäuschungen und Ärger führen kann. Interkulturelle Handlungskompetenz wird hier also durch den Zeitaspekt eines Auslandsaufenthaltes moderiert.

Diese Überlegungen lassen sich auf den TIHK übertragen. Die Aufgaben im TIHK beschreiben kritische Interaktionssituationen, die Erleben und Verhalten in Alltagssituationen thematisieren. Die dargestellte Situation im Test macht den Probanden, ähnlich wie in der realen interkulturellen Überschneidungssituation, auf die interkulturelle Thematik aufmerksam, so dass dieser auf die Andersartigkeit des fremdkulturellen Verhaltens vorbereitet ist. Die Probanden, die den TIHK ausfüllen, sind also auf die interkulturelle Situation eingestellt, dementsprechend sensibel für die dargestellten interkulturellen Konflikte und in der Lage, den interkulturellen Perspektivenwechsel zu vollziehen. Infolgedessen lösen sie die gestellten Aufgaben richtig und erzielen hohe Testergebnisse. Die Probanden, die den TIHK bestehen, haben realisiert, worum es in der dargestellten Situation geht. Wer demzufolge im TIHK gut abschneidet, hat eine Disposition zu interkultureller Handlungskompetenz, da er sich der interkulturellen Problematik bewusst ist. Aufgrund der oben dargelegten moderierenden Variable „Zeit“, kann es im Verlauf eines Auslandsaufenthaltes bei diesen Personen in realen kulturellen Überschneidungssituationen trotzdem zu kritischen Auseinandersetzungen oder Konflikten kommen. Wurde aber eine Disposition zu interkultureller Handlungskompetenz im TIHK diagnostiziert, gelten diese Probanden als besonders förderungswürdig in dem Sinne, dass sie z.B. durch interkulturelle Trainings auf einen Auslandsaufenthalt vorbereitet werden können. Diejenigen, die im TIHK die erforderliche Punktzahl nicht erreichen, gelten als interkulturell nicht sensibel genug und deswegen auch nicht als förderungswürdig.

7.3 Die Praktikabilität des Tests

Der TIHK, wie er in dieser Arbeit vorliegt, liefert ein praktikables Messinstrument zur Erfassung des interkulturellen Perspektivenwechsels als Indikator für interkulturelle Handlungskompetenz. Er zeichnet sich durch seine einfache Handhabung aus, die mit einem nur geringen technischen und ökonomischen Aufwand verbunden ist. Der Zeitaufwand pro Testform beträgt je nach Proband 10 bis 20 Minuten.

Darüber hinaus bietet der TIHK einen Auswertungsmodus der Ergebnisse, der wissenschaftlich fundiert und der kriteriumsorientierten Messung entnommen ist. Mit Hilfe eines kritischen Cut-off-Wertes kann eine Aussage darüber getroffen werden, ob die zu testende Person eine Disposition zu interkultureller Handlungskompetenz und damit zu kulturadäquatem Handeln besitzt oder nicht. Es wird eine eindeutige Interpretation der Ergebnisse der Probanden gewährleistet, was im Rahmen des Zusatzstudiums „Internationale Handlungskompetenz“ notwendig ist. Der Cut-off-Wert des TIHK kann je nach Zielsetzung vom Anwender höher oder niedriger angesetzt werden, was den Test flexibel und individuell einsetzbar macht. Es empfiehlt sich, die Instruktion zum Ausfüllen des Tests zusätzlich zur schriftlichen Form, mündlich vorzunehmen, da bei der vorliegenden Untersuchung an der Analytestichprobe immer wieder Schwierigkeiten hinsichtlich der Bildung einer Rangreihe der Antwortalternativen (jede Skalenausprägung darf pro Aufgabe nur einmal vergeben werden) aufgetreten sind. In diesem Zusammenhang ist jedoch auf die Standardisierung der mündlichen Testinstruktion im Sinne der Durchführungsobjektivität zu achten.

7.4 Ausblick auf weiterführende Forschung

Um das Konstrukt interkulturelle Handlungskompetenz inhaltlich greifbarer zu machen, ist es sinnvoll, einzelne Komponenten genauer zu betrachten. In Zusammenhang mit der vorliegenden Arbeit könnte deshalb untersucht werden, was der Faktor Perspektivenwechsel als eine Komponente zum Konstrukt interkulturelle Handlungskompetenz beiträgt. Damit ist gemeint, zu welchen Anteilen interkulturelle Handlungskompetenz auf die Fähigkeit des Perspektivenwechsels bzw. auf andere Eigenschaften zurückgeführt werden kann. Hierüber kann das statistische Verfahren der Varianzaufklärung Aufschluss geben. Der TIHK kann auf diese Weise einen großen Beitrag zur Charakterisierung des Konstrukts interkultureller Handlungskompetenz leisten.

Ein zweiter Aspekt der mit dem TIHK überprüft werden könnte, bezieht sich auf den Entwicklungsprozess von interkultureller Handlungskompetenz. Anhand von Verlaufsuntersuchungen könnte untersucht werden, wie sich das Konstrukt interkulturelle Handlungskompetenz im Laufe der Zeit verändert. Hier gäbe es z.B. unter anderem die Möglichkeit, interkulturelle Handlungskompetenz mittels des TIHK vor und nach einem Auslandsaufenthalt zu erfassen, um somit Veränderungen im Konstrukt zu identifizieren.

Neben weiterführenden Untersuchungen in Bezug auf das Konstrukt interkultureller Handlungskompetenz, könnten andere Forschungsaktivitäten auch bei der Weiterentwicklung des TIHK ansetzen. In diesem Zusammenhang wäre es durchaus denkbar, einen Attributionstests zu konstruieren, der kulturspezifische Subskalen aufweist. Damit ist gemeint, dass dieser Test zwar kulturallgemein aufgebaut wird, seine Subskalen allerdings kulturspezifisch sind. Konkret könnte er z.B. aus insgesamt 24 Aufgaben bestehen, wobei 8 Aufgaben jeweils einer bestimmten Kultur zugeordnet werden können. Die Subskalen sind so zu konstruieren, dass sie auch getrennt voneinander ausgewertet und untereinander verglichen werden können. Auf diese Weise kann eine besondere Disposition der getesteten Person für einen bestimmten Kulturraum diagnostiziert werden. Die Feststellung der Neigung für eine bestimmte Kultur ist von besonderer Bedeutung, denn nicht jeder Mensch ist für jede Kultur gleich gut geeignet, wie das folgende Zitat verdeutlicht.

„An individual who is extremely ambitious and task-oriented, an aggressive problemsolver, who lives out a „time-is-money“ philosophy - not infrequently the profile of success in North America - is arguably a person with a profile for failure in many cultures, particularly in the Third-World, where relationships and tradition are priorities“ (RUBEN & KEALEY, 1979).

Mit Hilfe eines derart konzipierten Tests könnte man demnach also zusätzlich die Sensibilität für eine spezifische Kultur erfassen und weitere wichtige Informationen über die zu bewertende Person gewinnen.

Um den TIHK einer größeren Zielgruppe zugänglich zu machen, ist es erforderlich, den TIHK an anderen Stichproben wie z.B. Fach- und Führungskräften eines Wirtschaftsunternehmens, zu erproben. Erst dann ist die Übertragbarkeit des TIHK auf andere Personengruppen gerechtfertigt. In diesem Zusammenhang ist auch eine Testeichung vorzunehmen, die unerlässlich ist, wenn der Test einen weiten Geltungsbereich in Anspruch nimmt und Gruppen-Normen vorgesehen sind (LIENERT & RAATZ, 1994).

Die generelle Erweiterung des Gültigkeitsbereichs des TIHK wäre in diesem Zusammenhang wünschenswert, da man dadurch den Anforderungen interkultureller Evaluation und Diagnostik besser gerecht werden kann.

7.5 Anwendungsmöglichkeiten in der Praxis

Ein Anliegen der vorliegenden Arbeit war es, zu Ergebnissen zu kommen, die nicht nur für die interkulturelle Forschung, sondern auch für die Praxis, nämlich die **Evaluation** des Zusatzstudienganges „Internationale Handlungskompetenz“ an der Universität Regensburg von Interesse sind. Im Rahmen des Zusatzstudiums wird von den Studierenden Testform A zu Beginn des ersten Semesters und Testform B am Ende des zweiten Semesters ausgefüllt. Auf der einen Seite erfährt der Studierende so, welche Fortschritte er im Laufe des Studiums gemacht hat. Auf der anderen Seite gibt der Test dem Dozenten Aufschluss darüber, wie gut die Lehre innerhalb des Studiums ist. Allgemein formuliert, ist es mit Hilfe des TIHK möglich, Personen zu 2 verschiedenen Messzeitpunkten zu testen, da der TIHK aus 2 parallelen Testformen besteht. Auf diese Weise können Veränderungen bei Personen erfassbar gemacht werden. Das hat den entscheidenden Vorteil den gleichen Test nicht zweimal hintereinander einsetzen zu müssen und damit Lerneffekte kontrollieren zu können. Dadurch ergibt sich zusätzlich auch ein Nutzen für die getestete Person: Anhand des Gesamtpunktwerts in den Testformen A und B kann sie die erreichten Punktwerte ablesen und daran persönliche Lernfortschritte erkennen.

Darüber hinaus können die Paralleltestformen des TIHK zu Evaluationszwecken auch in anderen Lernkontexten, wie z.B. interkulturellen Trainings angewandt werden. Dies gilt zumindest für die Gruppe der Studenten, die z.B. an einem Interkulturellen Training auf der Basis von Culture Assimilator Trainingsmaterial teilgenommen hat. Möchte man den TIHK im Rahmen von Evaluationszwecken an einer anderen Zielgruppe anwenden, muss dieser zunächst an einer repräsentativen Analysestichprobe adaptiert werden.

Einen weiteren Anwendungsbereich des TIHK stellt die **Diagnostik** des interkulturellen Perspektivenwechsels als Voraussetzung für interkulturelle Handlungskompetenz und interkulturelles Lernen dar. Mit gewissen Einschränkungen kann man mit Hilfe des entwickelten Testverfahrens dem Wunsch vieler Praktiker in Wirtschaft, Politik und Wissenschaft entsprechen, ein Instrument zu Potenzialanalyse und Personalauswahl zu erhalten. Für den Bereich der Beurteilung interkultureller Handlungskompetenz von Personen bedeutet dies, dass der

TIHK die Personen identifiziert, die aufgrund ihrer Fähigkeit zum Perspektivenwechsel grundsätzlich Potenzial für interkulturell kompetentes Handeln besitzen und demnach förderungswürdig sind. Da der TIHK auf eine bestimmte Zielgruppe, nämlich Studenten, zugeschnitten ist, ist analog zum Einsatz zu Evaluationszwecken eine Adaption des Tests (z. B. für Fach- und Führungskräfte) unerlässlich.

Es sei darauf hingewiesen, dass der TIHK nur einen, wenn auch bedeutenden Teilaspekt interkultureller Handlungskompetenz erfasst und somit einen Beitrag zur Diagnose von Personen leisten kann. Es wird empfohlen, den TIHK als ein Modul in Kombination mit anderen Analyseinstrumenten einzusetzen, um über eine größere Datenbasis zu verfügen und damit einen aussagekräftigen Eindruck der zu bewertenden Person gewinnen zu können. Da neben der Fähigkeit zum Perspektivenwechsel auch Persönlichkeitseigenschaften wie z.B. Offenheit oder Ambiguitätstoleranz für interkulturelle Handlungskompetenz eine gewisse Rolle spielen (vgl. Kap. 3.2.3), ist es sinnvoll, diese Aspekte zusätzlich mittels Persönlichkeitstests zu erfassen.

8 Zusammenfassung

In der vorliegenden Arbeit wurde im Rahmen der Evaluation des Zusatzstudiums „Internationale Handlungskompetenz“ an der Universität Regensburg ein Test zur Erfassung interkultureller Handlungskompetenz (TIHK) mit einer Parallelförmigkeit konstruiert. Der TIHK hatte zum Ziel, den Lernzuwachs der Studierenden im Zusatzstudiengang zu erfassen und Aufschluss über die Güte des Studienganges zu ermöglichen.

Als Indikator für interkulturelle Handlungskompetenz diente die Fähigkeit zum interkulturellen Perspektivenwechsel, die bei der Betrachtung verschiedener Untersuchungen zu interkultureller Handlungskompetenz als zentrale Voraussetzung für interkulturellen Handlungserfolg erachtet wird. Der Grundgedanke der isomorphen Attributionen von TRIANDIS (1975, 1977) bietet sich in diesem Zusammenhang als geeignetes Konzept zur Erfassung des Perspektivenwechsels an. Eine Methode, kulturisomorphe Attributionen zu fördern, stellt der Culture Assimilator dar. Aus diesem Grund wurden in der vorliegenden Arbeit kritische Interaktionssituationen der Culture Assimilator USA (MÜLLER & THOMAS, 1995), England (SCHMID, 1999) und Indonesien (MARTIN, 2000) zur Entwicklung des TIHK und der Parallelförmigkeit ausgewählt. Demzufolge wurde die im Theorieteil hervorgehobene Fähigkeit des Perspektivenwechsels als Indikator für interkulturelle Handlungskompetenz durch isomorphes Attribuieren messbar gemacht.

Der TIHK wurde in 2 Durchgängen an insgesamt 119 Studenten aller Fachrichtungen entwickelt. Das Resultat sind 2 parallele Testformen zu je 8 Aufgaben (kritische Interaktionssituationen). Nach testtheoretischen Gütekriterien (innere Konsistenz: Test A: $c = 0,75$; Test B = $0,74$ und Paralleltest-Reliabilität: $c = 0,98$) handelt es sich hierbei um einen einwandfreien Test, der hinsichtlich der Prüfung der Äquivalenzkriterien auf Parallelität zufrieden stellende Ergebnisse erzielen konnte. Damit liefert der TIHK wie er in dieser Arbeit vorliegt, ein praktisches Messinstrument zur Erfassung des interkulturellen Perspektivenwechsels als Indikator für interkulturelle Handlungskompetenz. Er zeichnet sich durch seine einfache Handhabung aus, die mit einem nur geringen technischen und ökonomischen Aufwand verbunden ist. Mit Hilfe eines kritischen Cut-off- Wertes kann eine Aussage getroffen werden, ob die zu testende Person eine Disposition zu interkultureller Handlungskompetenz und damit zu kulturadäquaten Handeln besitzt oder nicht. Die Ergebnisse der Validierung, wonach das Antwortverhalten im Test mit dem im Rollenspiel gezeigten Verhalten der Probanden übereinstimmt, sprechen für die Gültigkeit des Instruments, so dass das Resultat des dargestellten Entwicklungsprozesses eine hinreichend reliable und valide Testform zur Erfassung interkultureller Handlungskompetenz ist.

9 Literatur

- ABE, H. & WISEMAN, R. L. (1983). A cross-cultural confirmation of the dimensions of intercultural effectiveness. *International Journal of Intercultural Relations*, 7, 53-67.
- ALBERT, R.D. (1983). The Intercultural Sensitizer or Culture Assimilator: A Cognitive Approach. In D. Landis & R. W. Brislin (Eds.), *Handbook of intercultural training, Vol. 2: Issues in training methodology* (pp. 186-217). New York: Pergamon.
- APFELTHALER, G. (1999). *Interkulturelles Management. Die Bewältigung kultureller Differenzen in der internationalen Unternehmertätigkeit*. Wien: Manz-Verlag.
- BENNETT, M. J. (1986). A developmental approach to training for intercultural sensitivity. *International Journal of Intercultural Relations*, 10, 179-195.
- BENSON, P.G. (1978). Measuring cross-cultural adjustment: The problem of criteria. *International Journal of Intercultural Relations*, 12, 21-37.
- BHAWUK, D. P. S. (1998). The role of culture theory in cross-cultural training: A multimethod study of culture-specific, culture-general, and culture theory-based assimilators. *Journal of Cross-cultural Psychology*, 29 (5), 630-655.
- BHAWUK, D. P. S. & BRISLIN, R. (1992). The measurement of intercultural sensitivity using the concepts of individualism and collectivism. *International Journal of Intercultural Relations*, 16, 413-436.
- BIRD, A., HEINBUCH, S., DUNBAR, R. & MCNULTY, M. (1993). A Conceptual Model of the Effects of Area Studies Training Programs and a Preliminary Investigation of the Model's Hypothesized Relationships. *International Journal of Intercultural Relations*, 17, 415-435.
- BOESCH, E.E. (1980). *Kultur und Handlung - Eine Einführung in die Kulturpsychologie*. Bern: Huber.
- BLOOM, B.J. (Hrsg.). (1972). *Taxonomie von Lernzielen im kognitiven Bereich*. Weinheim: Beltz.
- BOLTEN, J. (2001). *Interkulturelle Kompetenz*. Erfurt: Landeszentrale für politische Bildung Thüringen.
- BORTZ, J. (1984). *Lehrbuch der empirischen Forschung. Für Sozialwissenschaftler*. Berlin: Springer.
- BORTZ, J. & DÖRING, N. (1995). *Forschungsmethoden und Evaluation*. Berlin: Springer.
- BREITENBACH, D. (1975). Interkulturelles Lernen und internationale Verständigung in der internationalen Jugendarbeit: Eine theoretische Einführung. In D. Breitenbach (Ed.), *Kommunikationsbarrieren in der internationalen Jugendarbeit*, Bd. 1, (S. 9-21). Saarbrücken: Breitenbach.
- BRISLIN, R. W. (1981). *Cross-Cultural encounters: Face-to-face interaction*. New York: Pergamon.

- BRISLIN, R. W. (1989). Intercultural communication training. In M. Asante & W. Gudykunst (Hrsg.), *Handbook of international and intercultural communication*, (pp. 441-457). Newsbury Park, CA: Sage Publications.
- BRISLIN, R. W. & PEDERSON, P. (1976). *Cross-cultural orientation programs*. New York: Gardner Press.
- BRISLIN, R. W., LANDIS, D. BRANDT, M. E. (1983). Conceptualizations of Intercultural Behavior and Training. In D. Landis & R. W. Brislin (Eds.), *Handbook of intercultural training, Vol. 1: Issues in theory and design* (S. 1-35). New York: Pergamon.
- BRISLIN, R. W., CUSHNER, K., CHERRIE, C. & YONG, M. (1986). *Intercultural interactions. A practical guide*. Beverly Hills: Sage.
- CHURCH, A. T. (1982). Sojourner adjustment. *Psychological Bulletin*, 91 (3), 540-572.
- CRONBACH, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- CUI, G. & AWA, N. E. (1992). Measuring intercultural effectiveness. An integrative approach. *International Journal of Intercultural Relations*, 16, 311-328.
- CUSHNER, K. (1989). Assessing the Impact of a Culture-General Assimilator. *International Journal of Intercultural Relations*, 13, 125-146.
- DADDER, R. (1987). *Interkulturelle Orientierung - Analyse ausgewählter interkultureller Trainingsprogramme*. Saarbrücken: Breitenbach.
- DEAN, O. & POPP, G. E. (1990). Intercultural Communication Effectiveness as Perceived by American Managers in Saudi Arabia and French Managers in the U.S. *International Journal of Intercultural Relations*, 14, 405-424.
- DELKESKAMP, J. (1991). Die Simulation als Medium zum interkulturellen Lernen. In R. Nestvogel (Hrsg.), *Interkulturelles Lernen oder verdeckte Dominanz*. (S. 139-151) Frankfurt/Main: Verlag Interkulturelle Kommunikation.
- DELLER, J. (1996). Interkulturelle Eignungsdiagnostik. In A. Thomas (Hrsg.), *Psychologie interkulturellen Handelns*. (S. 283-316). Göttingen: Hogrefe.
- DELLER, J. (2000). Interkulturelle Eignungsdiagnostik. Zur Verwendbarkeit von Persönlichkeitsskalen. Waldsteinberg: Verlag Heidrun Popp.
- DETWEILER, R. A. (1978). Culture category-width, and the attributions: A model-building approach to the reasons for cultural effects. *Journal of Cross-Cultural Psychology*, 9 (3), 259-284.
- DETWEILER, R. A., BRISLIN, R. W. & MCCORMACK, W. (1983). Situational Analysis. In D. Landis & R. W. Brislin (Eds.), *Handbook of intercultural training, Vol. 2: Issues in training methodology* (pp. 100-123). New York: Pergamon.
- DINGES, N. G. & BALDWIN, K. D. (1996). Intercultural competence. A research perspective. In D. Landis & R.S. Bhagat (Eds.), *Handbook of intercultural training* (2nd ed.) (pp. 106-123). Thousand Oaks, CA: Sage Publications.

- DINGES, N. & LIEBERMANN, D.A. (1989). Intercultural communication competence: Coping with stressful work situations. *International Journal of Intercultural Relations*, 13, (3), 371-385.
- EDER, G. (1996). „Soziale Handlungskompetenz“ als Bedingung und Wirkung interkultureller Begegnungen. In A. Thomas (Hrsg.), *Psychologie interkulturellen Lernens* (S. 411-422). Göttingen: Hogrefe.
- FAHRMEIR, L., KÜNSTLER, R., PIGEOT & I. TUTZ, G. (1997). *Statistik. Der Weg zur Datenanalyse*. Heidelberg: Springer.
- FIEDLER, F.E., MITCHELL & T., TRIANDIS, H.C. (1971). The Culture Assimilator: An approach to cross-cultural training. *Journal of Applied Psychology*, 55 (2), 95-102.
- FISSENI, H.-J. (1997). *Lehrbuch der psychologischen Diagnostik. Mit Hinweisen zur Intervention*. 2. überarb. und erw. Auflage. Göttingen: Hogrefe Verlag.
- FLANAGAN, J. C. (1954). The critical incident technique. *Psychological Bulletin*, 51, 327-358.
- FRICKE, R. (1972). Lehrzielorientierte Messung mit Hilfe stochastischer Messmodelle. In K.J. Klauer, R. Fricke, M. Herbig, H. Rupprecht & F. Schott (Hrsg.), *Lehrzielorientierte Tests. Beiträge zur Theorie, Konstruktion und Anwendung*. Düsseldorf: Schwann.
- FRICKE, R. (1974). Kriteriumsorientierte Leistungsmessung. Berlin: Verlag W. Kohlhammer.
- FURNHAM, A. & BOCHNER, S. (1986). *Culture shock. Psychological reactions to unfamiliar environments*. London: Methuen.
- GARDNER, G.H. (1962). Cross-Cultural communication. *The Journal of Social Psychology*, 58, 241-256.
- GEISINGER, K.F. (1994). Cross-cultural Normative Assessment: Translation and Adaptation Issues Influencing the Normative Interpretation of Assessment Instruments. *Psychological Assessment*, 6 (4), 304-312.
- GÖTZ, S. (2003). *Validierung eines Tests zur Erfassung interkultureller Handlungskompetenz (TIHK)*. Unveröffentlichte Diplomarbeit. Universität Regensburg.
- GUILFORD, J. P. (1954). *Psychometric methods*. (2nd ed.) New York: Mc Graw Hill.
- GUDYKUNST, W. B., GUZLEY, R. M. & HAMMER, M. R. (1996). Designing intercultural training. In D. Landis & R. S. Bhagat (Hrsg.), *Handbook of intercultural training* (2nd Ed.), (pp. 61-80). Thousand Oaks, CA: Sage Publications.
- GUDYKUNST, W.B. & HAMMER, M.R. (1983). Basic Training Design: Approaches to Intercultural Training. In D. Landis & R.W. Brislin (Eds.), *Handbook of intercultural training. Issues in Theory and Design, Vol. 1* (pp. 118-154). New York: Pergamon.
- HAMMER, M. R. (1987). Behavioural dimensions of intercultural effectiveness: A replication and extension. *International Journal of Intercultural Relations*, 11 (1), 65-88.
- HAMMER, M. R. (1989). Intercultural communication competence. In M. K. Asante & W. B. Gudykunst (Eds.), *Handbook of international and intercultural communication* (pp. 247-260). Newsbury Park, CA: Sage Publications.

- HAMMER, M. R., GUDYKUNST W. & WISEMAN, R. (1978). Dimensions of intercultural effectiveness: An exploratory study. *International Journal of Intercultural Relations*, 2, 382-393.
- HANNIGAN, T.P. (1990). Traits, attitudes, and skills that are related to intercultural effectiveness and their implications for cross-cultural training: a review of the literature. *International Journal of Intercultural Relations*, 14, 89-111.
- HARRIS, J. G. (1975). Identification of cross-cultural talent: the empirical approach of the Peace Corps. *Topics in Culture Learning*, 3, 66-78.
- HATZER, B. (2000). *Erfassung interkultureller Handlungskompetenz unter Berücksichtigung der lernpsychologischen Basisannahmen des Situated Learning*. Unveröffentlichte Diplomarbeit. Universität Regensburg.
- HELFRICH, H. (1996). Kulturstandards und individuelle Varianten. In A. Thomas (Hrsg.), *Psychologie interkulturellen Handelns* (S.199-208). Göttingen: Hogrefe.
- HENTSCHEL, S. (2001). Quantitative Evaluation des multimedialen Lernprogramms „Teamwork Across Cultures“ zur Sensibilisierung auf die Zusammenarbeit in internationalen Arbeitsgruppen. Unveröffentlichte Diplomarbeit. Universität Regensburg.
- HERBIG, M. (1972). Die Unzulänglichkeit der klassischen Testtheorie bei lehrzielorientierter Leistungsmessung. In K.J. Klauer, R. Fricke, M. Herbig, H. Rupperecht & F. Schott (Hrsg.), *Lehrzielorientierte Tests. Beiträge zur Theorie, Konstruktion und Anwendung*. (S. 117-123). Düsseldorf: Schwann.
- HERBIG, M. (1974). Ein lehrzielorientiertes Zensierungsmodell. *Zeitschrift für erziehungswissenschaftliche Forschung*, 8, S. 129-142.
- HERBRAND, F. (2000). *Interkulturelle Kompetenz: Wettbewerbsvorteil in einer globalisierenden Wirtschaft*. Bern: Haupt.
- HINZ-ROMMEL, (1994). *Interkulturelle Kompetenz. Ein neues Anforderungsprofil für die soziale Arbeit*. Münster: Waxmann.
- HOFSTEDE, G. (1980). *Culture's consequence: International differences in work related values*. Beverly Hills, CA: Sage Publications.
- HOOPES, D.S. (1981). Intercultural Communication Concepts and the Psychology of Intercultural Experience. In M.D. Pusch (Ed.), *Multicultural Education: A cross-cultural training approach*, (pp. 9-38). Chicago: Intercultural Press.
- HORST, P. (1971). *Messung und Vorhersage*. Weinheim: Beltz.
- JONES, E. E. & NISBETT, R. E. (1971). The Actor and the Observer: Divergent Perceptions of the causes of behavior. New York: General Learning Press.
- KAMMHUBER, S. (2000). *Interkulturelles Lernen und Lehren*. Inaugural-Dissertation. Universität Regensburg.
- KEALEY, D. J. (1989). A study of cross-cultural effectiveness: Theoretical issues, practical applications. *International Journal of Intercultural Relations*, 13 (3), 387-428.

- KEALEY, D. J. (1996). The Challenge of International Personnel selection. In D. Landis & R. S. Bhagat, (Eds.), *Handbook of intercultural training* (2nd Ed.), (pp. 81-105). Thousand Oaks, CA: Sage Publications.
- KEALEY, D. J. & RUBEN, B. D. (1979). Behavioral Assessment of Communication Competency and the Prediction of Cross-Cultural Adaptation. *International Journal of Intercultural Relations*, 3, 15-47.
- KEALEY, D. J. & RUBEN, B. D. (1983). Cross-Cultural Personnel Selection, Criteria, Issues, and Methods. In D. Landis & R. W. Brislin (Eds.), *Handbook of intercultural training, Vol. 1: Issues in theory and design* (pp. 155-175). New York: Pergamon.
- KELLY, C. & MEYERS, J.E. (1992). The cross-cultural adaptability inventory manual. In S. Fowler & M. Fowler (Eds.), *The intercultural sourcebook: Vol. 2*. Yarmouth, ME: Intercultural Press.
- KIM, Y. Y. (1991). Intercultural communication competence. In S. Ting-Toomey & F. Korzeny (Eds.), *Cross-cultural interpersonal communication* (pp. 259-275). Newberry Park, CA: Sage.
- KIM, Y. Y. & RUBEN, B. D. (1988). Intercultural Transformation. In Y. Y. Kim & W. B. Gudykunst (Eds.), *Theories in intercultural communication* (S. 299-321). London: Sage.
- KINAST, E.-U. (1998). *Evaluation interkultureller Trainings*. Lengerich: Pabst Science Publishers.
- KLAUER, K. J. (1972). Zur Theorie und Praxis des binomialen Modells lehrzielorientierter Tests. In K.J. Klauer, R. Fricke, M. Herbig, H. Rupperecht & F. Schott (Hrsg.), *Lehrzielorientierte Tests. Beiträge zur Theorie, Konstruktion und Anwendung*. Düsseldorf: Schwann.
- KRATHWOHL, D. R., BLOOM, B. S. & MASIA, B. B. (1975). *Taxonomie von Lernzielen im affektiven Bereich*. Weinheim: Beltz.
- KROEBER, A. & KLUCKHOHN, C. (1952). *Culture: A critical review of concepts and definitions*. Cambridge, MA: Addison-Wesley.
- KRUMING, M. (2000). *Anwendungsorientierte Analyse einer Methode zur Erfassung sozialer Perspektivenübernahme im interkulturellen Handlungsfeld*. Unveröffentlichte Diplomarbeit. Universität Regensburg.
- KÜHLMANN, T. M. (1995). Die Auslandsentsendung von Fach- und Führungskräften: Eine Einführung in die Schwerpunkte und Ergebnisse der Forschung. In T. M. Kühlmann (Hrsg.), *Mitarbeiterentsendung ins Ausland: Auswahl, Vorbereitung, Betreuung und Wiedereingliederung* (S. 1-30). Göttingen: Verlag für Angewandte Psychologie.
- KÜHLMANN, T. & STAHL, G. (1998). Diagnose interkultureller Kompetenz: Entwicklung und Evaluierung eines Assessment Centers. In C. I. Barmeyer, & J. Bolten (Hrsg.), *Interkulturelle Personalorganisation* (S. 213-224). Berlin: Verlag Wissenschaft und Praxis.
- LANDIS, D. & BHAGAT, R. S. (1996). A model of intercultural behavior and training. In D. Landis & R. S. Bhagat (Hrsg.), *Handbook of intercultural training* (2nd Ed.), (pp. 1-13). Thousand Oaks, CA: Sage Publications.

- LANDIS, D., DAY, H. R., MCGREW, P. L., THOMAS, J. A. & MILLER, A. B. (1976). Can a black „culture assimilator“ increase racial understanding? *Journal of Social Issues*, 32, 169-183.
- LANGE, C. (1994). *Interkulturelle Orientierung am Beispiel der Trainingsmethode „Cultural Assimilator“*. Beiträge zur interkulturellen Didaktik, Bd. 3. Göttingen: Zentrum für didaktische Studien e.V.
- LAYES, G. (1995). *Quantitative Evaluation eines interkulturellen Managementtrainings für deutsche Manager zur Vorbereitung auf die Kooperation mit Chinesen*. Unveröffentlichte Diplomarbeit. Universität Regensburg.
- LOO, R. & SHIOMI, K. (1999). A Structural and Cross-Cultural Evaluation of the Inventory of Cross-Cultural Sensitivity. *Journal of Social Behavior and Personality*, 14 (2), 267-278.
- LORD, F. M. & NOVICK, M. R. (1968). *Statistical theories of mental test scores*. Reading, Mass.: Addison-Wesley.
- LIENERT, G. & RAATZ, U. (1994). *Testaufbau und Testanalyse*. 5., erw. Auflage. Weinheim: Beltz Psychologie Verlags Union.
- MALPASS, R. S. & SALANCIK, G. R. (1977). Linear and branching formats in culture assimilator training. *International Journal of Intercultural Relations*, 1, 76-87.
- MARTIN, M. (2000). Entwicklung eines Culture Assimilator-Trainings für die Zusammenarbeit mit Indonesiern, Bd. 2. Unveröffentlichte Diplomarbeit. Universität Erlangen-Nürnberg.
- MESSICK, R. J. (1988). Validity. In Linn. R. L., *Educational Measurement*. (3rd ed.). New York: Am. Counc. Educ./Macmillan.
- MÜLLER, A. (1992). *Psychologische Voraussetzungen und Trainingsmaßnahmen für erfolgreiches interkulturelles Handeln - analysiert am Beispiel der Zielgruppe "deutsche Studenten, die ein Studienjahr in den USA verbringen"*. Unveröffentlichte Diplomarbeit. Universität Regensburg.
- MÜLLER, A. & THOMAS, A. (1991). *Interkulturelles Orientierungstraining für die USA. Übungsmaterial zur Vorbereitung auf ein Studium in den Vereinigten Staaten*. SSIP-Bulletin Nr. 62. Saarbrücken: Breitenbach.
- MÜLLER, A. & THOMAS, A. (1995). *Studienhalber in den USA – Interkulturelles Orientierungstraining für deutsche Studenten, Schüler und Praktikanten*. Heidelberg: Asanger.
- MUMMENDEY, H. D. (1995). *Die Fragebogen-Methode. Grundlagen und Anwendung in Persönlichkeits-, Einstellungs- und Selbstkonzeptforschung*. Göttingen: Hogrefe Verlag.
- OBERG, K. (1960). Cultural Shock: Adjustment to New Cultural Environments. *Practical Anthoropology*, 7, 177-182.
- POPHAM W. J. & HUSSEK T. R. (1969). Implications of Criterion-referenced Measurement. *Journal of educational Measurement*, 6, 1-9. Übersetzung in: P. Strittmatter (Ed.), 1973
- PRUEGGER, V.J. & ROGERS, T.B. (1993). Brief report: Development of a Scale to measure Cross-Cultural Sensitivity in the Canadian Context. *Canadian Journal of Behavioural Science*, 25 (4), 615-621.

- RANDOLPH, G., LANDIS, D. & TZENG, O. C. S. (1977). The Effects of Time and Practice upon Culture Assimilator Training. *International Journal of Intercultural Relations*, 1 (4), 105-119.
- ROGGE, K.-E. (1995). *Methodenatlas*. Berlin: Springer Verlag.
- ROHRMANN, B. (1978). Empirische Studien zur Entwicklung von Antwortskalen für die sozialwissenschaftliche Forschung. *Zeitschrift für Sozialpsychologie*, 9, 222-245.
- RUBEN, B. D. (1976). Assessing communication competency for intercultural adaptation. *Group and Organization Studies*, 1, 334-354.
- RUBEN, B. D. (1989). The study of cross-cultural competence: Traditions and Contemporary Issues. *International Journal of Intercultural Relations*, 13, 229-240.
- SANDHAAS, B. (1988). Interkulturelles Lernen - Zur Grundlegung eines didaktischen Prinzips interkultureller Begegnungen. *International Review of Education*, 34 (4), 415-438.
- SCHMICK, M. (2001). *Entwicklung eines Analyseinstrumentes zur Erfassung der individuellen Teamfähigkeit von Mitgliedern einer Arbeitsgruppe*. Unveröffentlichte Diplomarbeit. Universität Regensburg.
- SCHMID, H. (1992). *Psychologische Test. Theorie und Konstruktion*. Göttingen: Verlag Huber.
- SCHMID, S. (1999). *Entwicklung eines Culture Assimilator-Trainings für die Zusammenarbeit mit Engländern*, Bd. 2. Unveröffentlichte Diplomarbeit. Universität Regensburg.
- SIMON, P., THEIS, F. & METZNER, C. (in Vorbereitung). *Der c-Koeffizient als Ersatz für den Korrelationskoeffizienten der klassischen Testtheorie*.
- SPERBER, A. D., DEVELLIS, R. F. & BOEHLECKE, B. (1994). Cross-Cultural Translation: Methodology and Validation. *Journal of Cross-Cultural Psychology*, 25 (4), 501-524.
- Stahl, G.K. (1998). *Internationaler Einsatz von Führungskräften*. München: Oldenbourg.
- TAYLOR, E. (1994). A learning model for becoming interculturally competent. *International Journal for Intercultural Relations*, 16, 389-408.
- THOMAS, A. (Hrsg.) (1993). *Kulturvergleichende Psychologie. Eine Einführung*. Göttingen: Hogrefe.
- THOMAS, A. (1995). Die Vorbereitung von Mitarbeitern für den Auslandseinsatz: Wissenschaftliche Grundlagen. In T. M. Kühlmann (Hrsg.), *Mitarbeiterentsendung ins Ausland: Auswahl, Vorbereitung, Betreuung und Wiedereingliederung* (S. 85-115). Göttingen: Verlag für angewandte Psychologie.
- THOMAS, A. (1996). Analyse der Handlungswirksamkeit von Kulturstandards. In A. Thomas (Hrsg.), *Psychologie interkulturellen Handelns* (S. 107-135). Göttingen: Hogrefe.
- THOMAS, A. (in Druck). Interkulturelle Kompetenz: Grundlagen, Probleme und Konzepte. *Erwägen Wissen Ethik*, 14 (1).
- THOMAS, A., KAMMhuber, S. & LAYES, G. (1997). Interkulturelle Kompetenz. Ein Handbuch für internationale Einsätze der Bundeswehr. *Untersuchungen des Psychologischen Dienstes der Bundeswehr*, 32, 7-190.

- THOMAS, A. & SCHENK, E. (1996). *Handlungswirksamkeit zentraler Kulturstandards in der Interaktion zwischen Deutschen und Chinesen*. Unveröff. Abschlussbericht zum Forschungsprojekt, gefördert von der Volkswagenstiftung unter Aktenzeichen II/673621 im Rahmen des Schwerpunktes "China-Programm". Institut für Psychologie. Abteilung Sozialpsychologie. Universität Regensburg.
- TODD IMAHORI, T. & LANIGAN, M. L. (1989). Relational Model of Intercultural Communication Competence. *International Journal of Intercultural Relations*, 13, 269-286.
- TRIANDIS, H. C. (1972). *The analysis of subjective culture*. New York: Wiley.
- TRIANDIS, H. C. (1975). Culture Training, Cognitive Complexity and Interpersonal Attitudes. In R. W. Brislin, S. Bochner & W. J. Lonner (Eds.), *Cross-Cultural Perspectives on Learning* (pp. 39-77). New York: John Wiley & Sons.
- TRIANDIS, H. C. (1977). *Interpersonal behaviour*. Monterey, CA: Brooks/Cole.
- TRIANDIS, H.C. (1989). Intercultural Education and Training. In P. Funke (Ed.), *Understanding the USA: A cross-cultural perspective*, (pp. 305-322). Tübingen: Narr.
- TRIANDIS, H. C. (1994). *Culture and social behaviour*. New York: McGraw-Hill.
- TRIANDIS, H.C., CHEN, X.P. & CHAN, D.K.-S. (1998). Scenarios for the Measurement of Collectivism and Individualism. *Journal of Cross-Cultural Psychology*, 29 (2), 275-289.
- TUCKER, M. (1974). *Screening and selection for overseas assignment: assessment and recommendations to U.S. Navy*. Denver, Colorado: Centre for Research and Education.
- WELDON, D.E., CARSTON, C.E., RISSMAN, A.K., SLOBODIN, L. & TRIANDIS, H.C. (1975). A Laboratory Test of Effects of Culture Assimilator Training. *Journal of Personality and Social Psychology*, 32, 300-310.
- WINTER, G. (1988). Konzepte und Stadien interkulturellen Lernens. In A. Thomas (Ed.), *Interkulturelles Lernen im Schüleraustausch*, SSIP-Bulletin Nr.58 (S.151-178). Saarbrücken: Breitenbach.
- WINTER, G. (1994). Trainingskonzepte auf dem Prüfstand: Theoriebezug, Ethik, Evaluation. In Institut für Auslandsbeziehungen (Hrsg.), *Interkulturelle Kommunikation und Interkulturelles Training. Problemanalysen und Problemlösungen* (S. 42-54). Stuttgart: Institut für Auslandsbeziehungen.

10 Anhang

Verzeichnis der Anhänge

Anhang A: TIHK – Test A

Anhang B: TIHK – Test B

Anhang C: Lösungen Test A und Test B

Erklärung

Ich versichere hiermit, dass ich die anliegende Arbeit mit dem Thema:

„Entwicklung eines Tests zur Erfassung interkultureller Handlungskompetenz (TIHK)“

selbständig verfasst und keine anderen Hilfsmittel als die angegebenen benutzt habe. Die Stellen, die anderen Werken dem Wortlaut oder dem Sinn nach entnommen sind, habe ich in jedem einzelnen Fall durch Angabe der Quelle, auch der benutzten Sekundärliteratur, als Entlehnung kenntlich gemacht.

Regensburg, den 03. Februar 2003

Julia Loboda