

3 Empirische Untersuchung

In den vorangegangenen Kapiteln wurde der bisherige Forschungsstand der relevanten Theorien und wissenschaftlichen Themen dargelegt. Im Folgenden wird nun das Forschungsvorhaben konkret vorgestellt, die Operationalisierung dargelegt und die Hypothesen erläutert. Weiters werden die methodischen Grundlagen zu den empirischen Methoden angeführt, ein Überblick über die Stichprobe gegeben, auf einen Pretest des Autors eingegangen, sowie Aufbau und Ablauf des Experiments beschrieben.

3.1 Methodik

3.1.1 Forschungsdesign und Operationalisierung

Wie bereits eingangs verdeutlicht, möchte der Autor dieser Dissertation eine Forschungslücke schließen, indem konkret die Führungsfähigkeit und -eigenschaften von Nachwuchsführungskräften, also BerufseinsteigerInnen und Personen mit wenig Berufserfahrung, in homogenen und heterogenen Settings empirisch erforscht werden sollen. Als geeignete Methoden zur Beantwortung der forschungsrelevanten Fragen und Hypothesen (siehe Kapitel 3.1.2) und Datenerhebung (siehe dazu konkret Kapitel 3.2) wurden zwei Formen gewählt: einerseits wurde ein *Experiment* konstruiert, das im Mai 2011 an der Wirtschaftsuniversität Wien durchgeführt wurde. Dabei wurden 80 Studierende aus 4 unterschiedlichen Nationen / Kulturen (Österreich, Türkei, China und USA) eingeladen, an der wissenschaftlichen Untersuchung im Rahmen dieser Dissertation teilzunehmen (siehe Kapitel 3.3.3). Andererseits bedient sich der Verfasser zur Messung der Methodik der quantitativen Sozialforschung. Dazu wurde ein *Fragbogen* mit geschlossen Antwortkategorien entwickelt, der sowohl Fragen zur Führung als auch zur Gruppenarbeit beinhaltet. Die Items dafür wurden literaturgestützt, auf den Ebenen „Leader perception“ und „Group perception“ generiert.

3.1.2 Forschungsrelevante Fragen und Hypothesen

Ziel dieser Dissertation ist es, folgende Hauptforschungsfrage aus den Erkenntnissen dieser empirischen Untersuchung, unter Rückbindung an die Theorie, zu beantworten:

F1: „Welchen Einfluss haben die Nationalkulturen von Nachwuchsführungskräften und Geführten homogener und multikultureller Gruppen auf die Wahrnehmung dieser Geführten in Bezug auf die Führungsfähigkeit von Nachwuchsführungskräften?“

Ebenso wird folgende Unterfrage behandelt:

F1.1: „Inwieweit unterscheiden sich die Wahrnehmungen und Attributionen von Gruppenmitgliedern homogener und multikultureller Gruppen in Bezug auf die Arbeit in der Gruppe?“

Basierend auf den bisherigen Reflexionen und Überlegungen, können zentrale Hypothesen¹³ abgeleitet werden, die einer empirische Prüfung unterzogen werden. In der Darstellung der Theoriekonzepte und -modelle wurde verdeutlicht, dass Geführte implizite, kognitive Führungskraftprototypen aus ihren individuellen Erfahrungen entwickeln, die sie mit der jeweiligen Führung ihrer aktuellen Führungskraft vergleichen (Schyns / Schilling, 2011, S. 142f.; Epitropaki / Martin, 2004, S. 293f.; Chong / Thomas, 1997, S. 278; Offermann / Kennedy / Wirtz, 1994, S. 44f.). Desweiteren konnte festgehalten werden, dass ein wesentlicher Einfluss der Nationalkultur einer geführten Person auf ihre eigenen individuellen prototypischen Vorstellungen existiert (Yancey / Watanabe, 2009, S. 275f.; Chong / Thomas, 1997, S. 278f.; Thomas / Ravlin, 1995, S. 143; Gerstner / Day, 1994, S. 130). Ausgangspunkt für die erste Hypothese ist daher der jeweilige kulturelle Hintergrund eines Individuums, der einen Einfluss auf seine Wahrnehmung, Einschätzung und Zuschreibung einer Führungskraft hat. Bezogen auf die gewählte implizite Führungstheorie für die empirische Untersuchung im Rahmen dieser Dissertation nach Chong / Thomas (1997), ist nicht nur die Perzeption und Empfindung von Führung durch die Geführten von ihrer eigenen

13 Hypothesen sind Behauptungen eines Kausalzusammenhangs (Atteslander, 2003, S. 198f.; Stier, 1999, S. 12). Aus der wissenschaftlichen Betrachtung heraus lassen sich vier Kriterien für Hypothesen nennen: (1) Hypothesen beziehen sich auf reale Sachverhalte, die empirisch untersuchbar sind; (2) Hypothesen stellen Behauptungen dar, die zwar allgemeingültig sind, jedoch über ein bestimmtes Ereignis oder einen Einzelfall hinausgehen; (3) Implizit muss einer Hypothese die Formalstruktur eines sinnvollen Konditionalsatzes („Wenn-dann-Satz“ bzw. „Je-desto-Satz“) zugrunde liegen; (4) Konditionalsätze müssen falsifizierbar sein, d.h. es müssen Ereignisse möglich sein, die diesen Sätzen widersprechen (Schnell / Hill / Esser, 2005, S. 53f.; Bortz / Döring, 2003, S. 8f.)

Kultur getrieben. „In addition, followers‘ perceptions may be affected by the ethnicity of the leader.” (Chong / Thomas, 1997, S. 278) Die Erwartungen an eine Führungskraft sind demnach von den individuellen Führungskraftprototypen beeinflusst, die zumeist unbewusst unter Berücksichtigung des eigenen kulturellen Hintergrunds entwickelt wurden. Gleichzeitig steuert ebenso die Nationalkultur einer Führungskraft die Erwartungen und das Verhalten von Geführten (Chong / Thomas, 1997, S. 278f.). Dabei haben wissenschaftliche Erkenntnisse gezeigt, dass bei Übereinstimmung der Nationalkultur der Führungskraft und der Geführten Vertrautheit und Sicherheit aufkommen bzw. entstehen und dies eine positive Auswirkung auf die Leader-Follower-Beziehung hat (Festing / Maletzky, 2011, S. 187; Zander / Butler, 2010, S. 258f.; Bachmann / Wolf, 2007, S. 1056f.; House / Javidan, 2004, S. 17; Thomas, 1999, S. 257).

H1: Wenn eine Nachwuchsführungskraft dieselbe Nationalkultur wie Geführte einer homogenen Gruppe hat, dann werden dieser Führungsperson positivere Attribute hinsichtlich ihrer Führungsfähigkeit zugeschrieben als Nachwuchsführungskräften aus anderen Kulturen.

Nicht nur in den FührerInnen-Geführten-Beziehungen gibt es zahlreiche Annahmen, dass der gleiche kulturelle Hintergrund positive Auswirkungen auf die Zusammenarbeit zwischen Führungskräften und MitarbeiterInnen hat. Auch auf Gruppenebene gibt es zahlreiche Meinungen, dass sich bei gleicher Nationalkultur die Wahrnehmungen und Zuschreibungen von Situationen und Ereignissen ähneln und demzufolge auch bei den prototypischen Idealvorstellungen von Führung mögliche Gemeinsamkeiten vorhanden sind (Chatman / Flynn, 2001, S. 971; Thomas, 1999, S. 257f.; Baugh / Graen, 1997, S. 378ff.; Chong / Thomas, 1997, S. 279; McLeod / Lobel / Cox, 1996, S. 257).

H2: Geführte gleicher Nationalkultur haben ähnlichere Zuschreibungen und prototypische Einstellungen hinsichtlich der Führungsfähigkeit von Nachwuchsführungskräften, als Personen unterschiedlicher kultureller Herkunft.

Neben der Wahrnehmung von Führung spielt die Messung der individuellen Zuschreibung und Einschätzung der Effektivität einer Gruppe durch die Gruppenmitglieder eine entscheidende Rolle, um die Kohäsion und die Leistungsbereitschaft zu messen (Haas, 2010, S. 989ff.; Wageman / Hackman / Lehman, 2005, S. 376; Guzzo / Dickson, 1996, 310ff.). Dabei weisen zahlreiche Studien darauf hin, dass die Vielfalt einer multikulturellen Gruppe als möglicher Wettbewerbsvorteil gesehen werden kann (Ely / Thomas, 2001, S. 242; Richard, 2000, S.

165; Cox / Blake, 1991, S. 54). In der Praxis entstehen jedoch oftmals Schwierigkeiten innerhalb der Gruppe, die den eigentlichen Vorteil der Perspektiven- und Meinungsvielfalt gar nicht zulassen, sondern die intensive Auseinandersetzung mit den Problemen innerhalb der Gruppe zur eigentlichen Herausforderung machen (Jehn / Bezrukova, 2004, S. 703f.; von Glinow / Shapiro / Brett, 2004, S. 578; Watson / Kumar / Michaelsen, 1993, S. 596ff.). Mit der folgenden Hypothese möchte der Autor empirisch prüfen, ob die Mitglieder homogener Arbeitsgruppen der Ansicht sind, dass die Zusammensetzung ihrer Gruppe für die Aufgabenstellung im Rahmen des Experiments wirksamer ist oder nicht.

H3: *Der Grad der Effektivität einer Gruppe wird von Mitgliedern homogener Gruppen höher wahrgenommen als von Mitgliedern multikultureller Gruppen.*

Neben der Effektivität sind Zufriedenheit und Motivation der Gruppenmitglieder innerhalb der FührerInnen-Geführten-Beziehung und insbesondere der Arbeit in der Gruppe ein entscheidender Erfolgsfaktor der Gruppenarbeit (Wageman / Hackman / Lehman, 2005, S. 376; Hoegl / Gemuenden, 2001, S. 439; Chong / Thomas, 1997, S. 279f.). Vor allem die kulturelle Vielfalt der Individuen in heterogenen Gruppen stellt, wie bereits erläutert, einerseits einen möglichen Wettbewerbsvorteil dar. Andererseits können die auftretenden Probleme aber herausfordernd oder sogar kräfteraubend sein und können in weiterer Folge einen negativen Einfluss auf die Motivation und die Zufriedenheit der Gruppenmitglieder haben (van der Zee / Atsma / Brodbeck, 2004, S. 284; Ely / Thomas, 2001, S. 230; Cady / Valentine, 1999, S. 746; Pelled, 1996, S. 616). Gemäß diesen Behauptungen möchte der Verfasser dieser Dissertation die folgende Hypothese empirisch untersuchen.

H4: *Die Zufriedenheit und Motivation der Gruppenmitglieder ist in homogenen Gruppen größer als in multikulturellen Gruppen.*

Nach Erläuterung des Forschungsdesigns, Operationalisierung und Darstellung der forschungsrelevanten Fragen und Hypothesen werden nun die methodischen Grundlagen für das empirische Vorgehen verdeutlicht.

3.2 Methodische Grundlagen

3.2.1 Die empirische Gewinnung von Informationen mittels Experiment

Im Rahmen der Datengewinnung wird eine empirische Untersuchung unter experimentellen Bedingungen durchgeführt. Ziel ist es, zum einen die Wahrnehmung, Zuschreibung und Erwartungen der Geführten gegenüber der Führungskraft, im Sinne der impliziten Führungstheorien („Leader perception“) und zum anderen die Zusammenarbeit in der Gruppe („Group perception“) zu untersuchen.

Das Experiment als spezifische Form des Untersuchungsdesigns ermöglicht dabei das Variieren einzelner Bedingungsfaktoren, sogenannte unabhängige Variablen, um zu überprüfen und zu analysieren, welche Effekte, also abhängige Variablen, sich daraus erheben (Kühl, 2005, S. 213). Bei experimentellen Untersuchungen werden Untersuchungsobjekte mittels Randomisierung aufgeteilt. Unter Randomisierung wird die Zuteilung der Versuchspersonen zu experimentellen Gruppen nach einem Zufallsverfahren verstanden (Diekmann, 2009, S. 337; Schnell / Hill / Esser, 2005, S. 223; Bortz / Döring, 2003, S. 58; Manstead / Semin, 2002, S. 89). Ziel ist es, Informationen für kausale Schlüsse bereitzustellen (Manstead / Semin, 2002, S. 85). Durch die Änderung der abhängigen Variablen erhält der Forscher bzw. die Forscherin die Möglichkeit, mit Instrumenten der Datenerhebung, wie etwa Beobachtungen, Inhaltsanalysen, oder mündliche sowie schriftliche Befragungen, Veränderungen entsprechend zu messen (Kühl, 2005, S. 213; Croson / Anand / Agarwal, 2007, S. 174; Friedrichs, 1990, S. 333). Gleichzeitig können Störeinflüsse minimiert werden und dadurch eine hohe Güte des experimentellen Vorgehens erzielt werden (Puck, 2009, S. 110).

Demgegenüber stehen Quasi-Experimente zu unterscheiden. „Quasi-Experimente kann man definieren als Versuchsanordnungen, die dem Vorbild des Experiments nahekommen und der experimentellen Logik folgen, jedoch nicht die strengen Anforderungen an experimentelle Designs erfüllen.“ (Diekmann, 2009, S. 356) Beim Quasi-Experiment können die Testpersonen nicht zufällig auf die verschiedenen Experimentalbedingungen verteilt werden, Versuchs- und Kontrollgruppen werden somit nicht durch Randomisierung zugeordnet (Kühl, 2005, S. 215; Schnell / Hill / Esser, 2005, S. 229; Manstead / Semin, 2002, S. 89). Es wird daher mit natürlichen Gruppen gearbeitet (Bortz /

Döring, 2003, S. 58). Im klassischen Fall werden Quasi-Experimente in einer natürlichen, alltäglichen Umgebung durchgeführt. Bei dieser Bedingung hat der Forscher bzw. die Forscherin keine vollständige Kontrolle, da es sich um eine alltägliche Situation handelt (Manstead / Semin, 2002, S. 89).

Eine Möglichkeit, Vorteile beider Untersuchungsdesigns zu verbinden, stellt ein Feldexperiment dar. Feldexperimente sind grundsätzlich „echte“ Experimente, die in einer natürlichen bzw. gewohnten Form, d.h. in einer vom Untersucher möglichst unbeeinflussten, natürlichen Umgebung, durchgeführt werden (Kühl, 2005, S. 215; Atteslander, 2003, S. 200; Bortz / Döring, 2003, S. 60; Manstead / Semin, 2002, S. 91).

Neben den Felduntersuchungen gibt es Laboruntersuchungen, die in Umgebungen durchgeführt werden, die eine weitgehende Ausschaltung oder Kontrolle von Störgrößen ermöglichen. Bei diesem Untersuchungsdesign wird eine unnatürliche Umgebung konstruiert, wodurch es fraglich erscheint, ob die Ergebnisse auch auf andere, „natürlichere“ Situationen generalisierbar sind (Kühl, 2005, S. 214; Bortz / Döring, 2003, S. 60f.).

Das Experiment verschafft somit einige Vorteile. Einerseits bietet es die Möglichkeit, Personen und Gegenstände in einen „künstlich“ gestalteten Prozess einzubinden und damit soziale Prozesse kontrolliert zu untersuchen (Diekmann, 2009, S. 303; Atteslander, 2006, S. 166). Andererseits können in Experimenten „Extremsituationen“, die relativ selten vorkommen, konstruiert und empirisch überprüft werden. Aufgrund der Tatsache, dass Kausalbeziehungen im Bereich sozialer Phänomene feststellbar sind, gilt das Experiment außerdem als sicherste Methode der empirischen Sozialforschung (Atteslander, 2006, S. 166)

3.2.2 Der Fragebogen als Instrument quantitativer Datenerhebung

Für die Beantwortung der Forschungsfragen und Hypothesen werden Daten aus dem Experiment erhoben. Um dies zu gewährleisten, bedient sich der Autor der schriftlichen Befragung, konkret einem Fragebogen mit geschlossenen Antwortkategorien.

Grundsätzlich spricht man von einer schriftlichen Befragung als Methode der empirischen Sozialwissenschaft, wenn UntersuchungsteilnehmerInnen schriftlich vorgelegte Fragen, in Form eines Fragebogens, selbstständig schriftlich be-

antworten (Bortz / Döring, 2003, S. 253). „Ein Fragebogen ist ein Datenerhebungsinstrument, das aus einer Kollektion von Fragen besteht, mittels derer diejenigen Informationen gewonnen werden sollen, welche gemäß dem ‚theoretischen Vorspann‘ einer empirischen Untersuchung (Erkenntnis- bzw. Verwertungsinteresse, dimensionale und/oder semantische Analyse, Begriffsdefinitionen, Indikatorenwahl, Operationalisierung) benötigt werden.“ (Stier, 1999, S. 181) Diese Form erfordert eine hohe Strukturierbarkeit der Inhalte und verzichtet auf steuernde Eingriffe eines Interviewers oder einer Interviewerin (Bortz / Döring, 2003, S. 253).

Als wesentlicher Vorteil dieser Erhebungsform kann die rasche Datensammlung für aussagekräftige, statistische Auswertungen und Interpretationen genannt werden. In der Wissenschaft wird oftmals auf den Nachteil der unkontrollierten Erhebungssituation verwiesen. Dem lässt sich allerdings entgegenhalten, indem mehrere Befragte in Gruppen unter standardisierten Bedingungen bei Anwesenheit eines Untersuchungsleiters gleichzeitig befragt werden (Bortz / Döring, 2003, S. 253).

In Bezug auf die Auswahl der Fragen weisen Bortz / Döring (2003, S. 253f.) darauf hin, dass es ratsam ist zu überprüfen, ob bereits ein im Feld überprüfter Fragebogen anderer AutorInnen existiert, der für das eigene empirische Vorgehen geeignet ist. Wichtig ist dabei, die drei Testgütekriterien¹⁴, also Objektivität,

14 Mit den drei zentralen Kriterien der Testgüte (Objektivität, Reliabilität und Validität) lässt sich die Qualität eines Fragebogens festmachen (Diekmann, 2009, S. 247f.; Bortz / Döring, 2003, S. 193). Fragebögen und standardisierte Tests, die unter diesen Bedingungen entwickelt wurden bzw. sich auf diese Kriterien beziehen, werden als psychometrische Tests oder Fragebögen bezeichnet. Davon sind die projektiven Tests, sogenannte Entfaltungstests, abzugrenzen (Bortz / Döring, 2003, S. 193f.). Objektivität bezieht sich auf das Ausmaß der Unabhängigkeit der Ergebnisse von der Person, die den Fragebogen entwickelt hat (Diekmann, 2009, S. 249; Bortz / Döring, 2003, S. 194). Es ist zumeist ein unproblematisches Kriterium, das leicht realisierbar ist. Dazu ist jedenfalls standardisiert festzulegen, wie der Test durchzuführen, auszuwerten und die Ergebnisse zu interpretieren sind. Reliabilität, also die Zuverlässigkeit, ist eine Bezeichnung für den Grad der Messgenauigkeit oder Präzision eines Fragebogens (Bortz / Döring, 2003, S. 195), der die Reproduzierbarkeit eines Messinstruments zeigt (Diekmann, 2009, S. 250). Das bedeutet, eine Messung ist dann reliabel, wenn sie bei mehrfacher Durchführung (etwa zu unterschiedlichen Zeitpunkten und bei Erhebungen mit unterschiedlichen Befragten) zum gleichen Ergebnis führt (Manstead / Semin, 2002, S. 106). Die Validität ist das wichtigste Testgütekriterium und gibt die Gültigkeit eines Testin-

Reliabilität und Validität, zwar anzuerkennen, jedoch diese Faktoren in der eigenen Untersuchung zu überprüfen. Außerdem ist es entscheidend, Fragebögen auf die Sprachgewohnheiten der befragten Zielgruppe abzustimmen. Gibt es noch keine relevanten Voruntersuchungen und Fragebögen, hat der Forscher bzw. die Forscherin einen Fragebogen zu konstruieren. Da in dieser Dissertation ein Fragebogen eingesetzt wird, der auf zwei Theorien bzw. Instrumenten basiert, jedoch bereits mehrfach getestet und publiziert wurde, wird an dieser Stelle nicht auf die allgemeine Fragebogenkonstruktion eingegangen (Bortz / Döring, 2003, S. 253f.).

Mit der Methode „Fragebogen“ lassen sich die zu erhebenden Inhalte auch nach der Art der erfragten Informationen unterscheiden. Sie können sich dabei auf Folgendes beziehen: (Brake, 2005, S. 37)

- Einstellung oder Meinungen der Befragten,
- Überzeugungen der Befragten (subjektives Wissen über die Realität),
- Verhalten (retrospektiv / prospektiv) der Befragten,
- Eigenschaften der Befragten / soziodemographische Daten.

In den meisten Fällen enthält ein Fragebogen Items bzw. Fragen aus allen vier Bereichen, die soeben genannt wurden. Gleichzeitig müssen ForscherInnen das Layout des Fragebogens so übersichtlich wie möglich designen und auch den Inhalt und den Aufbau so strukturiert wie möglich gestalten (Brake, 2005, S. 37ff.). Dadurch wird garantiert, dass es den Befragten beim Ausfüllen so leicht wie möglich gemacht wird, sie zügig die Antworten auswählen können und nicht mitten in der Beantwortung der Fragen den Ausfüllprozess beenden (Schnell / Hill / Esser, 2005, S. 346).

Für die Beantwortung von Fragebögen mit geschlossenen Antwortkategorien werden unterschiedliche Arten von Techniken zur Messung von individuellen Einstellungen angewandt. Eine der häufigsten Verfahren stellt die Likert-Skala

struments an, also wie gut der Fragebogen in der Lage ist, das zu messen, was er zu messen vorgibt. Die drei Hauptarten von Validität, die die Präzision bestimmen, lauten Inhaltsvalidität, Kriteriumsvalidität und Konstruktvalidität (Diekmann, 2009, S. 256ff.; Bortz / Döring, 2003, S. 199f.; Manstead / Semin, 2002, S. 106).

dar, die auch im Rahmen der empirischen Untersuchung verwendet wird. Die Likert-Skalierung ist ein personenorientiertes, eindimensionales Skalierungsverfahren, das zur Selbsteinschätzung eingesetzt wird (Bortz / Döring, 2003, S. 222; Stier, 1999, S. 79). „Das Schlüsselmerkmal dieser Methode besteht darin, dass die Befragten gebeten werden, das Ausmaß ihrer Zustimmung oder Ablehnung zu einer Reihe von Aussagen über den Einstellungsgegenstand einzustufen.“ (Manstead / Semin, 2002, S. 111) In den meisten Fällen werden Rating Skalen mit fünf Antwortkategorien vorgelegt, die beispielsweise von „stimme stark zu“ bis zu „lehne stark ab“ lauten können. Unter Verwendung der 5-Rating-Skalen-Kategorien ergibt sich ein Testwert einer UntersuchungsteilnehmerIn als Summe der von der Person gewählten Skalenwerte. Daraufhin wird für jedes Item ein Trennschärfeindex ermittelt. Die Items mit den höchsten Werten bilden anschließend die Testskala (Bortz / Döring, 2003, S. 223; Stier, 1999, S80ff.). „Der Testwert einer mit der endgültigen Skala getesteten Person entspricht der Summe der angekreuzten, kategorienspezifischen Skalenwerte.“ (Bortz / Döring, 2003, S. 223f.) In der Praxis wird man häufig mit der Schwierigkeit konfrontiert, dass die Mittelkategorie nur schwer interpretierbar ist. Um dieser Herausforderung adäquat zu begegnen, weisen ForscherInnen explizit in der Instruktion darauf hin, wie die Mittelkategorie aufzufassen ist. Eine weitere Möglichkeit stellt das Ausweichen auf eine vierstufige Antwortskala dar (Bortz / Döring, S. 224).

3.2.3 Erstellung und Aufbau des Fragebogens

Für die Datenerhebung wurde, wie bereits ausführlich dargelegt, die Forschungsmethode der schriftlichen Befragung angewandt. Der Fragebogen, der in englischer Sprache verfasst wurde, bestand aus drei Teilen: dem ersten Teil, in dem sozioökonomische Daten erhoben wurden, dem zweiten Bereich, der 26 Items beinhaltete, die sich auf die „Leader perception“ bezogen und dem dritten Teil, der aus 38 Fragen bestand, die die „Group perception“ beleuchteten. Die eingesetzten 64 Items des zweiten und dritten Bereichs werden in Tabelle 6 dargestellt. Diese wurden, wie bereits erwähnt, aus bereits publizierten und auf die Kriterien der Testgüte (Objektivität, Reliabilität und Validität) überprüften Theorien und Instrumente (Chong / Thomas, 1997 bzw. Wageman / Hackman / Lehman, 2005) übernommen.

Im ersten Teil des Fragebogens wurden sozioökonomische Daten der Befragten erhoben. Dazu zählten:

- Alter
- Geschlecht (männlich / weiblich)
- Nationalität
- Aktuelle berufliche Tätigkeit (ja / nein)
- Wenn ja, in welchem Ausmaß (Vollzeit [mehr als 30 Stunden] / Teilzeit)
- Berufserfahrung (in Monaten)

In weiterer Folge wurden 64 Items im Fragebogen eingesetzt, wobei sich primär 26 Fragen auf die implizite Führungstheorie nach Chong / Thomas (1997) bezogen und somit die „Leader perception“-Ebene beleuchteten, während 38 Items aus dem „Team Diagnostic Survey“ nach Wageman / Hackman / Lehman (2005) entnommen wurden, um wertvolle Erkenntnisse der „Group perception“-Ebene zu erlangen. Die implizite Führungstheorie nach Chong / Thomas (1997) eignet sich insbesondere für die Untersuchung der Wahrnehmungen von Geführten mit unterschiedlichem kulturellen Hintergrund beziehungsweise Ethnie, die in multikulturellen und homogenen Gruppen arbeiten und die Führungsfähigkeit und die Effektivität einer Führungskraft beurteilen. Die gewählte Theorie gibt die Möglichkeit, Führung unter homogenen und multikulturellen Arbeitsbedingungen zu beleuchten und zu untersuchen. Außerdem nimmt sie sowohl auf den kulturellen Hintergrund der Geführten, wie auch auf die Führungsperson Bezug und stellt damit einen ausgezeichneten Bezugsrahmen für die empirische Erhebung dar (Gelfand / Erez / Aycan, 2007, S. 494; Yan / Hunt, 2005, S. 50; Lord / Emrich, 2001, S. 561; Hanges / Lord / Dickson, 2000, S. 135; Chong / Thomas, 1997, S. 278). Eine detaillierte Darstellung dieser impliziten Führungstheorie erfolgte im Kapitel 2.2.4.2 dieser Dissertation.

Der Team Diagnostic Survey nach Wageman / Hackman / Lehman (2005) wurde als Erhebungsinstrument konstruiert, das in kurzer Zeit von Untersuchungsobjekten ausgefüllt werden kann. Die psychometrischen Eigenschaften sind zufriedenstellend und wurden auf der Basis von Analysen von 2.474 Befragten in 321 Teams unterschiedlicher Organisationen erhoben (Wageman / Hackman / Lehman, 2005, S. 373f.).

Der Team Diagnostic Survey basiert auf bestehenden wissenschaftlichen Entwicklungen und Theorien über die Bedingungen, die die Effektivität einer

Gruppe oder eines Teams fördern. Das Instrument misst unterschiedliche leistungsrelevante Aspekte eines Teams unter Bezugnahme der Wahrnehmungen der befragten Gruppenmitglieder hinsichtlich soziostruktureller Merkmale, wie etwa Gruppenzusammensetzung, Arbeitsanordnung, Prozessindikatoren oder die Effektivität der Leistungserbringung (Higgins / Weiner / Young, 2012, S. 374). Dadurch können wertvolle Informationen über die Qualität der Gruppenarbeit, die Leader-Follower-Beziehung und mögliche Hebel und auch Schwachstellen gefunden werden, die die Effektivität der Gruppenarbeit verbessern können (Wageman / Hackman / Lehman, 2005, S. 395).

Grundsätzlich zielt der Team Diagnostic Survey in erster Linie auf die Untersuchung der Arbeit einer Gruppe von Personen, die mit der gemeinsamen Leistungserbringung und Zielerreichung vertraut sind und demnach als Team im engeren Sinn bezeichnet werden können (Wageman / Hackman / Lehman, 2005, S. 395). Zur Unterscheidung einer Gruppe von einem Team siehe Kapitel 2.5.1 dieser Dissertation. In der empirischen Untersuchung werden jedoch dezidiert Arbeitsgruppen untersucht, deren Gruppenmitglieder sich nicht kennen, noch nicht miteinander gearbeitet haben und daher nicht Teams im engeren Sinn sind. Der Autor dieser Dissertation übernahm daher ausgewählte Skalen aus dem Team Diagnostic Survey, adaptierte und entwickelte diese bezogen auf den Untersuchungskontext weiter. Im Fragebogen finden folgende Skalen aus dem Erhebungsinstrument von Wageman, Hackman und Lehman (2005) Anwendung: (Wageman / Hackman / Lehman, 2005, S. 383ff.)

- **Enabling structure:** Diese Skala zielt auf die strukturelle Ausrichtung von Gruppen aus und hinterfragt, ob Gruppen überhaupt richtig zusammengestellt wurden. Schließlich sind Gruppen oftmals überarbeitet, Aufgaben sind nicht ausreichend bestimmt oder die Strukturen lassen eine effektive Outputgenerierung gar nicht zu (Wageman / Hackman / Lehman, 2005, S. 377). Um dies adäquat zu messen, werden folgende Unterskalen in den Fragebogen eingebaut:
 - *Team composition.* Im Fragebogen wurde diese Subskala als „Group composition“ bezeichnet. Damit wird die Größe der Gruppen („Size“), die Vielfalt in den Gruppen („Diversity“) und die Fähigkeiten und Fertigkeiten („Skills“) gemessen.
 - Darüber hinaus bedient sich der Verfasser dieser Dissertation der Subskala „Group norms“.

- **Process criteria:** Die Skala der Prozesskriterien dient der Messung des Umfangs und der Ebenen, die Gruppenmitglieder bei der Leistungserstellung und -erbringung aufwenden („*Effort-related process criteria*“). Weiters wird die Qualität der Strategien, die im Prozess der Arbeitsleistung eingesetzt werden, untersucht („*Strategy-related process criteria*“). Außerdem können das Ausmaß und die Verteilung der gesamten Fähigkeiten und Fertigkeiten und das Wissen der Gruppenmitglieder überprüft werden („*Knowledge- and skill-related process criteria*“). Insgesamt ermöglicht diese Skala, die in diesem Fragebogen als „Process criteria (of Group effectiveness)“ bezeichnet wird, die Erhebung der Qualität der Aufgabenprozesse der Gruppen (Wageman / Hackman / Lehman, 2005, S. 387).
- **Social processes:** Die Skala „Social processes“ misst die Qualität der interpersonellen Beziehungen und fokussiert hier auf einer kollektiven Ebene. Mit der Subskala „*Quality of interaction*“ werden die Interaktionen der Mitglieder beleuchtet, während die Subskala „*Relationship satisfaction*“ die Zufriedenheit der Gruppenmitglieder mit den Beziehungen zueinander untersucht (Wageman / Hackman / Lehman, 2005, S. 388).
- **Individual well-being:** Mit dieser Skala wird die individuelle Perspektive jedes Gruppenmitglieds erhoben, die die Emotionen und Reaktionen der Gruppenmitglieder auf die Gruppe selbst und ihre Arbeit widerspiegelt. Dazu dienen die Subskalen „*Internal work motivation*“, „*Growth satisfaction*“ und „*General satisfaction*“ (Wageman / Hackman / Lehman, 2005, S. 388).

Das Theoriemodell wurde in zahlreichen Studien verwendet und zeigte, dass es ein valides und zuverlässiges Instrument ist, um soziokulturelle und -strukturelle Merkmale zu messen (Higgins / Weiner / Young, 2012, S. 374). Für die empirische Untersuchung dieser Dissertation stellen die verwendeten und weiterentwickelten Fragebogenbatterien des Team Diagnostic Survey ein geeignetes Erhebungsinstrument für die Ebene der „Group perception“ dar. Einen Überblick über die Skalen und Items gibt Tabelle 7, der eingesetzte Fragebogen ist im Anhang angeführt.

Items: Implicit Leadership Theory (Leader perception)

1	Treats me fairly
2	Tries to understand view
3	Treats all members as equal
4	Can talk easily about the task
5	Is friendly and approachable
6	Concerned about my personal problems with the task
7	Think the leader trusts me
8	Gives recognition for a good job
9	Tries to improve the working atmosphere
10	Asks opinion on problem-solving
11	Offers new approaches to problems
12	Strict observation of rules and regulations
13	Urges to complete work on time
14	Asks to follow rules and regulations
15	Makes me work to the max.
16	Requires progress reports
17	Sets clear deadlines
18	Talks about how much to do
19	Gives instructions/orders
20	Assigns members to tasks
21	Knowledgeable about equipment/systems
22	Lets you know about plans and tasks
23	Makes precise plans for goal achievement
24	Maintains definite standards
25	Doesn't waste working time
26	Improves facilities when asked

Items: Group Work (Group perception)

27			This work group is larger than it needs to be. (R)
28	Enabling structure	<i>Group composition - Size</i>	This work group has too few members for what it has to accomplish. (R)
29			This work group is just the right size to accomplish its purposes.

30		<i>Group composition - Diversity</i>	Members of this work group are too dissimilar to work together well. (R)
31			This work group does not have a broad enough range of experiences and perspectives to accomplish its purposes. (R)
32			This work group has a nearly ideal "mix" of members - a diverse set of people who bring different perspectives and experiences to the work.
33		<i>Group composition - Skills</i>	Members of this work group have more than enough talent and experience for the kind of work that we do.
34			Everyone in this work group has the special skills that are needed for team work.
35			Some members of this work group lack the knowledge and skills that they need to do their parts of the work group's work. (R)
36		<i>Group norms</i>	Standards for member behavior in this work group are vague and unclear. (R)
37			It is clear what is - and what is not - acceptable member behavior in this work group.
38			Members of this work group agree about how members are expected to behave.
39	Process criteria	<i>Effort</i>	Members demonstrate their commitment to our work group by putting in extra time and effort to help it succeed.
40			Everyone on this work group is motivated to have the work group succeed.
41			Some members of our work group do not carry their fair share of the overall workload. (R)
42		<i>Strategy</i>	Our work group often comes up with innovative ways of proceeding with the work that turn out to be just what is needed.
43			Our work group often falls into mindless routines, without noticing any changes that may have occurred in our situation. (R)
44		Our work group has a great deal of difficulty actually carrying out the plans we make for how we will proceed with the task. (R)	
45	Process criteria (Cont'd)	<i>Knowledge and Skill</i>	How seriously a member's ideas are taken by others on our work group often depends more on who the person is than on how much he or she actually knows. (R)
46			Members of our work group actively share their special knowledge and expertise with one another.
47			Our work group is quite skilled at capturing the lessons that can be learned from our work experiences.
48	Social processes	<i>Quality of interaction</i>	There is a lot of unpleasantness among members of this work group. (R)
49			The longer we work together as a work group, the less well we do. (R)

50			Working together energizes and uplifts members of our work group.	
51			Every time someone attempts to correct a member of our work group whose behavior is not acceptable, things seem to get worse rather than better. (R)	
52		<i>Relationship satisfaction</i>	My relations with other work group members are strained. (R)	
53			I very much enjoy talking and working with my teammates.	
54			The chance to get to know my teammates is one of the best parts of working on this work group.	
55	Individual well-being	<i>Internal work motivation</i>	I felt a real sense of personal satisfaction when our work group did well.	
56				I felt bad and unhappy when our work group performed poorly.
57				My own feelings are not affected one way or the other by how well our work group performs. (R)
58				When our work group did well, I did well.
59			<i>Growth satisfaction</i>	I learnt a great deal from my work on this work group.
60				My own creativity and initiative were suppressed by this work group. (R)
61				Working on this work group stretches my personal knowledge and skills.
62			<i>General satisfaction</i>	I enjoyed the kind of work we did in this work group.
63		Working on this work group was an exercise in frustration. (R)		
64		Generally speaking, I was very satisfied with this work group.		
<i>Note: (R) = "reverse coded"</i>				

Tabelle 7: Darstellung der Items des Fragebogens (Quelle: Eigene Darstellung)

Nach der Aufarbeitung der Grundlagen der Erhebungsmethoden und der Erläuterung des konkreten Fragebogens erfolgt nun im Kapitel 3.3 die Darstellung der konkreten Umsetzung des empirischen Vorhabens.

3.3 Konkrete Umsetzung des empirischen Vorhabens

Die Darstellung der konkreten Umsetzung des empirischen Vorhabens unterteilt sich in die Erklärung des Samples (Kapitel 3.3.1), den Verweis auf den Pretest (Kapitel 3.3.2), den Aufbau und Ablauf des Experiments (Kapitel 3.3.3) und den Methoden der Datenauswertung (Kapitel 3.3.4).

3.3.1 Sample

An den experimentellen Erhebungen im Mai 2011 nahmen jeweils 20 Personen aus Österreich, Türkei, China und den Vereinigten Staaten von Amerika teil (n=80).¹⁵ Alle TeilnehmerInnen waren zum Zeitpunkt des Experiments Vollzeit- oder Austauschstudierende der Wirtschaftsuniversität Wien. Grundsätzlich wurden für die Untersuchungen zwei Vergleichsgruppen, die jeweils aus vier Gruppen bestanden, gebildet. Jede Gruppe hatte eine männliche Führungsperson aus jeweils einer Nation / Kultur (Österreich, Türkei, China und USA) und neun Geführte. Die Zusammensetzungen erfolgten entweder homogen und multikulturell, was im vorliegenden Experiment vier unterschiedliche Nationalkulturen bedeuteten (Anm.: in der multikulturellen Zusammensetzung hat die Führungskraft eine bestimmte Nationalkultur, während die Geführten über unterschiedliche Kulturen, konkret drei verschiedene, verfügen). Der genaue Ablauf wird in Abbildung 13 verdeutlicht. Es wurde bewusst nur ein Geschlecht bei den Führungskräften berücksichtigt, um Gender-Effekte auszuschließen. Das Durchschnittsalter aller UntersuchungsteilnehmerInnen lag gesamt bei 22,57 Jahre (AT=22,83, TR=22,56, CN=24,22 und US=20,67) und wird in Abbildung 7 verdeutlicht.

15 Fortan werden zum Teil folgende Länderabkürzungen verwendet: AT (Österreich), TR (Türkei), CN (China) und US (Vereinigten Staaten von Amerika).

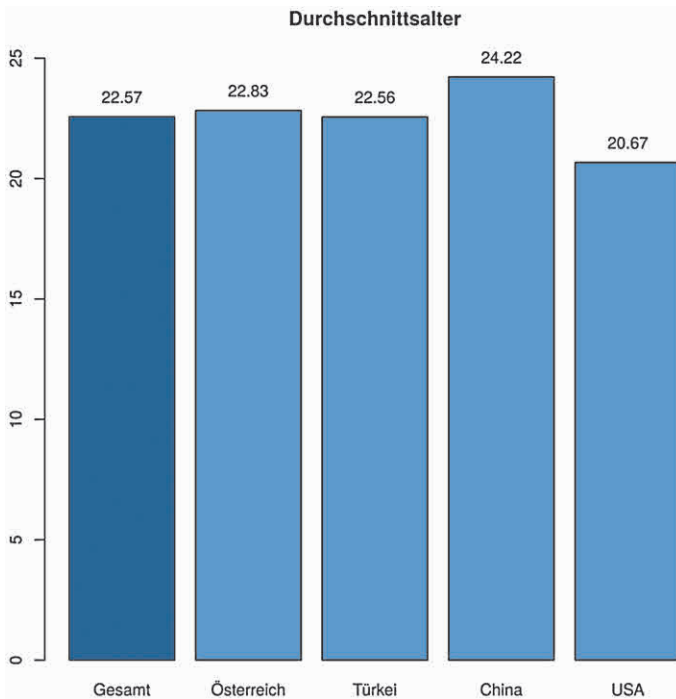


Abbildung 7: Durchschnittsalter der UntersuchungsteilnehmerInnen
(Quelle: Eigene Darstellung)

Tabelle 8 zeigt die Varianzunterschiede, sowie die Ergebnisse eines F-Tests in Bezug auf die Varianz des Alters der Studierenden, die an der wissenschaftlichen Erhebung teilgenommen haben. Dabei konnte ein signifikanter Unterschied in der Schwankungsbreite des Alters bei den österreichischen TeilnehmerInnen festgestellt werden.

	SD Homogen	SD Heterogen	p	Sig.
AT	4,14	1,67	0,019	*
TR	2,45	1,79	0,391	
CN	4,66	4,42	0,884	
US	0,78	0,44	0,126	

Tabelle 8: F-Test bezogen auf das Alter der TeilnehmerInnen
(Quelle: Eigene Darstellung)

Bezogen auf die Gender-Verteilung aller UntersuchungsteilnehmerInnen zeigt Abbildung 8, dass insgesamt 33 männliche und 39 weibliche Studierende als Geführte am Experiment mitwirkten. In den Gruppen fungierten ausschließlich männliche Studierende als Führungspersonen, da der Autor die Einflüsse genderspezifischer Unterschiede in der Führung vermeiden wollte. Die Verteilung bezogen auf Nationalitäten verlief relativ ausgeglichen (AT: m=7, w=11; TR: m=9, w=9; CN: m=8, w=10; US: m=9, w=9).

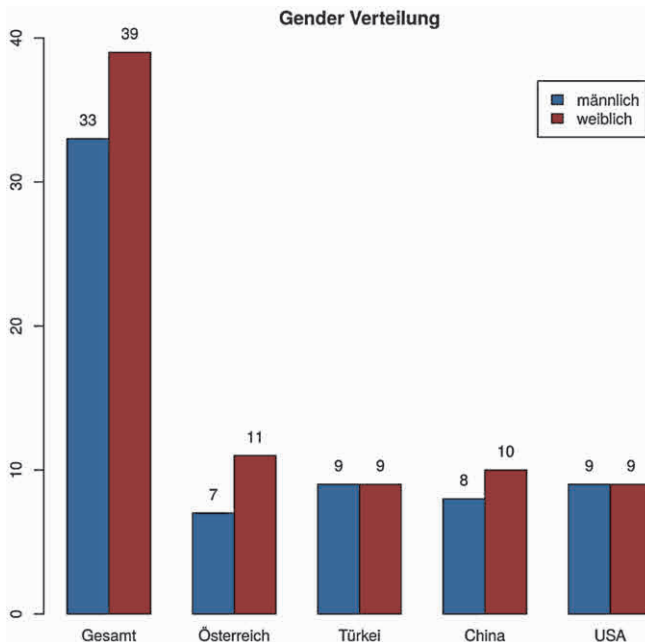


Abbildung 8: Gender-Verteilung der Geführten pro Nation
(Quelle: Eigene Darstellung)

Abbildung 9 stellt die durchschnittliche Arbeitserfahrung aller UntersuchungsteilnehmerInnen in Monaten, also Führungspersonen und Geführten, dar (gesamt=19,5 Monate). Bezogen auf die einzelnen Nationen ergeben sich folgende Erfahrungswerte: AT=25,2 Monate; TR=11,2 Monate; CN=18,1 Monate; US=23,7 Monate.

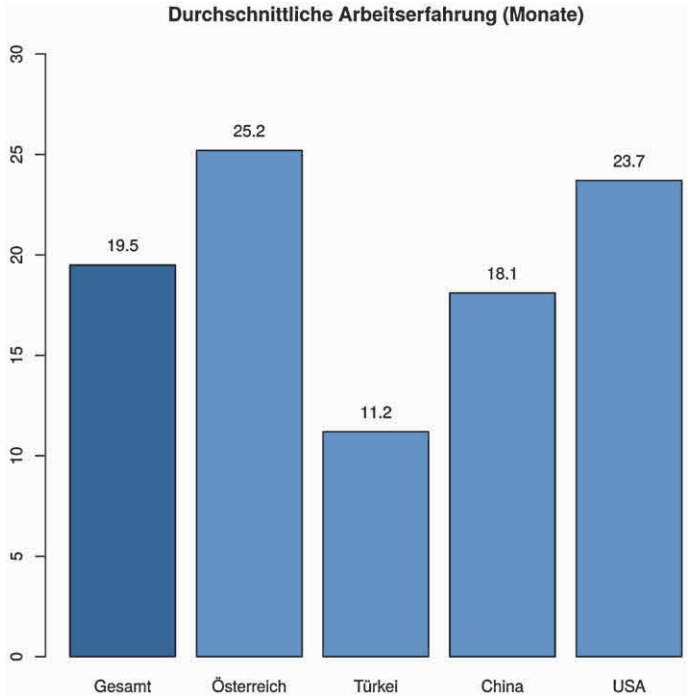


Abbildung 9: Durchschnittliche Arbeitserfahrung (in Monaten) der UntersuchungsteilnehmerInnen (Quelle: Eigene Darstellung)

Varianzunterschiede und die Ergebnisse eines F-Tests in Bezug auf die Varianz der Berufserfahrung der UntersuchungsteilnehmerInnen werden in Tabelle 9 gezeigt. In der Schwankungsbreite der Berufserfahrung konnte ein signifikanter Unterschied bei den türkischen TeilnehmerInnen festgestellt werden.

	SD Homogen	SD Heterogen	p	Sig.
AT	19,81	18,86	0,893	
TR	20,43	8,49	0,023	*
CN	23,1	31,36	0,405	
US	20,06	18,84	0,864	

Tabelle 9: F-Test bezogen auf die Berufserfahrung der UntersuchungsteilnehmerInnen
(Quelle: Eigene Darstellung)

3.3.2 Pretest

Um die Umsetzbarkeit der empirischen Untersuchung zu gewährleisten, wurde im Vorfeld ein Pretest durchgeführt. Ein Pretest dient der Überprüfung: (Schnell / Hill / Esser, 2005, S. 347):

- „der ausreichenden Variation der Antworten,
- des Verständnisses der Fragen durch den Befragten,
- der Schwierigkeit der Fragen für den Befragten,
- des Interesses und der Aufmerksamkeit des Befragten gegenüber den Fragen,
- der Kontinuität des Interviewablaufs („Fluss“),
- der Effekte der Fragenanordnung,
- der Güte der Filterführung,
- von Kontexteffekten,
- der Dauer der Befragung,
- des Interesses des Befragten gegenüber der gesamten Befragung,
- der Belastung des Befragten durch die Befragung.“

Außerdem ermöglicht ein Pretest herauszufinden, ob gewisse Fragen zu Themenkomplexen oder Theorien vergessen oder ungenügend berücksichtigt wurden (Bortz / Döring, 2003, S. 359; Stier, 1999, S. 184). Der konkrete Pretest bestand aus einer experimentellen Untersuchung mit Studierenden unterschiedli-

cher Nationalitäten, die an die Hauptuntersuchung der Dissertation angelehnt wurde. Für die Datenerhebung kam ein Fragebogen zum Einsatz, der sich primär auf die implizite Führungstheorie nach Chong / Thomas (1997) bezog. Der reibungslose Ablauf, die Analysen und Ergebnisse zeigten die Umsetzbarkeit des geplanten Vorhabens. Die Erkenntnisse wurden in zwei Konferenzbeiträgen verarbeitet, vor einer internationalen Community präsentiert und diskutiert (vgl. hierzu Kasper et al., 2011; Kodydek, 2011).

3.3.3 Aufbau und Ablauf des Experiments

Wie bereits im Kapitel 3.2.1 dargestellt, wurden wissenschaftliche Untersuchungen unter experimentellen Bedingungen geplant. Die TeilnehmerInnen waren alle Vollzeit- und Austauschstudierende der Wirtschaftsuniversität Wien. Die Untersuchungen wurden in Hörsälen der WU Wien im Mai 2011 durchgeführt. Somit konnte dieses echte Experiment mit Zufallszuweisung in einer für die UntersuchungsteilnehmerInnen gewohnten Umgebung durchgeführt werden (Manstead / Semin, 2002, S. 89), da alle Studierende hauptberuflich ihrem Studium nachgehen und die Experimente durchaus mit Gruppenaufgaben in Lehrveranstaltungen und Seminaren zu vergleichen sind.¹⁶

Den empirischen Untersuchungen sind detaillierte Planungsphasen vorausgegangen, wie Abbildung 10 zeigt.

16 Es ist außerdem kein Untersuchungsleiter im engeren Sinn während des Experiments anwesend, der praktisch jede Veränderung des Umfeldes kontrollieren könnte. Dies würde einer Laboruntersuchung entsprechen.

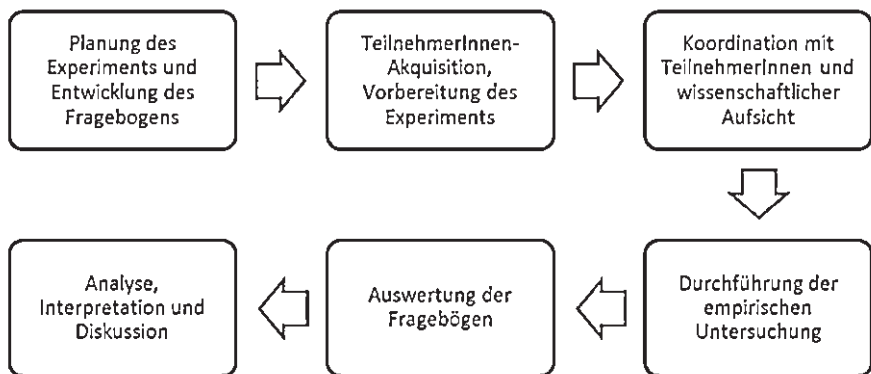


Abbildung 10: Ablauf des Forschungsvorhabens (Quelle: Eigene Darstellung)

Um Studierende unterschiedlicher Kulturen für das wissenschaftliche Vorhaben zu gewinnen, wurde mit MitarbeiterInnen des Zentrums für Auslandsstudien (ZAS) der Wirtschaftsuniversität Wien Kontakt aufgenommen. Dadurch konnten gezielt Austauschstudierende aus der Türkei, China und den USA angesprochen werden. Österreichische Studierende konnten aus dem Regelstudium für das Forschungsprojekt gefunden werden. Als kleiner Anreiz wurden den TeilnehmerInnen bei einer Mitwirkung Mensagutscheine zugesichert, die über ein Assistentenkleinprojekt von der Wirtschaftsuniversität Wien gefördert wurden. Die Studierenden wurden persönlich vom Verfasser dieser Dissertation angesprochen und über das Experiment und den Ablauf aufgeklärt. Sofern ein Studierender oder eine Studierende Interesse zeigte, wurden die E-Mailadresse notiert und das konkrete Datum der Erhebung, der Treffpunkt und die Uhrzeit bekannt gegeben. Jede(r) InteressentIn bekam zusätzlich noch ein Informationsblatt, das auch nochmals einige Tage vor dem Experiment als „Reminder“ ausgesandt wurde.

Der Verfasser dieser Dissertation wählte für das Experiment die gruppendynamische Übung „Turmbau“ nach Antons (2000, S. 131ff.). Dafür wurden die Arbeitsmaterialien angeschafft, die Räume an der Wirtschaftsuniversität Wien gebucht und Ansprechpersonen für etwaige Fragen, die im Umgang mit wissenschaftlichen Untersuchungen geschult und erfahren sind, akquiriert. Mit diesen KollegInnen wurde ein separater Besprechungstermin am Tag des Experiments vereinbart.

Die Gruppenexperimente fanden im Sommersemester 2011 in unterschiedlichen Räumen an der Wirtschaftsuniversität Wien mit 80 Studierenden aus den geplanten Ländern Österreich, Türkei, China und den USA statt. Kurz vor dem eigentlichen Beginn wurde der Ablauf nochmals mit KollegInnen besprochen, die als Ansprechpersonen während der Experimente in den Räumen zur Verfügung standen. Ihre grundlegenden Aufgaben waren die Führungspersonen detailliert über die Aufgabenstellungen und den Ablauf zu informieren, die Arbeitsmaterialien an diese Personen zu übergeben, die Fragebögen im Anschluss an die Erhebungsrunden auszuteilen und einzusammeln. Zusätzlich teilten sie am Ende die vereinbarten Mensagutscheine an die Studierenden aus.

Alle TeilnehmerInnen wurden gebeten, sich an den vorher festgelegten und per E-Mail kommunizierten Treffpunkten einzufinden. Die Studierenden wurden begrüßt und der Dank für ihre Teilnahme wurde ausgesprochen. Gleich darauf wurden sie darüber informiert, dass das Gesamtexperiment aus zwei Runden in unterschiedlichen Settings und Räumen stattfinden wird. Sie wurden gebeten, bis zum Ende zu bleiben und bei etwaigen Fragen die jeweilige Auskunftsperson, die als eine Art BetreuerIn in den Gruppenräumen verweilte, zu kontaktieren. Anschließend erfolgte die Zuteilung der TeilnehmerInnen auf die Versuchsgruppen.

Wie bereits erwähnt, wurden für das Experiment zwei Vergleichsgruppen, die jeweils aus vier Gruppen bestanden, gebildet. Jede Gruppe hatte eine Führungsperson aus jeweils einer Nation / Kultur (Österreich, Türkei, China und USA) und neun Geführte. Das Experiment bestand für jede teilnehmende Person aus zwei Prozessrunden. Entscheidend war, dass die einzig fixe Komponente die Führungskraft war, da diese über Runde 1 und Runde 2 konstant blieb. Die Geführten wechselten aber von einer homogenen Gruppe in eine heterogene bzw. multikulturelle in Vergleichsgruppe 1, während die TeilnehmerInnen der Vergleichsgruppe 2 die erste Runde in einem heterogenen bzw. multikulturellen Setting die Arbeitsaufgabe bewältigten und in der zweiten Runde mit Personen ihrer Kultur zusammenarbeiteten. Die unterschiedlichen Abläufe in den Vergleichsgruppen dienen dazu, Lerneffekte zu vermeiden und jede Gruppenform ohne konkrete Erfahrungswerte aus dem Gruppenprozess zu untersuchen und aussagekräftige Ergebnisse zu erhalten. Die Vergleichsgruppen arbeiteten darüber hinaus zur selben Zeit, sodass gewährleistet werden konnte, dass keine andere Gruppe über Möglichkeiten oder konkrete Abläufe informiert ist. Abbildung 11 zeigt die Zusammenstellung der Vergleichsgruppen mit Verweis auf die konträren Abläufe bei den Experimenten.

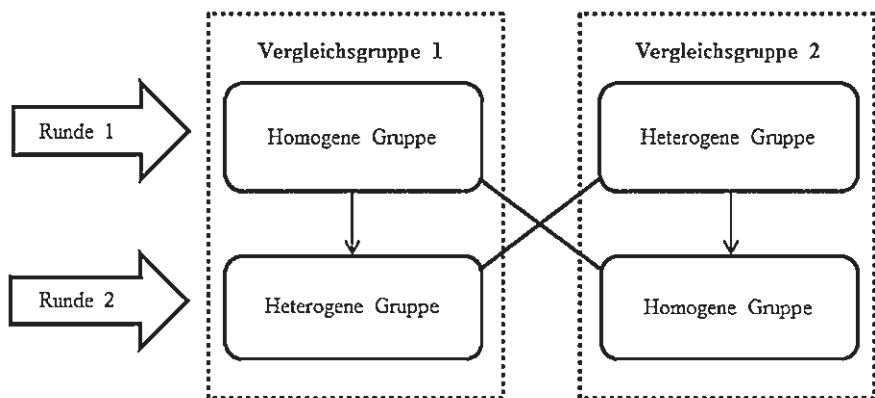


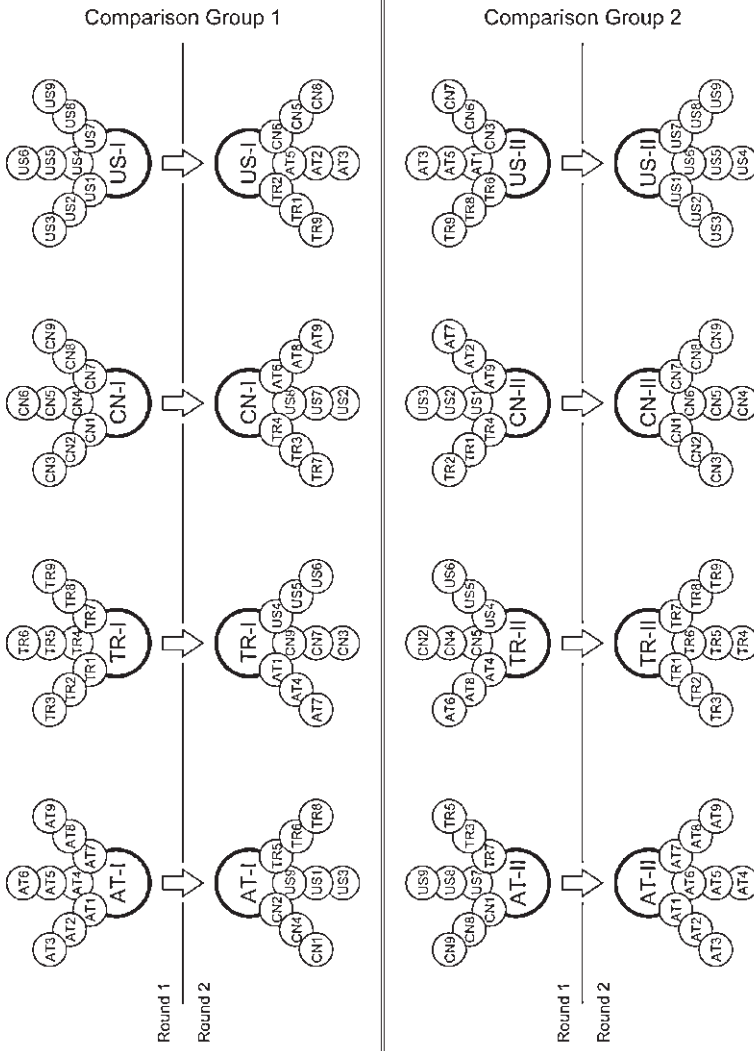
Abbildung 11: Die Vergleichsgruppen im Experiment unter Beibehaltung der Nachwuchsführungskräfte (Quelle: Eigene Darstellung)

Die Zuteilung erfolgte gemäß der Randomisierung, also zufällig und nicht kontrolliert¹⁷, indem die acht Aufsichtspersonen eine gewisse Anzahl an TeilnehmerInnen mit entsprechendem kulturellem Hintergrund aufriefen. Aufsichtspersonen, die die Vergleichsgruppen 1 beobachteten, baten jeweils zehn Personen gleicher Herkunft, sich zu ihnen zu begeben (zehn ÖsterreicherInnen, zehn TürkInnen, zehn ChinesInnen, zehn US-AmerikanerInnen). Die KollegInnen, die in weiterer Folge die Vergleichsgruppen 2 in der ersten Runde passiv beobachteten, forderten etwas konkreter drei Personen mit Herkunft A, drei Personen mit Herkunft B, drei Personen mit Herkunft C und eine Person mit Herkunft D auf, sich zu ihnen zu begeben. Anschließend gingen die Aufsichtspersonen mit den TeilnehmerInnen in die vorgesehenen Räume. Dort wurde jeweils ein männlicher Studierender, per Zufall, als Führungskraft ausgewählt und über seine künftige Rolle, die konstant über zwei Runden ging, aufgeklärt.

Anschließend gab es ein kurzes Briefing und eine schriftliche Aufgabenstellung für die Führungsperson und die Materialien wurden übergeben. Diese waren jeweils für jede Gruppe in Anlehnung an die Turmbau-Übung nach Antons (2000, S. 131): vier Bögen Karton, eine Tube Klebstoff, eine Schere, ein Lineal und

17 Durch die Randomisierung, also der zufälligen Zuteilung der TeilnehmerInnen zu Untersuchungsgruppen, wird unterstrichen, dass der Autor ein „echtes Experiment“ geplant und, wie bereits erwähnt, unter natürlichen Bedingungen im Feld durchführte.

vier Bögen Papier zum Entwerfen. Außerdem bekamen sie noch ein Informationsblatt, auf dem die Gruppenräume für die zweite Runde aufgelistet waren. Grundsätzlich war es die Aufgabe der Nachwuchsführungskräfte, eine Einteilung (TeilnehmerInnen arbeiteten in der zweiten Runde in heterogenen Gruppen) oder einen Verweis (TeilnehmerInnen arbeiteten in der zweiten Runde in homogenen Gruppen) gemäß der geplanten Anzahl an Personen pro kulturellen Hintergrund, durchzuführen, die TeilnehmerInnen dafür entsprechend auszuwählen und den jeweiligen Gruppenraum zu verlautbaren. Die Mitglieder aus den homogenen Gruppen arbeiteten nun in der zweiten Runde in einem multikulturellen Setting, während die Geführten aus der Vergleichsgruppe 2 sich in die homogenen Gruppen begaben. Es zeigt, dass die Aufteilung der UntersuchungsteilnehmerInnen in Runde 1 und 2, sowie die Ernennung der Führungspersonen per Zufall erfolgt ist (Randomisierung). Dadurch werden personenbezogene Störvariablen neutralisiert (Bortz / Döring, 2003, S. 58; Manstead / Semin, 2002, S. 89). Abbildung 12 zeigt die konkrete Zuteilung der TeilnehmerInnen zu den Vergleichsgruppen, also in welcher Gruppe welcher Geführte gearbeitet hat.



AT=Österreich, TR=Türkei, CN=China, US=Vereinigte Staaten von Amerika
 AT-I, AT-II, TR-I, TR-II, CN-I, CN-II, US-I, US-II = Leader (n=8); AT1 bis AT9, TR1 bis TR9, CN1 bis CN9, US1 bis US9 = Follower (n=72)

Abbildung 12: Darstellung der Zuteilung der UntersuchungsteilnehmerInnen zu den Vergleichsgruppen (Quelle: Eigene Darstellung)

Die konkrete Aufgabenstellung war es, innerhalb von 40 Minuten unter der Führung einer bestimmten, vorher definierten Person mit den vorhandenen Materialien einen Turm zu bauen, der auf einem eigenem Fundament ohne andere Unterstützung stehen kann, mit einer Standfestigkeit, die ein Lineal zu tragen vermag (Antons, 2000, S. 132). Nach Ablauf der Arbeitszeit war der Turm an die Führungskraft zu übergeben und gleich darauf ein Fragebogen auszufüllen. Nachdem alle Personen mit dem Ausfüllen fertig waren, informierte die Auskunftsperson alle Geführten über die jeweiligen Räume, in denen sie sich für die zweite Runde einzufinden hatten. Die Führungspersonen wurden informiert, dass sie keine Ortswechsel zu vollziehen haben, sondern die zweite Runde wieder im selben Raum zu absolvieren haben. Dadurch konnte gewährleistet werden, dass zügig und kontinuierlich in den Gruppen gearbeitet werden konnte. Tabelle 10 zeigt die kommunizierten Arbeitsaufträge der UntersuchungsteilnehmerInnen über den gesamten Verlauf.

Führungskraft	Geführte
<ul style="list-style-type: none"> • Bereitstellung der Informationen zur Aufgabenstellung an die Gruppe • Übergabe der Materialien an die Gruppe • Führung der Gruppe während der Bearbeitung der Aufgabenstellung • Verantwortung für die Zeiteinhaltung • Übergabe der Gruppenleistung (des Turms) an die Aufsichtsperson • Zuteilung der Gruppenmitglieder für die zweite Runde und Bereitstellung der Information über den jeweiligen Gruppenraum 	<ul style="list-style-type: none"> • Annahme des Arbeitsauftrags • Durchführung der Gruppenübung • Leistungserbringung als Gruppe und Anweisungen der Führungskraft Folge leisten • Gewissenhafte Ausfüllung eines Fragebogens nach jeder Runde • Pünktliches Einfinden in den kurzfristig zugeteilten Räumen

Tabelle 10: Arbeitsaufträge der Führungskräfte und Geführten im Rahmen des Experiments
(Quelle: Eigene Darstellung)

Nach kurzer Zeit fanden sich die TeilnehmerInnen in den vorgesehenen Gruppenräumen ein und führten ihren Gruppenauftrag wieder innerhalb von 40 Minuten durch. Auch hier wurde im Anschluss der vom Autor dieser Dissertation entwickelte Fragebogen ausgegeben (siehe auch Kapitel 3.2.3). Nach dem Einsammeln der letzten Fragebögen der Runde 2 wurden die Mensagutscheine ausgegeben und nochmals der Dank für die Teilnahme ausgesprochen. Abbildung 13 verdeutlicht nochmals den Ablauf des Experiments.

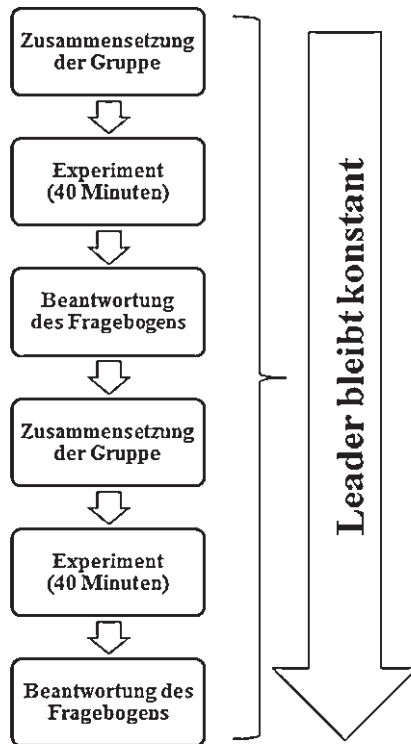


Abbildung 13: Ablauf des Experiments
(Quelle: Eigene Darstellung)

Die folgenden Fakten sind für die wissenschaftliche Erhebung unter experimentellen Bedingungen entscheidend und werden an dieser Stelle nochmals hervorgehoben:

- Grundsätzlich wurden je 2 Vergleichsgruppen pro Nationalkultur, bezogen auf die Führungsperson, gebildet. Jede Gruppe bestand aus neun Geführten und einer Führungskraft. Bei vier unterschiedlichen Nationen bzw. Kulturen nahmen insgesamt 72 Geführte und 8 Führungspersonen teil, die jeweils in bzw. mit homogenen und multikulturellen Gruppen arbeiteten.

- Es fungierten ausschließlich männliche Personen als Führungskraft. Um Gender-Effekte auszuschließen wurde bewusst nur ein Geschlecht beim Einsatz der Führungspersonen berücksichtigt. Als Geführte nahmen jedoch sowohl männliche als auch weibliche Studierende teil.
- Alle Studierende hatten einen gemeinsamen Hintergrund: sie waren zum Zeitpunkt des Experiments hauptberuflich Studierende an der WU Wien (Vollzeit- oder Austauschstudierende).
- Die jeweilige Führungskraft hatte ebenso zwei Runden zu absolvieren: einmal führte sie eine homogene Gruppe mit Personen, die über die gleiche Nationalkultur verfügten und anschließend leitete sie eine multikulturelle Gruppe, bei der alle Mitglieder einen anderen kulturellen Hintergrund als die Führungskraft hatten. Die Führungspersonen sind somit als Konstante zu betrachten.
- Um Lerneffekte zu vermeiden, wurde in der Vergleichsgruppe 1 zuerst eine homogene Gruppe und im Anschluss eine multikulturelle geführt. In der Vergleichsgruppe 2 arbeiteten die TeilnehmerInnen zuerst in einem heterogenen Setting, um anschließend in einer homogenen Gruppe den Arbeitsauftrag zu bewältigen.
- Die Aufgabenstellung blieb in allen Gruppen und jeder Runde gleich.
- Nach jeder absolvierten Runde füllten die Gruppenmitglieder einen Fragebogen aus.
- Alle Arbeitsgruppen konnten die Aufgabenstellung erfolgreich und zeitgerecht durchführen.

Im Folgenden werden nun die Methoden der Datenauswertung, die in dieser Dissertation angewandt werden, dargestellt.

3.3.4 Datenauswertung

Im Rahmen der Datenauswertung wurden unterschiedliche statistische Auswertungsverfahren mittels der Statistik- und Analyse-Software „R“ durchgeführt. Grundsätzlich ist R die state-of-the-art Statistiksoftware, Open Source und frei verfügbar (Hatzinger / Hornik / Nagel, 2011, S. 11). Neben dem Einsatz für statistische Herausforderungen wird sie vor allem als flexible (Software)Umgebung für Datenanalyse und Grafik verwendet. Aufgrund der unterschiedlichen Vortei-

le entwickelte sich diese Statistik- und Analysesoftware über die Jahre zum de facto Standard der statistischen Forschung an Universitäten (Hatzinger / Hornik / Nagel, 2011, S. 18). Im Folgenden werden nun die Faktorenanalyse, die Reliabilitätsanalyse, die Mittelwertvergleichstests (t-Tests) / Binomialtests und die Korrelationsanalyse als Auswertungsmethoden dieser Dissertation dargestellt.

3.3.4.1 Reliabilitätsanalyse

In dieser Dissertation wird eine der gebräuchlichsten Methoden der Reliabilitätsschätzung zur Berechnung der internen Konsistenz und Zuverlässigkeit von Tests verwendet: der Alpha-Koeffizient von Cronbach (1951) (Diekmann, 2009, S. 254). Alpha wird aus den Korrelationen aller Items untereinander berechnet und ist sowohl auf dichotome als auch auf polytome Items anwendbar (Bortz / Döring, 2003, S. 198).¹⁸ Die Anzahl der und die Interkorrelation der Items stellen eine Funktion Alphas dar (Schnell / Hill / Esser, 2005, S. 153).

Bei der Erfassung des Cronbach's Alpha zur Messung der internen Konsistenz von Fragebogenskalen ist entscheidend, dass Größen nahe Eins auf Reliabilität hindeuten. Dabei orientieren sich die meisten ForscherInnen an der Vorgabe von Nunnally (1978, S. 52ff.), der einen Wert von mindestens 0,7 fordert, der im Allgemeinen als akzeptabel gilt (Algesheimer / Herrmann / Dimpfel, 2006, S. 942). Empirische Werte über 0,8 werden als gut, Alpha Werte über 0,9 als exzellent bezeichnet (Schnell / Hill / Esser, 2005, S. 153).

3.3.4.2 Faktorenanalyse

Die Faktorenanalyse ist ein statistisches Verfahren, mit dem versucht wird, eine große Zahl von Untersuchungsvariablen durch eine kleinere Anzahl von Faktoren zu erklären. (Schnell / Hill / Esser, 2005, S. 161f.; Bortz / Döring, 2003, S. 383). Eine Faktorenanalyse wird insbesondere dann angewandt, wenn eine Vielzahl von Variablen zu einer bestimmten Fragestellung erhoben wird und im Anschluss die Reduktion oder Bündelung der Variablen im Vordergrund steht (Backhaus et al., 2001, S. 19). „Das allgemeine Ziel der Faktorenanalyse besteht

18 Dichotome Merkmale sind Merkmale, die zwei Abstufungen haben. So werden in empirischen Erhebungsinstrumenten bei dichotomen Items nur zwei Antwortmöglichkeiten gegeben. Polytome oder mehrkategoriale Merkmale beinhalten zumindest vier oder mehr Ausprägungen (Bortz / Döring, 2003, S. 139).

darin, korrelierende Variablen auf höherer Abstraktionsebene zu Faktoren zusammenzufassen.“ (Bortz / Döring, 2003, S. 383) Somit kann diese Analysemethode als datenreduzierendes Verfahren bezeichnet werden.

Bei der Faktorenanalyse werden Faktorwerte herangezogen, die die Variableninformationen abbilden. Neben der pragmatischen Funktion hat dieses statistische Auswertungsverfahren auch einen sogenannten heuristischen Vorteil: die Werte ermöglichen es, für die faktoriellen Variablenbündel sinnvolle Zusammenhänge und Interpretationen zu finden (Bortz / Döring, 2003, S. 383).

Grundsätzlich kann zwischen explorativen und konfirmatorischen Faktorenanalysen unterschieden werden. Bei der explorativen Analyse wird das Ziel verfolgt, einerseits die Anzahl der Faktoren und die Zusammenhänge zwischen diesen (Faktorenkorrelationen) und andererseits zwischen den Faktoren und den Variablen (Faktorenloadungen) aus dem Datensatz zu ermitteln (Schnell / Hill / Esser, 2005, S. 162). Die Faktorladung wird aus jeder Variable berechnet, die den Zusammenhang zwischen der Indikatorvariable und dem Faktor bzw. latentem Merkmal zeigt. Der Wertebereich von Faktorloadungen kann von -1 bis +1 betragen und dient auch als Gewichtungsfaktor (Bortz / Döring, 2003, S. 146f.). Bei den konfirmatorischen Faktorenanalysen ist die Festlegung einiger Größen, wie beispielsweise die Anzahl der Faktoren, die Beziehungen zwischen Faktoren oder etwa die Beziehungen zwischen Faktoren und beobachtbaren Variablen erforderlich (Schnell / Hill / Esser, 2005, S. 162). Sie dienen dazu, ein sogenanntes Messmodell, das Zusammenhänge zwischen Indikatoren und latenter Variable spezifiziert, zu überprüfen (Bortz / Döring, 2003, S. 147).

Zusammenfassend ist hervorzuheben, dass bei explorativen Faktorenanalysen Faktorloadungen als Gewichte aus den Daten errechnet werden. Im Unterschied dazu werden bei konfirmatorischen Analysen Gewichtungsfaktoren vorgegeben und auf ihre Gültigkeit geprüft (Bortz / Döring, 2003, S. 147).

3.3.4.3 t-Tests / Binomialtest / Varianzanalyse

Zur Hypothesenüberprüfung kommen neben Faktorenanalysen t-Tests, Binomialtests und Varianzanalysen zum Einsatz. Bei einem t-Test, also einem Mittelwertdifferenzentest, erfolgt die Verteilung der Differenzen von Stichprobenmittelwerten unabhängiger Stichproben aus derselben Grundgesamtheit (Schnell / Hill / Esser, 2005, S. 451). Bortz / Döring (2003, S. 498) verweisen darauf, dass

mit einem einseitigen Signifikanztest eine gerichtete und mit einem zweiseitigen Signifikanztest eine ungerichtete Alternativhypothese geprüft wird.¹⁹

Die statistische Auswertungsmethode Binomialtest wird zur Berechnung der Wahrscheinlichkeit für das Zutreffen von Null- bzw. Alternativhypothesen verwendet. Dieser Test ist grundsätzlich bei kleinen Stichproben zu bevorzugen, denn es werden jedenfalls die „exakten“ Wahrscheinlichkeiten berechnet (Hatzinger / Hornik / Nagel, 2001, S. 176ff.). Die Varianzanalyse oder ANOVA („analysis of variance“) wird dann eingesetzt, wenn die Unterschiede in den Mittelwerten zwischen zwei und mehr Gruppen von Interesse sind (Schnell / Hill / Esser, 2005, S. 457). „Die (einfache) Varianzanalyse ist eine Verallgemeinerung der t-Tests zur Signifikanzprüfung der Unterschiede zwischen Mittelwerten aus mehr als zwei Gruppen.“ (Diekmann, 2009, S. 695) Neben der einfachen Varianzanalyse gibt es noch weitere Formen, wie etwa die Kovarianzanalyse, die eine einfache Kombination aus multipler Regression und Varianzanalyse darstellt, oder auch die multivariate Varianzanalyse (MANOVA). Dabei werden mehrere abhängige Variablen gleichzeitig untersucht (Schnell / Hill / Esser, 2005, S. 457f.; Bortz / Döring, 2003, S. 530f.).

3.3.4.4 Korrelation

Die Berechnung von Korrelationen zwischen allen Variablen ermöglicht die Berechnung der Stärke der Beziehungszusammenhänge untereinander. Wird aufgrund der Datenstruktur ein Beziehungszusammenhang vermutet und möchte man diesen überprüfen, ist es erforderlich, die Quantifizierung von Beziehungen zwischen Variablen zu ermitteln. Das entsprechende statistische Maß stellt der

19 Grundsätzlich ist bei der Formulierung von Hypothesen in Bezug auf Variablenbeziehungen ein möglichst hoher Präzisionsgrad anzustreben. Eine fundamentale Form der Präzisierung von Variablenbeziehungen besteht in der Angabe der Richtung einer Unterschieds- oder Zusammenhangsrelation. Ungerichtete Hypothesen verdeutlichen, dass das Gegenstandsverständnis unter Umständen nicht ausreicht und verbessert werden muss. Schließlich soll die Aufstellung und Prüfung einer theoretisch und empirisch fundierten gerichteten Hypothese gelingen (Bortz / Döring, 2003, S. 11). Bei der statistischen Hypothesenprüfung spricht man von einem Hypothesenpaar, bestehend aus einer Nullhypothese (H0) und einer Alternativhypothese (H1). Die Alternativhypothese stellt die Forschungshypothese dar, während die Nullhypothese der Alternativhypothese widerspricht (Bortz / Döring, 2003, S. 28).

Korrelationskoeffizient dar (Backhaus et al., 2011, S. 332f.; Bortz / Döring, 2003, S. 508).

Nach der Darstellung der Methoden der Datenauswertung erfolgt nun die Auswertung, Darstellung und Interpretation der Ergebnisse der empirischen Untersuchung.