

Born, S., Loßnitzer, T. & Schmidt, B. (2006).
Lehrveranstaltungsevaluation an der Friedrich-Schiller-Universität
Jena – Eine Analyse der Dimensionalität der eingesetzten
Fragebögen. In B. Krause & P. Metzler (Hrsg.), *Empirische
Evaluationsmethoden* (Band 10, S. 99 - 116). Berlin: ZeE Verlag.

8. Lehrveranstaltungsevaluation an der Friedrich-Schiller-Universität Jena – Eine Analyse der Dimensionalität der eingesetzten Fragebögen

Sebastian Born, Tim Loßnitzer & Boris Schmidt (Jena)

8.1. Einleitung

Das Universitätsprojekt Lehrevaluation wurde im Jahre 1997 durch die damalige Universitätsleitung der Friedrich-Schiller-Universität Jena gegründet. Mit der Gründung war der Gedanke verbunden, eine einheitliche, systematische und wissenschaftlich fundierte Evaluation der Lehre zu etablieren. Dem Projekt kommt an der FSU Jena eine Vielfalt an Aufgabenbereichen zu: Evaluation von Lehrveranstaltungen, Entwicklung und Umsetzung von Konzepten für das Qualitätsmanagement in der Lehre, Befragungen zur Lehr- und Studiensituation, Forschungsaufgaben sowie die Unterstützung weiterer Evaluationsvorhaben. Der Schwerpunkt der Arbeit des Universitätsprojekts Lehrevaluation liegt hierbei in der zentralen Durchführung der Lehrveranstaltungsevaluation. In der ersten Projektphase nach der Einrichtung des Projekts lag die Zielrichtung daher zunächst darin, neben der Entwicklung geeigneter Fragebögen zur Lehrveranstaltungsevaluation, eine Projektstruktur aufzubauen und die regelmäßige Evaluation von Lehrveranstaltungen universitätsweit zu ermöglichen.

Mit aktuell rund 400 pro Semester evaluierten Veranstaltungen ist das Evaluationsangebot des Universitätsprojekts Lehrevaluation das universitätsweit meistgenutzte Verfahren zur Evaluation von Lehrveranstaltungen. Insgesamt werden auf diese Weise je Semester rund 11.000 Veranstaltungsbewertungen aus Studierenden- und Lehrendensicht erfasst und ausgewertet. Darüber hinaus nutzen mehrere Fachschaften bei den von ihnen organisierten Verfahren zur Lehrevaluation das technische System und die Fragebögen des Universitätsprojekts Lehrevaluation. Mit einer weiterhin steigenden Anzahl an teilnehmenden Lehrveranstaltungen leistet das Universitätsprojekt wichtige Beiträge zur Qualitätssicherung und -entwicklung an der Friedrich-Schiller-Universität Jena. Loßnitzer, Schmidt & Born (eingereicht) geben einen ausführlichen Überblick über die Aufgabenbereiche und die wichtigsten aktuellen Forschungsvorhaben des Universitätsprojekts Lehrevaluation.

8.1.1. Entwicklung der Fragebögen zur Lehrveranstaltungsevaluation

Bereits mit der Gründung des Universitätsprojekts Lehrevaluation wurde anstelle der Übernahme eines bereits existierenden Standardfragebogens zur Lehrevaluation die eigenständige (Weiter-)Entwicklung geeigneter Evaluationsinstrumente angestrebt. Grundlage für die Fragebogenkonstruktion zur Lehrveranstaltungsevaluation war die Konzeption eines tragfähigen Qualitätsmodells der Lehre. Das Qualitätsmodell des Universitätsprojekts Lehrevaluation definiert die Qualität einer Lehrveranstaltung multidimensional als Zusammenspiel von unmittelbar oder mittelbar durch die Beteiligten (Dozent und Studierende) veränderbaren

Prozessvariablen einerseits und den angestrebten bzw. erreichten Ergebnissen der Lehrveranstaltung andererseits (Buhl, 1999a).

Abbildung 1 stellt das Qualitätsmodell schematisch dar.

Qualität einer Lehrveranstaltung	
Zusammenwirken von Prozess- und Ergebnisvariablen	
Prozessvariablen	Ergebnisvariablen
Prinzipiell veränderbar, bilden den Prozess der Lehrveranstaltung ab	Veranstaltungsspezifisch, bilden Ergebnisse der Lehrveranstaltung ab
<ul style="list-style-type: none"> ▶ Rahmenbedingungen (Raum, Ausstattung, Lärm) ▶ Verhalten des Dozenten (Aufbereitung und Darstellung des Stoffes, Leitungsfunktion) ▶ Verhalten der Studierenden (Teilnahmeregelmäßigkeit, Aufmerksamkeit, Lerneinsatz) 	<ul style="list-style-type: none"> ▶ Gesamteindruck (Attraktivität der Lehrveranstaltung) ▶ Zuwachs an Wissen, Fertigkeiten, Kompetenzen (Fachwissen, praktische Anwendung, Fachübergreifendes Denken)

Abbildung 1: Qualitätsmodell der Lehre

Als Prozessvariablen einer Lehrveranstaltung werden beispielsweise die Rahmenbedingungen, das Verhalten der Studierenden und des Dozenten verstanden; sie bilden den Prozess einer Lehrveranstaltung ab und sind prinzipiell veränderbar. Demgegenüber beschreiben die Ergebnisvariablen einer Lehrveranstaltung die Wirkungen dieser Veranstaltung, wie beispielsweise den Zuwachs an Kompetenzen, Fertigkeiten und Wissen seitens der Studierenden, das Interesse an den in der Lehrveranstaltung behandelten Themen oder auch den Gesamteindruck.

Integraler Bestandteil des Qualitätsmodells ist der Gedanke eines Perspektivenaustausches zwischen Lehrenden und Studierenden. Vor diesem Hintergrund wurden neben den Fragebögen für Studierende auch Fragebögen für Dozenten entwickelt. Sowohl die Studierenden als auch der Dozent beurteilen den (bisherigen) Verlauf der Lehrveranstaltung anhand der im Fragebogen aufgeführten Bewertungsaspekte (IST-Perspektive Studierende und IST-Perspektive Dozent), zusätzlich definiert der Dozent das von ihm in der jeweiligen Lehrveranstaltung angestrebte Niveau der Ergebnisvariablen (SOLL-Perspektive des Dozenten), sodass insgesamt drei Perspektiven erhoben werden. Der Abgleich dieser drei Perspektiven stellt wiederum eine Grundlage für die Kommunikation zwischen dem Dozenten und der Studierenden über den Ablauf der Veranstaltung dar.

Basierend auf diesem Qualitätsmodell der Lehre wurden zu Beginn der Entwicklungsphase Items generiert, die in einem mehrzyklischen Prozess unter Beteili-

gung von Studierenden und Lehrenden verschiedener Fakultäten der FSU Jena selektiert und überarbeitet wurden.

Während der Pilotphase ab dem Wintersemester 1998/99 wurden kritische Items mehrfach modifiziert, bis im Sommersemester 2000 der Regelbetrieb mit einer Reihe von veranstaltungs- (z.B. Vorlesung vs. Übung) sowie fachspezifischen (z.B. Mathematik vs. Psychologie) Fragebogenversionen aufgenommen wurde. Im Wintersemester 2003/04 erfolgte unter anderem aus ökonomischen Erwägungen eine erste grundlegende Revision der eingesetzten Fragebögen mit dem Ziel, die Anzahl der unterschiedlichen Fragebogenvarianten auf möglichst wenige veranstaltungsspezifische Versionen zu reduzieren und diese so universell zu formulieren, dass keine fächerspezifische Auftrennung mehr erforderlich ist. Hierbei wurde nicht nur an dem ursprünglichen Konzept der Prozess- und Ergebnisvariablen festgehalten, sondern auch der erfolgreiche Gedanke des Perspektivenaustauschs beibehalten. Die modifizierten Itemformulierungen wurden von Lehrenden, Studierenden und Mitarbeitern des Universitätsprojekts Lehrvaluation einem umfangreichen Expertenreview unterzogen. Am Ende des Revisionsprozesses standen drei veranstaltungsspezifische Fragebogenversionen für Vorlesungen, für Seminare sowie für Übungen, die sich durch einen hohen Anteil an veranstaltungsunspezifischen, fächerübergreifende Items auszeichnen und darüber hinaus einige ergänzende veranstaltungsspezifische Aspekte speziell für die Evaluation von Vorlesungen, Seminaren bzw. Übungen enthalten. Für alle drei Fragebogenversionen wurden entsprechend des angestrebten Perspektivenaustausches neben den Studierendenbögen auch entsprechende Dozentenbögen entwickelt, welche dieselben Bewertungsaspekte in perspektivenspezifischer Formulierung enthalten.

Die Beantwortung der Items erfolgt auf einer fünfstufigen Likert-Skala mit den Antwortpolen „*stimme nicht zu*“ bis „*stimme zu*“. Die Bewertung der durch die Lehrveranstaltung erworbenen Kompetenzen (siehe Kapitel 8.1.2.2.) erfolgt auf einer ebenfalls fünfstufigen Likert-Skala mit den Antwortpolen „*wenig*“ bis „*viel*“. Zusätzlich besteht bei allen Items die Möglichkeit die Antwortalternative „*keine Angabe*“ (k. A.) zu wählen, wenn die Frage für den Studierenden nicht zu beantworten ist oder aber dieser keine Antwort geben möchte. Seit dem Sommersemester 2004 werden die revidierten Fragebögen des Universitätsprojekts Lehrvaluation eingesetzt; sie werden derzeit in mehreren Studien hinsichtlich ihrer psychometrischen Güte analysiert (Born, in Vorb.; Born & Loßnitzer, 2006; Loßnitzer, in Vorb.; Loßnitzer & Born, 2005).

Die in dieser Studie im Vordergrund stehende Analyse der Dimensionalität erfolgt exemplarisch am „Fragebogen für Seminare und Veranstaltungen mit Teilnehmerbeiträgen“, im Folgenden kurz „Seminarfragebogen“ genannt. Eine kurze Beschreibung dieses Fragebogens soll den grundlegenden Aufbau des Fragebogens im Detail verdeutlichen. Dabei werden sowohl die veranstaltungsübergreifenden als auch die veranstaltungsspezifischen Bewertungsaspekte dargestellt.

8.1.2. Bewertungsaspekte der Fragebögen zur Lehrveranstaltungsevaluation

Das zugrunde liegende Qualitätsmodell der Lehre postuliert fünf theoretische, veranstaltungsübergreifende Dimensionen, die sich in Prozess- und Ergebnisvariablen aufteilen lassen: (1) Gesamteindruck sowie (2) Kompetenzerwerb als Ergebnisvariablen, und (3) Rahmenbedingungen, (4) Dozentenverhalten und (5) Studierendenverhalten als Prozessvariablen (vgl. Abbildung 1). Je nach Veranstaltungstyp enthalten die Fragebögen zusätzliche, veranstaltungsspezifische Bewertungsaspekte. Im Falle des Seminarfragebogens werden damit einige für diesen Veranstaltungstyp spezifische Aspekte aufgegriffen, wie beispielsweise die thematische Einordnung von Referaten der Studierenden in Relation zu Inhalt und Zielen des Seminars („Der Dozent/die Dozentin macht Inhalte und Ziele der Studierendenbeiträge klar.“). Des Weiteren werden, wie in allen Fragebogenvarianten, demographische Daten der Studierenden, die Hauptgründe für den Veranstaltungsbesuch und der Arbeitsaufwand erfragt. Zudem beinhalten die Fragebögen für die Bewertungsdimensionen „Arbeitsaufwand“, „Gesamteindruck“, „Kompetenzerwerb“, „Dozentenverhalten“, „Studierendenverhalten“ sowie für „Rahmenbedingungen“ je ein zusammenfassendes Zufriedenheitsitem (z.B.: „Insgesamt gesehen bin ich mit den in dieser Veranstaltung erworbenen Kompetenzen zufrieden.“). Optional können zudem von der Veranstaltungsleitung individuell weitere Items („freie Items“) festgelegt werden. Tabelle 1 gibt am Beispiel des Seminarfragebogens eine Übersicht über die einzelnen veranstaltungsübergreifenden und veranstaltungsspezifischen Bewertungsaspekte.

Im Fokus dieses Beitrages stehen die durch das Qualitätsmodell postulierten fünf Bewertungsdimensionen: (1) Gesamteindruck, (2) Kompetenzerwerb, (3) Rahmenbedingungen, (4) Dozentenverhalten und (5) Studierendenverhalten. Diese werden im Folgenden anhand des Seminarfragebogens für Studierende näher beleuchtet. Der entsprechende Fragebogen für Dozenten weist dieselben fünf Bewertungsdimensionen in der perspektivenspezifischen Formulierung aus Sicht der Veranstaltungsleitung auf, wird jedoch an dieser Stelle nicht näher betrachtet (siehe Loßnitzer & Born, 2005).

Tabelle 1: Bewertungsaspekte des Seminarfragebogens und deren Operationalisierung anhand eines Beispielitems.

Veranstaltungsübergreifende Bewertungsaspekte		
▶ Demographische Daten	2 Items + 1 Gesamtitem	▶ Ich empfinde den von mir für diese Veranstaltung zu erbringenden Arbeitsaufwand als angemessen.
▶ Hauptgründe für Veranstaltungsbesuch		
▶ Arbeitsaufwand		
▶ Qualitative Bemerkungen (offene Antworten)		
▶ Optional: freie Items		
Postulierte Bewertungsdimensionen - <i>veranstaltungsübergreifend</i>		
▶ Gesamteindruck	5 Items + 1 Gesamtitem	▶ Der behandelte Stoff knüpft an meinen bisherigen Wissensstand an.
▶ Kompetenzerwerb	8 Items + 1 Gesamtitem	▶ Wissen über Theorien und Modelle
▶ Rahmenbedingungen	5 Items + 1 Gesamtitem	▶ Die räumlichen Gegebenheiten (Größe, bauliche Qualität, Lage, ...) sind für diese Veranstaltung ausreichend.
▶ Dozentenverhalten	8 Items + 1 Gesamtitem	▶ Der Dozent/die Dozentin hat Ziele und Struktur nachvollziehbar dargestellt.
▶ Studierendenverhalten	4 Items + 1 Gesamtitem	▶ Die meisten Teilnehmenden bereiten sich auf die einzelnen Termine angemessen vor.
Postulierte Bewertungsdimensionen - <i>veranstaltungsspezifisch für Seminare</i>		
▶ Dozentenverhalten	4 Items	▶ Der Dozent/Die Dozentin macht Inhalte und Ziele der Studierendenbeiträge klar.
▶ Studierendenbeiträge (Referate, Präsentationen, ...)	4 Items	▶ Die meisten Teilnehmerbeiträge (Referate, Präsentationen, etc.) sind didaktisch gut aufbereitet (Strukturierung, Anschaulichkeit, etc.).

8.1.2.1. Gesamteindruck

Die Dimension „Gesamteindruck“ beschreibt die Gesamtzufriedenheit der Studierenden mit der Lehrveranstaltung und umfasst insgesamt fünf Items. Seitens der Studierenden soll beispielsweise das eigene Interesse am Thema oder aber die wahrgenommene Angemessenheit des fachlichen Niveaus beurteilt werden.

8.1.2.2. Kompetenzerwerb

Mit Hilfe von insgesamt acht Items beurteilen die Studierenden das individuelle Ausmaß der durch den Veranstaltungsbesuch erworbenen Kompetenzen. Betrachtet man die Dimension „Kompetenzerwerb“ aus einer theoretischen Perspektive genauer, so wird deutlich, dass dem zugrunde liegenden Begriff „Kompetenz“ keine einheitliche Bedeutung zukommt, sondern von unterschiedlichen Autoren zahlreiche Differenzierungen erfährt (vgl. Braun, 2006; Braun & Gusy,

2005). Weinert (2002, S. 27) definiert Kompetenzen als „...die bei Individuen verfügbaren oder durch sie erlernbaren kognitiven Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten, um die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können.“ Stützt man sich auf diese Definition, so zeigt sich, dass es sich bei Kompetenzen um mehr als fachbezogenes Wissen zu einem bestimmten Thema handelt. Der Begriff „Kompetenz“ beinhaltet neben der fachlichen Seite auch methodische, personale und soziale Aspekte (Braun, 2006; Braun & Gusy, 2005). In der Literatur werden zahlreiche weitere, über das viergliedrige Kompetenzmodell hinaus gehende Facetten benannt, deren theoretischer Status jedoch bislang ungeklärt ist (wie z.B. „Innovationskompetenz“, „Führungskompetenz“, „Europa-kompetenz“, die auch als Kombinationen der vier Grundkompetenzen zu verstehen sind, vgl. Schmidt, in Vorb.). Das viergliedrige Modell wird daher auch an dieser Stelle als die theoretisch und empirisch am besten abgesicherte Konzeption herangezogen.

Fachkompetenz

Bloom (1956) unterteilt die Fachkompetenz in sechs hierarchisch geordneten Hauptklassen: (1) Wissen, (2) Verstehen, (3) Anwendung, (4) Analyse, (5) Synthese und (6) Bewertung. Mittels einer Validierungsstudie konnten Madaus, Woods und Nuttall (1973) die ersten vier Stufen dieses theoretischen Konzepts verifizieren. Braun und Gusy (2006) beschrieben diese vier Stufen wie folgt: *Wissen* als Kenntniserwerb und dessen Reproduktion, *Verstehen* als eine intensivere Auseinandersetzung mit dem Stoff und das Wiedergeben in eigenen Worten, *Anwenden* als Wissen um die Vorgehensweise in einer Situation, die für den Gegenstandsbereich typisch ist, und *Analysieren* als Abwägen des Nutzens und Auswahl des optimalen Gegenstandsbereichs für eine Fragestellung.

Methodenkompetenz

Unter Methodenkompetenz versteht man jenen Bereich der Kompetenzen einer Person, der sich auf die Fähigkeit bezieht, sich (Fach-)Wissen und Fertigkeiten anzueignen sowie diese zur Lösung konkreter Fragestellungen bzw. Problemstellungen anzuwenden, wie zum Beispiel im Falle von Methoden zur Arbeitsplanung, zur Präsentation oder zum Projektmanagement. Die Methodenkompetenz ist von der eigentlichen Fachkompetenz weitestgehend unabhängig; sie ist jedoch mitverantwortlich dafür, Fachkompetenz aufzubauen und sie erfolgreich zu nutzen (vgl. Krems, 2004).

Soziale Kompetenz

Soziale Kompetenz umfasst, ähnlich wie die vielfältige Fachkompetenz, ein breites Spektrum menschlicher Fähigkeiten und Fertigkeiten. Kanning (2003) definiert diese als die „Gesamtheit des Wissens, der Fähigkeiten und Fertigkeiten einer Person, welche die Qualität eigenen Sozialverhaltens - im Sinne der Definition sozial kompetenten Verhaltens - fördert.“ (S. 15). Auch bei näherer Betrachtung

der Definition von sozial kompetentem Verhalten bleibt der Begriff „Soziale Kompetenz“ unscharf; er bedarf einer weiteren Differenzierung und Konkretisierung. Braun (2006) unterteilt soziale Kompetenz beispielsweise weiter in Kommunikations-, Kooperations- und Konfliktfähigkeit. Im Rahmen dieses Beitrages sollen diese drei Facetten insgesamt als Sozialkompetenz verstanden werden.

Personale Kompetenz

Erpenbeck & von Rosenstiel (2003) beschreiben Personale Kompetenz „als die Disposition einer Person, reflexiv selbstorganisiert zu handeln.“ (S. XVI). Unter Personaler Kompetenz versteht man beispielsweise Selbstorganisation, Lernfähigkeit und Selbstreflexion, also die Fähigkeit, die eigene Handlungssteuerung situationsadäquat einzusetzen und sich in dieser Fähigkeit kontinuierlich weiter zu entwickeln. Die Abgrenzung zur Methodenkompetenz liegt darin begründet, dass es bei der personalen Kompetenz stärker um die Regulation der Motivation und der konkreten Handlungsausführung in Situationen, sowie um das Wissen einer Person über ihr eigenes (aufgabenbezogenes) Handeln geht, während der Schwerpunkt der Methodenkompetenz in der Verfügbarkeit von Arbeitstechniken liegt.

Die Bewertungsdimension „Kompetenzerwerb“ des an dieser Stelle im Fokus stehenden Fragebogens weist diese Unterteilung in spezifische Kompetenzarten auf einem höheren Aggregationsniveau auf, als dies beispielsweise im Berliner Evaluationsinstrument für selbsteingeschätzte, studentische Kompetenzen der Fall ist (BEvaKomp; Braun, 2006). Dieses Evaluationsinstrument bezieht sich fast ausschließlich auf die in Lehrveranstaltungen erworbenen Kompetenzen und lässt demgegenüber die im Qualitätsmodell des Universitätsprojekts Lehrevaluation benannten Prozessvariablen weitgehend unberücksichtigt. Dennoch lassen sich die genannten vier Kompetenzbereiche in zusammengefasster Form auch im Seminarfragebogen wieder finden. So lassen sich die Formulierungen der Items „Wissen über Theorien und Modelle“, „Wissen über Fakten, Begriffe und Konzepte“, „Wissen über Forschungsverfahren und wissenschaftliche Methoden“ und „Anwendung von Theorien, Methoden und Konzepten“ recht trennscharf den Fachkompetenzen zuordnen; entsprechend den dominierenden Zielen der Hochschulausbildung liegt somit ein Schwerpunkt des Fragebogens im Bereich der fachlichen Kompetenz. Die restlichen vier Items der Bewertungsdimension „Kompetenzerwerb“ stellen eine Mischung aus Methoden-, Sozial- und Personal-kompetenz dar. Insbesondere das Item „Schlüsselkompetenzen (Präsentieren, Arbeiten im Team, Recherchieren, ...)“ beinhaltet mit Präsentieren und Recherchieren Aspekte von Methodenkompetenz, sowie mit der Formulierung „Arbeiten im Team“ zusätzlich einen Aspekt der sozialen Kompetenz. Der theoretischen Differenzierung der verschiedenen Arten von Kompetenz wird bei der folgenden Analyse der Dimensionalität Rechnung getragen.

8.1.2.3. Rahmenbedingungen

Die Dimension „Rahmenbedingungen“ beschreibt den Rahmen, innerhalb dessen die Lehrveranstaltung stattfindet, dies betrifft die räumlichen Gegebenheiten (Größe, bauliche Qualität), die Ausstattung, wie vorhandene Medien und Technik, der zeitliche Rahmen und die Begleitmaterialien, wie Literatur und Skripte. Die Beurteilung der Rahmenbedingungen erfolgt an insgesamt fünf Items.

8.1.2.4. Dozentenverhalten

An insgesamt acht veranstaltungsübergreifenden Items wie beispielsweise „*Der Dozent/die Dozentin greift inhaltliche Anregungen und Fragen der Teilnehmenden auf.*“ bewerten die Studierenden das konkrete Verhalten des Dozenten in der Lehrveranstaltung. Im Falle des Seminarfragebogens wird die Einschätzung des Lehrendenverhaltens um vier weitere, veranstaltungsspezifische Items ergänzt (z.B. „*Der Dozent/die Dozentin formuliert Kritik in fairer und konstruktiver Weise.*“).

8.1.2.5. Studierendenverhalten

Neben dem Verhalten des Dozenten beurteilen die Studierenden auch das Verhalten ihrer Kommilitonen. Anhand von insgesamt vier veranstaltungsübergreifenden Fragen und weiteren veranstaltungsspezifischen Formulierungen sollen beispielsweise die Regelmäßigkeit des Veranstaltungsbesuches oder die Vorbereitung auf Einzelsitzungen seitens der Studierenden bewertet werden.

Diese auf der Grundlage des Qualitätsmodell der Lehre basierenden fünf Bewertungsdimensionen werden im Folgenden hinsichtlich ihrer multidimensionalen Struktur analysiert, dies erfolgt mit Hilfe von Strukturgleichungsmodellen. Es werden insgesamt zwei Modelle untersucht: (1) Ein Fünf-Faktoren-Modell entsprechend der fünf Bewertungsdimensionen und (2) ein um Faktoren zweiter Ordnung erweitertes Fünf-Faktoren-Modell, welches sowohl den seminarspezifischen Erweiterungen der Bewertungsdimensionen „Dozentenverhalten“ und „Studierendenverhalten“ als auch der theoretischen Differenzierung der Bewertungsdimension „Kompetenzerwerb“ Rechnung trägt.

8.2. Methode

Im Folgenden wird die dieser Untersuchung zugrunde liegende Stichprobe beschrieben und das Analyseverfahren zur Überprüfung der Dimensionalität der Fragebögen dargestellt.

8.2.1. Stichprobe

Die vorliegende Untersuchung basiert auf einem Datensatz von $N = 2774$ Fragebögen bzw. Studierenden. Insgesamt wurden $N = 367$ Lehrveranstaltungen im Zeitraum von Sommersemester 2004 bis Sommersemester 2005 evaluiert.

8.2.2. Analyse der Dimensionalität

Die durch das Qualitätsmodell der Lehre theoretisch zugrunde gelegte Dimensionalität der Fragebögen wird mittels einer konfirmatorischen Faktorenanalyse überprüft. Die konfirmatorische Faktorenanalyse ermöglicht auf der Grundlage der Kovarianzstruktur der manifesten Variablen die Überprüfung einer zuvor postulierten Faktorenstruktur. Hierbei wird in einem ersten Schritt ein theoretisches Strukturgleichungsmodell formuliert, welches die Anzahl der latenten Variablen, das Beziehungsgefüge zwischen latenten und manifesten Variablen und das Beziehungsgefüge der manifesten Variablen untereinander spezifiziert.

In Strukturgleichungsmodellen sind die quadrierten Ladungen (λ^2) der manifesten Variablen ein Indikator für den Varianzanteil, der durch die latenten Variablen an den manifesten Variablen erklärt wird. Bei Modellen, in denen jede manifeste Variable nur auf einer latenten Variablen lädt, liegt eine Einfachstruktur vor, die es ermöglicht, die erklärten Varianzanteile als Reliabilitäten (r_n) der Einzelitems zu interpretieren. Entsprechend ist die Rest- oder Fehlervarianz einer manifesten Variablen ein Indikator für den Anteil der nicht erklärten Varianz bzw. kann als Unreliabilität jedes Einzelitems verstanden werden.

Die Kovarianzen zwischen den einzelnen latenten Variablen bzw. zwischen den manifesten Variablen erlauben eine Schätzung der vom Modell implizierten Varianz-Kovarianz-Matrix, die dann mit der empirischen Varianz-Kovarianz-Matrix verglichen wird. Fit-Indizes geben Auskunft über den Grad der Übereinstimmung der beiden Matrizen und damit über die Güte des getesteten Modells. Standardmäßig wird bei Strukturgleichungsmodellen der Chi-Quadrat-Test genutzt, um die Nullhypothese zu überprüfen, dass die empirische und die vom Modell implizierte Varianz-Kovarianz-Matrix gleich sind. Jedoch führen gerade sehr große Stichproben und die damit einhergehende exzessive Teststärke zur Verwerfung von akzeptablen Modellen (Hayduk, 1996). Für die vorliegende Untersuchung wird der RMSEA zur Interpretation der Modellpassung herangezogen, da dieser stichprobenunabhängig ist.

8.2.3. Modellspezifikation

In der vorliegenden Untersuchung zur Dimensionalität werden zwei Modelle getestet. In Modell 1 werden die fünf Bewertungsdimensionen „Gesamteindruck“, „Kompetenzerwerb“, „Rahmenbedingungen“, „Dozentenverhalten“ und „Studierendenverhalten“ als latente Variablen bzw. Faktoren erster Ordnung modelliert. Die manifesten Variablen entsprechen den im Fragebogen formulierten und den Bewertungsdimensionen zugehörigen Items. Das Qualitätsmodell postuliert die Qualität der Lehre als Zusammenspiel von Prozess- und Ergebnisvariablen (vgl. Abbildung 1); entsprechend dürfen die latenten Variablen miteinander kovariieren. Abbildung 2 stellt die konzeptuellen Beziehungen der latenten und manifesten Variablen in Form des postulierten Fünf-Faktoren-Modells dar.

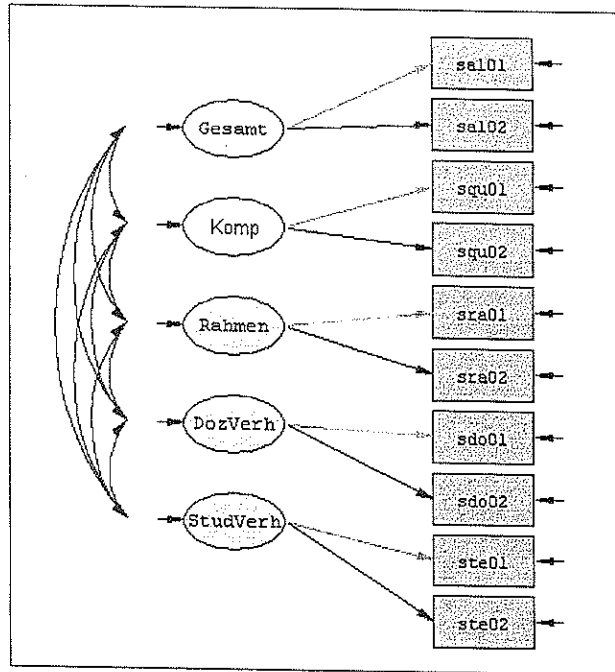


Abbildung 2: Lineares Strukturgleichungsmodell für das Fünf-Faktoren-Modell. Die latenten Variablen Gesamteindruck (Gesamt), Kompetenzerwerb (Komp), Rahmenbedingungen (Rahmen), Dozentenverhalten (DozVerh) und Studierendenverhalten (StudVerh) bilden die Bewertungsdimensionen des Fragebogens ab, auf denen die entsprechenden manifesten Variablen (sal, squ, sra, sdo und ste) laden.

Der Fokus der Untersuchung liegt auf dem erweiterten Modell 2, das sowohl den veranstaltungsspezifischen Bewertungsdimensionen „Dozentenverhalten“ und „Studierendenverhalten“ als auch der differenzierten Struktur der Dimensionen „Kompetenzerwerb“ Rechnung trägt (siehe Abbildung 3). Entsprechend werden die drei Bewertungsdimensionen „Kompetenzerwerb“, „Dozentenverhalten“ und „Studierendenverhalten“ durch je zwei Faktoren weiter spezifiziert. Für die beiden Dimensionen „Dozentenverhalten“ und „Studierendenverhalten“ werden jeweils ein allgemeiner (veranstaltungsübergreifend) und ein veranstaltungsspezifischer Faktor erster Ordnung angenommen (siehe Kapitel 8.1.2.4. und Kapitel 8.1.2.5.). Die Dimension „Kompetenzerwerb“ wird durch die beiden Faktoren erster Ordnung „Fachkompetenz“ (Kompetenz_1) und „sonstige Kompetenzen“ (Kompetenz_2) spezifiziert (siehe Kapitel 8.1.2.2.). In beiden Modellen werden die Fehler als unkorreliert angenommen.

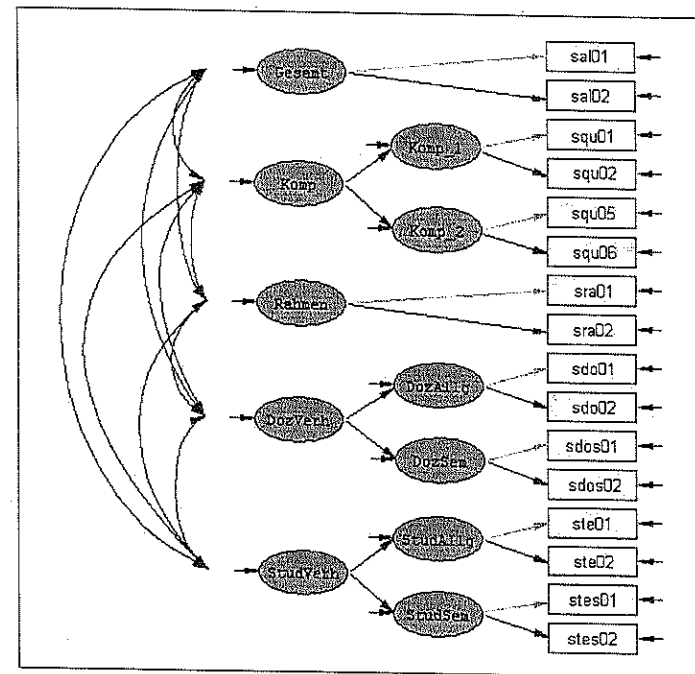


Abbildung 3: Lineares Strukturgleichungsmodell für das Modell mit Faktoren 2.Ordnung für den Seminarfragebogen. Die latenten Variablen Gesamteindruck (Gesamt), Kompetenzerwerb (Komp), Rahmenbedingungen (Rahmen), Dozentenverhalten (DozVerh) und Studierendenverhalten (StudVerh) bilden die Bewertungsdimensionen (Faktoren 2. Ordnung) des Fragebogens ab. Die latenten Variablen Fachkompetenz (Komp_1), sonstige Kompetenzen (Komp_2), allgemeines Dozentenverhalten (DozAllg), seminarspezifisches Dozentenverhalten (DozSem), allgemeines Studierendenverhalten (StudAllg) und seminarspezifisches Studierendenverhalten stellen die theoretisch begründeten Faktoren zweiter Ordnung dar. Auf den latenten Variablen laden die manifesten Variablen (sal, squ, sra, sdo und ste).

Alle postulierten Modelle werden mit Hilfe des Softwarepaketes LISREL 8.72 berechnet.

8.3. Ergebnisse

Die beiden in Kapitel 8.2.3. spezifizierten Modellen – Modell 1 und Modell 2 werden anhand der in Kapitel 8.2.1. beschriebenen empirischen Daten überprüft.

8.3.1. Modell 1: Fünf-Faktoren Modell mit Faktoren erster Ordnung

Für das Fünf-Faktoren-Modell ergibt sich ein Modell-Fit von $\chi^2 = 10739.53$, $df = 655$, $p = .000$, $RMSEA = 0.075$, der nach Browne und Cudeck (1993) als akzeptabel bezeichnet werden kann.

8.3.2. Modell 2: Fünf-Faktoren Modell mit Faktoren zweiter Ordnung

Für Modell 2, das Fünf-Faktoren-Modell mit Faktoren zweiter Ordnung () aufgrund des RMSEA eine akzeptable Passung ($\chi^2 = 6707.29$, $df = 652$, $p = 0.000$, $RMSEA = 0.058$).

Für die Dimension „Gesamteindruck“ liegen die Reliabilitäten der Items zwischen $r_u = 0.37$ und $r_u = 0.72$. Die Reliabilität des Items *sal02* („Der behandelte Stoff knüpft an meinen bisherigen Wissensstand an.“) ist im Vergleich hierzu mit $r_u = 0.20$ wesentlich geringer.

Für den Faktor „Fachkompetenz“ (Komp_1) liegen die Reliabilitäten der manifesten Variablen zwischen $0.34 \leq r_u \leq 0.48$. Die Reliabilitäten des Faktors „sonstige Kompetenzen“ (Komp_2) liegen mit $0.28 \leq \lambda \leq 0.55$ in einem ähnlichen Bereich. Beide Faktoren 1.Ordnung laden mit $\lambda = 0.74$ (Komp_1, Fachkompetenz) und $\lambda = 0.78$ (Komp_2, sonstige Kompetenzen) hoch auf dem Faktor 2. Ordnung „Kompetenzerwerb“.

Die Reliabilitäten der manifesten Variablen der Bewertungsdimension „Rahmenbedingungen“ stellen sich wesentlich differenzierter dar. Für die beiden Items *sra04* und *sra05* (Ausmaß und Brauchbarkeit der Begleitmaterialien) liegen die Reliabilitäten bei $0.64 \leq r_u \leq 0.71$. Dagegen weisen die Items *sra01*, *sra02* und *sra03* (räumliche Begebenheiten, Ausstattung und zeitlicher Rahmen) mit $0.06 \leq r_u \leq 0.14$ wesentlich niedrigere Reliabilitäten auf.

Betrachtet man die Reliabilitäten der Items für den Faktor „allgemeines Dozentenverhalten“ (DozAllg), so liegen diese zwischen $0.46 \leq r_u \leq 0.61$. Die Reliabilitäten des Faktors „seminarspezifisches Dozentenverhalten“ (DozSem) liegen mit Werten zwischen $r_u = 0.49$ und $r_u = 0.59$ in einem ähnlichen Bereich. Die Ladungen der beiden Faktoren 1. Ordnung laden nach Cohen (1977) mit $\lambda_{DozAllg} = 0.99$ und $\lambda_{DozSem} = 0.89$ sehr hoch auf dem Faktor 2. Ordnung „Dozentenverhalten“.

Die Reliabilitäten der manifesten Variablen für den Faktor „allgemeines Studierendenverhalten“ (StudAllg, Faktor 1. Ordnung) liegen im Bereich $0.49 \leq r_u \leq 0.69$. Lediglich das Item *ste01* (Regelmäßigkeit des Veranstaltungsbesuches) weist mit $r_u = 0,29$ eine vergleichsweise niedrige Reliabilität auf. Beim Faktor „seminarspezifisches Studierendenverhalten“ (StudSem, Faktor 1. Ordnung) liegen die Reliabilitäten zwischen $r_u = 0.56$ und $r_u = 0.69$. Bezogen auf die Ladungen der beiden Faktoren erster Ordnung auf den Faktor „Studierendenverhalten“ zeigt sich, dass der Faktor „allgemeines Studierendenverhalten“ (StudAllg) mit λ

= 0.88 wesentlich höher auf dem Faktor 2.Ordnung „Studierendenverhalten“ lädt als der Faktor „seminarspezifisches Studierendenverhalten“ (StudSem, $\lambda = 0.67$).

Die Interkorrelationen zwischen den fünf postulierten Dimensionen (Faktoren 2. Ordnungen) sind in Tabelle 2 aufgeführt, diese liegen $r = 0.40$ und $r = 0.79$.

Alle im Modell geschätzten Parameter sind signifikant auf einem Niveau von $\alpha = 0.05$.

Tabelle 2: Interkorrelationen der Faktoren zweiter Ordnung in Modell 2 (Fünf-Faktoren-Modell mit Faktoren zweiter Ordnung). Alle Interkorrelationen sind signifikant auf $\alpha = 0.05$.

Dimension	Gesamtein- druck	Kompetenz- erwerb	Rahmen- bedingungen	Dozenten- verhalten	Studieren- verhalten
Gesamtein- druck					
Kompetenz- erwerb	0.79				
Rahmen- bedingungen	0.48	0.42			
Dozenten- verhalten	0.74	0.73	0.56		
Studierenden- verhalten	0.61	0.73	0.40	0.65	

Abbildung 4: Pfaddiagramm für das Fünf-Faktoren-Modell mit Faktoren 2. Ordnung für den Seminarfragebogen, standardisierte Lösung. Die latenten Variablen Gesamteindruck (Gesamt), Kompetenzerwerb (Komp), Rahmenbedingungen (Rahmen), Dozentenverhalten (DozVerh) und Studierendenverhalten (StudVerh) bilden die Bewertungsdimensionen (Faktoren 2. Ordnung) des Fragebogens ab. Die latenten Variablen Fachkompetenz (Komp_1), sonstige Kompetenzen (Komp_2), allgemeines Dozentenverhalten (DozAllg), seminarspezifisches Dozentenverhalten (DozSem), allgemeines Studierendenverhalten (StudAllg) und seminarspezifisches Studierendenverhalten stellen die theoretisch begründeten Faktoren zweiter Ordnung dar. Auf den latenten Variablen laden die manifesten Variablen (*sal*, *sq*, *sra*, *sdo* und *ste*).

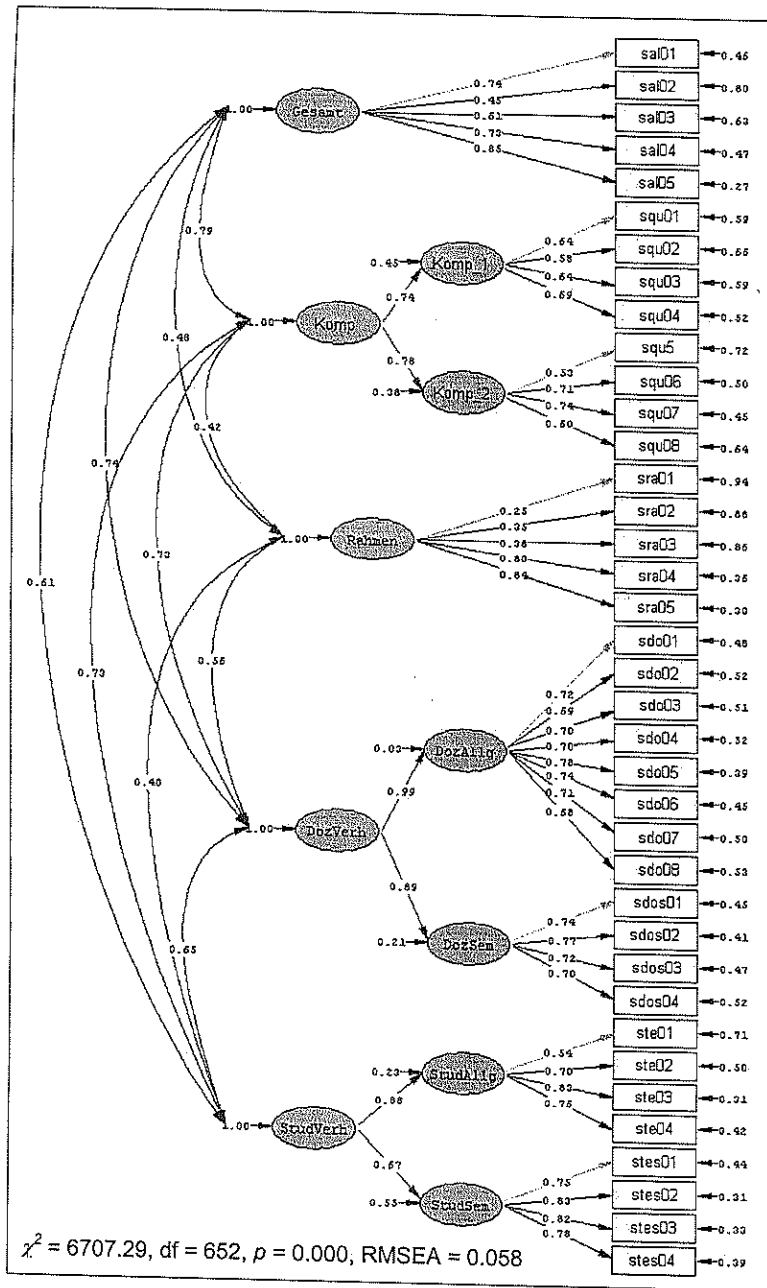


Abbildung 4

8.4. Diskussion

Ziel dieser Untersuchung war es erstens festzustellen, ob sich die durch das Qualitätsmodell der Lehre postulierten Dimensionen – „Gesamteindruck“, „Kompetenzerwerb“, „Rahmenbedingungen“, „Dozentenverhalten“ und „Studierendenverhalten“ – empirisch wieder finden lassen und zweitens zu überprüfen, ob die theoretisch abgeleitete und postulierte Multidimensionalität der Bewertungsdimensionen „Kompetenzerwerb“, „Dozentenverhalten“ und „Studierendenverhalten“ bestätigt werden kann.

Ausgehend von diesen Fragestellungen konnte gezeigt werden, dass sich die theoretisch zugrunde gelegte Fünf-Faktoren-Struktur in beiden Modellen wieder finden lässt. Der bessere Fit des Modells mit den Faktoren 2. Ordnung ($\chi^2 = 6707.29, df = 652, p = 0.000, RMSEA = 0.058$) gegenüber dem Fünf-Faktoren-Modell ($\chi^2 = 10739.53, df = 655, p = 0.000, RMSEA = 0.075$) und auch die höheren Reliabilitäten der Items für die Dimensionen „Kompetenzerwerb“, „Dozentenverhalten“ und „Studierendenverhalten“ im Modell mit den Faktoren 2. Ordnung sprechen für die Mehrdimensionalität dieser Dimensionen. Deshalb ist das Fünf-Faktoren-Modell zu Gunsten des Modells mit den Faktoren 2. Ordnung abzulehnen.

Des Weiteren gibt es Hinweise darauf, dass auch die Bewertungsdimension „Rahmenbedingungen“ mehrdimensional ist. Inhaltlich ist dieser Schluss plausibel, denn die Qualität der räumlichen Gegebenheiten und der Begleitmaterialien hängen nicht voneinander ab; beide sind lediglich Aspekte von Rahmenbedingungen. Dies zeigt sich auch bei den Reliabilitäten der Items zu den Begleitmaterialien und zur Literatur (sra04 und sra05), die mit $r_{ii} = 0.64$ $r_{ii} = 0.71$ wesentlich höher sind als die Reliabilitäten der Items zu den räumlichen und zeitlichen Gegebenheiten (sra01 mit $r_{ii} = 0.06$, sra02 $r_{ii} = 0.12$ und sra03 $r_{ii} = 0.14$).

Anzumerken bleibt noch, dass das Items saI02 „Der behandelte Stoff knüpft an meinen bisherigen Wissensstand an.“ der Dimension „Gesamteindruck“ mit $r_{ii} = 0.20$ eine wesentlich geringere Reliabilität aufweist als die anderen Items dieser Dimension. Dieser Sachverhalt könnte darauf zurückzuführen sein, dass dieses Item nur auf einen konkreten Aspekt - den behandelten Stoff - abzielt, wohingegen die anderen Items ein wesentlich globaleres Urteil über die Veranstaltung erfassen.

Abschließend bleibt festzuhalten, dass sich bei der Untersuchung zur Dimensionalität des Seminarfragebogens die durch das Qualitätsmodell der Lehre postulierten Bewertungsdimensionen wieder finden lassen. Übereinstimmend mit den theoretischen Vorüberlegungen zeigt sich aber auch, dass in diesem Zusammenhang nicht von einer Unidimensionalität der einzelnen Bewertungsdimensionen ausgegangen werden kann. Entsprechend den theoretischen Überlegungen konnte gezeigt werden, dass sowohl die Dimension „Kompetenzerwerb“ als auch die Bewertungsdimensionen „Dozentenverhalten“ und „Studierendenverhalten“ multidimensional sind.

8.5. Ausblick

Basierend auf den Ergebnissen dieser ersten Analyse zur Dimensionalität des Seminarfragebogens ergeben sich anschließende Fragestellungen für weitere Untersuchungen. Neben der Überprüfung der vermuteten Multidimensionalität der Bewertungsdimension „Rahmenbedingungen“ ist in den nächsten Schritten die dimensionale Struktur der beiden noch nicht untersuchten Fragebögen für Vorlesungen und Übungen und für die entsprechenden Versionen der Fragebögen für Dozenten zu untersuchen. Des Weiteren würden Modelle, in denen die Korrelationen zwischen den Fehler der Items erlaubt sind, Erkenntnisse über mögliche Zusammenhänge zwischen Items gleicher bzw. unterschiedlicher Dimensionen liefern. Solche Zusammenhänge bieten eine Grundlage für mögliche Erklärungen für die wesentlich geringeren Reliabilitäten einzelner Items

Aufgrund der universitären Struktur bezüglich der evaluierten Lehrveranstaltungen - Lehrveranstaltungen werden von Lehrenden angeboten, die bestimmten Lehrstühlen angehören, diese wiederum sind spezifischen Instituten und Fachbereichen/Fakultäten zugeordnet – bieten sich Multi-Level-Analysen an, die dieser Struktur auch methodisch Rechnung tragen. Lehrveranstaltungsbewertungen variieren erfahrungsgemäß aufgrund dieser strukturellen Zuordnung.

Auch regressionsanalytische Untersuchungen bezüglich des Einflusses der im Qualitätsmodell der Lehre postulierten Prozessvariablen auf die Ergebnisvariablen bzw. hinsichtlich der Vorhersagbarkeit der zusammenfassenden Items aus den Einzelitems der verschiedenen Bewertungsdimensionen stellen Anknüpfungspunkte für weitergehende Fragestellungen dar.

Des Weiteren bieten differenzierte Analysen des individuellen Antwortverhaltens interessante Fragestellungen, z.B.: Welche Art von Items werden bevorzugt mit „keine Angabe“ beantwortet oder aber ganz ausgelassen? Möglicherweise weisen jene Fälle, die auf Grund fehlender Werte von der Analyse ausgeschlossen wurden, eine bestimmte Systematik auf. Darüber hinaus sind aus Literatur so genannte Biasvariablen – wie beispielsweise der Besuchsgrund oder das Interesse am Thema - bekannt, welche die Bewertung einer Lehrveranstaltung beeinflussen (vgl. Rindermann, 2001).

Das Universitätsprojekt Lehrrevaluation führt in diesem Zusammenhang mehrere Studien zur Überprüfung der psychometrischen Güte der eingesetzten Fragebögen durch, auf deren Grundlage unter anderem auch zukünftig eine onlinebasierten Lehrveranstaltungsevaluation angeboten werden wird und die einen direkten Vergleich der Reliabilität und Validität der beiden unterschiedlichen Medien ermöglichen. Mit der vorliegenden Untersuchung wurde einer der ersten Schritte zur Überprüfung der Dimensionalität des Messinstrumentes vollzogen; das postulierte Qualitätsmodells der Lehre konnte bestätigt werden. Die nächsten Schritte werden die Beantwortung der sich an die vorliegenden Ergebnisse anschließenden Fragen sein.

8.6. Literatur

- Bloom, B. (1956). *Taxonomy of educational objectives. Handbook I: Cognitive domain*. New York: Davis McKay.
- Born, S. (in Vorb.). *Bestimmung der Gütekriterien eines Messinstrumentes zur Lehrveranstaltungsevaluation*. Friedrich-Schiller-Universität Jena (Diplomarbeit).
- Born, S. & Loßnitzer, T. (2006). Lehrveranstaltungsevaluation – Eine Analyse der Dimensionalität eines Fragebogens zur Evaluation von Lehrveranstaltungen. Vortrag auf Diplomanden-, Doktoranden- und Habilitandenkolloquium „Methoden- und Evaluationsforschung“, Spindler Mühle, Tschechische Republik, 09.03. – 12.03.2006.
- Braun, E. (2006). *Das Berliner Evaluationsinstrument für selbsteingeschätzte studentische Kompetenzen (BEvaKomp)*. Dissertation, Freie Universität Berlin.
- Braun, E. & Gusy, B. (2006). Perspektiven der Lehrrevaluation. In G. Krampen & H. Zayer (Hrsg.), *Psychologiedidaktik und Evaluation V*. Bonn: Deutscher Psychologen Verlag.
- Braun, E. & Gusy, B. (2005). *Perspektiven der Lehrrevaluation* [Online]. Verfügbar unter: http://www.ewi-psy.de/braun/Braun/braun_gusyAWF04.pdf (20.02.06).
- Browne, M. W. & Cudeck, R. (1993). Alternative ways of assessing model fit. In: K. A. Bollen & J. S. Long (Eds.) *Testing Structural Equation Models*. 136–162. Beverly Hills, CA: Sage.
- Buhl, Th. (1999a). Entwicklung eines Fragebogens zur Evaluation von Lehrveranstaltungen. Pilotphase im Wintersemester 1998/99 an der Friedrich-Schiller-Universität Jena. *metheval Report*, 1, Jena: Friedrich-Schiller-Universität, Institut für Psychologie, Lehrstuhl für Methodenlehre und Evaluationsforschung.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- Erpenbeck, J. & von Rosenstiel, L. (2003). *Handbuch Kompetenzmessung*. Stuttgart: Schäffer-Poeschel.
- Hayduk, L. A. (1996). *LISREL issues, debates and strategies*. Baltimore: Johns Hopkins: Univ. Press.
- Kanning, U.P.(2003). *Diagnostik sozialer Kompetenzen*. Göttingen: Hogrefe.
- Krems, B. (2004). *Methodenkompetenz*. In Online-Verwaltungslexikon [Online]. Verfügbar unter: <http://www.olev.de/m/methodenkompetenz.htm> (20.02.06).

- Loßnitzer, T. (in Vorb.). *Die prädiktive Validität eines Instruments zur Lehrveranstaltungsevaluation*. Friedrich-Schiller-Universität Jena (Dissertation).
- Loßnitzer, T., Schmidt, B. & Born, S. (eingereicht). *Zentrale Lehrveranstaltungsevaluation an der Friedrich-Schiller-Universität Jena – Qualitätsmodell und Messinstrument*. In: Krämer, M.: Psychologiedidaktik und Evaluation VI. Bonn: Deutscher Psychologen Verlag.
- Loßnitzer, T. & Born, S. (2005). Analyse der Messeigenschaften eines Instruments zur Lehrveranstaltungsevaluation. Vortrag auf der 7.Tagung der Fachgruppe Methoden und Evaluation, Münster, 29.09. – 01.10.2005.
- Madaus, G. F., Woods, E. M. & Nuttall, R. L. (1973). A Causal Model Analysis of Bloom's Taxonomy. *American Educational Research Journal*, 10 (4), 253-262.
- Rindermann, H. (2001). *Lehrevaluation. Einführung und Überblick zu Forschung und Praxis der Lehrevaluation an Hochschulen mit einem Beitrag zur Evaluation computerbasierten Unterrichts*. Landau: Verlag Empirische Psychologie.
- Schmidt, B. (in Vorb.). *Personalentwicklung an der Hochschule: Konzepte für junge wissenschaftliche Hochschulmitarbeiter/-innen*. Friedrich-Schiller-Universität Jena (Dissertation).
- Weinert, F.E. (2002). Vergleichende Leistungsmessung in Schulen - eine umstrittene Selbstverständlichkeit. In F.E. Weinert (Hrsg.). *Leistungsmessung in Schulen*. (S. 17-31). Weinheim und Basel: Beltz.