

# Zur Struktur von Sprachkompetenz

Dissertation  
zur Erlangung des akademischen Grades  
Doktorin der Philosophie (Dr. Phil.)

Vorgelegt dem Fachbereich  
Psychologie und Sportwissenschaften  
der Johann Wolfgang Goethe-Universität  
Frankfurt am Main

von

Dipl.-Psych. Nina Jude  
geboren am 21. September 1973 in Frankfurt am Main

1. Gutachter: Prof. Dr. Helfried Moosbrugger
2. Gutachter: Prof. Dr. Eckhard Klieme

Frankfurt am Main, im September 2008

## **Erklärung**

Hiermit erkläre ich,

- dass ich die vorliegende Dissertation selbständig verfasst und alle in Anspruch genommenen Hilfsmittel in der Dissertation angegeben habe,
- dass frühere Promotionsverfahren nicht erfolgt und insofern nicht erfolglos geblieben sind und
- dass mir die Ordnung zur Erlangung des akademischen Grades eines Doktors der Philosophie an der Johann Wolfgang Goethe-Universität vom 26.06.2001 bekannt ist.

Frankfurt am Main, den 01. September 2008

Nina Jude

## **Inhaltsverzeichnis**

Zur Struktur von Sprachkompetenz .....	7
1 Theorien und Modelle der Sprachkompetenz .....	10
1.1 Definitionen von Sprachkompetenz .....	11
1.2 Sprachkompetenz – Kompetenz oder Performanz? .....	13
1.3 Die Entwicklung von Sprachkompetenz in der Erstsprache .....	16
1.4 Institutionelle Vermittlung von Sprachkompetenz in der Erstsprache.....	18
1.5 Vermittlung von Sprachkompetenz in der Fremdsprache.....	21
1.6 Mehrsprachigkeit und bilingualer Unterricht.....	23
1.7 Kontextfaktoren und Sprachkompetenz.....	26
1.8 Sprachliche Kompetenzentwicklung – Konvergenz oder Divergenz?.....	28
1.9 Theorien und Modelle der Sprachkompetenz – Zusammenfassung .....	32
2 Empirische Befunde zur Struktur von Sprachkompetenz.....	33
2.1 Die vier Grunddimensionen von Sprachkompetenz.....	33
2.2 Empirische Ansätze.....	36
2.2.1 Schriftsprachliche Kompetenzen – Lesen und Schreiben .....	41
2.2.2 Rezeptive Kompetenzen – Hörverstehen und Leseverstehen .....	42
2.2.3 Auditive Kompetenzen – Hören und Sprechen.....	44
2.2.4 Spezifische Kompetenzstrukturen im Englischen als Fremdsprache.....	45
2.3 Zusammenhänge zwischen Erst- und Fremdsprachen .....	48
2.4 Gruppenunterschiede in der Dimensionalität von Sprachkompetenz .....	50
2.5 Dimensionalität als methodisches Artefakt? .....	52
2.6 Zusammenfassung.....	54

3	Die Erfassung von Sprachkompetenz im psychologisch-pädagogischen Kontext.....	56
3.1	Sprachbezogene Schulleistungstests .....	58
3.2	Die Erfassung von Sprachkompetenz in Large-Scale-Assessments in Deutschland.....	60
3.3	Die Erfassung von Sprachkompetenzen in der Studie Deutsch Englisch Schülerleistungen International.....	63
3.3.1	Stichprobe.....	64
3.3.2	Curriculare Verankerung der Testverfahren .....	66
3.3.3	DESI-Testkonstrukte und Testverfahren im Deutschen.....	67
3.3.4	DESI-Testkonstrukte und Testverfahren im Englischen.....	70
3.3.5	Datengrundlage – Personenschätzer und Stichprobengewichte.....	73
3.4	Zusammenfassung.....	76
4	Fragestellungen der Arbeit.....	77
4.1	Zur Struktur des Deutschen und des Englischen.....	78
4.2	Struktur von Sprachkompetenz unter Berücksichtigung der hierarchischen Schachtelung der Daten .....	80
4.3	Gruppenunterschiede – Erstsprache, Unterrichtsangebot und Geschlecht.....	82
4.4	Die Bedeutung kognitiver Grundfähigkeit für die Struktur von Sprachkompetenz .....	83
5	Methodische Grundlagen der Auswertung .....	85
5.1	Exploratorische Faktorenanalysen .....	85
5.2	Konfirmatorische Faktorenanalysen und Strukturgleichungsmodelle .....	89
5.3	Grundlagen der Mehrebenen-Analyse .....	94
5.3.1	Definition hierarchisch geschachtelter Datensätze .....	94

5.3.2	Methodische Grundlagen der Mehrebenenanalyse .....	98
5.4	Das methodische Vorgehen in der vorliegenden Arbeit .....	105
6	Ergebnisse .....	106
6.1	Deskriptive Ergebnisse - Beschreibung der Stichprobe .....	106
6.2	Struktur der Sprachkompetenz im Deutschen und im Englischen .....	109
6.2.1	Struktur im Deutschen .....	109
6.2.2	Struktur im Englischen .....	115
6.2.3	Exkurs: Modellierung eines auditiven Faktors der Sprachkompetenz .....	119
6.2.4	Vergleich der Modelle im Deutschen und im Englischen .....	120
6.2.5	Gemeinsames Modell des Deutschen und des Englischen .....	123
6.3	Struktur von Sprachkompetenz unter Berücksichtigung der hierarchischen Schachtelung der Daten .....	128
6.3.1	Analyse der Zusammenhänge auf Individual- und Klassenebene .....	128
6.3.1.1	Analyse der strukturellen Zusammenhänge auf Individual- und Klassenebene für das Deutsche .....	132
6.3.1.2	Analyse der strukturellen Zusammenhänge auf Individual- und auf Klassenebene für das Englische .....	136
6.3.2	Gemeinsames Mehrebenenmodell des Deutschen und des Englischen .....	138
6.3.3	Zusammenfassung der Ergebnisse der Mehrebenenanalysen .....	143
6.4	Mehrgruppenmodelle .....	147
6.4.1	Geschlechtsunterschiede .....	149
6.4.2	Unterschiede zwischen Schülerinnen und Schülern in bilingualen und nicht-bilingualen Bildungsgängen .....	154
6.4.3	Strukturelle Unterschiede in verschiedenen Sprachgruppen .....	160

6.4.4	Zusammenfassung der Ergebnisse der Mehrgruppenanalysen .....	165
6.5	Die Bedeutung kognitiver Grundfähigkeit für die Struktur von Sprachkompetenz .....	173
6.6	Zusammenfassung der Ergebnisse der Analysen zur Struktur von Sprachkompetenz .....	177
7	Diskussion.....	180
7.1	Allgemeine Befunde zur Struktur von Sprachkompetenz.....	181
7.2	Mehrebenenanalytische Betrachtung der Struktur von Sprachkompetenz .....	183
7.3	Zusammenhänge zwischen dem Deutschen und dem Englischen – spezifische oder allgemeine Transfereffekte? .....	186
7.4	Gruppenunterschiede in der Struktur von Sprachkompetenz.....	188
7.5	Kognitive Grundfähigkeit als Element der Struktur von Sprachkompetenz.....	193
7.6	Mehrebenenanalysen – Welche Ebenen sollten betrachtet werden?.....	195
7.7	Dimensionalität als methodisches Artefakt des diagnostischen Ansatzes? .....	197
7.8	Abschließender Ausblick .....	200
8	Literatur.....	202

## Zur Struktur von Sprachkompetenz

Sprachkompetenz stellt ein komplexes theoretisches Konstrukt dar. Die Diskussion darüber, was Sprachkompetenz ausmacht und wie diese Komplexität darzustellen sei, beschäftigt infolgedessen verschiedene Forschungsbereiche seit langem. Im Kontext der Psychologie sind es vor allem die Felder der Psycholinguistik und der Entwicklungspsychologie, aber auch der Diagnostik, die sich mit der Frage nach den Komponenten der Sprachkompetenz aus unterschiedlichen Blickwinkeln auseinandersetzen. Eine Hauptforschungsfrage dieser Wissenschaften ist die nach der *Struktur von Sprachkompetenz*, also die Frage danach, in welche Teilbereiche sich Sprachkompetenz unterteilen lässt und wie diese Bereiche zusammenhängen.

Je nach Fragestellung stehen die Unterscheidung zwischen Erst- und Fremdsprache, die Differenzierung von produktiven und rezeptiven Teilbereichen oder auch die Trennung von grundlegenden Sprachfertigkeiten wie Lesen, Schreiben, Sprechen und Hörverstehen im Vordergrund. Die Begriffe der Struktur oder auch Dimensionalität von Sprachkompetenz werden dabei durchaus unterschiedlich verwendet: In der Psycholinguistik bezieht sich die Struktur von Sprache eher auf sprachtheoretische und kognitive Bereiche. Struktur meint hier Elemente, die Worte konstituieren – wie Morpheme und Grapheme – aber auch die neuronalen Zusammenhänge zwischen den für die Sprachverarbeitung zuständigen Gehirnregionen. Die Entwicklungspsychologie beschäftigt sich mit Komponenten der Sprachkompetenz, die sich im Entwicklungsprozess verändern. Dazu zählen rein biologische Funktionen des Hörens und Artikulierens, die Analyse von Erwerbsbedingungen des Sprechens und Schreibens sowie Sprachentwicklungsstörungen, aber auch die Frage, ob beziehungsweise wie sehr sich die Fähigkeiten in einzelnen sprachlichen Teilbereichen im Prozess der Sprachentwicklung einander annähern. Die psychologische Diagnostik schließlich betrachtet die Frage nach der Struktur von Sprachkompetenz aus dem Blickwinkel der Erfassbarkeit und der empirischen Trennbarkeit unterschiedlicher Fähigkeitsbereiche.

Diese Fragen nach der Struktur von Sprachkompetenz sind aktuell in der pädagogisch-psychologischen Diagnostik von besonderer Bedeutung. Da sprachliche Kompetenzen als wesentlicher Bestandteil von Bildung angesehen werden, finden sie in empirischen Studien zu Schülerleistungen immer stärkere Beachtung. Neben der Feststellung von Ausprägungen der sprachlichen Kompetenzen, also der Einordnung von Schülerinnen und Schülern auf

Kompetenzniveaus, wird die Differenzierbarkeit von Teilbereichen der Sprachkompetenz betrachtet, wenn spezifische sprachliche Phänomene getrennt beobachtet und beschrieben werden sollen. Während zur Feststellung individueller Sprachleistungen und sprachlicher Entwicklung differenzierte Diagnoseinstrumente eingesetzt werden, bedienen sich repräsentative Schulleistungsstudien und Vergleichsarbeiten oft einiger weniger Indikatoren für Sprachkompetenz. Es liegt auf der Hand, danach zu fragen, welcher Auflösungsgrad der Betrachtung von Sprachkompetenz jeweils für spezifische Aussagen über sprachliche Kompetenzen notwendig und sinnvoll ist. Während beispielsweise ein Testwert der Lesekompetenz für internationale Vergleiche auf Länderebene ein valider Indikator für die Leistung des nationalen Bildungssystems sein kann, sollten individuelle Fördermaßnahmen sicherlich verschiedene sprachliche Teilkompetenzen betrachten, um Profile der Sprachkompetenz und eventueller sprachlicher Defizite auswerten zu können.

Die vorliegende Arbeit nähert sich der Frage nach der Struktur von Sprachkompetenz aus einer psychologisch-diagnostischen Perspektive. Auf der Grundlage von Daten einer deutschen Large-Scale-Studie und unter Bezug auf Theorien und Modellen der Psycholinguistik, der Entwicklungspsychologie und der psychologischen Diagnostik geht sie der Frage nach, welche Teilbereiche der Sprachkompetenz sich empirisch differenzieren lassen – wie sich also die Struktur von Sprachkompetenz aus Sicht der Diagnostik darstellen lässt. Darüber hinaus wird analysiert, ob sich spezifische Zusammenhänge zwischen einzelnen sprachlichen Teilbereichen empirisch auffinden lassen, die vor dem Hintergrund bekannter Theorien und Modelle der Struktur von Sprachkompetenz darzustellen und zu erklären sind. Schließlich wird untersucht, ob sich bestimmte Personengruppen hinsichtlich der strukturellen Zusammenhänge in ihrer Sprachkompetenz unterscheiden und inwiefern kognitive Grundfähigkeiten als Teil der Struktur von Sprachkompetenz zu betrachten sind.

Im Rahmen dieser Arbeit sind also auf der einen Seite die theoretischen Konstrukte zu betrachten, die aus inhaltlichen Perspektiven die Struktur von Sprachkompetenzen in unterschiedlicher Weise postulieren. Auf der anderen Seite stehen Testverfahren, deren empirische Ausgestaltung für den jeweiligen Verwendungskontext spezifisch ist und deren Differenziertheit in der Erfassung sprachlicher Kompetenzen nicht immer an die theoretischen Modellansprüche heranreichen kann. Trotzdem soll eine Integration versucht werden, mit dem Ziel, die Struktur von Sprachkompetenz abzubilden und auf der Grundlage bekannter Theorien und Befunde zu interpretieren.

Die vorliegende Arbeit gliedert sich in die folgenden Abschnitte: Im Anschluss an die Einleitung stellt Abschnitt 1 *Theorien und Modelle der Sprachkompetenz* dar, hierzu gehören grundlegende *theoretische Annahmen* über die Struktur von Sprachkompetenz in unterschiedlichen Fachdisziplinen. Beginnend mit Definitionsansätzen zur Sprachkompetenz allgemein, werden Modelle der Sprachentwicklung in Erst- und Fremdsprache sowie Ansätze der Lehr-Lern-Forschung dargestellt, die für die Analyse der Struktur von Sprachkompetenz in der vorliegenden Arbeit von Bedeutung sind. Der Abschnitt 2, *Empirische Befunde zur Struktur von Sprachkompetenz*, stellt speziell jene Ansätze dar, die im nationalen und internationalen Forschungsfeld die Struktur von Sprachkompetenz *empirisch untersucht haben*. Hierzu zählen Ergebnisse bezüglich der Zusammenhänge bestimmter sprachlicher Teilbereiche, der Beziehungen zwischen Erst- und Fremdsprache, zu Gruppenunterschieden sowie zur aktuellen Diskussion um methodische Herangehensweisen in diesem Forschungsfeld. Im Anschluss daran stellt Abschnitt 3 die Erfassung von Sprachkompetenz im psychologisch-pädagogischen Kontext dar. Die bisher angeführten Theorien und Modelle werden im *Kontext der empirischen Lehr-Lern-Forschung* spezifiziert und die Besonderheiten der sprachlichen Diagnostik in Schulleistungsstudien dargestellt. Speziell wird dabei auf die Datengrundlage der vorliegenden Arbeit, die Studie *Deutsch Englisch Schülerleistungen International (DESI)*, eingegangen. Vor diesem theoretischen Hintergrund werden im Abschnitt 4 die *Fragestellungen* dieser Arbeit aufgeführt. Abschnitt 5 stellt die *verwendeten Analysemethoden* dar. Abschließend werden in Abschnitt 6 die *Ergebnisse* der Analysen der Struktur von Sprachkompetenz erläutert und vor dem Hintergrund der angeführten Theorien und Modelle reflektiert. Eine abschließende *Diskussion* in Abschnitt 7 fasst die Befunde zusammen und diskutiert sie auch im Hinblick auf weiterführende Forschungsansätze.

## **1 Theorien und Modelle der Sprachkompetenz**

Sprachkompetenz ist ein komplexes Phänomen, dem ein lebenslanger Lernprozess zugrunde liegt. Während dieses Prozesses bilden sich nach Klein (2000) individuelle Strukturen der Sprachkompetenz aus, indem „der Lerner [...] sich die vielen phonologischen, lexikalischen und auch pragmatischen Charakteristika der Sprache“ (S. 549) aneignet. Das Forschungsfeld der psychologisch-diagnostischen Erfassung von Sprachkompetenz, in dem sich die vorliegende Arbeit bewegt, verbindet Fragestellungen aus unterschiedlichen psychologischen, pädagogischen und linguistischen Disziplinen. Daher ist es nicht verwunderlich, dass die theoretischen Definitionen von Sprachkompetenz ebenso variieren wie die empirischen Befunde zu ihrer Struktur.

Neben Modellen, die Sprache als universelle Kompetenz ansehen (Carroll, 1968; Lado, 1961) finden sich auch solche, die sprachliche Kompetenz situationsabhängig definieren (Hymes, 1967) oder zwischen grammatikalischen und kommunikativen Aspekten der Sprachverwendung unterscheiden (Bachman & Palmer, 1982). Psychophysiologische Modelle betrachten Strukturen von Sprachkompetenz rein unter dem Aspekt der strukturellen neuronalen Veränderungen im Prozess der Sprachentwicklung und des Spracherwerbs (Turkeltaub, Gareau, Flowers, Zeffiro & Eden, 2003). Und schließlich bestehen Unterschiede in der Definition dessen, was Sprachkompetenz ist und wie sie zu strukturieren sei, auch in Abhängigkeit davon, ob die Erstsprache, eine Fremdsprache oder bilinguale Sprachkompetenzen analysiert werden.

Es bedarf daher eines Überblicks über die verschiedenen Definitionen von Sprachkompetenz und die unterschiedlichen Modelle ihrer Struktur, um die vorliegende Arbeit in einen theoretischen Rahmen einordnen zu können. In diesem Kapitel werden unterschiedliche Definitionsansätze von Sprachkompetenz zusammenfassend dargestellt. Neben spezifischen Modellen von Sprachkompetenz und Sprachentwicklung in der Erstsprache wird dabei auf die institutionelle Vermittlung von Erst- und Fremdsprache sowie auf theoretische Ansätze der bilingualen Sprachkompetenz und des bilingualen Unterrichts eingegangen.

## 1.1 Definitionen von Sprachkompetenz

Sprachkompetenz – für diesen Begriff existieren bisher keine einheitlichen oder gar verbindlichen Definitionen. Vielmehr beziehen sich die spezifischen Auslegungen jeweils auf spezielle theoretische Ansätze aus den Bereichen der Linguistik, insbesondere der Psycholinguistik, der Neuropsychologie oder der sprachbezogenen Lehr-Lern-Forschung (Del Vecchio & Guerrero, 1995). Die *Linguistik* beschäftigt sich mit der Beschreibung und Erklärung sprachlicher Phänomene und betrachtet Sprache als ein System komplexer Strukturen, das die grammatikalischen Grundelemente der Sprache, nämlich Laute, Morpheme, Wörter, Sätze und Texte vereint. *Sprachkompetenz* wird handlungsbezogen in der Verwendung von Sprache als Kommunikationsmittel, unter Einbeziehung der grundsätzlichen Regeln des Sprachgebrauchs, untersucht (Linke, Nussbaumer & Portmann, 2001). Die *kognitive Psychologie* beschreibt Sprache auf Grundlage von Informationsverarbeitungstheorien durch Rekonstruktion derjenigen hierarchischen mentalen Prozesse, die zur Sprachentschlüsselung notwendig sind. Damit steht sie in engem Zusammenhang mit der *Neuropsychologie*, die neuronale Grundlagen von Sprache und sprachlichen Teilleistungen analysiert und Sprachkompetenz im Sinne eines Zugriffs auf sprachliche Informationen klassifiziert (List, 1989).

Auch *neurolinguistische* und *entwicklungspsychologische* Ansätze erforschen Sprachkompetenz im Hinblick auf zugrunde liegende neuronale Vernetzungen im Gehirn und deren Veränderungen im Prozess des Spracherwerbs (Jusczyk, 2003). Sie basieren unter anderem auf Annahmen der Linguistik und analysieren auf unterschiedlichen Ebenen phonologisches Verstehen und Produzieren, Mechanismen des Regelverstehens und der Assoziation sowie semantisches und syntaktisches Wissen (Plaut, 2003). Überschneidungen bestehen mit dem Forschungsbereich der *Entwicklungspsychologie*, welcher die Entwicklung des aktiven Sprachhandelns als psychologisches Element integriert (Szagun, 2000). Untersucht wird beispielsweise der Zusammenhang zwischen sprachlichen Strukturelementen und situationsspezifischem Sprachverhalten in konkreten (Test-)Situationen (Brown, Malmkjaer & Williams, 1996; Shohamy, 1996).

Die *Psycholinguistik* betrachtet ebenfalls die kognitiven Vorgänge bei der Rezeption und Produktion linguistischer Elemente sowie den Einfluss der Sprachstruktur auf das Denken, ein Schwerpunkt ist in diesem Forschungsfeld der Aspekt des Erwerbs einer Fremdsprache

(Hartig, 1999; Schwarz, 1996). *Soziolinguistische* Modelle rekurren stärker auf den vermittelnd-interaktiven Aspekt von Sprache und analysieren Funktionen und Ziele im kommunikativen Sprechakt (Bühler, 1992; Dittmar, 1989); hier finden sich wiederum Modelle der Linguistik wieder. *Differentialdiagnostische Ansätze* der Psychologie und Pädagogik erforschen Sprachkompetenzen vor dem Hintergrund des *Sprachlehrens und -lernens* und basieren auf der Annahme, dass sich sprachliche Gesamtkompetenz in spezifische Teilkompetenzen aufschlüsseln lässt. Deren empirische Erfassung ermöglicht Aussagen über individuelle Leistungsprofile von Sprachlernenden (Bachman, 1990b; Butler & Stevens, 2001), aus denen sich diagnostische Aussagen für die Förderung spezifischer Fähigkeiten ableiten lassen (Groeben & Hurrelmann, 2002).

Diese verschiedenen Forschungsrichtungen betrachten das Phänomen der Sprachkompetenz mit unterschiedlichen Schwerpunktsetzungen. Insgesamt liegt der Fokus von Linguistik und insbesondere Psycholinguistik stark auf einer theoretisch-kognitiven Definition von Sprachkompetenz. Ihre Modelle basieren auf theoretischen Strukturen, Grammatiken und sprachbezogenen Kognitionen und grenzen sich damit von der stärker handlungsbezogenen pädagogisch-psychologischen Theorie- und Modellbildung ab. Diese Frage nach dem Handlungsbezug von Sprachkompetenz, also dem Zusammenspiel von sprachlichem Wissen und tatsächlichem Sprachhandeln, wurde und wird besonders in der Linguistik als Problem der „Kompetenz und Performanz“ diskutiert und ist in Bezug auf die pädagogisch-psychologische Kompetenzerfassung von besonderer Relevanz. Aus diesem Grund soll im Folgenden kurz auf die Kompetenz-Performanz-Debatte eingegangen werden.

## 1.2 Sprachkompetenz – Kompetenz oder Performanz?

Die durchaus kontroverse Diskussion darüber, ob sich sprachliche Kompetenz als (potenzielle) Möglichkeit allen sprachlichen Handelns in ebendiesem eins zu eins abbildet oder ob das sprachliche Handeln als beobachtbare *Performanz* immer von der potenziell vorhandenen *Kompetenz* zu unterscheiden sei, zieht sich bis heute durch die Ansätze der empirischen Erfassung von sprachlichen Fähigkeiten. Zu Beginn des 20. Jahrhunderts legte Frederic de Saussure mit seinem Werk „Cours de linguistique générale“ den Grundstein zur Definition von Sprachkompetenz in der Linguistik (Saussure, 2001). Er unterschied zwischen „*Langue*“ als einem statischen System von Formen und Strukturen und „*Parole*“ als der Realisierung dieses Wissens im Sprechakt (Dresselhaus, 1979). Aus sprachpsychologischer Sicht deutete sich hier schon die Unterscheidung zwischen Sprache als Kognition und als Handlung an, und genau diese Sichtweise wurde mit dem Aufkommen der kognitiven Sichtweise in den 1960er Jahren verstärkt auch für die Definition von Sprachkompetenz übernommen.

Noam Chomsky als ein prominenter Vertreter der Sprachwissenschaften greift diese Unterteilung auf, benennt sie jedoch anders: Bei ihm wird die „*Langue*“ zur „Kompetenz (competence)“ und ist nun kein statisches Wissen mehr, sondern ein dynamisches, das Regeln und Operationen verbindet. Die „*Parole*“ wird bei Chomsky zur „Performanz (performance)“, zur Ausführung dieses sprachlichen Wissens (Chomsky, 2001).

Der *Aspekt der Sprachverwendung*, der bis dahin in der theoretischen Linguistik nur wenig Aufmerksamkeit fand, wurde besonders in der Soziolinguistik aufgegriffen. Ihre Vertreter sahen in der Kompetenz die Repräsentation vorwiegend grammatikalischen Wissens, während sich die sozio-pragmatischen Aspekte der Sprache in der Performanz widerspiegelten (Hymes, 1967; Taylor, 1988). Sie charakterisierten diese Unterscheidung folgendermaßen: „Competence is understood to be dependent on two things: (tacit) knowledge and (ability for) use“ (Hymes, 1971, S. 16). Diese Trennung von sprachlichem Wissen (language knowledge) und sprachlichem Verhalten (language use) als der Anwendung dieses Wissens (Coseriu, 1988) führte zu Debatten besonders in der empirischen Sprachforschung. Argumentiert wurde, dass die vorhandene *sprachliche Kompetenz* im Sinne Chomskys sich niemals komplett mit Testverfahren erfassen ließe, denn diese könnten nur das *sprachliche Handeln als Performanz*, also die Anwendung des Wissens, überprüfen (Shohamy, 1996). Die

Anwendung wiederum sei jedoch nicht nur beeinflusst vom Sprachwissen, sondern von vielfältigen Faktoren, der Motivation, der Situation oder auch der Interaktionsbedingungen im jeweiligen Kontext.

Diese strikte Trennung zwischen theoretischem Sprachwissen und praktischem Sprachhandeln manifestierte sich seit den 1960er Jahren in vielen Theorien der Sprachkompetenz. Prominenteste Vertreter jeweils unterschiedlicher Definitionen sprachlicher Kompetenz waren Bachman und Palmer (1982), die *grammatische* von *pragmatischer* Kompetenz unterschieden, sowie Canale und Swain (1980), die ihren Fokus auf die Sprachanwendung legten, diese jedoch in *kommunikative Kompetenz* und *kommunikative Performanz* unterteilten. Diese Fokussierung besonders auf die anwendungsbezogenen, kommunikativen Aspekte der Sprachbeherrschung ging einher mit neuen Entwicklungen der Lehr-Lern-Forschung. Diese stellte inzwischen nicht mehr nur die Vermittlung von Grammatik und Vokabular in den Vordergrund, sondern plädierte für eine kommunikative Sichtweise besonders beim Erwerb von Fremdsprachen (Widdowson, 1989). In den Sprachunterricht hielten Sprachlabore Einzug, in denen besonders Hörverstehen und Aussprache geübt wurden (Jung, 1978). In der psychologischen und soziologischen Sprachforschung entstanden integrative Ansätze, die das Verhältnis zwischen kognitiven Vorgängen und Sprachhandeln in Zusammenhang mit situationalen, motivationalen und affektiven Effekten untersuchten (Berns, 1990). Sprachkompetenz wurde zunehmend als ein Konstrukt angesehen, das auch von den Eigenschaften des Lernalters, wie beispielsweise dem kulturellen Hintergrund oder der individuellen Motivation beeinflusst wird (Gardner & MacIntyre, 1993; Kunnan, 1994). Definitionen von Sprachkompetenz im 20. Jahrhundert waren also wesentlich von der Unterscheidung zwischen *sprachlichem Wissen* und *sprachlichem Handeln* geprägt.

Aktuelle Definitionsansätze hingegen versuchen, mit dem Begriff der *Kompetenz* eine Brücke zu bauen zwischen deklarativem Wissen und seiner Anwendung. Die Kompetenz einer Person wird als das Bindeglied zwischen Wissen und Können definiert, als Befähigung zur Bewältigung unterschiedlicher Situationen (Klieme, 2004). In diesem Sinne bezieht sich der Begriff Kompetenz auf allgemeine Fähigkeiten im Sinne von Dispositionen ebenso wie auf funktional spezifische Kenntnisse und bereichsspezifische Fertigkeiten, die kognitive und motivationale Aspekte im Handlungskontext integrieren (Weinert, 1999).

Diese allgemeine Definition von Kompetenz lässt sich dementsprechend auf den Begriff der *Sprachkompetenz* übertragen und findet sich auch in aktuellen sprachdiagnostischen Ansätzen wieder. Sprachkompetenz wird dabei allerdings stärker funktional definiert und als kognitive Disposition angesehen, die dazu befähigt, situative Anforderungen erfolgreich zu bewältigen: „Sprachliche Kompetenzen sind die Summe des (deklarativen) Wissens, der (prozeduralen) Fertigkeiten und der persönlichkeitsbezogenen Kompetenzen und allgemeinen kognitiven Fähigkeiten, die es einem Menschen erlauben, Handlungen auszuführen. [...] Kommunikative Sprachkompetenzen befähigen Menschen zum Handeln mit Hilfe spezifischer sprachlicher Mittel“ (Council of Europe, 2001, S. 21).

Das Konzept der Sprachkompetenz als Befähigung zur Handlung in sprachbezogenen Situationen steht in direkter Beziehung zum Begriff der „literacy“ im englischen Sprachraum (Verhoeven, 1994). Literacy kann auf unterschiedliche Kompetenzbereiche bezogen werden und bezeichnet die Anwendung von Wissen und Fähigkeiten in einem alltäglichen Kontext: „Reading literacy is defined as: the ability to understand and use those written language forms required by society and/or valued by the individual“ (Mullis, Martin, Gonzalez & Kennedy, 2003). Auf dieses Konzept beziehen sich auch die internationalen Schulleistungsstudien der Organisation for Economic and Cultural Development (OECD). *Literacy* charakterisiert Sprachkompetenz in einer Doppelfunktion: Einerseits referiert der Begriff auf Sprachkönnen in dem Sinne, dass eine bestimmte Sprache verwendet werden kann, andererseits bezeichnet er Sprache als eine unabdingbare Voraussetzung und als Instrumentarium für die Aneignung von neuem Wissen, sodass sich Sprachkompetenz auf sich selbst aufbauend entwickelt (Bos, Lankes, Prenzel et al., 2003). Vor diesem Hintergrund wird in der vorliegenden Arbeit von der Struktur von Sprachkompetenz gesprochen, die als eben jene Kompetenz anzusehen ist, die sich im Sprachhandeln manifestiert und sich über dieses beobachten und messen lässt.

### **1.3 Die Entwicklung von Sprachkompetenz in der Erstsprache**

Sprache kann als ein humanspezifisches Phänomen angesehen werden, das zentral für das menschliche Leben ist (Grimm, 2000). Der Erwerb von Sprachkompetenz in der Erstsprache ist ein langjähriger, komplexer Prozess, der auf angeborenen Prädispositionen zum Spracherwerb basiert (Hennon, Hirsh-Pasek & Golinkoff, 2000). Sprachentwicklung ist nicht als isolierter Vorgang zu verstehen, sondern Teil einer umfassenden biologischen Gesamtentwicklung des Menschen, die wiederum in einen Sozialisationsprozess eingebettet ist. Der Erwerb sprachlicher Kompetenzen ist ein Teil der universellen Entwicklung des Menschen, der sensorische, motorische, kognitive, emotionale und sozial-kommunikative Funktionsbereiche umfasst, die sich in ihrer Wirkungsweise gegenseitig beeinflussen (Grohnfeldt, 1993). Die biophysiologischen Grundlagen hierfür bilden das differenzierte Wahrnehmungsvermögen des Menschen sowie seine (fein-)motorischen Fähigkeiten, die zur sprachlichen Artikulation nötig sind (Jusczyk, 2003). Im Entwicklungsprozess werden dabei die unterschiedlichsten Teilbereiche von Sprache, wie Grammatik, Lexik und Semantik, und spezifische Fähigkeiten des Hörverstehens, des Sprechens und später auch des Lesens und Schreibens erworben (Grimm, 2000).

Über die Art und Weise des Erwerbs der Erstsprache existieren unterschiedliche Theorien: Lange Zeit galt der Spracherwerb als ein Prozess der reinen Imitation und Verstärkung, in dem das Kind Sprache durch Nachahmung erlernt sowie durch die Reaktion, die es auf die korrekte Sprachanwendung erhält. Spezielle Theorien hierzu finden sich im Behaviorismus (Skinner, 1957) und Interaktionismus (Bruner, 1997). Demgegenüber stellten nativistische Theorien die Hypothese auf, dass Menschen einen angeborenen Mechanismus zum Spracherwerb besitzen, welcher die enorme Geschwindigkeit erklären soll, mit der Kinder ihre Muttersprache erlernen (Chomsky, 1987). Ob es sich bei diesem Mechanismus um eine Art Wissen über sprachliche Universalien handelt oder ob lediglich allgemeine Vorgehensweisen zum Spracherwerb in der Natur des Menschen liegen, ist nach wie vor umstritten (Szagun, 2000). Der Kognitivismus sieht die Entwicklung sprachlicher Strukturen als Produkt neuronaler intellektueller Reifung an und argumentiert, dass bestimmte sprachliche Phänomene erst dann auftreten, wenn die kognitive Grundlage hierfür ausgebildet ist. Ein Beispiel ist die Verwendung von Vergleichsstrukturen („Dieses Auto ist größer als jenes“), für welche die kognitive Fähigkeit zur Beurteilung von Größenunterschieden ausgebildet sein muss (Crystal, 1993; Haberzettl & Wegener, 2003; Piaget, 1983). Neuere

Forschungsansätze – besonders in der Neurobiologie – greifen diese kognitiven und entwicklungspsychologischen Modelle auf, um linguistische Phänomene differenziert zu untersuchen. Im Mittelpunkt stehen neurokognitive Korrelate von Sprache und Spracherwerb, beispielsweise der Aufbau des mentalen Lexikons und die Anwendung grammatikalischer Regeln (Plaut, 2003; Ullman, 2001).

Insgesamt lässt sich der Spracherwerb als ein Prozess beschreiben, bei dem viele Mechanismen ineinandergreifen und der eng mit der allgemeinen Entwicklung und Sozialisation des Kindes zum handlungsfähigen Subjekt in seiner Gesellschaft verbunden ist. Auf der Basis sensumotorischer Schemata laufen kognitive Prozesse der Wahrnehmungsverarbeitung ab, die mit emotional-motivationalen Faktoren und sozial-kommunikativen Bedingungen verknüpft sind. Sprachentwicklung kann als ein Wechselspiel von biologisch-neurophysiologischen Voraussetzungen und soziokulturell beeinflussten Umwelterfahrungen verstanden werden. Eine wesentliche Rolle spielt dabei der geleitete Spracherwerb in Kindergarten und Schule. Die Sprachverarbeitung verändert sich mit dem Schuleintritt insofern, als zur automatisierten, intuitiven Sprachverarbeitung die geleitete, kontrollierte Sprachverarbeitung hinzukommt. Besonders in der Schule wird Sprache als Objekt thematisiert und metasprachliches Wissen, also Wissen über Sprache, auf- und ausgebaut. Sprache kann dann auch auf einer Metaebene als formales System analysiert werden (Friederici & Hahne, 2000; Karmiloff-Smith, 1979).

#### 1.4 Institutionelle Vermittlung von Sprachkompetenz in der Erstsprache

Die *institutionelle Vermittlung von Sprachkompetenz* in der Erstsprache beginnt mit dem Eintritt in Vorschule oder Schule, in Deutschland übernimmt in der Regel der Deutschunterricht diese Aufgabe. Dies bedeutet auch, dass *Schülerinnen und Schüler mit einer anderen Erstsprache als Deutsch* die institutionelle Vermittlung von Sprachkompetenz in einer Fremdsprache erfahren, es sei denn, sie besuchen zusätzliche schulische Angebote wie den so genannten „Muttersprachlichen Unterricht“ (Hessisches Kultusministerium, 1995). Im Folgenden wird unter *Unterricht in der Erstsprache* der Deutschunterricht an deutschen Schulen verstanden.

Der Unterricht in der Erstsprache baut auf einer breiten Grundlage von Vorkenntnissen auf Seiten der Schülerinnen und Schüler auf. Die Vermittlung von Sprachkompetenz verläuft jetzt aktiv geleitet, das heißt, Sprache wird nicht mehr im freien sozialen Kontext zwangsläufig erworben, sondern gezielt vermittelt. In der Grundschule hat der Unterricht in der Erstsprache die Aufgabe, die sprachlich-kommunikative Handlungsfähigkeit der Schüler differenziert aufzugreifen und weiterzuentwickeln, um insbesondere die Ausdrucks- und Kommunikationsfähigkeit sowie das Sprachbewusstsein der Schüler zu fördern; ein besonderer Schwerpunkt liegt dabei auf der Hinführung zur Schriftkultur (Hessisches Kultusministerium, 1995). Der Deutschunterricht der Sekundarstufe I und II fokussiert darauf, Grundlagen in der ästhetischen und kreativen Sprachkompetenz herzustellen und zu erweitern. Dabei sollen Schülerinnen und Schüler in persönlichen, beruflichen und öffentlichen Zusammenhängen situationsangemessen und adressatengerecht sprechen und zuhören sowie reflektierend, kommunikativ und gestalterisch schreiben können (Ständige Konferenz der Kulturminister der Länder in der Bundesrepublik Deutschland, 2004).

Die Sprachvermittlung im Deutschunterricht erfolgt in gewisser Weise immer holistisch – beim Üben der Lesekompetenz werden realitätsnahe, kommunikative Texte verwendet, die Grammatik und Soziopragmatik gleichermaßen ansprechen. Aufgaben der Textproduktion beziehen neben dem Inhalt immer auch Aspekte des Stils und der Rechtschreibung mit ein, und mündliche Äußerungen der Schüler werden auf ihre sprachliche Korrektheit hin bewertet. Trotzdem werden bestimmte sprachliche Kompetenzen im didaktischen Kontext auch getrennt fokussiert: Besondere Bedeutung in der erstsprachlichen Bildung besitzt die Lese- und Rechtschreibkompetenz als Grundlage sowohl für schulischen Erfolg als auch für den

beruflichen Werdegang (Schneider, 2001). Auch repräsentative Schulleistungsstudien wie das Programme for International Student Assessment (PISA) und die Internationale Grundschule Lese Untersuchung (IGLU) fokussieren auf eine funktionale Lesekompetenz als Basiskompetenz von Sprachfähigkeit (Artelt, Stanat, Schneider & Schiefele, 2001; Bos, Lankes, Prenzel et al., 2003). Lesen ist eine unabdingbare Kulturtechnik, um im Alltag selbst einfachste Situationen erfolgreich zu meistern und sich neues Wissen anzueignen (Nickel, 2002). Dem Deutschunterricht kommt dabei die Aufgabe zu, Lesekompetenz angesichts der vorhandenen interindividuellen Unterschiede gezielt zu fördern (Jenkins, Broek, Espin & Deno, 2003; Richter & Christmann, 2002).

Auch auf Rechtschreibung als schulisch vermitteltes, wesentliches Element einer Schriftsprache wird im schulischen Unterricht mit eigenen Lehr-Lern-Inhalten eingegangen. Die Methoden des Buchstabierens und Diktierens sollen die Wahrnehmung für die korrekte Schreibweise von Wörtern schulen, sodass diese im Idealfall automatisiert wird (Thomé, 1998; Turkeltaub et al., 2003). Das richtige Schreiben von Wörtern als Grundlage für die schriftsprachliche Verständigung wird als ein wesentlicher Bestandteil des Deutschunterrichts, insbesondere in der Grundschule, angesehen.

Ein weiteres spezifisches Element des Deutschunterrichts ist die Vermittlung von Wissen über die Strukturen und Regeln der Grammatik (Hartig, 1999). In den letzten Jahren werden hier aktuelle Konzepte der Vermittlung der sogenannten *Sprachbewusstheit* im Deutschunterricht der reinen Wissensvermittlung vorgezogen (Ingendahl, 1999; Neuland, 2002). Sprachbewusstheit als Lernziel beinhaltet unter anderem die Reflexion über Sprachverwendung, besonders über sprachliche Angemessenheit in Abhängigkeit von funktionalen und situativen Gegebenheiten. Während bei der mündlichen Sprachproduktion im Unterricht im Wesentlichen Inhalt und kommunikative Angemessenheit des Gesagten beurteilt werden, spielen bei der schriftlichen Sprachproduktion die Kompetenzen in Rechtschreibung und Grammatik eine wesentliche Rolle (Käsermann & Foppa, 2003; Kellogg, 2003).

Die institutionelle Vermittlung von Sprachkompetenz zielt also einerseits immer auf Sprache als Ganzes ab, initiiert jedoch auch Lehr-Lern-Anlässe, in denen spezielle Teilbereiche des Sprechens, Schreibens, Lesens oder des Hörverstehens und der Vermittlung sprachlichen Meta-Wissens fokussiert vermittelt werden. Es kann angenommen werden, dass die Art und

Weise der Sprachvermittlung auch die Struktur der Kompetenz beeinflusst. Ein Beispiel hierfür wären kommunikativ orientierte Lehransätze, die auf Sprechen und Hörverstehen gemeinsam abzielen. In diesem Fall wäre zu vermuten, dass die Kompetenzen in den so vermittelten Teilbereichen parallel anwachsen und in etwa ein gleiches Niveau erreichen. Ein anderes Beispiel stellt ein stark schriftbasierter Unterricht dar, bei dem das Lesen von Texten und das Wissen um grammatikalische Formen im Mittelpunkt stehen, weniger jedoch die eigene Ausdrucksfähigkeit. In einem solchen Fall wäre anzunehmen, dass die rezeptiven Fähigkeiten Lesen und Hörverstehen stärker ausgeprägt sind als die produktiven. Diese Schwerpunktsetzung auf entweder kommunikative oder schriftsprachliche Aspekte findet sich allerdings weniger im Unterricht in der Erstsprache als häufiger bei der institutionellen Vermittlung von Fremdsprachenkompetenz.

## 1.5 Vermittlung von Sprachkompetenz in der Fremdsprache

Ob eine Sprache als Erst- oder Fremdsprache zu betrachten ist, lässt sich über ihren Erwerbskontext definieren. Als *Muttersprache oder Erstsprache* wird jene Sprache bezeichnet, die ein Kind als erste mündliche Sprache von Bezugspersonen sprechen und verstehen lernt und in der Regel als erste Schriftsprache institutionell vermittelt erwirbt. Diese wird unterschieden von einer oder mehreren Fremdsprachen. Als *Zweitsprache* wird eine Sprache bezeichnet, die nach der Erstsprache erworben und im sozialen Alltag verwendet wird. In Deutschland wäre Deutsch für jene Schülerinnen und Schüler als *Zweitsprache* zu bezeichnen, die eine nicht-deutsche Erstsprache erworben haben (Apeltauer, 1997). Einen Spezialfall bilden hierbei Personen, die mit mehr als einer Sprache aufwachsen und meistens als bilinguale oder multilinguale Sprecher bezeichnet werden (vgl. Abschnitt 1.6). Als *Fremdsprache* werden Sprachen bezeichnet, die meist durch institutionelle Vermittlung, oft erst im späteren Entwicklungsverlauf, erworben werden (Zydatið, 2002). Für Schülerinnen und Schüler in Deutschland ist das Englische in den meisten Fällen die erste institutionell vermittelte Fremdsprache (Statistisches Bundesamt, 2004).

Fremdsprachenkompetenz besonders im Englischen ist heute eine Basisqualifikation im alltäglichen und beruflichen Leben und rückt nicht erst seit neuestem in das Blickfeld der schulischen Bildung (Meyer, 2001; Tenorth, 2001). Die institutionelle Vermittlung von Sprachkompetenz in der Fremdsprache baut in der Regel auf den Kompetenzen der Erstsprache auf. So muss in den häufigsten Fremdsprachen an deutschen Schulen kein neues Alphabet erlernt werden. Setzte der Fremdsprachenunterricht in Deutschland noch bis vor wenigen Jahren traditionell in der fünften Jahrgangsstufe ein, so zeichnet sich seit einigen Jahren die Tendenz ab, die Fremdsprache Englisch in fast allen deutschen Bundesländern bereits in der Grundschule zu unterrichten und auch in Kindergärten finden sich zunehmend Ansätze der frühen Fremdsprachvermittlung (Brusch, 2000; Wode, Burmeister, Daniel, Kickler & Knust, 1996; Wode, Devich-Henningsen, Fischer, Franzen & Pasternak, 2001).

Eine wesentliche Unterteilung besteht im institutionellen Fremdsprachenerwerb zwischen den grammatikbasierten, linguistisch orientierten Lehransätzen und den kommunikativ-funktionalen Herangehensweisen der Sprachvermittlung, seltener existieren auch situative Ansätze (Canale & Swain, 1980; Diehl, Christen, Leuenberger, Pelvat & Studer, 2000; Vollmer et al., 2001). Meistens beginnt der Unterricht auf einer schriftfreien,

kommunikativen Ebene, um einen einfachen und effektiven Einstieg in die neue Sprache zu ermöglichen (Brusch, 2000). Dabei werden anfangs besonders verbale kommunikative Sprachkompetenzen, also das Hörverstehen und das Sprechen in der Fremdsprache, trainiert. Die institutionalisierte Vermittlung von Fremdsprachen ist besonders in höheren Jahrgangsstufen jedoch auch immer noch charakterisiert durch akademische Diskurse im Unterricht und eine Fokussierung auf schriftsprachliche Fähigkeiten, eine Einbeziehung von kommunikativen, alltäglichen Situationen in den Fremdsprachenunterricht ist dabei oft noch die Ausnahme (Kramsch, 2002). Entsprechend finden sich vermehrt Plädoyers für die Zusammenführung dieser kommunikativen und kognitiven Ansätze (Henrici, 2000; Riemer, 2000). Aktuelle didaktische Ansätze heben hervor, dass der Fremdspracherwerbsprozess nicht linear verläuft, sondern dass Teilbereiche der Fremdsprachenkompetenz unterschiedlich schnell erworben werden. Wie auch beim Erwerb der Erstsprache bildet eine ausgedehnte Phase des Hörverstehens die Grundlage des Fremdspracherwerbs, dem – individuell und situativ unterschiedlich – die Sprachproduktion nachfolgt (Bleyhl, 2005). Spezifische Teilkompetenzen, wie eine ausgefeilte Sprachbewusstheit und Schriftsprachproduktion auf einem höheren Kompetenzniveau, können erst beobachtet werden, wenn ein grundlegender Wortschatz vorhanden ist (Ingendahl, 1999).

So ist bezüglich der Struktur von Fremdsprachenkompetenz zu vermuten, dass sie, bedingt durch einen später beginnenden Erwerb, differenzierter sein könnte als die der Kompetenz in der Erstsprache, sodass eine parallele Sprachbeherrschung in allen Teilbereichen nicht zum gleichen Zeitpunkt wie in der Erstsprache erreicht werden kann (vgl. Abschnitt 1.8). Es ist anzunehmen, dass der Fremdsprachenunterricht je nach didaktischer Ausrichtung gewisse Teilbereiche stärker in den Blick nimmt, seien es die kommunikativen oder die eher sprachwissensbezogenen, und sich dadurch ein differenzierteres Kompetenzprofil ergeben könnte, in dem die Teilbereiche nicht durchgängig hoch miteinander zusammenhängen. Im Kontext der vorliegenden Arbeit wird diesen Phänomenen dementsprechend differenziert betrachtet, indem die Struktur von Sprachkompetenz auf Individualebene und auf Klassenebene parallel modelliert wird (vgl. Abschnitt 5.3). Ein etwas anderes Bild könnte sich bei jenen Personen ergeben, die mit zwei oder mehreren Sprachen aufwachsen, also über bilinguale Sprachkompetenzen verfügen. Hier ist zu vermuten, dass sich je nach Erwerbkontext in verschiedenen Sprachen unterschiedliche Kompetenzstrukturen finden lassen.

## 1.6 Mehrsprachigkeit und bilingualer Unterricht

Mehrsprachigkeit, also die Fähigkeit, mehr als eine Sprache zu erwerben und zu verwenden, ist ein normaler Bestandteil menschlichen Sprachvermögens (Dehaene, 1999; Reich & Roth, 2002). Der Spracherwerb kann entweder parallel verlaufen, was auch als Bilingualität bezeichnet wird, oder zeitlich versetzt und somit aufeinander aufbauend, wobei Transfereffekte der Erst- auf die Zweitsprache sowohl förderliche als auch hinderliche Bedingungen darstellen (Faerch & Kasper, 1987). Der simultane Bilingualismus tritt oft als natürlicher Zweitspracherwerb in der frühen Kindheit auf, ist meist ungeleitet und führt dazu, dass beide Sprachen wie eine Erstsprache erworben werden (Mohanty & Perregaux, 1997).

Diskutiert werden unterschiedliche Theorien der Mehrsprachentwicklung: Nach der *Interdependenzhypothese* ist die Entwicklung der Kompetenz in der Zweitsprache abhängig von den Fähigkeiten in der Erstsprache, sodass beide Sprachen gleichermaßen zu fördern seien, um negative Auswirkungen zu vermeiden (Cummins, 1980). Demgegenüber stehen Annahmen, die davon ausgehen, dass erst in einer Sprache ein gewisses Kompetenzniveau als *Schwelle* erreicht werden muss, bevor sinnvoll eine weitere Sprache erlernt werden kann. Die bereits erlernten Regeln und Strukturen der ersten Sprache könnten dann auf die nachfolgende übertragen werden (Wolff, 2001).

Inzwischen finden sich empirische Hinweise dafür, dass bilinguale Kinder schon im Vorschulalter – also auch bei noch nicht perfekter Sprachkompetenz – über ein stärker ausgeprägtes meta-linguistisches Wissen verfügen, welches den weiteren Spracherwerb in der Schule positiv beeinflusst. In späteren Entwicklungsphasen wirken sich besonders schriftsprachliche Kenntnisse in einer oder mehreren Sprachen positiv auf den Erwerb von weiteren Sprachen aus (Cummins, 1991, 1993). Von wesentlicher Bedeutung für eine umfassende mehrsprachige Kompetenz ist das Alter, in dem eine Sprache erworben wird, die Sprache als solche sowie ihr soziokultureller Status (Bialystok & Miller, 1999; Hamers & Blanc, 2000). Gegen die weitverbreitete Annahme, dass Fremdsprachen möglichst in einem frühen Alter erworben werden sollten, sprechen die uneinheitlichen empirischen Befunde zu Alterseffekten beim Fremdspracherwerb (Chiswick & Miller, 2007; Flege, 1987). Zwar fällt es jüngeren Kinder oft leichter, phonologische und syntaktische Systeme einer neuen Sprache praktisch intuitiv zu erlernen, allerdings verfügen sie oft noch nicht über schriftsprachliche Kenntnisse, die beim Fremdsprachenlernen im höheren Alter eine wesentliche Unterstützung des Spracherwerbsprozesses darstellen (Mohanty & Perregaux,

1997). Eine frühe Konfrontation mit einer Fremdsprache kann deren Erwerbsprozess beschleunigen und erleichtern (Wode, Devich-Henningsen, Fischer, Franzen & Pasternak, 2001). Doch Fremdsprache ist nicht gleich Fremdsprache und deshalb ist Bilingualismus kein einheitliches Phänomen (Hufeisen, 2003): Die Dekodierfähigkeit kann beispielsweise eine Funktion einer speziellen Sprache sein, die sich nicht automatisch auf die Zweitsprache übertragen lässt: Lesen muss in Abhängigkeit vom Alphabet unter Umständen zweifach erlernt werden, da lateinische, kyrillische oder asiatische Schriftzeichen ein jeweils eigenes System darstellen (Bialystok, McBride-Chang & Luk, 2005). Aktuelle neuropsychologische Studien zeigen, dass Erst- und Zweitsprache bei Bilingualen zwar primär in gemeinsamen Hirnarealen repräsentiert sind, dass aber je nach spezifischer Sprachanwendung durchaus auf unterschiedliche neuronale Netzwerke zurückgegriffen wird (Dehaene et al., 1997).

Im Kontext der vorliegenden Arbeit ist es von besonderem Interesse, bei der Analyse des Phänomens der Mehrsprachigkeit die Struktur von Sprachkompetenz in den einzelnen Sprachen miteinander zu vergleichen. So kann festgestellt werden, ob verschiedene Teilbereiche in mehreren Sprachen gleich beherrscht werden, oder ob bestimmte rezeptive oder produktive Faktoren voneinander zu trennen sind und wie diese zwischen Erst- und Fremdsprache miteinander zusammenhängen.

Da Mehrsprachigkeit als durchaus anzustrebende Fähigkeit angesehen wird und nicht zuletzt im Rahmen der europäischen Bildungspolitik ein wesentliches Ziel darstellt, bedienen sich fremdsprachendidaktische Ansätze vermehrt des bilingualen Sachfachunterrichts, von dem sie sich eine Verbesserung der Fremdsprachenkompetenz erhoffen (Europäische Kommission, 2002; Nelde, 2001; Zydati, 2000). Der bilinguale Sachfachunterricht ist ein erweiterter Fremdsprachenunterricht, bei dem Sachfächer ganz oder teilweise in der Fremdsprache unterrichtet werden. Diese Unterrichtsform existiert in Deutschland seit den 1960er Jahren und betrifft in der Regel die Fremdsprachen Englisch und/oder Französisch (Wolff, 1997). Bilingualer Unterricht wird überwiegend in den gesellschaftswissenschaftlichen Fächern wie Erdkunde oder Geschichte angeboten, in den letzten Jahren auch zunehmend in naturwissenschaftlichen oder musischen Fächern (Wenderott, 2005).

Die vorhandenen empirischen Untersuchungen zur Effizienz des bilingualen Sachfachunterrichts weisen auf positive Auswirkungen hin: Schülerinnen und Schüler, die am bilingualen Unterrichtsangebot teilnehmen, weisen stärker fundierte

Fremdsprachenkenntnisse, eine höhere allgemeine Sprachbewusstheit sowie mehr fachliches und politisches Wissen über den eigenen Kulturkreis hinaus auf (Fehling, 2003; Wode, Burmeister, Daniel, Kickler & Knust, 1996). Besonders im Hörverstehen, einer wesentlichen Voraussetzung für fremdsprachliche Kommunikation, zeigt sich ein klarer Vorsprung bilingual unterrichteter Schülerinnen und Schüler (Klieme et al., 2006). Die sprachlichen Kompetenzen von Schülerinnen und Schülern in bilingualen Unterrichtsangeboten gleichen zwar damit nicht jenen, die beim natürlichen Mehrsprachenerwerb aufgebaut werden, übersteigen aber die Fremdsprachenkompetenzen, die im normalen Fremdsprachenunterricht vermittelt werden können. Als eine Ursache hierfür wird die quantitativ höhere Lernzeit angesehen, da die Lernenden mit der fremden Sprache häufiger und länger konfrontiert werden als in Regelklassen. Auch die Fachinhalte stellen eine Lernumgebung her, in der Sprache authentisch und nicht durch pseudoreale Lerninhalte vermittelt wird und in der Kommunikation und Interaktion stärker gefördert werden (Wolff, 1997). Dieses fördert Interesse und Motivation für die bilingualen Fächer, wie in Befragungen von Lernenden festgestellt werden konnte (Abendroth-Timmer, 2004). Es ist daher zu vermuten, dass sich nicht nur die Ausprägung, sondern auch die Struktur von Fremdsprachenkompetenz zwischen bilingual und nicht-bilingual unterrichteten Schülerinnen und Schülern unterscheidet.

## 1.7 Kontextfaktoren und Sprachkompetenz

Kontextfaktoren haben einen nicht unerheblichen Einfluss auf Lernleistungen und den Kompetenzerwerb im sprachlichen Bereich. Als Kontextfaktor kann die Zusammensetzung von Lerngruppen angesehen werden, wie sie sich beispielsweise in Schulklassen darstellt. Obwohl das individuelle Leistungsniveau von Schülerinnen und Schülern in einer Schulklasse variieren kann, stellen Schulklassen eine je spezifische, unterschiedlich anregende Lernumwelt dar (Hosenfeld, Helmke, Ridder & Schrader, 2001). Der Einfluss der Zusammensetzung dieser Lerngruppen auf Leistung, insbesondere den Spracherwerb und Sprachstand, wird in vielen Untersuchungen aufgezeigt. So geht ein höherer Durchschnitt in Tests zu kognitiven Fähigkeiten einher mit höherer sprachlicher Leistung auf Klassenebene (De Fraine, Van Damme, Van Landeghem, Opendakker & Onghena, 2003). Auch auf Individualebene lassen sich Unterschiede im (Fremd-)Spracherwerb zwischen Personen auf allgemeine kognitive Fähigkeiten (Rost, 1989) oder kognitive Verarbeitungsgeschwindigkeit (Robinson, 2001) zurückführen. Empirisch zeigen sich positive Zusammenhänge zwischen allgemeinen kognitiven Grundfähigkeiten und dem Erwerb von Fremdsprachen, die im mittleren Bereich liegen (Sasaki, 1993, 1996). Über differenzierte spezifische Zusammenhänge zwischen Intelligenz und Sprachkompetenz berichten Teepen (2004) und Genesee (1976): Während sich hohe Zusammenhänge zwischen Intelligenz und Lesekompetenz finden lassen, spielen die kognitiven Grundfähigkeiten bei kommunikativen Sprachkompetenzen eine geringere Rolle.

Aber nicht nur kognitive Grundfähigkeiten sind im Zusammenhang mit dem Erwerb von Sprachkompetenz zu betrachten, sondern auch das Geschlecht der Lernenden. In vielen empirischen Studien finden sich Unterschiede in der Sprachkompetenz, die oft zugunsten der weiblichen Personengruppen ausfallen. Mädchen beginnen früher zu sprechen und weisen in der Grundschule höhere Leistungswerte in der Lesekompetenz und der Rechtschreibung auf als Jungen in der gleichen Jahrgangsstufe (Bornstein, Hahn & Haynes, 2004; Bos, Lankes, Prenzel et al., 2003). Vorteile der Jungen hingegen finden sich oft in einem größeren Fachwortschatz auf bestimmten Gebieten (Mais, 2006). Neuere Studien deuten jedoch darauf hin, dass sich diese Unterschiede im Entwicklungsverlauf verändern können. Für den Deutsch- und Fremdsprachenunterricht lässt sich feststellen, dass der Vorteil der Mädchen in diesen Fächern auch in höheren Jahrgangsstufen vorhanden ist, diese Geschlechtsunterschiede jedoch zwischen Klassen und Bildungsgängen stark variieren (Dresel, Stöger & Ziegler,

2005), ein Phänomen, das auch mit dem Anteil von Mädchen in einer Klasse in Zusammenhang stehen kann (De Fraine, Van Damme, Van Landeghem, Opendakker & Onghena, 2003). Bei der Auswahl des Testmaterials ist der Aspekt der Testfairness zu berücksichtigen; besonders Lesetests verwenden oft Material, das thematisch näher an den Interessen von Mädchen liege und diese dadurch bevorzugen könnte (Hyde & Linn, 1988). Über Effekte von Motivation und Interesse, die zu einem besseren Lernergebnis bei Schülerinnen führen könnten, wird besonders auch für den Bereich der Fremdsprachenkompetenz berichtet (Coleman, 1997; Dresel et al., 2005).

Zur Frage, ob sich die Struktur von Sprachkompetenz zwischen Jungen und Mädchen unterscheidet, lassen sich vor dem Hintergrund aktueller Forschungsliteratur keine eindeutigen Hypothesen formulieren. Weisen Jungen in Sprachtests in verschiedenen Bereichen durchweg niedrigere Leistungen auf als Mädchen, so wird sich dies nicht zwingend in der Struktur der Kompetenz niederschlagen, da keine zwangsläufigen Profilunterschiede bestehen. Verhält es sich jedoch so, dass in bestimmten Teilbereichen Entwicklungsvorsprünge zu beobachten sind, so könnte dies durchaus dazu führen, dass in einer bestimmten Altersgruppe sich die Kompetenzstruktur der Mädchen homogener darstellt als die der Jungen. Dabei sollte jedoch berücksichtigt werden, dass nach Ergebnissen der Study of Reading Literacy (Elley, 1994) Entwicklungsvorsprünge im sprachlichen Bereich auch noch in der Adoleszenz aufgeholt werden können (Wagemaker, 1996).

## **1.8 Sprachliche Kompetenzentwicklung – Konvergenz oder Divergenz?**

Wenn sich sprachliche Kompetenzen in verschiedenen Teilbereichen natürlich oder institutionell vermittelt entwickeln, ist zu untersuchen, inwieweit diese Entwicklungen parallel verlaufen: Entwickeln sich Kompetenzbereiche gemeinsam und beeinflussen sie sich gegenseitig derart, dass bei insgesamt hoher Kompetenz alle Bereiche gleichermaßen beherrscht werden? Oder verhält es sich so, dass sich lediglich in bestimmten Facetten eine Expertise herausdifferenziert, in anderen hingegen nur eine durchschnittliche Kompetenz entwickelt wird?

Diese Fragen sind besonders dann relevant, wenn standardisierte Testverfahren zwar differenziert messen, jedoch zur Beurteilung oder Einstufung von Personen einen Gesamtwert bilden und verwenden. Je nach allgemeinem Kompetenzgrad würde ein Gesamtwert eventuell vorhandenen Profilen in Teilkompetenzen nicht gerecht werden, die Validität eines solchen Tests wäre zu hinterfragen. Besonders bei sehr leistungsschwachen oder sehr leistungsstarken Personen könnte ein Gesamtwert spezifische Fähigkeiten zu wenig berücksichtigen, auf die weder in der Diagnose noch bei der Förderung eingegangen werden könnte. Auch die Vergleichbarkeit von Testwerten zwischen Gruppen auf unterschiedlichem Leistungsniveau wäre nicht mehr gegeben, stattdessen müssten gruppenspezifische Auswertungen erfolgen (Kunnan, 1992).

Besonders wichtig ist der Aspekt der Divergenz beim Einsatz von Testverfahren, die zwar verschiedene Bereiche von Sprachkompetenz erfassen, den diagnostischen Entscheidungen jedoch einen Gesamtwert zugrunde legen. Empirische Ergebnisse legen nahe, dass solche Gesamtwerte, bei denen Leistungen in unterschiedlichen Teilbereichen letztendlich gleich gewertet werden, besonders Personen auf einem niedrigen Niveau der Sprachkompetenz benachteiligen (Kunnan, 1992). Bei solchen diagnostischen Gesamtentscheidungen werden spezifische Kenntnisse nicht berücksichtigt, Personen mit niedriger Sprachkompetenz daher eher falsch eingestuft, wenn es beispielsweise um die Zuordnung zu Sprachkursen geht.

Einige Befunde legen die Annahme nahe, dass sich mit zunehmender Gesamtkompetenz die Ausprägung in unterschiedlichen Teilkompetenzen einander annähert, sodass die schriftsprachlichen Leistungen in der Muttersprache dem Niveau der auditiven Leistungen zunehmend entsprechen (Oerter & Montada, 2002). Die Strukturen von Sprachkompetenz in der Muttersprache würden sich in diesem Fall bei Erwachsenen in einem homogenen,

globalen Faktor besser abbilden lassen als in einem mehrdimensionalen Modell. Entsprechende Modelle finden sich in linguistischen Ansätzen der 1970er Jahre, die von einer singulären übergreifenden Sprachkompetenz ausgehen und alle spezifischen sprachlichen Fähigkeiten hierauf beziehen (Kunnan, 1992; Oller, 1976). Eine Begründung für homogenere Strukturen bei hohem Sprachniveau wird im Transfer zwischen sprachlichen Kompetenzbereichen gesehen. Wer in seiner Erstsprache auf einem funktionalen Niveau lesen und schreiben kann, wird sich auch mündlich dementsprechend ausdrücken können. Verschiedene aktuelle Studien finden diese theoretische Annahme zumindest teilweise bestätigt und beschreiben einen Zusammenhang zwischen sprachlichen Teilbereichen, die sich über die Zeit hin zu einem homogenen Kompetenzfaktor entwickeln:

Schülerinnen und Schüler, deren Lesekompetenz hoch ist, erzielen gute Leistungen in produktiven Schreibaufgaben und umgekehrt; dieser Zusammenhang wird über die Zeit stärker, worauf die Ergebnisse verschiedener Längsschnittstudien hinweisen (Heller, 1999; Langer & Flihan, 2000). Der Zusammenhang zwischen Lesekompetenz und Rechtschreibung zeigt sich sowohl bei der Erstsprache als auch bei Fremdsprachen: Schülerinnen und Schüler, die über eine hohe Lesekompetenz verfügen, weisen bessere Leistungen in der Rechtschreibung auf, unabhängig davon, ob diese Sprache ihre Erst- oder eine Fremdsprache ist (Wade-Wooley & Siegel, 1997). Über eine Angleichung der sprachlichen Teilfähigkeiten mit zunehmender Gesamtkompetenz ist auch für die Bereiche des Lese- und des Hörverstehens berichtet worden, welche nach Rost und Hartmann (1992) im Erwachsenenalter etwa vergleichbar gut ausgeprägt sind, während im Grundschulalter das Hörverstehen noch dominiert: „Offensichtlich nimmt die Korrelation mit dem Alter, d. h. mit zunehmender Automatisierung des Leseprozesses [...] zu“ (S. 348). Zu entsprechenden Ergebnissen kommt auch eine Querschnittsstudie in den Jahrgangsstufen zwei, vier, sechs und acht, die feststellt: „[...] the overall relationship between listening and reading comprehension becomes significantly stronger after second grade, when word decoding skills can be assumed to have been mastered to a satisfactory degree“ (Diakidoy, Stylianou, Kerefillidou & Papageorgiou, 2005, S. 67).

Trotz der offensichtlichen Gemeinsamkeiten von bestimmten sprachlichen Teilbereichen existieren Befunde, die auf eine Mehrdimensionalität von Sprachkompetenz auch auf einem hohen Gesamtniveau, also eher auf eine *Divergenz* von Kompetenz hinweisen. In Abhängigkeit von der Differenziertheit der Kompetenzerfassung lassen sich durchaus

mehrdimensionale Modelle empirisch bestätigen. In ihren Studien zeigen beispielsweise Shanahan (1984) sowie Shanahan und Lomax (1986), dass Lesekompetenz und Schreibkompetenz zwar durchaus gemeinsame Varianz aufweisen, jedoch Einzeltests zum Textverstehen sowie zum lesespezifischen Wortschatz und zur Rechtschreibung nicht auf einen gemeinsamen Faktor zurückzuführen sind. Ergebnisse neuerer Studien weisen darauf hin, dass sich die Beziehung zwischen Lesen und Schreiben im Entwicklungsprozess generell ändert und in verschiedenen Stadien der Kompetenz gegenseitige Einflüsse unterschiedlich stark wirken (Fitzgerald & Shanahan, 2000). Teilbereiche sprachlicher Kompetenzen können besonders bei allgemein hoher Sprachkompetenz besser differenziert werden, während Testwerte in sprachlichen Teilkomponenten bei Personen auf einem niedrigen Kompetenzniveau stärker korrelieren (Ginther & Stevens, 1998).

Exemplarisch sei hier eine Studie angeführt, welche die Faktorenstruktur des Test of English as a Foreign Language (TOEFL) bei Personen mit unterschiedlicher Erstsprache untersucht hat (Swinton & Powers, 1980). In der Gruppe mit der durchschnittlich höchsten Englischkompetenz (Erstsprache Deutsch) ließen sich vier verschiedene Faktoren differenzieren: Hörverstehen, Schreiben, Lesen und ein unbenannter Faktor. In der nächstbesten Gruppe (Erstsprache Spanisch) konnte eine dreifaktorielle Lösung erzielt werden, während in der Gruppe mit der niedrigsten Englischkompetenz (Erstsprache Farsi) alle Tests auf einem einzigen Faktor luden. Diese Ergebnisse wurden damit begründet, dass Personen mit wenigen Sprachkenntnissen eben weder gut lesen noch gut sprechen oder schreiben können (Powers, 1982). Dementsprechend korrelieren Sprachtestwerte auf einem niedrigen Fähigkeitsniveau so sehr miteinander, dass sie sich auf einem Faktor abbilden lassen. Anders formuliert: Bei höherer Sprachkompetenz ist die Varianz zwischen Items auf einen höheren Anteil an wahrer Varianz zurückzuführen, und dementsprechend differenzierter können faktorielle Strukturen abgebildet werden.

Besonders bei Personen, die eine Fremdsprache lernen, wird diese zusätzliche Sprache oft nur partiell beherrscht: Die mündliche Kommunikation fällt meist wesentlich leichter als das Lesen von Schriftstücken, und das Verfassen fremdsprachlicher Briefe oder Aufsätze muss nicht zwingend mit der verbalen Ausdrucksfähigkeit zusammenhängen. Auch bei bilingual aufgewachsenen Personen ist die schriftsprachliche Kompetenz in beiden Sprachen nicht unbedingt gleich gut. Insgesamt lässt sich feststellen, dass die Struktur von Sprachkompetenz in Abhängigkeit von Alter, allgemeiner Sprachfähigkeit und auch vom Sprachhintergrund der

untersuchten Personengruppen durchaus variieren kann. Diese Hintergrundfaktoren sollten bei Analysen zur Struktur von Sprachkompetenz soweit möglich berücksichtigt werden (North, 2000).

## 1.9 Theorien und Modelle der Sprachkompetenz – Zusammenfassung

Sprachkompetenz kann als eine spezifisch menschliche Fähigkeit angesehen werden, die sich im Laufe der individuellen Entwicklung in vielen verschiedenen Teilbereichen ausformt. Dementsprechend sollten die Kompetenzen in der Erstsprache und in den Fremdsprachen als unterschiedlich ausdifferenzierte Strukturen von Sprachkompetenz beschrieben werden können. Nicht alle sprachlichen Teilfähigkeiten werden simultan ausgebildet, vielmehr werden schriftsprachliche Kompetenzen wie Lesen und Schreiben in der Regel erst bei vorhandenen grundlegenden mündlichen Fähigkeiten erworben und zu einem großen Teil institutionell vermittelt. Auch Erwerbsprozesse in der Fremdsprache folgen unterschiedlichen Mustern – je nach Erwerbkontext erfolgt hier die Vermittlung sprachlicher Teilkompetenzen simultan oder auf spezifische kommunikative oder rezeptive Bereiche fokussiert. Die Fähigkeitsniveaus in sprachlichen Teilbereichen der Erst- und der Fremdsprache nähern sich im Entwicklungsverlauf zwar einander an, es wird jedoch nicht unbedingt in allen Bereichen das gleiche Niveau erreicht. Dies gilt besonders für den Fremdspracherwerb, der je nach Vermittlungskontext zu durchaus heterogenen Kompetenzprofilen in den Teilbereichen führen kann. Mit zunehmender Gesamtkompetenz ist daher einerseits eine Tendenz hin zu einer Eindimensionalität von Sprachkompetenz anzunehmen, andererseits können sich auch ausdifferenzierte Fähigkeiten in spezifischen Teilkompetenzen entwickeln, was wiederum zu einem eher heterogenen Kompetenzkonstrukt führen kann. Die existierenden *Theorien und Modelle* der Struktur von Sprachkompetenz weisen diesbezüglich durchaus heterogenes Bild auf. Im nächsten Abschnitt wird daher vertieft und differenziert auf *aktuelle empirische Befunde* zur Struktur von Sprachkompetenz eingegangen.

## **2 Empirische Befunde zur Struktur von Sprachkompetenz**

Das Konstrukt der Sprachkompetenz ist vielschichtig und seine Definitionen sind, wie bereits aufgezeigt wurde, zahlreich. Im folgenden Abschnitt werden empirische Studien vorgestellt, die auf verschiedene Weisen Sprachkompetenz in der Erst- und Fremdsprache erfassen und deren Struktur analysieren. Speziell wird dargestellt, welche Teilbereiche differenziert werden können und ob diese in spezifischer Weise miteinander zusammenhängen. Neben umfassenden Modellen, die viele verschiedene Teilbereiche von Sprachkompetenz gleichzeitig betrachten, werden Zusammenhänge zwischen einzelnen Teilbereichen gesondert dargestellt. Ein weiterer Abschnitt geht auf Befunde zu strukturellen Zusammenhängen zwischen Erst- und Fremdsprachen ein.

### **2.1 Die vier Grunddimensionen von Sprachkompetenz**

Bezüglich der Struktur von Sprachkompetenz sind verschiedene konkurrierende theoretische Ansätze zu verzeichnen. Strukturalistische Modelle für den Bereich der Fremdsprachenkompetenz gehen davon aus, dass diese sich aus verschiedenen Komponenten zusammensetzt. Das Modell von Lado (1961) führt die sechs Elemente Aussprache, Intonation, Betonung, Grammatik, Wortschatz und Übersetzung an, die sich jeweils in unterschiedlichem Ausmaß in den vier Grunddimensionen von Sprachkompetenz niederschlagen: im Lesen, Schreiben, Sprechen und Hörverstehen. Von Harris (1969) werden diesen vier Grunddimensionen die Komponenten der Phonologie, der Struktur, des Wortschatzes und der Sprachflüssigkeit zugeordnet. Er unterscheidet darüber hinaus die sprachlichen Prozesse des Enkodierens und des Dekodierens sowie die geschriebene und die gesprochene Sprache voneinander. Je nach Prozess und Modalität lassen sich die Fähigkeiten und Komponenten unterschiedliche klassifizieren (vgl. Abbildung 1).

		Modalität	
		Gesprochen auditiv	Geschrieben visuell
		Sprechen	Schreiben
Prozess	Enkodieren produktiv	Grammatik Wortschatz Lautsystem Flüssigkeit	Grammatik Wortschatz Rechtschreibung
		Hörverstehen	Lesen
	Dekodieren rezeptiv	Grammatik Wortschatz Lautsystem	Grammatik Wortschatz Geschwindigkeit

**Abbildung 1: Schematisches Modell der vier Grunddimensionen von Sprachkompetenz**

Auf der Unterscheidung der vier Grunddimensionen des *Sprechens*, des *Schreibens*, des *Hörverstehens* und des *Lesens* basieren viele Theorien und empirische Ansätze zur Struktur von Sprachkompetenz. Diese vier Grunddimensionen können entweder noch stärker differenziert oder zu übergeordneten Faktoren zusammengefasst werden. Sie beziehen sich sowohl auf die Erst- als auch auf Fremdsprachen und lassen sich in vielen empirischen Studien auffinden, manchmal als grundlegende vier Dimensionen, manchmal auch als übergeordnete latente Konstrukte, denen Testaufgaben mit unterschiedlichen Inhalten zugeordnet werden (vgl. Abschnitt 2.2). Diese vier Dimensionen bilden die Grundlage für übergeordnete Faktoren: *Sprachmodalitäten*, die sich in auditive Kompetenzen (Hörverstehen und Sprechen) sowie schriftsprachlichen Kompetenzen (Lesen und Schreiben) unterteilen lassen, und *Prozessdimensionen* der produktiven (Sprechen und Schreiben) und rezeptiven (Hören und Lesen) Sprachkompetenz.

Die Analyse der Zusammenhänge dieser vier Grunddimensionen ist Bestandteil der angewandten Linguistik, welche sich explizit für eine mehrdimensionale Struktur der Kompetenz ausspricht (Spolsky, 1989). Dabei werden die Grunddimensionen selbst noch stärker differenziert und beispielsweise unterschiedliche Teilkompetenzen der kommunikativen Sprachfähigkeit postuliert. Die grammatikalische, die soziolinguistische, die strategische und die Diskurskompetenz (Canale & Swain, 1980) oder auch die organisatorische Kompetenz, die formale Aspekte der Sprache wie Lexik und Syntax umfasst, sowie die pragmatische Kompetenz, die sich auf soziolinguistische und strategische Bereiche der Sprachverwendung bezieht (Bachman, 1990b).

Diesen verschiedenen Theorien ist gemeinsam, dass sie Fremdsprachenkompetenz zwar als mehrdimensionales Konstrukt ansehen, jedoch jeweils unterschiedliche Zusammenhänge der Einzelkomponenten annehmen und so schließlich wieder zu übergeordneten Dimensionen kommen. In diesen drückt sich das Zusammenspiel vieler Einzelelemente aus: Im Kommunikationsprozess müssen Dekodierung und Enkodierung auf der Grundlage von Sprachwissen und Sprachkönnen aneinander anschließen. Die Beherrschung lediglich einer Grundfähigkeit, beispielsweise des Hörverstehens, würde zu einer unproduktiven Unterbrechung einer potenziellen sprachlichen Verständigung führen. Es verwundert daher nicht, dass frühe empirische Ansätze die Annahme einer Mehrdimensionalität von Sprachkompetenz zunächst nicht bestätigen konnten.

## 2.2 Empirische Ansätze

Empirische Studien, die Zusammenhänge zwischen sprachlichen Teilkomponenten analysieren, lassen sich grob in zwei Kategorien einteilen: Solche Studien, die nur *einen* übergeordneten Faktor empirisch auffinden, und solche, deren Ergebnisse *mehrdimensionale* Modelle unterstützen. Einschränkend hinzuweisen ist hierbei auf die oft mangelnde Vergleichbarkeit von Ergebnissen aus unterschiedlichen Untersuchungen, da diese Studien selten repliziert oder re-analysiert werden. Auch sind die verwendeten Testverfahren von Studie zu Studie unterschiedlich, ebenso wie die untersuchten Alters- und Sprachgruppen. Ohnehin sind empirische Studien, die sich mit einer *umfassenden Charakterisierung* der Struktur von Sprachkompetenz befassen und entsprechend viele verschiedene Teilbereiche erfassen und ihre Zusammenhänge analysieren, kaum zu finden. Aufgrund des hohen Aufwandes der betrieben werden müsste, um an einer repräsentative Stichprobe eine Vielzahl unterschiedlicher Sprachbereiche zu erheben, finden sich in der empirischen Forschung meist lediglich Befunde zum Zusammenhang zwischen zwei oder maximal drei sprachlichen Teilbereichen.

Ein Schwerpunkt ist zu verzeichnen bei solchen Studien, die Lesekompetenz und Schreibkompetenz erheben. Erklären lässt sich dies dadurch, dass sowohl Lesen als auch Schreiben in der Regel institutionell vermittelt werden und die Ergebnisse dieser Lernprozesse oft auch evaluiert werden. Im Gegensatz dazu sind der Erwerb von Sprechen und Hörverstehen zumindest in der Erstsprache Entwicklungsprozesse, die im frühesten Kindesalter ablaufen und mit der Einschulung größtenteils ein funktionales Niveau erreicht haben, das selten empirisch hinterfragt wird. Diese auditiv-kommunikativen Bereiche von Sprachkompetenz spielen dann erst wieder beim Fremdspracherwerb eine Rolle. Empirische Studien, die mündlich-produktive Kompetenzen erfassen, sind auf Grund des höheren Testaufwands generell seltener zu finden, ohnehin beschränken sich die Testformate vieler Studien auf geschlossene Antworten.

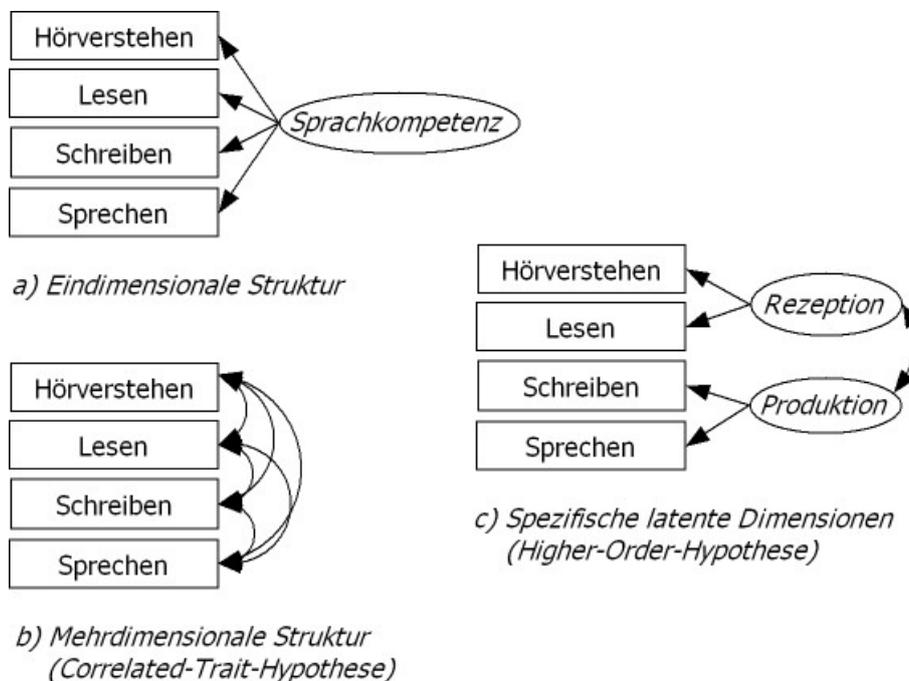
Welche Hypothesen bezüglich der faktoriellen Struktur gestützt werden – die eindimensionale oder die mehrdimensionale –, hängt dann zusätzlich von der Auswahl der Testinstrumente ab: So besteht eher eine Tendenz zum Auffinden eines Generalfaktors, wenn mit ähnlichen Aufgabenformaten ähnliche Kompetenzen erhoben werden, besonders wenn dabei Testwerte zur (freien) mündlichen und schriftlichen Sprachproduktion fehlen (Gräfe-Bentzien, 2000).

Viele Studien, die über einen Generalfaktor der Sprachfähigkeit berichten, basieren dementsprechend auf solchen Tests, die nur wenige sprachliche Teilbereiche erfassen und führen zu Ergebnissen, in denen der berichtete Generalfaktor nur etwa 50 % der Varianz erklären kann.

Die wohl bekannteste und immer noch häufig zitierte Studie, die einen Generalfaktor der Sprachkompetenz postuliert, stammt von Oller (1976). Sie basiert auf der Annahme, dass der sprachliche Informationsverarbeitungsprozess auf einem grundlegenden Faktor basiert und bei der Sprachverwendung simultan auf verschiedene Teilkompetenzen zurückgegriffen wird. Oller nutzte die damals noch relativ neue Technik der exploratorischen Faktorenanalyse, um Datensätze aus vier Stichproben und zwei verschiedenen Testformen des Testverfahrens English as a Second Language Placement Examination (ESLPE) hinsichtlich ihrer strukturellen Zusammenhänge zu analysieren. Er fasste seine Ergebnisse dahingehend zusammen, dass die Annahme eines universalen Traits im Sinne einer globalen Sprachfähigkeit – der *expectancy grammar* – zu bestätigen sei, da ein einziger Faktor 74,4 % der Varianz im Modell erklärt (Oller, 1976). Oller analysierte seine Ergebnisse nicht konfirmatorisch, trotzdem gingen sie als bis heute noch viel zitierte Belege für die Eindimensionalität der Sprachkompetenz in die Forschung ein. Als Folge dieser Befunde wurden verstärkt Sprachtests eingesetzt, die Sprachkompetenz umfassend und als Ganzes erfassen sollten und deren Testformate dementsprechend keine Differenzierung von sprachlichen Teilbereichen mehr ermöglichten (Feldmann, Grotjahn & Stemmer, 1986; Klein-Braley, 1985; Raatz, 1985).

Die Ergebnisse von Oller wurden in der Folge jedoch kritisiert, besonders die Tatsache, dass die ausschließliche Verwendung von exploratorischen Faktorenanalysen, die immer einen starken ersten Faktor ergeben, in der Tat zu einem Artefakt geführt haben könnte (Sang & Vollmer, 1980). So beschäftigte sich auch die empirische Forschung zur Fremdsprachenkompetenz wieder zunehmend mit mehrdimensionalen Modellen, in denen Sprachkompetenz stärker differenziert wurde. Besonders zwei konkurrierende Hypothesen wurden dabei in verschiedenen Studien überprüft: Die *Correlated-Trait*-Hypothese und die *Higher-Order*-Hypothese. Erstere geht davon aus, dass Sprachkompetenz auf mehrere spezielle Traits zurückzuführen sei, die sich als separate, jedoch interkorrelierende Faktoren abbilden ließen. In der *Higher-Order*-Hypothese sind diese speziellen Traits ebenfalls

differenzierbar, lassen sich jedoch zu einem oder mehreren latenten Faktoren zusammenfassen (vgl. Abbildung 2).



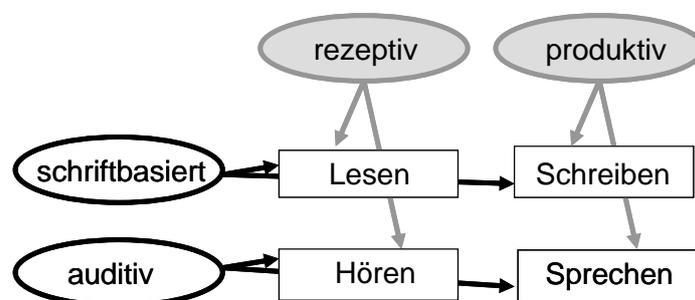
**Abbildung 2: Schematische Darstellung der eindimensionalen (a) und der mehrdimensionalen (b) Struktur von Sprachkompetenz sowie eines mehrdimensionalen Modells mit spezifischen latenten Dimensionen (c)**

Empirische Belege existieren dabei für beide Modelle, wobei jeweils unterschiedliche Teilbereiche erfasst und verschiedene übergeordnete Faktoren gefunden werden (Bachman & Palmer, 1981, 1982). Zumeist finden sich unter den empirisch erfassten sprachlichen Teilbereichen die bereits erwähnten vier Grunddimensionen der Sprachkompetenz wieder. Carroll führt die Daten aus 22 unterschiedlichen Fremdsprachtestskalen auf fünf korrelierende Dimensionen zurück: Sprechen, Hörverstehen, Lesen und Schreiben, mündliches Sprachverständnis sowie eine nicht interpretierbare Dimension (Carroll, 1975). Ähnliche Dimensionalität berichten auch Bachman, Davidson, Ryan und Choi (1995) für eine Re-Analyse von 13 Sprachtestskalen: Die Ergebnisse weisen auf vier korrelierende Dimensionen hin, die als Sprechen, Hörverstehen und zwei weitere Faktoren klassifiziert werden, welche Items zum Lesen und Schreiben gleichermaßen beinhalten.

In dem von Roeder und Treumann (1974) erarbeiteten Überblick über verschiedene Untersuchungen von Schulleistungen im Unterrichtsfach Englisch als Fremdsprache werden

je nach Studie klar voneinander abgrenzbare Leistungsdimensionen berichtet, die sich hinsichtlich der *Sprachverwendung* (mündliche versus schriftliche Kommunikation), der *Habitualisierung* (Flüssigkeit des Sprachgebrauchs), der *linguistischen Ebenen* (Phonologie, Syntax, Wortschatz) und der *Kommunikationssituation* (aktive versus passive Verwendung von Sprache) differenzieren lassen. Auch ein Drei-Faktor-Modell der Fremdsprachenkompetenz mit den Dimensionen *Elementares Wissen* (Aussprache, Orthographie, Wortschatz), *Komplexeres Wissen* (Grammatik, Leseverständnis) und *Kommunikationsfertigkeiten* (Hörverständnis, Sprachproduktion) kann empirisch belegt werden (Sang, Schmitz, Baumert & Roeder, 1985). Neben diesen Befunden für die empirische Trennung unterschiedlicher Teilbereiche, also die *Correlated-Trait-Hypothese*, finden sich genug Hinweise auf Zusammenhänge der Teildimensionen untereinander, die gegebenenfalls durch hierarchisch übergeordnete latente Dimensionen erklärt werden können, die *Higher-Order-Hypothese*.

Felix (1977) findet Belege für die beiden übergeordneten Faktoren der *reproduktiven* und der  *kreativen* Sprachkompetenz. Über ein zweidimensionales Modell mit den modalspezifischen Fähigkeiten der *schriftsprachlichen* Kompetenz und der *auditiven* Kompetenz berichtet Kunnan (1992), die schriftsprachliche Kompetenz bildet dabei einen übergeordneten Faktor für die beiden Teilbereiche des Lesen und Schreibens, die auditive Kompetenz vereint die Fähigkeiten im Hören und Sprechen. Über gegensätzliche Zusammenhangsmuster der vier Basiskompetenzen Lesen, Schreiben, Hören und Sprechen berichtet hingegen North (2000). In seinen Untersuchungen zeigen sich ebenfalls zwei latente Dimensionen, die sich jedoch als *Sprachrezeption und -reflexion* (Lesen und Hören) und *Sprachproduktion* (Sprechen und Schreiben) inhaltlich beschreiben lassen (vgl. Abbildung 3).



**Abbildung 3: Graphische Veranschaulichung unterschiedlicher Modelle der Higher-Order-Hypothese**

Die Entscheidung darüber, wie die Zusammenhänge zwischen den einzelnen Dimensionen interpretiert werden, basiert auf theoretischen Modellannahmen ebenso wie auf den empirisch gefundenen Korrelationen. Diese variieren jedoch in Abhängigkeit vom eingesetzten Testverfahren. So ist es nicht verwunderlich, dass sich für die beiden Modelle von Kunnan und North, die theoretisch gegensätzliche Zusammenhänge postulieren, je nach Studie durchaus empirische Bestätigungen finden lassen.

Generell wird anhand der in diesem Abschnitt aufgeführten Ergebnisse deutlich, dass die Dimensionalität oder faktorielle Struktur von Sprachkompetenz auf Grund der Vielzahl von Theorien und empirischen Befunde keineswegs eindeutig interpretiert werden kann. Offensichtlich lässt sich eine Binnenstruktur ausdifferenzieren, deren Teilkomponenten untereinander oft stark zusammenhängen. Diese Zusammenhänge können auf verschiedene Ursachen zurückgeführt werden: Auf neuronaler Basis werden für bestimmte sprachliche Bereiche gemeinsame Repräsentationen oder geteilte sprachliche Verarbeitungseinheiten angenommen, ebenso wie geteilte sprachliche Prozesse oder symmetrische Systeme der Input-Verarbeitung (McKay, 1987; Monsell, 1987; Venezky & Massaro, 1987). So könnten Lese- und Hörverstehen jeweils als sprachlicher Input in lediglich unterschiedlicher Form angesehen werden, deren gemeinsamem Verstehensprozess gleiche Systematiken zugrunde liegen. Aber auch unter der Annahme getrennter sprachlicher Systeme für Rezeption und Produktion ist von einer gegenseitigen Beeinflussung und Anpassung sprachlicher Systeme auszugehen: Sprachverstehen und Sprechen bedienen sich der gleichen Syntax und eines gemeinsamen Wortschatzes (Cutler, 1987), und im Sprachentwicklungsprozess entwickelt sich das Sprechen durch das Nachahmen gehörter Laute (Studdert-Kennedy, 1987). Aus Perspektive der Lehr-Lern-Forschung können faktorielle Zusammenhänge Rückschlüsse auf spezifische Lehr-Lern-Ansätze ermöglichen: Ein stark kommunikativ ausgerichteter Sprachunterricht könnte dazu führen, dass die auditiven sprachlichen Teilkompetenzen Sprechen und Hörverstehen ähnlich stark ausgeprägt sind. Dementsprechend könnte ein eher schriftbasierter Unterricht die schriftsprachlichen Subkomponenten parallel fördern. Denn auch die Art der Sprachvermittlung fördert bestimmte Kompetenzen simultan und fokussiert einige stärker als andere.

Je nach theoretischem oder anwendungsbezogenem Hintergrund der Studien werden also entweder die vier Basiskomponenten von Sprachkompetenz oder aber stärker die kommunikativ orientierten Bereiche als Indikatoren für sprachliche Fähigkeiten fokussiert.

Eine eindeutige Definition dessen, welche Teilbereiche Sprachkompetenz letztendlich ausmachen und welche faktorielle Struktur diese bilden, existiert bisher nicht, wie es Bachman und Cohen (1998) formulieren, welche die folgende Frage als weiterhin ungeklärt betrachten: „How many component abilities are there, and what are they?“ (S. 179).

Im Folgenden werden empirische Ergebnisse zum Zusammenhang zwischen bestimmten sprachlichen Teilbereichen ausschnittsweise dargestellt, dabei werden im Wesentlichen die vier Basiskomponenten von Sprachkompetenz – Lesen, Schreiben, Sprechen und Hörverstehen – berücksichtigt. Hier werden sowohl Befunde für die Struktur der Kompetenz in der Erstsprache als auch solche für verschiedene Fremdsprachen zusammengefasst.

### **2.2.1 Schriftsprachliche Kompetenzen – Lesen und Schreiben**

Der Schriftspracherwerb beginnt in der Regel mit der Einschulung im Alter zwischen fünf und sieben Jahren. Lesen und Schreiben können daher als in der Regel institutionell vermittelte, schriftsprachliche Grundfähigkeiten angesehen werden. Die Analyse von Lernbedingungen in diesem Bereich steht im Mittelpunkt vieler empirischer Untersuchungen, wobei Lesen und Schreiben nicht nur auf Grund ihres simultanen Erwerbs oft gemeinsam betrachtet werden. Auch linguistische Theorien gehen davon aus, dass Lesen und Schreiben auf Grund ihrer kommunikativen Zielsetzung rhetorische Gemeinsamkeiten haben, ihnen ähnliche kognitive Prozesse zugrunde liegen und nicht zuletzt auf einen geteilten Wissensinhalt, wie beispielsweise ein gemeinsames Lexikon, zurückgegriffen wird (Coltheart & Funnell, 1987; Fitzgerald & Shanahan, 2000). Dementsprechend lassen sich viele Studien finden, die einen starken positiven Zusammenhang zwischen Lesekompetenz und Schreibkompetenz verzeichnen (Heller, 1999).

Die korrelativen Zusammenhänge zwischen Lesekompetenz und Schreibkompetenz liegen je nach Studie und Testverfahren in mittleren Größenordnungen. So berichtet Parodi (2007) von 51 % gemeinsamer Varianz in Gesamttestwerten für Lesen und Schreiben. Die vorhandenen Korrelationen zwischen Verstehens- und Produktionsleistungen führt Parodi auf gemeinsame Strategien und geteiltes prozedurales Wissen im Lese- und Schreibprozess, in seinem Sinne eine Makrokompetenz, die Lesen und Schreiben beeinflusst, zurück: „In other words, the present findings suggest both a basic general common competence and the possibility of having mode-specific, diversified discourse competence“ (S. 13). Ähnliche Interpretationen legen auch die Studien von Shanahan nahe, die verschiedene theoretische Modelle der

Schreib-Lese-Relationen mit unterschiedlichen Methoden analysierten (Shanahan, 1984; Shanahan & Lomax, 1986). Demnach finden sich in Abhängigkeit von der Differenziertheit der Testinstrumente unterschiedlich starke Partialzusammenhänge: Hohe Korrelationen finden sich zwischen phonetischen Teilkompetenzen und orthografischer Richtigkeit ( $r=.64$ ), geringere hingegen zwischen Leseverstehen und orthografischer Richtigkeit ( $r=.49$ ) (Shanahan, 1984). In mehrfaktoriellen Strukturgleichungsmodellen finden sich durchaus unterschiedliche Zusammenhänge zwischen phonetischer Kompetenz und Rechtschreibung ( $r=.50$ ), Vokabelwissen im Lesen und Schreiben ( $r=.23$ ) und zwischen Textverstehen und produktiver Textstrukturierung ( $r=.14$ ) (Shanahan & Lomax, 1986). Die Tatsache, dass in verschiedenen Studien über unterschiedlich hohe Zusammenhänge zwischen Lesen und Schreiben berichtet wird, könnte mit Unterschieden bei den untersuchten Altersgruppen zusammenhängen (Fitzgerald & Shanahan, 2000). Zu vermuten ist jedoch auch, dass das Testformat einen nicht unwesentlichen Einfluss auf die gefundenen Zusammenhänge hat. So lassen sich starke Korrelationen beobachten, wenn offene Antwortformate sowohl zur Erfassung von Leseverstehen als auch bei der Testproduktion verwendet werden, diese Zusammenhänge verschwinden jedoch, wenn Lesetests mit Multiple-Choice-Itemformat verwendet werden (Jenkins, Johnson & Hileman, 2004). Erste Untersuchungen über die Unterschiede der Zusammenhänge auf Individual- und Klassenebene weisen darauf hin, dass Lesen, Schreiben und auch Wortschatz im Grundschulalter auf Individualebene durchaus divergente individuelle Kompetenzen darstellen, auf Klassenebene jedoch durch einen allgemeinen „Literacy-Faktor“ beschrieben werden können (Mehta, Foorman, Branum-Martin & Taylor, 2005).

### **2.2.2 Rezeptive Kompetenzen – Hörverstehen und Leseverstehen**

Auch für den Bereich der rezeptiven Sprachkompetenzen des Hör- und Leseverstehens zeigen vielfältige empirische Befunde – besonders aus der Neurolinguistik – starke Gemeinsamkeiten auf: In einer Studie an erwachsenen Muttersprachlern stellt Sticht (1968) fest, dass die Fähigkeit des Sprachverstehens sich nicht signifikant zwischen gelesenen und gehörtem Text unterscheidet, unabhängig vom allgemeinen Intelligenzniveau der Personengruppen. In einer experimentellen Studie zur Wahrnehmungsgeschwindigkeit kommt Sinatra (1990) zu dem Schluss, dass Lesen und Hörverstehen auf der Ebene der Wortverarbeitung auf gemeinsame Wahrnehmungsprozesse zurückzuführen und gesehene

und gehörte Wörter Teile eines gemeinsamen Lexikons sind. Auch aktuelle neuropsychologische Studien kommen unter Verwendung bildgebender Verfahren zu der Aussage, dass bei der Verarbeitung gelesener und gesprochener Sprache bestimmte Hirnareale gleichermaßen aktiviert werden (Homae, Haschimoto, Nakajima, Miyashita & Sakai, 2002).

Eine Schlüsselposition nimmt dabei die allgemeine Dekodierfähigkeit ein, die sowohl beim Hörverstehen als auch beim Lesen benötigt wird: Besonders die phonologische Bewusstheit und die Fähigkeit, gesprochene Wörter zu dekodieren, stehen in Zusammenhang mit dem Erkennen und Verstehen geschriebener Sprache (Engen & Høien, 2002). Bei Schülerinnen und Schülern bis zum sechsten Schuljahr kann die allgemeine Dekodierfähigkeit als wichtiger Prädiktor für das Lese- wie das Hörverstehen in der Erstsprache angesehen werden (Carver, 1998). Hoover und Gough (1990) fanden in ihren Untersuchungen Hinweise darauf, dass sich auch bei bilingualen Kindern die Lesekompetenz in beiden Sprachen – in diesem Fall Englisch und Spanisch – durch die beiden Prädiktoren Hörverstehen und Dekodierfähigkeit gut vorhersagen lässt. Die Bedeutung phonologischer Dekodierfähigkeit für das Leseverstehen zeigt sich auch bei Personengruppen mit Defiziten in diesem Bereich – unter Kontrolle von Intelligenz – sodass von einer starken Interdependenz zwischen Hörverstehen, Leseverstehen und phonologischer Dekodierfähigkeit ausgegangen werden kann (Hagtvet, 2003). Personen mit Defiziten im Leseverstehen weisen oft auch ein vermindertes Hörverständnis auf. Förderansätze konnten Transfereffekte von Hörverstehenstrainings auf das Leseverstehen nachweisen (Urban, 1977).

Die Beziehungen zwischen diesen beiden rezeptiven Teilbereichen liegen in vielen empirischen Studien im mittleren Bereich, berichtet wird je nach Altersgruppe und allgemeiner Sprachkompetenz über Koeffizienten zwischen  $r=.45$  und  $r=.75$  (Diakidoy et al., 2005; Rost & Hartmann, 1992). In einer Studie an Erstklässern, die Zusammenhänge verschiedener sprachlicher Teilbereiche mittels Strukturgleichungsmodellierung analysiert, erklärten die Leistungen in der phonologischen Bewusstheit 35 % der Varianz in Leseverstehensaufgaben (Engen & Høien, 2002). Dabei wird der Zusammenhang zwischen Lese- und Hörverstehen mit dem Anstieg der allgemeinen sprachlichen Kompetenz stärker, besonders wenn ein bestimmtes Niveau der allgemeinen Dekodierfähigkeit erreicht ist.

Während in den frühen Grundschuljahren das Hörverstehen noch besser ausgeprägt ist als das Leseverstehen, da hier die Lesefertigkeit erst erworben werden muss, ändern sich die Ausprägungen im Laufe der Entwicklung und des Wissenserwerbs und nähern sich dabei tendenziell an. Jedoch zeigen verschiedene Studien, dass „die Relation ‚Hörverstehen – Leseverstehen‘ [...] von der erreichten Lesefertigkeit bzw. von der Schwierigkeit des Textes“ abhängt (Rost, D. H. & Hartmann, 1992, S. 347). So kann für Kinder im Grundschulalter bei schwierigen Texten zwar das Hörverstehen akkurater sein als das Leseverstehen, bei einfachen Texten jedoch können umgekehrte Effekte auftreten (Royer, Sinatra & Schumer, 1990). Interaktionen der Textsorte zeigen sich für die gleiche Altersgruppe wie folgt: Während das Leseverstehen von erzählenden Texten in der zweiten Jahrgangsstufe noch am höchsten mit dem Leseverstehen von Sachtexten korreliert, wird in höheren Klassen das Hörverstehen von Erzähltexten zum einzigen Prädiktor für das Lesen von beiden Textsorten (Diakidoy et al., 2005).

Diese Erkenntnisse scheinen auch auf die Fremdsprache übertragbar zu sein, allerdings sind die differenziellen Ergebnisse auch abhängig vom Sprachhintergrund: Für eine Gruppe von Englischlernenden mit Japanisch als Erstsprache berichtet Hirai (1999), dass das Leseverstehen bei geringer Gesamtkompetenz leichter fiel als das Hörverstehen und sich die erwarteten Zusammenhänge zwischen beiden Teilbereichen erst bei höherer allgemeiner Fremdsprachenfähigkeit zeigten. Die Bedeutung des Hörverstehens für Kompetenzen im Leseverstehen ist also offensichtlich, wenn auch differenzierte Interaktionen angenommen werden müssen. Im Folgenden wird auf die Zusammenhänge zwischen Hörverstehen und Sprechen eingegangen.

### **2.2.3 Auditive Kompetenzen – Hören und Sprechen**

Hörverstehen als rezeptiver und mündliche Sprachproduktion als produktiver Kompetenzbereich bilden zusammen die *auditiven sprachlichen Kompetenzen* und sind von den schriftsprachlichen Bereichen zu differenzieren. So ist es möglich, kommunikative Kompetenzen im Sprechen und Hörverstehen in einer Sprache zu erwerben, ohne diese lesen oder schreiben zu können, was oft beim nicht-institutionellen Zweitspracherwerb der Fall ist (Rubin, 1994). Die Beziehung zwischen Hörverstehen und Sprechen gründet nicht zuletzt in einer neurobiologischen Voraussetzung: der neuronalen Verbindung zwischen dem Broca-Areal, das die mündliche Sprachproduktion steuert, und dem Wernecke-Areal, in welchem

sich das Wortverstehen abspielt (Golinski & Doetjes, 2005). In den Prozess des Verstehens von gesprochenen Worten sind beide Areale eingebunden. Beim Hörverstehen werden die Signale vom primären auditiven Cortex an das Wernicke-Areal weitergeleitet. Soll ein Wort gesprochen werden, gelangen die Informationen von dort zum Broca-Areal, das es an die motorischen Areale weiterleitet (Nord, 1980). Im Prinzip kann also eine funktionierende Sprachrezeption als unabdingbare Grundlage für die mündliche Sprachproduktion angesehen werden (Cutler, 1987).

Diese Zusammenhänge zeigen sich nicht nur für die Erstsprache, sondern betreffen auch die Fremdsprachenkompetenz. In einer Studie mit englischsprachigen Studenten findet Feyten (1991) für die Fremdsprachen Französisch und Spanisch einen positiven Transfer vom Hörverstehen in der Muttersprache sowohl zum Hörverstehen als auch zur Sprachproduktion in der jeweiligen Fremdsprache. Hörverstehen steht auch insofern im Zusammenhang mit Sprechen, als im Prozess die eigene Sprachproduktion evaluiert und reflektiert werden muss, um Fehler wahrzunehmen oder sogar zu verbessern. Dieses Zusammenspiel funktioniert für die Erstsprache in der Regel automatisch, für die Fremdsprache muss dieser Prozess bewusstgemacht und gezielt eingeübt werden (Van Patten, 2004). Die gehörte Aussprache spielt auch beim Fremdspracherwerb eine Rolle: Auf die akustisch aufgenommenen Informationen über die Phonologie eines Wortes wird bei der Aussprache zurückgegriffen (Flege, 1987). Diese Fähigkeit, über das Hörverstehen neue phonetische Kategorien aufzubauen, variiert dabei zwischen den Sprachen und Personen: Nicht jeder wird eine akzentfreie Aussprache in einer Fremdsprache erreichen. Die Bedeutung bewusster phonologischer Wahrnehmung für die mündliche Sprachproduktion ist auch bei bilingualen Personen zu verzeichnen (Sebastián-Gallés & Kroll, 2003). Sogar in experimentellen Studien konnten Einflüsse des Hörverstehens auf die Sprachproduktion – speziell auf grammatikalische Kompetenzen – nachgewiesen werden. So kann durch gezieltes Training im Hörverstehen eine fremdsprachliche Wissensbasis aufgebaut werden, auf die bei der mündlichen Sprachproduktion zurückgegriffen wird (De Jong, 2005).

#### **2.2.4 Spezifische Kompetenzstrukturen im Englischen als Fremdsprache**

Die in den vorangegangenen Abschnitten dargestellten Befunde zu spezifischen Zusammenhängen zwischen sprachlichen Teilkompetenzen beziehen sich sowohl auf die Kompetenzen in der Erstsprache als auch auf fremdsprachliche Kompetenzen. Speziell für das

Englische als Fremdsprache finden sich empirische Belege dafür, dass sprachliche Teilbereiche systematische Zusammenhänge aufweisen. Diese variieren durchaus in Abhängigkeit von der Art des Fremdsprachenerwerbs: Die Vermittlung einzelner sprachlicher Teilbereiche führt nicht automatisch zu Kompetenzen in anderen Teilbereichen. So führt der Erwerb von Lesekompetenz nicht per se zu einer besseren mündlichen Sprachfähigkeit, wohingegen ein Unterricht, der alle vier Basiskompetenzen gleichermaßen berücksichtigt, durchaus zu einem allgemeinen Sprachfähigkeitsniveau in allen Bereichen führen kann. Beeinflussende Effekte sind dabei Vorwissen, Transfer und sprachliche Redundanzen sowie das Vokabular, auf das beim Lesen, Sprechen, Hören und Schreiben gleichermaßen zurückgegriffen werden muss (Larson, 1983).

Für den Bereich des Englischen bieten die Daten des Test of English as a Foreign Language (TOEFL) eine ideale Grundlage für Analysen und Re-Analysen der Dimensionalität von Sprachkompetenz. Der TOEFL ist ein Placement-Test für Englisch als Fremdsprache und wird seit den 1980er Jahren in stets aktualisierten Formen dazu verwendet, über die Zulassung ausländischer Studierender zu englischsprachigen Universitäten zu entscheiden (Educational Testing Service, 2003; Hamp-Lyons & Kroll, 1997). Die aktuelle Version des TOEFL umfasst drei getrennte Skalen, die folgende Teilbereiche der Sprachkompetenz gesondert erfassen: Hörverständnis, Kenntnisse in Orthographie und Grammatik sowie Leseverständnis und Wortschatz. Durch die getrennte Erfassung dieser drei Teilbereiche lassen sich die Zusammenhänge zwischen diesen Kompetenzfacetten analysieren. In einer Studie von Swinton und Powers (1980) zeigt sich nach exploratorischer Faktorenanalyse entsprechend die dreifaktorielle Struktur, es wird allerdings über Doppelladungen einzelner Testitems berichtet. In konfirmatorischen Faktorenanalysen des TOEFL zeigen sich jedoch hohe gemeinsame Korrelationen der Subtests, sodass je nach Zielsetzung der Diagnostik das Vorgehen gerechtfertigt erscheint, zur Beurteilung der Sprachkompetenz mit dem TOEFL letztendlich immer einen Gesamtestwert zu bilden (Hale, Rock & Jirele, 1989; Jang & Roussos, 2007).

Dementsprechend lassen sich auf der Grundlage dieser Fremdsprachentestbatterie auch Strukturanalysen interpretieren, die je nach untersuchter Population entweder sehr hohe Gesamtkorrelationen einzelner Testteile finden oder spezifische Kompetenzprofile ermitteln (Hosley & Meredith, 1979). Zusammenhänge zwischen Teilkompetenzen finden sich dabei unabhängig von der spezifischen Fremdsprache: Larson (1983) findet hohe Korrelationen

zwischen Lesekompetenz und Grammatik sowie Lesekompetenz und Textproduktion bei Studenten des Französischen als Fremdsprache, höhere Korrelationen hingegen zwischen Lese- und Hörverstehen bei Studenten des Deutschen als Fremdsprache. Weitere strukturelle Unterschiede in der Struktur der Sprachkompetenz im Englischen in Abhängigkeit von der Erstsprache, über die auf der Grundlage der TOEFL-Daten berichtet wurde, sind in den Abschnitten 1.8 und 2.4 dargestellt.

### 2.3 Zusammenhänge zwischen Erst- und Fremdsprachen

Praktisch alle linguistischen Theorien gehen davon aus, dass Wissen in der Erstsprache auch im Prozess des Fremdsprachenerwerbs zum Tragen kommt, dass es also Transfereffekte zwischen den Sprachen gibt. Auf welche Weise dieser Transfer erfolgt, ist jedoch noch nicht endgültig geklärt. Es ist anzunehmen, dass sowohl *übergreifende grundlegende* als auch *spezifische Transfereffekte* existieren.

Zu den übergreifenden Faktoren gehören neben den allgemeinen kognitiven Grundfähigkeiten (Csapó 2003; Csapó & Nikolov, 2002) auch die allgemeine Sprachbewusstheit (Lightbow, 2000; Luchtenberg, 1995) sowie neurologisch-biologische Grundlagen (Hufeisen, 2003; Ullmann, 2001). Zu letzteren zählen beispielsweise grundlegende physiologische Funktionen, wie das akustische Sprachverstehen oder das auf motorischen Grundfähigkeiten basierende Schreiben, die – einmal erlernt – auch in anderen Sprachen praktisch automatisch aktiviert werden. Die allgemeine Sprachbewusstheit tritt in Form von deklarativen Wissenseffekten auf, es wird auf Kenntnisse in der Erstsprache zurückgegriffen, um Zugang zum System der Zweitsprache zu bekommen. Hierzu zählt das Wissen über den Aufbau von Sprache allgemein, beispielsweise über die Funktionen von Grammatik und Pragmatik (Diehl, 2000; Eckerth, 2000; Eckerth & Riemer, 2000). Darüber hinaus spielen beim Transfer übergreifende sprachliche Mechanismen im Sinne der allgemeinen Dekodierfähigkeit (Meschyan & Hernandez, 2002) und der kognitiven Grundfähigkeiten eine wesentliche Rolle (vgl. Abschnitt 1.7). In diese bereits vorhandenen sprachlichen Strukturen werden die neuen Elemente der Fremdsprache wie Wortschatz, Betonung und Aussprache soweit als möglich aufgenommen.

Darüber hinaus wird davon ausgegangen, dass *spezifische Kompetenzen* die im Zusammenhang mit der erstsprachlichen Entwicklung erworben wurden, auf die Fremdsprache übertragen werden, dies gilt insbesondere für die Lesekompetenz (Riemer, 2000). Hier sind direkte Transfereffekte nachzuweisen, und zwar nicht nur, wenn es sich bei Erst- und Fremdsprache um dasselbe Schriftsystem handelt (Carrell, 1991). So konnten Kahn-Horwitz, Joseph und Sparks (2005) aufzeigen, dass die Lesekompetenz im Hebräischen und im Englischen bei den Grundschulern ihrer Studie hohe Zusammenhänge aufwies. Die Autoren führen dies auf gemeinsame neuronale Prozesse zurück, auf die beim Lesen zurückgegriffen wird.

In der vorliegenden Arbeit ist insbesondere zu untersuchen, inwieweit allgemeine und *spezifische* Transfereffekte zwischen *sprachlichen Teilbereichen* vorliegen, hierzu zählen der *sprachreflexiv-rezeptive* und der *sprachproduktive* Kompetenzbereich. Die Annahme eines spezifische Transfers der rezeptiven Sprachkompetenz liegt nahe, da Sprachrezeption sowohl im auditiven wie auch im schriftsprachlichen Bereich in allen sprachlichen Systemen darauf basiert, dass Laute entziffert und aus schriftlichen Elementen zusammengesetzte Worte und Sätze dekodiert werden. In der Erst- wie auch in der Fremdsprache wird beim Rezeptionsprozess auf eine allgemeine Dekodierfähigkeit zurückgegriffen (Sparks et al., 1998). Die allgemeine Sprachbewusstheit rekurriert auf einen sprachlichen Teilbereich, der im Sinne einer grammatikalischen Reflexionsfähigkeit – mentales Lexikon – auch Grundlage der Sprachrezeption in der Fremdsprache ist (Ullman, 2001).

Wesentliche Aspekte der Sprachproduktion basieren in Erst- und Fremdsprache auf vergleichbaren grundlegenden Fertigkeiten: auf dem sprachlichen Ausdruck und der Schreibfähigkeit. Die Bedeutung der Rechtschreibfähigkeit für verständliche schriftsprachliche Kommunikation in jeder Sprache ist offensichtlich (Thomé, 2003). Das Wissen um die und die Berücksichtigung der Korrektheit der Schreibweise von Wörtern kann als eine Kompetenz angesehen werden, die, wenn sie in der Erstsprache vorhanden ist, auf weitere Sprachen übertragen wird. Auch auf die Fähigkeit, zielgerichtete aussagekräftige Texte frei zu formulieren, kann, wenn sie in der Erstsprache erworben wurde, durchaus in allen weiteren Sprachen zurückgegriffen werden (Kellogg, 2003). Denn hierzu gehören viele Bereiche des Sprachwissens, die wenn sie jeweils um einen ausreichenden Wortschatz in der Fremdsprache ergänzt werden, praktisch universell für jede Sprache einsetzbar sind (Selinker & Douglas, 1988).

Weitere Einflüsse ergeben sich aus der Art des Fremdspracherwerbs, der ungeleitet oder institutionell erfolgen kann (Asher, 1972). Werden Unterschiede und Gemeinsamkeiten von verschiedenen Sprachen während des Sprachenlernens bewusstgemacht, lassen sich positive Transfereffekt fördern und negative verringern (Lightbown & Spada, 2000; Timmermann, 2000). Bei jeder zusätzlich zu erwerbenden Fremdsprache wird dann oft auf bereits vorhandene Sprachlernstrategien zurückgegriffen (Wolff, 2001). Dies führt in den meisten Fällen zu einem beschleunigten Erwerb einer zweiten wie auch jeder weiteren Fremdsprache (Hufeisen, 2003).

## **2.4 Gruppenunterschiede in der Dimensionalität von Sprachkompetenz**

Unterschiede in der Struktur von Sprachkompetenz könnten darauf zurückgeführt werden, ob es sich bei der untersuchten Sprache um eine Erst- oder Fremdsprache handelt, und darauf, welchen sprachlichen Hintergrund die betrachteten Personengruppen haben (vgl. Abschnitt 1.7). Über unterschiedliche Strukturen in der Fremdsprache, die je nach Sprachhintergrund der untersuchten Personengruppen variieren, berichten Swinton und Powers (1980) in ihren umfassenden Analysen des TOEFL anhand von Daten von sieben Sprachgruppen. Sie kommen zu dem Schluss, dass die Anzahl der Dimensionen, mit denen sich die fremdsprachliche Kompetenz empirisch beschreiben lässt, sich zwischen Personengruppen mit verschiedenen Erstsprachen durchaus unterscheidet.

Die Kompetenz von Personen mit Spanisch oder Deutsch als Erstsprache ließ sich durch zwei Faktoren beschreiben: Der erste bündelte die Aufgaben zum Wortschatz und Leseverstehen, der zweite solche zur Sprachstrukturierung und zum schriftlichem Ausdruck. In Gruppen mit afrikanischem, arabischem oder asiatischem Sprachhintergrund jedoch bündelte der erste Faktor die Aufgaben zum Leseverstehen, zur Sprachstrukturierung und zum schriftlichen Ausdruck, wohingegen der zweite Faktor ausschließlich Aufgaben zum Wortschatz widerspiegelte. Insgesamt stellten die Autoren fest, dass bei Gruppen mit niedriger Sprachkompetenz hohe Korrelationen zwischen Lesen, Schreiben und Wortschatz bestehen, für Gruppen mit höherer Sprachkompetenz hingegen die Teilbereiche empirisch gut separierbar sind. Sie kommen daher zu dem Schluss, dass die Zusammenhänge zwischen sprachlichen Teilkompetenzen nicht unabhängig vom sprachlichen Hintergrund der Stichprobe betrachtet werden dürfen und Gesamtestwerte nicht in allen Fällen zuverlässige, vergleichbare Aussagen über die Sprachkompetenz verschiedener Sprachgruppen ermöglichen. Oltman, Stricker und Barrows (1988) kommentieren diese Ergebnisse dahingehend, dass die Konstruktvalidität des TOEFL über Sprachgruppen hinweg nicht anzuzweifeln sei, sondern dass die strukturellen Zusammenhänge der Teiltests vom allgemeinen Kompetenzniveau statt von der Erstsprache der zu Testenden beeinflusst seien.

Aktuellere Analysen der Daten des TOEFL kommen zu anderen Ergebnissen: Kunnan (1994) berichtet über stabile, interkorrelierende Dimensionen, die sich zwischen Personen mit indoeuropäischem Sprachhintergrund und solchen mit anderem Sprachhintergrund nicht unterscheiden. Ähnlich interpretieren auch Hale, Rock und Jirele (1989) die Ergebnisse ihrer konfirmatorischen Faktorenanalysen mit TOEFL-Daten, die auf der Grundlage von vier

Sprachtests zwei latente Dimensionen über alle Sprachgruppen hinweg ausmachen, Hörverstehen als ein Faktor und alle anderen Teiltests als weiterer Faktor. Auch Boldt (1988) berichtet für jeden der drei Teilbereiche des TOEFL über eine klare eindimensionale Struktur über Sprachgruppen hinweg, wobei er jedoch nur die Eindimensionalität innerhalb der einzelnen Testteile mit Modellen der Item-Response-Theorie (IRT), nicht jedoch die Korrelationen verschiedener Testbereiche untereinander überprüfte.

Insgesamt scheint es bei der Analyse der Struktur von Sprachkompetenz durchaus sinnvoll zu sein, die Faktorenstruktur innerhalb verschiedener Personengruppen zu vergleichen. Merkmale, die verschiedene Gruppen charakterisieren können, sind zum einen die Erstsprache, zum anderen aber auch Lehr-Lern-Bedingungen, die besonders für die Fremdsprache als Einflussfaktoren auf die Struktur der Kompetenz angesehen werden können.

## 2.5 Dimensionalität als methodisches Artefakt?

Wie bereits mehrfach erwähnt, ist es offensichtlich, dass sich nicht nur die Theorien über die Struktur von Sprachkompetenz hinsichtlich der Anzahl und Zusammenhänge von Faktoren und Dimensionen unterscheiden, auch die empirischen Ergebnisse sind vielfältig und bei Weitem nicht immer einheitlich. Hierfür werden in der Literatur unterschiedliche Gründe als Erklärung angeführt, aus Sicht der Diagnostik sind besonders Effekte der Analysemethoden und der Testinstrumente interessant.

Verschiedene Analysemethoden stellen eine Vergleichbarkeit der Ergebnisse durchaus in Frage. So berichten diejenigen Studien, die ausschließlich explorative Faktorenanalysen anwenden, in einer überwiegenden Anzahl der Fälle von einer oder nur wenigen Dimensionen von Sprachkompetenz (North, 2000). Jene Studien hingegen, die konfirmatorische Faktorenanalysen oder Strukturgleichungsmodelle verwenden, finden meist eine Bestätigung für die komplexeren, mehrdimensionalen Modelle. So erklären Stricker, Rock und Lee (2005) die in Abschnitt 2.4 dargestellten Ergebnisse zu Gruppenunterschieden damit, dass Hale und andere (1989) nur explorative Faktorenanalysen verwendet hatten, Swinton & Powers (1980) hingegen konfirmatorische Analysen. Auch Re-Analysen der Ergebnisse von Oller (1976), die eben nicht mehr nur explorative Verfahren verwenden, berichten durchweg über eine differenzierte, mehrdimensionale Struktur von Sprachkompetenz. In letzter Zeit werden traditionelle Modelle der Strukturanalysen ergänzt durch differenziertere Mehrgruppenmodelle oder Mehrebenenanalysen, welche Effekte des Sprachhintergrunds oder der Lerngruppenzusammensetzung berücksichtigen können (vgl. Abschnitt 1.6).

Auch die Methoden der Kompetenzerhebung, besonders die Testformate, scheinen die Ergebnisse entscheidend zu beeinflussen. So lässt sich feststellen, dass Studien eher über einen Generalfaktor berichten, wenn die Analysen auf Testverfahren beruhen, die hauptsächlich rezeptive Kompetenzen fokussieren und stärker geschlossene Antwortformate verwenden. Hierzu gehören in erster Linie Testformate mit Richtig-Falsch-Antworten, die mit Vorliebe sowohl bei der Erfassung des Leseverstehens als auch des Hörverstehens oder des grammatikalischen Wissen eingesetzt werden (Gräfe-Bentzien, 2000). Der Einsatz von Testformaten mit schriftbasiertem, aber offenem Antwortmodus führt dementsprechend meist zu stärker differenzierten Modellen. Besonders bei solchen Studien, die mündliche Sprachkompetenz erfassen, zeigen sich klare mehrdimensionale Strukturen von

Sprachkompetenz. Bei der Interpretation von hohen Korrelationen zwischen Teilkompetenzen sollte also auch in Betracht gezogen werden, dass ein nicht unerheblicher Teil diese Zusammenhänge auf die Art der Kompetenzerfassung zurückzuführen sein könnte (Grotjahn, 2000b). So könnte zu kritisieren sein, dass viele Testansätze komplizierter sind als das Konstrukt, welches sie erfassen sollen, und dass die theoretische Differenzierung und die testmethodische Umsetzung von sprachlichen Konstrukten oft wenig Übereinstimmung zeigen (Bachman, 1990b).

## 2.6 Zusammenfassung

In diesem Kapitel wurde eine Auswahl von Studien dargestellt, die der Frage nach der Struktur und Dimensionalität von Sprachkompetenz empirisch nachgegangen sind. Deutlich wird, wie bereits im vorangegangenen Kapitel zu den theoretischen Modellen, dass auch die vorliegenden empirischen Ergebnisse vielfältig, uneinheitlich und vor allem widersprüchlich sind. Insgesamt lässt sich feststellen, dass sich Sprachkompetenz durchaus empirisch in unterschiedliche Teilbereiche differenzieren lässt. Wie viele Teilbereiche jedoch gefunden werden, hängt dabei zum einen vom Fokus der Forschungsrichtung und der Differenziertheit des theoretischen Modells von Sprachkompetenz ab, zum anderen auch von der Fragestellung der jeweiligen empirischen Studie und ihrer methodischen Umsetzung der diagnostischen Erfassung dieser Teilbereiche.

Weiterhin lassen sich spezifische Zusammenhänge zwischen diesen Teilbereichen auffinden, die auf übergeordnete latente Konstrukte hinweisen. Die Natur und Gestalt dieser latenten Konstrukte verbleiben dabei jedoch unklar: Es kann sich um grundlegende kognitive Prozesse, geteilte Wissensinhalte oder übergreifende neuronale Grundlagen handeln. Darüber hinaus wird dem natürlichen und dem geleiteten Erwerbskontext eine nicht unwesentliche Rolle bei der parallel ablaufenden Entwicklung von spezifischen sprachlichen Teilfähigkeiten zugeschrieben. Dies betrifft das Hörverstehen und das Sprechen als aufeinander abfolgende, natürliche Schritte der Sprachentwicklung und das Lesen und Schreiben als parallel institutionell vermittelte, schriftsprachliche Kompetenzen. Korrelationen zwischen Teilbereichen der Sprachkompetenz können allerdings auch auf Ähnlichkeiten in den verwendeten Testverfahren zurückgeführt werden. Hohe empirische Zusammenhänge zwischen sprachlichen Teilbereichen werden besonders dann berichtet, wenn geschlossene Antwortformate zum Einsatz kommen, während diagnostische Instrumente mit alternativen Testformaten durchaus zu differenzierten Ergebnissen kommen.

Je nachdem, ob die Erst- oder eine Fremdsprache im Mittelpunkt der Analyse der Struktur von Sprachkompetenz steht, werden unterschiedliche strukturelle Zusammenhänge zwischen sprachlichen Teilkompetenzen berichtet. Und die Struktur von Sprachkompetenz in einer bestimmten Fremdsprache unterscheidet sich darüber hinaus in Abhängigkeit davon, welche Erstsprache die Sprachlernenden aufweisen.

Erwähnt werden muss in diesem Zusammenhang auch die Bedeutung der Auswertungsmethoden, die zumindest konfirmatorischen Charakter haben sollten, um eine Prüfung unterschiedlich stark differenzierter Modelle von Sprachkompetenz zu ermöglichen. Das hier dargestellte, eher breite Spektrum der empirischen Befunde über die Struktur von Sprachkompetenz wird im folgenden Kapitel auf die Erfassung von Sprachkompetenz im pädagogisch-psychologischen Kontext eingengt, indem speziell die Fragen der mehrdimensionalen Erfassung von Sprachkompetenz behandelt werden.

### **3 Die Erfassung von Sprachkompetenz im psychologisch-pädagogischen Kontext**

Die Diskussion um die Erfassung und Bewertung von Schülerleistungen ist seit der Veröffentlichung der Ergebnisse von PISA zumindest in Deutschland in ein neues Zeitalter eingetreten. Die ersten empirischen Wellen der Leistungsdiagnostik in den 1970er Jahren wurden durch eine Vielzahl von Verfahren in allen Schulleistungsbereichen und für alle Jahrgangsstufen unterstützt (Borchert, Knopf-Jerichow & Dahbashi, 1991; Brickenkamp, 1997; Ingenkamp, 1985). Im Gegensatz zur kontinuierlichen Zunahme der Testentwicklung bis 1990 gingen Verkauf und Anwendung der Testverfahren jedoch seit Mitte der 1980er Jahre wieder stark zurück: 1990 machte der Testumsatz gerade noch 10 % des Umsatzes von 1974 aus (Schöler, 2001). Ursache dafür war nicht zuletzt die Kritik auf Seiten derjenigen Vertreter in Pädagogik und Fachdidaktik, die sich mit Argumenten des „nicht zu vermessenden Menschen“ und der Gefahr der unfairen Selektion gegen eine standardisierte Erfassung von Leistung verwahrten. Im deutschen Sprachraum ist aktuell und nicht zuletzt nach den Ergebnissen der ersten PISA-Studie im Jahr 2000 wieder ein verstärktes Interesse an standardisierten Leistungsstudien im schulischen Kontext zu verzeichnen. Diese werden inzwischen regional, national und international vergleichend durchgeführt (Helmke & Schrader, 2001).

Bezüglich der Kompetenzdiagnostik im schulischen Bereich gilt es zu unterscheiden zwischen Tests, die mit dem Ziel der Diagnose von *individuellen Leistungen* angewendet werden, und solchen Testinstrumenten, die der *Evaluation des Bildungssystems* dienen. Letztere werden als nationales Monitoring in Form regelmäßiger, meist national repräsentativer Studien durchgeführt, die bestimmte inhaltliche Bereiche umfassen. Im Kontext der vorliegenden Arbeit sind Modelle und Testverfahren von Interesse, die *Sprachkompetenz* von Schülerinnen und Schülern erfassen. Dabei interessieren besonders diejenigen Ansätze, die Sprache nicht als holistisches, sondern als mehrfaktorielles Konstrukt ansehen und dementsprechend differenzierte Messinstrumente einsetzen. Solche Verfahren zur Erfassung von Sprachkompetenz verfolgen im psychologisch-pädagogischen Kontext unterschiedliche Zielsetzungen. Dazu gehören:

- die Differenzialdiagnostik von *Sprachentwicklung und Sprachauffälligkeiten* unter Einsatz von Testverfahren, die auf spezifische Problembereiche eingehen und die Erstellung individuelle Profile ermöglichen,
- die individualisierte Prüfung *schulischer Sprachleistungen* im Erwerbsprozess zum Zweck der Intervention im Sinne eines formativen Assessments sowie
- die stark bildungspolitisch gesteuerte, summative Erfassung von Sprachleistungen als *Erreichung bestimmter Standards*, um repräsentative Aussagen über ein Bildungssystem machen zu können.

Die vorliegende Arbeit bewegt sich im Kontext der letztgenannten Vorgehensweisen, welche die sprachlichen Kompetenzen als Resultate schulischen Unterrichts erfassen und beurteilen sollen. Unbestritten ist, dass diagnostische Ansätze jeweils nur Ausschnitte aus einer angenommenen Ganzheit von Kompetenz betrachten können, also auf Grund einer bestimmten Fragestellung auf spezielle Kompetenzbereiche fokussieren. Um den Bereich der sprachlichen Kompetenzen für die unterschiedlichen Zielsetzungen psychologisch-pädagogischer Diagnostik möglichst repräsentativ abbilden zu können, ist daher zu entscheiden, welche und wie viele Facetten von Sprachkompetenz erfasst werden müssen, um sinnvolle Aussagen über die Kompetenzen von einzelnen Schülerinnen und Schülern, von ganzen Schulklassen oder Schulen sowie nicht zuletzt von einzelnen Bildungssystemen insgesamt treffen zu können.

In den vorangegangenen Kapiteln wurde darauf eingegangen, welche theoretischen Annahmen über die Struktur von Sprachkompetenz existieren und welche empirisch fundierten Erkenntnisse darüber vorliegen, in welche Teilkompetenzen sich diese ausdifferenzieren lässt. Auch wurden die Zusammenhänge der Teilkompetenzen untereinander dargestellt und die Struktur von Sprachkompetenz diskutiert. In diesem Kapitel nun wird differenzierter auf die Frage eingegangen, wie sprachliche Kompetenzen im Kontext von Schule und Unterricht erfasst werden und welche Schlussfolgerungen sich aus den Ergebnissen für die vorliegende Arbeit ziehen lassen.

### 3.1 Sprachbezogene Schulleistungstests

Die meisten Schulleistungstests zur individuellen Diagnostik sprachlicher Kompetenzen können als standardisierte Papier-und-Bleistift-Verfahren im Rahmen einer Schulstunde als Gruppentests durchgeführt werden (Lukesch, 1994). Ihre messtheoretischen Grundlagen basieren in der Regel auf Konzepten der klassischen Testtheorie, dazu gehören oft Normierungen für Bezugspopulationen von Schultypen, Jahrgangsstufen und Altersgruppen (vgl. zu Testtheorien Moosbrugger & Kelava, 2007). Die Unschärfen dieser Verfahren liegen in der Regel bei der Bestimmung der Gültigkeit bzw. der Zuordnung der Tests zu theoretischen Modellen von Sprachkompetenz, da oft nur die logische Validität deklariert wird und sich Leistungstests meist an den Curricula als Zielstandards orientieren, weniger hingegen an Kompetenzmodellen des Spracherwerbs.

Solche Sprachtests beziehen differenzierte Aspekte der Sprachkompetenz mit ein. Beispiele hierfür sind der Allgemeine Schulleistungstest für dritte Klassen (AST 3) (Fippinger, 1991), der die sprachlichen Bereiche Satzverständnis und Rechtschreibung beinhaltet, sowie der Allgemeine deutsche Sprachtest (ADST) (Steinert, 1978), der eine linguistisch-didaktische Analyse des Sprachleistungsstandes beabsichtigt. Der ADST unterscheidet die vier Basiskompetenzen Hören, Lesen, Sprechen und Schreiben und kombiniert diese mit den sechs Sprachebenen Textematik, Lexematik, Morphematik, Syntagmatik, Phonematik und Prosodie, sodass sich insgesamt 24 Unterskalen ergeben. Konzipiert als Gruppentest für die dritte bis zehnte Klasse des Gymnasiums, sind Normen für jede Jahrgangsstufe und Schulart verfügbar. Bei einer Faktorenanalyse über Daten des ADST lassen sich zwei Faktoren extrahieren, die insgesamt etwa 62 % der Gesamtvarianz erklären, wobei der erste Faktor allein etwa 56 % auf sich vereinigt. Er umfasst Aspekte der „auditiven Perzeption sowie das Differenzieren semantischer und grammatikalischer Aspekte auf der Wortebene“, während der zweite Faktor Aufgaben zur „Diskrimination abstraktester Sprachsegmente sowie begrifflichen Abstraktion“ auf sich vereint (Messelken, 1971). Deutlich werden in dieser Beschreibung die linguistischen Termini, die auf eine entsprechende theoretische Fundierung dieses Testinstruments schließen lassen.

Oft werden im schulischen Bereich auch sogenannte Screening-Verfahren zur Erfassung individueller Leistungen eingesetzt, die einen Gesamtindikator für Sprachkompetenz erfassen. Hierzu zählt der Schulleistungstest Deutsch für vierte Klassen CT-D4 (Ratz & Klein-Braley,

1992). Er geht auf das Konzept des *cloze procedure* zurück, das vor allem in der Fremdsprachdiagnostik eingesetzt wird und Sprachkompetenz ganzheitlich erfassen soll (Oller, 1973; Raatz & Klein-Braley, 1983). Dabei werden in einem Text zufällig oder systematisch eine bestimmte Anzahl von Wörtern oder Worthälften gestrichen. Aufgabe ist es, diese Lücken zu füllen, wozu Wortschatzwissen, Rechtschreibkenntnisse und semantisches Inhaltsverstehen vorhanden sein müssen. Sind die Texte ausreichend lang, so wird angenommen, dass die Textlücken alle Teilbereiche der schriftbasierten sprachlichen Leistung repräsentieren und daher valide für die Erfassung der jeweiligen Sprachkompetenz sind (Stenzel, 1989). Einzelne Teilbereiche wie Wortschatz, Syntax oder Morphologie können dabei allerdings nicht separat bewertet werden, da die Sprachleistung holistisch erfasst wird. Obwohl also keine Teilbereiche der schriftsprachlichen Leistung differenziert werden, wird mit dem CT-D4 der Anspruch erhoben, die allgemeine schriftliche Beherrschung der deutschen Sprache erfassen und prüfen zu können.

Offensichtlich wird, dass nur wenige Verfahren existieren, die unterschiedliche sprachliche Bereiche abdecken. Vielmehr werden meist ein oder zwei Kompetenzen erfasst, oft jene, in denen sprachliche Defizite vermutet werden. Neben Tests zur Erfassung von Lese- und Rechtschreibleistungen liegen solche zur Prüfung phonologisch-phonetischer Leistungen, des Wortschatzes und der Grammatik vor. Neue kommunikative Ansätze in der Leistungserhebung, die stärker offene und mündliche Testformate verwenden, unterstützen die facettenreiche Messung von Sprachkompetenz ebenso wie statistische Analysemethoden, welche die Überprüfung solcher mehrdimensionaler Modelle ermöglichen (Kuhlemeier, van den Bergh & Rijlaarsdam, 2002; McNamara, 1995). So können methodische Artefakte reduziert und differenzierte Strukturen von Sprachkompetenz konfirmatorisch überprüft werden (Gräfe-Bentzien, 2000).

### **3.2 Die Erfassung von Sprachkompetenz in Large-Scale-Assessments in Deutschland**

Die Erfassung sprachlicher Kompetenzen in sogenannten *Large-Scale-Assessments*, also Studien mit einer großen Stichprobe und dem Anspruch der Repräsentativität oft für ein ganzes Bildungssystem, weicht von den Zielsetzungen der individualdiagnostisch ausgerichteten Schulleistungstests ab: Sie unterscheiden sich in der Art der Erhebungsinstrumente und der Art von Aussagen, die aus den Daten abgeleitet werden können. Im Rahmen von derartigen Schulleistungstudien erfolgen eine Auswahl und oft auch eine Einschränkung der zu erfassenden sprachlichen Kompetenzbereiche, die mit angemessenen, theoretisch fundierten, diagnostischen Verfahren erfasst werden. Eine wesentliche Rolle bei der Auswahl von Testinhalten spielen definierte Lernziele in Form von Curricula oder von Bildungsstandards, über die Schülerinnen und Schüler zu einem bestimmten Zeitpunkt ihres Bildungslebenslaufes verfügen sollen (Ackeren, 2005; Brindley, 2001; Cascallar & Cascallar, 2003; Ständige Konferenz der Kulturminister der Länder in der Bundesrepublik Deutschland, 2003).

Die wohl bekanntesten und letztendlich auch wegweisenden repräsentativen Studien zur Erfassung von sprachlichen Schülerleistungen in Deutschland sind PISA (Deutsches PISA-Konsortium, 2001) und die Internationale Grundschul-Lese-Untersuchung (IGLU) (Bos, Lankes, Schwippert et al., 2003). Beide Studien sind als internationale Vergleichsstudien konzipiert und erfassen die Leseleistung als Indikator für sprachliche Kompetenz. Die Zielgruppe von PISA sind die 15-jährigen Schüler und Schülerinnen, in IGLU wird die Lesekompetenz in der dritten Jahrgangsstufe erhoben. Beiden Studien liegt das theoretische Konzept einer alltagsnahen, angewandten Sprachkompetenz zugrunde, sie orientieren sich nicht an linguistischen Modellen von Sprachkompetenz oder nationalen Lehrplänen. Vielmehr sollen die Tests jene sprachlichen Kompetenzen erfassen, die für Schülerinnen und Schüler der jeweiligen Altersgruppe in aller Welt die Voraussetzung für eine Partizipation in der Gesellschaft sowie für einen weiteren Erwerb institutioneller Bildung darstellen. Die Lesekompetenz kann dafür als ein wesentlicher und grundlegender Bedingungsfaktor angesehen werden.

Aktuell lassen sich nur wenige nationale Large-Scale-Studien auffinden, die sprachliche Leistungen deutscher Schülerinnen und Schüler differenziert und repräsentativ erheben. Dazu

gehört die Hamburger Studie Aspekte der Lernausgangslage und Lernentwicklung (LAU) (Lehmann, Peek, Gänsfuß & Husfeldt, 2001). LAU beabsichtigt eine längsschnittliche Analyse von Lernbedingungen und Lernentwicklungen in Hamburger Schulen von Beginn der Sekundarstufe I an und erfasst Lernstände am Ende der Jahrgangsstufe fünf, sieben, neun und elf mittels standardisierter Schulleistungstests. Differenziert untersucht werden die sprachlichen Teilbereiche Leseverstehen, Rechtschreibung, Textproduktion und Sprachverständnis sowie die Fremdsprachenleistungen in Englisch, Französisch und Latein. Wesentliche aktuelle Ergebnisse in Bezug auf die Weiterentwicklung sprachlicher Fähigkeiten sind der positive Effekt leistungsheterogener Lerngruppen und der tendenziell negative Zusammenhang zwischen Migrationshintergrund und Leistungen in Deutsch und der Fremdsprache Englisch.

Auch für den Bereich der Fremdsprachenkompetenz kann in Deutschland nur auf wenige repräsentative Studien verwiesen werden. Denn obwohl diagnostische Sprachtests wie der TOEFL oder der Test Deutsch als Fremdsprache (TestDaF) Sprachkompetenz in verschiedenen Facetten erfassen und inzwischen standardisiert zur Diagnostik, Selektion und Beurteilung von Studienanfängern eingesetzt werden (Arras & Grotjahn, 2002; Hamp-Lyons & Kroll, 1997), beruhen diese Verfahren nicht auf repräsentativen Stichproben oder erfassen Fremdsprachenkompetenz ebenfalls nur spezialisiert auf bestimmte Teilbereiche (Bos & Schwippert, 2002). Als internationale, vergleichende Studien zur Sprachkompetenz im Englischen mit deutscher Beteiligung ist die bereits in den 1970er Jahren durchgeführte Study of English as a Foreign Language der International Association for the Evaluation of Educational Achievement (IEA) zu nennen, in der Lese- und Hörverstehen sowie Schreiben im Englischen bei Schülerinnen und Schülern der neunten und der 13. Jahrgangsstufe erfasst wurden. Als relevante Bedingungsfaktoren für die (fremd-)sprachlichen Fähigkeiten zeigten sich die Erstsprache, die allgemeinen kognitiven Fähigkeiten, der sozioökonomische Hintergrund und das Geschlecht der Schülerinnen und Schüler (Elley, 1994; Purves, 1992; Walker, 1976). Für die internationale Gesamtstichprobe zeigten sich starke positive Zusammenhänge zwischen den beiden Bereichen der Fremdsprachenkompetenz und des Wortschatzes in der Erstsprache (Lewis & Massad, 1975).

Zu den neueren Ansätzen der Erfassung von Sprachkompetenz in Deutschland zählen die kürzlich eingeführten zentralen Vergleichsarbeiten der Länder sowie die Bildungsstandards, über die jedoch noch keine empirischen Ergebnisse vorliegen (Helmke & Hosenfeld, 2004;

Hovestadt & Keßler, 2005). In Zukunft werden sprachliche Kompetenzen in der Erst- und Fremdsprache auch im Rahmen der nationalen Überprüfung der Bildungsstandards in verschiedenen Jahrgangsstufen erfasst werden, entsprechende Instrumente werden zurzeit entwickelt (Köller, 2007, 2008).

Die vorliegende Arbeit verwendet Daten der nationalen Schulleistungsstudie Deutsch Englisch Schülerleistungen International (DESI), die als ein in Deutschland bisher wohl einmaliges Projekt sprachliche Kompetenzen von Schülerinnen und Schülern der neunten Jahrgangsstufe facettenreich – nämlich unter Berücksichtigung von zwölf sprachlichen Teilbereichen in Deutsch und Englisch – erhoben hat (Klieme, 2006). Die Fragestellungen der vorliegenden Arbeit sind im Kontext der DESI-Studie entstanden und beruhen auf Daten, die im Rahmen dieses Projektes erhoben wurden. Das folgende Kapitel erläutert Zielsetzung und Inhalt sowie Testverfahren, Stichprobe und Datengrundlage von DESI.

### **3.3 Die Erfassung von Sprachkompetenzen in der Studie Deutsch Englisch Schülerleistungen International**

Die Studie Deutsch Englisch Schülerleistungen International untersuchte im Schuljahr 2003/2004 die sprachlichen Leistungen von Schülerinnen und Schülern der neunten Jahrgangsstufe in Deutschland. Sie wurde von der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (KMK) im Jahre 2000 in Auftrag gegeben, um das nationale Bildungsmonitoring, das bis dahin nur durch auf internationalen Vergleich ausgelegte Large-Scale-Studien wie PISA oder IGLU abgedeckt wurde, um Aussagen zu sprachlichen Kompetenzen von Schülerinnen und Schülern zu ergänzen. Aufgrund ihrer bundesweit repräsentativen Stichprobe, des komplexen Designs und der facettenreichen Erhebung von sprachlichen Kompetenzen und unterrichtlichen Kontextfaktoren ermöglicht die DESI-Studie differenzierte Aussagen über Lehr-Lern-Prozesse und den Erwerb sprachlicher Kompetenzen, die für Unterrichtspraxis und Bildungspolitik gleichermaßen wichtig sind. DESI wurde in den Jahren 2001 bis 2006 von einem interdisziplinären Konsortium durchgeführt, dem unter Federführung des Deutschen Instituts für Internationale Pädagogische Forschung (DIPF) sowohl Bildungsforscher als auch Fachdidaktiker für Deutsch und Englisch angehörten.

Durch DESI sollten differenzierte Aussagen über die Fähigkeiten von Schülerinnen und Schülern der neunten Jahrgangsstufe im Deutschen und im Englischen ermöglicht werden. Die Fokussierung auf diese beiden Sprachen ist begründet in der Rolle des Deutschen als Erst- oder zumindest als Verkehrssprache der Schülerinnen und Schüler in Deutschland und des Englischen als „Lingua franca“ in der westlichen Welt (Beck & Klieme, 2003). DESI hatte darüber hinaus das Ziel, neben der Erfassung der sprachlichen Kompetenzen auch analytisches Wissen über die Wirksamkeit von Schule und Unterricht zu gewinnen, um daraus Erklärungsansätze für unterschiedliche Leistungsniveaus sowie Grundlagenwissen für bildungs- und schulpolitische Aktivitäten, die Revision von Curricula, Lehrtexten und Unterrichtsmaterialien, die Lehreraus- und -fortbildung und vor allem für die Unterrichtsgestaltung ableiten zu können. Die geprüften sprachlichen Leistungen wurden vor dem Hintergrund individueller, schulischer und familiärer Faktoren analysiert, um Optimierungsansätze aufzeigen zu können.

Um im Sinne eines solchen Rahmenmodells gezielt analytisches Wissen über die Wirksamkeit von Schule und Unterricht zu gewinnen, war ein Längsschnittdesign erforderlich, das es ermöglichte, Leistungsveränderungen valide zu erfassen. In DESI wurde dementsprechend eine kleine Auswahl der Leistungstests und Fragebögen zu zwei Messzeitpunkten, nämlich am Anfang und Ende der neunten Jahrgangsstufe, eingesetzt. Die erhobenen Daten können dazu herangezogen werden, Erklärungsansätze für unterschiedliche Leistungsniveaus abzuleiten und die Unterrichtswirksamkeit im Verlauf eines Schuljahres zu analysieren. Die Ergebnisse bezüglich der Leistungsversänderung im Laufe des neunten Schuljahres sind inzwischen vom DESI-Konsortium (2008) veröffentlicht worden.

Entsprechend der aktuellen Forschungslage zur Dimensionalität von erst- und fremdsprachlichen Kompetenzen wurden in den DESI-Tests jeweils mehrere Kompetenzbereiche unterschieden: Sowohl im Deutschen als auch im Englischen wurden die Sprachdomänen Lesen und Schreiben abgedeckt, während die Bereiche Hören und Sprechen aus erhebungstechnischen Gründen nur im Englischen berücksichtigt werden konnten. In beiden Sprachen wurden zudem sprachpragmatische und grammatikalische Aspekte der Sprachbewusstheit erfasst. Eine zusätzliche Betonung erfuhr der kommunikative Aspekt sprachlichen Handelns dadurch, dass ein schriftlicher Test zur Argumentation im Deutschen eingesetzt wurde. Auch lexikalische und orthographische Aspekte von Sprachkompetenz wurden im Deutschen in gesonderten Tests abgebildet und im Englischen als Komponenten von Lückentexten (C-Tests) erfasst. Zusätzlich zu den insgesamt zwölf Testverfahren ergänzten Fragebögen zu Lehr-Lern-Bedingungen, Hintergrundfaktoren und motivationalen Aspekten sowohl für Schülerinnen und Schüler als auch für Lehrpersonen, Fachkollegien, Schulleitung und Eltern die Erhebung. Die einzelnen Testverfahren sind in Abschnitt 3.3.3 differenziert dargestellt.

### **3.3.1 Stichprobe**

Die Zielpopulation der DESI-Studie stellten alle Schülerinnen und Schüler in Deutschland dar, die sich zum Zeitpunkt der Erhebung in der neunten Jahrgangsstufe an allgemeinbildenden Schulen befanden. Hierzu gehören in Deutschland die Schulformen Hauptschule, Realschule, Integrierte Gesamtschule, Schulen mit mehreren Bildungsgängen und Gymnasium. Förderschüler wurden auf Grund der inhaltlichen Ausrichtung der Studie nicht berücksichtigt. Um die Repräsentativität der erhobenen Daten nicht nur für die

Gesamtheit der Schülerinnen und Schüler in Deutschland, sondern auch für die unterschiedlichen Bildungsgänge gewährleisten zu können, erfolgte die Stichprobenziehung nach einem stratifizierten Verfahren. Strata stellten dabei die Bundesländer und die Bildungsgänge in den Bundesländern dar. Schulen aller Bildungsgänge wurden proportional zu ihrem Anteil in den Bundesländern gezogen. Als drittes Stratum wurden in den gezogenen Schulen jeweils zufällig zwei komplette neunte Klassen in die Stichprobe aufgenommen (Beck, Bundt & Gomolka, 2008). Durch diese Schichtung der Stichprobenziehung und die damit verbundenen Stichprobengewichte wurde eine Stichprobe erreicht, in der die getesteten Schülerinnen und Schüler repräsentativ für alle Schülerinnen und Schüler der neunten Klassen in Deutschland waren.

Eine Sonderfragestellung stellten in DESI die Leistungen und Unterrichtsbedingungen der Schülerinnen und Schüler in bilingualen Unterrichtsangeboten dar. Um eine Mindeststichprobengröße für diese Personengruppe zu erzielen, gingen Schulen mit bilingual unterrichteten Schülern zusätzlich überproportional häufig in die Stichprobe ein. Dabei wurde darauf geachtet, dass sich die Anzahl der auszuwählenden Schulen proportional zum Anteil der Schulen mit bilingualem Unterrichtsangebot im jeweiligen Bundesland verhielt. Generell gibt es bilingualen Unterricht in Deutschland in erster Linie an Gymnasien, sodass vor allem diese Schulform in der DESI-Stichprobe der bilingualen Schulen vertreten ist.

Die Stratifizierung als Voraussetzung für eine repräsentative Stichprobenziehung wurde ergänzt durch das Stichprobengewicht für jede getestete Person (Foy, 1999). Dieses Gewicht legt fest, wie viele Personen in der Population durch eine getestete Person repräsentiert werden sollen. Durch die Gewichtung wird ausgeglichen, dass nicht alle Schülerinnen und Schüler mit der gleichen Wahrscheinlichkeit in die Stichprobe gelangen (Johnson, 2004). Um einzelne Ziehungsgruppen bei den Analysen nicht über- oder unterrepräsentiert eingehen zu lassen, erhalten sie ein entsprechendes Gewicht, das sowohl die verschiedenen Ebenen der Stichprobenziehung (Schulen, Klassen sowie Schüler und Schülerinnen) als auch die unterschiedlichen Beteiligungsraten auf diesen Ebenen berücksichtigt. Die Güte der bei DESI vorgenommenen Gewichtung zeigt sich daran, dass durch die Gewichte die Populationsverhältnisse entsprechend internationalen Standards sehr gut abgebildet wurden.

An der Haupterhebung von DESI nahmen in der Zeit von September 2003 bis Juni 2004 Schülerinnen und Schüler aus 219 Schulen teil, darunter 40 Schulen mit bilingualem

Sachfachunterricht. Von insgesamt 10 639 gezogenen Schülerinnen und Schülern zum ersten Messzeitpunkt, d. h. zu Beginn des neunten Schuljahres, beteiligten sich 94,4 % an den Leistungstests und 91,5 % an der Beantwortung der Fragebögen. Von 10 632 Schülerinnen und Schülern zum zweiten Messzeitpunkt, d. h. am Ende des neunten Schuljahres, beteiligten sich 94,6 % an den Leistungstests und 92,1 % an der Beantwortung der Fragebögen. Diese Schülerinnen und Schüler waren repräsentativ für die geschätzte Population von 916 430 Schülerinnen und Schülern der neunten Jahrgangsstufe im Schuljahr 2003/2004 in Deutschland (Statistisches Bundesamt, 2004).

### **3.3.2 Curriculare Verankerung der Testverfahren**

Für Schulleistungsstudien ist die Frage der inhaltlichen Gültigkeit, das heißt, inwieweit die verwendeten Tests die in den Curricula festgehaltene Lernziele und Lerninhalte repräsentieren, von zentraler Bedeutung. Denn „die inhaltliche Validität eines Tests hängt nicht zuletzt davon ab, inwieweit es gelingt, die zu erfassende latente Kompetenz als theoretisches Konstrukt vorab zu definieren oder zumindest post hoc zu rekonstruieren“ (Baumert, Klieme, Lehrke & Savelsbergh, 2000, S. 108). Die Testentwicklung für die DESI-Studie verfolgte daher das Ziel, nicht nur den aktuellen Forschungsstand der Linguistik und Fachdidaktik im Englischen wie im Deutschen zu berücksichtigen, sondern wies, den bildungspolitischen Fragestellungen entsprechend, in der Konstruktion von Testaufgaben einen engen Bezug zum deutschen Schulsystem, insbesondere den Lehrplänen und Lehrwerken der deutschen Länder für die Fächer Deutsch und Englisch in der neunten Jahrgangsstufe auf. DESI erfasste einen breiten Bereich sprachlicher Fähigkeiten, die von Schülerinnen und Schülern im neunten Schuljahr erwartet werden können (DESI-Konsortium, 2003). Darüber hinaus wurden die DESI-Tests zur Englischkompetenz an den Gemeinsamen Europäischen Referenzrahmen für Sprachen (GER) (Council of Europe, 2001) angebunden. Ein Großteil der Testskalen wurde mit den Niveaus des GER abgeglichen (Takala, 2007), Testaufgaben aus DESI werden darüber hinaus in internationalen Folgeprojekten verwendet, die der Implementation des GER in die Praxis dienen (Jurecka, 2007).

Dieses Vorgehen erforderte konzeptuelle und testmethodische Innovationen sowie die Neuentwicklung spezifischer Testkomponenten. Bei der Bestimmung der Güte der DESI-Testinstrumente stand ihre Inhalts- und Konstruktvalidität im Vordergrund. Um diesen inhaltlichen Bezug nachzuweisen, wurden im Rahmen von DESI umfangreiche

Curriculumanalysen für die Fächer Deutsch und Englisch durchgeführt, sodass die Testverfahren als curricular valide angesehen werden können (Dubberke & Harks, 2008). Gezeigt werden konnte auch die konvergente Validität mit sprachlichen Testverfahren von PISA 2000 und IGLU 2001, mit dem Rechtschreibtest RST (Bulheller & Häcker, 2001) und dem ADST für das Deutsche (Steinert, 1978), sowie weiterhin mit Unterskalen aus LAU (Lehmann et al., 2001) und dem EU-Test für das Englische (Bonnet, 2004). Alle DESI-Instrumente wurden im Rahmen einer Pilotierungsstudie einer testtheoretischen Qualitätsprüfung auf der Grundlage der Item-Response-Theorie unterzogen und für die Hauptuntersuchung optimiert (DESI-Konsortium, 2003). Im Folgenden sind die theoretischen Grundlagen für die Testmodelle im Deutschen und im Englischen differenzierter dargestellt.

### **3.3.3 DESI-Testkonstrukte und Testverfahren im Deutschen**

Die Aufgabenbereiche des Sprachunterrichts in der ersten Unterrichtssprache lassen sich nach Ingendahl (1999) wie folgt definieren: Der Deutschunterricht soll den Schülerinnen und Schülern *sprachpragmatische Kompetenz*, das heißt situationsspezifisches Sprechen und Schreiben und *sprachliches Wissen* und dessen Anwendung vermitteln sowie im Sprachlernprozess *Sprachbewusstheit* dadurch wecken, dass die Reflexion über das Gehörte angeregt wird und eigenständige Kommunikationsprozesse ermöglicht werden. DESI erfasste diese im Deutschunterricht zu vermittelnden Kompetenzen mit Testinstrumenten für die Bereiche Lesekompetenz, Argumentation, Sprachbewusstheit, Rechtschreibung, Textproduktion und Wortschatz.

*Lesekompetenz* stellt die Grundlage für alltägliche Informationsaufnahme dar und kann als elementare Voraussetzung der gesellschaftlichen Partizipation angesehen werden (Artelt et al., 2001). Die Definition von Lesekompetenz variiert je nach spezifischer Fragestellung, und auch ihre Erfassung ist kein einfaches Vorhaben: Die Schwierigkeit eines Lesetextes ist nicht für alle Leser gleich groß, vielmehr entsteht Textverständnis erst in der individuellen Interaktion zwischen Leser und Text (Grotjahn, 2000a). Zu berücksichtigen sind dabei nicht nur Vorwissen und Textsortenspezifik, sondern auch das Testformat (Alderson & Benerjee, 2002). Bei DESI bezieht sich das Konzept der Lesekompetenz auf drei Aspekte: Die Decodierung des Textes und die Konstruktion eines eigenen Sinnverständnisses, die interne Schwierigkeit des Textes sowie das Vorwissen der Leser. Der Test zur Lesekompetenz erfasste das Produkt eines Leseprozesses, in dessen zeitlichem Verlauf Texte allmählich tiefer

und komplexer verstanden werden (Willenberg, 2007a). In DESI wurden vier Texte unterschiedlicher Textsorte und Länge – Sach- und literarische Texte – verwendet, welche die Schülerinnen und Schüler lesen sollten, um anschließend pro Text zwischen drei und sieben Multiple-Choice-Fragen mit jeweils vier Antwortalternativen zu beantworten (Gailberger & Willenberg, 2008).

Die Fähigkeit der *Argumentation* beinhaltet das Beherrschen der Standardsprache, gegenseitiges Zuhören und das Sammeln von Argumenten für den eigenen Standpunkt mit der Zielsetzung, andere Menschen mit stichhaltigen Argumenten zu überzeugen oder zu einem anerkannten Konsens zu kommen (Götter, 1978). Die Testaufgaben zur Argumentation in DESI bestehen aus drei Aufgabenstämmen in Form von unvollendeten Dialogen, die von den Schülerinnen und Schülern unter Berücksichtigung eines vorgegebenen Kommunikationsziels schriftlich fortgeführt werden sollten. Dazu war es nötig, Sprechakte und fremde Argumente einzuschätzen, eigene Argumentationen zielführend zu produzieren und Einsicht in die Strukturen der vorgetragenen Argumentationen nachzuweisen (Krelle & Willenberg, 2008; Willenberg, Gailberger & Krelle, 2007).

Das Konzept der *Sprachbewusstheit* im Deutschen mit seinem Fokus auf „sprachbewusstem Sprachhandeln“ entspricht dem aktuellen Trend der Sprachdiagnostik, das Testen grammatikalischer Kenntnisse durch die Berücksichtigung sprachproduktiver Fähigkeiten zu ergänzen (Rea-Dickins, 1997, 2001). Denn klassische Tests zur Erfassung reiner grammatikalischer Kenntnisse leisten nur wenig zusätzliche Erklärung in Sprachtestbatterien und korrelieren redundant hoch mit anderen Testverfahren (Alderson, 1993). Darin zeigt sich die umfassende Bedeutung des sprachlichen Regelsystems: Grammatikalische Kenntnisse bilden die Grundlage jeder sprachlichen Testaufgabe und sprachlich korrektes Verstehen, Schreiben und Sprechen ist nur unter Berücksichtigung grammatikalischer Regeln möglich (Jude, 1975). So bezieht sich das entsprechende Testmodul Sprachbewusstheit Deutsch in DESI erweitert auf die Fähigkeit, gesprochene Sprache von Schriftsprache zu unterscheiden, grammatikalische und semantische Fehler in Sätzen zu erkennen und zu korrigieren, Konjunktivformen zu erkennen sowie mit direkter und indirekter Rede korrekt umzugehen, Stilformen zu erkennen und schriftsprachlich zu umschreiben (Eichler, 2007, 2008).

Die *Rechtschreibleistungen* im Deutschen wurden in DESI durch ein Diktat mit 68 Wörtern erhoben, das genügend orthographische Anforderungen enthielt, um auch zwischen

Schülerinnen und Schülern im mittleren und oberen Leistungsbereich noch zu differenzieren (Eichler & Thomé, 2008; Thomé & Gomolka, 2007). Neben der Berechnung quantitativer Kompetenzwerte wurden die Fehlschreibungen analog zu den Phasenmodellen des Schriftspracherwerbs (Thomé, 2003) nach ihrer inneren Struktur klassifiziert. Diese qualitative Fehleranalyse wurde in Anlehnung an die Oldenburger Fehleranalyse (OLFA) (Thomé & Thomé, 2004) durchgeführt, wonach bestimmte Fehlertypen nur auf bestimmten Kompetenzniveaus auftreten.

Unter *produktiver Schreibkompetenz* ist die Fähigkeit zu verstehen, Texte adressatengerecht zu formulieren und, je nach Zielsetzung, präzise zu informieren, überzeugend zu argumentieren oder Sprache ästhetisch ansprechend und kreativ einzusetzen. Im DESI-Testmodul *Textproduktion* Deutsch wurde die Schreibkompetenz von Schülerinnen und Schülern der Jahrgangsstufe neun mit Hilfe eines Aufsatzes zu einem festgesetzten Thema – einem Beschwerdebrief – erfasst (Neumann & Lehmann, 2008). Die Kompetenzen wurden in einem sehr aufwendigen Bewertungsprozess beurteilt (Harsch, Neumann, Lehmann & Schröder, 2007), dabei ergaben sich die beiden differenzierbaren Leistungsdimensionen der Textproduktion Pragmatik und der Textproduktion Systematik. In die Beurteilungen auf der Skala Pragmatik flossen angemessene Wortwahl, Strukturierung und Aufbau eines freien Textes ein, der Bereich der Systematik umfasste die grammatikalische Richtigkeit und Orthographie von Satzkonstruktionen.

Die Ausbildung eines *Alltagswortschatzes* ebenso wie die Kenntnis von Fremdwörtern und Metaphern kann als unabdingbare Voraussetzung für jegliche Art gesprochener und geschriebener Sprache angesehen werden (Willenberg, 2007b). Die Vermittlung eines umfassenden Vokabulars ist dementsprechend Bestandteil der Lehrpläne in Jahrgangsstufe neun. Die DESI-Testaufgaben zum Wortschatz testeten die Kenntnis von Basiswörtern, Fach- und Fremdwörtern sowie Abstrakta, die von den Schülerinnen und Schülern frei produziert, in Text- und Satzzusammenhängen rezeptiv auf ihre Richtigkeit überprüft oder in Abbildungen beschriftet werden mussten (Willenberg, 2008).

Die verwendeten Testverfahren im Deutschen integrierten unterschiedliche Itemformate, hierzu gehörten Multiple-Choice-Anwortalternativen, schriftliche Kurzantworten sowie offene Antwortformate zur freien Textproduktion (Tabelle 1).

**Tabelle 1: Testformate der Deutsch-Testverfahren in DESI**

Deutstests	Antwortformat	Itemanzahl
Leseverstehen	Multiple-Choice- und kurze schriftliche Antworten	38
Textproduktion	Offene schriftliche Antworten	8 Aufgaben
Sprachbewusstheit Grammatik	Fehler ankreuzen und kurze schriftliche Antworten	34
Rechtschreibung	Diktat, offene schriftliche Antworten	28
Argumentation	Multiple-Choice- und kurze schriftliche Antworten	16
Wortschatz	Kurze schriftliche Antworten	44

### 3.3.4 DESI-Testkonstrukte und Testverfahren im Englischen

Bei der Erfassung der Englischkompetenz in DESI galt das besondere Interesse der Sprachbeherrschung im Sinne aktiver kommunikativer Kompetenzen. Die vorliegende Arbeit integriert in die Analysen die Daten der Englisch-Sprachtests zum Hörverstehen und Leseverstehen, zur Textrekonstruktion, zur Sprachbewusstheit und zur freien schriftlichen und mündlichen Sprachproduktion.

Die Tatsache, dass sich zur Erforschung fremdsprachlichen *Hörverstehens* bisher nur wenige empirische Studien finden lassen, weist auf die Schwierigkeiten hin, diese Fähigkeit zu erfassen (Brindley, 1998). Zwar kann davon ausgegangen werden, dass es sich hierbei um ein eigenes Kompetenzkonstrukt handelt, jedoch lässt sich dieses nicht durch auditiven Input alleine operationalisieren (Buck, 2001). Zur Erfassung des Hörverstehens muss vielmehr auf weitere Fähigkeiten zurückgegriffen werden: Dazu gehören entweder das schriftliche oder mündliche Darlegen des Verstandenen oder das Erlesen von Antwortalternativen bei Testformaten mit Mehrfachantworten. Diese Einflüsse des Testformats können zu einer systematischen Verzerrung der Ergebnisse und Interkorrelationen mit jeweils anderen sprachlichen Teilkompetenzen führen (Alderson & Benerjee, 2002). Der DESI-Test zum *Hörverstehen* im Englischen konzentriert sich auf die grundlegenden Kompetenzen der Wahrnehmung, des Verstehens, der Interpretation und der Reflexion von fremdsprachlichen Äußerungen. Die Aufgaben beziehen sich auf theoretische Modelle von Buck (2001) und Rost (1990) sowie auf die Lernziele der Curricula zur Entwicklung des fremdsprachlichen Hörverstehens im Englischunterricht an deutschen Schulen (Nold & Rossa, 2007a). Der Hörverstehenstest besteht aus zwei verschiedenen Arten von auditiven Sprachproben in authentischer Sprechgeschwindigkeit, kurzen Dialogen und längeren Textpassagen. Zu diesen mussten schriftlich vorgegebene Multiple-Choice-Aufgaben mit jeweils drei Antwortmöglichkeiten gelöst werden (Nold & Rossa, 2008a).

Die Fähigkeit des *Leseverstehens* ist wie die des Hörverstehens eine reflexiv-rezeptive Sprachkompetenz, die auf Informationsverarbeitungs- und Verstehensprozesse ausgerichtet ist. Unterschiede finden sich jedoch in der Art des lesespezifischen Dekodierungsprozesses, dessen Grundlage die visuelle Verarbeitung graphischer Zeichen darstellt. Darüber hinaus müssen thematisch-inhaltliche Aspekte des allgemeinen und domänenspezifischen Weltwissens aktiviert und im Verstehensprozess mit den sprachlichen Anteilen des Leseprozesses koordiniert werden (Nold & Rossa, 2007b). Die Testaufgaben zum englischen Leseverstehen in DESI bestehen aus vier authentischen Texten der Gattungen Erzählung, Sachtext und Dramenausschnitt von einer Textlänge zwischen 270 und 400 Wörtern, zu denen pro Text jeweils zwischen neun bis zwölf Fragen im Multiple-Choice-Format beantwortet werden mussten (Nold, Rossa & Chatzivassiliadou, 2008).

*Sprachbewusstheit* im Englischen wird in DESI definiert als eine Fähigkeit, die sich in der Auseinandersetzung mit einer neu zu lernenden Sprache entwickelt und Lernende dazu befähigt, die Regeln der Fremdsprache in der eigenen Sprachproduktion und -rezeption mit Hilfe von implizit und explizit verfügbarem Sprach- und Handlungswissen zu erkennen und gegebenenfalls zu steuern (Nold & Rossa, 2007c). Diese Definition lehnt sich an das britische Konzept der *language awareness* an; neben der kognitiven Domäne wird in den DESI-Testaufgaben auch die soziale Domäne von Sprachbewusstheit angesprochen (Garrett & James, 1991; Gnutzmann, 1997). Die Testinhalte fordern von den Schülern das Auffinden und Verbessern sprachlicher Fehler in Kurztexten und Sätzen und das Auswählen situativ angemessener sprachlicher Reaktionen (Nold & Rossa, 2008b).

Auch in der Fremdsprache zielt der Unterricht auf die Vermittlung des freien, kommunikativen Schreibens ab, wobei das (semi)-kreative Schreiben explizit als Aufgabenstellung und Übungsform in den Lehrplänen erwähnt wird. Um in DESI die Fähigkeit zur Textproduktion im Englischen zu erfassen, sollten die Schülerinnen und Schüler einen Aufsatz zum Thema Klassenfahrt beziehungsweise einen persönlichen Brief schreiben (Harsch et al., 2007). Die Texte wurden anhand unterschiedlicher Kriterien beurteilt, die letztendlich zu einem Gesamtleistungswert zusammengefasst wurden (Harsch, Schröder & Neumann, 2008).

Der in DESI eingesetzte *Textrekonstruktionstest* schließlich orientierte sich am C-Test (Grotjahn, 1987; Grotjahn & Stemmer, 1985), einem gut erforschten Testinstrument zur

Erfassung verschiedener Teilkomponenten der Sprachfähigkeit. Dieser Lückentext, bei dem fehlende Wortteile zu ergänzen sind, erfasst die Fähigkeit, Wortschatz, Grammatik und textuelles Wissen integriert anzuwenden (Harsch & Schröder, 2007). In DESI wurde der C-Test als ein integratives Testformat eingesetzt, mit dem Kenntnisse des Wortschatzes und der Grammatik sowie Lesestrategien und Interpolationsfähigkeit erfasst wurden (Harsch & Schröder, 2008; Raatz, 1985). Den Schülern wurden jeweils vier kurze Lückentexte mit je 25 auszufüllenden Lücken vorgegeben.

Zur Erfassung der mündlichen Sprachproduktion im Englischen wurde auf das computergestützte Testverfahren SET-10 zurückgegriffen, das speziell für DESI modifiziert wurde (Ordinate, 2004). Eine Teilstichprobe von 1 870 Schülerinnen und Schülern beantwortete mündlich auf Englisch offen formulierte Fragen, in denen sprachliches Verständnis, die Verfügbarkeit lexikalischer Redemittel sowie Korrektheit und Aussprache bewertet wurden (Nold & Rossa, 2008c).

Auch die Antwortformate der Tests zur Sprachkompetenz im Englischen variieren zwischen den Teilbereichen (Tabelle 2).

**Tabelle 2: Testformate der Englisch-Testverfahren in DESI**

Englishtests	Antwortformat	Itemanzahl
Leseverstehen	Multiple-Choice	46
Schreiben/Textproduktion	Offene schriftliche Antworten	8 Aufgaben
Sprachbewusstheit Grammatik	Multiple-Choice- und offene Antworten	29
Textrekonstruktion	Offene schriftliche Antworten	300
Hörverständnis	Multiple-Choice	51
SET-10	Offene gesprochene Antworten	53

In zwei Modulen im Englischen wurden die Antworten ausschließlich durch das Ankreuzen von Mehrfachwahl-Antwortalternativen erfasst, dies betrifft das Leseverstehen und das Hörverstehen. Damit lehnen sich die DESI-Aufgaben an traditionelle Testformate zum Hör- und Leseverstehen an, die das Multiple-Choice-Format wählen, um mögliche Effekte der Schriftsprachkompetenz auf das Antwortverhalten zu vermeiden. In allen anderen Testmodulen waren die Antworten auf einzelne Fragen schriftlich zu geben, teilweise wurden sie durch Multiple-Choice-Formate ergänzt. Besonders die offenen Aufgaben zur freien Textproduktion stellten große Herausforderungen an die Auswertung und Beurteilung der

Schülerleistungen. Objektivität und Reliabilität wurden daher durch geschulte Rater sichergestellt.

### **3.3.5 Datengrundlage – Personenschätzer und Stichprobengewichte**

Die getesteten Schülerinnen und Schüler bearbeiteten in DESI eine Vielfalt verschiedener Aufgaben, die nach vorher definierten Kodierschemata auf ihre Qualität und Korrektheit hin bewertet wurden. Von den erfassten Testantworten wurde dann auf die individuelle Kompetenzausprägung in den einzelnen Testbereichen im Deutschen und im Englischen geschlossen. Eine Besonderheit der Erhebung in DESI war, dass die Vorgabe der Leistungstests in einem so genannten Matrix-Design erfolgte, bei dem jeder Schüler und jede Schülerin nur einen Teil der Aufgaben eines jeden Testbereichs bearbeitete (Mislevy, Beaton, Kaplan & Sheehan, 1992). Dieses bei Large-Scale-Studien übliche Vorgehen hat den untersuchungsökonomischen Vorteil, dass insgesamt eine breite Menge von Testaufgaben eingesetzt werden kann, die individuelle Testzeit aber dennoch begrenzt bleibt. Der einzelnen Testperson wird jeweils nur ein Ausschnitt aus der Gesamtheit aller Aufgaben vorgelegt. Für DESI war das Matrix-Design noch aus einem weiteren Grund unerlässlich: Sprachkompetenzen wurden zu Beginn und am Ende des neunten Schuljahres getestet, um den Zuwachs während dieses Schuljahres ermitteln zu können. Um Erinnerungseffekte zu vermeiden, können in einem solchen Längsschnittdesign dieselben Personen nicht zu zwei Zeitpunkten mit denselben Aufgaben getestet werden. Die Vorgabe der Tests im Matrix-Design erlaubte es, jeder Schülerin und jedem Schüler zu beiden Zeitpunkten verschiedene Aufgaben vorzugeben. Es ist daher nicht möglich, die bloße Anzahl gelöster Aufgaben in jedem Testmodul als Indikator für die jeweils interessierende Kompetenz zu verwenden, weil ein solches Vorgehen die unterschiedlichen Schwierigkeiten der Aufgaben nicht berücksichtigen würde.

Zur Schätzung der Fähigkeitswerte für jede Person wurden in DESI daher Modelle der Item-Response-Theorie (IRT) verwendet (Rasch, 1960; Wright, 1999). In diesen Modellen werden die Schwierigkeiten der Testaufgaben und die Kompetenzen der getesteten Personen auf derselben Kompetenzskala beschrieben; zwischen der Kompetenz der Schülerinnen und Schüler und den erwarteten Lösungshäufigkeiten der jeweiligen Aufgaben werden statistische Zusammenhänge formuliert (Allerup, 1999). Die Auswertung mit IRT-Modellen erlaubt eine gemeinsame Schätzung der Kompetenzen aller Schüler auf derselben Skala, auch wenn diese

*unterschiedliche* Test-Aufgaben bearbeitet haben (Rost, 2004b). Das in DESI eingesetzte spezifische Skalierungsmodell war ein generalisiertes Rasch-Modell, welches in der Analysesoftware ConQuest implementiert ist (Wu, Adams & Wilson, 1998). In diesem Modell lassen sich sowohl dichotome als auch ordinale Auswertungsformate innerhalb desselben Tests modellieren (Masters & Wright, 1997). Diese Skalierung ermöglicht die kriterienorientierte Beschreibung von Fähigkeiten in Bezug auf die Aufgabenschwierigkeiten auf einer einzigen Kompetenzskala (Wilson, 2003).

Diese Aufgabenschwierigkeit ist im Rasch-Modell als der Punkt definiert, an dem die Aufgabe von Personen mit einer gleich hohen Kompetenz zu 50 % gelöst wird. Umgekehrt können die Kompetenzen von Personen in Form von Lösungswahrscheinlichkeiten interpretiert werden: Personen, deren Kompetenz der Schwierigkeit einer Aufgabe entspricht, sollten diese Aufgabe mit einer Wahrscheinlichkeit von 50 % lösen, wenn das Modell gilt. Allerdings kann argumentiert werden, dass eine Lösungswahrscheinlichkeit von 50 % in einem diagnostischen Kontext ein eher niedriges Kriterium darstellt, typischerweise werden daher höhere Schwellen verwendet, beispielsweise 62 % Lösungswahrscheinlichkeit in PISA oder 65 % in der Third International Mathematics and Science Study (TIMSS) (Baumert, Bos & Lehmann, 2000) und auch für DESI wurde das 65-Prozent-Kriterium gewählt (Hartig, Jude & Wagner, 2008).

Für die Ermittlung von *individuellen numerischen Kompetenzwerten* der getesteten Personen stehen innerhalb der IRT verschiedene Techniken zur Verfügung. Alle diese Schätzverfahren beruhen auf iterativen Verfahren, deren Ziel es ist, die individuellen Fähigkeiten so zu schätzen, dass die Wahrscheinlichkeit der tatsächlich beobachteten Antworten maximiert wird. Dieses Vorgehen ist auch als Prinzip der Maximum-Likelihood-Schätzung (ML) (Rupp, 2005) bekannt. Die nach dem ML-Prinzip geschätzten individuellen Fähigkeiten oder auch *Personenparameter* sind im Rasch-Modell in erster Linie eine Funktion der Anzahl gelöster Aufgaben: Je mehr Aufgaben eine Person gelöst hat, desto höher wird ihre Kompetenz eingeschätzt. Bei der Auswertung von Daten aus einem Matrixdesign spielt zusätzlich noch eine Rolle, welche Aufgabenmenge eine Person bearbeitet hat, sodass sich berücksichtigen lässt, ob die Menge gelöster Aufgaben auf leichteren oder schwierigeren Aufgaben basiert (Van der Linden, Veldkamp & Carlson, 2004).

In DESI wurden so genannte Weighted Likelihood Estimates (WLEs) (Warm, 1989) als ML-basierte Kompetenzwerte geschätzt, also Werte, die jeweils auf den individuellen gelösten Aufgaben der Personen basieren. Zur Berechnung der WLEs werden die Schätzwerte mittels individuell verfügbarer Testinformationen (A-priori-Verteilung) gewichtet, dadurch wird eine Reduktion der messwertbedingten Verzerrung der Populationsverteilung erreicht. WLE-basierte Kompetenzschätzungen stellen für die individuelle Messung optimale Personenschätzer dar, allerdings sind sie nicht frei von Messfehlern.

Es existieren alternative Personenschätzer, die verwendet werden, um die Verteilung der Kompetenzen und Zusammenhänge mit Hintergrundmerkmalen in der interessierenden Grundgesamtheit unverzerrt zu beschreiben (Mislevy et al., 1992). Hierfür kann die so genannte Plausible-Value-Technik verwendet werden, bei der zur Schätzung der Kompetenzen Hintergrundvariablen der Personen einbezogen werden (Hartig & Kühnbach, 2006). Eine gemeinsame Erzeugung von Plausible-Value-Schätzern für alle in DESI erfassten sprachlichen Kompetenzbereiche in einer simultanen mehrdimensionalen Analyse war auf Grund der großen Anzahl verschiedener Kompetenzen technisch nicht möglich. Für die Untersuchung von Zusammenhängen zwischen den verschiedenen Kompetenzen, wie sie Inhalt der vorliegenden Arbeit sind, werden daher die WLEs verwendet. Die Messfehler, die bei Analysen mit WLEs unvermeidlich sind, werden durch die Modellierung der Kompetenzstrukturen mit latenten Variablen aufgefangen (Jude & Hartig, 2005).

Wie schon in Abschnitt 3.3.1 dargestellt, wurde die DESI-Stichprobe als geclusterte Stichprobe gezogen, um die Repräsentativität der Daten bezüglich der anvisierten Population zu gewährleisten. Um darüber hinaus die Repräsentativität bestimmter Gruppen von Schulen, in diesem Fall solche mit bilingualen Bildungsgängen, zu gewährleisten, wurden diese überzufällig häufig in die Stichprobe aufgenommen. Da in diesem Fall die Stichprobenverteilung und das Ausmaß des Oversampling bekannt ist, können entsprechende Stichprobengewichte eingesetzt werden, die ungewollte Effekte und statistische Artefakte verhindern (Asparouhov, 2004; Foy, 1999). Die Verwendung von Stichprobengewichten in den Analysen korrigiert die Schätzung von Modellparametern und Stichprobenvarianzen (Stapleton, 2002).

### **3.4 Zusammenfassung**

Nachdem in Abschnitt 1 verschiedene Theorien der Sprachkompetenz und ihres Erwerbs und in Abschnitt 2 eine Vielfalt an empirischen Befunden zur Struktur von Sprachkompetenz dargestellt wurden, konzentrierte sich Abschnitt 3 speziell auf die Erfassung von Sprachkompetenz bei Schülerinnen und Schülern sowie in Schulleistungsstudien in Deutschland. Deutlich wurde, dass bisher nur wenige verallgemeinerbare empirische Befunde zu den sprachlichen Kompetenzen deutscher Schülerinnen und Schüler vorliegen. Dies liegt vor allem daran, dass die Durchführung vergleichender Studien über viele Jahrzehnte im deutschen Bildungssystem unüblich war und sich die Erfassung von Sprachkompetenz oft auf den Bereich der Defizitdiagnostik beschränkte. Dort, wo repräsentative Daten verfügbar sind, decken sie in der Regel nur einige wenige sprachliche Teilbereiche ab.

Dies ist ein Indikator dafür, dass die empirischen Befunde zur Sprachkompetenz und ihrer Struktur immer auch vor dem Hintergrund der Zielsetzung der jeweiligen Studie und des diagnostischen Auftrags zu bewerten sind: Defizitorientierte diagnostische Fragestellungen im Schulbereich richten den Blick zumeist auf nur einen spezifischen sprachlichen Kompetenzbereich, repräsentative Leistungsstudien erfassen eher umfassende Kompetenzkonzepte, während standardbasierte Testungen nicht auf theoretischen Modellen der Sprachkompetenz beruhen und statt theoretisch definierten Teilbereichen von Sprachkompetenz spezielle lehrplanorientierte Kompetenzkriterien erfassen.

Eine solch umfassende und differenzierte Erhebung von Sprachkompetenzen im Deutschen und im Englischen, wie sie die hier beschriebene DESI-Studie leistete, ist aktuell im deutschen Bildungssystem einmalig. Ausblickend ist anzumerken, dass die facettenreiche Erfassung von sprachlichen Kompetenzen mit curriculumbasierten, mehrdimensionalen Testverfahren sich in den kommenden Jahren in Deutschland über das Instrument der Bildungsstandards etablieren wird. Aktuell bieten die Daten der DESI-Studie eine ideale Grundlage für die Analyse der Struktur von Sprachkompetenz bei Schülerinnen und Schülern. Die grundlegenden Fragestellungen der vorliegenden Arbeit werden im folgenden Kapitel expliziert.

#### **4 Fragestellungen der Arbeit**

Die vorliegende Arbeit bietet auf Grund der außergewöhnlichen Datenlage die Möglichkeit, die Struktur von Sprachkompetenz im Deutschen und Englischen differenziert unter Verwendung einer gemeinsamen Datenbasis zu überprüfen. Wie bereits in Abschnitt 3.3.1 dargestellt, liegen Daten von 10 543 Personen zu insgesamt zwölf Sprachtestbereichen vor. Teilstichproben von Schülerinnen und Schülern mit nichtdeutscher Muttersprache und bilinguaem schulischem Bildungshintergrund ermöglichen zudem differenzierte Gruppenvergleiche. Die nachfolgend dargestellten Fragestellungen bauen in ihrer Komplexität aufeinander auf und lassen sich wie folgt gliedern:

Die ersten Analysen zielen darauf ab, die Struktur von Sprachkompetenz innerhalb der beiden Sprachen wie auch für beide gemeinsam zu beschreiben. Betrachtet wird dabei die Gesamtstichprobe unter Berücksichtigung der hierarchischen Schachtelung der Daten. Diese Analysen dienen dazu, grundlegende theoretische Modelle, die in den Abschnitten 1 und 2 dargestellt wurden, miteinander zu vergleichen. Ziel ist es, ein theoriegeleitetes Modell der Struktur der Sprachkompetenz darzustellen, das die Struktur in den empirischen Daten für das Deutsche und das Englische abbilden lässt. In einem nächsten Schritt werden die Zusammenhänge zwischen den beiden Sprachen in einem gemeinsamen Modell untersucht.

Darauf aufbauend wird analysiert, wie dieses Modell speziell auf die Mehrebenenstruktur der Daten anzupassen ist, ob also die Struktur von Sprachkompetenz auf Individualebene und auf Klassenebene für beide Sprachen getrennt und gemeinsam unter Beibehaltung des Modells für die Gesamtstichprobe dargestellt werden kann. In einem weiteren Schritt wird analysiert, ob das gefundene Modell der Struktur von Sprachkompetenz sich in spezifischen Untergruppen unterscheidet. Abschließend wird untersucht, welche Bedeutung den kognitiven Grundfähigkeiten im Gesamtmodell der Struktur von Sprachkompetenz zukommt.

#### 4.1 Zur Struktur des Deutschen und des Englischen

Sprachkompetenz wird in DESI in zwei Bereichen erfasst, im Deutschen als Unterrichtssprache und im Englischen als Fremdsprache. Dabei wird jeder Bereich mit sechs verschiedenen Testverfahren abgedeckt. Zu analysieren ist in einem ersten Schritt, in welcher Art und Weise die verwendeten Testverfahren den jeweiligen sprachlichen Bereich repräsentieren, dem sie zugeordnet wurden. Analysiert werden zunächst also die Zusammenhänge der Leistungen in den Testmodulen jeweils für das Deutsche und das Englische mit dem Ziel, Aussagen über die Binnenstruktur der Sprachkompetenz zu machen.

Wie bereits dargestellt wurde, existieren verschiedene theoretische Modelle und empirische Befunde, welche die Binnenstruktur von Sprachkompetenz durchaus widersprüchlich darstellen (vgl. Kapitel 2):

- *eindimensionale* Modelle, welche die sprachlichen Leistungen in verschiedenen Teilbereichen letztendlich auf einen gemeinsamen latenten Faktor zurückführen,
- *mehrdimensionale* Modelle, die differenzierbare Teilbereiche von Sprachkompetenz postulieren, sowie
- *hierarchische* Modelle, die latente Faktoren zweiter Ordnung annehmen und davon ausgehen, dass die Zusammenhänge zwischen sprachlichen Teilbereichen auf spezifische übergeordnete Faktoren zurückzuführen sind.

Sowohl die mehrdimensionalen als auch die hierarchischen Modelle postulieren, wie bereits dargestellt, durchaus unterschiedliche latente Faktoren, exemplarisch sei hier nur erneut auf die gegensätzlichen Annahmen von *rezeptiven* und *produktiven* Dimensionen sowie einer *schriftbasierten* und einer *auditiven* Dimension hingewiesen (vgl. Abbildung 3). Auch die Höhe des Zusammenhangs zwischen bestimmten sprachlichen Teilbereichen, welche dann als gemeinsame latente Dimensionen angesehen werden könnten, unterscheidet sich, wie bereits berichtet, von Studie zu Studie.

Um der Frage nach der Binnenstruktur von Sprachkompetenz in der vorliegenden Arbeit nachzugehen, werden die drei unterschiedlichen Arten von Modellen auf der Grundlage der in Abschnitt 2.1 dargestellten, theoretisch fundierten Modelle der Struktur von Sprachkompetenz verglichen. Allerdings sollen an dieser Stelle keine Hypothesen darüber formuliert werden, welches dieser Modelle die Datenlage in den jeweiligen Sprachen am

besten abbilden kann. Jedoch ist auf Grund der in Abschnitt 2.2.4 dargestellten bekannten Unterschiede zwischen der Struktur in der Erst- und in der Fremdsprache und der teilweise unterschiedlichen verwendeten Testverfahren davon auszugehen, dass sich die Binnenstrukturen der beiden sprachlichen Bereiche unterscheiden. So lässt sich folgende allgemeine Hypothese 1 formulieren:

***Hypothese 1:*** Die darstellbare Binnenstruktur der Sprachkompetenz weist für das Deutsche eine andere latente Faktorenstruktur auf als für das Englische.

Die Ergebnisse dieser Analysen bilden die Grundlage für die darauf folgende Analyse der Zusammenhänge zwischen dem Deutschen und dem Englischen. Während die Analysen unter Hypothese 1 nach den strukturellen Zusammenhängen *innerhalb der beiden Sprachen* fragten, bezieht sich folgende Fragestellung darauf, inwieweit spezifische Zusammenhänge *zwischen den beiden Sprachen* angenommen werden können.

Auf der Grundlage von bisherigen Erkenntnissen ist davon auszugehen, dass spezifische Zusammenhänge zwischen den Kompetenzen in der Erstsprache und in weiteren (Fremd-)Sprachen bestehen, die sich besonders auf ähnliche sprachliche Teilbereiche beziehen (vgl. Abschnitt 1.6). So ist davon auszugehen, dass Zusammenhänge zwischen korrespondierenden sprachlichen Teilbereichen, also beispielsweise dem Schreiben im Deutschen und dem Schreiben im Englischen, höher ausfallen, als zwischen unterschiedlichen Teilbereichen, wie dem Schreiben im Deutschen und dem Hörverstehen im Englischen. Hieraus leitet sich die folgende Hypothese 2 ab:

***Hypothese 2:*** Die speziellen Zusammenhänge zwischen den korrespondierenden latenten Faktoren, die im Deutschen und im Englischen erfasst wurden, sind höher als die Zusammenhänge zwischen nicht korrespondierenden sprachlichen Teilbereichen im Deutschen und Englischen.

Auf der Grundlage dieser Analysen, die im Wesentlichen unterschiedliche Modelle und Vorgehensweisen bereits vorhandener Studien aufgreifen und anhand der DESI-Daten soweit als möglich replizieren, werden im Folgenden die für dieses Arbeit spezifischen, auf erweiterten Modellen basierenden Fragestellungen dargestellt. Hierzu gehören Annahmen und Methoden der *Mehrebenenanalyse*, deren Grundlagen in Abschnitt 5.3 ausführlich dargestellt sind.

#### 4.2 Struktur von Sprachkompetenz unter Berücksichtigung der hierarchischen Schachtelung der Daten

Wie in Abschnitt 5.3 dargestellt, ist bei Vorliegen von hierarchisch geschachtelten Daten die Bedeutsamkeit eben dieser Clusterung zu analysieren und gegebenenfalls in statistischen Analysen zu berücksichtigen, um differenzielle Effekte der verschiedenen Ebenen analysieren zu können. Daraus folgend wird Hypothese 3 formuliert:

**Hypothese 3:** *Die in der DESI-Studie erhobenen Daten weisen eine hohe Abhängigkeit zwischen den Klassen, das heißt eine bedeutsame hierarchische Schachtelung auf.*

Unter der Annahme, dass hohe Clustereffekte bestehen, wird im Weiteren danach gefragt, wie sich die Struktur von Sprachkompetenz beschreiben lässt, wenn Individual- und Klassenebene getrennt betrachtet werden. Die Analysen in diesem Teil der Arbeit untersuchen daher die Struktur von Sprachkompetenz im Englischen und Deutschen auf zwei Ebenen: Die Zusammenhänge zwischen den sprachlichen Teilkomponenten werden für jede der beiden Ebenen differenziert analysiert und in einem gemeinsamen Mehrebenenmodell für jede Sprache dargestellt.

Auf der Grundlage bisheriger Befunde zu strukturellen Zusammenhängen in Mehrebenenmodellen (vgl. 5.3.1) ist davon auszugehen, dass sich Strukturen auf übergeordneten Analyseebenen – in diesem Fall auf der Ebene der Klassen – homogener darstellen als auf der untergeordneten Analyseebene – in diesem Fall auf Ebene der Individuen in einer Klasse. Dies bedeutet, dass zwischen den manifesten Variablen auf Individualebene meist niedrigere Korrelationen bestehen, als dies für die geschätzten Zusammenhänge auf aggregierter Ebene der Fall ist. Die Hypothese zur Binnenstruktur von Sprachkompetenz im Deutschen und im Englischen bei mehrebenenanalytischer Betrachtung wird daher wie folgt formuliert:

**Hypothese 4:** *Die Zusammenhänge der sprachlichen Teilbereiche sind auf Klassenebene höher als auf Individualebene.*

Es kann weiterhin davon ausgegangen werden, dass sich nicht nur die quantitative Höhe der Zusammenhänge auf den beiden Ebenen unterscheidet, sondern dass sich spezielle Unterschiede auch in der Binnenstruktur von Sprachkompetenz auf Klassen- und auf Individualebene beschreiben lassen. Dementsprechend werden Modelle mit gleichen Strukturen auf beiden Ebenen mit solchen Modellen verglichen, die spezifische strukturelle

Zusammenhänge für jede der beiden Ebenen annehmen. Hierzu werden folgende Hypothesen formuliert:

***Hypothese 5:*** *Die faktorielle Struktur von Sprachkompetenz ist auf Klassenebene homogener als auf Individualebene.*

Aufbauend auf diesen Mehrebenenmodellen für das Deutsche und das Englische ergibt sich dann wiederum die Frage nach den Zusammenhängen zwischen den Leistungen in beiden Sprachen (vgl. Hypothese 2), die in diesem Fall ebenfalls in einer Mehrebenenstruktur zu beschreiben sind. Die Zusammenhänge sind sowohl für die Individual- als auch für die Klassenebene getrennt zu beschreiben, jedoch dann in einem gemeinsamen Modell zu schätzen.

Es kann angenommen werden, dass sich für die spezifischen Zusammenhänge zwischen den Kompetenzen im Deutschen und im Englischen analog zum Modell für die Gesamtstichprobe auch ein Mehrebenenmodell abbilden lässt. Aufgrund dessen, dass die institutionelle Vermittlung der Fremdsprache in Schulen in der Regel auf den Kompetenzen in der Erstsprache aufbaut (vgl. Abschnitt 1.5) und besonders im Fall des Deutschen und des Englischen sprachübergreifende Gemeinsamkeiten wie beispielsweise ein gleiches Schriftsystem existieren, ist die Annahme zu treffen, dass die allgemeine Sprachkompetenz in der Fremdsprache von der sprachlichen Kompetenz in der Erstsprache beeinflusst wird. Darüber hinaus ist davon auszugehen, dass die spezifischen sprachlichen Teilbereiche beider Sprachen systematisch zusammenhängen. Daraus ergibt sich Hypothese 6:

***Hypothese 6:*** *Die Kompetenzen im Englischen lassen sich auf Individual- und auf Klassenebene durch die Kompetenzen im Deutschen vorhersagen.*

Das Modell der Struktur von Sprachkompetenz im Deutschen und Englischen, das auf Basis der Ergebnisse der Hypothesen 1 bis 6 spezifiziert werden kann, dient als Ausgangspunkt für die weiteren Analysen.

### 4.3 Gruppenunterschiede – Erstsprache, Unterrichtsangebot und Geschlecht

Aktuelle Befunde zur Dimensionalität von Sprachkompetenz weisen darauf hin, dass sich die Strukturen von Kompetenz in Abhängigkeit von der Gruppenzugehörigkeit unterscheiden können (vgl. Abschnitt 1.7). Hierzu gehören Unterschiede zwischen Mädchen und Jungen, und zwischen Personen mit unterschiedlicher Erstsprache. Darüber hinaus ist es in der vorliegenden Arbeit von Interesse, ob sich die Struktur besonders der Fremdsprachenkompetenz in Abhängigkeit davon unterscheidet, unter welchen Unterrichtsbedingungen die Fremdsprache erworben wurde. Dies betrifft in der vorliegenden Arbeit die Unterscheidung zwischen Schülerinnen und Schülern, die ein bilinguales Unterrichtsangebot im Englischen wahrnehmen, und solchen, die in der Fremdsprache Englisch den Regelunterricht besuchen.

Mit Mehrgruppenmodellen wird daher geprüft, ob sich die Struktur von Sprachkompetenz für die verschiedenen Gruppen gleich beschreiben lässt, oder ob unterschiedliche Binnenstrukturen in Abhängigkeit von der Gruppenzugehörigkeit angenommen werden müssen. Ausgangspunkt sind dabei jeweils die in den vorangegangenen Analysen für die Gesamtstichprobe berichteten Modelle. Folgende Hypothesen werden formuliert:

***Hypothese 7:** Die Struktur von Sprachkompetenz im Deutschen und im Englischen unterscheidet sich zwischen Mädchen und Jungen.*

***Hypothese 8:** Die Struktur von Sprachkompetenz im Deutschen und Englischen unterscheidet sich zwischen den Schülerinnen und Schülern, die ein bilinguales Unterrichtsangebot wahrnehmen, und denjenigen, die den Regelunterricht in der Fremdsprache besuchen.*

***Hypothese 9:** Die Struktur von Sprachkompetenz im Deutschen und im Englischen unterscheidet sich zwischen Schülerinnen und Schülern mit deutscher Erstsprache und denjenigen mit einer anderen Erstsprache.*

#### 4.4 Die Bedeutung kognitiver Grundfähigkeit für die Struktur von Sprachkompetenz

Wie berichtet sind die kognitiven Grundfähigkeiten als eine Grundbedingung für sprachliche Leistungen anzusehen (vgl. Abschnitt 1.7). Anzunehmen ist, dass Zusammenhänge zwischen den kognitiven Grundfähigkeiten und den Leistungen in den sprachlichen Teilbereichen im Deutschen und im Englischen vorhanden sind. Hieraus wird Hypothese 10 wie folgt abgeleitet:

*Hypothese 10: Es bestehen unterschiedlich hohe Zusammenhänge zwischen den kognitiven Grundfähigkeiten und sprachlichen Teilbereichen auf Individual- und Klassenebene.*

Darüber hinaus ist danach zu fragen, welche allgemeine Bedeutung den kognitiven Grundfähigkeiten zukommt. Unter der Annahme, dass sie eine Grundbedingung für menschliche Sprachkompetenz darstellen kann die Hypothese aufgestellt werden, dass die Zusammenhänge zwischen sprachlichen Teilbereichen im Deutschen und im Englischen *ausschließlich* durch gemeinsame kognitive Grundfähigkeiten zustande kommen, diese quasi als Mittler die Transfereffekte zwischen den Sprachen bedingen. In diesem Fall sollte unter Einbeziehung der kognitiven Grundfähigkeiten der systematische Zusammenhang zwischen den sprachlichen Teilbereichen ganz verschwinden. Existieren jedoch über die kognitiven Grundfähigkeiten hinausgehende, spezifisch sprachliche Transfereffekte, so sollten sich die Zusammenhänge zwischen den sprachlichen Kompetenzen im Deutschen und im Englischen weiterhin zeigen, lediglich korrigiert um den Anteil der allgemeinen kognitiven Fähigkeit, der gleichermaßen in die Lösung von Sprachtestaufgaben einfließt. Die Struktur der Sprachkompetenz würde sich dann nicht ändern. Darüber hinaus könnten differenzielle Effekte der kognitiven Grundfähigkeit auf die Struktur von Sprachkompetenz beobachtet werden, wenn sich unterschiedlich hohe Zusammenhänge zwischen kognitiver Grundfähigkeit und Leistung zeigen. Daher lautet schließlich Hypothese 11:

***Hypothese 11:*** *Unter Einbeziehung der individuellen kognitiven Grundfähigkeiten in ein gemeinsames Modell der Struktur der Sprachkompetenz im Deutschen und im Englischen sind die gleichen strukturellen Zusammenhänge darstellbar wie ohne Berücksichtigung der kognitiven Grundfähigkeiten.*

Die hier dargestellten elf Hypothesen bauen in ihrer Logik aufeinander auf. Beginnend mit Beschreibungen der Struktur von Sprachkompetenz in den jeweiligen Sprachen Deutsch und Englisch werden Zusammenhänge zwischen beiden Sprachen analysiert. Aufbauend darauf werden die Strukturen auf Individual- und Klassenebene in beiden Sprachen differenzierter und in einem gemeinsamen Modell zu beschreiben sein. Darüber hinaus wird geprüft, inwieweit Unterschiede zwischen Personengruppen bestehen, und schließlich wird nach der Rolle der kognitiven Grundfähigkeiten in der Struktur von Sprachkompetenz gefragt. Grundlage für die Überprüfung dieser Fragestellungen sind die im folgenden Abschnitt dargestellten statistischen Analysemethoden.

## 5 Methodische Grundlagen der Auswertung

Die Analyse der komplexen und umfangreichen Datenlage, wie sie aus DESI resultiert, ist mit einer Reihe methodischer Herausforderungen verbunden. Wie in Abschnitt 2.5 dargestellt wurde, sind die gefundenen Ergebnisse auch immer vor dem Hintergrund der verwendeten Analysemethoden zu bewerten. In dieser Arbeit kommen daher je nach Fragestellung unterschiedliche statistische Verfahren zum Einsatz, um eine möglichst gesicherte Interpretation der Ergebnisse zu gewährleisten. Neben der obligatorischen deskriptiven Analyse der Daten wird die Struktur von Sprachkompetenz in der vorliegenden Arbeit durch *exploratorische Faktorenanalysen* (Abschnitt 5.1), *konfirmatorische Faktorenanalysen und Strukturgleichungsmodelle* (Abschnitt 5.2), *Mehrebenen-Strukturgleichungsmodelle* (Abschnitt 5.3.2) sowie durch *Mehrgruppenvergleiche* analysiert.

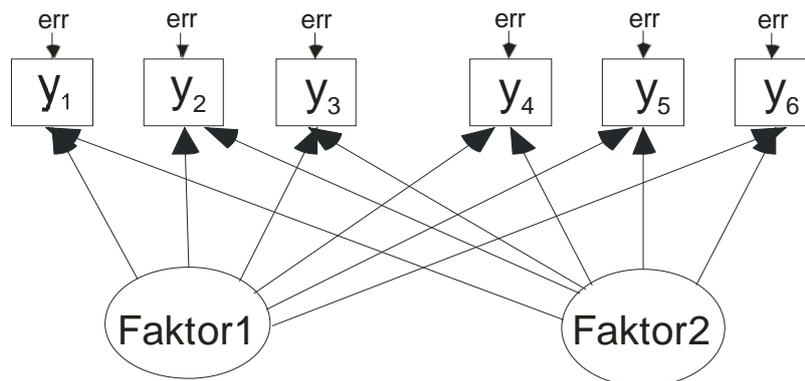
### 5.1 Exploratorische Faktorenanalysen

Unter Faktorenanalysen versteht man eine Sammlung von statistischen Verfahren, die dazu dienen, Variablen anhand ihrer gemeinsamen Korrelationen auf diesen Korrelationen zugrundeliegende Faktoren zurückzuführen. Ein solcher nicht beobachtbarer Faktor steht stellvertretend für ein theoretisch angenommenes latentes Konstrukt, welches den gemeinsam hoch korrelierenden, gemessenen Variablen zugrunde liegt. Mittels Faktorenanalysen können aus vielen Variablen einige wenige Faktoren extrahiert werden, welche die Kovarianz der Variablen gut erklären und für weitere Analysen anstelle der einzelnen Variablen verwendet werden können. Daher wird die Faktorenanalyse auch als datenreduzierendes Verfahren bezeichnet (Moosbrugger & Hartig, 2003).

In der vorliegenden Arbeit werden exploratorische Faktorenanalysen verwendet, um die Zusammenhänge in den Daten grob zu analysieren. Diese Methode kommt darüber hinaus – durchaus nicht unkritisiert – in vielen empirischen Studien zur Struktur von Sprachkompetenz zum Einsatz (Bachman, 1990a). Dementsprechend werden die Verfahren der exploratorischen Faktorenanalyse auch in der vorliegenden Arbeit in einem ersten Auswertungsschritt verwendet. Zum einen, um die so gewonnenen Ergebnisse mit denen anderer Studien vergleichen zu können, zum anderen, um auch die empirisch vorhandenen Zusammenhänge in den Daten der Stichprobe einzubeziehen. Die Verfahren der exploratorischen Faktorenanalyse sind auch im Rahmen der Mehrebenenanalysen (vgl. Abschnitt 5.3.2) in der

Regel der erste Schritt, um die strukturellen Zusammenhänge in der Datenmatrix zu analysieren.

Für die Berechnung der Faktoren existiert eine Vielzahl von Verfahren, von denen die Hauptkomponentenanalyse als Extraktionsverfahren am häufigsten verwendet wird und sukzessiv jene Faktoren ermittelt, die einen maximalen Anteil der gemeinsamen Varianz erklären und somit hohe *Eigenwerte* aufweisen. Die Extraktion von Faktoren erfolgt auf der Grundlage der Korrelationen zwischen den einzelnen Variablen, die Höhe der Korrelation zwischen zwei Variablen bestimmt deren Zuordnung zu einem gemeinsamen Faktor (Rost, 2004a). Kann ein einzelner Faktor nicht alle Korrelationen der Variablen untereinander erklären, muss ein weiterer Faktor aus den Restkorrelationen partialisiert werden. Dies erfolgt so lange, bis keine Restkorrelationen mehr vorliegen. Dabei werden in der exploratorischen Faktorenanalyse für alle Variablen Ladungen auf allen Faktoren zugelassen (vgl. Abbildung 4)



**Abbildung 4: Schematische Darstellung einer exploratorischen Faktorenanalyse (die beobachteten Variablen sind mit  $y$  gekennzeichnet)**

Die Abbildung zeigt eine schematische Darstellung für eine exploratorische Faktorenanalyse, auf die im Folgenden immer wieder Bezug genommen wird. Die rechteckigen Kästchen stellen die beobachteten Variablen dar, die auch als *manifeste Indikatoren* bezeichnet werden; die ovalen Formen stellen die zugrunde liegenden, nicht direkt beobachtbaren Faktoren dar. Die gerichteten Pfeile zwischen Faktoren und Indikatoren stellen die im Modell zugelassenen Faktorladungen dar, also die Korrelationen zwischen der jeweiligen Variable und dem Faktor. In diesem Beispiel der exploratorischen Faktorenanalyse laden alle Indikatoren auf allen Faktoren. Die nicht durch diese Faktorenlösung erklärte Varianz der Indikatoren wird durch Fehlerterme dargestellt.

Die Methoden der Faktorenanalysen sind explorative Verfahren, die keine Vorannahmen über die Anzahl der Faktoren treffen. Diese können vor der Analyse zwar theoriegeleitet formuliert werden, werden in der exploratorischen Faktorenanalyse jedoch nicht überprüft. Daher wird die exploratorische Faktorenanalyse in dieser Arbeit nur als ein Teilschritt verwendet, der durch theoriegeleitete, konfirmatorische Modelle ergänzt wird. Es gibt unterschiedliche Kriterien, anhand derer über die *Anzahl der zu verwendenden Faktoren* entschieden werden kann, in der vorliegenden Arbeit werden der graphische *Scree-Test* (Cattell, 1966) und die *Parallelanalyse* (Horn, 1965) eingesetzt. Beim Scree-Test werden die Eigenwerte der Faktoren abfallend sortiert graphisch abgetragen und die Faktoren mit hoher Varianzaufklärung als relevante Faktoren betrachtet. Die Parallelanalyse erweitert dieses Vorgehen, indem sie das Eigenwertediagramm der korrelierten Daten mit einem Eigenwertediagramm von Zufallsdaten vergleicht; nur die Faktoren, deren Eigenwerte höher sind als die Zufallseigenwerte, gelten dann als bedeutsam.

Unterstützend können theoretische Annahmen über die Anzahl der Faktoren für eine Entscheidung hinzugezogen werden, auch kann anhand der aufzuklärenden Gesamtvarianz über die erforderliche Faktorenzahl entschieden werden. Berücksichtigt werden sollte bei einer Entscheidung über die Anzahl der Faktoren auch die sinnvolle Interpretierbarkeit des Ladungsmusters. *Faktorladungen* sind Indexzahlen, welche die Beziehung einer Variablen zum jeweiligen Faktor bezeichnen, das heißt, je höher eine Faktorladung, desto enger hängt eine Variable mit den anderen Variablen dieses Faktors zusammen. Die Faktorladungen dienen nach der Extraktion dazu, die Faktoren *inhaltlich zu interpretieren*. Die inhaltliche Interpretation wird erleichtert, wenn Variablen jeweils nur auf einem Faktor hoch laden, auf anderen Faktoren dagegen nicht. Ein solches Ladungsmuster wird als Einfachstruktur bezeichnet (Thurstone, 1945). Um eine solche Einfachstruktur zu erreichen, bedient man sich zusätzliche der *Faktorenrotation*. Die Interpretation der Bedeutung der Faktoren erfolgt dann danach, welche der Variablen auf ihnen laden und wie hoch die Faktorladungen sind. Bei der Rotation können auch Korrelationen zwischen den Faktoren zugelassen werden, indem *oblique Rotationsverfahren* gewählt werden. Diese schiefwinkligen Rotationsverfahren sind inhaltlich oft angemessener, wenn davon ausgegangen werden muss, dass die Faktoren Konstrukte repräsentieren, die in Wirklichkeit korreliert sind. In der vorliegenden Arbeit wird auf Grund der inhaltlichen Fragestellungen von der Korreliertheit der Faktoren ausgegangen,

da anzunehmen ist, dass sprachliche Kompetenzen sich gegenseitig bedingen und beeinflussen.

Insgesamt lassen sich die Methoden der Faktorenanalysen als ein heuristisches, hypothesengenerierendes Verfahren bezeichnen. Sie dienen darüber hinaus auch zur Überprüfung der Dimensionalität komplexer Merkmale (Bortz, 2004). Zu diesem Zweck werden sie, wie bereits dargestellt, in vielen Studien zur Dimensionalität von Sprachkompetenz verwendet. Aufgrund ihrer methodischen Begrenztheit wurden sie jedoch häufig kritisiert. Die exploratorischen Faktorenanalysen führen zu Ergebnissen, bei denen der erste Faktor den weitaus größten Anteil an Varianz erklärt, was in vielen empirischen Studien zur Sprachkompetenz oft fälschlicherweise als Beleg für einen Generalfaktor interpretiert wurde (Stricker et al., 2005). Auch sind die gefundenen Modelle nicht überprüfbar, da die Methoden der exploratorischen Faktorenanalysen keine Prüfung der Modellgüte erlauben. Dies ist im Rahmen der konfirmatorischen Faktorenanalysen und darüber hinaus mit Hilfe von Strukturgleichungsmodellen möglich, auf die im folgenden Abschnitt eingegangen wird.

## 5.2 Konfirmatorische Faktorenanalysen und Strukturgleichungsmodelle

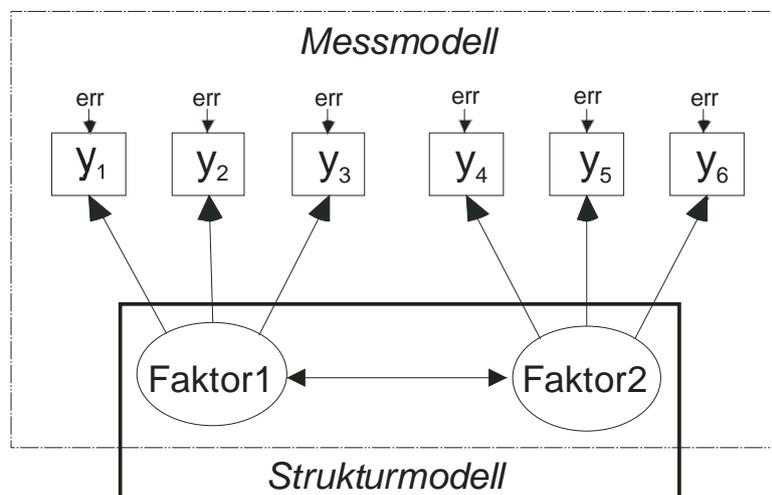
Strukturgleichungsmodelle sind ein inzwischen häufig angewendetes methodisches Standardverfahren, bei dem Methoden der Faktorenanalysen, Pfadanalysen und Regressionsanalysen verbunden werden. Sie werden oft dazu verwendet, komplexe theoretische Konstrukte auf der Grundlage empirischer Daten zu analysieren (Eid, 1999; Kaplan, 2000). Strukturgleichungsmodelle überprüfen die angenommenen Beziehungen zwischen gemessenen Variablen – den *manifesten Indikatoren* – und nicht direkt beobachtbaren Faktoren – den *latenten Variablen*. Darüber hinaus können mit Strukturgleichungsmodellen auch Beziehungen zwischen verschiedenen latenten Variablen analysiert werden (Bollen & Long, 1993; Muthén, 2002).

Ein Strukturgleichungsmodell kann je nach Anzahl der gemessenen manifesten Indikatoren und der angenommenen latenten Variablen sehr komplex werden. Das zu prüfende Modell ergibt sich normalerweise auf Grund von theoretischen Annahmen oder aus ersten bereits vorhandenen Analysen, beispielsweise aus exploratorischen Faktorenanalysen (Jöreskog, 1993). Ein komplettes Strukturgleichungsmodell besteht aus dem *Messmodell* und dem *Strukturmodell*. Im *Messmodell* werden die Beziehungen zwischen den angenommenen latenten Variablen und den ihnen jeweils zugeordneten manifesten Indikatoren beschrieben. Die Variation der Indikatoren wird im Messmodell in zwei Teile zerlegt, nämlich in die durch die latenten Variablen erklärte Varianz und in die durch das Modell nicht aufgeklärte spezifische oder Fehlervarianz. Diese Fehlerterme der Indikatoren sind in der Regel sowohl voneinander als auch von den latenten Faktoren unabhängig. Bei der in der Analysesoftware LISREL (Jöreskog & Sörbom, 1996) und verschiedenen Lehrbüchern verwendeten Notation für Strukturgleichungsmodelle wird zwischen einem Messmodell für *exogene* und *endogene* Variablen unterschieden. Diese Unterscheidung bezieht sich darauf, ob die latenten Variablen im Modell durch gerichtete Effekte erklärt werden (endogene Variablen) oder eben nicht (exogene Variablen) (Backhaus, Erichson, Plinke & Weiber, 2003). Diese Unterscheidung wird in der Notation von Mplus (Muthén & Muthén, 2004) nicht verwendet und in der vorliegenden Arbeit dementsprechend auch nicht vorgenommen.

Im *Strukturmodell* werden die Beziehungen der latenten Variablen untereinander beschrieben. Zwischen den latenten Variablen können gerichtete und ungerichtete Beziehungen angenommen werden. Wie in der Pfadanalyse können auch im Strukturmodell komplexe

Zusammenhänge mit direkten und indirekten Effekten zwischen den latenten Variablen formuliert werden. Die im Strukturmodell geschätzten Zusammenhänge zwischen den latenten Variablen werden als messfehlerfrei betrachtet (Eid, 1999; Skrandal & Rabe-Hesketh, 2004). Wenn das Strukturmodell nur ungerichtete Zusammenhänge zwischen den latenten Variablen enthält, wird auch von einer *konfirmatorischen Faktorenanalyse* gesprochen (Backhaus et al., 2003). In diesem Fall werden nur die spezifischen Ladungsmuster der Indikatoren, nicht aber die Zusammenhänge zwischen den latenten Variablen überprüft.

Zur besseren Darstellung und Interpretierbarkeit werden Strukturgleichungsmodelle oft graphisch in Pfaddiagrammen dargestellt, um die Beziehungen zwischen den manifesten und den latenten Variablen besser veranschaulichen zu können. Beobachtete Variablen werden dabei in rechteckigen Kästchen dargestellt, latente Faktoren durch Kreise oder Ovale. Pfeile repräsentieren die Beziehungen zwischen den manifesten und den latenten Variablen sowie zwischen den latenten Variablen untereinander. In Abbildung 5 ist ein komplettes Strukturgleichungsmodell schematisch dargestellt.



**Abbildung 5:** Schematische Darstellung eines Strukturgleichungsmodells mit zwei interkorrelierten, latenten Faktoren und sechs meßfehlerbehafteten, manifesten Indikatoren  $y_1$  bis  $y_6$ .

Diese schematische Abbildung zeigt zwei latente Faktoren im Strukturmodell, die miteinander korreliert sind. Beide Faktoren werden als latente Variable für jeweils drei Indikatoren angesehen, dementsprechend wird im Messmodell ein gerichteter Zusammenhang zwischen ihnen angenommen. Die Indikatoren sind in diesem Beispiel – im Gegensatz zur

exploratorischen Faktorenanalyse – eindeutig einem Faktor zugeordnet. Die manifesten Indikatoren sind jeweils mit einem eigenen Fehlerterm versehen, während die latenten Faktoren als messfehlerfrei angenommen werden.

Eine solche schematische Darstellung hilft bei der Formulierung der Hypothesen über die Zusammenhangsmuster und bei der Umsetzung des Modells in der entsprechenden Analysesoftware. Wenn das angenommene Modell spezifiziert ist, erfolgt die Schätzung der Parameter des Modells auf der Grundlage der empirischen Datenbasis. Die empirische Analyse des Modells verfolgt dabei zwei Zielsetzungen: Zum einen werden die Modellparameter in den Messmodellen und im Strukturmodell, das heißt die Faktorladungen, die Varianzen und Kovarianzen der Faktoren und die Residualvarianz der manifesten Indikatoren, geschätzt. Zum anderen wird die Modellanpassung bestimmt, das heißt es wird geschätzt, wie gut das theoretische Modell die empirisch vorhandenen Kovarianzen in den Daten abbilden kann.

Die Kovarianzen der empirischen Datenbasis sind die Grundlage für die Parameterschätzung (Bollen, 1989). Bei der Spezifizierung des theoretischen Modells wird nur ein Teil der angenommenen Beziehungen explizit festgelegt. Die freien Parameter des Modells werden iterativ so geschätzt, dass sich eine möglichst hohe Übereinstimmung zwischen den angenommenen und den beobachteten Kovarianzen ergibt. Für die Parameterschätzung stehen unterschiedliche iterative Schätzverfahren zur Verfügung, für die vorliegende Arbeit ist die Maximum-Likelihood-Methode (ML) von Bedeutung. Hierbei sollen die Parameter so geschätzt werden, dass die Wahrscheinlichkeit maximiert wird, dass die empirischen Daten aus einer Population stammen, in der das Modell gilt. Praktisch bedeutet dies, dass die Differenz zwischen der modelltheoretischen Varianz-Kovarianz-Matrix und der empirischen Varianz-Kovarianz-Matrix der Stichprobe minimiert wird. Nach diesem Kriterium werden alle Parameter geschätzt (Muthén, 1989, 1990, 1994; Yuan & Hayashi, 2005). Wenn eine Optimierung der Modellparameter nach dem ML-Kriterium vorgenommen ist, kann die verbleibende Abweichung zwischen modelltheoretischer und empirischer Kovarianzmatrix dazu herangezogen werden, die Güte der Anpassung des Modells an die empirischen Daten einzuschätzen.

Zur *Einschätzung der Modellanpassung* stehen verschiedene Gütekriterien zur Verfügung. Über die Modellgüte von Strukturgleichungsmodellen sollte nicht pauschal anhand nur eines

Wertes entschieden werden (Schermelleh-Engel & Moosbrugger, 2002). Vielmehr existiert eine Vielzahl von globalen und spezifischen Gütekriterien, anhand derer bestimmt werden kann, wie gut die empirische Kovarianzmatrix mit der modelltheoretischen übereinstimmt. Modellgütekriterien bestimmen die Gesamtgüte, dienen aber auch zum Vergleich zwischen verschiedenen Modellen, wobei die Komplexität des Modells berücksichtigt werden kann (Schumacker & Lomax, 1996). Ein gängiges Gütekriterium, das die Anpassung des Modells an die beobachtete Datenstruktur beurteilt, ist der Chi-Quadrat-Wert (Bearden, Sharma & Teel, 1982). Er resultiert aus einem Chi-Quadrat-Anpassungstest, der die Nullhypothese: „die empirische Kovarianzmatrix entspricht der modelltheoretischen“ gegen die Alternativhypothese: „die empirische Kovarianzmatrix entspricht einer beliebigen“ prüft. Um ein theoretisches Modell beizubehalten, sollte der Chi-Quadrat-Wert im Verhältnis zu den Freiheitsgraden möglichst klein sein, in der Regel kleiner als 2.5 (Homburg & Baumgartner, 1995). Da der Chi-Quadrat-Wert jedoch von der Stichprobengröße abhängt, ist er bei großen Stichproben nur mit Vorsicht zu verwenden. Als Index für die Gesamtgüte weist er auch dann hohe Werte auf, indiziert also ein unpassendes Modell, wenn nur Teile eines komplexen Modells von der empirischen Struktur abweichen (Hox, 2002).

Im Vergleich von Modellen ist es daher immer sinnvoll, zusätzliche und verschiedene Modellgütekriterien bei der Beurteilung einzubeziehen (Tanaka, 1993). Ein häufig verwendetes Kriterium ist der Goodness of Fit Index (GFI), der die relative Menge an Varianz und Kovarianz anzeigt, die durch das angenommene Modell insgesamt erklärt werden kann. Wünschenswert ist ein GFI über 0.90, um das Modell beizubehalten (Jöreskog & Sörbom, 1989). Der Comparative Fit Index (CFI) berücksichtigt zusätzlich die Zahl der Freiheitsgrade und vergleicht das existierende Modell mit einem Nullmodell, das unkorrelierte latente Faktoren annimmt. Um ein Modell beizubehalten, sollte der CFI einen Wert größer als 0.90 aufweisen, was besagt, dass mehr als 90 Prozent der Varianz in den Daten durch das Modell erklärt werden können (Bentler, 1990). Der CFI wird besonders für den Vergleich alternativer Modelle genutzt, ebenso wie das Akaike-Informations-Kriterium (AIC). Das AIC berücksichtigt die Komplexität des Modells, je niedriger das AIC, desto besser ist das Modell. Steigt der AIC-Wert während der Modifikation von Modellen wieder an, so bildet das Modell mit dem niedrigsten AIC die empirischen Daten am besten ab (Burnham & Anderson, 2004).

Mit Hilfe der hier dargestellten Modellgütekriterien kann beurteilt werden, wie gut sich das vorhandene Modell dem wahren Modell annähert, wie gering also der Fehler in der

Modellannahme ist (vgl. Tabelle 3). Mit dem Root Mean Square Error of Approximation (RMSEA) kann geprüft werden, ob das angenommene Modell die Realität hinreichend gut approximiert; je besser die Annäherung an die wahre Datenstruktur, desto geringer ist der RMSEA, angestrebt wird ein Wert von  $\leq 0.05$  (Browne & Cudeck, 1992; Hu & Bentler, 1999). Zur Beurteilung der Modellgüte von Strukturgleichungsmodellen empfiehlt Hox (1995) die gemeinsame Berücksichtigung von Chi-Quadrat/df, CFI, RMSEA und des GFI. Beim Vergleich von Mehrebenenmodellen ist zusätzlich die separate Varianzaufklärung für die Between-Ebene ( $SRMR_B$ ) und die Within-Ebene ( $SRMR_W$ ) zu berücksichtigen (Heck, 2001).

**Tabelle 3: Beurteilung der Güte von Modellen mit latenten Variablen anhand ausgewählter Gütekriterien**

Gütekriterium	Akzeptable Modellanpassung	Gute Modellanpassung
$\chi^2 / df$ (Chi-Quadrat-Prüfstatistik geteilt durch Freiheitsgrade)	$\leq 3$	$\leq 2$
SRMR (Standardized Root Mean Square Residual)	$\leq 0.08$	$\leq 0.05$
RMSEA (Root Mean Square Error of Approximation)	$\leq 0.08$	$\leq 0.05$
GFI (Goodness of Fit Index)	$\geq 0.90$	$\geq 0.95$
NNFI (Non-Normed Fit Index)	$\geq 0.90$	$\geq 0.95$
CFI (Comparative Fit Index)	$\geq 0.90$	$\geq 0.95$

In der vorliegenden Arbeit dienen Strukturgleichungsmodelle dazu, die Zusammenhänge zwischen sprachlichen Teilkompetenzen zu analysieren, dieses Vorgehen entspricht den Analysen in den meisten aktuellen Studien zur Struktur von Sprachkompetenz (Chalhoub-Deville, 1997; Gardner, Lalonde & Pierson, 1983; Kunnan, 1994). So soll versucht werden, die in Abschnitt 2.2 dargestellten vorhandenen empirischen Befunde in Teilen zu replizieren und darüber hinausgehend die spezifischen Fragestellungen dieser Arbeit zu beantworten. Die angeführten Modellgütekriterien werden verwendet, um zu bestimmen, welche Modelle die Struktur der empirischen Daten im Vergleich am besten abbilden können. Die Fragestellungen dieser Arbeit gehen jedoch teilweise über das hinaus, was mit einfachen Strukturgleichungsmodellen zu überprüfen ist. Analysiert werden soll die Struktur von Sprachkompetenz unter Berücksichtigung der hierarchischen Clusterung von Daten und unter Einbeziehung von Gruppenvergleichen. Zur Analyse dieser Fragestellungen sind daher Methoden der Mehrebenenanalyse zu verwenden, die im folgenden Abschnitt dargestellt werden.

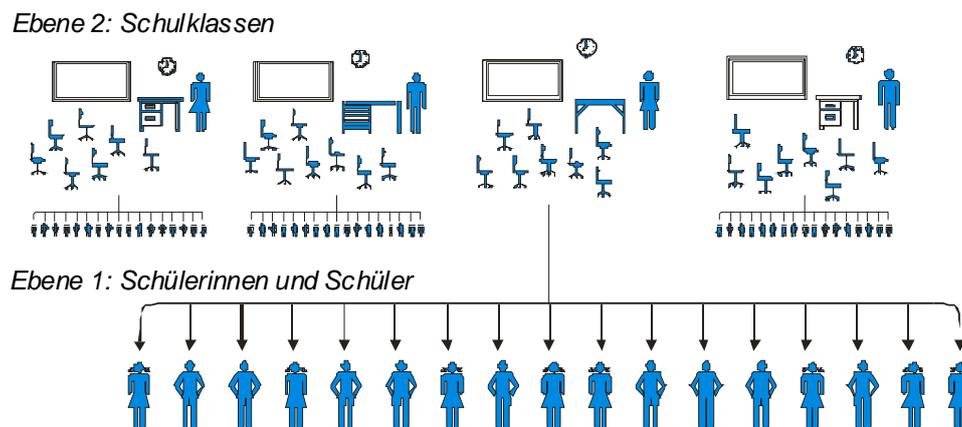
## 5.3 Grundlagen der Mehrebenen-Analyse

### 5.3.1 Definition hierarchisch geschachtelter Datensätze

Hierarchisch geschachtelte Datensätze weisen eine spezifische Struktur auf, eine solche Datenbasis ist in mehrere hierarchisch geordnete Ebenen unterteilbar. Auf jeder Ebene existieren beobachtbare, klar definierte Messeinheiten. Jede Messeinheit einer niedrigeren Ebene ist eindeutig einer Einheit auf der jeweils höheren Ebene zugeordnet (Raudenbush & Bryk, 2002). Solche hierarchischen Datensätze können durch die Art der Stichprobenziehung verursacht werden, fallen aber auch dann an, wenn sich Einheiten verschiedener Ebenen einander natürlich zuordnen lassen. Beispiele hierfür sind Tiere in Herden, Bäume in Wäldern oder Personen in Institutionen (Hox & Bechger, 1998). In solchen Fällen sind einzelne Messeinheiten innerhalb einer Gruppe *nicht unabhängig von einander*. Dadurch können sich Beobachtungen in einer Gruppe untereinander stärker ähneln als Beobachtungen zwischen verschiedenen Gruppen.

Besonders in den Human- und Sozialwissenschaften fallen hierarchische Datensätze an und können dann genutzt werden, wenn sich Forschungsfragestellungen nicht nur auf ein Individuum, sondern auf Personen in sozialen Kontexten beziehen (Ditton, 1998). Auf jeder Ebene eines solchen hierarchischen Systems sind potenziell spezifische Einflussfaktoren zu vermuten, die sich auf die individuelle Ebene auswirken können. Auch viele Fragestellungen der Bildungsforschung beziehen sich auf Effekte, die sich nicht nur mit dem einzelnen Menschen, sondern mit Menschen als Teil eines Systems beschäftigen. Dazu gehört die Schulleistungsforschung, welche zum Beispiel die Leistungen von *Schülerinnen und Schülern* auf Individuumsebene analysiert, dabei jedoch Effekte des *Klassenkontextes* als hierarchisch übergeordneter Ebene ebenso berücksichtigen muss wie potenzielle Zusammenhänge zwischen individueller Leistung und dem institutionellen Rahmen der *Schule* oder dem Bildungssystem eines *Landes* (Rowe & Hill, 1998). In solchen Fällen ist es inhaltlich wichtig, die hierarchische Datenstruktur zu berücksichtigen, um mögliche Einflussfaktoren ihrer jeweiligen Wirkungsebene zuordnen zu können. So ist davon auszugehen, dass die Leistungen von Schülerinnen und Schülern nicht nur durch individuelle Merkmale, sondern auch durch die Zugehörigkeit zu einer bestimmten Schulklasse als übergeordneter Ebene beeinflusst werden (Jerusalem, 1997; Lundberg & Rosen, 1995). Es könnte beispielsweise angenommen werden, dass der gemeinsame Unterricht bei der gleichen Lehrperson die

individuellen Leistungen aller Schülerinnen und Schüler in einer Klasse in ähnlicher Weise beeinflusst (Helmke & Weinert, 1997). Als eine zusätzliche, hierarchisch übergeordnete Ebene kann die Zugehörigkeit der Schulklasse zu einer eindeutig zuzuordnenden Schule betrachtet werden. Eine Schule in einem sozialen Brennpunkt könnte spezielle Einflussfaktoren auf Schulebene aufweisen, die in anderen Schulen unter Umständen nicht zu finden sind. Abbildung 6 stellt schematisch die hierarchische Schachtelung von Daten am Beispiel von Schülerinnen und Schülern in Schulklassen dar.

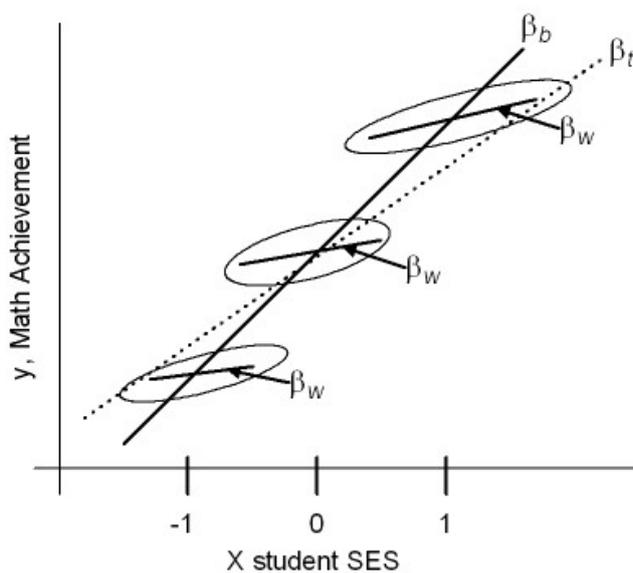


**Abbildung 6:** Schematische Darstellung einer Mehrebenenstruktur eines Datensatzes aus der Bildungsforschung, mit Schülerinnen und Schülern auf Ebene 1, die eindeutig einer Klasse auf Ebene 2 zugeordnet sind

Welche Wirkungsfaktoren auf welchen Ebenen in die Analysen einbezogen werden, hängt von den spezifischen Forschungsfragestellungen und der vorhandenen Datenlage ab (Schneider & Helmke, 1986). So können in der Schulforschung neben Klassen und Schulen zusätzlich der Bildungsgang, eine regionale politische Ebene oder ein Staat in Mehrebenenanalysen berücksichtigt werden, sofern ausreichende Daten vorliegen und begründet werden kann, warum spezifische Effekte auf diesen Ebenen vermutet werden.

Die hierarchische Schachtelung des Datensatzes ist bei der Analyse jedoch auch und vor allem zu berücksichtigen, um methodische Artefakte und damit Fehlschlüsse zu vermeiden (Lüdtke, Robitzsch & Köller, 2002). Denn die Effekte, die eine solche Clusterung von Daten mit sich bringt, sind nicht zu unterschätzen. Bei der Verwendung klassischer Analysemethoden, welche die spezifischen Effekte der Clusterung nicht berücksichtigen können, besteht die Gefahr, statistische Artefakte zu erzeugen oder Zusammenhänge fehlerhaft zu interpretieren (Cronbach, 1976). Statistische Standardverfahren sind nicht robust

gegenüber der Verletzung der Unabhängigkeitsannahme der Daten, welche in hierarchisch geschichteten Datensätzen vorliegt. Vielmehr wird bei der Bestimmung von Standardfehlern und Signifikanzen von homogenen Zufallsstichproben ausgegangen; dies ist bei Daten aus Mehrebenenstrukturen, bei denen Klassen oder Schulen und nicht Individuen gezogen werden – sogenannten Klumpenstichproben –, jedoch nicht der Fall. Besonders wichtig ist die Beachtung von geclusterten Datensätzen dann, wenn inhaltliche Schlussfolgerungen aus Zusammenhangsanalysen gezogen werden, deren Ergebnisse bei Beachtung der verschiedenen Ebenen durchaus anders ausfallen könnten (vgl. Abbildung 7).



**Abbildung 7: Zusammenhang zwischen X und Y: Zusammenhang innerhalb der Gruppen  $\beta_w$ , zwischen den Gruppen  $\beta_b$  und in der Gesamtstichprobe  $\beta_t$ , unter Vernachlässigung der hierarchischen Schachtelung der Daten (Quelle: Raudenbush & Bryk, 2002, S. 137)**

Die Abbildung stellt eine fiktive Stichprobe von drei Schulen dar, die jeweils eine niedrige Leistung und einen niedrigen sozioökonomischen Hintergrund (SES), eine mittlere Leistung und einen mittleren SES sowie eine hohe Leistung bei hohem SES aufweisen. Die gestrichelte Linie  $\beta_t$  symbolisiert den einfachen Regressionszusammenhang unter Vernachlässigung der hierarchischen Schachtelung der Daten. Sie lässt sich zerlegen in die Regressionsgerade innerhalb der Schulen  $\beta_w$  und zwischen den Schulen  $\beta_b$ . Diese zeigen sichtlich andere Zusammenhänge auf: Regressionsgerade  $\beta_w$  stellt den eher niedrigen Zusammenhang zwischen Leistung  $y$  und SES  $x$  innerhalb der Schulen dar, die Regressionsgerade  $\beta_b$  den

Zusammenhang zwischen mittlerer Leistung und mittlerem SES über die Schulen hinweg. (Raudenbush & Bryk, 2002).

Die Messungen innerhalb einer Gruppe können in solchen Fällen also nicht als unabhängig voneinander angesehen werden. Das bedeutet, dass sich Beobachtungen innerhalb einer Gruppe ähnlicher sein können als die Beobachtungen zwischen den Gruppen. Werden solche Daten wie unabhängige Daten auf Individualebene behandelt, besteht die Gefahr von Verzerrungen (Muthén, 1991). Je ähnlicher sich Individuen einer Gruppe sind, desto homogener also die Ausprägung eines Merkmals innerhalb einer Gruppe. Mit üblichen statistischen Verfahren werden die Standardfehler stark unterschätzt, Modellparameter dadurch eher signifikant (Heck, 2001; Stevens, 1996). Während bei der Parameterschätzung in klassischen Modellen von zufälligen Fehlerparametern ausgegangen wird, die unabhängig voneinander und normalverteilt sind, wird diese Annahme in hierarchisch strukturierten Datensätzen verletzt (Bryk & Raudenbush, 1992). Der Grund hierfür liegt in der Tatsache, dass die Fehlerkomponenten in Mehrebenenstrukturen innerhalb jeder Einheit miteinander korreliert sind, weil sie nicht durch Zufallseinflüsse, sondern durch systematische, gruppenspezifische Merkmale zustande kommen. Wird diese Struktur nicht beachtet, mischen sich in den Fehlerkomponenten auf Individualebene die Fehlereinflüsse aller übergeordneter Ebenen (Kreft & de Leeuw, 1998).

Um solche hierarchisch strukturierte Daten trotzdem mit klassischen Verfahren analysieren zu können, werden oft vor Beginn der Analysen die Indikatoren von unterschiedlichen Ebenen ineinander überführt. Dies geschieht durch die *Disaggregation* aller hierarchisch übergeordneten Daten auf die Individualebene oder durch die *Aggregation* aller Daten einer niedrigeren Ebene auf eine gemeinsame höhere Ebene (Haertel & Wiley, 1986). Ergebnisse, die man aus aggregierten Daten gewinnt, dürfen jedoch nicht als Zusammenhänge auf Individualebene interpretiert werden. Ein bekanntes Beispiel für Fehlschlüsse auf Grund der Nichtbeachtung hierarchischer Schachtelung ist der sogenannte *ökologische Trugschluss* (Robinson, 1950).

Um sowohl methodische Artefakte zu vermeiden als auch möglichst vielschichtige Informationen über die Einflussfaktoren auf unterschiedlichen Ebenen analysieren zu können, existieren verschiedene Vorgehensweisen. Eine Möglichkeit zum Umgang mit hierarchisch geschachtelten Datensätzen ist eine Korrektur der Standardfehler mit geeigneten

Schätzverfahren, eine weitere ist das Verwenden von Methoden, welche die Struktur der Daten explizit berücksichtigen. Diese speziellen Verfahren berücksichtigen die hierarchische Clusterung von Datensätzen, beziehen simultan Indikatoren mehrerer Ebenen in die Analysen ein und werden unter dem Begriff der *Mehrebenenanalysen* zusammengefasst (Hox, 2002). Auf die Grundlagen und Vorgehensweisen der Mehrebenenanalyse wird im folgenden Abschnitt eingegangen.

### **5.3.2 Methodische Grundlagen der Mehrebenenanalyse**

Für Verfahren, die auf Modelle der Mehrebenenanalyse zurückgreifen, existieren verschiedene Bezeichnungen: In der Soziologie meist als „multilevel linear models“ zu finden, sind auch „mixed-effect models“, „random-effect models“ (Biometrik), „random-coefficient regression models“ (Volkswirtschaftslehre) oder „covariance component models“ (Statistik) präsent (Bryk & Raudenbush, 1992). In den unterschiedlichen Forschungsrichtungen haben sich jeweils eigene Softwareanwendungen durchgesetzt. Zur Analyse von Regressionszusammenhängen zwischen unterschiedlichen Ebenen existiert zum Beispiel das Programm Hierarchical Linear Modeling (HLM) (Raudenbush, Bryk & Congdon, 2007). Daten unterschiedlicher Analyseebenen können partiell auch von der Software Linear Structural Relations (LISREL) (Jöreskog, 1993; Jöreskog & Sörbom, 1989) berücksichtigt werden. Ein Programm, welches speziell auf unterschiedliche Analysemethoden für hierarchisch geschachtelte Datensätze ausgerichtet ist und mehrere Ebenen simultan in die Analysen einbezieht, ist Mplus (Muthén & Muthén, 2004). Mit dieser Softwareanwendung können nicht nur unterschiedliche Modelle der Pfadanalysen, Mehrebenenregressionen und Strukturgleichungsmodelle auf mehreren Ebenen analysiert werden, es verfügt auch über eigene Schätzverfahren für Modellparameter, Standardfehler und Modellgüte speziell für hierarchisch geschachtelte Datensätze. Darüber hinaus können Stichprobengewichte in die Schätzung einbezogen werden. Mplus erlaubt die Schätzung von Ein- und Mehrebenen-Strukturgleichungsmodellen auch für mehrere Gruppen mit so genannten Mehrebenen-Mehrgruppen-Modellen.

Um darüber zu entscheiden, ob eine hierarchische Datenstruktur die erwähnten Unabhängigkeitsannahmen verletzt und in den Analysen berücksichtigt werden muss, kann ein spezieller Indikator verwendet werden: Die *Intraklassenkorrelation* ( $\rho$ ) beschreibt das Ausmaß der Ähnlichkeit der abhängigen Variablen innerhalb einer Gruppe, gibt also an, wie stark der Einfluss der Clusterung auf die Individualmesswerte ist. Sie wird wie folgt berechnet:

$$\rho = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2}$$

**Formel 1: Berechnung der Intraklassenkorrelation  $\rho$ ,  $\sigma_b^2 = \text{Varianz zwischen den Gruppen}$ ,  $\sigma_w^2 = \text{Varianz innerhalb der Gruppen}$**

Der Koeffizient der Intraklassenkorrelation zeigt an, wie groß der Anteil an Varianz in der abhängigen Variable ist, der auf die Zugehörigkeit zu einer übergeordneten Gruppe zurückgeführt werden kann (Ditton, 1998). Die Intraklassenkorrelation ist dann gleich null, wenn sich die geschätzten Mittelwerte der Ebene-2-Kriterien nicht voneinander unterscheiden, und wird umso größer, je mehr diese voneinander abweichen. Eine geringe Intraklassenkorrelation, also ein geringer Varianzanteil auf höherer Ebene, würde besagen, dass keine gruppenspezifischen Effekte vorliegen. Die Mehrebenenstruktur könnte in statistischen Analysen in einem solchen Fall vernachlässigt werden. Wäre  $\rho = 1$  würde dies jedoch bedeuten, dass alle Schüler innerhalb einer Klasse den gleichen Wert in einem Kriterium aufweisen, alle Klassen sich jedoch in diesem Kriterium unterscheiden. In diesem Fall könnte die Vernachlässigung der Mehrebenenstruktur zu sehr großen methodischen Artefakten führen. Denn statistische Standardverfahren der Regressions- und Varianzanalyse gehen von unabhängigen Beobachtungen, also  $\rho = 0$  aus, die Verletzung dieser Unabhängigkeitsannahme kann zu einer Inflation des  $\alpha$ -Fehler-Niveaus und einer Überschätzung der Signifikanz von Zusammenhängen führen (Stevens, 1996). Daher ist eine Varianzanalyse, die alle zu analysierenden Ebenen einschließt, der erste Schritt, um die zu erwartenden Gruppeneffekte abzuschätzen. Als Richtwerte, ab wann die Methoden der Mehrebenenanalysen zu verwenden sind, werden Werte von  $\rho \geq .05$  (Heck, 2001), aber auch  $\rho \geq .12$  (James, 1982) genannt. Liegt die Intraklassenkorrelation unter diesen Werten, können also weniger als 5 % der Varianz in der abhängigen Variablen auf die Zugehörigkeit zu einer

Gruppe zurückgeführt werden, wird davon ausgegangen, dass auch klassische Analysen auf Individualebene zu korrekten Schätzern für Modellparameter und Standardfehler gelangen (Julian, 2001). Die Berechnung der Intraklassenkorrelation ist daher der erste Schritt in einer Mehrebenenanalyse, um über die Notwendigkeit dieser Methode zu entscheiden.

Unter der Annahme, dass die Individuen einer Population jeweils Gruppen in dieser Population zuzuordnen sind, kann jeder beobachtete Wert auf der Individualebene in zwei Komponenten zerlegt werden (Cronbach & Webb, 1979; Muthén, 1991, 1994): In eine gruppenspezifische Komponente (*Between-Komponente*) welche dem aggregierten Gruppenmittelwert entspricht und in eine individuelle Komponente (*Within-Komponente*), welche die Abweichung des Messwerts vom Gruppenmittelwert repräsentiert (vgl. Formel 2).

$$y_i = y_b + y_w$$

**Formel 2: Zerlegung des individuellen Messwerts  $y_i$  in eine Between-Komponente  $y_b$  und eine Within-Komponente  $y_w$**

Da  $y_b$  und  $y_w$  als unkorreliert angesehen werden (Hox & Maas, 2005; Searle, Casella & McCulloch, 1992), kann diese Zerlegung auch auf die Gesamtkovarianzmatrix übertragen werden, die somit in eine Kovarianzmatrix *innerhalb der Gruppen* ( $S_{pw}$ , pooled within-group) und eine Kovarianzmatrix *zwischen den Gruppen* ( $S_B$ , between-group) zerlegt werden (Hox, 1995).

$$S = S_B + S_{pw}$$

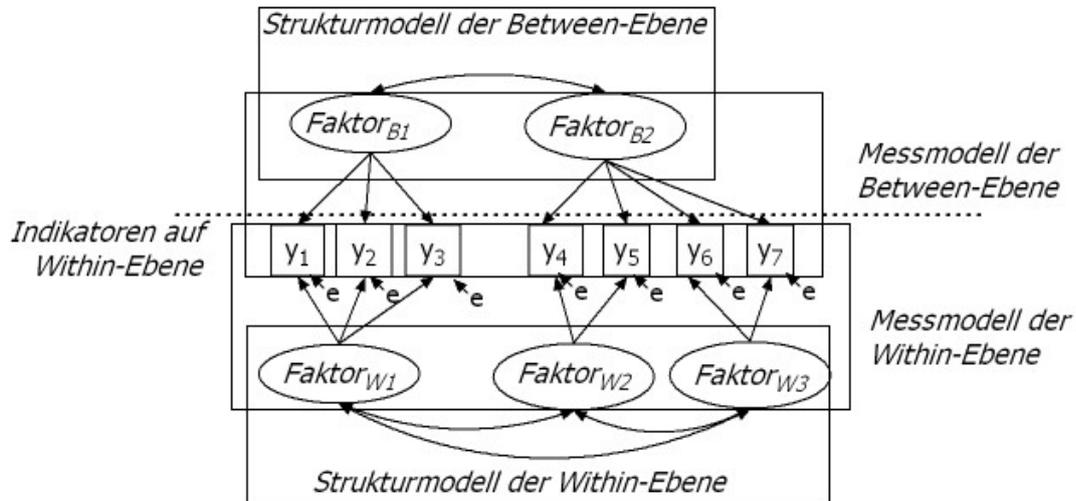
**Formel 3: Zerlegung der Gesamtkovarianzmatrix  $S$  in eine Between-Komponente  $S_B$  und eine Within-Komponente  $S_{pw}$**

Auf der Basis dieser zwei resultierenden Kovarianzmatrizen innerhalb und zwischen den Gruppen können Mehrebenenfaktorenanalysen und Mehrebenen-Strukturgleichungsmodelle modelliert werden (Muthén, 1991). Die beiden Kovarianzmatrizen können durchaus auch mit statistischen Standardverfahren berechnet und beispielsweise durch die Verwendung von Mehrgruppenmodellen verglichen werden. Allerdings erlaubt diese Vorgehensweise nicht den simultanen Einbezug beider Kovarianzmatrizen in die Schätzung von Parametern und Standardfehlern. Die Verwendung von Mplus ermöglicht es, die Variation in den abhängigen Variablen auf Individualebene in Varianzkomponenten auf jeder Ebene der hierarchischen

Struktur zu zerlegen und die Prädiktoren auf allen Ebenen simultan in die Analyse einzubeziehen (Muthén & Muthén, 1998).

Bei der Analyse von Mehrebenen-Strukturgleichungsmodellen wird in mehreren, aufeinander aufbauenden Analyseschritten vorgegangen (Cheung & Au, 2005). Im ersten Schritt wird die Gesamtkovarianzmatrix dazu verwendet, Parameter und Modellanpassung eines klassischen Strukturgleichungsmodells unter Vernachlässigung der hierarchischen Clusterung der Daten zu berechnen, um auffällige Fehlspezifikationen des Modells zu erkennen. In einem zweiten Schritt wird die Intraklassenkorrelation berechnet, um die Stärke des Einflusses der Clusterung zu beurteilen. Danach erfolgt die Schätzung der Pooled-within-Kovarianzmatrix und der Between-Kovarianzmatrix.

Die Passung des zu überprüfenden Strukturgleichungsmodells wird in einem weiteren Schritt für beide Kovarianzmatrizen *separat* analysiert. Dies ist notwendig, da davon ausgegangen wird, dass die aggregierten Variablen, die der Between-Kovarianzmatrix zugrunde liegen, eine andere Bedeutung haben können als Indikatoren auf Individualebene (Klein & Kozlowski, 2000). In Mehrebenen-Strukturgleichungsmodellen können sich die Strukturen von  $S_B$  und  $S_W$  daher durchaus unterscheiden (Härnquist, Gustafsson, Muthén & Nelson, 1994). Die Herausforderung besteht darin, im abschließenden Schritt ein gemeinsames Modell anzupassen, das die Zusammenhänge auf beiden Ebenen am besten wiedergibt. Dieses wird dann im letzten Schritt unter *simultaner Einbeziehung* beider Kovarianzmatrizen geschätzt und stellt das vollständige Mehrebenen-Strukturgleichungsmodell dar. Ein solches Mehrebenen-Strukturgleichungsmodell kann analog zu einfachen Strukturgleichungsmodellen graphisch in einem Pfaddiagramm abgebildet werden (vgl. Abbildung 8).



**Abbildung 8:** Schematische Darstellung eines Mehrebenen-Strukturgleichungsmodells mit sieben manifesten Indikatoren  $y_1$  bis  $y_7$  und den zugehörigen Messfehlern  $e$ . Auf der Within-Ebene werden drei latente Faktoren  $W_1$  bis  $W_3$  angenommen, auf der Between-Ebene die beiden latenten Faktoren  $B_1$  und  $B_2$

Für jede Ebene des Mehrebenen-Strukturgleichungsmodells werden ein Messmodell und ein Strukturmodell angepasst, wobei die manifesten, meßfehlerbehafteten Indikatoren immer auf der untersten Ebene gemessen werden. Im Falle des dargestellten Zwei-Ebenen-Modells sind für die Individualebene die Zusammenhänge zwischen den manifesten Indikatoren  $y_1$  bis  $y_7$  und den latenten Faktoren  $Faktor_{w1}$ ,  $Faktor_{w2}$  und  $Faktor_{w3}$  im Messmodell der Within-Ebene dargestellt. Die drei beobachteten Variablen  $y_1$  bis  $y_3$  dienen als Indikatoren für  $Faktor_{w1}$ , während auf  $Faktor_{w2}$  und  $Faktor_{w3}$  jeweils zwei manifeste Indikatoren laden. Im Strukturgleichungsmodell der Within-Ebene sind die Zusammenhänge zwischen den drei latenten Faktoren auf Individualebene abgebildet, es werden in diesem Beispiel ungerichtete Zusammenhänge angenommen. Auf der aggregierten Ebene dienen die manifesten Variablen  $y_1$  bis  $y_7$  ebenfalls als Indikatoren für die beiden latenten Faktoren  $Faktor_{B1}$  und  $Faktor_{B2}$  im Messmodell der Between-Ebene, deren ungerichteter Zusammenhang im Strukturgleichungsmodell der Between-Ebene dargestellt ist. Deutlich wird bei dieser schematischen Darstellung, dass sich die strukturellen Zusammenhänge zwischen den Ebenen durchaus unterscheiden können. Besonders auf aggregierter Ebene kommt es oft zu hohen Interkorrelationen der latenten Faktoren, sodass Modelle auf dieser Ebene in vielen Fällen eine geringere Dimensionalität aufweisen. Anhand der in Abschnitt 5.2 dargestellten

Modellgütekriterien kann über die Passung und mögliche Modifikationen des Modells entschieden werden.

Wie dargestellt wurde, ist die Anpassung von Mehrebenen-Strukturgleichungsmodellen an die empirischen Daten ein mehrstufiger Prozess, der im Idealfall auf theoretischen Vorannahmen über die Dimensionalität der Daten basiert, der also eine hypothesenprüfende Vorgehensweise unterstützt. Dieses rein *konfirmatorische Vorgehen* (Jöreskog, 1993), bei dem ein spezielles Modell bestätigt oder verworfen wird, soll in der vorliegenden Arbeit durch den *Vergleich alternativer Modelle* ergänzt werden. Diese alternativen Modelle, die unterschiedliche Zusammenhänge von sprachlichen Teilkompetenzen postulieren, werden dabei anhand ihrer empirischen Modellpassung verglichen. Darüber hinausgehend ist es durchaus zulässig und sinnvoll, auf dieser Basis im Weiteren zu einem *modellgenerierenden Vorgehen* zu gelangen, bei dem theoretische Annahmen und empirische Ergebnisse zu einer Optimierung des Modells genutzt werden, und so im Verlauf der Analysen die Modelle zu modifizieren.

Eine gängige Erweiterung von Strukturgleichungsmodellen auch auf mehreren Ebenen sind *Mehrgruppenmodelle*, in denen die Strukturgleichungsmodelle für unterschiedliche Gruppen von Personen simultan geschätzt werden. Dieses Vorgehen erlaubt es, zu analysieren, ob theoretisch angenommene Strukturen sich auf unterschiedliche Personengruppen generalisieren lassen, oder ob strukturelle Zusammenhänge in den Gruppen abweichen. Bedeutung erlangen solche Analysen zum einen, wenn es darum geht, theoretisch postulierte Unterschiede zwischen Gruppen zu überprüfen, zum anderen aber auch, um die Messgüte von Verfahren zu bestätigen, die für interkulturell vergleichende Studien eingesetzt werden sollen. Mehrgruppenvergleiche sind immer dann ratsam, wenn davon ausgegangen werden muss, dass eine Gruppenzugehörigkeit die Zusammenhänge in Messmodell oder Strukturmodell beeinflussen könnte (Pellegrini & Scandura, 2005; Wicherts, Dolan & Hessen, 2005).

Für solche Mehrgruppenmodelle werden anhand eines Kriteriums, beispielsweise des Geschlechts oder des sprachlichen Hintergrunds, mehrere Gruppen gebildet, denen jedes Individuum eindeutig zugeordnet wird. Geprüft wird die Hypothese, dass das theoretische Modell für alle Gruppen gilt. Im strengsten Fall werden nicht nur gleiche strukturelle Zusammenhänge, sondern auch äquivalente Modellparameter für alle Gruppen angenommen. Zum Vergleich können Modelle geschätzt werden, in denen verschiedene Parameter frei gesetzt werden. Dies können die Faktorladungen sein, die sich dann zwischen den Gruppen

unterscheiden dürfen, aber auch die Anzahl der latenten Faktoren kann zwischen den Gruppen variieren. Anhand von Modellgütekriterien kann schließlich verglichen werden, welche Modelle den empirischen Daten eher angemessen sind, ob sich also die Strukturen zwischen den Gruppen unterscheiden oder nicht.

#### **5.4 Das methodische Vorgehen in der vorliegenden Arbeit**

Die vorliegende Arbeit analysiert die Daten von insgesamt zwölf Sprachtestverfahren in zwei sprachlichen Bereichen, dem Deutschen und dem Englischen. Ziel der Analysen ist es, die Struktur von Sprachkompetenz, also die Zusammenhänge zwischen den Sprachtestverfahren untereinander, unter Rückgriff auf theoretische Modelle und empirische Ergebnisse zu überprüfen. Die dabei verwendeten methodischen Analyseschritte orientieren sich in ihrer Abfolge an den in Abschnitt 4 dargestellten Fragestellungen. Diese Fragestellungen bauen in ihrer Komplexität aufeinander auf, sodass mit zunehmender Differenziertheit der Hypothesen die verwendeten Methoden komplexer werden.

Der Ergebnisteil der Arbeit beginnt dementsprechend mit den deskriptiven Auswertungen der Stichprobendarstellung. Darauf folgen Analysen, welche die Struktur von Sprachkompetenz auf der Grundlage verschiedener theoretischer Modell konfirmatorisch überprüfen und deren empirische Passung auf der Grundlage der vorliegenden Stichprobe miteinander vergleichen. Solche konfirmatorischen Faktorenanalysen und Strukturgleichungsmodelle gehören in der pädagogisch-psychologischen Forschung zur Sprachkompetenz inzwischen zu den häufig verwendeten Standardverfahren, um Einflussfaktoren von Sprachkompetenz oder auch Zusammenhänge zwischen sprachlichen Teilkompetenzen zu untersuchen (Chalhoub-Deville, 1997; Gardner et al., 1983; Kunnan, 1994). Dieser Teil der Auswertungen dient dabei auch der Replizierung bereits bekannter Modellannahmen. Um jedoch über herkömmliche Analysemethoden hinauszugehen und im Rahmen dieser Arbeit nicht nur einen methodisch neuen Ansatz zu wählen, sondern auch zu weiterführenden Erkenntnissen über die Struktur von Sprachkompetenz zu kommen, werden darauf aufbauend die Methoden der Mehrebenenanalyse verwendet. Die Einbeziehung der hierarchischen Schachtelung des vorliegenden Datensatzes ist wie in Abschnitt 5.3.1 dargestellt wurde, unabdingbar, um der Komplexität der Beziehungen zwischen Indikatoren im Rahmen der vorliegenden Fragestellungen gerecht zu werden und die Daten auf der Grundlage aktueller methodischer Vorgehensweisen zu analysieren. Insgesamt geht die vorliegende Arbeit damit in ihrer Komplexität über bisherige empirische Ansätze zur Struktur der Sprachkompetenz hinaus.

## 6 Ergebnisse

### 6.1 Deskriptive Ergebnisse - Beschreibung der Stichprobe

Die Zielpopulation der DESI-Studie stellten alle Schülerinnen und Schüler in Deutschland dar, die sich zum Zeitpunkt der Erhebung im Schuljahr 2003/2004 in der neunten Jahrgangsstufe an allgemeinbildenden Schulen befanden. Aus der Gesamtheit aller Schülerinnen und Schüler, die dieses Kriterium erfüllten, wurde eine für Deutschland repräsentative Stichprobe gezogen. Die Datenbasis, auf der die vorliegende Arbeit basiert, umfasst Leistungswerte von insgesamt 10 543 Schülerinnen und Schülern der neunten Jahrgangsstufe aus 219 Schulen, darunter 40 Schulen mit einem bilingualen Unterrichtsangebot. Diese stehen repräsentativ für die Gesamtpopulation der Schülerinnen und Schüler der neunten Jahrgangsstufe zum Testzeitpunkt in Deutschland. Zur Gewährleistungen dieser Repräsentativität werden die Daten in den Analysen gewichtet (vgl. Abschnitt 3.3.1).

Deutlich sichtbar wird der Effekt der Gewichtung durch den Vergleich zwischen den ungewichteten und den gewichteten deskriptiven Häufigkeiten für die Stichprobencharakteristik (vgl. Tabelle 4). So sind beispielsweise die Schülerinnen und Schüler des Bildungsgangs Hauptschule in der Stichprobe unterrepräsentiert – ein Anteil von 18 % in der Stichprobe steht für ca. 27 % in der Gesamtpopulation. Bilingual unterrichtete Schülerinnen und Schüler hingegen wurden bewusst überproportional häufig gezogen, sie sind in der Stichprobe mit 12 % der getesteten Personen vertreten, in der Gesamtpopulation der deutschen Schülerinnen und Schüler machen sie jedoch nur etwa 1 % aus (Statistisches Bundesamt, 2004). Um also auf Grund der dargestellten Stichprobe repräsentative und vor allem unverzerrte Aussagen über die Dimensionalität von Sprachkompetenz machen zu können, ist die Gewichtung in alle weiteren Analysen einzubeziehen.

**Tabelle 4: Beschreibung der Stichprobe (gewichtete und ungewichtete Stichprobenkennwerte, die fehlenden zu 100 % sind fehlende Angaben)**

Strata	Gesamt		gewichtet	Schülerinnen		gewichtet	Schüler		gewichtet
<i>Bildungsgang</i>	<i>Häufigkeit</i>	%	%	<i>Häufigkeit</i>	%	%	<i>Häufigkeit</i>	%	%
Hauptschule	1898	18	27	846	15	25	1052	21	29
Realschule	3622	34	36	1919	35	37	1703	34	34
Gesamtschule	715	7	10	330	6	9	385	8	11
Gymnasium	4308	41	27	2441	44	30	1867	37	25
Bilingual unterrichtet	1233	12	1	767	14	1	466	9	1
Stichprobenbeschreibung									
	Gesamt		gewichtet	Schülerinnen		gewichtet	Schüler		gewichtet
<i>Erstsprache</i>	<i>Häufigkeit</i>	%	%	<i>Häufigkeit</i>	%	%	<i>Häufigkeit</i>	%	%
Deutsch	7959	76	74	4244	82	81	3715	82	81
Deutsch und andere	601	6	6	356	7	7	245	5	5
andere	1145	11	12	563	11	12	582	13	14
<i>Geschlecht</i>	<i>Häufigkeit</i>	%	%						
Schülerinnen	5536	52	51						
Schüler	5007	48	49						
<i>Geburtsjahr</i>	<i>Häufigkeit</i>	%	%	<i>Häufigkeit</i>	%	%	<i>Häufigkeit</i>	%	%
1984	1	-	-	0	-	-	1	-	-
1985	3	-	-	1	-	-	2	-	-
1986	177	2	2	79	1	2	91	2	2
1987	1423	14	16	615	11	13	808	16	18
1988	5253	50	50	2692	49	49	2561	51	51
1989	3659	34	33	2136	39	37	1523	31	29
1990	20	-	-	12	-	-	8	-	-

In der Stichprobe befinden sich 52,5 % Mädchen und 47,5 % Jungen im Alter zwischen 14 und 20 Jahren, wobei die Mehrzahl der Schülerinnen und Schüler zum Zeitpunkt der Testung 15 oder 16 Jahre alt war. Mit der Stichprobe wurden alle Bildungsgänge in den Ländern der Bundesrepublik Deutschland als explizite Strata abgedeckt. Hierzu gehörten die Hauptschule, die Realschule, die integrierte Gesamtschule und die Schulen mit mehreren Bildungsgängen sowie das Gymnasium. Die überrepräsentative Anzahl von Gymnasiasten in der Stichprobe resultiert aus der hohen Anzahl von Schulen mit bilinguaem Unterrichtsangebot, die gezielt überproportional gezogen wurden, um in DESI spezielle Effekte dieser Unterrichtsform analysieren zu können. Auch die Erstsprache der befragten Schülerinnen und Schüler wurde

erhoben, wobei die Mehrheit der Schülerinnen und Schüler das Deutsche als Erstsprache angab.

In DESI erfolgten die Skalierung der Testaufgaben und die Schätzungen der individuellen Kompetenzwerte auf der Grundlage der Item-Response-Theorie (vgl. Abschnitt 3.3.5). Die Personenfähigkeitswerte, welche in der vorliegenden Arbeit verwendet werden, sind Weighted Likelihood Estimates (WLEs), die für jeden der sprachlichen Bereiche auf einer separaten eindimensional skalierten Kompetenzskala geschätzt wurden (vgl. Kapitel 3.3.5). Tabelle 5 gibt einen Überblick über die Anzahl der vorhandenen Leistungswerte für jeden der gemessenen Kompetenzbereiche im Deutschen und Englischen.

**Tabelle 5: Häufigkeit der vorhandenen Leistungswerte für die Kompetenzbereiche im Deutschen und Englischen**

Deutsch	Bewusstheit	Lesekompetenz	Textproduktion Pragmatik	Textproduktion Systematik
N	9772	9767	9353	9348
	Argumentation	Wortschatz	Rechtschreibung	
N	9643	9622	9397	
Englisch	C-Test	Hörverstehen	Leseverstehen	Sprachbewusstheit
N	9744	9714	9575	9703
	Textproduktion	Mündliche Sprachproduktion		
N	8991	1885		

Die Daten der hier dargestellten Stichprobe und Testverfahren bilden die Basis für alle folgenden Analysen, welche die Zusammenhänge dieser sprachlichen Teilbereiche untereinander analog zu den Fragestellungen in Kapitel 4 analysieren. Ergebnisse zu den Leistungen von Schülerinnen und Schülern in den jeweiligen Kompetenzbereichen, die Verteilung auf einzelne Kompetenzniveaus und die kriterienorientierte Interpretationen dieser Leistungswerte sind nicht Inhalt dieser Arbeit, sie werden jedoch in den Veröffentlichungen zum DESI-Projekt ausführlich dargestellt (DESI-Konsortium, 2008).

## 6.2 Struktur der Sprachkompetenz im Deutschen und im Englischen

Sprachkompetenz wurde in DESI für das Deutsche als Unterrichtssprache und das Englische als Fremdsprache erfasst. Dabei wird der Bereich des Deutschen mit sechs Testverfahren abgedeckt, wobei für eines, den Test zur Sprachproduktion, zwei Leistungswerte zur Verfügung stehen. Auch im Bereich des Englischen bearbeiteten die Schülerinnen und Schüler sechs verschiedene Testverfahren, wobei der Test zur mündlichen Sprachproduktion jedoch nur von einer Teilstichprobe bearbeitet wurde (vgl. Kapitel 3.3.3 und Kapitel 3.3.4). Zur differenziellen Analyse der Zusammenhänge in den Einzelsprachen werden im Folgenden die Testmodule im Deutschen und im Englischen separat betrachtet. Wie in Fragestellung 1 dargestellt ist es das Ziel, für jeden der beiden Sprachbereiche Deutsch und Englisch den Binnenzusammenhang zwischen den Testmodulen als Struktur von Sprachkompetenz zu beschreiben.

**Hypothese 1:** *Die darstellbare Binnenstruktur der Sprachkompetenz weist für das Deutsche eine andere latente Faktorenstruktur auf als für das Englische.*

Zur Überprüfung dieser Hypothese werden unterschiedliche theoretische Modelle von Sprachkompetenz mit jeweils spezifischen Annahmen von Sprachstrukturen auf ihre empirische Passung hin analysiert. Dazu gehören für jeden Bereich ein Modell mit einem Generalfaktor, Modelle mit unterschiedlichen latenten Faktoren und hierarchische Modelle.

### 6.2.1 Struktur im Deutschen

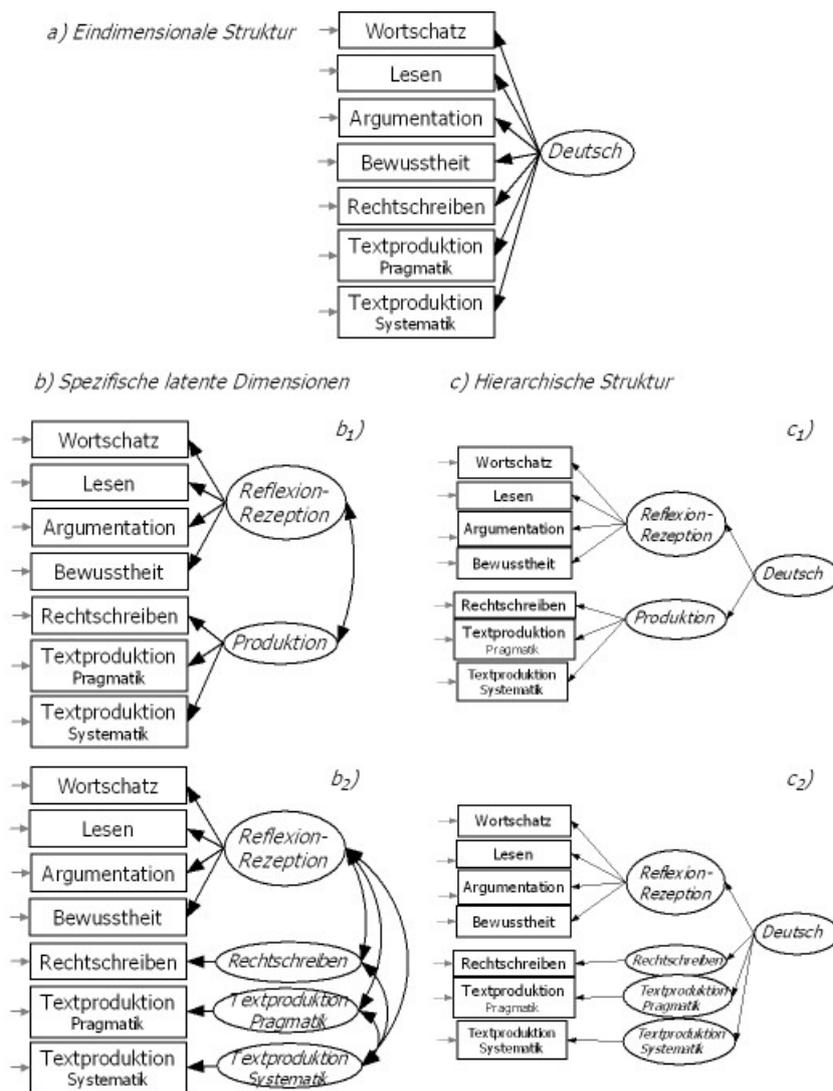
Die sprachlichen Kompetenzen im Deutschen wurden in DESI mit schriftbasierten Testverfahren für die reflexiv-rezeptiven und produktiven Bereiche der Sprachkompetenz erfasst. Für den Bereich des Deutschen können mit den vorhandenen Testdaten die in Abbildung 9 dargestellten Modelle analysiert werden:

- Modell a), das eine eindimensionale Struktur der Sprachkompetenz postuliert. In diesem in den Sprachwissenschaften prominenten Modell wird angenommen, dass die Zusammenhänge zwischen allen Testbereichen durch einen gemeinsamen Faktor erklärt werden können (vgl. Abschnitt 2.1).
- Die Modelle b<sub>1</sub>) und b<sub>2</sub>), die spezifische latente Dimensionen annehmen, basieren auf denjenigen Theorien der Sprachkompetenz, nach denen sich die sprachlichen Bereiche der *reflexiv-rezeptiven* und der *produktiven* Sprachkompetenz durch jeweils

entsprechende Testverfahren repräsentieren lassen. Mit Modell b<sub>2</sub>) wird analysiert, inwieweit der produktiven Sprachkompetenz noch weiter in die Bereiche der Sprachproduktion und der Rechtschreibung differenziert werden kann (vgl. Abschnitt 2.2.1).

- Die dritte Gruppe von Modellen unter c) stellen hierarchische Modelle dar. Sie erweitern die Modelle aus b) um jeweils einen übergeordneten Faktor, auf den die gesamtsprachliche Kompetenz zurückgeführt wird (vgl. Abschnitt 1.8).

Rein rechnerisch implizieren die Modelle b und c *die gleiche Zusammenhangsstruktur der beobachteten Variablen*: Die Zusammenhänge, die in Modell b<sub>1</sub> als Korrelation zwischen den latenten Faktoren modelliert werden, entsprechen in Modell c<sub>1</sub> dem Pfad zwischen dem latenten Faktor der Sprachproduktion und dem übergeordneten Faktor Deutsch – der Pfad zwischen dem latenten Faktor Reflexion-Rezeption und dem übergeordneten Faktor Deutsch ist auf 1 fixiert. Auch die empirisch zu schätzende Zusammenhangsstruktur in Modell c<sub>2</sub> entspricht modelltheoretisch denjenigen in den Modellen b<sub>1</sub> und c<sub>1</sub>, weil die latenten Variablen für die Bereiche Rechtschreiben, Textproduktion Pragmatik und Textproduktion Systematik jeweils nur durch einen Indikator repräsentiert sind und diese Pfade daher ebenfalls fixiert werden. Die Varianzen dieser latenten Faktoren sind dann nicht mehr frei schätzbar. Bei der Analyse der jeweils ineinander geschachtelten Modelle b) und c) werden dementsprechend *die gleichen Modellanpassungen* erreicht, es ist also nicht möglich, anhand der Modellanpassung zwischen den Modellen zu entscheiden. Der Vollständigkeit halber sind diese theoretischen Modelle an dieser Stelle jedoch mit aufgeführt.



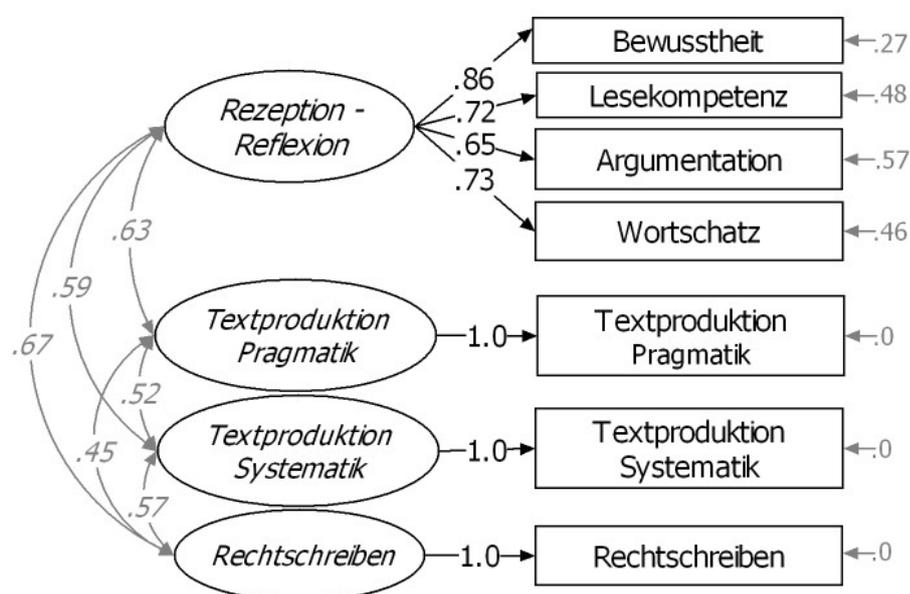
**Abbildung 9: Zu vergleichende Modelle der Sprachkompetenz im Deutschen**

Zur Analyse und zum Vergleich der verschiedenen Modelle wurden konfirmatorische Faktorenanalysen durchgeführt, explizit wurde dabei die hierarchische Schachtelung der Daten berücksichtigt (Muthén & Muthén, 2001). Diese geht in die Schätzung der Modellparameter und der Modellanpassung ein, sodass die Standardfehler unter Berücksichtigung der Datenclusterung geschätzt werden (vgl. Kapitel 5.3.2). Die resultierenden Modellanpassungen wurden anhand der in Kapitel 5.2 dargestellten Kriterien der Modellanpassungsgüte verglichen. Die Modellanpassung der verschiedenen Modelle ist in Tabelle 6 dargestellt.

**Tabelle 6: Modellanpassungen der verschiedenen konfirmatorischen Faktorenanalysen im Deutschen**

Modell	N	$\chi^2$	df	p	CFI	RMSEA	SRMR	AIC
a) Eindimensionale Struktur	9062	888.17	14	$\leq .01$	.97	.08	.04	150669.66
b <sub>1</sub> ) Zweidimensionale Struktur	9062	495.60	13	$\leq .01$	.98	.06	.03	150089.84
b <sub>2</sub> ) Vierdimensionale Struktur	9062	269.88	11	$\leq .01$	.99	.05	.02	149746.58
c <sub>1</sub> ) / c <sub>2</sub> ) hierarchische Struktur	Entspricht b <sub>1</sub> ) Zweidimensionale Struktur							

Die dargestellten Modelle lassen sich auf der Grundlage der verwendeten Modellgütekriterien vergleichen, die Modelle c<sub>1</sub> und c<sub>2</sub> weisen die gleiche Anpassung auf wie Modell b<sub>1</sub>. Die beste Erklärung der Zusammenhänge der Testverfahren für die sprachlichen Teilbereiche im Deutschen erreicht das in Abbildung 10 dargestellte Modell b<sub>2</sub> der Sprachkompetenz mit vier latenten Faktoren, es weist die beste Modellanpassung an die Daten auf.



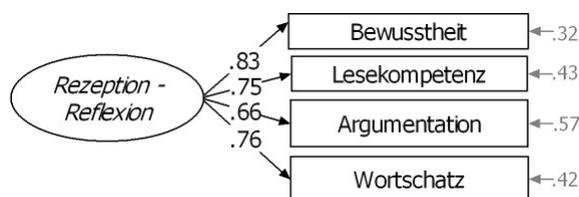
**Abbildung 10: Modell b<sub>2</sub> der Struktur der Deutschkompetenz mit dem latenten Faktor Reflexion-Rezeption und den drei manifesten Indikatoren Textproduktion Pragmatik, Textproduktion Systematik und Rechtschreiben (N=9062,  $\chi^2=269.88$ , df=11, p $\leq$ .00, CFI=.99, AIC=149746.58, RMSEA=.05, SRMR=.02)**

Im hier dargestellten Modell wird die Struktur der Sprachkompetenz im Deutschen durch vier latente Faktoren modelliert. Der erste Faktor repräsentiert die Teilbereiche der Sprachbewusstheit, des Lesens, der Argumentation und des Wortschatzes, die Faktorladungen betragen zwischen  $\lambda=.65$  und  $\lambda=.86$ . Allen diesen vier sprachlichen Bereichen liegen Fähigkeiten der rezeptiven, aber auch der reflexiven Sprachkompetenz zugrunde, daher wird dieser latente Faktor als „*reflexiv-rezeptive*“ *Sprachkompetenz* bezeichnet. Dadurch soll zum Ausdruck gebracht werden, dass bei der Bearbeitung von Testaufgaben im Lesen, in der Sprachbewusstheit, der Argumentation und des Wortschatzes nicht nur passive rezeptive Fähigkeiten, also etwa solche der Worterkennung (Lesen), benötigt werden, sondern ebenso kognitiv komplexere reflexive Fähigkeiten gefragt sind, etwa bei der Benennung von Gegenständen (Wortschatz) oder der Verwendung grammatikalisch korrekter Regeln. Obwohl jeder dieser vier Bereiche also eigene Anforderungen an die sprachliche Kompetenz stellt, lassen sich wesentliche Anteile der Leistungsvariation in den Einzeltests auf einen gemeinsamen Faktor zurückführen.

Deutlich von diesem Faktor zu unterscheiden sind hingegen die drei Indikatoren für die sprachlichen Kompetenzen der Textproduktion Pragmatik, der Textproduktion Systematik und der Rechtschreibung. Obwohl alle diese drei Teilkompetenzen als *produktive Fähigkeiten* charakterisiert werden können, zeigt das Modell  $b_1$ , das ihnen einen gemeinsamen latenten Faktor zuweist, eine schlechtere Anpassung an die Daten auf als das hier präferierte Modell  $b_2$ . Diesem liegt die Annahme zugrunde, dass diese drei produktiven Teilbereiche zu differenzieren sind, weil sie unterschiedliche Anforderungen an die Kompetenzen von Personen stellen. Anders gesagt, es wird davon ausgegangen, dass die Leistung von Personen in diesen drei Teilbereichen durchaus stark variieren kann. So beziehen sich die beiden Werte für die Beurteilung der Leistung im Test zur Textproduktion – der Wert für Pragmatik und derjenigen für Systematik – auf unterschiedliche Bereiche sprachproduktiver Leistungen. Dies lässt den Schluss zu, dass die freie Textproduktion, wie sie in DESI testmethodisch umgesetzt wurde, durchaus ein vielschichtiges Konstrukt ist, dem eine mehrdimensionale Beurteilung angemessen zu sein erscheint. Rechtschreiben als die Fähigkeit, richtig zu schreiben, ist ebenfalls als eigenständiges Kompetenzkonstrukt anzusehen, das jedoch mit reflexiv-rezeptiven Prozessen in einem starken Zusammenhang steht. Die beiden Bereiche weisen die in diesem Modell höchste latente Korrelation von  $r=.67$  auf, was sich dadurch erklären lässt, dass korrektes Schreiben nur durch ständige Sprachreflexion möglich ist, die auf einem

gefestigten Wortschatz aufbaut und im Leseprozess evaluiert wird. Richtig schreiben kann nur, wer die Schreibweise von Wörtern beim Lesen erlernt und als Wortschatz merkt. Rechtschreibung und Sprachreflexion fließen dann wiederum in den Prozess der Textproduktion ein, allerdings können sie für einen verständlichen und zielführenden freien Text nur Grundlage sein. Dies drückt sich in den latenten Korrelationen der beiden Bereiche zur Textproduktion mit den anderen Teilkompetenzen aus, die im mittleren Bereich liegen. Sie sind auch daher als eigenständige Indikatoren zu modellieren, weil bei der Gestaltung eines freien Textes nicht nur die grundlegenden Fähigkeiten der anderen Teilbereiche wesentlich sind, sondern auch textsortenspezifisches Wissen, freie kommunikative Zielsetzung und die individuelle sprachliche Ausdruckform von wesentlicher Bedeutung sind, die durch die übrigen Textverfahren nicht erfasst wurden.

Die Annahme, dass der reflexiv-rezeptive Kompetenzbereich homogener ist als der produktive, lässt sich durch ein separates Teilmodell bestärken, das lediglich die vier reflexiv-rezeptiven Testmodule der Lesekompetenz, der Sprachbewusstheit, der Argumentation und des Wortschatzes modelliert (vgl. Abbildung 11).



**Abbildung 11: Teilmodell mit latentem Faktor der reflexiv-rezeptiven Kompetenz im Deutschen (N=9062,  $\chi^2=4.35$ ,  $df=2$ ,  $p=.11$ , CFI=1.00, AIC=92683.303, RMSEA=.01, SRMR=.00)**

Das Teilmodell kann mit einem latenten Faktor die Varianz in den vier sprachlichen Teilbereichen praktisch perfekt erklären, sogar der Chi-Quadrat-Wert zeigt keine signifikante Abweichung zwischen modellimplizierter und beobachteter Struktur. Dieses Teilmodell zeigt, dass sich die gemeinsame latente Dimension der reflexiv-rezeptiven Sprachkompetenz im Deutschen separat sehr gut modellieren lässt. Bei gleichzeitiger Betrachtung des reflexiv-rezeptiven und des produktiven Teilbereichs können die Zusammenhänge in den Daten nicht im gleichen Ausmaß erklärt werden, trotzdem kann aufgrund der Modellgütekriterien zugunsten eines mehrdimensionalen Modells entschieden werden.

## 6.2.2 Struktur im Englischen

Zur Überprüfung der Zusammenhänge zwischen den Testmodulen im Englischen wird das gleiche Vorgehen gewählt wie für den Bereich des Deutschen: Mittels konfirmatorischer Faktorenanalysen unter Berücksichtigung der hierarchischen Schachtelung der Daten werden die bereits im Theorieteil der Arbeit dargestellten theoretisch fundierten, alternativen Modelle der Struktur von Sprachkompetenz im Englischen analysiert. Verglichen werden dabei a) ein eindimensionales Modell, b) Modelle mit verschiedenen latenten Faktoren und c) hierarchische Modelle, die spezifische latente Faktoren postulieren, denen ein gemeinsamer Faktor der Englischkompetenz zugrunde liegt (vgl. Abbildung 12).

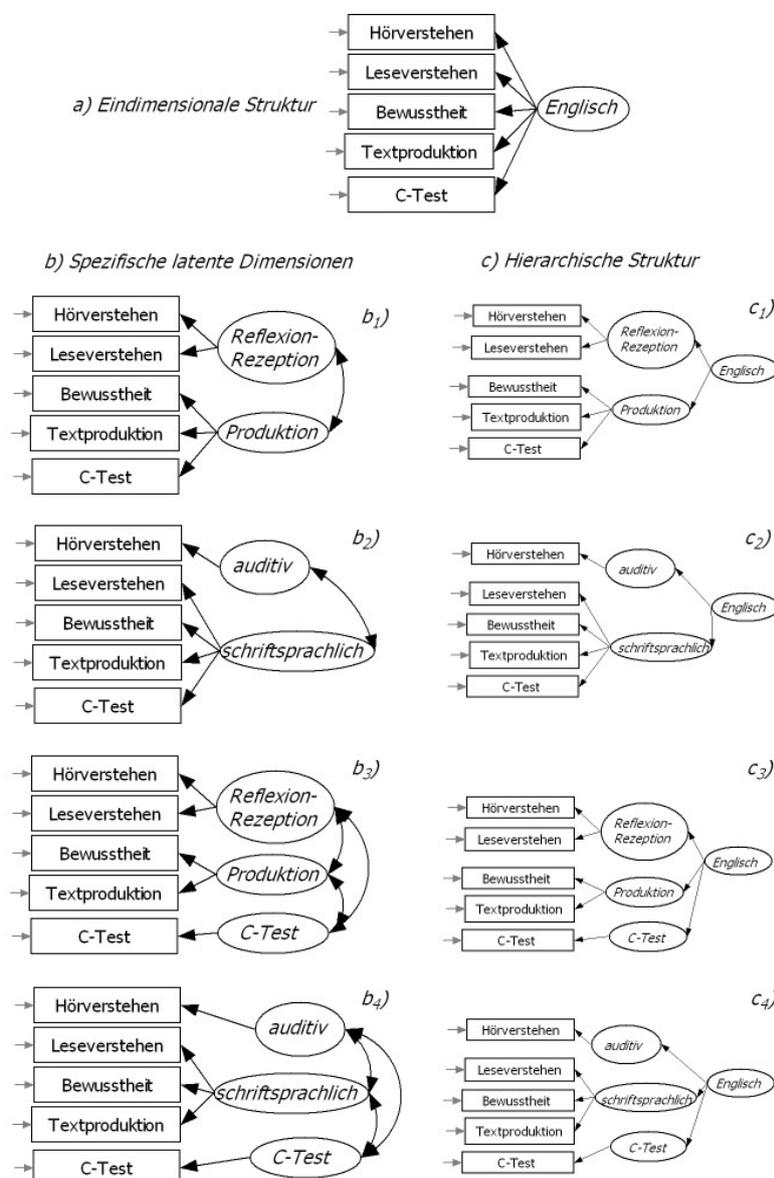


Abbildung 12: Theoretische Modelle der Sprachkompetenz im Englischen

Wie aus den Abbildungen 9 und 12 ersichtlich wird, können für die Sprachstruktur im Englischen auf der Grundlage der verwendeten Testwerte mehr theoretisch fundierte Modelle abgebildet und analysiert werden, als dies für das Deutsche möglich ist. So erfasst der DESI-Test zum Hörverstehen im Englischen die *auditive* Teilfähigkeit und hebt sich damit von den ansonsten *schriftsprachlichen* Testmodulen ab. Gleichzeitig jedoch kann der Test zum Hörverstehen entsprechend dem Modell der vier sprachlichen Teilkompetenzen auch dem *reflexiv-rezeptiven* Kompetenzbereich zugeordnet werden. Dementsprechend können die so gebildeten Modelle mit unterschiedlichen latenten Faktoren – einmal mit *auditiven* und *schriftsprachlichen* Faktoren, einmal mit *reflexiv-rezeptiven* und *produktiven* Faktoren – anhand der vorliegenden Daten analytisch gegenübergestellt werden.

Hinzuweisen ist auch an dieser Stelle jedoch darauf, dass die Modellstrukturen der Modelle b und c bedingt durch die Anzahl der vorhandenen manifesten Indikatoren statistisch gesehen gleich sind. Eine Entscheidung zwischen solchen Modellen kann hier nicht auf Grund statistischer Modellgütekriterien getroffen werden, sondern muss auf inhaltlichen Überlegungen basieren. Trotzdem sind die jeweiligen hierarchischen Modelle der Vollständigkeit halber hier dargestellt, da sie eine vorhandene theoretische Annahme über die Struktur von Sprachkompetenz darstellen. Die entsprechenden Modellanpassungen sind in Tabelle 7 dargestellt.

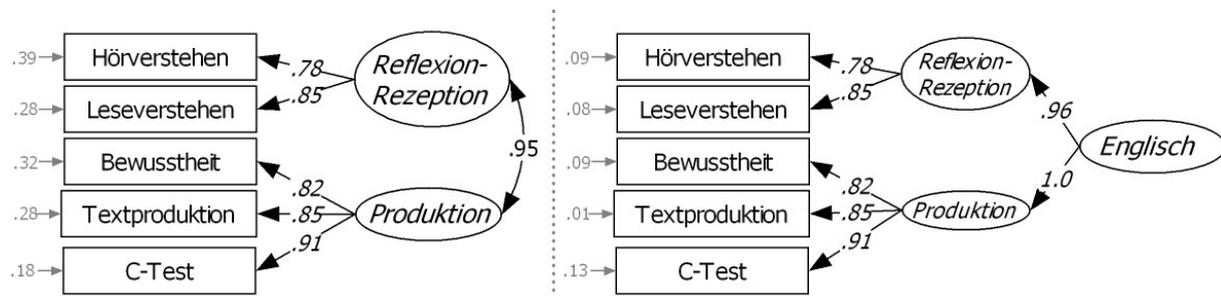
**Tabelle 7: Modellanpassung der theoretischen Modelle zur Sprachkompetenz im Englischen**

Modell	N	$\chi^2$	df	p	CFI	RMSEA	SRMR	AIC
a) Eindimensionale Struktur	8964	121.44	5	≤.01	.99	.05	.01	94469.73
b) Latente Faktoren								
b <sub>1</sub> ) Zweidimensionale Struktur reflexiv-rezeptive und produktive Kompetenzen	8964	28.04	4	≤.01	.99	.03	.01	94272.80
b <sub>2</sub> ) Zweidimensionale Struktur schriftsprachliche und auditive Kompetenzen	Entspricht a) Eindimensionale Struktur							
b <sub>3</sub> ) Dreidimensionale Struktur reflexiv-rezeptive Kompetenz, produktive Kompetenz und C-Test	8964	27.93	3	≤.01	.99	.03	.01	94274.46
b <sub>4</sub> ) Dreidimensionale Struktur schriftsprachliche und auditive Kompetenzen und C-Test	8964	94.86	4	≤.01	.99	.05	.01	94426.56
c) hierarchische Modelle								
c <sub>1</sub> ) Zweidimensionale Struktur reflexiv-rezeptive und produktive Kompetenzen	Entspricht b <sub>1</sub> ) Zweidimensionale Struktur reflexiv-rezeptive und produktive Kompetenzen							
c <sub>2</sub> ) Zweidimensionale Struktur schriftsprachliche und auditive Kompetenzen	Entspricht a) Eindimensionale Struktur							
c <sub>3</sub> ) Dreidimensionale Struktur reflexiv-rezeptive Kompetenz, produktive Kompetenz und C-Test	8964	119.10	4	≤.01	.99	.06	.01	94470.78
c <sub>4</sub> ) Dreidimensionale Struktur schriftsprachliche und auditive Kompetenzen und C-Test	Entspricht a) Eindimensionale Struktur							

Für die Beurteilung der Modellanpassung im Englischen sind alle Modellgütekriterien heranzuziehen und kritisch zu vergleichen. Deutlich wird auch hier, dass auf Grund der Modellspezifikationen hierarchisch ineinander geschachtelte, vergleichbare Modelle die gleiche Anpassung an die Daten aufweisen. Dies ist dadurch bedingt, dass auch für das Englische nur ein Indikator für jeden Kompetenzbereich vorliegt, sodass in den entsprechenden Fällen die gleichen Pfade geschätzt werden, auch wenn sich die Modelle theoretisch unterscheiden.

Auch wenn der CFI und der SRMR-Wert für alle Modelle vergleichbare Werte aufweisen, unterscheiden sich die Modelle untereinander doch bezüglich ihres Chi-Quadrat-Wertes, der Freiheitsgrade und des RMSEA-Werts. Das Modell mit der besten Anpassung an die Daten - gemessen am Chi-Quadrat/df-Verhältnis, dem RMSEA-Wert und dem AIC - das sich darüber hinaus auch am stärksten vom eindimensionalen Modell unterscheidet, ist das Modell b<sub>1</sub>) mit den zwei latenten Dimensionen der *reflexiv-rezeptiven* und der *produktiven* Kompetenz. Eine

identische Modellanpassung wird mit dem entsprechenden Modell  $c_1$ ) mit einem hierarchisch übergeordneten Gesamtfaktor der Sprachkompetenz erreicht (vgl. Abbildung 13).



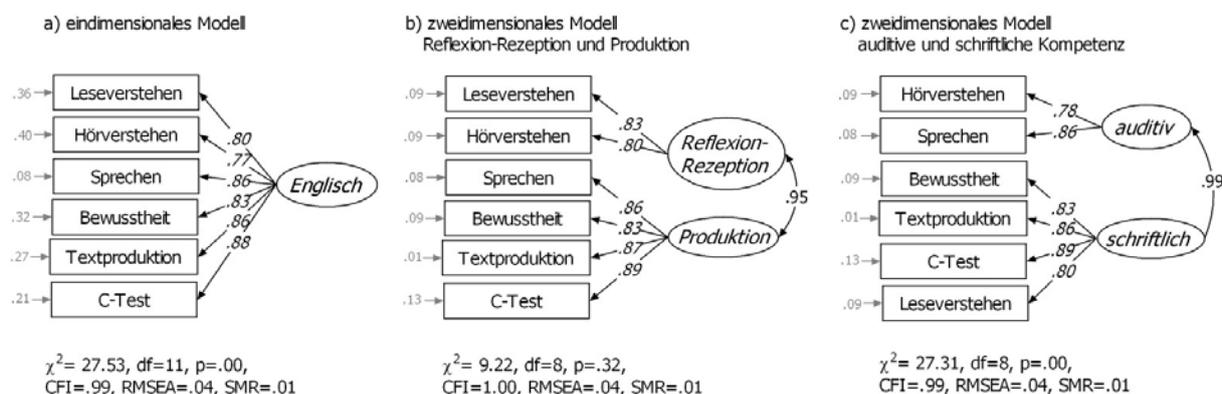
**Abbildung 13: Faktorenstruktur der Sprachkompetenz im Englischen (N=8964,  $\chi^2=28.04$ ,  $df=4$ ,  $p \leq .01$ , CFI=.99, AIC=94272.80, RMSEA=.03, SRMR=.01)**

Deutlich wird in dieser Darstellung die Schachtelung der Modelle: Der latente Gesamtfaktor zweiter Ordnung „Englisch“ ist definiert durch die beiden latenten Faktoren Reflexion-Rezeption und Produktion. Die Ladungen dieser beiden Faktoren spiegeln in diesem Fall – eine latente Variable mit nur zwei Indikatoren – die Interkorrelation zwischen beiden wider. Dementsprechend unterscheiden sich diese beiden Modelle untereinander nicht in ihrer Anpassung an die Daten.

Interessant ist im Vergleich der verschiedenen Modelle die Tatsache, dass die Modelle, welche speziell das Hörverstehen als auditive Kompetenz separat modellieren, keine bessere Modellanpassung aufweisen als diejenigen Modelle, die das Hörverstehen zusammen mit dem Leseverstehen auf einen gemeinsamen latenten Faktor, die reflexiv-rezeptive Sprachkompetenz, zurückführen. Hier liegen unterschiedliche Erklärungen nahe: Zum einen ist die Modellierung eines Kompetenzbereichs, der lediglich durch einen Indikator repräsentiert wird, in Messmodellen nicht optimal zu lösen, da Unreliabilitäten in diesem Indikator sich automatisch negativ auf die Anpassung des Gesamtmodells auswirken. Zum anderen ist zu bemerken, dass auch das Testformat des Tests zum Hörverstehen im Englischen starke schriftbasierte Anteile aufweist: Die Schülerinnen und Schüler müssen sowohl die Fragestellung als auch die Antwortalternativen im Multiple-Choice-Format lesen und diese schriftlichen Informationen mit den auditiv erfassten in Zusammenhang bringen. Daher ist es nicht verwunderlich, dass der Test zum Hörverstehen – wenn auch nicht ganz so hoch – zusammen mit dem Leseverstehen auf einem gemeinsamen latenten Faktor lädt.

### 6.2.3 Exkurs: Modellierung eines auditiven Faktors der Sprachkompetenz

Wie bereits beschrieben, wurde in DESI auch ein Test zur mündlichen Sprachkompetenz im Englischen durchgeführt, an dem auf Grund des Studiendesigns jedoch nur eine Teilstichprobe von Schülerinnen und Schülern teilnahmen. In den Gesamtauswertungen der vorliegenden Arbeit soll daher lediglich in diesem Exkurs auf Modelle der englischen Sprachkompetenz eingegangen werden, die diesen Test berücksichtigen, während alle anderen Analysen auf der Gesamtstichprobe der DESI-Studie basieren. Im Folgenden wird die Sprachkompetenz daher anhand der Daten der reduzierten Stichprobe von N=1692 Schülerinnen und Schüler, die am Test zur mündlichen Sprachkompetenz (SET-10) (Ordinate, 2004) teilnahmen, gemäß der drei in Abbildung 14 dargestellten alternativen Modelle der Sprachkompetenz nach dem Modell von Kunnan (1992) in die zwei latenten Dimensionen der *auditiven* und der *schriftsprachlichen* Kompetenz unterteilt (vgl. Abschnitt 2.2). Die Annahme, dass sich in der Fremdsprache ein eigener, auditiver Faktor abbilden lässt, liegt in der der institutionellen Praxis begründet, im fremdsprachlichen Anfangsunterricht zunächst auf verbale kommunikative Sprachkompetenzen zu fokussieren. Auch basieren Beziehung zwischen Hörverstehen und Sprechen auf gemeinsamen neuronalen Strukturen (vgl. Abschnitte 1.5 und 2.2.3).



**Abbildung 14: Vergleich der drei alternativen Modelle der Sprachkompetenz im Englischen unter Berücksichtigung des Tests zur mündlichen Sprachkompetenz (Set-10) (N=1692)**

Im Vergleich zwischen den drei Modellen wird deutlich, dass auch unter Hinzunahme des Tests zur Erfassung der mündlichen Sprachkompetenz das Modell mit den beiden latenten Faktoren Reflexion-Rezeption und Produktion die Struktur der Daten am besten abbilden kann. Bei diesem Modell wird sogar der Chi-Quadrat-Wert nicht signifikant. Das alternative

Modell, bei dem mit den beiden Testwerten zum Sprechen und zum Hörverstehen der latente Faktor „auditive Sprachkompetenz“ modelliert wird, kann die Zusammenhänge in den empirischen Daten nicht besser erklären als das eindimensionale Modell. Zudem korrelieren in diesem Modell die beiden latenten Faktoren auditive und schriftsprachliche Kompetenz zu nahezu eins.

Für die Analysen anhand der reduzierten Stichprobe, aber auch für die weiteren Analysen ohne den Test zum Hörverstehen kann festgestellt werden, dass ein Faktor der auditiven Sprachkompetenz mit den vorliegenden Testverfahren nicht separat modelliert werden kann. Das Modell mit den beiden latenten Faktoren der reflexiv-rezeptiven und produktiven Sprachkompetenz sowohl mit als auch ohne den zusätzlichen Test zur mündlichen Sprachkompetenz kann die Struktur der Daten hingegen sehr gut abbilden.

#### **6.2.4 Vergleich der Modelle im Deutschen und im Englischen**

Zur Analyse der Struktur von Sprachkompetenz im Deutschen und im Englischen wurden jeweils Modelle mit eindimensionaler, mehrdimensionaler und hierarchischer Struktur verglichen. Sowohl für das Deutsche als auch für das Englische zeigte sich, dass mehrdimensionale Modelle der Struktur der empirischen Daten am ehesten angemessen sind. Damit kann festgestellt werden, dass die in DESI verwendeten Testverfahren in der Lage sind, Sprachkompetenz im Deutschen und im Englischen differenziert zu erfassen. Leistungsunterschiede und Leistungszuwächse können entsprechend für bestimmte sprachliche Teilbereiche analysiert werden. Auf der Grundlage dieser einzelnen Teilbereiche wiederum ist es möglich, in beiden Sprachen latente Dimensionen zu modellieren, welche *reflexiv-rezeptive* und *produktive* Kompetenzbereiche repräsentieren. Diese Ergebnisse bestätigen beispielsweise die Befunde von North (2000).

Während im Deutschen die produktiven Teilbereiche stärker ausdifferenziert sind und sich in den bisher analysierten Modellen nicht optimal auf einen gemeinsamen latenten Faktor zurückführen lassen, weisen die produktiven Tests im Englischen eine homogenere Struktur auf. Das bedeutet, dass die Leistungen in Tests, welche im Deutschen Lesen, Wortschatz, Sprachbewusstheit und Argumentation erfassen, sehr hoch zusammenhängen, im Englischen ist dies bei den Testinstrumenten für die Messung des Leseverstehens und des Hörverstehens der Fall. Unterschiede zwischen den beiden Sprachen zeigen sich in den produktiven Sprachkompetenzen: Während im Deutschen zwischen der Rechtschreibung, der

Textproduktion Pragmatik und der Textproduktion Systematik zu differenzieren ist, lässt sich für das Englische die gemeinsame Varianz in den produktiven Testmodulen des C-Tests, der freien Textproduktion und der Bewusstheit Englisch durch eine gemeinsame latente Dimension der Sprachproduktion gut erklären. Für das Englische war es auf der Grundlage einer kleineren Stichprobe darüber hinaus möglich, ein konkurrierendes Modell mit den latenten Faktoren der *auditiven* und der *schriftsprachlichen* Kompetenz zu modellieren (vgl. Kunnan, 1994 und Abschnitt 2.2.3), mit welchem sich jedoch die Struktur in den Daten nicht so gut abbilden lässt.

Die Gemeinsamkeiten der beiden Sprachen – die *reflexiv-rezeptiven* und *produktiven* Kompetenzbereiche – deuten auf eine Struktur von Sprachkompetenz hin, die sprachübergreifend ist. Die reflexiv-rezeptive Sprachkompetenz ist differenzierbar vom produktiven Umgang mit Sprache, sowohl im Deutschen als auch im Englischen. Sprachverstehen kann als eine eigene, wesentliche Komponente in der Gesamtstruktur von Sprachkompetenz angesehen werden, deren Teilkomponenten auf gemeinsame Wahrnehmungsprozesse, wie beispielsweise die allgemeine Dekodierfähigkeit, zurückgeführt werden können (vgl. Abschnitt 2.2.2).

Dass für das Deutsche, anders als im Englischen, die Teilfähigkeiten der Rechtschreibung und Textproduktion Pragmatik getrennt zu betrachten sind, mag auch an der Differenziertheit der Testinstrumente liegen. Im Englischen könnte die stärkere Überschneidung der Testinhalte, insbesondere des C-Tests als gebundener Erfassung von Textproduktionsleistung mit dem Test zur freien Textproduktion, dazu führen, dass unter Einbezug dieser manifesten Indikatoren ein gemeinsamer Faktor der Sprachproduktion in der Fremdsprache modelliert werden kann.

Die in der Literatur zu findende Annahme, die Struktur der Erstsprache sei homogener als die der Fremdsprache (vgl. Abschnitte 1.5 und 1.8), lässt sich mit den vorliegenden Ergebnissen nicht stützen. Auch für die Hypothese, Sprachkompetenz sei letztlich ein homogenes, auf einen Faktor reduzierbares Konstrukt, findet sich in diesen Analysen keine Bestätigung. Vielmehr scheint gerade diese Annahme auf Grund der vielfältigen in DESI eingesetzten Testinstrumente widerlegbar zu sein. Am Beispiel der reflexiv-rezeptiven Testverfahren im Deutschen wurde darauf eingegangen: Betrachtet man lediglich diese vier Testverfahren, die ähnliche Kompetenzbereiche ansprechen, so lassen sie sich sehr gut auf eine Dimension

zurückführen. Die rezeptiven Kompetenzen in der Erstsprache sind bei den getesteten Schülerinnen und Schülern jeweils ähnlich ausgeprägt: Wer über eine hohe Lesekompetenz verfügt, beherrscht auch den entsprechenden Wortschatz und kann über das Gelesene reflektieren, hier zeigen sich also keine auffälligen Profilunterschiede zwischen den Personen. Sobald jedoch produktive Testverfahren hinzugenommen werden, zeigt sich ein differenzierteres Bild der Sprachkompetenz: Wer über eine hohe Lesekompetenz verfügt, kann nicht automatisch auch fehlerfreie Texte schreiben und umgekehrt. Ähnlich verhält es sich im Englischen. Dass die rezeptiven Testverfahren zum Lese- und Hörverstehen stark miteinander korrelieren, könnte auf gemeinsame sprachliche Verarbeitungseinheiten zurückgeführt werden, wie dies von McKay (1987) und Monsell (1987) angenommen wird (vgl. Abschnitt 2.2). Es ist jedoch auch zu berücksichtigen, dass in diesem Fall mit Multiple-Choice-Testformaten gleiche Testmethoden zur Erfassung der beiden Kompetenzen verwendet wurden, sodass ein gemeinsamer Methodeneffekt zu berücksichtigen ist.

Bei näherer Betrachtung der reflexiv-rezeptiven und der produktiven Teilkompetenzen in den beiden Sprachen wird deutlich, dass diese Bereiche von durchaus unterschiedlichen Testmodulen repräsentiert werden. Ursache hierfür sind zum einen die in DESI eingesetzten Testverfahren, die für das Englische einen Test im Hörverstehen und mündlicher produktiver Sprachkompetenz einschlossen, im Deutschen dafür die Testbereiche der Rechtschreibung und des Wortschatzes abdeckten (vgl. Abschnitt 3.3.3 und 3.3.4). Über diese Effekte des Testdesigns hinaus soll hier ergänzend darüber reflektiert werden, warum der Test zur „Sprachbewusstheit“ im Deutschen ein Element der reflexiv-rezeptiven Sprachkompetenz darstellt, während der entsprechende Test im Englischen dem produktiven Teilbereich zugeordnet wird. Zunächst ist festzuhalten, dass die gleichlautende Benennung der Testverfahren keinen direkten Rückschluss auf Inhalt oder Format der Testhefte ermöglicht. Auch wenn die Tests zur Erfassung der Sprachbewusstheit sowohl im Deutschen als auch im Englischen viele Items mit offenem Aufgabenformat enthalten, so liegt doch im Deutschen der Schwerpunkt auf der *Reflexion von sprachlichem Wissen*, sowohl im Bereich der Pragmatik als auch in dem der Grammatik (Eichler, 2007, 2008). Dementsprechend ist zu argumentieren, dass ein eindimensionales Teilmodell der reflexiv-rezeptiven Sprachkompetenz unter Einbeziehung der Sprachbewusstheit im Deutschen, wie in Abbildung 11 dargestellt, eine sehr gute Anpassung an die Daten erreicht.

Im Englischen hingegen wird bei Betrachtung der Testaufgaben deutlich, dass bei der Beantwortung der Testaufgaben zur Sprachbewusstheit neben deklarativem grammatikalischen Wissen auch Wortschatzwissen und Rechtschreibung eine wesentlich Rolle spielen (Nold & Rossa, 2007, 2008). Die beiden letzten Kompetenzbereiche werden im Englischen auch vom C-Test erfasst, der ebenfalls als Indikator stellvertretend für die produktive Sprachkompetenz steht. Auch empirisch zeigt sich in einer alternativen Modellierung, dass ein Modell, das den Sprachbewusstheitstest als Teil der reflexiv-rezeptiven Kompetenz modelliert, im Vergleich eine schlechtere Modellanpassung erreicht (N=8964,  $\chi^2=84.40$ ,  $df=4$ ,  $p \leq .01$ , CFI=.99, AIC=94387.87, RMSEA=.05, SRMR=.01) als Modell b<sub>1</sub>).

*Im Bezug auf **Hypothese 1** kann festgestellt werden, dass die Struktur von Sprachkompetenz im Deutschen und auch im Englischen gleichermaßen reflexiv-rezeptive und produktive Bereiche aufweist, so dass eine strukturelle Vergleichbarkeit zwischen der Erst- und der Fremdsprache durchaus gegeben ist.*

Diese vergleichbare, mehrdimensionale Struktur von Sprachkompetenz kann zur Beschreibung von Unterschieden zwischen Personen herangezogen werden, die einen Mittelweg darstellt zwischen einer eindimensionalen Betrachtungsweise – der Bildung von Mittelwerten über die gesamten Teiltests – und einer stark differenzierten Betrachtung einzelner Testmodule. Die so theoretisch-inhaltlich interpretierbaren Strukturgleichungsmodelle für die beiden Sprachen werden im Folgenden in einem gemeinsamen Strukturgleichungsmodell der Erst- und Fremdsprachenkompetenz analysiert.

### **6.2.5 Gemeinsames Modell des Deutschen und des Englischen**

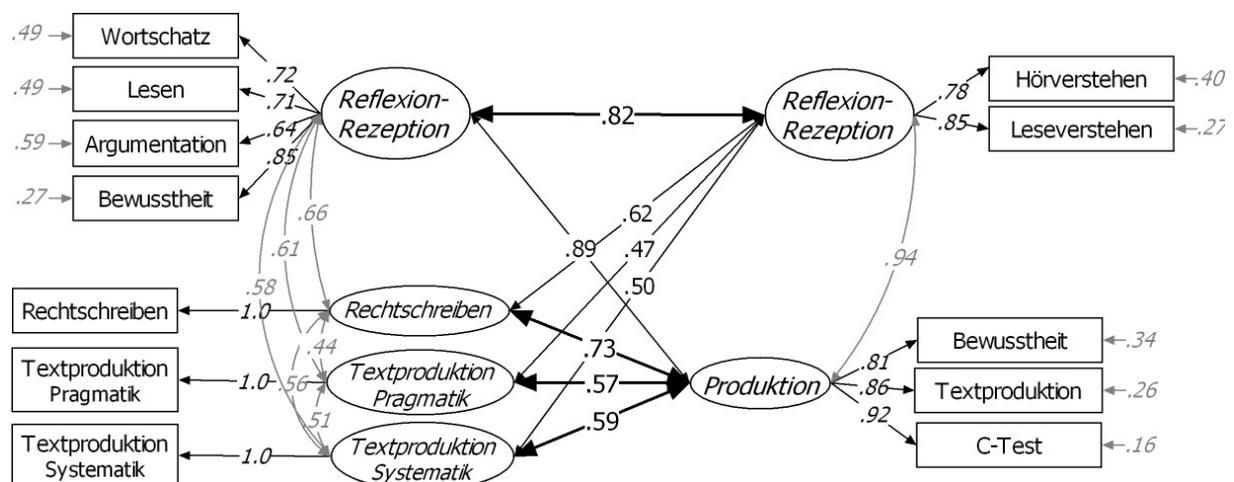
Auf der Grundlage der separaten Analysen der Zusammenhänge zwischen den sprachlichen Teilbereichen im Deutschen und im Englischen wurden in einem nächsten Schritt die Zusammenhänge zwischen den beiden Sprachen in einem gemeinsamen Strukturgleichungsmodell analysiert. Hierzu wurden für die beiden Messmodelle im Deutschen und im Englischen jeweils diejenigen Modelle verwendet, die in den Einzelanalysen die beste Anpassung an die Daten aufweisen konnten: Im Deutschen das Modell mit vier Faktoren, im Englischen das Modell mit zwei latenten Faktoren. Es wurde ein gemeinsames Modell analysiert, in dem die Zusammenhänge zwischen allen latenten

Dimensionen frei geschätzt wurden. Entsprechend der Fragestellungen 2 in Abschnitt 4.1 ist zu analysieren, welche differenzierten Zusammenhänge zwischen den spezifischen sprachlichen Kompetenzbereichen im Deutschen und im Englischen bestehen.

**Hypothese 2:** Die speziellen Zusammenhänge zwischen den korrespondierenden latenten Faktoren, die im Deutschen und im Englischen erfasst wurden, sind höher als die Zusammenhänge zwischen nicht korrespondierenden sprachlichen Teilbereichen im Deutschen und Englischen.

Angenommen wird, dass die Zusammenhänge zwischen dem Bereich der Sprachreflexion-Rezeption im Deutschen und im Englischen sowie zwischen den produktiven Teilbereichen im Deutschen und der Sprachproduktion im Englischen jeweils höher sind als die Zusammenhänge verschiedener sprachlicher Teilbereiche untereinander. Diese Annahme basiert auf Befunden zu *spezifischen Transfereffekten* zwischen Erst- und Fremdsprache (vgl. Abschnitt 2.3).

Auf der Grundlage der Daten der Gesamtstichprobe lassen sich die Zusammenhänge zwischen dem Englischen und dem Deutschen wie folgt darstellen (vgl. Abbildung 15).



**Abbildung 15: Modell der sprachlichen Kompetenzen im Deutschen und im Englischen mit Zusammenhängen zwischen allen sprachlichen Teilbereichen auf der Grundlage der Gesamtstichprobe (N=8695,  $\chi^2=780.40$ ,  $df=42$ ,  $p \leq .01$ , CFI=.99, RMSEA=.05, SRMR=.02)**

Das dargestellte Modell weist insgesamt eine sehr gute Modellanpassung auf, die nicht zuletzt durch die vorab gut angepassten separaten Messmodelle für das Deutsche und das Englische bedingt ist. Es zeigen sich signifikante Korrelationen zwischen allen sprachlichen Teilbereichen. Zu beachten ist hierbei, dass die *absolute Höhe der Korrelationen* für die latent modellierten Faktoren und die manifesten Indikatoren *jeweils separat zu betrachten* ist und nicht direkt verglichen werden kann.

*Der latente Faktor der Reflexion-Rezeption im Deutschen korreliert entgegen der Annahmen in **Hypothese 2** höher mit dem sprachproduktiven als mit dem reflexiv-rezeptiven Faktor im Englischen. Hypothesenkonform korrelieren die sprachproduktiven Teilbereiche im Deutschen mit den latenten Faktoren im Englischen.*

Für den Bereich der rezeptiven Sprachkompetenz muss die Annahme der spezifischen Zusammenhänge verworfen werden: Die Sprachreflexion-Rezeption im Deutschen hängt höher mit der Sprachproduktion im Englischen zusammen als mit der Sprachreflexion-Rezeption in dieser Fremdsprache. Die produktiven Teilbereiche im Deutschen weisen allerdings hypothesenkonform höhere Korrelationen mit dem produktiven Teilbereich des Englischen auf als mit dem rezeptiven.

In der Forschung zum Fremdsprachenerwerb wird oft auf allgemeine und spezifische Transfereffekte zwischen Erst- und Fremdsprache hingewiesen. Es wird davon ausgegangen, dass kognitive Grundlagen wie die Dekodierfähigkeit für allgemeine Zusammenhänge zwischen den Kompetenzen verantwortlich sind, darüber hinaus jedoch auch spezifische Fähigkeiten wie grammatikalisches Regelwissen und sprachliche Ausdrucksfähigkeit von der Erstsprache auf die Fremdsprachen übertragen werden. In der hier dargestellten Analyse lassen die starken Zusammenhänge zwischen dem Deutschen und dem Englischen für alle Teilbereiche auf übergreifende allgemeine Transfereffekte schließen, während die unterschiedliche Höhe derselben auf spezifischen Transfer bestimmter Komponenten schließen lässt: Der Bereich der Reflexion-Rezeption im Deutschen korreliert hoch mit beiden latenten Faktoren im Englischen: Es kann vermutet werden, dass die in diesem Faktor zusammengefassten Fähigkeiten – Lesen, Wortschatz, Argumentation und Bewusstheit – eine Art Basissprachkompetenz der Erstsprache darstellen, die zu einem gewissen Anteil die Grundlage für alle sprachlichen Kompetenzen, auch jene in der Fremdsprache, bildet. Besonders auf das Lesen als institutionell vermittelte, grundlegende Fähigkeit, die stark von

der Dekodierungsfähigkeit abhängt, wird bei der Bearbeitung jeglicher Testverfahren im Deutschen wie im Englischen – im reflexiv-rezeptiven wie auch im produktiven Bereich – zurückgegriffen (vgl. Abschnitt 2.2.1).

Für den Bereich der reflexiv-rezeptiven Sprachkompetenz lässt sich festhalten, dass diese für die Erst- und die Fremdsprache im Rezeptionsprozess auf eine allgemeine Dekodierfähigkeit zurückgreift (Meschyan & Hernandez, 2002; Sparks et al., 1998). Darüber hinaus kann davon ausgegangen werden, dass auch die allgemeine Lesefähigkeit im Sinne von Worterkennung und Dechiffrierung in beiden Sprachen gleichermaßen zum Tragen kommt, wenn die Lesetestaufgaben in beiden Sprachen bearbeitet werden (Riemer, 2000). Letztendlich rekurriert die Sprachbewusstheit, die im Deutschen als reflexive Komponente erfasst wurde, auf einen sprachlichen Teilbereich, der im Sinne einer grammatikalischen Reflexionsfähigkeit – mentales Lexikon – auch Grundlage der fremdsprachlichen Sprachrezeption ist (Ullman, 2001). Und schließlich sollte auch berücksichtigt werden, dass zur Erfassung der Lesekompetenz im Deutschen und im Englischen ein Multiple-Choice-Testformat verwendet wurde, wodurch sich ebenfalls ein Anteil der gemeinsamen Varianz der Leistungen erklären lässt (vgl. auch Rost & Sparfeldt, 2007).

Stärker differenzierte spezifische Zusammenhänge zeigen sich zwischen den produktiven Kompetenzen in beiden Sprachen. Festgestellt werden können Transfereffekte in den spezifischen Bereichen der Rechtschreibung im Sinne einer Bewusstheit für das richtige Schreiben von Wörtern. Die Bedeutung der Rechtschreibfähigkeit lässt sich dadurch erklären, dass diese sowohl für korrekte Antworten im C-Test als auch für das freie Schreiben im Englischen eine Rolle spielt und in deren Bewertung einfließt. Das Wissen um und die Berücksichtigung der Korrektheit der Schreibweise von Wörtern kann als eine Kompetenz angesehen werden, die wenn sie in der Erstsprache vorhanden ist, auf weitere Sprachen übertragen wird. Auch auf die Fähigkeit, zielgerichtete aussagekräftige Texte frei zu formulieren, kann, wenn sie in der Erstsprache erworben wurde, durchaus in allen weiteren Sprachen zurückgegriffen werden. Denn hierzu gehören viele Bereiche des Sprachwissens, die – wenn sie jeweils um einen ausreichenden Wortschatz in der Fremdsprache ergänzt werden – praktisch für jede Sprache einsetzbar sind (Selinker & Douglas, 1988).

Zusammenfassend ist schon an dieser Stelle festzuhalten, dass ein beträchtlicher Anteil der gemeinsamen Varianz von Erst- und Fremdsprachenkompetenz auf gemeinsame

grundlegende Faktoren zurückzuführen ist, die im Rahmen der DESI-Studie nur teilweise identifiziert werden konnten. Hierzu gehören neben den allgemeinen kognitiven Grundfähigkeiten (Csapó 2003; Csapó & Nikolov, 2002) sicherlich auch die allgemeine Sprachbewusstheit (Lightbow, 2000; Luchtenberg, 1995) sowie neurologische Grundlagen (Hufeisen, 2003; Ullmann, 2001). Auch wenn an dieser Stelle auf Grund des Querschnittsdesigns nicht von kausalen Transfereffekten gesprochen werden soll, so ist doch anzunehmen, dass auf erlernte Fertigkeiten in der Erstsprache zurückgegriffen wird, wenn ähnliche Anforderungen in der Fremdsprache gestellt werden.

## 6.3 Struktur von Sprachkompetenz unter Berücksichtigung der hierarchischen Schachtelung der Daten

### 6.3.1 Analyse der Zusammenhänge auf Individual- und Klassenebene

Wie im Methodenkapitel der vorliegenden Arbeit bereits ausgeführt wurde, basieren die hier dargestellten Analysen auf einem hierarchisch geschachtelten Datensatz (vgl. Abschnitt 5.3.1), das heißt, sie weisen eine Mehrebenenstruktur mit zwei Ebenen auf: Die getesteten Schülerinnen und Schüler auf Ebene eins, der Individualebene, sind in Klassen auf Ebene zwei, der Klassenebene, geschachtelt. Es kann davon ausgegangen werden, dass in solchermaßen hierarchischen Datenstrukturen die Merkmale der Schülerinnen und Schüler im Sinne von Leistungen oder Einstellungen nicht unabhängig voneinander sind, sondern durch die Klassenzugehörigkeit systematisch beeinflusst werden können (Muthén, 1994; Raudenbush & Bryk, 2002). Die daraus resultierende Korreliertheit von individuellen Beobachtungen führt bei der Verwendung von klassischen Analyseverfahren zu einer Unterschätzung der Standardfehler.

Darüber hinaus ist bei einer solchen Datenstruktur zu berücksichtigen, dass sich die interindividuellen Unterschiede zwischen den Schülerinnen und Schülern aus zwei Variationsquellen zusammensetzen, der Variation auf der individuellen Ebene und der Variation auf der Klassenebene. Diese können *inhaltlich interpretiert* werden. Dementsprechend werden im Folgenden die Strukturen von Sprachkompetenz in Mehrebenenmodellen analysiert, welche die Zusammenhänge zwischen den Testmodulen für beide Ebenen separat, jedoch in einem gemeinsamen Modell schätzen.

Um die Struktur der Sprachkompetenz unter Berücksichtigung der hierarchischen Schachtelung der Daten analysieren zu können, wurde das von Muthén (1994) beschriebene schrittweise Vorgehen gewählt, bei dem die Zusammenhänge zwischen den Teilkompetenzen für die beiden Ebenen der Individuen und der Klassen *zuerst getrennt analysiert* werden, bevor sie *in einem gemeinsamen Modell* zu schätzen sind. In einem ersten Schritt ist entsprechend der Hypothese 3 das Ausmaß der Ähnlichkeit der abhängigen Variablen innerhalb einer Gruppe zu bestimmen, also zu schätzen, wie stark der Einfluss der Clusterung auf die Individualmesswerte ist.

**Hypothese 3:** Die in der DESI-Studie erhobenen Daten weisen eine hohe Abhängigkeit zwischen den Klassen, das heißt eine bedeutsame hierarchische Schachtelung auf.

Dieser Anteil der Varianz, die *Intraklassenkorrelation* (ICC) (vgl. Abschnitt 5.3.1), ist für alle Testmodule in Tabelle 8 dargestellt.

**Tabelle 8: Intraklassenkorrelationen der Tests im Deutschen und im Englischen.**

Deutsch							
Test	Bewusstheit	Lesen	Textproduktion Pragmatik	Textproduktion Systematik	Argumentation	Wortschatz	Rechtschreiben
$\rho$	0.53	0.34	0.27	0.28	0.29	0.40	0.43

Englisch					
Test	Bewusstheit	Lesen	Textproduktion	C-Test	Hörverstehen
$\rho$	0.59	0.45	0.60	0.60	0.49

Der Koeffizient der Intraklassenkorrelation zeigt an, wie groß der Anteil an Varianz in der abhängigen Variable ist, der auf die Zugehörigkeit zu einer übergeordneten Gruppe, in diesem Fall einer Schulklasse, zurückgeführt werden kann. Als Richtwerte, ab wann die Methoden der Mehrebenenanalysen zu verwenden sind, werden Werte von  $\rho \geq .05$  (Heck, 2001), aber auch  $\rho \geq .12$  (James, 1982) genannt. Deutlich wird anhand der Daten in Tabelle 8, dass der Varianzanteil, der auf die Zugehörigkeit zu einer Klasse zurückzuführen ist, insgesamt über dem Wert liegt, ab dem die Berücksichtigung der Schachtelung in statistischen Analysen geboten ist. Darüber hinaus unterscheidet er sich zwischen den Testmodulen: Besonders die individuellen Leistungen im Englischen sind stark von der jeweiligen Klassenzugehörigkeit geprägt. Erklärt werden kann dies dadurch, dass sich im gemeinsamen Datensatz auch Klassen mit bilinguaem Unterricht in Englisch befinden, von denen angenommen werden kann, dass sie sich von den Regelklassen stark unterscheiden, sodass sich dieser stichprobenimmanente Unterschied zwischen den Gruppen entsprechend in der Intraklassenkorrelation niederschlägt.

Analog zu **Hypothese 3** kann festgestellt werden, dass die in der DESI-Studie erhobenen Daten eine bedeutsame hierarchische Schachtelung aufweisen.

Unter Berücksichtigung dieses Ergebnisses sind in den folgenden Analysen nicht nur die beiden Analyseebenen zu berücksichtigen, sondern es ist auch danach zu fragen, ob sich die

strukturellen Zusammenhänge der sprachlichen Teilbereiche im Deutschen und im Englischen zwischen den Ebenen unterscheiden, wie in Hypothese 4 postuliert wird.

**Hypothese 4:** Die Zusammenhänge der sprachlichen Teilbereiche sind auf Klassenebene höher als auf Individualebene.

Zur Überprüfung dieser Hypothese werden die Interkorrelationen der Testskalen auf Individualebene und auf Klassenebene für die Testmodule im Deutschen (vgl. Tabelle 9) und im Englischen (vgl. Tabelle 10) bestimmt.

**Tabelle 9: Korrelationen der Testmodule Deutsch auf Individualebene (N=9062) links unterhalb der Diagonale und auf Klassenebene (N=422) rechts oberhalb der Diagonale**

Between / Within	Bewusstheit	Lesen	Textproduktion Pragmatik	Textproduktion Systematik	Argumentation	Wortschatz	Rechtschreiben
Bewusstheit		0.97	0.94	0.95	0.96	0.95	0.95
Lesen	0.35		0.91	0.91	0.96	0.95	0.90
Textproduktion Pragmatik	0.31	0.24		0.92	0.94	0.89	0.90
Textproduktion Systematik	0.30	0.17	0.37		0.92	0.90	0.97
Argumentation	0.29	0.28	0.28	0.18		0.92	0.92
Wortschatz	0.36	0.35	0.20	0.17	0.26		0.89
Rechtschreiben	0.29	0.15	0.22	0.36	0.18	0.18	

**Tabelle 10: Korrelationen der Testmodule Englisch auf Individualebene (N=8964) links unterhalb der Diagonale und auf Klassenebene (N=422) rechts oberhalb der Diagonale**

Between / Within	C-Test	Hörverstehen	Leseverstehen	Sprachbewusstheit	Semikreatives Schreiben
C-Test		0.93	0.97	0.98	0.96
Hörverstehen	0.39		0.94	0.93	0.91
Leseverstehen	0.51	0.43		0.97	0.95
Sprachbewusstheit	0.43	0.30	0.36		0.95
Semikreatives Schreiben	0.52	0.28	0.42	0.35	

Aus den in diesen Tabellen dargestellten Ergebnissen lässt sich folgender Schluss im Bezug auf Hypothese 4 ziehen:

*Die Zusammenhänge zwischen den Testmodulen im Deutschen wie im Englischen sind auf Individualebene deutlich geringer als die Zusammenhänge zwischen den Mittelwerten der Klassen, womit **Hypothese 4** gestützt werden kann.*

Dieses Phänomen ist ein inzwischen häufig berichtetes und kann zum einen darauf zurückgeführt werden, dass die auf Klassenebene aggregierten Variablen eine höhere Reliabilität aufweisen (Burstein, 1980; Snijders & Bosker, 1999). Werden die meßfehlerbehafteten Testwerte auf Individualebene auf Klassenebene aggregiert betrachtet, so stellen sie einen ungleich reliableren Schätzer für die mittlere Kompetenz in der Klasse dar. Zum anderen unterscheiden sich die Klassenmittelwerte nicht so stark voneinander wie die individuellen Leistungswerte von Personen einer Klasse. Also wird Hypothese 4 durch die vorliegenden Daten gestützt.

Sowohl die *hohen Intraklassenkorrelationen* für alle Testmodule, als auch die *starken Unterschiede zwischen den Korrelationen auf Individual- und auf Klassenebene* begründen die Notwendigkeit, die Struktur von Sprachkompetenz im Weiteren in Mehrebenenfaktorenanalysen beziehungsweise Mehrebenen-Strukturgleichungsmodellen zu analysieren, um potenzielle Unterschiede in der Struktur auf beiden Ebenen darstellen zu können (vgl. Abschnitt 5.3.2.).

***Hypothese 5:** Die faktorielle Struktur von Sprachkompetenz ist auf Klassenebene homogener als auf Individualebene.*

In den nächsten Abschnitten werden entsprechend der Hypothese 5 zunächst die Strukturen in den sprachlichen Bereichen Deutsch und Englisch getrennt analysiert, bevor ein gemeinsames Mehrebenen-Strukturgleichungsmodell analysiert wird.

### **6.3.1.1 Analyse der strukturellen Zusammenhänge auf Individual- und Klassenebene für das Deutsche**

Wie schon anhand der Korrelationsmatrizen in Tabelle 9 deutlich wurde, unterscheiden sich die Korrelationen der Testmodule im Deutschen zwischen der Individual- und der Klassenebene stark: Die Zusammenhänge auf Klassenebene sind durchweg hoch, während die Zusammenhänge auf Individualebene auf eine stärker differenzierbare Struktur hinweisen. Um in einem integrierten Mehrebenenmodell die Struktur auf beiden Ebenen korrekt darstellen zu können, wurde daher das mehrschrittige Vorgehen von Muthén (1994) gewählt: Zunächst wird die Faktorenstruktur mittels exploratorischer und konfirmatorischer Faktorenanalysen für beide Ebenen getrennt analysiert, bevor ein gemeinsames Modell für beide Ebenen angepasst wird. Allerdings ist darauf hinzuweisen, dass exploratorische Faktorenanalysen ein Verfahren sind, das zur Anwendung kommt, wenn keine spezifischen Hypothesen über die Anzahl und Struktur der den Korrelationen potenziell zugrunde liegenden Faktoren vorliegen (Moosbrugger & Schermelleh-Engel, 2007). Dieser Analyseschritt dient hier zum einen der Illustration des idealen Vorgehens bei der Analyse von Mehrebenenstrukturen, zum anderen gehen die exploratorisch gefundenen Modelle im Sinne einer empiriegeleiteten Hypothesengenerierung als Alternativmodelle in den Vergleich der theoriebasierten Modelle ein.

Die entsprechenden exploratorischen Faktorenanalysen auf der Grundlage der Within-Group-Kovarianzmatrix für die Individualebene und der geschätzten Between-Group-Kovarianzmatrix für die Klassenebene weisen für die Sprachkompetenz im Deutschen für beide Ebenen nach Parallelanalyse eine dreidimensionale Faktorenlösung aus. In einem nächsten Schritt wurden konfirmatorische Faktorenanalysen auf Individual- und Klassenebene gerechnet, die zum einen auf den in Abschnitt 6.2.1 analysierten, theoretisch fundierten Modellen basieren, zum anderen das Ergebnis der exploratorischen Faktorenanalysen einbeziehen:

- ein eindimensionales Modell mit einem Generalfaktor,
- ein zweidimensionales Modell mit den latenten Faktoren Reflexion-Rezeption und Produktion,
- ein dreidimensionales Modell analog zur exploratorischen Faktorenanalyse, in dem der Bereich der Textproduktion in zwei Faktoren zerlegt wird, sowie

- ein vierdimensionales Modell, das anhand der Ergebnisse für die Gesamtstichprobe bestätigt werden konnte.

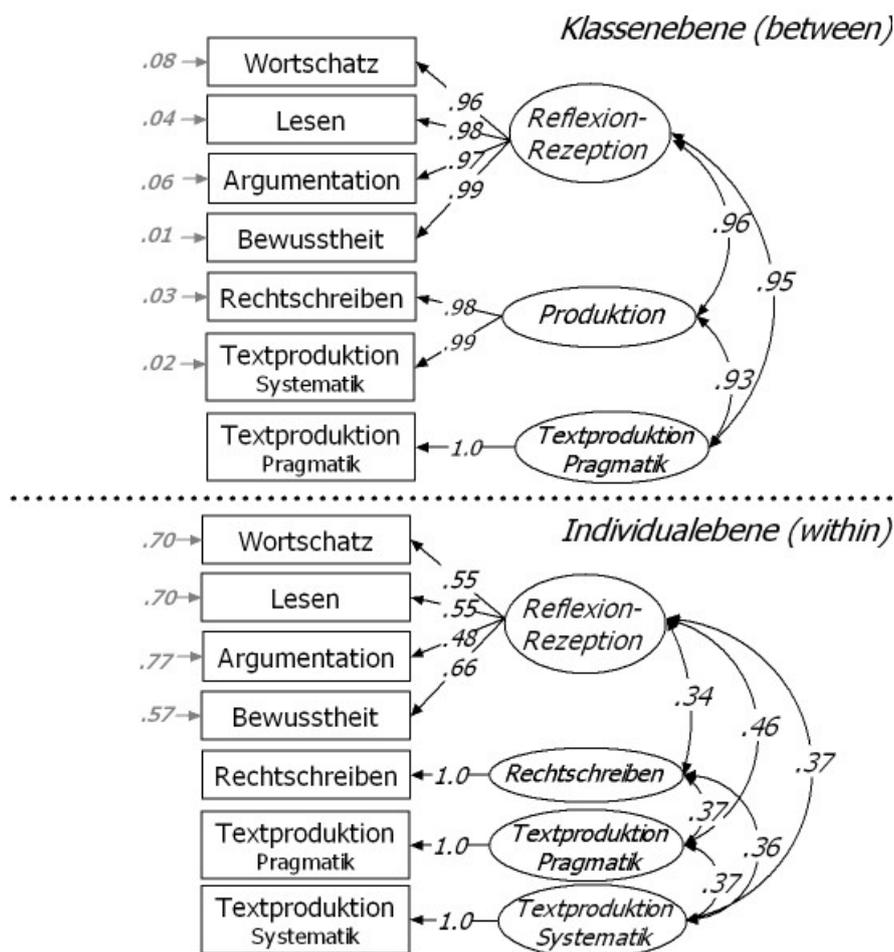
Analysiert wurden also insgesamt 16 verschiedene Modelle, welche die oben dargestellten vier Modelle auf beiden Ebenen kombinierten. Die Ergebnisse der konfirmatorischen Faktorenanalysen sind in Tabelle 11 dargestellt.

**Tabelle 11: Modellanpassung der konfirmatorischen Mehrebenen-Faktorenanalysen für die Struktur der Deutschkompetenz auf Individual- und Klassenebene**

Individualebene (within)	Klassenebene (between)	$\chi^2$	df	p	CFI	AIC	RMSEA	SRMR <sub>b</sub>	SRMR <sub>w</sub>
Eindimensional	Eindimensional	846.08	28	≤.01	0.90	142686.54	0.06	0.02	0.05
Eindimensional	Zweidimensional	785.96	27	≤.01	0.91	142604.69	0.06	0.02	0.05
Eindimensional	Dreidimensional	776.62	26	≤.01	0.91	142569.55	0.06	0.01	0.05
Eindimensional	Vierdimensional	761.19	25	≤.01	0.91	142556.41	0.06	0.01	0.05
Zweidimensional	Eindimensional	449.64	27	≤.01	0.95	142090.80	0.04	0.02	0.04
Zweidimensional	Zweidimensional	427.55	26	≤.01	0.95	142055.75	0.04	0.01	0.04
Zweidimensional	Dreidimensional	406.16	25	≤.01	0.95	142013.43	0.04	0.01	0.04
Zweidimensional	Vierdimensional	397.34	24	≤.01	0.96	142005.44	0.04	0.01	0.04
Dreidimensional	Eindimensional	356.35	26	≤.01	0.96	141942.11	0.04	0.02	0.03
Dreidimensional	Zweidimensional	337.48	25	≤.01	0.96	141910.62	0.04	0.02	0.03
Dreidimensional	Dreidimensional	323.32	24	≤.01	0.96	141884.42	0.04	0.01	0.03
Dreidimensional	Vierdimensional	315.03	23	≤.01	0.96	141876.56	0.04	0.01	0.03
Vierdimensional	Eindimensional	281.83	25	≤.01	0.97	141832.83	0.03	0.02	0.03
Vierdimensional	Zweidimensional	261.30	24	≤.01	0.97	141799.50	0.03	0.01	0.03
Vierdimensional	Dreidimensional	245.76	23	≤.01	0.97	141772.96	0.03	0.01	0.03
Vierdimensional	Vierdimensional	242.24	22	≤.01	0.97	141770.87	0.03	0.01	0.03

Deutlich wird im Vergleich der Modelle untereinander, dass die Modellanpassung der stärker differenzierten Modelle sowohl für die Individual- als auch für die Klassenebene besser ausfällt. Die insgesamt beste Anpassung an die Daten erreichen die Modelle, die eine vierdimensionale Struktur der Sprachkompetenz auf Individualebene annehmen – dies entspricht der Struktur für das Deutsche, die bereits in den Analysen in Abschnitt 6.2.1 bestätigt wurde. Für die Klassenebene weist das Modell mit der dreidimensionalen Struktur – die analog zu den Befunden der exploratorischen Faktorenanalyse modelliert wurde – die beste Anpassung auf, als Kriterium hierfür dient das Chi-Quadrat/df-Verhältnis, das bei diesem Modell im Vergleich am besten ist. Für die Mehrebenen-Struktur der Sprachkompetenz im Deutschen lässt sich das Modell, welches auf Individualebene eine vierdimensionale Struktur und auf Klassenebene eine dreidimensionale Struktur annimmt, als

angemessen bezeichnen. Die entsprechenden Modellgütekriterien sind in der Tabelle grau unterlegt. Das resultierende Modell ist in Abbildung 16 dargestellt.



**Abbildung 16: Konfirmatorische Mehrebenenfaktorenanalyse für die Sprachkompetenz im Deutschen (N=9062,  $\chi^2=245.76$ ,  $df=23$ ,  $p \leq .01$ , CFI=.97, RMSEA=.03,  $SRMR_{within}=.03$ ,  $SRMR_{between}=.01$ )**

Bei der Betrachtung dieses Modells sind zwei wesentliche Aspekte zu reflektieren: Die unterschiedliche Anzahl der latenten Faktoren auf den beiden Ebenen und die Zusammenhänge zwischen diesen Dimensionen der Sprachkompetenz.

*Für das Deutsche finden sich Hinweise auf die in **Hypothese 5** postulierte Annahme der homogeneren Strukturen von Sprachkompetenz auf Klassenebene.*

Auf Individualebene entspricht das Modell demjenigen, das in Abschnitt 6.2.1 im Vergleich am besten abschnitt. Offensichtlich wird, dass die Zusammenhänge zwischen den vier latenten Dimensionen auf Individualebene wesentlich niedriger ausfallen als im

Gesamtmodell. Der Zusammenhang zwischen dem Faktor Sprachreflexion-Rezeption und dem Modul Rechtschreibung, der im Gesamtmodell mit  $\phi=.67$  am stärksten ausfiel, ist auf Individualebene am geringsten. Die höchste Korrelation findet sich nun zwischen der Sprachreflexion-Rezeption und der Textproduktion Pragmatik. Dies lässt sich insofern leicht nachvollziehen, als in die Bewertung der Leistung im Testmodul Textproduktion Pragmatik hermeneutisch Aspekte der Wortwahl, des Ausdrucks und der sprachbewussten Produktion frei zu schreibender Texte eingehen, die in den separaten Testverfahren, welche durch den Faktor der Sprachreflexion-Rezeption repräsentiert werden, ebenfalls erfasst wurden. Trotzdem stellen die Textproduktion Pragmatik ebenso wie die Textproduktion Systematik und die Rechtschreibung separat zu interpretierende Kompetenzkomponenten dar, weshalb sie auf Individualebene auch nicht durch ein eindimensionales Konstrukt zu repräsentieren sind. Vielmehr variieren die individuellen Leistungen von Schülerinnen und Schülern in einer Klasse durchaus unabhängig in diesen vier Dimensionen, sodass differenzierte Leistungsprofile auf der Grundlage der separaten Testleistungen erstellt werden können und bezüglich der Unterschiede in den sprachlichen Teilbereichen sinnvoll zu interpretieren sind.

Auf Klassenebene hingegen zeigt sich eine homogenere Struktur der Teilbereiche: Die latenten Korrelationen zwischen den latenten Faktoren sind sehr hoch, trotzdem kann das dreidimensionale Modell die Zusammenhänge in den empirischen Daten besser abbilden als ein eindimensionales Modell. Als Besonderheit der Struktur auf Klassenebene hervorzuheben ist die Struktur der produktiven Teilkomponenten: Die mittleren Leistungen in den Testmodulen Rechtschreibung und Textproduktion Systematik korrelieren über alle Klassen hinweg so hoch, dass die gemeinsame Varianz durch einen latenten Faktor abgebildet werden kann. Separat zu modellieren ist jedoch die Leistung in der Textproduktion Pragmatik – diese scheint zwischen den Klassen und unabhängig von den Leistungen in den anderen beiden sprachproduktiven Teilbereichen so stark zu variieren, dass sie nicht mit diesen in einem gemeinsamen Faktor abgebildet werden kann. Anders gesagt, es scheint Klassen zu geben, in denen alle sprachproduktiven Kompetenzaspekte gleichermaßen beherrscht werden, während in anderen Klassen der Leistungsfokus entweder auf sprachsystematischen *oder* sprachpragmatischen Aspekten liegt. Dies ist durchaus denkbar, da sich der Deutschunterricht der DESI-Klassen besonders zwischen Bildungsgängen sowohl hinsichtlich der eingesetzten Lernmaterialien als auch der spezifischen sprachlichen Lernsituationen durchaus

unterscheidet (Klieme et al., 2008). Im Folgenden ist zu analysieren, welche entsprechenden Sprachstrukturen im Bereich des Englischen vorliegen.

### 6.3.1.2 Analyse der strukturellen Zusammenhänge auf Individual- und auf Klassenebene für das Englische

Analog zur schrittweisen Vorgehensweise im Deutschen wurden auch für die Zusammenhänge zwischen den Testmodulen im Englischen zunächst separate exploratorische Faktorenanalysen für die Individual- und die Klassenebene berechnet. Dabei zeigte sich für die fünf Testmodule im Englischen eine deutlich eindimensionale Struktur auf beiden Ebenen. In einem nächsten Schritt wurden entsprechend dem Vorgehen in Abschnitt 6.3.1.1 konfirmatorische Mehrebenen-Faktorenanalysen berechnet, bei denen permutierend jeweils drei Modellstrukturen auf beiden Ebenen verglichen wurden:

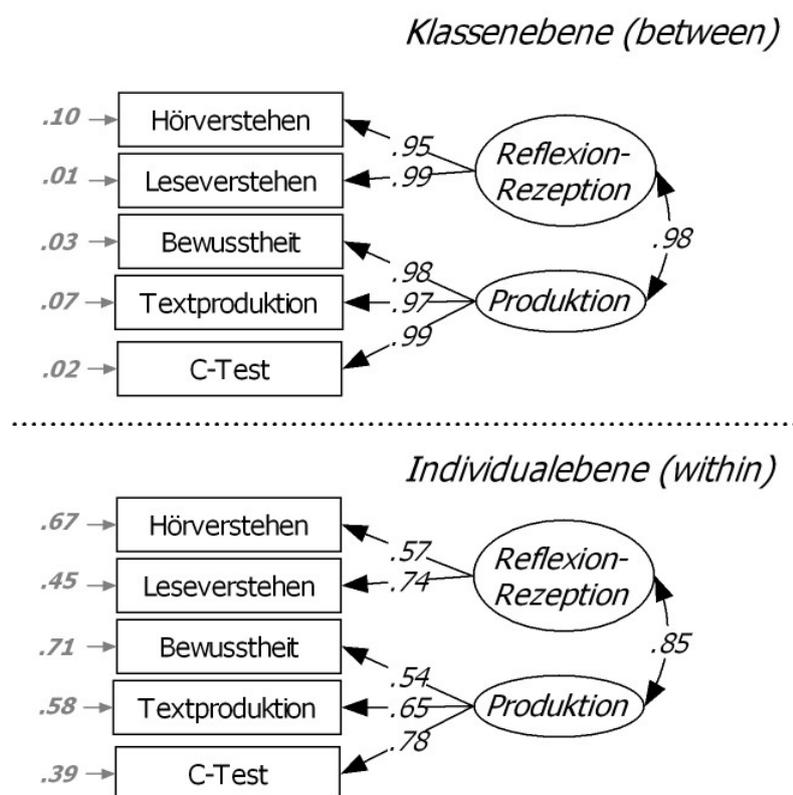
- ein eindimensionales Modell,
- ein zweidimensionales Modell mit den latenten Dimensionen der *Sprachreflexion-Rezeption* und der *Sprachproduktion* sowie
- ein zweidimensionales Modell mit den latenten Dimensionen der *auditiven* und der *schriftsprachlichen* Kompetenz.

Die Modellanpassungen sind in Tabelle 12 dargestellt.

**Tabelle 12: Modellanpassung der konfirmatorischen Mehrebenen-Faktorenanalysen für die Struktur der Englischkompetenz auf Individual- und Klassenebene**

Individualebene (within)	Klassenebene (between)	$\chi^2$	df	p	CFI	AIC	RMSEA	SRMR <sub>b</sub>	SRMR <sub>w</sub>
Eindimensional	Eindimensional	168.41	10	≤.01	0.97	83910.55	0.04	0.01	0.03
Eindimensional	Zweidimensional Rezeption-Produktion	153.81	9	≤.01	0.97	83890.28	0.04	0.01	0.03
Eindimensional	Zweidimensional Auditiv-Schriftsprachlich	Entspricht dem Modell: Eindimensionale Struktur auf beiden Ebenen							
Zweidimensional Rezeption-Produktion	Eindimensional	59.01	9	≤.01	0.99	83726.66	0.03	0.01	0.02
Zweidimensional Rezeption-Produktion	Zweidimensional Rezeption-Produktion	49.67	8	≤.01	0.99	83715.52	0.02	0.00	0.02
Zweidimensional Rezeption-Produktion	Zweidimensional Auditiv-Schriftsprachlich	Entspricht dem Modell: Eindimensionale Struktur auf beiden Ebenen							
Zweidimensional Auditiv-Schriftsprachlich	Eindimensional	Entspricht dem Modell: Eindimensionale Struktur auf beiden Ebenen							
Zweidimensional Auditiv-Schriftsprachlich	Zweidimensional Rezeption-Produktion	153.81	9	≤.01	0.97	83890.28	0.04	0.01	0.03
Zweidimensional Auditiv-Schriftsprachlich	Zweidimensional Auditiv-Schriftsprachlich	Entspricht dem Modell: Eindimensionale Struktur auf beiden Ebenen							

Beim Vergleich der Modellgütekriterien fällt zum einen auf, dass die Ergebnisse der Eindimensionalität nach exploratorischer Faktorenanalyse die Datenlage keineswegs am besten abbilden, wenn sie durch konfirmatorische Analysen den alternativen Modellen gegenüber gestellt werden. Zum anderen lässt sich feststellen, dass die zweidimensionalen Modelle mit den latenten Dimensionen der *auditiven* und der *schriftsprachlichen* Kompetenz im Vergleich den Daten weniger angemessen sind bzw. sich in ihrer Modellanpassung rechnerisch nicht von Modellen mit eindimensionaler Struktur unterscheiden lassen. Diejenigen beiden Modelle, welche auf Individualebene eine zweidimensionale Struktur mit den latenten Dimensionen der Sprachreflexion-Rezeption und der Sprachproduktion postulieren, zeigen eine bessere Modellanpassung als alle anderen Modelle, sofern auch auf der Klassenebene das gleiche Zusammenhangsmuster oder eine eindimensionale Struktur angenommen wird. Eine Entscheidung zwischen diesen beiden Modellen kann auf der Grundlage der Gütekriterien zugunsten des Modells fallen, das auf beiden Ebenen eine zweidimensionale Struktur mit den latenten Dimensionen der Sprachreflexion-Rezeption und der Sprachproduktion annimmt (vgl. Abbildung 17).



**Abbildung 17: Konfirmatorische Mehrebenen-Faktorenanalyse für die Sprachkompetenz im Englischen (N=8964,  $\chi^2=49.67$ ,  $df=8$ ,  $p \leq .01$ , CFI=.99, RMSEA=.03, SRMR<sub>within</sub>=.02, SRMR<sub>between</sub>=.00)**

Das dargestellte Modell weist auf beiden Ebenen die gleiche Faktorenstruktur mit jeweils zwei latenten Dimensionen auf.

*Für den Bereich des Englischen muss die Annahme unterschiedlicher struktureller Zusammenhänge auf Individual- und Klassenebene aus **Hypothese 5** verworfen werden.*

Die beiden latenten Faktoren Reflexion-Rezeption und Produktion, die sehr hoch miteinander korrelieren, können die differenziellen Zusammenhänge zwischen den fünf Testmodulen erklären. Das Modell passt besser als diejenigen Modelle, welche die Varianzen der Testmodule auf lediglich einen gemeinsamen Faktor zurückführen, auch wenn die beiden latenten Variablen besonders auf der Klassenebene hoch korrelieren. Daher bietet es sich an, bei einem mehrdimensionalen Modell zu bleiben, auch wenn die latenten Faktoren stark korreliert sind. Dies ermöglicht es, die Sprachstruktur im Deutschen wie im Englischen gemeinsam differenziert darzustellen, um gezielt allgemeine und spezifische Transfereffekte zwischen den produktiven und den reflexiv-rezeptiven Teilbereichen zu analysieren. Dies wäre bei Betrachtung eines eindimensionalen Modells nicht möglich.

### **6.3.2 Gemeinsames Mehrebenenmodell des Deutschen und des Englischen**

Das gemeinsame Mehrebenenmodell der Struktur der sprachlichen Kompetenzen im Deutschen und im Englischen integriert die jeweils separat geschätzten Zwei-Ebenen-Modelle in einem gemeinsamen Modell. Analysiert werden der allgemeine Zusammenhang zwischen der Sprachkompetenz im Deutschen und im Englischen, sowie im mehrdimensionalen Modell die spezifischen Pfade zwischen den unterschiedlichen Teilbereichen sprachlicher Kompetenz. Speziell interessieren dabei die gerichteten Zusammenhänge zwischen den latenten Dimensionen der sprachlichen Reflexion-Rezeption und der Produktion in beiden Sprachen.

***Hypothese 6:** Die Kompetenzen im Englischen lassen sich auf Individual- und auf Klassenebene durch die Kompetenzen im Deutschen vorhersagen.*

Diese Hypothese beruht zum einen auf Befunden über die Entwicklung von Sprachkompetenz in der Erst- und Fremdsprache, die speziell für die vorliegende Arbeit relevant sind und spezifische Transfereffekte zwischen sprachlichen Teilbereichen belegen (vgl. Abschnitt 2.3).

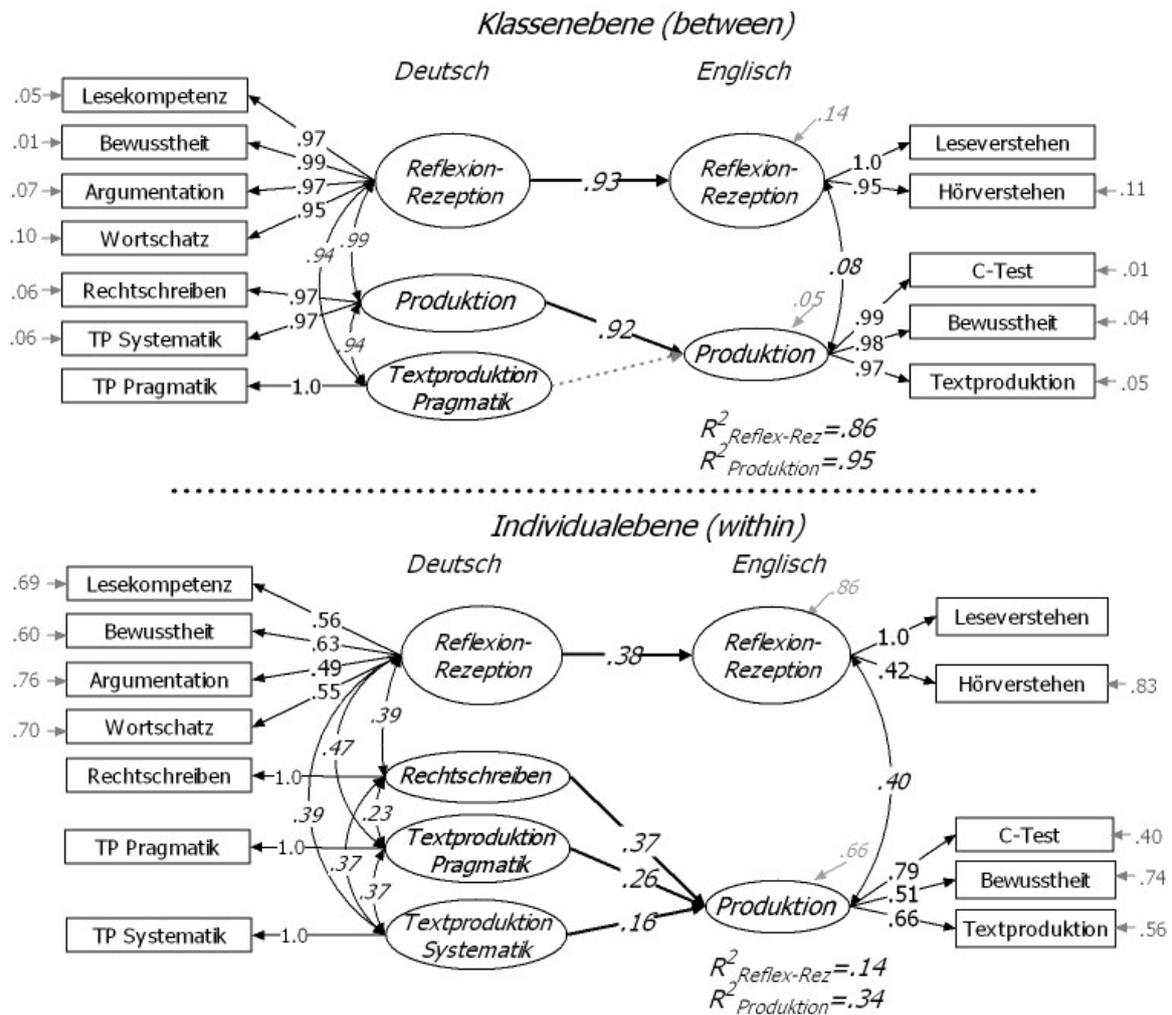
Zum anderen baut die institutionelle Vermittlung von Fremdsprachenkompetenz in der Regel grundlegend auf den bereits vorhandenen Kompetenzen in der Erstsprache auf, sodass von spezifischen Zusammenhängen zwischen den sprachlichen Teilbereichen auszugehen ist (vgl. Abschnitt 1.5).

Zur Überprüfung von Hypothese 6 werden zwei Strukturgleichungsmodelle analysiert, die unterschiedliche gerichtete Beziehungen zwischen den Kompetenzen im Deutschen und im Englischen annehmen:

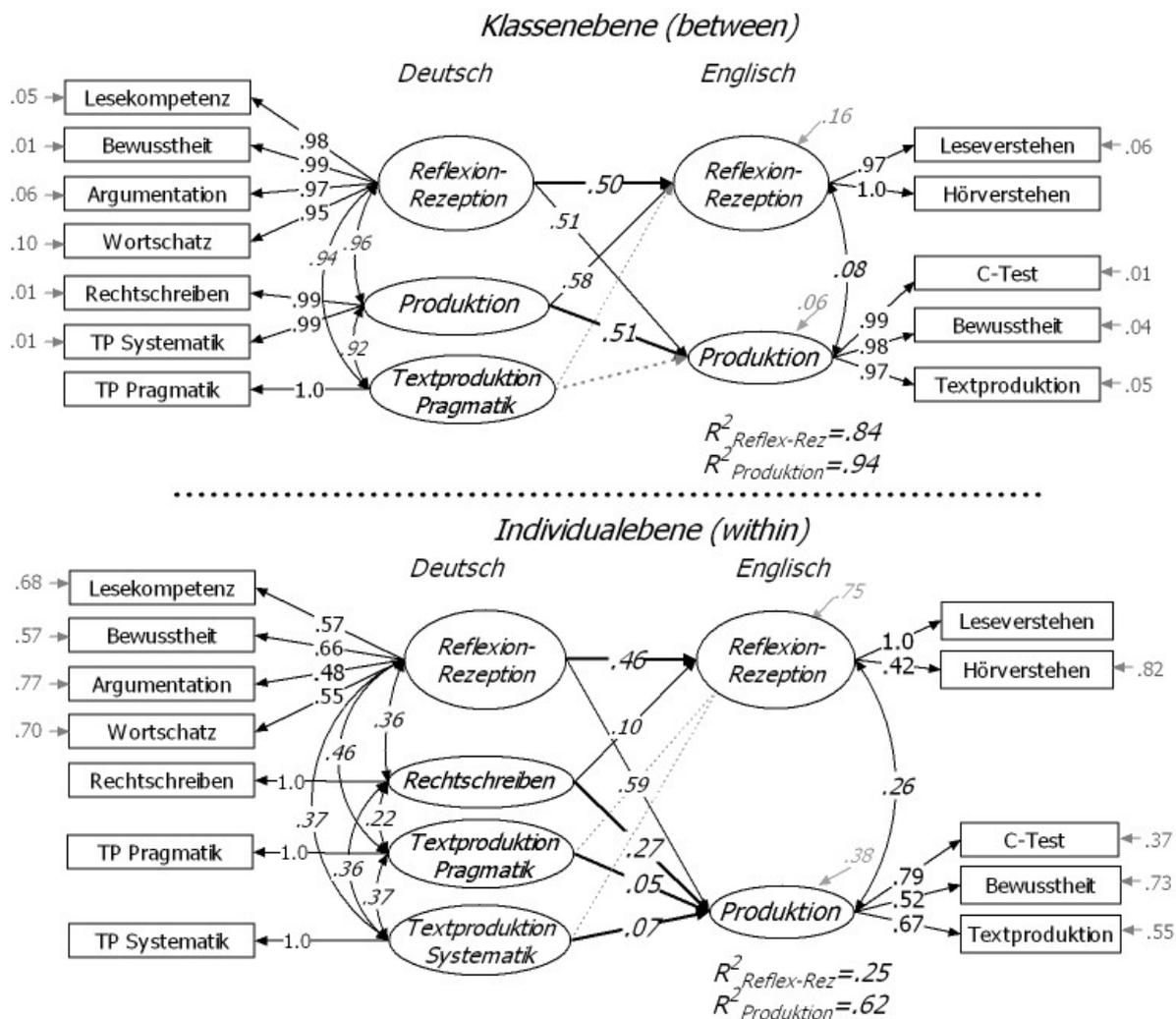
- Spezifische gerichtete Beziehungen zwischen den reflexiv-rezeptiven und den produktiven Kompetenzdimensionen im Deutschen und im Englischen.
- Gerichtete Beziehungen jeweils zwischen den vier Kompetenzdimensionen im Deutschen und den beiden Kompetenzdimensionen im Englischen.

Verglichen wird, wie viel Varianz die Leistung im Deutschen in den Teilbereichen im Englischen erklären kann und welches der beiden Modelle die Varianz in den empirischen Daten insgesamt besser abbilden kann. Beide Modelle sind im Vergleich in

Abbildung 18 und Abbildung 19 dargestellt.



**Abbildung 18: Spezifische gerichtete Zusammenhänge zwischen den reflexiv-rezeptiven und den produktiven Kompetenzen im Deutschen und im Englischen (N=8695,  $\chi^2= 2315.27$ ,  $df=96$ ,  $p \leq .01$ , CFI=.89, AIC= 212060.86, RMSEA=.05,  $SRMR_{\text{within}}=.09$ ,  $SRMR_{\text{between}}=.02$ )**



**Abbildung 19: Gerichtete Zusammenhänge zwischen den latenten Dimensionen im Deutschen und im Englischen, nicht signifikante Pfade sind mit gestrichelten Linien dargestellt (N=8695,  $\chi^2=1384.75$ ,  $df=89$ ,  $p \leq .01$ , CFI=.93, AIC= 210662.14, RMSEA=.04, SRMR<sub>within</sub>=.05, SRMR<sub>between</sub>=.02)**

Im Vergleich der beiden Modelle ist auf folgende Aspekte hinzuweisen: Die Modellanpassung an die Daten gelingt mit dem Modell, das gerichtete Zusammenhänge zwischen den Kompetenzen im Deutschen und *beiden* latenten Dimensionen im Englischen annimmt, besser. Die Varianzaufklärung in den abhängigen Variablen im Englischen unterscheidet sich in beiden Modellen jedoch nur auf Individualebene, hier kann durch die zusätzlichen Prädiktoren im Modell in Abbildung 19 mehr Varianz in den individuellen Leistungen in den Bereichen der Reflexion-Rezeption und der Produktion erklärt werden.

*Wie in **Hypothese 6** postuliert kann durch die Kompetenzen im Deutschen ein gewisser Anteil an Varianz der Leistungen im Englischen auf Individual- und auf Klassenebene vorhergesagt werden.*

Im ersten Modell (vgl.

Abbildung 18) erweisen sich auf der Individualebene alle postulierten Zusammenhänge zwischen den spezifischen Teilkompetenzen als signifikant, auf Klassenebene kann jedoch der Prädiktor Textproduktion Pragmatik keinen zusätzlichen erklärenden Beitrag zur Varianz des Faktors Produktion im Englischen leisten; von der Leistung im latenten Faktor Produktion im Deutschen allein kann sehr gut auf die entsprechende Leistung im Englischen geschlossen werden. In diesem Modell kann auf Klassenebene durch zwei Prädiktoren der Kompetenz im Deutschen der größte Teil der Kovarianzen zwischen den beiden abhängigen Variablen erklärt werden.

Die gleiche Erklärungskraft ist im Modell in Abbildung 19 für die Klassenebene zu beobachten: Der Faktor Produktion im Deutschen dient gleichermaßen als Prädiktor für die reflexiv-rezeptiven und die produktiven Kompetenzen im Englischen. Allerdings kann in diesem Modell der Faktor Textproduktion Pragmatik – anders als im Vergleichsmodell – einen eigenen Beitrag zur Varianzerklärung in der abhängigen Variable der Textproduktion Pragmatik leisten, nicht jedoch für die abhängige Variable der reflexiv-rezeptiven Kompetenz. Auf Klassenebene ist in diesem Modell zusätzlich auf den Einfluss der Sprachreflexion-Rezeption im Deutschen auf die Sprachproduktion im Englischen hinzuweisen. Wie bereits erwähnt wurde, gelingt die Erklärung der Varianz in den abhängigen Variablen auf Klassenebene in beiden Modellen praktisch gleich gut. Auf Individualebene gelingt es, durch die Hinzunahme von Prädiktoren mehr Varianz in den abhängigen Variablen zu erklären, dies gilt insbesondere für die Leistung im Faktor Produktion im Englischen. Durch die Hinzunahme des Prädiktors Sprachreflexion-Rezeption im Deutschen wird fast das Doppelte an Varianz in diesem Faktor erklärt. Umgekehrt kann lediglich der produktive Leistungsbereich der Textproduktion Systematik noch zusätzliche Varianz in der Reflexion-Rezeption im Englischen erklären.

### 6.3.3 Zusammenfassung der Ergebnisse der Mehrebenenanalysen

Die Methoden der Mehrebenenanalysen, welche explizit die hierarchische Schachtelung der Daten berücksichtigen, ermöglichen es, die Varianz zwischen Individuen einer Gruppe und zwischen Gruppen getrennt zu betrachten. Ziel solcher Analysen ist es, Effekte auf verschiedenen Ebenen genauer analysieren zu können. Dass diese Betrachtungsweise durchaus sinnvoll ist, zeigte sich in den Analysen zu Hypothese 3, die nach dem Einfluss der Gruppenzugehörigkeit auf die Zusammenhänge in den Daten fragte und den Schluss zulässt, dass Mehrebenenanalysen die Methode der Wahl sein sollten, wenn – wie in DESI – Schülerinnen und Schüler im Klassenverbund getestet werden und relevante Bedingungsfaktoren auf Klassenebene zu verorten sind. Bei differenzierter Betrachtung der Zusammenhänge auf Individual- und Klassenebene wird deutlich, dass die Zusammenhänge zwischen den sprachlichen Kompetenzen auf Individualebene bei Berücksichtigung der Mehrebenenstruktur von der Struktur her zwar prinzipiell gleich bleiben, jedoch insgesamt niedriger sind. Dies kann abermals als Hinweis darauf gedeutet werden, dass zwar bestimmte sprachliche Bereiche auf gemeinsame latente Dimensionen zurückführbar sind, diese jedoch differenziert zu betrachten sind und Sprache nicht als homogenes, eindimensionales Konstrukt betrachtet werden sollte. Vielmehr wurde analysiert, ob und wie die Strukturen von Sprachkompetenz sich zwischen der Individual- und der Klassenebene unterscheiden.

Auf der Grundlage von exploratorischen und konfirmatorischen Faktorenanalysen wurden sowohl für das Englische als auch für das Deutsche verschiedene Modelle verglichen: Im Deutschen zeigte sich dabei die bereits bekannte vierdimensionale Struktur auf Individualebene, während die Varianz auf Klassenebene in einem dreidimensionalen Modell besser abgebildet werden konnte, das die beiden Testmodule Rechtschreiben und Textproduktion Systematik auf einen gemeinsamen latenten Faktor zurückführte. Die Varianz in den Testmodulen im Englischen lässt sich durch das zweidimensionale Modell auf Individual- wie auch auf Klassenebene gut darstellen. Obwohl auf Klassenebene die latenten Dimensionen in beiden Sprachen sehr hoch korrelieren, erreichen die differenzierten Modelle eine bessere Anpassung als diejenigen, die für die Klassenebene ein homogenes, eindimensionales Kompetenzkonstrukt postulieren. Auch für die Klassebene scheint es daher sinnvoll, Sprachkompetenz in einem mehrdimensionalen Modell darzustellen und eben nicht mit einem aggregierten Messwert auszudrücken. Wie schon für die Gesamtstichprobe ist auch hier festzuhalten, dass diagnostische Aussagen über *Kompetenzen von Klassen* dann einen

Informationsmehrwert liefern, wenn hinsichtlich der reflexiv-rezeptiven und der produktiven Bereiche unterschieden wird. Diese Differenzierung bei Aussagen über das Niveau einer Klasse kann besonders im Vergleich zwischen Gruppen interessantere Aufschlüsse liefern, als dies durch einen Gesamtwert möglich wäre.

Unter testökonomischen Gesichtspunkten erscheint es bezogen auf das Deutsche für Aussagen über die Kompetenz von Klassen durchaus ausreichend, für die produktiv-systematischen Aspekte der Sprachkompetenz einen auf Gesamtwert zu verwenden, dessen Hauptmerkmal die Rechtschreibleistung ist. Ob diese in der formal eingegrenzten, sprachlichen Handlungssituation des Diktats erfasst wird, oder ob die Orthographie eines frei verfassten Texts zu bewerten ist, muss auf Klassenebene nicht mehr unterschieden werden. Jedoch ist deutlich hervorzuheben, dass ein Leistungswert für den Bereich der Sprachpragmatik in der freien Textproduktion zusätzliche differenzierte Information über Leistungsprofile der Schülerinnen und Schüler einer Klasse liefern kann. Dieser Faktor ist jedoch bei der Erklärung von Leistungen in der Fremdsprache auf Klassenebene zu vernachlässigen.

Auf der Grundlage der Mehrebenenmodelle für beide Sprachen wurde analysiert, inwieweit die Annahme spezifischer, gerichteter Transfereffekte zwischen den Dimensionen im Deutschen als Erstsprache und dem Englischen als Fremdsprache auf der Grundlage der vorliegenden Daten als zutreffend bezeichnet werden kann. Hierzu wurden zwei Modelle verglichen, die das reine Korrelationsmodell durch gerichtete Pfade zwischen den sprachlichen Teilbereichen auf beiden Ebenen erweiterten: Ein Modell nahm dabei lediglich Pfade zwischen den spezifischen Teilbereiche an, das Alternativmodell postulierte gerichtete Zusammenhänge zwischen allen Bereichen im Deutschen und im Englischen. Obwohl im zweiten Modell auf Individual- und auf Klassenebene nur jeweils zwei der zusätzlichen Pfade zwischen den nicht korrespondierenden Teilkompetenzen signifikant wurden, zeigt sich – unter Annahme von gerichteten Zusammenhängen zwischen allen Teilkompetenzen – insgesamt eine bessere Modellanpassung.

Die stärkste zusätzliche Erklärung der Varianz der produktiven Kompetenzen im Englischen auf Individual- und Klassenebene erreicht der Faktor der Reflexion-Rezeption im Deutschen. Dieser kann, wie bereits diskutiert wurde, als Repräsentation eines allgemeinen grundlegenden Sprachkompetenzfaktors in der Erstsprache angesehen werden. Als Indikator

für diese Annahme könnte der Umstand gewertet werden, das sich unter Einbeziehung dieses Faktors die Zusammenhänge zwischen den spezifischen produktiven Sprachkompetenzen im Deutschen und im Englischen zwar verringern, aber dennoch signifikant bleiben. Schriftsprachliche Kenntnisse in der Erstsprache können demnach, wie schon in anderen Studien berichtet wurde, durchaus als Grundlage für den erfolgreichen Erwerb einer Fremdsprache sowohl im reflexiv-rezeptiven als auch im produktiven Bereich angesehen werden (vgl. Abschnitt 1.6).

Weiterhin können durchaus differenzielle Transfereffekte zwischen den spezifischen sprachproduktiven Kompetenzbereichen angenommen werden, die auf Individual- wie auf Klassenebene existieren. Darüber hinaus lassen sich auf der Grundlage dieses Modells Einflüsse der Fähigkeiten der Textproduktion im Deutschen, speziell der Systematik, auf die Sprachreflexion-Rezeption im Englischen annehmen: Sowohl auf Individual- als auch auf Klassenebene kann dieser Bereich Varianz in der Englischleistung erklären. Der Schluss liegt nahe, dass Wissen über den systematischen Aufbau von Texten, welches sich im Deutschen im freien Verfassen von Texten zeigt, die reflexiv-rezeptive Erschließung von fremdsprachlichen Texten durchaus erleichtern kann (vgl. Abschnitt 2.3).

Insgesamt lassen sich für den Bereich der mehrbenenanalytischen Betrachtung von Sprachkompetenz folgende Aussagen treffen:

- Tendenziell sind die gleichen strukturellen Zusammenhänge auf Individual- und auf Klassenebene zu finden, welche bereits auf der Grundlage der Gesamtstichprobe dargestellt wurden.
- Die Strukturen auf Klassenebene sind in beiden Sprachen homogener, die latenten Dimensionen korrelieren hoch miteinander.
- Trotzdem weist ein Modell mit einer eindimensionalen Struktur auf Klassenebene für keine der beiden Sprachen eine bessere Anpassung an die Daten auf als die jeweils mehrdimensionalen Modelle.
- Die spezifischen Zusammenhänge zwischen korrespondierenden sprachlichen Bereichen in der Unterrichtssprache Deutsch und der Fremdsprache Englisch sind auf Individual- und auf Klassenebene zu betrachten.

- Zusätzlich zu den spezifischen Zusammenhängen zeigt sich insbesondere der Faktor der Sprachreflexion-Rezeption als wichtiger Prädiktor für die produktive Sprachkompetenz im Englischen auf beiden Ebenen
- Der Bereich der Textproduktion Systematik im Deutschen kann zusätzliche Varianz in der reflexiv-rezeptiven Sprachkompetenz im Englischen erklären.

Auch wenn diese Analysen darauf hindeuten, dass Zusammenhänge zwischen den beiden Sprachen nicht nur zwischen den spezifischen latenten Dimensionen bestehen, so sind doch gerade diese zu fokussieren, wenn es darum geht, spezifische Transfereffekte zwischen den beiden Sprachen zu analysieren. Dass die Sprachrezeption in der Erstsprache eine wesentliche Grundlage für alle weiteren sprachlichen Kompetenzen in anderen Teilbereichen oder auch in anderen Sprachen darstellt, zeigt sich in allen bisher dargestellten Modellen und ist ein aus der Literatur zur Sprachentwicklung bekanntes Phänomen. Im Folgenden interessieren daher die spezifischen Zusammenhänge zwischen den differenzierbaren latenten Dimensionen, um die Frage der differenziellen Transfereffekte nicht nur in der Gesamtstichprobe, sondern auch in einzelnen Gruppen zu analysieren.

## 6.4 Mehrgruppenmodelle

Wie bereits in Abschnitt 2.4 dargestellt, kann sich die Struktur von Sprachkompetenz, also die strukturellen Zusammenhänge zwischen sprachlichen Teilbereichen, durchaus unterscheiden, je nachdem, welche Personengruppen betrachtet werden. Die wichtigsten Bedingungsfaktoren, bei deren Berücksichtigung unterschiedliche Kompetenzstrukturen zu erwarten sind, sind das Geschlecht, der sprachliche Hintergrund und die Bildungserfahrung der Schülerinnen und Schüler (Wicherts, Dolan & Hessen, 2005; Swinton & Powers, 1980). Auch für die DESI-Stichprobe soll im Weiteren der Frage nach differenziellen Modellen für bestimmte Gruppen von Schülerinnen und Schülern nachgegangen werden.

Überprüft wird mit der Methode der Mehrgruppenanalyse, ob das für die Gesamtstichprobe gefundene Modell der Struktur von Sprachkompetenz für unterschiedliche Personengruppen bestätigt werden kann. Mehrgruppenmodelle sind eine spezielle Form von Strukturgleichungsmodellen (vgl. Abschnitt 5.4). Bei diesem Vorgehen wird vorausgesetzt, dass das jeweilige Modell mit seinen spezifischen strukturellen Zusammenhängen für die gesamte Population gültig ist; dies wird mittels unterschiedlich restriktiven Vergleichmodellen überprüft (Eid, 1999). Hierzu werden wie im einfachen Mehrgruppenmodell die Parameter über die Gruppen gleichgesetzt, und die Annahme überprüft, dass die festgesetzten Parameter in allen Gruppen gleich sind. Stimmt diese Annahme nicht mit den Strukturen der empirischen Daten überein, so sind sukzessive Parameter in den Gruppen freizusetzen. In den vorliegenden Analysen sind dies die Varianzen der latenten Faktoren, die Korrelationen zwischen den latenten Faktoren und die Pfade im Strukturgleichungsmodell. Je mehr Parameter freizusetzen sind, also in den Gruppen individuell geschätzt werden müssen, um eine gute Modellanpassung zu erreichen, desto stärker unterscheiden sich die Modelle und es kann nicht mehr davon ausgegangen werden, dass ein gemeinsames Strukturgleichungsmodell die Zusammenhänge in allen Subpopulationen zutreffend beschreibt.

Im Folgenden werden Mehrgruppenmodelle für folgende Gruppen analysiert:

- a) Mädchen und Jungen (weibliche und männliche Subgruppe),
- b) Schülerinnen und Schüler mit unterschiedlicher Erstsprache (deutscher Sprachhintergrund, bilingualer Sprachhintergrund und anderer Sprachhintergrund) sowie

## c) Schülerinnen und Schüler in bilingualen Unterrichtsangeboten.

Ausgangspunkt für alle Analysen ist dabei das für die Gesamtstichprobe gültige Mehrebenenmodell der Struktur der Sprachkompetenz mit gerichteten Zusammenhängen zwischen den spezifischen sprachlichen Teilkompetenzen im Deutschen und im Englischen (vgl.

Abbildung 18).

Die in dieser Arbeit verwendeten Mehrgruppenmodelle betrachten dabei ausschließlich die Zusammenhänge auf Individualebene bei gleichzeitiger Korrektur der Standardfehler (TYPE=Complex). Dies geschieht aus mehreren Gründen: Für die Modelle a) zu Geschlechtsunterschieden und b) zu unterschiedlichen Erstsprachen ist auf Klassenebene keine sinnvolle Gruppenbildung möglich. Das Geschlecht als individuelles konstantes Merkmal kann höchstens im quantitativen Vergleich auf Klassenebene interpretiert werden – Klassen mit einem hohen Anteil an Mädchen weisen einen größeren Lernfortschritt auf – Mädchen und Jungen befinden sich in der Regel jedoch eben gemeinsam in den gleichen Klassen, eine Unterteilung auf Klassenebene zwischen diesen Gruppen wäre hier künstlich und inhaltlich nicht interpretierbar. Ein entsprechendes Modell würde Aussagen über die Struktur von Sprachkompetenz der Mädchen respektive der Jungen auf Klassenebene differenzieren, die in der Realität so nicht vorliegen. Noch deutlicher wird dieses Phänomen für Modell b) der unterschiedlichen Erstsprachen, hier verteilten sich vereinzelte Schülerinnen und Schüler mit unterschiedlichen Herkunftssprachen auf die Klassen, so dass spezifische strukturelle Zusammenhänge bei Einzelnen in keinsten Weise als gemeinsames Phänomen des Unterrichts interpretiert werden können. Für beide Modelle werden daher die strukturellen Zusammenhänge auf Individualebene betrachtet.

Anders verhält es sich mit den Analysen zum Vergleich zwischen Schülerinnen und Schülern, die den Regelunterricht im Englischen besuchen und solchen, die ein bilinguales Unterrichtsangebot im Englischen wahrnehmen. Hier existieren zwar vergleichbare Klassen – jedoch nur für das Englische. Die Stichprobe ist so aufgebaut, dass jeweils zwei Deutschklassen in den teilnehmenden Schulen gezogen wurden. Die Schülerinnen und Schüler in diesen Klassen besuchten in der Regel auch den Englischunterricht im gleichen Klassenverbund. In vielen der teilnehmenden Schulen mit bilingualem Unterrichtsangebot jedoch nahmen die Schülerinnen und Schüler in den Deutschklassen nicht alle am bilingualen

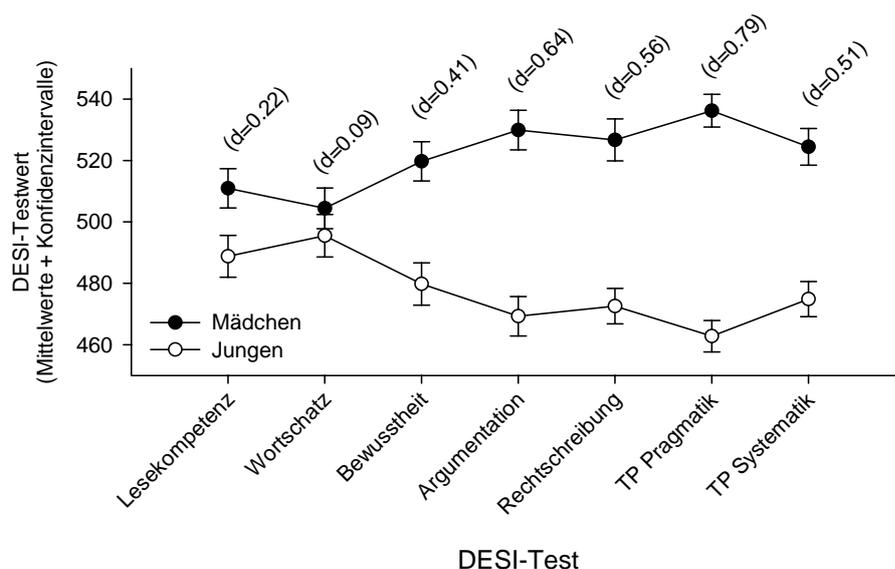
Unterrichtsangebot teil, das heißt sie wurden nicht im gleichen Klassenverbund im Englischen unterrichtet: Einige Schülerinnen und Schüler besuchten das bilinguale Unterrichtsangebot, andere den Regelunterricht im Englischen. Es ist daher auf Klassenebene kein Abgleich zwischen diesen Klassen möglich, auch hier beschränkt sich die Analyse auf die Betrachtung der Individualebene. Trotzdem sind gerade in diesem Fall beim Vergleich des bilingualen Unterrichtsangebots mit dem Regelunterricht im Englischen interpretative Rückschlüsse auf den Unterricht möglich und sinnvoll, wenn davon auszugehen ist, dass individuelle Kompetenz durch den Unterricht beeinflusst wird.

#### **6.4.1 Geschlechtsunterschiede**

Wie bereits in Abschnitt 1.7 dargestellt wurde, wird in vielen aktuellen Leistungsstudien über unterschiedliche Leistungsniveaus von Jungen und Mädchen in den sprachlichen Fächern berichtet, meist weisen die Mädchen dabei insgesamt bessere Leistungen auf (Deutsches PISA-Konsortium 2001; OECD 2004; PISA-Konsortium Deutschland 2004). Auf der Grundlage der DESI-Daten lassen sich mittels Mehrebenen-Strukturgleichungsmodelle diese beiden Gruppen dahingehend vergleichen, ob sich die Zusammenhänge zwischen den Teilkompetenzen unterscheiden.

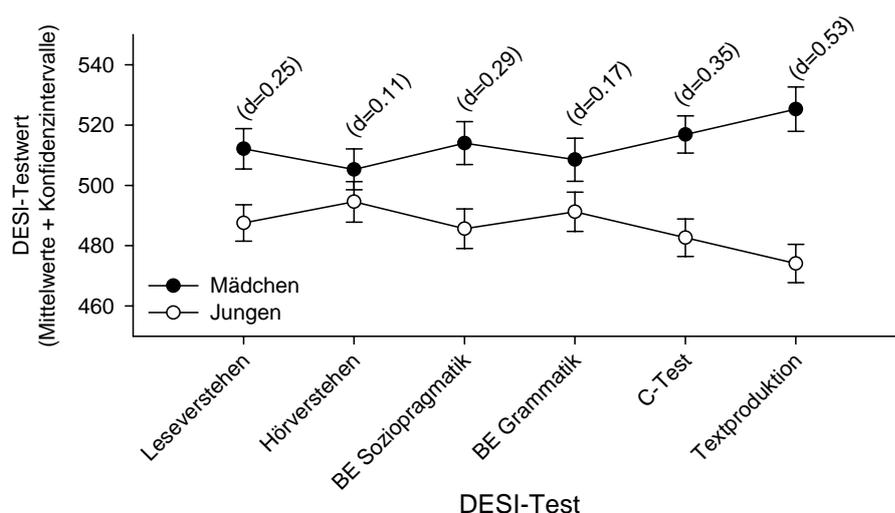
***Hypothese 7:** Die Struktur von Sprachkompetenz im Deutschen und im Englischen unterscheidet sich zwischen Mädchen und Jungen.*

Ein erster Vergleich zwischen den beiden Gruppen zeigt, dass sich die mittleren Leistungen in den sprachlichen Teilkompetenzen zwischen Mädchen und Jungen zum Teil deutlich unterscheiden (vgl. Abbildung 20).



**Abbildung 20: Leistungen für Mädchen und Jungen in den Teilkompetenzen im Deutschen Ende der neunten Jahrgangsstufe (TP = Textproduktion) (Quelle: Hartig & Jude, 2008, S. 203)**

Die Mädchen weisen in Ihren Leistungen in fast allen sprachlichen Teilbereichen im Deutschen einen statistisch bedeutsamen Vorsprung vor den Jungen auf (vgl. Hartig & Jude, 2008). Diese Differenz ist im Wortschatz am geringsten, in der Textproduktion Pragmatik am größten. Auch im Englischen zeigen die Schülerinnen durchgehend bessere Leistungen als die Schüler, allerdings ist der Leistungsunterschied insgesamt geringer.



**Abbildung 21: Leistungen für Mädchen und Jungen in den Teilkompetenzen im Englischen Ende der neunten Jahrgangsstufe (BE = Bewusstheit) (Quelle: Hartig & Jude, 2008, S. 204)**

Während die Unterschiede in der Englischleistung im Hörverstehen praktisch zu vernachlässigen sind, zeigen sich die größten Differenzen auch im Englischen im Bereich der produktiven Komponenten der Textproduktion und des C-Tests. Insgesamt lässt sich also feststellen, dass die Leistungen in den Tests zur Sprachproduktion in beiden Sprachen bei den Mädchen deutlich besser ausfallen als bei den Jungen, sich aber auch in den Test zu den reflexiv-rezeptiven Kompetenzen deutliche Unterschiede im Englischen wie auch im Deutschen zugunsten der Mädchen finden lassen.

Diese Mittelwertsunterschiede alleine müssen jedoch nicht bedeuten, dass sich auch die strukturellen Zusammenhänge innerhalb der beiden sprachlichen Bereiche zwischen den Gruppen unterscheiden. Dieser Frage wird im Folgenden mittels einer Mehrebenen-Mehrgruppen-Analyse nachgegangen. Die allgemeine Modellstruktur entspricht dabei dem für die Gesamtstichprobe angenommenen Modell mit gerichteten Beziehungen zwischen den spezifischen sprachlichen Teilbereichen (vgl. Abschnitt 6.2.5).

Verglichen werden vier Modelle, die jeweils unterschiedlich starke restriktive Annahmen treffen. Im ersten Modell werden im Strukturgleichungsmodell die Varianzen der latenten Faktoren, deren Interkorrelationen sowie die Pfade zwischen den Prädiktoren (UV) und den abhängigen Variablen (AV) zwischen beiden Gruppen gleichgesetzt. Im zweiten Modell wird angenommen, dass die Varianz der latenten Faktoren und der Zusammenhang zwischen den Prädiktoren und den abhängigen Variablen in beiden Gruppen gleich sind. Im dritten Modell schließlich werden lediglich die Pfade zwischen UV und AV zwischen den Gruppen gleichgesetzt, während im vierten Modell alle Varianzen frei geschätzt werden. Bei allen diesen Modellen wird davon ausgegangen, dass die latenten Variablen in beiden Gruppen in gleicher Weise durch die manifesten Konstrukte repräsentiert werden. Die Modellanpassungen aller vier Modelle sind in

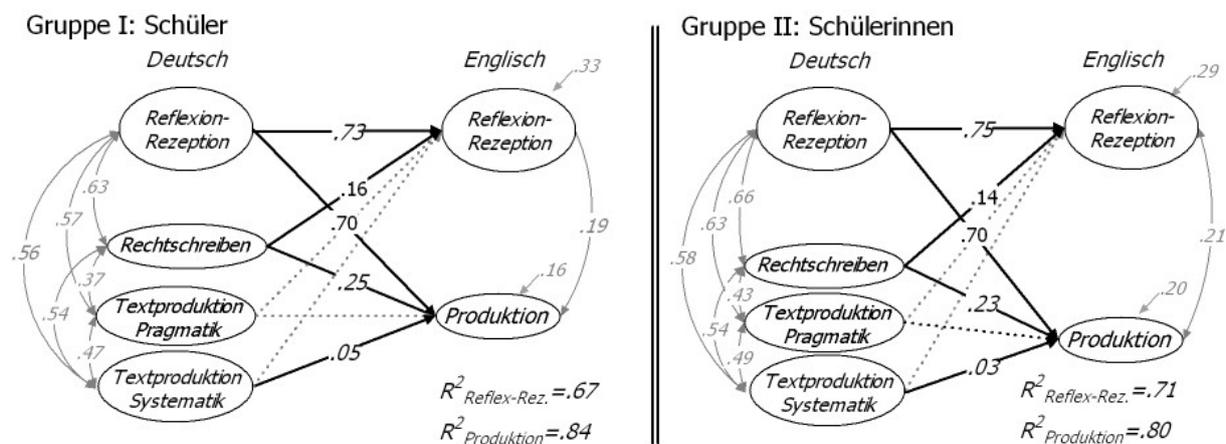
Tabelle 13 dargestellt.

**Tabelle 13: Modellanpassung des Mehrgruppenvergleichs zwischen Schülerinnen und Schülern in unterschiedlich stark restringierten Modellen**

Modell	$\chi^2$	df	p	CFI	AIC	RMSEA	SRMR
1) Latente Korrelationen, Faktorvarianz und Pfade fixiert	1324.26	114	$\leq .01$	0.98	220414.96	0.05	0.03
2) Faktorvarianz und Pfade fixiert	1317.80	107	$\leq .01$	0.98	220408.56	0.05	0.03
3) Pfade fixiert	1310.62	104	$\leq .01$	0.98	220379.90	0.05	0.03
4) Strukturmodell frei geschätzt	1320.49	96	$\leq .01$	0.98	220381.10	0.05	0.03

Im Modellvergleich wird ersichtlich, dass alle vier Modelle sehr gute Anpassungen an die Daten aufweisen. Modell 3), welches lediglich die Pfade zwischen den latenten Dimensionen im Deutschen und im Englischen in beiden Gruppen gleichhält, weist gemessen am Chi-Quadrat/df-Verhältnis und dem AIC sogar die beste Modellanpassung auf.

Zur Analyse etwaiger vorhandener Unterschiede in den Zusammenhängen zwischen den beiden Sprachen, dem Deutschen und dem Englischen, ist Modell 4 mit frei geschätztem Strukturmodell in Abbildung 22 dargestellt.



**Abbildung 22: Mehrgruppen-Mehrebenen-Strukturgleichungsmodelle (Modell 4) für die beiden Gruppen der Mädchen und Jungen, dargestellt ist nur das Strukturmodell ( $N_m=3975$ ,  $N_w=4720$ ,  $\chi^2=1320.49$ ,  $df=96$ ,  $p \leq .01$ ,  $CFI=.98$ ,  $RMSEA=.05$ ,  $SRMR=.03$ )**

Wie schon auf Grund der oben dargestellten Modellanpassungen zu vermuten war, lassen sich unter der Annahme allgemein gleicher Strukturen in beiden Gruppen *nur noch geringe Unterschiede* zwischen den Gruppen der Schülerinnen und der Schüler finden: Für die Gruppe der Schülerinnen zeigt sich ein höherer Zusammenhang zwischen *allen* Teilbereichen in der deutschen Sprache, und auch die beiden latenten Faktoren im Englischen weisen unter

Berücksichtigung aller Prädiktoren in dieser Gruppe noch einen etwas höheren Zusammenhang auf, der nicht auf die Sprachkompetenzen im Deutschen zurückgeführt werden kann. Die Zusammenhänge zwischen den sprachlichen Teilkompetenzen jeweils im Deutschen und im Englischen sind in der Gruppe der Jungen geringfügig niedriger.

Auch die Zusammenhänge zwischen dem Deutschen und dem Englischen unterscheiden sich zwischen den Gruppen – die Varianz der reflexiv-rezeptiven Sprachkompetenzen im Englischen kann in der Gruppe der Mädchen zu einem größeren Teil durch die Sprachkompetenzen im Deutschen erklärt werden, während in der Gruppe der Jungen die produktiven Leistungen im Englischen einen stärkeren Bezug zu den Sprachkompetenzen im Deutschen aufweisen.

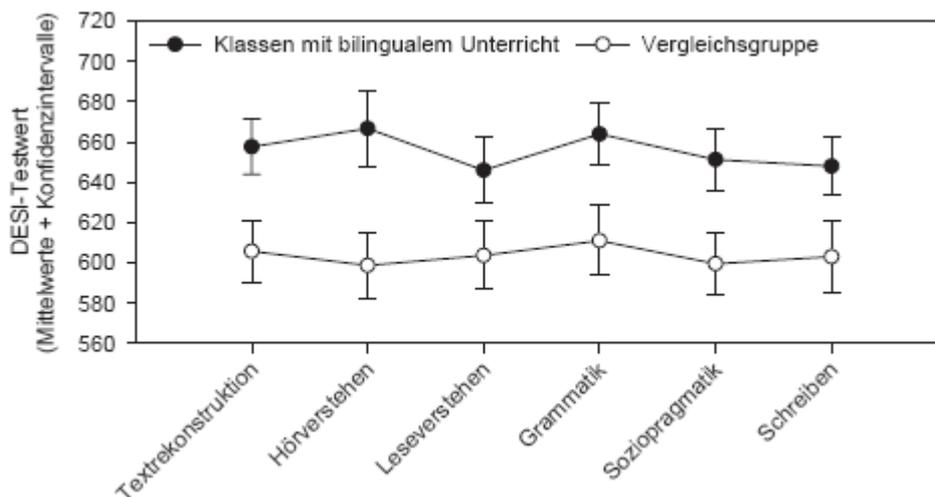
*Zur Annahme in **Hypothese 7** kann festgestellt werden, dass sich die grundlegende Struktur der Sprachkompetenzen im Deutschen und im Englischen mit den latenten Dimensionen der rezeptiv-reflexiven und der produktiven Sprachkompetenz nicht zwischen den Gruppen der Mädchen und der Jungen unterscheidet.*

*Marginale Unterschiede bestehen lediglich in den Zusammenhängen zwischen dem Deutschen und dem Englischen.*

#### **6.4.2 Unterschiede zwischen Schülerinnen und Schülern in bilingualen und nicht-bilingualen Bildungsgängen**

In DESI wurden zusätzlich zu den für Deutschland als repräsentativ gezogenen Schulen und Schulklassen solche Klassen getestet, in denen Schülerinnen und Schüler am bilingualen Sachfachunterricht in Englisch teilnahmen (vgl. Abschnitt 3.3.5). Insgesamt wurden in DESI N=1233 Schülerinnen und Schüler getestet, die in verschiedenen Fächern und mit unterschiedlicher Wochenstundenzahl bilingual unterrichtet wurden. Dieses Unterrichtsangebot verwendet das Englische als Unterrichtssprache in Sachfächern, wobei allerdings der Gegenstand des Sachfaches im Mittelpunkt steht, nicht das Fremdsprachenlernen. Trotzdem ist davon auszugehen, dass Schülerinnen und Schüler, die den bilingualen Unterricht besuchen, bessere Leistungen in der Fremdsprache zeigen als Schülerinnen und Schüler in nicht-bilingualen Unterrichtsgängen (vgl. Abschnitt 1.6). Zu entsprechenden Ergebnissen kommen die diesbezüglichen Analysen im Rahmen von DESI,

welche die Englischleistungen von Schülerinnen und Schülern in beiden Unterrichtsformen vergleichen (Nold, Hartig, Hinz & Rossa 2008) (vgl. Abbildung 23).



**Abbildung 23: Mittlere Englisch-Kompetenzen zum Ende der neunten Jahrgangsstufe in Klassen mit bilinguaalem Unterricht und in einer hinsichtlich Deutsch Gesamtleistung, Bildungsgang, sozioökonomischem Status, kognitiver Grundfähigkeit, Erstsprache und Geschlecht parallelen Vergleichsgruppe (Quelle: Nold, Hartig, Hinz & Rossa 2008, S. 455)**

Für die in dieser Arbeit zu analysierenden Fragestellungen ist im Weiteren zu untersuchen, ob sich nicht nur die absoluten Kompetenzwerte der beiden Gruppen unterscheiden – bilingual unterrichtete Schülerinnen und Schüler sind wie ersichtlich in allen Teilbereichen des Englischen besser als die Vergleichsgruppe –, sondern ob von einer vergleichbaren Struktur der Sprachkompetenz im Englischen in beiden Gruppen ausgegangen werden kann. Aufgrund der höheren Kompetenz in allen sprachlichen Teilbereichen könnten die bilingual unterrichteten Schülerinnen und Schüler eine homogenere Kompetenzstruktur im Englischen aufweisen als die Gruppe der nicht-bilingual Unterrichteten, wenn davon auszugehen wäre, dass sich die Struktur mit zunehmender Gesamtkompetenz verändert (vgl. Abschnitt 1.8). Zusätzlich ist zu analysieren, ob der Zusammenhang zwischen den Sprachen Deutsch und Englisch in den beiden Gruppen jeweils unterschiedlich ist. Unterschiede in den Zusammenhängen in einer der beiden Gruppen könnten auf Unterschiede im Spracherwerbsprozess hinweisen oder auf allgemeine Erwerbsbedingungen, welche Kompetenz über beide Sprachen hinweg vermitteln, fördern und integrieren.

**Hypothese 8:** *Die Struktur von Sprachkompetenz im Deutschen und Englischen unterscheidet sich zwischen den Schülerinnen und Schülern, die ein bilinguales Unterrichtsangebot wahrnehmen, und denjenigen, die den Regelunterricht in der Fremdsprache besuchen*

Dieser Frage wird methodisch ebenfalls mittels Mehrgruppenmodellen nachgegangen. Allerdings werden in diesem Fall nicht alle Schülerinnen und Schüler in bilingualen Unterrichtsangeboten mit dem Rest der Stichprobe verglichen, sondern es werden sogenannte Vergleichsklassen ausgewählt, die sich hinsichtlich Bildungsgang und sozialen Hintergrundmerkmalen ähneln. Dies erscheint notwendig, weil die bilingualen Unterrichtsangebote sich in der Stichprobe auf die Gymnasien und wenige Realschulen beschränken. Es ist davon auszugehen, dass diese Einschränkung einhergeht mit weiteren Merkmalen wie beispielsweise dem sozioökonomischen Hintergrund der Schülerinnen und Schüler, die bilinguale Unterrichtsangebote wahrnehmen. Durch die Bildung von Vergleichsgruppen wird die Stichprobe geringfügig reduziert. Verglichen wird die Struktur der Sprachkompetenz von 893 Schülerinnen und Schülern im bilingualen Englischunterricht mit der von 882 Schülerinnen und Schülern im regulären Englischunterricht.

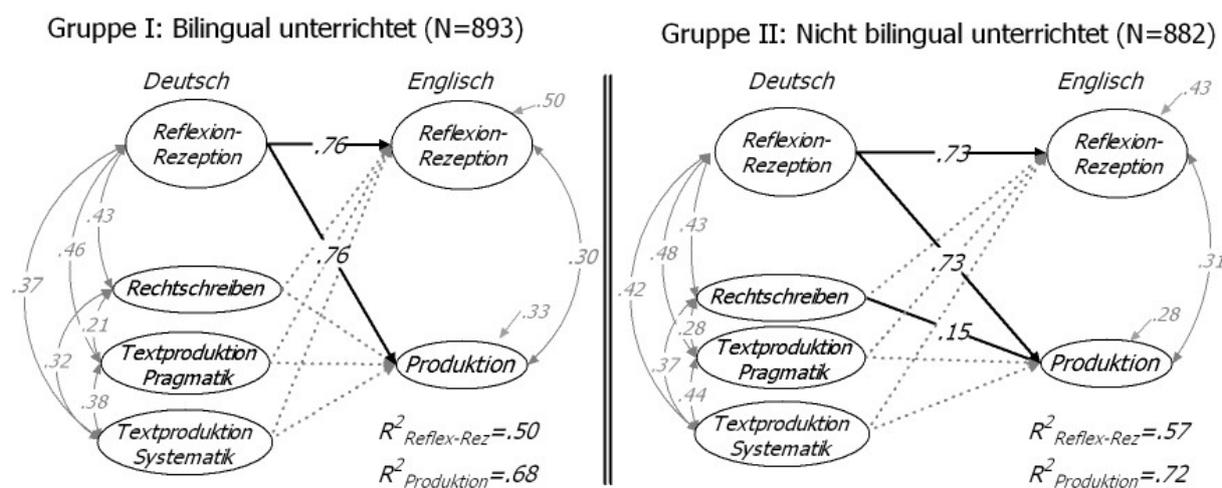
In einem ersten Schritt wird ein Modell analysiert, das eine gleiche Zusammenhangsstruktur in beiden Gruppen annimmt. Diese Struktur entspricht dem Modell, das für die Gesamtstichprobe angepasst werden konnte. Überprüft wird die Annahme, dass diese Struktur sich in beiden Gruppen – der bilingual unterrichteten und der Vergleichsgruppe – gleichermaßen auf der Grundlage der empirischen Daten abbilden lässt. Hierzu werden die unstandardisierten Koeffizienten zwischen den Gruppen restringiert. Dies erlaubt die Analyse potenziell bestehender Unterschiede in den Zusammenhängen zwischen den sprachlichen Teilbereichen in den beiden Gruppen auf Grund der standardisierten Regressionskoeffizienten und Korrelationskoeffizienten. Schrittweise werden die latenten Korrelationen der Faktoren, ihre Varianzen und schließlich auch die Regressionspfade freigegeben, so dass diese alternativen Modelle verglichen werden können (vgl.

Tabelle 14).

**Tabelle 14: Modellanpassung des Mehrgruppenvergleichs zwischen Schülerinnen und Schülern in bilingualen und nicht-bilingualen Unterrichtsangeboten**

Modell	$\chi^2$	df	P	CFI	AIC	RMSEA	SRMR
1) Latente Korrelationen, Faktorvarianz und Pfade fixiert	1157.04	114	$\leq .01$	0.99	221622.72	0.05	0.10
2) Faktorvarianz und Pfade fixiert	1091.81	107	$\leq .01$	0.99	221570.98	0.05	0.09
3) Pfade fixiert	985.82	104	$\leq .01$	0.99	221317.67	0.04	0.03
4) Strukturmodell frei geschätzt	962.36	96	$\leq .01$	0.99	221285.84	0.05	0.03

Auch im Vergleich dieser Modelle zeigt sich ähnlich wie bei den Analysen zu den Geschlechtsunterschieden, dass die Annahme gleicher Sprachstrukturen auf der Grundlage der Modellanpassungen in Modell 1 durchaus berechtigt ist. Lediglich der SRMR ist in diesem Modell und auch in Modell 2 zu hoch. Die beste Modellanpassung wird erreicht, wenn bei Annahme gleicher latenter Strukturen die Zusammenhänge zwischen den beiden Sprachen Deutsch und Englisch in den Gruppen frei geschätzt werden (Modell 4, vgl. Abbildung 24).



**Abbildung 24: Mehrgruppen-Strukturmodelle mit frei geschätzten Parametern für die beiden Gruppen der bilingual und nicht-bilingual unterrichteten Schülerinnen und Schüler, dargestellt ist nur das Strukturmodell ( $N_{\text{bil}}=893$ ,  $N_{\text{norm}}=882$ ,  $\chi^2=962.3$ ,  $df=96$ ,  $p \leq .01$ ,  $CFI=.99$ ,  $RMSEA=.05$ ,  $SRMR=.03$ )**

Im Modell mit frei geschätzten Regressionszusammenhängen zwischen dem Deutschen und dem Englischen zeigt sich ein wesentlicher Unterschied zwischen beiden Gruppen: In der bilingual unterrichteten Gruppe können die produktiven Sprachkompetenzen im Deutschen bei gleichzeitiger Berücksichtigung aller drei produktiven Prädiktoren und dem Faktor Sprachreflexion-Rezeption keinen signifikanten Anteil der Varianz der Sprachproduktion im

Englischen erklären. Für Schülerinnen und Schüler mit Regelunterricht im Englischen hingegen kann der Prädiktor der Rechtschreibung im Deutschen signifikant einen Anteil der Varianz der Sprachproduktionsleistung im Englischen aufklären. Vergleichbare Zusammenhänge bestehen in beiden Gruppen zwischen dem Faktor Sprachrezeption im Deutschen und der Rezeption und Produktion im Englischen.

Besonders hinsichtlich der sprachproduktiven Kompetenzen im Englischen zeigen sich jeweils differenzielle Zusammenhangsmuster in den beiden Gruppen: Während die Leistungen in der Rechtschreibung Deutsch und der Sprachproduktion im Englischen bei Schülerinnen und Schülern im Regelunterricht systematische gemeinsame Variationen aufweisen, ist dies in der bilingual unterrichteten Gruppe nicht der Fall. Es könnte vermutet werden, dass sprachproduktive Aspekte zu denen auch das Regelwissen gehört – in diesem Fall die Rechtschreibung – in der bilingual unterrichteten Gruppe als sprachspezifische Phänomene vermittelt und reflektiert werden. Vorhandene Kompetenzen der Sprachproduktion in der Erstsprache werden dann weniger stark transferiert, vielmehr werden die sprachsystematischen und sprachpragmatischen Eigenheiten der Fremdsprache bei der eigenen Textproduktion in Betracht gezogen.

***Hypothese 8** wird teilweise durch die empirischen Befunde gestützt: Die grundlegende mehrdimensionale Struktur von Sprachkompetenz mit latenten Dimensionen der reflexiv-rezeptiven und der produktiven Sprachkompetenz ist zwischen Schülerinnen und Schülern, die ein bilinguales Unterrichtsangebot wahrnehmen, und denjenigen, die den Regelunterricht in der Fremdsprache besuchen vergleichbar. Unterschiede zwischen den Gruppen zeigen sich im Anteil der Varianz im Englischen, der durch die Deutschleistung erklärt werden kann.*

Um einer weiteren möglichen Erklärung für die unterschiedlichen Zusammenhänge zwischen dem Deutschen und dem Englischen nachzugehen, wurde der Sprachhintergrund der Schülerinnen und Schüler in den beiden Gruppen verglichen. Hier sind keine signifikanten Unterschiede zwischen den Gruppen festzustellen (vgl. Tabelle 15).

**Tabelle 15: Kreuztabelle der Erstsprache und der Unterrichtsbedingungen in der Gesamtstichprobe**

Teilnahme am bilingualen Unterricht	Erstsprache			Gesamt
	Deutsch	Deutsch und andere	Andere als Deutsch	
nein	81.5%	6.0%	12.4%	100.0%
ja	85.4%	7.2%	7.4%	100.0%

Die beiden Gruppen unterschieden sich hinsichtlich der Erstsprachen der Schülerinnen und Schüler nicht wesentlich, lediglich der Anteil jener, die Deutsch nicht als Erstsprache angeben, ist in der bilingual unterrichteten Gruppe niedriger. Bilinguale Schülerinnen und Schüler finden sich hingegen nur unwesentlich häufiger in der bilingual unterrichteten Gruppe.

Das für die Gesamtstichprobe passende Modell der Struktur von Sprachkompetenz mit der Unterscheidung der rezeptiv-reflexiven und der produktiven sprachlichen Teilbereiche lässt sich also sowohl für die beiden Gruppen der Mädchen und der Jungen, als auch für die beiden Gruppen der bilingual Unterrichteten und der nicht-bilingual Unterrichteten anhand der empirischen Daten insgesamt stützen. Zwischen den Gruppen können dabei jedoch spezifische Unterschiede festgestellt werden, die sich vor allem in den unterschiedlich starken Zusammenhängen zwischen den Kompetenzen im Deutschen und im Englischen zeigen, die auf unterschiedliche Transfereffekte hinweisen. Im Folgenden soll in einem dritten Schritt analysiert werden, ob sich das bisher bestätigte Modell der Struktur von Sprachkompetenz auch für die Gruppen anpassen lässt, welche eine andere Erstsprache haben als Deutsch.

### 6.4.3 Strukturelle Unterschiede in verschiedenen Sprachgruppen

Neben den möglichen Unterschieden in der Struktur von Sprachkompetenz, die sich auf Grund des Geschlechts oder der Unterrichtsbedingungen ergeben, stellt sich die vielleicht noch interessantere Frage, ob es strukturelle Unterschiede in den Sprachkompetenzen gibt, die durch die Erstsprache bedingt sind.

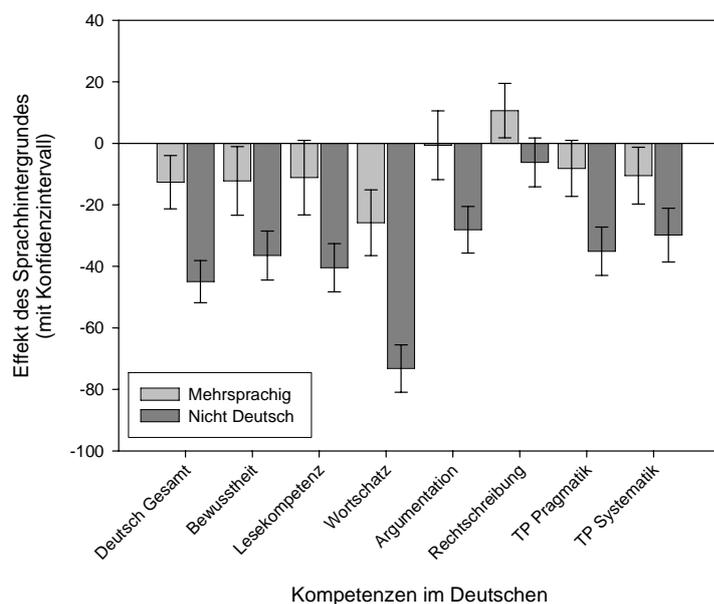
**Hypothese 9:** *Die Struktur von Sprachkompetenz im Deutschen und im Englischen unterscheidet sich zwischen Schülerinnen und Schülern mit deutscher Erstsprache und denjenigen mit einer anderen Erstsprache*

Strukturelle Unterschiede könnten dabei in den folgenden Bereichen vorliegen:

- in der Struktur des Deutschen, das besonders für die dritte genannte Gruppe eine Fremdsprache darstellt,
- in der Struktur des Englischen, das für die deutschen Schülerinnen und Schüler in der Regel die erste Fremdsprache, für die anderen Gruppen jedoch die zweite oder dritte Fremdsprache sein kann, und
- in den Zusammenhängen zwischen den beiden Sprachen, die sich zwischen den Gruppen durchaus unterscheiden könnten.

In der DESI-Stichprobe wurden die Schülerinnen und Schüler nach ihrer Erstsprache gefragt, dabei gaben 74 % Deutsch, 5,5 % Deutsch und eine andere Sprache (deutsch-bilingual) sowie 12 % eine andere Sprache als Deutsch als ihre Erstsprache an. 8,5 % der befragten Schülerinnen und Schüler gaben ihre Erstsprache nicht an. Im Vergleich der reinen Leistungsmittelwerte erreichen diejenigen Schülerinnen und Schüler, deren Erstsprache Deutsch ist, das höchste Ergebnis, während die Leistungen der Deutsch-Bilingualen signifikant geringer sind als die der monolingual Deutschsprachigen, aber signifikant höher als die der Schülerinnen und Schüler mit nicht-deutscher Erstsprache (Hesse, Göbel & Hartig, 2008).

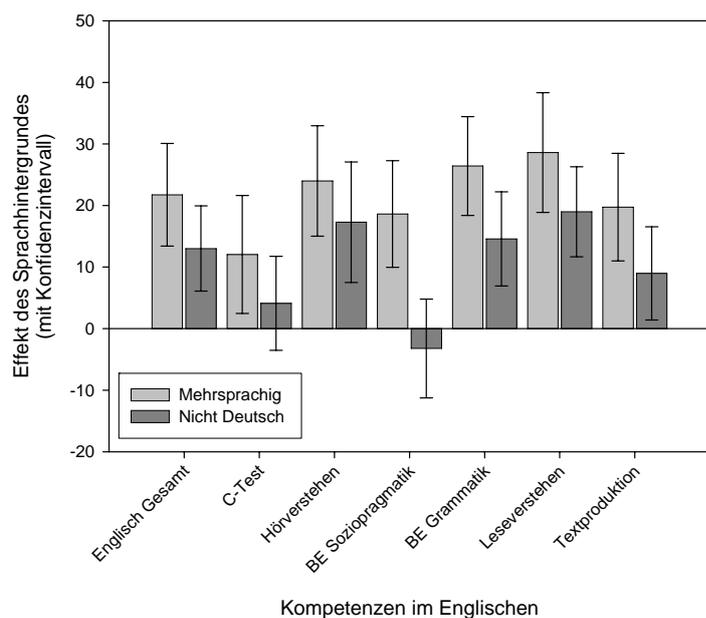
Bei differenzierter Betrachtung der unterschiedlichen Testmodule fallen darüber hinausgehende interessante Effekte auf, die bei der Analyse der strukturellen Zusammenhänge der Sprachkompetenz zwischen den Gruppen zu berücksichtigen sind (vgl. Abbildung 25).



**Abbildung 25: Effekte des sprachlichen Hintergrundes auf Teilkompetenzen im Deutschen zum Ende der neunten Jahrgangsstufe unter Kontrolle von Bildungsgang, sozioökonomischem Hintergrund, kognitiver Grundfähigkeit und Geschlecht: Kontraste der Leistungen von Mehrsprachigen oder monolingual nicht-deutsch Erstsprachigen gegenüber monolingual deutsch Erstsprachigen (TP = Textproduktion) (Quelle: Hesse, Göbel & Hartig, 2008, S. 217)**

Unter Kontrolle relevanter Bedingungsfaktoren zeigen sich die dargestellten Unterschiede in der Leistung zwischen den drei Gruppen mit unterschiedlichem Sprachhintergrund auch für die einzelnen Teilkompetenzen im Deutschen: Die Leistungen der deutsch-bilingualen Schülerinnen und Schüler, hier als mehrsprachig bezeichnet, sind in fast allen Bereichen schlechter als die derjenigen mit Deutsch als Erstsprache – aber in allen Bereichen besser als die Leistungen der Schülerinnen und Schüler mit einer anderen Erstsprache als Deutsch. In der Rechtschreibung weisen die Bilingualen sogar tendenziell bessere Leistungen auf als die monolingual deutschen Schülerinnen und Schüler, allerdings ist dieser Unterschied nicht signifikant. Die stärksten Defizite beider Gruppen, der deutsch-bilingualen und der Schülerinnen und Schülern mit einer anderen Sprache als Deutsch als Erstsprache bestehen im Deutschen im Bereich Wortschatz. Es kann vermutet werden, dass dieses Defizit sich auch auf alle anderen Leistungsbereiche auswirkt.

Für die sprachlichen Kompetenzen im Englischen zeigt sich ein deutlich anderes Bild als für die Bereiche des Deutschen: deutsch-bilinguale Schülerinnen und Schüler weisen hohe Testleistungen in den Testmodulen Grammatik, Sprachbewusstheit und Hörverstehen auf. In diesen Teilkompetenzen zeigen auch die Schülerinnen und Schüler mit einer anderen Erstsprache als Deutsch deutlich bessere Testleistungen als diejenigen mit Deutsch als Erstsprache, wenn die konfundierenden Variablen des Bildungsgangs, des sozioökonomischen Hintergrunds, der kognitiven Grundfähigkeit und des Geschlechts kontrolliert werden (vgl. Abbildung 26).



**Abbildung 26: Effekte des sprachlichen Hintergrundes auf Teilkompetenzen im Englischen zum Ende der neunten Jahrgangsstufe unter Kontrolle von Bildungsgang, sozioökonomischem Hintergrund, kognitiver Grundfähigkeit und Geschlecht: Kontraste der Leistungen von Mehrsprachigen oder monolingual nicht-deutsch Erstsprachigen gegenüber monolingual deutsch Erstsprachigen (Bewusstheit Soziopragmatik wurde aus technischen Gründen zu Beginn des Schuljahres erfasst) (Quelle: Hesse, Göbel & Hartig, 2008, S. 218)**

Allerdings beziehen diese Analysen die genannten Bedingungsfaktoren mit ein, unterscheiden sich damit also vom rein deskriptiven Leistungsvergleich ohne Berücksichtigung von Kontexteinflüssen, bei dem die Gruppe der Schülerinnen und Schüler mit Deutsch als Erstsprache in allen Teilbereichen signifikant am besten abschneidet.

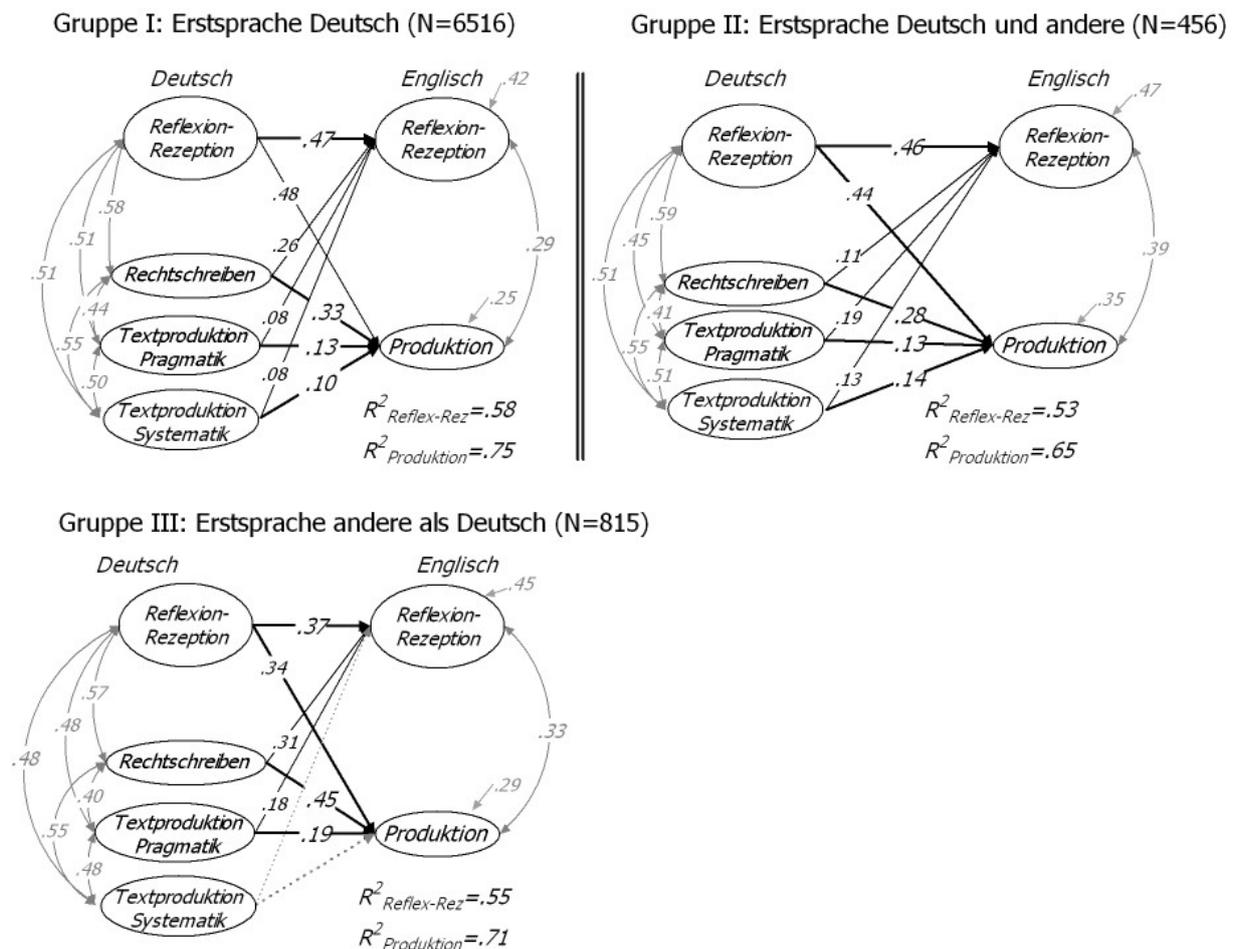
Zur Analyse der Frage, ob die Sprachkompetenzen in den drei vorhandenen Sprachgruppen die gleiche Struktur aufweisen, wurden wie schon bei der Analyse der Geschlechtsunterschiede vier gemeinsame Mehrgruppenmodelle mit unterschiedlich starken Restriktionen geschätzt. Die Modellanpassungen sind in Tabelle 16 dargestellt.

**Tabelle 16: Modellanpassung des Mehrgruppenvergleichs zwischen Schülerinnen und Schülern unterschiedlicher Erstsprache**

Modell	$\chi^2$	df	p	CFI	AIC	RMSEA	SRMR
1) Latente Korrelationen, Faktorvarianz und Pfade fixiert	1057.23	186	$\leq .01$	0.98	197386.76	0.04	0.03
2) Faktorvarianz und Pfade fixiert	1038.83	172	$\leq .01$	0.98	197387.47	0.04	0.03
3) Pfade fixiert	1034.98	166	$\leq .01$	0.98	197383.79	0.05	0.03

4) Strukturmodell frei geschätzt	990.13	150	$\leq .01$	0.98	197358.29	0.05	0.03
----------------------------------	--------	-----	------------	------	-----------	------	------

Analog zu den bisherigen Mehrgruppenvergleichen zeigen sich wenig Unterschiede zwischen den Modellanpassungen in den vier Modellen, es finden sich also keine Hinweise darauf, dass sich die Strukturen der Sprachkompetenz zwischen den drei Gruppen unterscheiden. Die nachfolgende Abbildung stellen die frei geschätzten Zusammenhänge zwischen den sprachlichen Teilkompetenzen für Modell 4 in den drei Gruppen dar.



**Abbildung 27: Zusammenhänge zwischen den sprachlichen Teilkompetenzen in den Gruppen mit unterschiedlicher Erstsprache ( $N_I=6516$ ,  $N_{II}=456$ ,  $N_{III}=815$ ,  $\chi^2=1002.62$ ,  $df=154$ ,  $p \leq .01$ ,  $CFI=.98$ ,  $RMSEA=.05$ ,  $SRMR=.03$ )**

Auch unter der Annahme gleicher struktureller Zusammenhänge variiert die Höhe eben dieser zwischen den drei Gruppen im Modell mit den frei geschätzten Parametern. Insgesamt kann in Gruppe I, den Schülerinnen und Schülern mit deutscher Erstsprache, ein größerer Anteil der Varianz in der Fremdsprache Englisch auf die Leistungen im Deutschen zurückgeführt

werden, als dies für die beiden anderen Gruppen möglich ist. Auch die Zusammenhänge zwischen den Teilkompetenzen im Deutschen und den reflexiv-rezeptiven und produktiven Faktoren im Englischen unterscheiden sich zwischen den Gruppen. Durch die hohe Varianzaufklärung in Gruppe I verbleibt zwischen Reflexion-Rezeption und Produktion im Englischen nur noch eine geringe Restkorrelation, während diese beiden Faktoren in Gruppe II stärker zusammenhängen.

Spezifische Effekte finden sich besonders in der Bedeutung der sprachproduktiven Komponenten im Deutschen, diese hängen in den drei Gruppen unterschiedlich hoch mit dem reflexiv-rezeptiven Faktor im Englischen zusammen. In allen Gruppen kann durch die Kompetenz Rechtschreiben im Deutschen Varianz in der Sprachproduktion im Englischen erklärt werden, dieser Zusammenhang ist am höchsten in der Gruppe III, den Schülerinnen und Schüler ohne deutsche Erstsprache. In dieser Gruppe kann durch den Prädiktor Textproduktion Systematik keine zusätzliche Varianz der Kompetenz im Englischen erklärt werden, wie es in den beiden anderen Gruppen der Fall ist. Für alle drei Gruppen zeigt sich die Bedeutung der Sprachreflexion-Rezeption im Deutschen auch für die produktiven Leistungen im Englischen, diese Komponente erklärt in allen drei Gruppen am meisten Varianz in der Fremdsprache. Insgesamt bleibt auch für diesen Mehrgruppenvergleich festzuhalten, dass ein Modell mit reflexiv-rezeptiven und produktiven Teilbereichen die grundlegende Struktur der Sprachkompetenz in allen drei Gruppen gut abbilden kann, wobei allerdings differenzierte Zusammenhänge in den Gruppen zu betrachten sind.

*Hypothese 9 kann dahingehend unterstützt werden, als sich bei Annahme gleicher mehrdimensionaler Strukturen von Sprachkompetenz in den Gruppen der Schülerinnen und Schülern mit deutscher Erstsprache und denjenigen mit einer anderen Erstsprache in diesen Gruppen spezifische Unterschiede in den Zusammenhängen zwischen dem Deutschen und dem Englischen zeigen.*

#### **6.4.4 Zusammenfassung der Ergebnisse der Mehrgruppenanalysen**

Dieser Abschnitt der Ergebnisanalyse befasste sich mit der Frage, ob die Struktur von Sprachkompetenz für spezielle Personengruppen von der Struktur der Gesamtstichprobe mit ihrer Unterteilung reflexiv-rezeptiver und produktiver Faktoren in den beiden Sprachen abweicht. Methodisch wurden Mehrgruppenmodelle analysiert, die unterschiedlich restriktiv

waren. Für alle drei Vergleiche – Geschlecht, Bilingualer Bildungsgang und Erstsprache – wiesen die analysierten Modelle unter Restriktion der grundlegenden Struktur eine gute bis sehr gute Anpassung an die empirischen Daten auf. Vorhandene Unterschiede zwischen den Gruppen konnten auf der Grundlage derjenigen Modelle näher betrachtet werden, welche zwar die Struktur restringierten, die differenziellen Zusammenhänge zwischen den Teilkomponenten jedoch frei schätzen ließen.

*Unterschiede zwischen Mädchen und Jungen* in der Struktur der Kompetenz sind zu analysieren, weil in vielen empirischen Studien über Leistungsvorteile von Mädchen in sprachlichen Fähigkeiten berichtet wird, und zwar nicht nur im Lesen, sondern auch in anderen Teilbereichen, wobei sich auch der Mädchenanteil in einer Klasse positiv auf die Gesamtkompetenz der Klasse auswirkt (De Fraine, Van Damme, Van Landeghem, Opendakker & Onghena, 2003). Allerdings weisen aktuelle Ergebnisse darauf hin, dass sich Kompetenzvorteile im Entwicklungsverlauf verringern (Dresel, Stöger, & Ziegler, 2005). Darüber ob diese rein quantitativen Leistungsvorsprünge sich auch in den Zusammenhangsstrukturen zeigen, liegen bisher keine Erkenntnisse vor. Die Analysen der vorliegenden Arbeit in Abschnitt 6.4.1 lassen den Schluss zu, dass die Binnenstruktur der Sprachkompetenz in beiden Gruppen absolut vergleichbar ist. Die Varianz der Sprachtests in beiden Gruppen lässt sich gut durch das mehrdimensionale Gesamtmodell erklären, auch die restringierten Modelle weisen eine sehr gute Anpassung an die empirischen Daten auf.

Marginale Unterschiede zeigen sich in der Bedeutung der Sprachreflexion-Rezeption im Deutschen, die bei den Mädchen höhere Zusammenhänge mit allen anderen sprachlichen Kompetenzen sowohl im Deutschen als auch im Englischen aufweist, als dies für die Gruppe der Jungen der Fall ist. Interpretiert werden könnte dieses Ergebnis dahingehend, dass sich der rein quantitativ vorhandene Vorsprung der Mädchen im Bereich des Lesens als Grundfähigkeit und Voraussetzung für jegliche Bearbeitung schriftsprachbezogenen Aufgabenmaterials auf die Erfassung sprachlichen Materials in allen anderen Bereichen auswirkt und sich diese Gruppe auch bei der Bearbeitung von Sprachtestaufgaben stärker auf die Fähigkeit der Sprachreflexion, der Sprachbewusstheit bezieht. Hingegen finden sich tendenziell höhere Zusammenhänge zwischen den sprachproduktiven Bereichen bei den Jungen, bei denen die sprachsystematischen Aspekte als das Bindeglied zwischen den beiden Sprachen Deutsch und Englisch angesehen werden können.

Insgesamt jedoch sind den hier durchgeführten Analysen keine strukturellen Geschlechtsunterschiede feststellbar. Zu vermuten wäre, dass diese sich auch weniger in der Struktur als in der absoluten Höhe der Kompetenz zeigen. Die Veränderung struktureller Unterschiede im Prozess der Sprachentwicklung, wie sie von verschiedenen Forschern postuliert werden (vgl. Abschnitt 1.7), sind mit den hier vorliegenden Methoden jedoch nicht analysierbar. Ebenso einschränkend ist darauf hinzuweisen, dass DESI nicht daraufhin angelegt war, spezifische Geschlechtsunterschiede zu diagnostizieren. Die Testverfahren wurden auf größtmögliche Fairness hinsichtlich potenzieller Einflussfaktoren wie Geschlecht entwickelt und zielten auf Kompetenzen ab, die sowohl bei Mädchen als auch bei Jungen am Ende der neunten Jahrgangsstufe vorhanden sein sollten.

Um strukturelle Entwicklungsunterschiede zwischen den Geschlechtern zu vergleichen, wäre auf die Ergebnisse aktueller neurobiologischer Studien hinzuweisen, die auf Unterschiede im Prozess der sprachlichen Informationsverarbeitung zwischen Mädchen und Jungen hinweisen. Während Mädchen anscheinend auf ein universelles Sprachnetzwerk zurückgreifen, verarbeiten Jungen auditive und visuelle sprachliche Reize unterschiedlich (Burman, 2008). Allerdings untersuchen diese Studien Wahrnehmungsprozesse und strukturelle Unterschiede unter neurobiologischem Aspekt, sodass diese Ergebnisse mit den hier verwendeten Instrumenten und Methoden nicht vergleichbar zu analysieren sind. Rein deskriptiv können auf der Grundlage der DESI-Tests keine wesentlichen Unterschiede in der Struktur von Sprachkompetenz festgestellt werden, für weiterführende Forschungen wäre hier auf individualdiagnostische, gegebenenfalls durch bildgebende Verfahren unterstützte Studien hinzuweisen, die nach strukturellen Unterschieden auch im Entwicklungsverlauf suchen (Hartshorne, 2006).

Die mögliche Bedeutung der *Unterrichtsumgebung* für potenzielle Unterschiede in der Struktur von Sprachkompetenz *zwischen Schülerinnen und Schülern in bilingualen und regulären Unterrichtsangeboten* wurde in Abschnitt 6.4.2 dargestellt. Quantitativ kann, wie dargestellt wurde, ein Leistungsvorsprung der bilingual unterrichteten Schülerinnen und Schüler in der Fremdsprache festgestellt werden. Inwieweit dieser auf die intensivere Beschulung in der Fremdsprache im Vergleich zu Schülerinnen und Schülern in regulären Unterrichtsangeboten zurückzuführen oder aber mit weiteren Bedingungsfaktoren verknüpft ist, diskutieren Nold, Hartig, Hinz und Rossa (2008) auf der Grundlage der DESI-Daten.

Der Mehrgruppenvergleich in der vorliegenden Arbeit kommt auf der Grundlage der guten Modellanpassung der restringierten Modelle zu dem Schluss, dass von vergleichbaren mehrdimensionalen Strukturen der Sprachkompetenz in beiden Gruppen ausgegangen werden kann. Die Annahme, dass bilingualer Unterricht insgesamt zu einer Veränderung der Struktur von Sprache führen könnte, wird von der vorliegenden Arbeit nicht gestützt. Vielmehr ist davon auszugehen, dass die sprachlichen Teilkompetenzen alle in stärkerem Maße gefördert werden. So verfügen bilingual unterrichtete Schülerinnen und Schüler in allen Teilbereichen des Englischen über eine signifikant höhere Kompetenz, zeigen jedoch keinen spezifischen Vorsprung bei einzelnen Fähigkeiten. Eine Ausnahme bildet hier der Bereich des Hörverstehens: Durch die quantitativ höheren Stundenzahlen und das damit verbundene häufigere Hören des Englischen als Unterrichtssprache sind die bilingual unterrichteten Schülerinnen und Schüler deutlich im Vorteil. Dies gilt insbesondere deshalb, da Befunde aus DESI darauf hinweisen, dass im regulären Englischunterricht auch das Deutsche als Unterrichtssprache häufig zum Einsatz kommt (Helmke et al., 2008).

Dieser Leistungsvorteil geht jedoch in der vorliegenden Stichprobe nicht einher mit einer veränderten Kompetenzstruktur. Es findet sich keine Bestätigung für die Annahme, dass eine insgesamt höhere Kompetenz zu einer homogeneren Struktur der Sprachkompetenz führt (vgl. Abschnitt 1.8) – im Gegenteil: Die Korrelationen der sprachlichen Teilbereiche innerhalb des Deutschen sind bei bilingual unterrichteten Schülerinnen und Schüler tendenziell niedriger. Dieses Ergebnis deutet darauf hin, dass bei gleicher Kompetenzstruktur in der leistungsstärkeren Gruppe der bilingual unterrichteten marginal stärker zwischen einzelnen Teilkompetenzen differenziert werden kann, diagnostisch also differenzielle Kompetenzprofile zutage treten.

Interpretiert werden könnten diese Unterschiede dahingehend, dass ein durchaus differenziertes Bild von Sprache in der Gruppe der bilingual unterrichteten Schülerinnen und Schüler vorzuherrschen scheint. Sprachliches Wissen und Können im Deutschen wird nicht unreflektiert auf den Bereich des Englischen übertragen, vielmehr kommen hier spezifische Fähigkeitskomponenten ins Spiel. Dementsprechend bestehen auch deutlich unterschiedliche Zusammenhänge in den sprachlichen Teilbereichen zwischen den beiden Gruppen: In der Gruppe der bilingual unterrichteten erklären die Kompetenzen im Deutschen wesentlich weniger Varianz der Kompetenzen im Englischen als in der Gruppe der Schülerinnen und Schüler mit regulärem Fremdsprachenunterricht.

Die geringeren Zusammenhänge zwischen den beiden Sprachen, dem Deutschen und dem Englischen, in der Gruppe der bilingual unterrichteten Schülerinnen und Schüler könnte zum einen als Hinweis darauf gedeutet werden, dass in dieser Gruppe weitere Bedingungsfaktoren diese Zusammenhänge beeinflussen, wie beispielsweise die Schulart oder die Dauer der bilingualen Beschulung, zum anderen lässt sich vermuten, dass die Schülerinnen und Schüler im bilingualen Unterrichtsangebot durch die stärkere Auseinandersetzung mit der Fremdsprache zu einem differenzierteren Bild davon gelangen, dass Sprachen sich in vielen Aspekten unterscheiden. Unterstützt werden könnte diese Annahme mit Theorien über Mehrsprachigkeit und Sprachbewusstheit, die postulieren, dass mit zunehmendem Einblick in verschiedene sprachliche Systeme deren Unterschiede stärker reflektiert werden (Hufeisen, 2003). Auf höheren Kompetenzstufen geht dies einher mit einer geringeren Häufigkeit von Transferfehlern, wenn Kenntnisse in der einen Sprache nicht unreflektiert auf eine andere übertragen werden. Für die hier getesteten Schülerinnen und Schüler stellt – bezogen auf die Struktur von Kompetenz – das Englische nach wie vor eine Fremdsprache dar, die sie allerdings insgesamt besser beherrschen als Schülerinnen und Schüler im Regelunterricht.

Zu berücksichtigen ist dabei auch, dass der bilinguale Unterricht in der Regel (und so auch in der vorliegenden Stichprobe) nur an Gymnasien, in Einzelfällen auch an Realschulen angeboten wird. Bei der Interpretation der Ergebnisse ist darüber hinaus zu bedenken, dass die Bedingungen des bilingualen Sachfachunterrichts in der untersuchten Stichprobe sehr heterogen waren; neben der reinen Stundenanzahl variierten die bilingualen Unterrichtsfächer und die Jahrgangsstufe, in dem mit dem bilingualen Unterricht begonnen wurde. Unter Berücksichtigung dieser einschränkenden Bedingungen sind auf der Grundlage der doch im Vergleich eher kleinen DESI-Stichprobe zudem keine Aussagen über die längerfristigen Entwicklungen der Kompetenzentwicklung in dieser Gruppe zu machen. Um aussagekräftige Unterschiede in der Struktur der Sprachkompetenz zu analysieren, wären Leistungsbeobachtungen in Vergleichsklassen mit bilingualem und Regelunterricht notwendig, die über mehrere Schuljahre hinweg die Entwicklung der Sprachkompetenz verfolgen.

Unterschiede zwischen Gruppen von *Schülerinnen und Schülern mit unterschiedlicher Erstsprache* wurden in einem dritten Schritt analysiert. DESI erfragte die Sprache, die als Erst- und Alltagssprache im Elternhaus gesprochen wird, wodurch drei Personengruppen unterschieden werden könnten: Die weitaus größte Gruppe gab an, zu Hause Deutsch zu sprechen, einer weitere Gruppe wuchs mit einer anderen Erstsprache als Deutsch zu Hause

auf und eine dritte, sehr kleine Gruppe gab an, bilingual mit Deutsch und einer anderen Erstsprache aufgewachsen zu sein. Auch hier zeigten die analysierten Mehrgruppenmodelle, dass prinzipiell von einer vergleichbaren Struktur der Sprachkompetenz in allen drei Gruppen ausgegangen werden kann. Auffällig und systematisch sind jedoch die Unterschiede in den Zusammenhängen zwischen den Kompetenzen im Deutschen und im Englischen zwischen den verglichenen Gruppen: Die Zusammenhänge sind am höchsten in Gruppe I, bei Schülerinnen und Schülern mit deutscher Erstsprache, etwas geringer bei bilingual deutschsprachigen Personen und am geringsten in Gruppe III, der Gruppe mit nicht-deutscher Erstsprache - hier werden lediglich drei Regressionspfade signifikant: Am meisten Varianz der Leistung im Englischen erklärt die Sprachreflexion-Rezeption im Deutschen, zusätzlich erklärt die Rechtschreibleistung im Deutschen Varianz in der produktiven Englischleistung. Insgesamt lassen sich in dieser Gruppe die Leistungen im Englischen etwas weniger stark erklären als in der Gruppe I.

Es kann vermutet werden, dass Schülerinnen und Schüler, für die sowohl das Deutsche als auch das Englische eine Fremdsprache darstellt, Wissen im Bereich der freien Sprachproduktion nicht unbedingt von einer auf die andere Sprache übertragen, lediglich die Rechtschreibung stellt eine Ausnahme dar. Dies ist derjenige Bereich, in dem diese Gruppe die geringsten Leistungsunterschiede im Vergleich zur Gruppe I aufweist (vgl. Abbildung 25). Anzunehmen ist, dass bei der Produktion von freien Texten in der jeweiligen Fremdsprache auf die Erstsprache zurückgegriffen wird, sodass direkte Transfereffekte zwischen den beiden Sprachen kaum zum Tragen kommen. Vermutet werden könnte auch, dass sowohl das Deutsche als auch das Englische von diesen Schülerinnen und Schülern als unterschiedliche Fremdsprachen erlernt wurden: Das Deutsche als Alltagssprache auch in nicht-institutionellen Umgebungen, das Englische hingegen als institutionell vermittelte Sprache, die von der deutschen gegebenenfalls auch funktional unterschieden wird. Diese Ergebnisse bestätigen Befunde aus der Fremdsprachenforschung, die darauf hinweisen, dass Fremdsprache nicht gleich Fremdsprache ist, zwischen Fremdsprachen nicht unbedingt ein Transfer stattfindet und der Erwerbskontext sowie die Lerndauer eine relevante Rolle beim Kompetenzerwerb spielen.

Etwas anders gestaltet sich der Zusammenhang in Gruppe II, den deutsch-bilingualen Schülerinnen und Schülern. Die Bedeutung des Faktors Sprachrezeption-Rexflexion entspricht praktisch derjenigen in Gruppe I, lediglich die sprachproduktiven Leistungen sind

distinkter. Die Ergebnisse weisen darauf hin, dass allgemeine und spezifische Transfereffekte in dieser Gruppe gegebenenfalls stärker zum Tragen kommen. Es kann vermutet werden, dass bilingual aufgewachsene Schülerinnen und Schüler sowohl die Kompetenzen in der deutschen Erstsprache als auch ihr durch die Zweitsprache erlangtes Sprachwissen auf die Fremdsprache Englisch differenziell stärker übertragen, als dies bei Gruppe III der Fall ist. Es ist zu vermuten, dass in dieser Gruppe stärker auf universelle, sprachliche Kompetenzen zurückgegriffen werden kann, die jedoch differenzierter und reflektierter übertragen werden als in Gruppe I. Dieses Ergebnis passt zu Befunden, die bilingual aufwachsenden Personen eine höhere Sprachbewusstheit nachweisen (vgl. Abschnitt 1.6).

Zusammenfassend ist festzustellen, dass das theoriegeleitete Modell der reflexiv-rezeptiven und produktiven Sprachstrukturen in allen hier verglichenen Gruppen aufzufinden ist. Trotzdem bleibt festzuhalten, dass sich differenzielle Unterschiede in den Zusammenhängen der Teilkompetenzen besonders zwischen dem Englischen und dem Deutschen aufzeigen lassen. Um der Frage, ob sich die Struktur der Kompetenz in unterschiedlichen Fremdsprachen bei Schülerinnen und Schülern nichtdeutscher Erstsprache unterscheidet, differenzierter nachzugehen, wären allerdings größere Stichproben von Schülerinnen und Schülern mit verschiedenen Erstsprachen nötig. Diese Datengrundlage konnte DESI nicht liefern. In der hier dargestellten Gruppe befinden sich verschiedene kleine Teilgruppen von Schülerinnen und Schülern, deren Erstsprache beispielsweise slawischen, asiatischen oder romanischen Ursprungs ist. Das gleiche gilt für diejenigen, die mit den beiden Erstsprachen Deutsch und Englisch zugleich aufgewachsen sind. Im Einklang mit bisherigen Befunden ist daher zu vermuten, dass sich die Struktur der Sprachkompetenz in Fremdsprachen anderen linguistischen Ursprungs durchaus zwischen diesen Teilgruppen unterscheiden könnte (vgl. Abschnitt 1.6). Auch spezifische Transfereffekte werden in Abhängigkeit von der Herkunftssprache wahrscheinlich variieren. Vertiefende Analysen bedürften jedoch einer speziell zu diesem Zweck ausgewählten Stichprobe, die im Idealfall auch noch Leistungstests in ihrer jeweiligen Erstsprache bearbeiten würde, um spezifische Transfereffekte und unterschiedliche Strukturen der Sprachkompetenz differenziert analysieren zu können.

Für die Analysen der Gruppenunterschiede bleibt zusammenfassend festzuhalten, dass quantitative Leistungsunterschiede oder Entwicklungsvorsprünge nicht zwangsläufig einhergehen mit einer anderen Struktur der Sprachkompetenz, soweit sie auf der Grundlage der verwendeten Testinstrumente deskriptiv beschreibbar ist. Um gezielt und

differenzialdiagnostisch Unterschiede zwischen Personengruppen analysieren zu können, bedarf es speziellerer Testdesigns und methodischer Zugänge, als dies auf der Grundlage eines Large-Scale-Assessments möglich ist. Immerhin deuten die hier vorgestellten Ergebnisse darauf hin, dass:

- Schülerinnen und Schüler der neunten Jahrgangsstufe sehr wohl als gemeinsame Gruppe hinsichtlich ihrer sprachlichen Fähigkeiten zu beschreiben sind und dass es gelingt, ein gemeinsames Strukturgleichungsmodell der Sprachkompetenz darzustellen,
- sich differenzielle Unterschiede auffinden lassen, die weniger mit dem Geschlecht oder dem Unterrichtsangebot als tendenziell mit unterschiedlichen Erstsprachen im Zusammenhang stehen, und
- die Ergebnisse als Hinweise auf Unterschiede in den Transfereffekten zwischen spezifischen Teilkompetenzen in unterschiedlichen Sprachen gedeutet werden können, die durch weiterführende Studien differenzierter untersucht werden sollten.

## 6.5 Die Bedeutung kognitiver Grundfähigkeit für die Struktur von Sprachkompetenz

Wie bereits in Abschnitt 1.7 dargestellt wurde existieren unterschiedliche Hypothesen über den Zusammenhang zwischen der Sprachkompetenz von Schülerinnen und Schülern und deren kognitiven Grundfähigkeiten. Abschließend wird in der vorliegenden Arbeit daher das bisher dargestellte Modell der Struktur von Sprachkompetenz *um einen Indikator der kognitiven Grundfähigkeiten erweitert*: Es wird analysiert, welche Bedeutung der kognitiven Grundfähigkeit von Schülerinnen und Schülern bei der Modellierung der Struktur von Sprachkompetenz zukommt.

Die kognitive Grundfähigkeit wurde in DESI mit dem Modul N2 des nonverbalen Teils des Kognitiven Fähigkeitstests (KFT 4-12+R) (Perleth & Heller, 2000) erfasst. Der KFT 4-12+R erfasst differenzielle kognitive Fähigkeitsdimensionen mit besonderer Relevanz für schulisches Lernen und beinhaltet die drei Dimensionen *Sprachliches Denken* (V), *Quantitative Fähigkeiten* (Q) und *Figurales Denken* (N), die mit jeweils drei unterschiedlichen Testteilen erfasst werden. Der Testteil N2 – Figurenanalogien - repräsentiert in den hier dargestellten Analysen den Bereich der *sprachunabhängigen Fähigkeiten* im Sinne fluider Intelligenz. Durch die Verwendung des figuralen Testmoduls im Gegensatz zu einem Modul aus dem Bereich sprachliches Denken sollen methodische Artefakte vermieden werden: Die Tests zum sprachlichen Denken des KFT ähneln von Inhalt und Testformat stark den DESI-Tests im Bereich Wortschatz. Die dadurch entstehenden hohen Korrelationen zwischen den kognitiven Grundfähigkeiten und einzelnen sprachlichen Tests würden jedoch keinen interpretierbaren Erklärungsbeitrag für das hier zu analysierende Modell leisten.

Entsprechend der Hypothesen in Abschnitt 4.4 wird nachfolgend untersucht, welche Zusammenhänge zwischen den kognitiven Grundfähigkeiten und den Kompetenzen in den sprachlichen Teilbereichen bestehen und wie sich die Struktur von Sprachkompetenz unter Berücksichtigung der kognitiven Grundfähigkeiten beschreiben lässt.

**Hypothese 10:** *Es bestehen unterschiedlich hohe Zusammenhänge zwischen den kognitiven Grundfähigkeiten und sprachlichen Teilbereichen auf Individual- und Klassenebene.*

Zur Analyse von Hypothese 10 werden die Korrelationen zwischen den sprachlichen Teilbereichen im Deutschen und im Englischen und den kognitiven Grundfähigkeiten auf Individual- und auf Klassenebene berechnet (vgl. Tabelle 17).

**Tabelle 17: Korrelative Zusammenhänge zwischen kognitiven Grundfähigkeiten und den latenten Dimensionen im Deutschen und im Englischen auf Individual- und Klassenebene**

Individualebene Deutsch	Bewusstheit	Lesen	Textproduktion Systematik	Textproduktion Pragmatik	Argumentation	Wortschatz	Rechtschreiben
KFT_n	0.22	0.25	0.13	0.12	0.16	0.23	0.10
Klassenebene Deutsch	Bewusstheit	Lesen	Textproduktion Systematik	Textproduktion Pragmatik	Argumentation	Wortschatz	Rechtschreiben
KFT_n	0.92	0.93	0.85	0.87	0.92	0.93	0.86

Individualebene Englisch	C-Test	Hörverstehen	Lesen	Bewusstheit	Textproduktion
KFT_n	0.24	0.16	0.19	0.17	0.14
Klassenebene Englisch	C-Test	Hörverstehen	Lesen	Bewusstheit	Textproduktion
KFT_n	0.87	0.78	0.84	0.83	0.86

Deutlich wird bei Betrachtung der korrelativen Analysen, dass unterschiedlich starke Zusammenhänge zwischen den kognitiven Grundfähigkeiten und den einzelnen sprachlichen Teilkompetenzen im Deutschen und im Englischen bestehen. Die höchsten Zusammenhänge auf Individualebene zeigen sich zwischen dem KFT und den Testwerten der Sprachbewusstheit, des Lesens und des Wortschatzes im Deutschen sowie dem C-Test und dem Leseverstehen im Englischen. Am geringsten fallen die Korrelationen zwischen dem KFT und den produktiven Teilbereichen im Deutschen und Englischen aus. Besonders die Rechtschreibung kann als eine Leistung angesehen werden, die nur schwach mit kognitiven Grundfähigkeiten zusammenhängt.

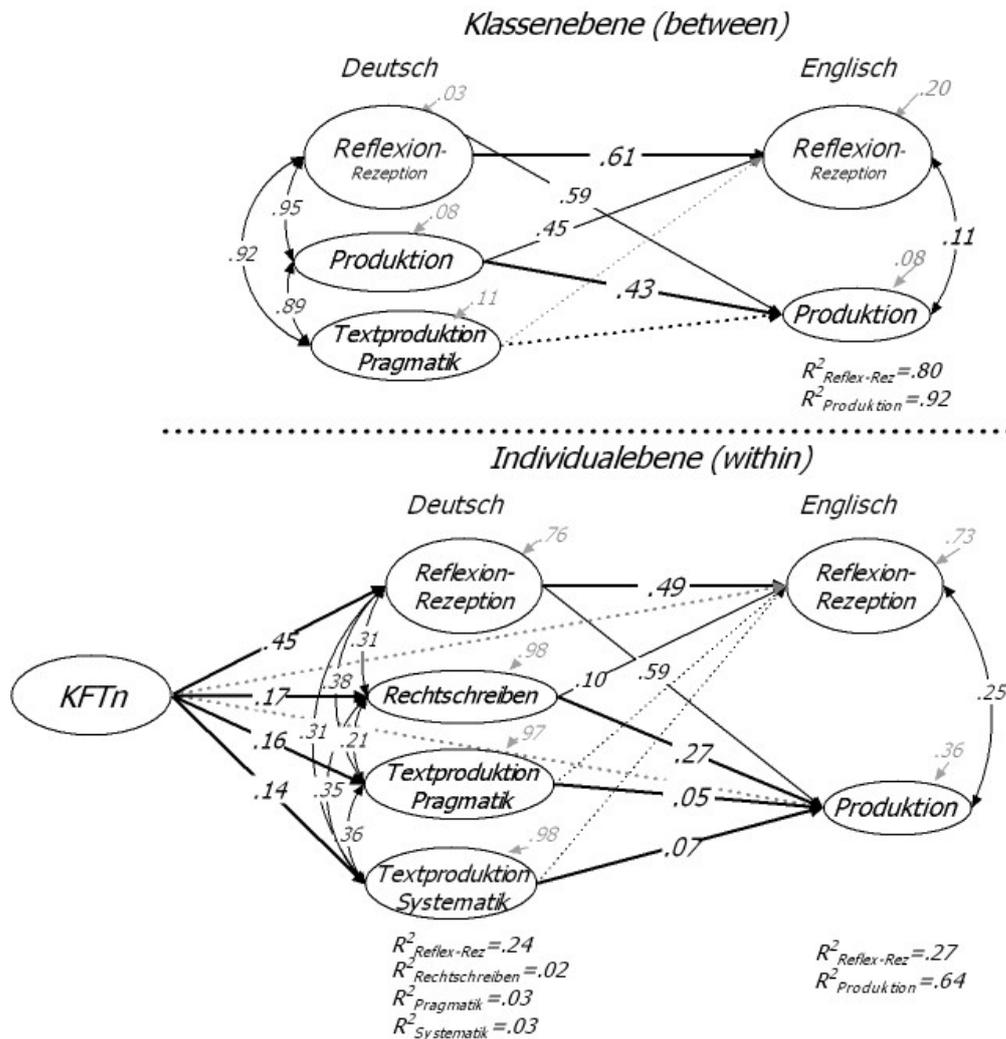
Auf Klassenebene sind die Korrelationen zwischen dem mittleren kognitiven Leistungsniveau innerhalb der Klasse und den durchschnittlichen Leistungen in den sprachlichen Teilbereichen durchweg sehr hoch. Diese zwischen den Bereichen kaum variierenden Zusammenhänge der Werte auf Klassenebene mit dem KFT spiegeln die Selektionswirkung des deutschen Schulsystems wider: Sie weisen auf den sehr hohen Zusammenhänge zwischen Bildungsgang, sozioökonomischem Hintergrund und kognitiven Fähigkeitswerten hin, die in

Schulleistungsstudien immer wieder berichtet werden. Diese Einflüsse sind im vorliegenden Modell nicht zu differenzieren und daher nicht substantiell im Bezug auf die Struktur von Sprachkompetenz interpretierbar.

Entsprechend werden die kognitiven Grundfähigkeiten im Folgenden als Prädiktoren auf Individualebene einbezogen, um so Hypothese 11 zu prüfen.

***Hypothese 11:** Unter Einbeziehung der individuellen kognitiven Grundfähigkeiten in ein gemeinsames Modell der Struktur der Sprachkompetenz im Deutschen und im Englischen sind die gleichen strukturellen Zusammenhänge darstellbar wie ohne Berücksichtigung der kognitiven Grundfähigkeiten.*

Hierzu wird das Maß der kognitiven Grundfähigkeit auf Individualebene als Prädiktor für die latenten Faktoren der Sprachrezeption-Reflexion und der Sprachproduktion im Deutschen und im Englischen in das Mehrebenen-Modell der Sprachkompetenz aufgenommen (vgl. Abbildung 28).



**Abbildung 28: Strukturgleichungsmodelle der Sprachkompetenz auf Individual- und Klassenebene unter Berücksichtigung der kognitiven Grundfähigkeiten auf Ebene 1. Gestrichelte Linien stellen nicht signifikante Regressionspfade dar. (N=8679,  $\chi^2=1439.44$ ,  $df=96$ ,  $p \leq .01$ , CFI=.94, RMSEA=.04, SRMR<sub>B</sub>=.03, SRMR<sub>W</sub>=.05)**

Im Modell der Struktur der Sprachkompetenz, das um den KFTn als Prädiktor für die latenten Faktoren der Reflexion-Rezeption und der Produktion erweitert wurde, lassen sich spezifische Regressionspfade der kognitiven Grundfähigkeiten darstellen. Der KFT kann auf Individualebene als Prädiktor zur Varianzaufklärung für alle vier Kompetenzbereiche im Deutschen beitragen, allerdings nicht für die beiden Bereiche des Englischen. Der Erklärungsbeitrag ist dabei für die Reflexion-Rezeption im Deutschen eher hoch, für die produktiven Teilbereiche im Deutschen vergleichsweise niedrig.

Wie in **Hypothese II** postuliert ist im Vergleich zum Gesamtmodell ohne Berücksichtigung des KFTn (vgl. Abbildung 19) festzustellen, dass sich die latenten Korrelationen zwischen sprachlichen Faktoren im Deutschen zwar leicht verringern, die strukturellen Zusammenhänge jedoch weiterhin vorhanden sind. Die gerichteten Zusammenhänge zwischen dem Deutschen und dem Englischen bleiben praktisch unverändert.

Es lässt sich schlussfolgern, dass auch unter Berücksichtigung der kognitiven Grundfähigkeiten das bisher postulierte Modell der Struktur von Sprachkompetenz haltbar ist. Die Bedeutung des Faktors der Reflexion-Rezeption in der Erstsprache Deutsch, der schon in den vorangegangenen Analysen deutlich wurde, zeigt sich hier abermals: Dieser Faktor repräsentiert zu einem guten Teil Fähigkeiten, die sehr stark auf die kognitiven Grundfähigkeiten zurückzuführen sind, während die produktiven Fähigkeiten davon unabhängig stärker variieren. Das Phänomen, dass die kognitiven Grundfähigkeiten als Prädiktoren für die Leistung im Englischen keine zusätzliche Varianz erklären, kann darauf zurückgeführt werden, dass die Zusammenhänge zwischen den kognitiven Grundfähigkeiten und den Kompetenzen in der Fremdsprache durch die Erstsprache mediiert werden.

## **6.6 Zusammenfassung der Ergebnisse der Analysen zur Struktur von Sprachkompetenz**

Die in diesem Teil der Arbeit dargestellten Ergebnisse bauen in der Logik ihrer Analysen systematisch aufeinander auf. Da für viele Bereiche der Struktur von Sprachkompetenz eher allgemeine Hypothesen existierten, stellen die jeweils vorhergehenden Ergebnisse die Ausgangsbasis für weitergehende, differenziertere Analysen dar. Zu Beginn war die jeweils bestehende Binnenstruktur für die beiden Sprachen des Deutschen und des Englischen zu beschreiben. Hierzu wurden auf Basis der Gesamtstichprobe verschiedene existierende, theoretisch fundierte Modelle verglichen. Die resultierenden Ergebnisse zeigen, dass die in Abschnitt 2.1 dargestellten Dimensionen der *rezeptiv-reflexiven* und der *produktiven* Sprachkompetenz, die in vielen Studien berichtet werden, auch in der Struktur der hier untersuchten Gruppe der deutschen Schülerinnen und Schüler differenzierbar ist, und zwar sowohl im Deutschen als auch im Englischen. Dieses mehrdimensionale Modell stellte die Grundlage für die weiteren Analysen dar. Bei jedem nachfolgenden Analyseschritt wurden

Alternativmodelle getestet, die entweder theoretisch begründet waren oder auf den Ergebnissen exploratorischer Faktorenanalysen beruhten.

Die in den Abschnitten 6.2.1 und 6.2.2 dargestellten Ergebnisse können als Argument für die Notwendigkeit einer mehrdimensionalen Betrachtung der Struktur von Sprachkompetenz angesehen werden. Gleichzeitig weisen sie darauf hin, dass die Rückführung der Varianz in verschiedenen Testmodulen auf theoretisch fundierte, latente Faktor gut gelingen kann. Bestes Beispiel hierfür ist der Bereich der *Reflexion-Rezeption im Deutschen*, dessen vier Testmodule sich in einem sehr gut passenden Modell konfirmatorisch auf eben diesen einen latenten Faktor zurückführen lassen (vgl. Abbildung 11). Neben der Tatsache, dass die vier hier zusammengeführten Teilkompetenzen des Lesens, der Sprachbewusstheit, der Argumentation und des Wortschatzes vom Konstrukt her ineinandergreifen und ähnliche Sprachverarbeitungsprozesse erfordern, ist allerdings darauf hinzuweisen, dass gerade für diese latente Dimension mit vier Tests vergleichsweise viele manifeste Indikatoren zur Verfügung standen, was eine Schätzung des Modells statistisch gesehen erleichtert haben könnte. Für jene latenten Dimensionen mit nur wenigen manifesten Indikatoren gelang die Modellanpassung im Vergleich weniger gut. Alle anderen Modellgütekriterien auch der sehr komplexen Modelle können jedoch als gut bis sehr gut bezeichnet werden und dienen im Weiteren dazu, zwischen alternativen Modellen zu entscheiden (vgl. Tabelle 3).

In einem nächsten Schritt wurden die jeweils getrennt analysierten Modelle des Deutschen und des Englischen in ein gemeinsames Modell integriert, das ungerichtete korrelative Beziehungen zwischen den sprachlichen Teilbereichen annimmt (vgl. Abbildung 15). Hier zeigen sich die bestehenden Zusammenhänge zwischen den latenten Dimensionen der Rezeption-Reflexion und der Produktion im Deutschen wie auch im Englischen, die in den darauf folgenden Analysen hinsichtlich möglicher Transfereffekte differenzierter betrachtet wurden.

In Abschnitt 6.3 des Ergebnisteils wurde analysiert, inwieweit sich die Struktur von Sprachkompetenz zwischen der Individual- und der Klassenebene unterscheidet. Unter der Annahme, dass spezifische institutionelle Effekte durchaus zu Varianz zwischen Schulklassen führen können, wurden mittels Methoden der Mehrebenenanalysen konfirmatorische Faktorenanalysen auf zwei Ebenen durchgeführt. Nachdem sich für das Deutsche und Englische jeweils die mehrdimensionale Unterscheidung der rezeptiv-reflektiven und der

produktiven Kompetenzbereiche aufrechterhalten ließ (vgl. Abschnitte 6.3.1.1 und 6.3.1.2), wurde in einem gemeinsamen Mehrebenen-Strukturgleichungsmodell der Frage nach vorhandenen Transfereffekten nachgegangen. Dabei wurde deutlich, dass Transfereffekte nicht nur spezifisch zwischen den jeweils vergleichbaren latenten Dimensionen vorhanden sind, sondern darüber hinaus die rezeptiv-reflektiven und die produktiven Leistungen im Deutschen durchaus Varianz in beiden Dimensionen im Englischen erklären können, also gerichtete Zusammenhänge zwischen allen Teilbereichen anzunehmen sind.

Entsprechend diente im Weiteren genau dieses Modell, das gerichtete Zusammenhänge zwischen allen Teilbereichen im Deutschen und Englischen annimmt, als Ausgangsmodell für die nachfolgenden Gruppenvergleiche. Mit Mehrgruppenmodellen wurde die Struktur der Sprachkompetenz zwischen Mädchen und Jungen, zwischen Schülerinnen und Schülern unterschiedlicher Herkunftssprachen und in bilingualen und regulären Unterrichtsformen verglichen (vgl. Abschnitt 6.4.4).

Abschließend wurde in den hier dargestellten Analysen der Frage nachgegangen, welche Bedeutung den kognitiven Grundfähigkeiten im Modell der Struktur von Sprachkompetenz zukommt (vgl. Abschnitt 6.5). Schon in den rein korrelativen Analysen zeigte sich, dass die Zusammenhänge zwischen kognitiven Grundfähigkeiten und sprachlichen Teilbereichen auf Individual- und auf Klassenebene vorhanden sind, jedoch nur auf Individualebene die Höhe dieser Zusammenhänge variiert. Die Annahme einer mehrdimensionalen Struktur von Sprachkompetenz konnte auch unter Berücksichtigung des KFTn-Wertes auf Individualebene des Mehrebenen-Strukturgleichungsmodells bestätigt werden. Signifikante prädiktive Effekte der kognitiven Grundfähigkeiten auf die latenten Dimensionen ließen sich in diesem Modell lediglich für den Bereich des Deutschen nachweisen.

Insgesamt ermöglichen es die in diesem Kapitel dargestellten Analysen, die Struktur von Sprachkompetenz im Deutschen und Englischen auf Basis der vorliegenden Daten differenziert abzubilden. Im nachfolgenden Kapitel 7 werden nach einer kurzen Zusammenfassung der Arbeit die hier dargestellten Ergebnisse im Gesamtzusammenhang der Forschung zur Struktur von Sprachkompetenz vor dem Hintergrund der verwendeten Modelle und Methoden diskutiert. Abschließend wird ein Ausblick über anknüpfende Forschungsmöglichkeiten gegeben.

## 7 Diskussion

Wie in den einleitenden Abschnitten dieser Arbeit dargestellt wurde, ist das Konstrukt der *Sprachkompetenz* sowohl konzeptionell als auch empirisch als sehr komplex anzusehen. Deutlich wird bei näherer Betrachtung der entsprechenden theoretischen Ansätzen und empirischen Befunde sehr schnell, dass zwar von *Sprachkompetenz* im Singular gesprochen wird, wenn die biologisch vorhandene Disposition zum Erwerb einer oder mehrerer Sprachen des Menschen gemeint ist, dass jedoch in Wahrheit *sprachliche Kompetenzen* im Plural zu betrachten sind, um der Komplexität der Struktur von Sprache gerecht zu werden. Die vorliegende Arbeit analysiert entsprechend Fragestellungen, die sich damit beschäftigen, in welcher Art und Weise die differenzierbaren und diagnostizierbaren Facetten von Sprachkompetenz zusammenhängen.

Anhand des in Abschnitt 1.1 einleitend dargestellten Überblicks über existierende Definitionen des Konstrukts „Sprachkompetenz“ wird ersichtlich, dass je nach theoretischem Hintergrund andere Aspekte dieses komplexen Phänomens in den Fokus rücken. Beispielfhaft seien die Unterschiede zwischen den rekonstruierenden Modellen von Grundelementen der Sprache - wie sie die Linguistik schafft - und der Erforschung der biologischen Grundlagen für die neuronale Verarbeitung von sprachlicher Information – im neurobiologischen Sinne – hervorgehoben, um das breite Spektrum der Forschungsrichtungen zu skizzieren, die jeweils eigene Ansätze zur Erforschung der Struktur von Sprachkompetenz verfolgen. Die vorliegende Arbeit verortet sich bei jenen differentialdiagnostischen Ansätzen, die von Sprachkompetenz als einem *empirisch messbaren, mehrdimensionalen Konstrukt* ausgehen. Die *Struktur* dieses Konstruktes ist dann analysierbar und zu beschreiben, wenn möglichst viele der Teilbereiche, aus denen es sich zusammensetzt, differenziert erfasst werden.

Die Analysen dieser Arbeit basieren auf zwei Ausgangspunkten:

- Auf den Theorien zur Struktur von Sprachkompetenz und den empirischen Befunde zu Zusammenhängen zwischen sprachlichen Teilbereichen aus vorangegangenen Studien unterschiedlicher Forschungsrichtungen, wie sie in den Abschnitten 1 und 2 dargestellt werden sowie
- auf den empirisch vorhandenen Daten der DESI-Erhebung, die es durch die differenzierte Erfassung von Sprachkompetenz im Deutschen und Englischen

überhaupt erst ermöglichen, die in den Abschnitten 1 und 2 dargestellten Modelle als Ausgangspunkt der eigenen Analysen zu nutzen.

### 7.1 Allgemeine Befunde zur Struktur von Sprachkompetenz

Das grundlegende Ergebnis der vorliegenden Arbeit ist die Aussage, dass *Sprachkompetenz als mehrdimensionales Konstrukt* anzusehen ist. Entgegen der Befunde einiger früherer Studien, welche potenziell vorhandene Teilbereichen von Sprachkompetenz letztendlich auf einen gemeinsamen, übergeordneten Faktor zurückführen (vgl. Abschnitt 2.2), zeigen sich in den vorliegenden Ergebnissen deutlich unterscheidbare Teilbereiche in den beiden Sprachen Deutsch und Englisch. Im Vergleich unterschiedlicher Modelle erweist sich das eindimensionale Modell, das jegliche gemeinsame Varianz in den Teilbereichen auf einen einzigen, gemeinsamen Faktor zurückführt, als unzureichend. Vielmehr ist zu beobachten, dass *mehrere latente Dimensionen* von Sprachkompetenz existieren, die jeweils spezifische Teilbereiche auf sich vereinen und darüber hinaus systematisch miteinander korrelieren.

Besonders hervorzuheben ist, dass diese latenten Dimensionen *im Englischen und Deutschen in vergleichbarer Weise vorhanden sind*: Sie bilden die beiden Bereiche der *reflexiv-rezeptiven* und der *produktiven* Sprachkompetenz ab. Die Tatsache, dass reflexiv-rezeptive und produktive Sprachkompetenzen in beiden Sprachen gleichermaßen modelliert werden können, bestätigt die Befunde von North (2000). Das in der vorliegenden Arbeit ebenfalls überprüfte Gegenmodell, welches eine *schriftsprachliche* und eine *auditive* Dimension trennt (Kunnan, 1992), erreicht auf Basis der vorliegenden Daten für das Englische hingegen keine bessere Modellanpassung (vgl. Abschnitt 6.2.3).

Gestützt wird der Befund der Mehrdimensionalität auch dadurch, dass diesen latenten Dimensionen im Deutschen und Englischen zwar teilweise unterschiedliche manifeste Indikatoren zugrunde liegen, *die Struktur in beiden Sprachen jedoch trotzdem vergleichbar ist*. Auch unter Verwendung verschiedener Testverfahren für die spezifischen sprachlichen Teilbereiche sind die latenten Faktoren modellierbar. Diese Tatsache spricht für die Robustheit dieses theoretische fundierten Modells der Struktur von Sprachkompetenz.

Der Ansatz der DESI-Studie, Sprachkompetenz in differenzierten Facetten zu erfassen, kann damit als erfolgreich beurteilt werden. Die Tests für die verschiedenen sprachlichen Bereiche erfassen unterschiedliche Aspekte von Sprachkompetenz. Dies ist umso positiver hervorzuheben, als die Testentwicklung sich nur zum Teil an theoretischen Modellen der Sprachkompetenz orientierte, im Fokus vielmehr die curriculare Validität der Instrumente stand. In diesem Sinne stellt die DESI-Studie Testverfahren zur Verfügung, die - obgleich nicht als individualdiagnostische Instrumente entwickelt – nicht nur vergleichende Aussagen über den Kompetenzstand in verschiedenen Bereichen des schulischen Lehrplans ermöglichen, sondern darüber hinaus verwendet werden können, um die Struktur von Sprachkompetenz differenziert abzubilden. Durch die Verwendung der verschiedenen Testformate mit vielfältigen Stimuli und mit offenem sowie geschlossenem Antwortmodus gelingt es, die unterschiedlichen Aspekte von Sprachkompetenz zu erfassen und abzubilden.

Diese mehrdimensionale Modellierung von Sprachkompetenz und das resultierende Modell mit den übergeordneten, latenten Dimensionen der Reflexion-Rezeption und der Produktion bildet die Basis, um Phänomene auf Individual- und auf Klassenebene, vorhandene Zusammenhänge zwischen den beiden Sprachen sowie Gruppenunterschiede und die Bedeutung der kognitiven Grundfähigkeiten für die Struktur von Sprachkompetenz zu analysieren.

## 7.2 Mehrebenenanalytische Betrachtung der Struktur von Sprachkompetenz

Die vorhandenen Daten der DESI-Studie basieren auf einer repräsentativen Stichprobe der Schülerinnen und Schüler in der neunten Jahrgangsstufe in Deutschland zum Zeitpunkt der Testerhebung. Hierzu wurden die teilnehmenden Schulen, Klassen und Personen in einem stratifizierten Verfahren aus der Gesamtheit aller allgemeinbildenden Schulen in Deutschland gezogen. Daraus resultiert eine hierarchische geschachtelte Struktur des verwendeten Datensatzes: Betrachtet wird die Leistung von Schülerinnen und Schülern in Klassen in Schulen. Wie im Methodenteil dieser Arbeit ausgeführt wird, ist diese Schachtelung zum einen bei der Modellanalyse zu berücksichtigen, um methodische Artefakte zu vermeiden. Zum anderen können aber genau diese verschiedenen Ebenen getrennt betrachtet werden, um Effekte auf den unterschiedlichen Ebenen zu modellieren und zu analysieren. Eine solche simultane Betrachtung der Analyseebenen ist in der vorliegenden Arbeit gewählt worden, um die Angemessenheit des postulierten Konstrukts – der Struktur von Sprachkompetenz – für beide Ebenen zu überprüfen und um Aussagen über dessen diagnostische Relevanz auf der Klassenebene zu ermöglichen.

Überprüft wurde die Annahme, dass von einer vergleichbaren Struktur von Sprachkompetenz auf Individual- und auf Klassenebene ausgegangen werden kann (vgl. Abschnitt 6.3). Hierzu wurde ein mehrschrittiges Vorgehen mit separaten Teilanalysen der Strukturen für das Deutsche und Englische auf Individual- und Klassenebene verwendet, das in einem gemeinsamen Mehrebenenstrukturgleichungsmodell resultierte (vgl. Abbildung 19). Als Kernaussage dieses Modells kann festgehalten werden, dass sich die latenten Dimensionen der reflexiv-rezeptiven und der produktiven Sprachkompetenz *auch auf Klassenebene* abbilden lassen. Die Struktur von Sprachkompetenz auf der Klassenebene ist marginal homogener, das heißt für das Deutsche werden auf Klassenebene nicht vier, sondern drei Faktoren gebildet. Der Faktor *Sprachproduktion auf Klassenebene* vereint die aggregierten Werte der beiden Indikatoren für die Leistung im Bereich Rechtschreibung und der freien Sprachproduktion Systematik, während die *Pragmatik auf Klassenebene* nach wie vor als separater Faktor darzustellen ist.

Die Darstellung im Zwei-Ebenen-Modell erweitert den Blick auf die Struktur von Sprachkompetenz: Die individuell differenzierbare Struktur der Sprachkompetenz kann auf die Klassenebene übertragen werden, die mittleren Leistungen von Klassen sind dadurch aussagekräftig nicht nur in verschiedenen Teilkompetenzen beschreibbar, sondern unterscheiden sich *in systematischer Weise*: Hinsichtlich der latenten Faktoren der Reflexion-Rezeption und der Produktion. Diese können unter Annahme des hier gewählten Modells genutzt werden, um Kompetenzunterschiede zwischen Klassen zu beschreiben. Eine solche Darstellungsweise geht über die üblicherweise verwendeten, eindimensionalen Gesamtmittelwerte für den Vergleich von Leistungsniveaus zwischen Klassen hinaus. Auch wenn die latenten Korrelationen zwischen den Faktoren auf Klassenebene hoch sind, so ermöglicht die differenzierte Darstellung zusätzliche Aussagen über die Kompetenz von Schulklassen: Statt Klassen hinsichtlich eines Kompetenzwerts der „allgemeinen Sprachkompetenz“ zu beschreiben, können die reflexiv-rezeptiven oder die produktiven Sprachkompetenzen durchaus getrennt betrachtet werden.

Die Leistungen in diesen latenten Faktoren können in weiteren Analysen mit Merkmalen des Unterrichts in Verbindung gebracht werden. Hierdurch ergeben sich vielfältige Ansätze für die Unterrichtsforschung und die Fachdidaktik: Zu fragen wäre beispielsweise nach Unterschieden zwischen den Bildungsgängen. Fokussieren bestimmte Bildungsgänge eher darauf, produktive Sprachkompetenzen zu vermitteln? In welcher Weise beziehen sie dabei die reflexiv-rezeptiven Bereiche ein, die ein starker Prädiktor für die produktiven Bereiche sind? Es ist davon auszugehen, dass sich besonders bei der Vermittlung der Fremdsprache die didaktischen Ansätze zwischen den Bildungsgängen unterscheiden. Eine weiterführende, praxisrelevante Fragestellung könnte darin bestehen zu untersuchen, welche Rolle verschiedene fachdidaktische Ansätze für die Entwicklung in den latenten Kompetenzbereichen spielen. In diesem Kontext von Interesse wäre sicherlich auch die Frage nach Unterschieden zwischen der Vermittlung der Erstsprache Deutsch und des Englischen im Regelunterricht und in bilingualen Schulen. Unter Einbezug dieser zusätzlichen Unterrichtsmerkmale in Regressionsmodelle könnte nicht nur das durchschnittliche Leistungsniveau dargestellt, sondern darüber hinaus auch der Einfluss von Unterricht auf die speziellen latenten Komponenten betrachtet werden.

Insgesamt bestätigt sich mit dem hier dargestellten Mehrebenenmodell die Annahme, dass *eine differenzierte Struktur von Sprachkompetenz ein Phänomen ist, das nicht nur auf Individualebene verbleibt*. Neben der Bildung von eindimensionalen Klassenmittelwerten als übergreifender Indikator, auf dem viele Vergleiche basieren, oder der Herausnahme eines speziellen Leistungswertes – beispielsweise der Lesekompetenz – bietet die mehrdimensionale Betrachtungsweise den Vorteil, dass das Zusammenspiel der sprachlichen Teilkompetenzen auf Klassenebene betrachtet werden kann. Dazu gehört auch die Frage nach den Zusammenhängen zwischen dem Deutschen und dem Englischen.

### 7.3 Zusammenhänge zwischen dem Deutschen und dem Englischen – spezifische oder allgemeine Transfereffekte?

Neben der gesonderten Darstellung der Struktur von Sprachkompetenz für das Deutsche und das Englische beschäftigt sich die vorliegende Arbeit mit den *strukturellen Zusammenhängen zwischen diesen beiden Sprachen* (vgl. Abschnitt 6.2.5). Generell ist davon auszugehen, dass Transferleistungen in den meisten Fällen von der Erstsprache auf die Fremdsprache(n) erfolgen, seltener umgekehrt (Lightbown & Spada, 2000; Timmermann, 2000). Die Zielsprache, in der Transfereffekte zum Tragen kommen sollten, ist in der vorliegenden Arbeit das Englische. Untersucht werden daher gerichtete, sowohl allgemeine als auch spezifische Zusammenhänge zwischen den latenten Faktoren im Deutschen und Englischen, die vor dem Hintergrund der aktuellen Erkenntnisse zum Transfer zwischen Sprachen zu erwarten sind (vgl. Abschnitt 2.3.).

Der Transfer von Kompetenzen zwischen Sprachen kann auf verschiedene Weisen erfolgen: Zum einen wird bei der Verarbeitung fremdsprachlicher Informationen auf grundlegende gemeinsame kognitive Prozesse zurückgegriffen - beispielsweise bei der Dekodierung von Schriftsprache -, die in der Erstsprache erworben wurden und als übergreifende sprachbezogene Fähigkeiten angesehen werden können. Darüber hinaus ist von spezifischen Transfereffekten zwischen sprachlichen Teilbereichen auszugehen, wenn beispielsweise grammatikalische Regeln der Erstsprache – ob angemessen oder nicht – auf die Fremdsprache übertragen werden. In diesem Kontext kommt es auch zu Fehlern beim Transfer im Sinne einer Übergeneralisierung. Die vorliegende Arbeit prüft Modelle, die von Transfereffekten zwischen dem Deutschen und dem Englischen ausgehen und stellt darüber hinaus die Frage, ob diese Transfereffekte nur zwischen den jeweils *korrespondierenden*, latenten Dimensionen der Reflexion-Rezeption und der Produktion oder zwischen allen Dimensionen im Deutschen und Englischen vorhanden sind.

Anhand der beiden gegenübergestellten Mehrebenen-Modelle in Abbildung 18 und Abbildung 19 kann festgestellt werden, dass *sowohl allgemeine als auch spezifische Transfereffekte* zwischen dem Deutschen und dem Englischen zu beobachten sind. Die Annahme, dass lediglich spezifische Transfereffekte zwischen den jeweiligen korrespondierenden latenten Faktoren existieren, ist nicht haltbar. Von besonderer Relevanz ist hierbei der Faktor der *Sprachreflexion-Rezeption*, der die Fähigkeit des Lesens, der

Sprachbewusstheit und auch der Argumentation und des Wortschatzes in der Erstsprache repräsentiert. Durch den Einbezug dieses Faktors als Prädiktor nicht nur für die korrespondierende reflexiv-rezeptive Dimension im Englischen, sondern auch für die produktive Sprachkompetenz steigt die erklärte Varianz in dieser auf Individualebene auf fast das Doppelte an. Im Gegensatz dazu trägt lediglich die Rechtschreibung als einer der drei sprachproduktiven Bereiche im Deutschen signifikant zur Erklärung der Varianz der Reflexion-Rezeption im Englischen bei. Diese auf Individualebene zu beobachtenden Transfereffekte zeigen sich auch auf Klassenebene.

Dies kann zum einen darauf zurückgeführt werden, dass eben diese Kompetenzen auf grundlegenden kognitiven Strukturen beruhen, die – einmal in der Erstsprache erworben – auch für den Sprachgebrauch in der Fremdsprache herangezogen werden. Zum anderen beinhaltet der Faktor der Reflexion-Rezeption im Deutschen mit dem manifesten Indikator „Sprachbewusstheit“ ein Testmodul, dessen korrespondierendes Element im Englischen auf den Faktor der Produktion zurückgeführt wird. Insgesamt fällt auf, dass die Sprachreflexion-Rezeption in der Erstsprache einen nicht zu unterschätzenden Beitrag zur Erklärung beider Bereiche der Fremdsprachkompetenz zu leisten vermag. Die Tatsache, dass in diesem Faktor die beiden Testmodule der Lesekompetenz und der Sprachbewusstheit enthalten sind, legt den Schluss nahe, dass diese in der Erstsprache erworben Kompetenzen eine wesentliche Grundlage für den Erwerb jeglicher fremdsprachlicher Kompetenzen darstellen. Dies bestätigt Befunde von Carrell (1991), die Lesekompetenz sprachübergreifend als grundlegendes Element von Sprachkompetenz identifiziert, und von Lightbown und Spada (2000), die für den Bereich der Sprachbewusstheit deutliche Übertragungseffekte zwischen Sprachen identifizieren.

Zu berücksichtigen ist bei der Interpretation der Modelle, dass in dieser Studie kein Prozess des Transfers abgebildet werden konnte. Die dargestellten Ergebnisse basieren auf nur einem Messzeitpunkt zum Ende der neunten Klassenstufe. Trotzdem erscheint es aus den dargelegten Gründen legitim, die vorhandenen Leistungen in den beiden Sprachen in einen gerichteten Zusammenhang zu stellen. In weiterführenden Auswertungen wären für die zu beiden Messzeitpunkten der DESI-Studie erhobenen Teilbereiche Effekte im Längsschnitt über ein Schuljahr zu modellieren.

#### 7.4 Gruppenunterschiede in der Struktur von Sprachkompetenz

Nachdem die Struktur von Sprachkompetenz für das Deutsche und das Englische in den ersten Analyseschritten der vorliegenden Arbeit in einem gemeinsamen Modell mehrdimensional beschrieben werden konnte, war zu analysieren, ob diese Struktur auch für verschiedene Personengruppen gilt. Ausgangspunkt dieser Hypothesen sind Befunde, die Kompetenzunterschiede zwischen Mädchen und Jungen und zwischen Personen unterschiedlicher Herkunftssprache berichten (vgl. Abschnitte 1.7 und 2.4). Diese Unterschiede beziehen sich jedoch lediglich auf die *quantitative Ausprägung* von sprachlichen Fähigkeiten im Entwicklungsprozess.

Die hier zu analysierende Frage nach potenziellen strukturellen Unterschieden basiert auf der *Annahme der konvergenten Entwicklung von Sprachkompetenz* (vgl. Abschnitt 1.8): Wenn anzunehmen ist, dass mit einer höheren Kompetenz eine stärkere Homogenität der Sprachkompetenz einhergeht, dann sollte sich die Sprachstruktur zwischen Mädchen und Jungen und zwischen Schülerinnen und Schülern unterschiedlicher Erstsprache unterscheiden, eine Vergleichbarkeit der mehrdimensionalen Struktur zwischen den beiden Gruppen wäre nicht mehr gegeben. In der vorliegenden Arbeit kann die Frage, ob sich Sprachkompetenz divergent oder konvergent *entwickelt*, allerdings nicht analysiert werden, da wie bereits dargestellt summative Testleistungen zum Ende des 9. Schuljahres erhoben wurden. Überprüfbar ist lediglich, ob die Struktur von Sprachkompetenz sich in bestimmten Gruppen homogener darstellt.

Auch wenn aufgrund der Modellanpassung des Modells für die Gesamtstichprobe (vgl. Abbildung 19) davon ausgegangen werden kann, dass eben dieses mehrdimensionale Modell die Struktur der Sprachkompetenz in der Gesamtstichprobe gut abbildet, erscheint es trotzdem legitim, im direkten Vergleich der verschiedenen Gruppen zu überprüfen, ob die Annahme gleicher Strukturen in den Gruppen gerechtfertigt ist, oder ob bei der gegebenen mehrdimensionalen Struktur Unterschiede in den Varianzen der latenten Dimensionen und den Regressionspfaden zwischen dem Deutschen und dem Englischen anzunehmen sind.

Auf Basis der durchgeführten Mehrgruppenanalyse zeigte sich, dass durchaus von einer *Vergleichbarkeit der Struktur von Sprachkompetenz bei Mädchen und Jungen* ausgegangen werden kann (vgl. Abschnitt 6.4.1) und zumindest unter Verwendung der vorliegenden Testinstrumente keine strukturellen Unterschiede in den Kompetenzen von Mädchen und Jungen festzustellen sind. Auch bezüglich des Transfers zwischen der Erst- und der Fremdsprache sind keine markanten Unterschiede zwischen den Gruppen aufzufinden. Gleichzeitig legt dieses Ergebnis nahe, dass auch bei höherer Gesamtkompetenz – in diesem Fall bei den Mädchen - a) sprachliche Kompetenzen durchaus differenziert diagnostiziert werden können und b) nicht automatisch eine Konvergenz der Kompetenzen, also eine homogenere Struktur, resultiert.

Hinzuweisen ist jedoch darauf, dass sich diese Vergleichbarkeit der beiden Gruppen lediglich auf die mit den vorliegenden Testverfahren diagnostizierbaren Sprachkompetenzen bei der hier untersuchten Altersgruppe bezieht: Durchaus denkbar ist es daher, dass in jüngeren Kohorten strukturelle Unterschiede zwischen Mädchen und Jungen bestehen, sich diese aber in der Entwicklung und vor allem im Prozess der institutionellen Vermittlung von Sprachkompetenz ausgleichen, wie dies beispielsweise von Elley (1994) und Wagemaker (1996) berichtet wird. Interessant wären in diesem Zusammenhang längsschnittliche Analysen der Zusammenhänge verschiedener Bereiche der sprachlichen Teilkompetenz in ihrer Entwicklung vom Anfang der Grundschule ab der 1. Klasse, die den Beginn der Beschulung in der Fremdsprache ab der 3. Klasse einbeziehen und die Relation zwischen Erst- und Fremdsprache über mehrere Jahre hinweg betrachten könnten.

Ein etwas differenzierteres Bild zeigt sich beim Vergleich der Personen unterschiedlicher Herkunftssprache (vgl. Abschnitt 6.4.3). Auch wenn das restringierte Mehrgruppenmodell eine gute Anpassung an die Daten aufweist, also von einer *Vergleichbarkeit der Sprachstruktur in den Gruppen* auszugehen ist, lassen sich bei frei geschätzten Regressionspfaden unterschiedliche Transfereffekte in den Gruppen abbilden: Bei Schülerinnen und Schülern mit der Erstsprache Deutsch, mit deutsch-bilinguaem Sprachhintergrund und jenen mit einer anderen Erstsprache als Deutsch unterscheiden sich die Zusammenhänge zwischen dem Deutschen und dem Englischen. Dieses Ergebnis weist darauf hin, dass der Stellenwert einer Sprache – Erstsprache, Zweitsprache oder Fremdsprache – darüber entscheiden kann, inwieweit diese Sprache als Referenzsystem für den Fremdsprachenerwerb dient.

In der Gruppe der Schülerinnen und Schüler mit *Deutsch als Erstsprache* sind Transfereffekte zwischen allen sprachlichen Bereichen im Deutschen und Englischen zu beobachten, stärkste prädiktive Erklärungskraft besitzt der Faktor der Reflexion-Rezeption im Deutschen. Die hohen Zusammenhänge zwischen diesen beiden Sprachen legen die Vermutung nahe, dass diese Gruppe von Schülerinnen und Schülern bei der Verwendung der Fremdsprache Englisch in starkem Maße auf die Kenntnisse in der Erstsprache Deutsch zurückgreift. Diese Interpretation kann unter Bezug auf die institutionellen Vermittlung der Fremdsprache Englische in deutschen Schulen wie folgt begründet werden: Deutsch ist auch im Englischunterricht die Referenzsprache, Übersetzungen erfolgen vom Deutschen ins Englische und häufig wird auch im Englischunterricht die Deutsche Sprache verwendet (vgl. Helmke et al. 2008). Der enge Bezug zwischen dem Deutschen und dem Englischen ergibt sich für diese Personengruppe also zwangsläufig, besonders wenn der Erwerbskontext rein auf die institutionelle Vermittlung in der Schule beschränkt bleibt und nicht durch Auslandsaufenthalte – die meistens erst in einer höheren Klassenstufe erfolgen – ergänzt wird. Die Bedeutung des Zugangs zur ersten Fremdsprache Englisch über das Deutsche ist nicht zu unterschätzen, denn positive und negative Transfereffekte sind im Prozess des Fremdsprachenerwerbs zu berücksichtigen.

In der Gruppe der *deutsch-bilingualen Schülerinnen und Schüler* zeigen sich ähnlich hohe Zusammenhänge zwischen den beiden Sprachen, der einzige Unterschied zu den Schülerinnen und Schülern mit deutscher Erstsprache ist lediglich die geringere Erklärungskraft der Rechtschreibleistung im Deutschen für die Kompetenzen im Englischen. Im Prinzip besteht daher keine Notwendigkeit, diese ohnehin kleine Personengruppe im Hinblick auf die Beschreibung der Struktur von Sprachkompetenz von den Schülerinnen und Schülern mit Deutsch als Erstsprache zu trennen. Bilingualität alleine scheint bei der differenzierten Diagnostik von Sprachkompetenz kein Unterscheidungskriterium zu sein. Ein interessanter weitergehender Untersuchungsgegenstand wäre daher eine Strukturanalyse der Sprachkompetenz von deutsch-englisch bilingualen Schülerinnen und Schülern. Sollten sich bei bilingualen Schülern tatsächlich eine *Interlanguage* schon im vorschulischen Spracherwerb entwickeln, so wäre anzunehmen, dass sich das Zusammenhangsmuster zwischen den Sprachen deutlich unterscheidet (Reich & Roth, 2002; Selinker, 1987). So jedoch ist davon auszugehen, dass das Englische für die in dieser Studie untersuchten, *nicht*

englisch-bilingualen Schülerinnen und Schüler eine institutionell vermittelte Fremdsprache darstellt und sich daher weder Erwerbskontext noch Transfereffekte strukturell unterscheiden.

Deutliche Unterschiede – nämlich prinzipiell geringere Zusammenhänge zwischen dem Deutschen und dem Englischen - sind allerdings bei der Gruppe von *Schülerinnen und Schülern mit einer anderen Erstsprache* als Deutsch zu beobachten. Dies gilt auch für den Faktor der Reflexion-Rezeption im Deutschen, der in dieser Gruppe nicht die gleiche Bedeutung für die Varianzerklärung in der Fremdsprache erlangt wie in den anderen beiden Gruppen. Hinweisen könnte dies auf ein jenseits des Deutschen und Englischen existierendes Referenzsystem, nämlich jenes der individuellen Muttersprache der Personen dieser Stichprobe. Die vorhandenen Zusammenhänge zwischen den beiden (Fremd-)sprachen Deutsch und Englisch wären dann weniger Indikatoren für spezifische Transferleistungen, sondern könnten vielmehr auf die Bedeutung der vorhandenen Sprachkompetenz in der jeweiligen Fremdsprache hinweisen. Die interessante Tatsache, dass in dieser Gruppe jedoch die Rechtschreibleistung im Deutschen in hohem Maße Varianz im Englischen erklärt, könnte darauf zurückgeführt werden, dass Regelwissen und dessen Anwendung bei der Sprachvermittlung in der Schule eine besondere Rolle spielen. Dies ist hingegen nicht der Fall bei einer nicht-institutionell erworbenen Erstsprache, die im schulischen Unterricht nicht weiter vertieft wird. Erfolgt daher der Schriftspracherwerb im Deutschen und Englischen quasi parallel im schulischen Kontext, so könnte dieses Merkmal – die Regelmäßigkeit von Sprache – als verbindendes Element beim Transfer zwischen den beiden Fremdsprachen angewandt werden.

Die *Bedeutung des institutionellen Erwerbskontexts* betrachtet auch das dritte Mehrgruppenmodell in der vorliegenden Arbeit, mit dem die Unterschiede in der Struktur von Sprachkompetenz zwischen englisch-bilingual unterrichteten und Schülerinnen und Schülern im Regelunterricht analysiert werden (vgl. Abschnitt 6.4.2). Unter der zutreffenden Annahme gleicher Sprachstrukturen in beiden Gruppen tragen die produktiven Kompetenzen im Deutschen in der Gruppe der bilingual Unterrichteten nicht signifikant zur Erklärung der Leistung im Englischen bei. Die institutionelle Vermittlung speziell der Sprachproduktion im Englischen scheint sich im bilingualen Sachfachunterricht nicht unbedingt an der Vorgehensweise des Deutschunterrichts oder des Regelunterrichts zu orientieren. Vielmehr ist davon auszugehen, dass die Sprachproduktion in den vielfältigen Kontexten des Sachfachunterrichts – der in der bilingual unterrichteten Gruppe auf Englisch abgehalten wird

- eine eigene Komponente darstellt. Englisch wird hier als tatsächliche Gebrauchssprache für alltägliche Schularbeiten in verschiedenen Inhaltsbereichen verwendet, was dazu führen könnte, dass am Ende der 9. Jahrgangsstufe bei den bilingual unterrichteten Schülerinnen und Schülern eine durchaus eigenständige Dimension der Sprachproduktion in der Fremdsprache vorhanden ist.

## 7.5 Kognitive Grundfähigkeit als Element der Struktur von Sprachkompetenz

Wie bereits in Abschnitt 1.7 dargestellt wurde, wird die Rolle der kognitiven Grundfähigkeiten in der Struktur von Sprachkompetenz durchaus kontrovers diskutiert. Im Bezug auf die Zusammensetzung von Lerngruppen existieren beispielsweise Befunde, die einen positiven Zusammenhang zwischen den durchschnittlichen kognitiven Fähigkeiten und den sprachlichen Leistungen auf Klassenebene berichten (De Fraine, Van Damme, Van Landeghem, Opdenakker & Onghena, 2003). Ebenfalls festzustellen sind Unterschiede in den (fremd-) sprachlichen Kompetenzen in Abhängigkeit von den individuellen kognitiven Fähigkeiten (Rost, 1989; Robinson, 2001). Bemerkenswert ist hierbei, dass je nach sprachlicher Teilfähigkeit durchaus verschiedene Effekte existieren: Die kognitiven Grundfähigkeiten haben für die Vorhersage der Leistungen im Lesen eine höhere Bedeutung als für die kommunikative Sprachfähigkeit (Teepen, 2004; Genesee, 1976).

In der vorliegenden Arbeit werden die kognitiven Grundfähigkeiten in ihrer Funktion als Element in der Struktur von Sprachkompetenz betrachtet. Zu analysieren war, inwieweit die strukturellen Zusammenhänge zwischen den sprachlichen Teilbereichen und zwischen dem Deutschen und dem Englischen auf die kognitiven Grundfähigkeiten zurückgeführt werden können. Dazu kann insgesamt festgestellt werden:

- Ein Teil der gemeinsamen Varianz der sprachlichen Teilfähigkeiten im Deutschen lässt sich auf Unterschiede in den kognitiven Grundfähigkeiten zurückführen,
- besonders relevant sind diese für die sprachlichen Kompetenzen der Reflexion-Rezeption,
- für die Leistung in den sprachproduktiven Bereiche ist die zusätzliche Erklärungskraft der kognitiven Grundfähigkeiten – über die ohnehin vorhandenen Zusammenhänge dieser mit dem Faktor Reflexion-Rezeption hinaus – eher gering, und
- es besteht keine zusätzliche signifikante Erklärungskraft der kognitiven Grundfähigkeiten für die Leistungen in der Fremdsprache. *Der korrelative Zusammenhang zwischen den kognitiven Grundfähigkeiten und den Leistungen im Englischen wird vermittelt über die Fähigkeit in der Erstsprache.*

Die Zusammenhänge zwischen den kognitiven Grundfähigkeiten und den Kompetenzen in der Fremdsprache werden also durch die Erstsprache mediiert, während die Transfereffekten zwischen der Erstsprache Deutsch und der Fremdsprache Englisch in diesem Modell unverändert bestehen bleiben.

Dieses Ergebnis deckt sich mit existierenden Befunden, die über eine geringere Bedeutung der kognitiven Grundfähigkeiten für kommunikative sprachliche Bereiche als für das Leseverstehen berichten. Hier scheinen sprachproduktionsspezifische Phänomene, wie erlerntes Regelwissen, aber auch sprachpragmatische Handlungsfähigkeit unabhängig von den kognitiven Grundfähigkeiten vorhanden zu sein. Sieht man letztere als fixe, eher unveränderliche individuelle Bedingungsfaktoren an, so liegt die Schlussfolgerung nahe, dass in sprachproduktiven Bereichen ein zusätzlicher Raum für Leistungsförderung besteht.

## 7.6 Mehrebenenanalysen – Welche Ebenen sollten betrachtet werden?

Die vorliegende Arbeit berücksichtigt in ihren Analysen die hierarchische Schachtelung der Daten und bezieht explizit unterschiedliche Betrachtungsebenen ein. Für die Analysen zur Struktur von Sprachkompetenz im Allgemeinen (vgl. Abschnitt 6.2) und für den Vergleich zwischen den unterschiedlichen Personengruppen (vgl. Abschnitt 6.4) wurden Modelle gewählt, welche die Zuordnung der Schülerinnen und Schüler zu ihren Klassen statistisch berücksichtigen, die strukturellen Zusammenhänge auf Klassenebene jedoch nicht explizit modellieren. Für die Beschreibung der Struktur von Sprachkompetenz in einem Gesamtmodell, zur Analyse der Transfereffekte und zur Beantwortung der Frage nach der Bedeutung der kognitiven Grundfähigkeiten wurden Mehrebenenstrukturgleichungsmodelle verwendet. Die Entscheidung darüber, welche Modelle zu verwenden sind, ist jedoch keineswegs beliebig, sondern hängt von den interessierenden Fragestellungen und den vorhandenen Daten ab.

Modelle, welche *Strukturen auf der Individualebene* unter Kontrolle der hierarchischen Schachtelung analysieren, sind die Methode der Wahl, wenn, wie in der vorliegenden Arbeit, individuelle Zusammenhänge abgebildet werden sollen, über die auf Klassenebene entweder keine theoretischen Annahmen vorliegen oder die auf Klassenebene nicht sinnvoll zu interpretieren sind. Durch die Kontrolle der Standardfehler wird die Gefahr von statistischen Artefakten reduziert, die Zusammenhänge auf Individualebene können so differenzierter abgebildet werden. Interpretiert werden dann die Effekte auf der Individualebene:

Der rein quantitative Leistungsunterschied zwischen Mädchen und Jungen, der wie gezeigt auch auf Individualebene nicht mit strukturellen Unterschieden zwischen den beiden Gruppen einhergeht, wäre schwerlich auf die Klassenebene zu übertragen. Dem müsste die Annahme zugrunde liegen, dass die Struktur der Sprachkompetenz aller Mädchen (oder aller Jungen) in einer Klasse von gleichen systematischen Effekten beeinflusst wird – ohne dass diese definierbar wären. Ein Mehrgruppenmodell für Mädchen und Jungen auf der Klassenebene würde voraussetzen, dass es sinnvoll wäre, für Mädchen und Jungen in der gleichen Klasse jeweils einen eigenen geschlechtsspezifischen Kontext zu modellieren, der die Struktur von Sprachkompetenz prägt – dieses Annahme ist nach vorliegenden Erkenntnissen nicht haltbar.

Statt einer Strukturmodellierung von Gruppenunterschieden auf Klassenebene kann es hingegen sinnvoll sein, das Geschlecht als Moderatorvariable in ein Modell aufzunehmen. Ein

solches Modell könnte die Annahme überprüfen, dass der Anteil der Mädchen in einer Klasse Unterschiede in der sprachlichen Struktur in dieser Gruppe bedingt, oder die Zusammenhänge zwischen sprachlichen Teilbereichen in unterschiedlicher Weise beeinflusst. Das Geschlecht wird in diesem Fall als Moderatorvariable auf Klassenebene in das Modell aufgenommen.

In gleicher Weise kann argumentiert werden, dass auch die *individuelle Erstsprache kein auf Klassenebene analysierbares Merkmal* darstellt, wenn es um den Vergleich der Struktur von Sprachkompetenz zwischen Gruppen von Personen geht. Denn wie sich Schülerinnen und Schüler mit unterschiedlicher Erstsprache zwischen den in dieser Stichprobe erhobenen Klassen verteilen, ist abhängig von Hintergrundmerkmalen und Sozialstruktur. In den meisten Klassen lernen Schülerinnen und Schüler mit unterschiedlicher Erstsprache, die Definition getrennter Lernkontexte für diese Gruppen innerhalb derselben Klasse verbliebe künstlich und wäre nicht interpretierbar.

In den hier vorgestellten Gruppenvergleichen zielten die Fragestellungen jeweils auf Phänomene der Individualebene ab, so dass die Analysen entsprechende Modelle verwendeten. Eine *mehrebenenanalytische Betrachtung* ist immer dann notwendig und sinnvoll, wenn Effekte der Klassenebene im Fokus stehen, die inhaltlich interpretierbar sind. Mit den hier verwendeten *Mehrebenenmodellen*, die Zusammenhänge simultan auf Individual- und auf Klassenebene analysieren, konnte der Frage nachgegangen werden, ob bei der Diagnostik von Sprachkompetenz die gleiche Anzahl von latenten Faktoren auf Individual- und auf Klassenebene relevant ist, ob also das bisher nur für die Individualebene postulierte theoretische Modell auch auf aggregierte Werte übertragen werden kann. Die Überprüfung der Hypothesen, dass die Struktur von Kompetenz auf aggregierter Ebene homogener sei als auf Individualebene, kann in diesem Fall nur durch ein Mehrebenenmodell erfolgen: Wie die dargestellten Ergebnisse zeigen, lässt sich die mehrdimensionale Struktur von Sprachkompetenz auch auf Klassenebene abbilden (vgl. Abschnitt 6.3.2).

In diesem Fall reicht eine Kontrolle der hierarchischen Schachtelung der Daten nicht aus, vielmehr steht der Vergleich der Struktur von Sprachkompetenz auf den beiden Ebenen im Mittelpunkt der Analysen. Dieser ist inhaltlich interpretierbar - hohe korrelative Zusammenhänge auf Klassenebene weisen auf eine simultane Vermittlung verschiedener Teilbereiche der Sprachkompetenz im Unterricht hin -, und ist aus Sicht der Unterrichtsforschung ein wichtige Indiz dafür, dass auch auf Klassenebene die

Mehrdimensionalität des Konstrukts Sprache zu berücksichtigen ist: Für differenzierte Rückschlüsse auf das Unterrichtsgeschehen ist es demnach nicht zulässig, von den Klassenmittelwerten in einem Testbereich auf die durchschnittliche Klassenleistung in anderen sprachlichen Teilbereichen zu schließen. Hier geht die DESI-Studie mit ihrer differenzierten Erfassung von Sprachkompetenz über die Erkenntnisse bisheriger Schulleistungsstudien hinaus, so dass die berichtete Dimensionalität keineswegs – wie in früheren Studien diskutiert – als methodisches Artefakt anzusehen ist.

### **7.7 Dimensionalität als methodisches Artefakt des diagnostischen Ansatzes?**

Die Tatsache, dass Erfassungs- und Auswertungsmethoden durchaus zu Artefakten führen können (vgl. Abschnitt 2.5), ist auch in der vorliegenden Arbeit zu berücksichtigen und soll im Folgenden kritisch hinterfragt werden:

Die differenzierte Modellierung der Struktur von Sprachkompetenz in der vorliegenden Arbeit wird erst durch das facettenreiche Testdesign der DESI-Studie ermöglicht. Die Auswahl der zu testenden Bereiche und die Umsetzung in Testverfahren, die sich verschiedener Testformate bedienen, erlaubt eine differenzierte Betrachtung von strukturellen Zusammenhängen vor dem Hintergrund verschiedener theoretischer Modellen der Struktur von Sprachkompetenz. Durch die Verwendung des Tests zur freien mündlichen Sprachproduktion können für das Englische sogar zwei konkurrierende Zusammenhangsmodelle verglichen werden (vgl. Abschnitt 6.2.3), im Deutschen können durch die detaillierte Kodierung der Aufgabe zur freien schriftlichen Sprachproduktion mit der Rechtschreibung, der Pragmatik und der Systematik sogar drei Aspekte der produktiven Deutschkompetenz abgebildet werden.

Hervorzuheben ist, dass die *Methode der exploratorischen Faktorenanalyse* auch für die Daten der DESI-Studie in vielen Fällen auf eine eindimensionale Struktur hinweist. Besonders das Eigenwertkriterium kann zu der Annahme verleiten, einen eindimensionalen Faktor der Sprachkompetenz anzunehmen. Erst der Vergleich der theoriegeleiteten, konfirmatorischen Modelle zeigt deutlich, dass die *Annahme der Eindimensionalität* von Sprachkompetenz zumindest für die Daten der vorliegenden Arbeit einen Fehlschluss darstellen würde. Neben der statistischen Analyseverfahren kann jedoch auch die Art der Erfassung ein Grund für die vielen widersprüchlichen Befunde zur Struktur von Sprachkompetenz sein:

Mit den vier Testmodulen der Lesekompetenz, der Argumentation, des Wortschatzes und der Sprachbewusstheit kann im Deutschen der gemeinsame latente Faktor der *Sprachreflexion-Rezeption* modelliert werden, der ein eindimensionales Teilmodell darstellt (vgl. Abbildung 11). Die Testverfahren, welche dieser Faktor repräsentiert, erfassen mit einander ähnlichen Testformaten solche Aspekte der Sprachkompetenz, die im vorliegenden Fall stark interkorrelieren. Diese Korrelationen können einen Effekt der Testformate darstellen oder aber darauf zurückzuführen sein, dass diese Aspekte der Sprachkompetenz in der hier getesteten Personengruppe in homogener Weise ausgebildet sind. In jedem der beiden Fälle lässt sich mit diesen Tests ein eigener Faktor der *reflexiv-rezeptiven* Sprachkompetenz abbilden. Wird dieser jedoch nicht – wie in DESI geschehen – durch Test zur Erfassung *produktiver Kompetenzen* ergänzt, kann der Eindruck entstehen, dass Sprachkompetenz ein eindimensionales Konstrukt sei. Werden also in empirischen Studien lediglich solche sprachliche Teilbereiche erfasst, die ohnehin von ihrem Konstrukt her korrelieren und darüber hinaus auch noch ähnliche Testformate verwendet, so ist eine Differenzierung zwischen diesen sprachlichen Aspekten kaum möglich: Das Fazit lautet dann oft fälschlicher Weise, dass die Kompetenz in der Erstsprache eine eindimensionale Struktur aufweise. Erst die *Erweiterung des diagnostischen Rahmens* um Aspekte der produktiven Sprachkompetenz zeigt die methodische Problematik auf, der sich viele Studien aus Kosten- und Zeitgründen nicht stellen können.

Wie zu Beginn dargestellt (vgl. Abschnitt 3.2), erfolgt die Erfassung von Sprachkompetenz entweder vor dem Hintergrund differenzialdiagnostischer Fragestellungen mit dem Fokus auf nur einen sprachlichen Teilbereich oder in repräsentativen Schulleistungstudien in eher holistischer Weise durch die Berücksichtigung nur weniger Indikatoren für sprachliche Kompetenzen. Prominentes Beispiel für einen solchen Indikator ist die Lesekompetenz, welche aufgrund ihrer Relevanz für das eigenständige Erschließen von Wissen in keiner Schulleistungstudie fehlen darf, oder die Verwendung von Screening-Verfahren wie dem C-Test. Sicherlich ist die Feststellung der Lesekompetenz ein wichtiger Indikator für den Kompetenzstand der Schülerinnen und Schüler in einem Bildungssystem – als alleiniger Indikator erscheint sie jedoch nicht ausreichend. Um ein umfassendes Bild sprachlicher Kompetenzen abbilden zu können, müssen differenziertere Testverfahren verwendet werden, die sowohl reflexiv-rezeptive als auch produktive Bereiche der Sprachkompetenz erfassen. Optimaler Weise sollten dabei neben der schriftbasierten Erfassung auch auditive Testformate

verwendet werden; mit der Erfassung der mündlichen Sprachproduktion im Englischen wurden in DESI wesentliche Erkenntnisse zur aktiven Fremdsprachkompetenz von Schülerinnen und Schülern gewonnen (Nold & Rossa, 2008c).

Diese Möglichkeiten der Erfassung bieten sich nicht in jeder Studie. Die Tatsache, dass auch im Rahmen von DESI für viele weiterführende Analysen je ein Gesamtwert der Kompetenz im Deutschen und Englischen gebildet wurde, zeigt, dass sich viele Fragestellungen zur Unterrichtswirklichkeit im Sprachunterricht an deutschen Schulen nicht auf differenzierte sprachliche Teilkompetenzen, sondern auf einen *Gesamtkompetenzstand* beziehen. Die Analysen der vorliegenden Arbeit sollten gezeigt haben, dass hierbei viel Information verschenkt wird. Besonders die Tatsache, dass sowohl im Deutschen als auch im Englischen *vergleichbare latente Dimensionen von Sprachkompetenz* modelliert werden können und dass darüber hinaus *spezifische Transfereffekte* zwischen diesen abgebildet werden können, liefert wertvolle zusätzliche Informationen – und gelingt nicht bei Verwendung von jeweils einem Summenwert für das Deutsche und das Englische.

Natürlich ist darauf hinzuweisen, dass mit allen verwendeten Methoden, auch den konfirmatorischen Faktorenanalysen und Strukturgleichungsmodellen, durchaus die Abbildung *alternativer Modelle* der Struktur von Sprachkompetenz möglich ist. In der hier vorliegenden Arbeit wurden solche Modelle zur Überprüfung ausgewählt, die eine fundierte theoretische Basis aufweisen und auch in vorangegangenen empirischen Studien – zumindest für bestimmte Teilbereiche – empirisch gestützt wurden. Daher kann geschlussfolgert werden, dass unter Verwendung entsprechender diagnostischer Verfahren und statistischer Analysemethoden eine mehrdimensionale Darstellung der Struktur von Sprachkompetenz auch in zukünftigen Studien möglich und aussagekräftig ist.

## 7.8 Abschließender Ausblick

Mit der DESI-Studie, deren Daten die Grundlage für viele Publikationen bilden, wurde gezeigt, dass eine differenzierte Erfassung von Sprachkompetenz weit mehr Aussagen zum Kompetenzstand von Schülerinnen und Schülern in Deutschland ermöglicht, als dies mit einem allgemeinen Kompetenzindikator zu leisten wäre. Die Erkenntnis, dass sprachliche Kompetenzen in vielen Facetten zu erfassen sind, wenn aus den Ergebnissen praxisrelevante Hinweise für ihre institutionelle Vermittlung abgeleitet werden soll, wurde inzwischen auch auf bildungspolitischer Ebene aufgegriffen. Bei der Normierung der Bildungsstandards für den mittleren Schulabschluss werden im Jahr 2009 differenzierte Testverfahren für rezeptive und produktive Aspekte der Erstsprache Deutsch sowie der Fremdsprachen Englisch und Französisch eingesetzt, die es ermöglichen werden, Kompetenzniveaus und Zusammenhänge zwischen den sprachlichen Teilbereichen zu analysieren.

Auch die Daten der DESI-Studie bieten weitere Anknüpfungspunkte für vertiefende Analysen. Wie in den vorangegangenen Abschnitten dargestellt wurde, können neben der Gesamtstruktur von Sprachkompetenz, welche im Fokus der vorliegenden Arbeit steht – verschiedene Personengruppen betrachtet werden. Von besonderem Interesse sind hierbei weiterhin die Gruppen der Schülerinnen und Schüler mit deutsch-bilingualer Erstsprache beziehungsweise mit nicht-deutschem Sprachhintergrund (Hesse, Göbel & Hartig 2008). Trotz der kleinen Stichprobe könnte in weiterführenden Analysen nach systematischen Unterschieden in den sprachlichen Kompetenzen dieser Gruppen gesucht werden. Dabei muss nicht zwangsläufig die Struktur im Vordergrund stehen, vielmehr wäre danach fragen, welche Zusammenhänge zwischen den sprachlichen Kompetenzen und den von Schülerinnen und Schülern wahrgenommenen Unterrichtsprozessen bestehen. Weiterführend wären auch die Einflüsse institutioneller Bedingungen – speziell der Unterrichtsform des bilingualen Unterrichts – zu analysieren, die über rein quantitative Kompetenzbeschreibungen hinausgehen (Nold, Hartig, Hinz & Rossa 2008).

In eingeschränkten Modellen mit jenen Testmodulen, die zu beiden Messzeitpunkten verwendet wurden, könnten Veränderungen im Zusammenhang zwischen dem Deutschen und dem Englischen über den Zeitraum von einem Schuljahr analysiert werden. Dabei stellt sich besonders die Frage nach Unterschieden im Transfer zwischen den Sprachen: Zwischen Bildungsgängen, zwischen Personengruppen und unter Einbezug von Merkmalen des Unterrichts. Die in dieser Arbeit dargestellte Struktur von Sprachkompetenz mit den latenten Faktoren der Reflexion-Rezeption und der Produktion kann damit auch in zukünftigen Forschungsvorhaben als Basis dienen, um Unterschiede in der sprachlichen Kompetenzen von Schülerinnen und Schülern auf Effekte des Individuums und des Unterrichts zu beziehen.

## 8 Literatur

- Abendroth-Timmer, D. (2004). Evaluation bilingualer Module aus Schülerperspektive: Zur Lernbewusstheit und ihrer motivationalen Wirkung. *Zeitschrift für Interkulturellen Fremdsprachenunterricht [Online]*, 9(2), 27pp.
- Ackeren, I. von (2005). Vom Daten- zum Informationsreichtum? Erfahrungen mit standardisierten Vergleichstests in ausgewählten Nachbarländern. *Pädagogik*, 57(5), 24-28.
- Alderson, J. C. (1993). The relationship between grammar and reading in an English for academic purposes test battery. In D. Douglas & C. Chapelle (Hrsg.), *A new decade of language testing research: Selected papers from the 1990 Language Testing Research Colloquium* (S. 203–219). Alexandria, VA: TESOL.
- Alderson, J. C. & Benerjee, J. (2002). Language testing and assessment (Part 2). *Language Teaching*, 35(1), 79-113.
- Allerup, P. (1999). Rasch measurement theory. In G. N. Masters & J. P. Keeves (Hrsg.), *Advances in measurement in educational research and assessment* (S. 64-84). Oxford: Pergamon Press.
- Apeltauer, E. (1997). *Grundlagen des Erst- und Fremdspracherwerbs*. Berlin/München: Langenscheidt.
- Arras, U. & Grotjahn, R. (2002). TestDaF: Aktuelle Entwicklungen. *Fremdsprachen und Hochschule*, 66, 65-88.
- Artelt, C., Stanat, P., Schneider, W. & Schiefele, U. (2001). Lesekompetenz: Testkonzeption und Ergebnisse. In Deutsches PISA-Konsortium (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 69-137). Opladen: Leske und Budrich.
- Asher, J. J. (1972). Children's first language as a model for second language learning. *The Modern Language Journal*, 56(3), 133-139.
- Asparouhov, T. (2004). *Weighting for unequal probability of selection in multilevel modeling*. – Verfügbar unter:  
<http://www.statmodel.com/download/webnotes/MplusNote81.pdf> [22.08.2007]
- Bachman, L. F. (1990a). Constructing measures and measuring constructs. In B. Harley, P. Allen, J. Cummins & M. Swain (Hrsg.), *The development of second language proficiency* (S. 26-38). Cambridge: Cambridge University Press.
- Bachman, L. F. (1990b). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F. & Cohen, A. D. (1998). *Interfaces between second language acquisition and language testing research*. Cambridge: Cambridge University Press.
- Bachman, L. F., Davidson, F., Ryan, K. & Choi, I.-C. (1995). *An investigation into the comparability of two tests of English as a foreign language*. Cambridge: Cambridge University Press.
- Bachman, L. F. & Palmer, A. S. (1981). The construct validation of the FSI oral interview. *Language Learning*, 31(1), 67-86.
- Bachman, L. F. & Palmer, A. S. (1982). The construct validation of some components of communicative proficiency. *TESOL Quarterly*, 16(4), 449-465.
- Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (2003). *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung*. Berlin: Springer.
- Baumert, J., Bos, W. & Lehmann, R. (Hrsg.). (2000). *TIMSS/III. Dritte internationale Mathematik- und Naturwissenschaftsstudie. Mathematische und*

- naturwissenschaftliche Bildung am Ende der Schullaufbahn*. Opladen: Leske und Budrich.
- Baumert, J., Klieme, E., Lehrke, M. & Savelsbergh, E. (2000). Konzeption und Aussagekraft der TIMSS-Leistungstests. Zur Diskussion um TIMSS-Aufgaben aus der Mittelstufenphysik. *Die Deutsche Schule*, 92(1), 102-115.
- Bearden, W. O., Sharma, S. & Teel, J. E. (1982). Sample size effects on chi square and other statistics used in evaluating causal models. *Journal of Marketing Research*, 19, 425-430.
- Beck, B., Bundt, S. & Gomolka, J. E. (2008). Ziele und Anlage der DESI-Studie. In DESI-Konsortium (Hrsg.), *Unterricht und Kompetenzerwerb zu Deutsch und Englisch. Ergebnisse der DESI-Studie* (S. 11-25). Weinheim: Beltz.
- Beck, B. & Klieme, E. (2003). DESI – Eine Längsschnittstudie zur Untersuchung des Sprachunterrichts an deutschen Schulen. *Empirische Pädagogik*, 17(3), 380-393.
- Beck, B. & Klieme, E. (Hrsg.). (2007). *Sprachliche Kompetenzen. Konzepte und Messung. DESI-Studie (Deutsch-Englisch-Schülerleistungen-International)*. Weinheim: Beltz.
- Bentler, P. M. (1990). Comparative fit indices in structural models. *Psychological Bulletin*, 107, 238-246.
- Berns, M. (1990). *Contexts of competence. Social and cultural considerations in communicative language teaching*. New York: Plenum Press.
- Bialystok, E., McBride-Chang, C. & Luk, G. (2005). Bilingualism, language proficiency, and learning to read on two writing systems. *Journal of Educational Psychology*, 97(4), 580-590.
- Bialystok, E. & Miller, B. (1999). The problem of age in second-language acquisition: Influences from language, structure, and task. *Bilingualism: Language and Cognition*, 2, 127-145.
- Bleyhl, W. (2005). Fremdsprachenlernen – „gesteuert“ oder nach den Prinzipien des Mutterspracherwerbs? *PRAXIS Fremdsprachenunterricht*, 2(3), 2-7.
- Boldt, R. F. (1988). *Latent structure analysis of the test of English as a foreign language*. Princeton, NJ: Educational Testing Service.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Bollen, K. A. & Long, J. S. (1993). Introduction. In K. A. Bollen & J. S. Long (Hrsg.), *Testing structural equation models* (S. 1-9). Newbury Park, CA: Sage.
- Bonnet, G. (Hrsg.). (2004). *The assessment of pupils' skills in English in eight European countries*. Paris: Ministère de la Jeunesse, de l'Éducation et de la Recherche.
- Borchert, J., Knopf-Jerichow, H. & Dahbashi, A. (1991). *Testdiagnostische Verfahren in Vor-, Sonder- und Regelschulen*. Heidelberg: Asanger.
- Bornstein, M. H., Hahn, C.-S. & Haynes, O. M. (2004). Specific and general language performance across early childhood: Stability and gender considerations. *First Language*, 24(3), 267-304.
- Bortz, J. (2004). *Statistik*. Berlin: Springer.
- Bos, W., Lankes, E.-M., Prenzel, M., Schwippert, K., Walther, G. & Valtin, R. (Hrsg.). (2003). *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich*. Münster: Waxmann.
- Bos, W., Lankes, E.-M., Schwippert, K., Valtin, R., Voss, A., Badel, I. et al. (2003). Lesekompetenz deutscher Grundschülerinnen und Grundschüler am Ende der vierten Jahrgangsstufe im internationalen Vergleich. In W. Bos, E.-M. Lankes, M. Prenzel, K. Schwippert, G. Walther & R. Valtin (Hrsg.), *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich* (S. 69-142). Münster: Waxmann.

- Bos, W. & Schwippert, K. (2002). TIMSS, PISA, IGLU & Co. Vom Sinn und Unsinn internationaler Schulleistungsuntersuchungen. *Bildung und Erziehung*, 55(1), 5-24.
- Brickenkamp, R. (1997). *Handbuch psychologischer und pädagogischer Tests*. Göttingen: Hogrefe.
- Brindley, G. (1998). Assessing listening abilities. *Annual Review of Applied Linguistics*, 18, 171-191.
- Brindley, G. (2001). Outcome-based assessment in practice: Some examples and emerging insights. *Language Testing*, 18(4), 393-407.
- Brown, G., Malmkjaer, K. & Williams, J. (Hrsg.). (1996). *Performance and competence in second language acquisition*. Cambridge: Cambridge University Press.
- Browne, M. W. & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research*, 21, 230-258.
- Bruner, J. (1997). *Wie das Kind sprechen lernt*. Bern: Huber.
- Brusch, W. (2000). Englisch vom ersten Schultag an. *Grundschule*, 32(6), 34-42.
- Bryk, A. S. & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.
- Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press.
- Bühler, K. (1992). *Sprachtheorie. Die Darstellungsfunktion der Sprache* (UTB. Bd. 1159). Stuttgart: Lucius und Lucius.
- Bulheller, S. & Häcker, H. (2001). *Rechtschreibtest RST Neue Rechtschreibung*. Frankfurt: Swets & Zeitlinger.
- Burnham, K. P. & Anderson, D. R. (2004). Multimodel Inference. Understanding AIC and BIC in Model Selection. *Sociological Methods & Research*, 33(2), 261-304.
- Butler, F. A. & Stevens, R. (2001). Standardized assessment of the content knowledge of English language learners K-12: Current trends and old dilemmas. *Language Testing*, 18(4), 409-427.
- Canale, M. & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1-47.
- Carroll, J. B. (1975). *The teaching of French as a foreign language in eight countries*. New York: Wiley.
- Carrell, P. L. (1991). Second language reading: Reading ability or language proficiency? *Applied Linguistics*, 12(2), 159-179.
- Carroll, J. B. (1968). The psychology of language testing. In A. Davies (Hrsg.), *Language testing symposium. A psycholinguistic approach* (S. 46-70). London: Oxford University Press.
- Carver, R. P. (1998). Predicting reading level in grades 1 to 6 from listening level and decoding level: Testing theory relevant to the simple view of reading. *Reading and Writing*, 10, 121-154.
- Cascallar, A. & Cascallar, E. (2003). Setting standards in the assessment complex performances: The optimized extended-response standard setting method. In M. Segers, P. Dochy & E. Cascallar (Hrsg.), *Optimizing new modes of assessment: In search of qualities and standards* (S. 247-266). Dordrecht: Kluwer.
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1, 245-276.
- Chalhoub-Deville, M. (1997). Theoretical models, assessment frameworks and test construction. *Language Testing*, 14(1), 3-22.
- Cheung, M. W.-L. & Au, K. (2005). Applications of multilevel structural equation modeling to cross-cultural research. *Structural Equation Modeling*, 12(4), 598-619.

- Chiswick, B. R. & Miller, P. W. (2007). *The critical period hypothesis for language learning: What the 2000 US census says*. Bonn: Forschungsinstitut zur Zukunft der Arbeit.
- Chomsky, N. (1987). *Aspekte der Syntax-Theorie*. Frankfurt am Main: Suhrkamp.
- Chomsky, N. (2001). *The architecture of language*. New Delhi: Oxford University Press.
- Coleman, J. (1997). Residence abroad within language study. *Language Teaching*, 30(1), 1-20.
- Coltheart, M. & Funnell, E. (1987). Reading and writing: One lexicon or two? In A. Allport, D. MacKay, W. Prinz & E. Scheerer (Hrsg.), *Language perception and production: Relationships between listening, speaking, reading, and writing*. London: Academic Press.
- Coseriu, E. (1988). *Sprachkompetenz: Grundzüge der Theorie des Sprechens*. Tübingen: Francke.
- Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Cronbach, L. J. (1976). *Research on classrooms and schools: Formulation of questions, designs and analysis*. Stanford: Stanford Evaluation Consortium.
- Cronbach, L. J. & Webb, N. (1979). Between class and within class effects in a reported aptitude x treatment interaction: A reanalysis of a study by G. L. Anderson. *Journal of Educational Psychology*, 67, 717-724.
- Crystal, D. (1993). *Die Cambridge-Enzyklopädie der Sprache*. Frankfurt am Main: Zweitausendeins-Verlag.
- Csapó, B. (2003). *Cognitive factors of the development of foreign language skills*. Paper presented at the 10th biennial conference of the European Association for Research on Learning and Instruction (26.-30. August 2003), Padova, Italy.
- Csapó, B. & Nikolov, M. (2002). *The relationship between students' foreign language achievement and general thinking skills*. Paper presented at the Annual meeting of the American Educational Research Association (1.-5. April 2002), New Orleans, USA.
- Cummins, J. (1980). The cross-lingual dimension of language proficiency: Implications for bilingual education and the optimal age issue. *TESOL Quarterly*, 14(2), 175-187.
- Cummins, J. (1991). Interdependence of first and second language proficiency in bilingual children. In E. Bialystok (Hrsg.), *Language Processing in Bilingual Children* (S. 70-89). New York: Cambridge University Press.
- Cummins, J. (1993). Biligualism and second language learning. *Annual Review of Applied Linguistics*, 13, 51-70.
- Cutler, A. (1987). Speaking for listening. In A. Allport, D. Mackay, W. Prinz & E. Scheerer (Hrsg.), *Language perception and production: Relationships between listening, speaking, reading, and writing* (S. 23-40). London: Academic Press.
- De Fraine, B., Van Damme, J., Van Landeghem, G., Opdenakker, M.-C. & Onghena, P. (2003). The effect of schools and classes on language achievement. *British Educational Research Journal*, 29(6), 841-859.
- De Jong, N. (2005). Can second language grammar be learned through listening? An experimental study. *Studies in Second Language Acquisition*, 27, 205-234
- Dehaene, S. (1999). Fitting two languages into one brain. *Brain*, 122(12), 2207-2208.
- Dehaene, S., Dupoux, E., Mehler, J., Cohen, L., Paulesu, E., Perani, D. et al. (1997). Anatomical variability in the cortical representation of first and second language. *Neuroreport*, 8(17), 3809-3815.
- Del Vecchio, A. & Guerrero, M. (1995). *Handbook of English language proficiency tests*. Albuquerque, NM: Highlands University.

- DESI-Konsortium. (2003). *DESI: Bericht über die Entwicklung und Erprobung der Erhebungsinstrumente. Interner Bericht für die Kultusministerkonferenz*. Frankfurt am Main: Deutsches Institut für Internationale Pädagogische Forschung.
- DESI-Konsortium (Hrsg.). (2008). *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie*. Weinheim: Beltz.
- Deutsches PISA-Konsortium (Hrsg.). (2001). *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske und Budrich.
- Diakidoy, I.-A. N., Stylianou, P., Kerefillidou, C. & Papageorgiou, P. (2005). The relationship between listening and reading comprehension of different types of text at increasing grade levels *Reading Psychology*, 26(1), 55-80.
- Diehl, E., Christen, H., Leuenberger, S., Pelvat, I. & Studer, T. (2000). *Grammatikunterricht: Alles für der Katz? Untersuchungen zum Zweitspracherwerb*. Tübingen: Niemeyer.
- Dittmar, N. (1989). Soziolinguistik. In K.-R. Bausch, H. Christ, W. Hüllen & H.-J. Krumm (Hrsg.), *Handbuch Fremdsprachenunterricht* (S. 27-34). Tübingen: Francke.
- Ditton, H. (1998). *Mehrebenenanalyse. Grundlagen und Anwendungen des Hierarchisch Linearen Modells*. Weinheim: Juventa-Verlag.
- Dresel, M., Stöger, H. & Ziegler, A. (2005). *Klassen- und Schuleffekte bei Geschlechtsunterschieden im schulischen Leistungsbereich*. Ulm: Universität Ulm.
- Dresselhaus, G. (1979). *Langue/Parole und Kompetenz/Performanz. Zur Klärung der Begriffspaare bei Saussure und Chomsky, ihre Vorgeschichte und ihre Bedeutung für die moderne Linguistik*. Frankfurt am Main: Lang.
- Dubberke, T. & Harks, B. (2008). Zur curricularen Validität der DESI-Aufgaben: Ergebnisse eines Expertenratings. In DESI-Konsortium (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie*. Weinheim: Beltz.
- Eckerth, J. (2000). Zielsprachliche Kommunikation über Grammatik im Fremdsprachenunterricht. *Zeitschrift für Fremdsprachenforschung*, 11(1), 9-30.
- Eckerth, J. & Riemer, C. (2000). Awareness und Motivation: Noticing als Bindeglied zwischen kognitiven und affektiven Faktoren des Fremdsprachenlernens. In C. Riemer (Hrsg.), *Kognitive Aspekte des Lehrens und Lernens von Fremdsprache* (S. 228-246). Tübingen: Narr.
- Educational Testing Service. (2003). *TOEFL*. – Verfügbar unter: <http://www.toefl.org>.
- Eichler, W. (2007). Sprachbewusstheit. In B. Beck & E. Klieme (Hrsg.), *Sprachliche Kompetenzen. Konzepte und Messung. DESI-Studie (Deutsch-Englisch-Schülerleistungen-International)* (S. 147-157). Weinheim: Beltz.
- Eichler, W. (2008). Sprachbewusstheit Deutsch. In DESI-Konsortium (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (S. 112-119). Weinheim: Beltz.
- Eichler, W. & Thomé, G. (2008). Rechtschreiben Deutsch. In DESI-Konsortium (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (S. 104-111). Weinheim: Beltz.
- Eid, M. (1999). Lineare Strukturgleichungsmodelle. In B. Strauß, H. Haag & M. Kolb (Hrsg.), *Datenanalyse in der Sportwissenschaft* (S. 427-454). Schorndorf: Hofmann.
- Elley, W. B. (Hrsg.). (1994). *The IEA-Study of reading literacy: Achievement and instruction in thirty-two school systems*. Oxford: Pergamon Press.
- Engen, L. & Høien, T. (2002). Phonological skills and reading comprehension. *Reading and Writing*, 15, 613-631.

- Europäische Kommission (2002). *Sprachunterricht und Spracherwerb. Maßnahmen der Europäischen Gemeinschaft.* – Verfügbar unter: [ec.europa.eu/education/policies/lang/policy/consult/action\\_de.pdf](http://ec.europa.eu/education/policies/lang/policy/consult/action_de.pdf)
- Faerch, C. & Kasper, G. (1987). Perspectives on Language Transfer. *Applied Linguistics*, 8(2), 111-136.
- Fehling, S. (2003). Der Einfluss von bilinguaem Unterricht auf die Entwicklung von Language Awareness. *Empirische Pädagogik*, 17(3), 332-346.
- Feldmann, U., Grotjahn, R. & Stemmer, B. (1986). Was messen Sprachtests eigentlich? Überlegungen zur introspektiven Validierung von C-Tests. In Bausch, K.-R. (Hrsg.). *Probleme und Perspektiven der Sprachlehrforschung. Bochumer Beiträge zum Fremdsprachenunterricht in Forschung und Lehre* (S. 325-338). Frankfurt am Main: Scriptor-Verlag.
- Felix, S. W. (1977). Kreative und reproduktive Kompetenz im Zweitspracherwerb. In H. Hunfeld (Hrsg.), *Neue Perspektiven der Fremdsprachendidaktik. Eichstätter Kolloquium zum Fremdsprachenunterricht* (S. 25-34). Kronberg/Ts. : Scriptor-Verlag.
- Feyten, C. M. (1991). The power of listening ability: An overlooked dimension in language acquisition. *The Modern Language Journal*, 75(2), 173-180.
- Fippinger, F. (1991). *Allgemeiner Schulleistungstest für 3. Klassen AST 3*. Weinheim: Beltz.
- Fitzgerald, J. & Shanahan, T. (2000). Reading and writing relations and their development. *Educational Psychologist*, 35(1), 39-50.
- Flege, J. E. (1987). A critical period for learning to pronounce foreign languages? *Applied Linguistics*, 8(1), 162-177.
- Foy, P. (1999). Sampling weights. In M. O. Martin, K. D. Gregory & S. E. Stemler (Hrsg.), *TIMSS 1999 Technical Report* (S. 189-200). Chesnut Hill: International Study Center, Lynch School of Education, Boston College.
- Friederici, A. D. & Hahne, A. (2000). Neurokognitive Aspekte der Sprachentwicklung. In H. Grimm (Hrsg.), *Sprachentwicklung* (Enzyklopädie der Psychologie. Bd. C/III/3, S. 273-310). Göttingen: Hogrefe.
- Gailberger, S. & Willenberg, H. (2008). Leseverstehen Deutsch. In DESI-Konsortium (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (S. 60-71). Weinheim: Beltz.
- Gardner, R. C., Lalonde, R. N. & Pierson, R. (1983). The socio-educational model of second language acquisition: An investigation using LISREL causal modeling. *Journal of Language and Social Psychology*, 2(1), 1-15.
- Gardner, R. C. & MacIntyre, P. D. (1993). On the measurement of affective variables in second language learning. *Language Learning*, 42(2), 157-194.
- Garrett, P. & James, C. (1991). Language awareness: A way ahead. In C. James & P. Garrett (Hrsg.), *Language awareness in the classroom* (S. 306-318). London: Longman.
- Genesee, F. (1976). The role of intelligence in second language learning. *Language Learning*, 26(2), 267-280.
- Ginther, A. & Stevens, J. (1998). Language background, ethnicity, and the internal construct validity of the advanced placement Spanish language examination. In A. J. Kunnan (Hrsg.), *Validation in language assessment* (S. 169-194). Mahawah, NJ: Erlbaum.
- Gnutzmann, C. (1997). Language awareness. Geschichte, Grundlagen, Anwendungen. *Praxis des neusprachlichen Unterrichts*, 44(3), 227-236.
- Golinski, B. & Doetjes, G. (2005). *Sprachverstehensuntersuchungen im semikommunikativen Kontext*. Hamburg: Universität Hamburg.
- Götter, K.-H. (1978). *Argumentation. Grundzüge ihrer Theorie im Bereich theoretischen Wissens und praktischen Handelns*. Tübingen: Niemeyer.

- Gräfe-Bentzien, S. (2000). *Evaluierung bilingualer Kompetenz. Eine Pilotstudie zur Entwicklung der deutschen und italienischen Sprachfähigkeiten in der Primarstufe beim Schulversuch der Staatlichen Europa-Schule Berlin (SESB)*. Berlin: Freie Universität Berlin.
- Grimm, H. (Hrsg.). (2000). *Sprachentwicklung* (Enzyklopädie der Psychologie. Bd. C/III/3). Göttingen: Hogrefe.
- Groeben, N. & Hurrelmann, B. (Hrsg.). (2002). *Lesekompetenz: Bedingungen, Dimensionen, Funktionen*. Weinheim: Juventa-Verlag 2002.
- Grohnfeldt, M. (1993). *Störungen der Sprachentwicklung*. Berlin: Marhold.
- Grotjahn, R. (1987). How to construct and evaluate a C-Test: A discussion of some problems and some statistical analyses. In R. Grotjahn, C. Klein-Braley & D. K. Stevenson (Hrsg.), *Taking their measures: The validity and validation of language tests* (S. 219-253). Bochum: Brockmeyer.
- Grotjahn, R. (2000a). Determinanten der Schwierigkeit von Leseverstehensaufgaben: Theoretische Grundlagen und Konsequenzen für die Entwicklung des TESTDAF. In S. Bolton (Hrsg.), *TESTDAF: Grundlagen für die Entwicklung eines neuen Sprachtests. Beiträge aus einem Expertenseminar* (S. 7-55). Köln: VUB Printmedia.
- Grotjahn, R. (2000b). Kognitive Determinanten der Schwierigkeit von fremdsprachlichen Lese- und Hörverstehensaufgaben. Zur Prognose der Aufgabenschwierigkeit von Sprachtests. In C. Riemer (Hrsg.), *Kognitive Aspekte des Lehrens und Lernens von Fremdsprachen* (S. 338-353). Tübingen: Narr.
- Grotjahn, R. & Stemmer, B. (1985). On the development and evaluation of a C-Test for French. *Fremdsprachen und Hochschule*, 13/14, 101-120.
- Haberzettl, S. & Wegener, H. (Hrsg.). (2003). *Spracherwerb und Konzeptualisierung*. Frankfurt am Main: Lang.
- Haertel, E. & Wiley, D. E. (1986). Die Schätzung von Beziehungen auf Individualebene durch Daten auf Aggregatebene: ein Beispiel. In M. von Saldern (Hrsg.), *Mehrebenenanalyse. Beiträge zur Erfassung hierarchisch strukturierter Realität* (S. 135-145). Weinheim: Psychologie-Verlags-Union.
- Hagtvet, B. (2003). Listening comprehension and reading comprehension in poor decoders: Evidence for the importance of syntactic and semantic skills as well as phonological skills. *Reading and Writing*, 16, 505-539.
- Hale, G. A., Rock, D. A. & Jirele, T. J. (1989). *Confirmatory factor analysis of the test of English as a foreign language*. Princeton, NJ: Educational Testing Service.
- Hamers, J. F. & Blanc, M. H. A. (2000). *Bilinguality and bilingualism*. Cambridge: Cambridge University Press.
- Hamp-Lyons, L. & Kroll, B. (1997). *TOEFL 2000 writing: composition, community and assessment*. Princeton, NJ: Educational Testing Service.
- Härnquist, K., Gustafsson, J.-E., Muthén, B. & Nelson, G. (1994). Hierarchical models of ability at individual and class levels. *Intelligence*, 18, 165-187.
- Harris, D. P. (1969). *Testing English as a second language*. New York: McGraw-Hill.
- Harsch, C., Neumann, A., Lehmann, R. & Schröder, K. (2007). Schreibfähigkeit. In B. Beck & E. Klieme (Hrsg.), *Sprachliche Kompetenzen. Konzepte und Messung. DESI-Studie (Deutsch-Englisch-Schülerleistungen-International)* (S. 42-62). Weinheim: Beltz.
- Harsch, C. & Schröder, K. (2007). Textrekonstruktion: C-Test. In B. Beck & E. Klieme (Hrsg.), *Sprachliche Kompetenzen. Konzepte und Messung. DESI-Studie (Deutsch-Englisch-Schülerleistungen-International)* (S. 212-225). Weinheim: Beltz.

- Harsch, C. & Schröder, K. (2008). Textrekonstruktion Englisch. In DESI-Konsortium (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (S. 149-156). Weinheim: Beltz.
- Harsch, C., Schröder, K. & Neumann, A. (2008). Schreiben Englisch. In DESI-Konsortium (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (S. 130-138). Weinheim: Beltz.
- Hartig, J., Jude, N. & Wagner, W. (2008). Methodische Grundlagen der Messung und Erklärung sprachlicher Kompetenzen. In DESI-Konsortium (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (S. 34-54). Weinheim: Beltz.
- Hartig, J. & Jude, N. (2008). Sprachkompetenzen von Mädchen und Jungen. In DESI-Konsortium (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (S. 202-207). Weinheim: Beltz.
- Hartig, J. & Kühnbach, O. (2006). Schätzung von Veränderung mit „plausible values“ in mehrdimensionalen Rasch-Modellen. In A. Ittel & H. Merckens (Hrsg.), *Veränderungsmessung und Längsschnittstudien in der empirischen Erziehungswissenschaft* (S. 27-44). Wiesbaden: VS, Verlag für Sozialwissenschaften.
- Hartig, M. (1999). *Psycholinguistik des Deutschen* (Germanistische Lehrbuchsammlung. Bd. 17). Berlin: Weidler.
- Heck, R. H. (2001). Multilevel modeling with SEM. In G. A. Marcoulides & R. E. Schumacker (Hrsg.), *New development and techniques in structural equation modeling* (S. 89-128). Mahwah, NJ: Erlbaum.
- Heller, M. F. (1999). *Reading-writing connections: From theory to practice*. Mahwah, NJ: Erlbaum.
- Helmke, A., Helmke, T., Schrader, F.-W., Wagner, W., Nold, G. & Schröder, K. (2008). Alltagspraxis des Englischunterrichts. In DESI-Konsortium (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (S. 371-181). Weinheim: Beltz.
- Helmke, A. & Hosenfeld, I. (2004). Vergleichsarbeiten – Standards – Kompetenzstufen: Begriffliche Klärung und Perspektiven. In R. S. Jäger, A. Frey & M. Wosnitza (Hrsg.), *Lernprozesse, Lernumgebung und Lerndiagnostik* (S. 56-76). Landau: Verlag Empirische Pädagogik.
- Helmke, A. & Schrader, F.-W. (2001). Jenseits von TIMSS: Messungen sprachlicher Kompetenzen, komplexe Längsschnittstudien und kulturvergleichende Analysen. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 237-250). Weinheim: Beltz.
- Helmke, A. & Weinert, F. E. (1997). Bedingungsfaktoren schulischer Leistungen. In F. E. Weinert (Hrsg.), *Psychologie des Unterrichts und der Schule* (Enzyklopädie der Psychologie. Bd. D/I/3, S. 71-176). Göttingen: Hogrefe.
- Hennon, E., Hirsh-Pasek, K. & Golinkoff, R. M. (2000). Die besondere Reise vom Fötus zum spracherwerbenden Kind. In H. Grimm (Hrsg.), *Sprachentwicklung* (Enzyklopädie der Psychologie. Bd. C/III/3, S. 41-103). Göttingen: Hogrefe.
- Henrici, G. (2000). Solo für Kognitives? Für eine mehrperspektivische Erforschung des Fremdsprachenwerberbs. In C. Riemer (Hrsg.), *Kognitive Aspekte des Lehrens und Lernens von Fremdsprache* (S. 171-185). Tübingen: Narr.
- Hesse, G., Göbel, K. & Hartig, H. (2008). Sprachliche Kompetenzen von mehrsprachigen Jugendlichen und Jugendlichen nicht-deutscher Erstsprache. In DESI-Konsortium (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (S. 208-230). Weinheim: Beltz.

- Hessisches Kultusministerium. (1995). *Rahmenplan Grundschule*. Frankfurt am Main: Diesterweg.
- Hirai, A. (1999). The relationship between listening and reading rates of Japanese EFL learners. *The Modern Language Journal*, 83(3), 367-384.
- Homae, F., Haschimoto, R., Nakajama, K., Miyashita, Y. & Sakai, K. L. (2002). From perception to sentence comprehension: The convergence of auditory and visual information of language in the left inferior frontal cortex. *NeuroImage*, (16), 883-900.
- Homburg, C. & Baumgartner, H. (1995). Beurteilung von Kausalmodellen. *Marketing: Zeitschrift für Forschung und Praxis*, 17, 162-176.
- Hoover, W. A. & Gough, P. B. (1990). The simple view of reading. *Reading and Writing*, 2, 127-160.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30, 179-185.
- Hosenfeld, I., Helmke, A., Ridder, A. & Schrader, F.-W. (2001). Eine mehrbenenanalytische Betrachtung von Schul- und Klasseneffekten. *Empirische Pädagogik*, 15(4), 513-534.
- Hosley, D. & Meredith, K. (1979). Inter- and intra-test correlates of the TOEFL. *TESOL Quarterly*, 13(2), 209-217.
- Hovestadt, G. & Keßler, N. (2005). Sechzehn Bundesländer. Eine Übersicht zu Bildungsstandards und Evaluationen. In Becker, G. (Hrsg.), *Standards. Unterrichten zwischen Kompetenzen, zentralen Prüfungen und Vergleichsarbeiten* (Friedrich-Jahresheft. Bd. 23, S. 8-10). Seelze: Friedrich.
- Hox, J. J. (1995). *Applied multilevel analysis*. Amsterdam: TT-Publikaties.
- Hox, J. J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Erlbaum.
- Hox, J. J. & Bechger, T. M. (1998). An introduction to structural equation modeling. *Family Science Review*, 11, 354-373.
- Hox, J. J. & Maas, C. J. M. (2005). Multilevel analysis. In K. Kempf-Leonard (Hrsg.), *The encyclopedia of social measurement* (Bd. 2, S. 785-793). San Diego: Elsevier.
- Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1-55.
- Hufeisen, B. (2003). L1, L2, L3, L4, Lx – alle gleich? Linguistische, lernerinterne und lernerexterne Faktoren in Modellen zum multiplen Spracherwerb. *Zeitschrift für Interkulturellen Fremdsprachenunterricht*, 8(2/3), 1-13.
- Hyde, J. S. & Linn, M. C. (1988). Gender differences in verbal ability: A meta analysis. *Psychological Bulletin*, 102(1), 53-69.
- Hymes, D. (1967). Models of the interaction of language and social setting. *Journal of Social Issues*, 13(2), 8-28.
- Hymes, D. (1971). Competence and performance in linguistic theory. In R. Huxley & E. Ingram (Hrsg.), *Language Acquisition and Methods*. New York: Academic Press.
- Ingendahl, W. (1999). *Sprachreflexion statt Grammatik*. Tübingen: Niemeyer.
- Ingenkamp, K. (1985). 1955 – 1985: Drei Dekaden pädagogischer Diagnostik in Deutschland: Entwicklungen, Kontroversen und Perspektiven. *Tests und Trends*, 4, 13-40.
- James, L. R. (1982). Aggregation bias in the estimates of perceptual agreement. *Journal of Applied Psychology*, 67, 219-229.
- Jang, E. E. & Roussos, L. (2007). An Investigation into the dimensionality of TOEFL using conditional covariance-based nonparametric approach. *Journal of Educational Measurement*, 44(1), 1-21.

- Jenkins, J. R., Broek, P. van den, Espin, C. & Deno, S. L. (2003). Sources of individual differences in reading comprehension and reading fluency. *Journal of Educational Psychology*, 95(4), 719-729.
- Jenkins, J. R., Johnson, E. & Hileman, J. (2004). When is reading also writing: Sources of individual differences on the new reading performance assessments. *Scientific Studies on Reading*, 8(2), 125-151.
- Jerusalem, M. (1997). Schulklasseneffekte. In F. E. Weinert (Hrsg.), *Psychologie des Unterrichts und der Schule* (Enzyklopädie der Psychologie. Bd. D/I/3, S. 253-278). Göttingen: Hogrefe.
- Johnson, J. W. (2004). Factors affecting relative weights: The influence of sampling and measurement error. *Organizational Research Methods*, 7(3), 283-299.
- Jöreskog, K. G. (1993). Testing structural equation models. In K. A. Bollen & J. S. Long (Hrsg.), *Testing structural equation models* (S. 294-316). Newbury Park, CA: Sage.
- Jöreskog, K. G. & Sörbom, D. (1989). *LISREL 7. A guide to the program and applications*. Chicago: SPSS Inc.
- Jöreskog, K. G. & Sörbom, D. (1996). *LISREL 8*. Uppsala: Scientific Software International.
- Jude, N. & Hartig, J. (2005). *Weighted likelihood oder plausible values? Die Verwendung unterschiedlicher Personenschätzer in mehrdimensionalen Mehrebenenmodellen*. Vortrag auf der 7. Tagung der Fachgruppe Methoden und Evaluation der DGPs, Münster i. W.
- Jude, W. K. (1975). *Deutsche Grammatik*. Braunschweig: Westermann.
- Julian, M. W. (2001). The consequences of ignoring multilevel data structures in non-hierarchical covariance modeling. *Structural Equation Modeling*, 8, 325-352.
- Jung, U. O. H. (1978). *Das Sprachlabor. Möglichkeiten und Grenzen technischer Medien im Unterricht*. Königstein: Scriptor.
- Jurecka, A. (2007). Vergleichbarkeit der Messung von Fremdsprachenkompetenzen im europäischen Raum. *Dipf Informiert*, 11, 12-15.
- Jusczyk, P. W. (2003). The role of speech perception capacities in early language acquisition. In M. T. Banich & M. Mack (Hrsg.), *Mind, brain, and language* (S. 61-86). Mahwah: Erlbaum.
- Käsermann, M.-L. & Foppa, K. (2003). Sprachproduktion im Gespräch. In T. Herrmann & J. Grabowski (Hrsg.), *Sprachproduktion* (Enzyklopädie der Psychologie. Bd. C/III/1, S. 767-800). Göttingen: Hogrefe.
- Kahn-Horwitz, J., Joseph, S. & Sparks, R. L. (2005). Predicting foreign language reading achievement in elementary school students. *Reading and Writing*, 18, 527-558.
- Kaplan, D. (2000). *Structural equation modeling: Foundations and extensions*. Thousand Oaks, CA: Sage.
- Karmiloff-Smith, A. (1979). *A functional approach to child language: A study of determiners and reference*. New York: Cambridge University Press.
- Kellogg, R. T. (2003). Schriftliche Sprachproduktion. In T. Herrmann & J. Grabowski (Hrsg.), *Sprachproduktion* (Enzyklopädie der Psychologie. Bd. C/III/1, S. 531-559). Göttingen: Hogrefe.
- Klein, K. J. & Kozlowski, S. W. (2000). From micro to meso: critical steps in conceptualizing and conducting multilevel research. *Organizational Research Methods*, 3(3), 211-236.
- Klein, W. (2000). Prozesse des Zweitspracherwerbs. In H. Grimm (Hrsg.), *Sprachentwicklung* (Enzyklopädie der Psychologie. Bd. C/III/3, S. 537-570). Göttingen: Hogrefe.
- Klein-Braley, C. (1985). Reduced redundancy as an approach to language testing. *Fremdsprachen und Hochschule*, 13/14, 1-13.

- Klieme, E. (2004). Was sind Kompetenzen und wie lassen sie sich messen? *Pädagogik*, 56(6), 10-13.
- Klieme, E. (2006). Kompetenzerwerb in Deutsch und Englisch: Ergebnisse der DESI-Studie. *Schulverwaltung. Ausgabe Hessen und Rheinland-Pfalz*, 11(5), 137-140.
- Klieme, E., Eichler, W., Helmke, A., Lehmann, R. H., Nold, G., Rolff, H.-G. et al. (2006). *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Zentrale Befunde der Studie Deutsch Englisch Schülerleistungen International*. Frankfurt am Main: Deutsches Institut für Internationale Pädagogische Forschung.
- Klieme, E., Jude, N., Rauch, D., Ehlers, H., Helmke, A., Eichler, W., et al. (2008). Alltagspraxis, Qualität und Wirksamkeit des Deutschunterrichts. In DESI-Konsortium (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (S. 319-144). Weinheim: Beltz.
- Köller, O. (2007). Bildungsstandards, einheitliche Prüfungsanforderungen und Qualitätssicherung in der Sekundarstufe II. In D. Benner (Hrsg.), *Bildungsstandards. Chancen und Grenzen, Beispiele und Perspektiven* (S. 13-28). Paderborn: Schöningh.
- Köller, O. (2008). Operationalisierung und Überprüfung von Bildungsstandards: Ist Bildung messbar? In F. Hofmann, C. Schreiner & J. Thonhauser (Hrsg.), *Qualitative und quantitative Aspekte: Zu ihrer Komplementarität in der erziehungswissenschaftlichen Forschung* (S. 281-298). Münster: Waxmann.
- Kramsch, C. (2002). Standard, norm, and variability in language learning. In S. Gass, K. Bardovi-Harlig, S. Sieloff Magnan & J. Walz (Hrsg.), *Paedagogical norms for second and foreign language learning and teaching* (Language learning and language teaching. Bd. 5, S. 59-80). Amsterdam: Benjamins.
- Kreft, I. & Leeuw, J. de (1998). *Introducing multilevel modeling* (ISM, Introducing statistical methods). London: Sage.
- Krelle, M. & Willenberg, H. (2008). Argumentation Deutsch. In DESI-Konsortium (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (S. 81-88). Weinheim: Beltz.
- Kuhlemeier, H., van den Bergh, H. & Rijlaarsdam, G. (2002). The dimensionality of speaking and writing: A multilevel factor analysis of situational, task and school effects. *British Journal of Educational Psychology*, Vol. 72, 467-482.
- Kunnan, A. J. (1992). An investigation of a criterion-referenced test using G-theory, and factor and cluster analyses. *Language Testing*, 9(1), 30-44.
- Kunnan, A. J. (1994). Modelling relationships among some tests-taker characteristics and performance on EFL tests: An approach to construct validation. *Language Testing*, 11, 225-251.
- Lado, R. (1961). *Language testing: The construction and the use of foreign language tests*. London: Longman.
- Langer, J. & Flihan, S. (2000). Writing and reading relationships: Constructive tasks. In R. Indrisano & J. R. Squire (Hrsg.), *Writing: Research / Theory / Practice*. Newark, DE: International Reading Association.
- Larson, J. W. (1983). Skill correlations: A study of three final examinations. *The Modern Language Journal*, 67(3), 228-234.
- Lehmann, R. H., Peek, R., Gänsfuß, R. & Husfeldt, V. (2001). *LAU 9. Aspekte der Lernausgangslage und der Lernentwicklung – Klassenstufe 9*. Hamburg: Behörde für Bildung und Sport.
- Lewis, E. G. & Massad, C. E. (1975). *The teaching of English as a foreign language in ten countries*. Stockholm: Almqvist & Wiksell.

- Lightbown, P. M. & Spada, N. (2000). Do they know what they're doing? L2 learners' awareness of L1 influence. *Language Awareness*, 9(4), 198-217.
- Linke, A., Nussbaumer, M. & Portmann, P. R. (2001). *Studienbuch Linguistik*. Tübingen: Niemeyer.
- List, G. (1989). Psycholinguistik und Sprachpsychologie. In K.-R. Bausch, H. Christ, W. Hüllen & H.-J. Krumm (Hrsg.), *Handbuch Fremdsprachenunterricht* (S. 20-27). Tübingen: Francke.
- Luchtenberg, S. (1995). Language Awareness-Konzeptionen. Ein Weg zur Aktualisierung des Lernbereichs „Reflexion über Sprache“. *Der Deutschunterricht*, 47(4), 93-108.
- Lüdtke, O., Robitzsch, A. & Köller, O. (2002). Statistische Artefakte bei Kontexteffekten in der pädagogisch-psychologischen Forschung. *Zeitschrift für Pädagogische Psychologie*, 16(4), 217-231.
- Lukesch, H. (1994). *Einführung in die pädagogisch-psychologische Diagnostik* (Psychologie in der Lehrerbildung. Bd. 3). Regensburg: CH-Verlag.
- Lundberg, I. & Rosen, M. (1995). *Two-level structural modeling of reading achievement as a basis for evaluationg teaching effects*. Paper presented at the annual meeting of the American Educational Research Association (April 18-22, 1995) San Fransisco, California.
- Mais, S. (2006). Die Schule – ein jugenfeindliches Biotop? Geschlechterbewusste Pädagogik in der Schule. *Pädagogische Beiträge: Unterricht und Schulleben in Rheinland-Pfalz*, (1), 12-15.
- Masters, G. N. & Wright, B. D. (1997). The partial credit model. In W. van der Linden & R. K. Hambleton (Hrsg.), *Handbook of modern Item Response Theorie* (S. 101-122). New York: Springer.
- McKay, D. G. (1987). Introduction. In A. Allport, D. MacKay, W. Prinz & E. Scheerer (Hrsg.), *Language perception and production. Relationships between listening, speaking, reading, and writing* (S. 403-406). London: Academic Press.
- McNamara, T. F. (1995). Modelling performance: Opening Pandora's box. *Applied Linguistics*, 16(2), 159-175.
- Mehta, P. D., Foorman, B. R., Branum-Martin, L. & Taylor, W. P. (2005). Literacy as a unidimensional multilevel construct: Validation, sources of influence, and implications in a logitudinal study in grades 1 to 4. *Scientific Studies on Reading*, 9(2), 85-116.
- Meschyan, G. & Hernandez, A. (2002). Is native-language decoding skill related to second-language learning? *Journal of Educational Psychology*, 94(1), 14-22.
- Messelken, H. (1971). *Empirische Sprachdidaktik*. Heidelberg: Quelle & Meyer.
- Meyer, M. A. (2001). Englischunterricht aus allgemeindidaktischer Sicht. In H.-E. Tenorth (Hrsg.), *Kerncurriculum Oberstufe. Mathematik – Deutsch – Englisch. Expertise im Auftrag der Ständigen Konferenz der Kultusminister* (S. 230-251). Weinheim: Beltz.
- Mislevy, R. J., Beaton, A. E., Kaplan, R. B. & Sheehan, K. M. (1992). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement*, 29(2), 133-161.
- Mohanty, A. K. & Perregaux, C. (1997). Language acquisition and bilingualism. In J. W. Berry, P. R. Dasen & T. S. Saraswathi (Hrsg.), *Handbook of cross-cultural psychology: Basic processes and human development* (Bd. 2, S. 217-254). New-York: Allyn & Bacon.
- Monsell, S. (1987). On the relation between lexical input and output pathways for speech. In A. Allport, D. MacKay, W. Prinz & E. Scheerer (Hrsg.), *Language perception and*

- production: Relationships between listening, speaking, reading, and writing* (S. 273-312). London: Academic Press.
- Moosbrugger, H. & Hartig, J. (2003). Faktorenanalyse. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 137-145). Weinheim: Beltz.
- Moosbrugger, H. & Kelava, A. (Hrsg.). (2007). *Testtheorie und Fragebogenkonstruktion*. Heidelberg: Springer.
- Moosbrugger, H. & Schermelleh-Engel, K. (2007). Exploratorische (EFA) und Konfirmatorische Faktorenanalyse (CFA). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 307-324). Heidelberg: Springer.
- Mullis, I. V. S., Martin, M. O., Gonzalez, E. J. & Kennedy, A. M. (2003). *PIRLS 2001 international report*. Chestnut Hill, MA: Boston College.
- Muthén, B. O. (1989). Latent variable modelling in heterogeneous populations. *Psychometrika*, 54, 557-585.
- Muthén, B. O. (1990). *Means and covariance structure analysis of hierarchical data* (UCLA statistics series. Bd. 62). Los Angeles, CA: University of California at Los Angeles.
- Muthén, B. O. (1991). Multilevel factor analysis of class and student achievement components. *Journal of Educational Measurement*, 28(4), 338-354.
- Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods & Research*, 22, 376-398.
- Muthén, B. O. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika*, 29(1), 81-117.
- Muthén, B. O. & Muthén, L. (1998). *Mplus user's guide*. Los Angeles, CA: Muthén & Muthén.
- Muthén, B. O. & Muthén, L. (2004). *Mplus. Statistical analysis with latent variables. User's guide*. Los Angeles, CA: Muthén & Muthén.
- Nelde, P. H. (2001). Eine neue Sprachpolitik für Europa? *Akademie-Journal*, 2, 25-30.
- Neuland, E. (2002). Sprachbewusstsein – eine zentrale Kategorie für den Sprachunterricht. *Der Deutschunterricht*, 54(3), 4-10.
- Neumann, A. & Lehmann, R. (2008). Schreiben Deutsch. In DESI-Konsortium (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (S. 89-103). Weinheim: Beltz.
- Nickel, S. (2002). *Funktionaler Analphabetismus – Ursachen und Lösungsansätze hier und anderswo*. Vortrag auf der IV. Internationalen Woche des Fachbereiches Erziehungs- und Bildungswissenschaften der Universität Bremen. – Verfügbar unter: [http://elib.suub.uni-bremen.de/publications/ELibD890\\_Nickel-Analphabetismus.pdf](http://elib.suub.uni-bremen.de/publications/ELibD890_Nickel-Analphabetismus.pdf) [13.12.2005]
- Nold, G., Hartig, J., Hinz, S. & Rossa, H. (2008). Klassen mit bilinguaem Sachfachunterricht: Englisch als Arbeitssprache. In DESI-Konsortium (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (S. 451-457). Weinheim: Beltz.
- Nold, G. & Rossa, H. (2007a). Hörverstehen. In K. Beck & E. Klieme (Hrsg.), *Sprachliche Kompetenzen. Konzepte und Messung. DESI-Studie (Deutsch-Englisch-Schülerleistungen-International)* (S. 178-196). Weinheim: Beltz.
- Nold, G. & Rossa, H. (2007b). Leseverstehen. In B. Beck & E. Klieme (Hrsg.), *Sprachliche Kompetenzen. Konzepte und Messung. DESI-Studie (Deutsch-Englisch-Schülerleistungen-International)* (S. 197-211). Weinheim: Beltz.

- Nold, G. & Rossa, H. (2007c). Sprachbewusstheit. In B. Beck & E. Klieme (Hrsg.), *Sprachliche Kompetenzen. Konzepte und Messung. DESI-Studie (Deutsch-Englisch-Schülerleistungen-International)* (S. 226-244). Weinheim: Beltz.
- Nold, G. & Rossa, H. (2008a). Hörverstehen Englisch. In DESI-Konsortium (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (S. 120-129). Weinheim: Beltz.
- Nold, G. & Rossa, H. (2008b). Sprachbewusstheit Englisch. In DESI-Konsortium (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (S. 157-169). Weinheim: Beltz.
- Nold, G. & Rossa, H. (2008c). Sprechen Englisch. In DESI-Konsortium (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (S. 170-179). Weinheim: Beltz.
- Nold, G., Rossa, H. & Chatzivassiliadou, K. (2008). Leseverstehen Englisch. In DESI-Konsortium (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (S. 130-138). Weinheim: Beltz.
- Nord, J. R. (1980). Developing listening fluency before speaking: An alternative paradigm *System*, 8(1), 1-22.
- North, B. (2000). *The development of a common framework scale of language proficiency* (Theoretical studies in second language acquisition. Bd. 8). New York: Lang.
- Oerter, R. & Montada, L. (2002). *Entwicklungspsychologie*. Weinheim: Psychologie-Verlags-Union.
- Oller, J. W. jr. (1973). Cloze tests of second language proficiency and what they measure. *Language Learning*, 23, 105-118.
- Oller, J. W. jr. (1976). Evidence for a general language proficiency factor: An expectancy grammar. *Die Neueren Sprachen*, 75(2), 165-174.
- Oltman, P. K., Stricker, L. J. & Barrows, T. (1988). *Native language, English proficiency, and the structure of the test of English as a foreign language*. Princeton, NJ: Educational Testing Service.
- Ordinate. (2004). *Spoken English Test-10*. Menlo Park, CA: Ordinate.
- Parodi, G. (2007). Reading-writing connections: Discourse-oriented research. *Reading and Writing*, 20(3), 225-250.
- Pellegrini, E. K. & Scandura, T. A. (2005). Construct equivalence across groups: An unexplored issue in mentoring research. *Educational and Psychological Measurement*, 65(2), 323-335.
- Perleth, C. & Heller, K. A. (2000). *Kognitiver Fähigkeitstest für 4.-12. Klassen, Revision (KFT 4-12+ R)*. Göttingen: Hogrefe.
- Piaget, J. (1983). *Sprechen und Denken des Kindes*. Frankfurt am Main: Ullstein.
- Plaut, D. C. (2003). Connectionist modelling of language: Examples and implications. In M. T. Banich & M. Mack (Hrsg.), *Mind, brain, and language* (S. 143-167). Mahwah, NJ: Erlbaum.
- Powers, D. E. (1982). Electing samples for testing the hypothesis of divisible versus unitary competence in language proficiency. *Language Learning*, 32(2), 331-335.
- Purves, A. C. (Hrsg.). (1992). *The IEA study of written composition II: Education and performance in fourteen countries*. Oxford: Pergamon Press.
- Raatz, U. (1985). Test of reduced redundancy – the C-test, a practical example. *Fremdsprachen und Hochschule*, 13/14, 66-71.
- Raatz, U. & Klein-Braley, C. (1983). Ein neuer Ansatz zur Messung der Sprachleistung. Der C-Test: Theorie und Praxis. *Tests und Trends*, 3, 107-138.

- Raatz, U. & Klein-Braley, C. (1992). *CT-D4 Schulleistungstest Deutsch für 4. Klassen. Beiheft mit Anleitungen und Normentabellen*. Weinheim: Beltz.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.
- Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. London: Sage.
- Raudenbush, S. W., Bryk, A. & Congdon, R. (2007). HLM 6 (Version 6.0). Lincolnwood: Scientific Software International.
- Rea-Dickins, P. (1997). The testing of grammar In C. Clapham & D. Corson (Hrsg.), *Language testing and assessment* (Encyclopedia of language and education. Bd. 7, S. 87-97). Dordrecht: Kluwer.
- Rea-Dickins, P. (2001). Mirror, mirror on the wall: Identifying processes of classroom assessment. *Language Testing*, 18(4), 429-462.
- Reich, H. H. & Roth, H.-J. (2002). *Spracherwerb zweisprachig aufwachsender Kinder und Jugendlicher. Ein Überblick über den Stand der nationalen und internationalen Forschung*. Hamburg: Behörde für Bildung und Sport, Amt für Schule.
- Richter, T. & Christmann, U. (2002). Lesekompetenz: Prozessebenen und interindividuelle Unterschiede. In N. Groeben & B. Hurrelmann (Hrsg.), *Lesekompetenz. Bedingungen, Dimensionen, Funktionen* (S. 25-58). Weinheim: Juventa-Verlag.
- Riemer, C. (Hrsg.). (2000). *Kognitive Aspekte des Lehrens und Lernens von Fremdsprachen*. Tübingen: Narr.
- Robinson, P. (2001). Individual differences, cognitive abilities, aptitude complexes and learning conditions in second language acquisition. *Second Language Research*, 17(4), 368-392.
- Robinson, W. S. (1950). Ecological correlations and the behaviour of individuals. *American Sociological Review*, 15, 351-357.
- Roeder, P. M. & Treumann, K.-P. (1974). *Dimensionen der Schulleistung*. Stuttgart: Klett.
- Rost, D. H. (1989). Reading comprehension: Skill or skills? *Journal of Research in Reading*, 12(2), 87-113.
- Rost, D. H. & Hartmann, A. (1992). Lesen, Hören, Verstehen. *Zeitschrift für Psychologie*, 200, 345-361.
- Rost, J. (2004a). *Lehrbuch Testtheorie / Testkonstruktion*. Bern: Huber.
- Rost, J. (2004b). Psychometrische Modelle zur Überprüfung von Bildungsstandards anhand von Kompetenzmodellen. *Zeitschrift für Pädagogik*, 50(5), 662-678.
- Rost, M. (1990). *Listening in Language Learning*. New York: Longman.
- Rowe, K. J. & Hill, P. W. (1998). Modeling educational effectiveness in classrooms: The use of multi-level structural equations to model students' progress. *Educational Research and Evaluation*, 4(4), 307-347.
- Royer, J. M., Sinatra, G. M. & Schumer, H. (1990). Patterns of individual differences in the development of listening and reading comprehension. *Contemporary Educational Psychology*, 15(2), 183-196.
- Rubin, J. (1994). A review of second language listening comprehension research. *The Modern Language Journal*, 78(2), 199-221.
- Rupp, A. A. (2005). Maximum likelihood item response theory estimation. In B. Everitt & D. Howell (Hrsg.), *Encyclopedia of statistics in behavioral science*. New York: Wiley.
- Sang, F., Schmitz, B., Baumert, J. & Roeder, P. M. (1985). Die Struktur der Fremdsprachenfähigkeit in Abhängigkeit von kognitiven Voraussetzungen und Lernbedingungen. *Zeitschrift für Empirische Pädagogik und Pädagogische Psychologie*, 9, 119-138.

- Sang, F. & Vollmer, H. J. (1980). Modelle linguistischer Kompetenz und ihre empirische Fundierung. In R. Grotjahn & E. Hopkins (Hrsg.), *Empirical Research on Language Teaching and Language Acquisition* (Quantitative linguistics. Bd. 6, S. 3-80). Bochum: Brockmeyer.
- Sasaki, M. (1993). Relationships among second language proficiency, foreign language aptitude, and intelligence: A structural equation modeling approach. *Language Learning*, 43(3), 313–344.
- Sasaki, M. (1996). *Second language proficiency, foreign language aptitude, and intelligence: Quantitative and qualitative analyses* (Theoretical studies in second language acquisition. Bd. 6). New York: Lang.
- Saussure, F. de (2001). *Grundfragen der allgemeinen Sprachwissenschaft*. Berlin: de Gruyter.
- Schermelleh-Engel, K. & Moosbrugger, H. (2002). *Beurteilung der Modellgüte von Strukturgleichungsmodellen*. Frankfurt am Main: Universität, Institut für Psychologie.
- Schneider, W. (2001). Schulleistungen im Bereich muttersprachlicher Bildung. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 143-152). Weinheim: Beltz.
- Schneider, W. & Helmke, A. (1986). Mehrebenenanalytische Ansätze zur Erklärung von Schulleistungen. In M. von Saldern (Hrsg.), *Mehrebenenanalyse. Beiträge zur Erfassung hierarchisch strukturierter Realität* (S. 170-193). Weinheim: Psychologie-Verlags-Union.
- Schöler, H. (2001). *Sprachleistungsmessung im Schulalter. Ein Überblick*. – Verfügbar unter: [http://www.ph-heidelberg.de/wp/schoeler/bericht\\_11.htm](http://www.ph-heidelberg.de/wp/schoeler/bericht_11.htm) [14.02.2008]
- Schumacker, R. E. & Lomax, R. G. (1996). *A beginner's guide to structural equation modeling*. Mahwah, NJ: Erlbaum.
- Schwarz, M. (1996). *Einführung in die kognitive Linguistik*. Tübingen: Francke.
- Searle, S. R., Casella, G. & McCulloch, C. E. (1992). *Variance components*. New York: Wiley.
- Sebastián-Gallés, N. & Kroll, J. (2003). Phonology in bilingual language processing: Acquisition, perception and production. In A. Meyer & N. Schiller (Hrsg.), *Phonetics and phonology in language comprehension and production: Differences and similarities*. Berlin: Mouton de Gruyter.
- Selinker, L. Douglas, D. (1987). LSP and interlanguages: Some empirical studies. *English for Specific Purposes*, 6(2), 75-85.
- Selinker, L. & Douglas, D. (1988). Using discourse domains in creating interlanguage: Context, theory and research methodology. *Studies in Descriptive Linguistics*, 18, 357-379.
- Shanahan, T. (1984). Nature of the reading-writing relation: An exploratory multivariate analysis. *Journal of Educational Psychology*, 76(8), 466-477.
- Shanahan, T. & Lomax, R. G. (1986). An analysis and comparison of theoretical models of the reading-writing relationship. *Journal of Educational Psychology*, 78(2), 116-123.
- Shohamy, E. (1996). Competence and performance in language testing. In G. Brown, K. Malmkjaer & J. Williams (Hrsg.), *Performance and competence in second language acquisition* (S. 138-151). Cambridge: Cambridge University Press.
- Sinatra, G. M. (1990). Convergence of listening and reading processing. *Reading Research Quarterly*, 25(2), 115-130.
- Skinner, B. F. (1957). *Verbal behavior*. Englewood Cliffs, NJ: Prentice-Hall.
- Skrondal, A. & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling – multilevel, longitudinal, and structural equation models*. Boca Raton: Chapman & Hall/CRC.
- Sparks, R. L., Artzer, M., Ganschow, L., Siebenhar, D., Plageman, M. & Patton, J. (1998). Differences in native-language skills, foreign-language aptitude, and foreign-language

- grades among high-, average-, and low-proficiency foreign-language learners: Two studies. *Language Testing*, 15(2), 181-216.
- Spolsky, B. (1989). Communicative competence, language proficiency, and beyond. *Applied Linguistics*, 10(2), 138-156.
- Ständige Konferenz der Kulturminister der Länder in der Bundesrepublik Deutschland (2003). *Entwicklung und Implementation von Bildungsstandards*. Bonn: Sekretariat der KMK.
- Ständige Konferenz der Kulturminister der Länder in der Bundesrepublik Deutschland (2004). *Bildungsstandards im Fach Deutsch für den Mittleren Schulabschluss*. Neuwied: Luchterhand.
- Stapleton, L. M. (2002). The incorporation of sample weights into multilevel structural equation models. *Structural Equation Modeling*, 9(4), 475-502.
- Statistisches Bundesamt (2004). *Bildung und Kultur. Allgemeinbildende Schulen. Schuljahr 2003/2004*. Wiesbaden: Statistisches Bundesamt.
- Steinert, J. (1978). Allgemeiner Deutscher Sprachtest. Göttingen: Hogrefe.
- Stenzel, K. (1989). Cloze-Tests und C-Tests in der Schule. *Praxis des neusprachlichen Unterrichts*, 38, 311-314.
- Stevens, J. (1996). *Applied multivariate statistics for the social sciences* (3. Auflage). Mahwah, NJ: Erlbaum.
- Sticht, T. G. (1968). Some relationships of mental aptitude, reading ability, and listening ability using normal and time-compressed speech. *The Journal of Communication*, 18(3), 243-258.
- Stricker, L. J., Rock, D. A. & Lee, Y.-W. (2005). *Factor structure of the LanguEdge™ test across language groups*. Princeton, NJ: Educational Testing Service.
- Studdert-Kennedy, M. (1987). The phoneme as a perceptumotor structure. In A. Allport, D. Mackay, W. Prinz & E. Scheerer (Hrsg.), *Language perception and production: Relationships between listening, speaking, reading, and writing* (S. 67-84). London: Academic Press.
- Swinton, S. E. & Powers, D. E. (1980). *Factor analysis of the test of English as a foreign language for several language groups*. Princeton, NJ: Educational Testing Service.
- Szagan, G. (2000). *Sprachentwicklung beim Kind*. Weinheim: Beltz.
- Takala, S. (2007). Relating examinations to the Common European Framework. In B. Beck & E. Klieme (Hrsg.), *Sprachliche Kompetenzen. Konzepte und Messung. DESI-Studie (Deutsch-Englisch-Schülerleistungen-International)*. Weinheim: Beltz.
- Tanaka, J. S. (1993). Multifaceted conceptions of fit in structural equation models. In K. A. Bollen & J. S. Long (Hrsg.), *Testing structural equation models* (S. 10-39). Newbury Park, CA: Sage.
- Taylor, D. S. (1988). The meaning and use of the term „competence“ in linguistics and applied linguistics. *Applied Linguistics*, 9(2), 148-168.
- Teepen, J. (2004). On the relationship between aptitude and intelligence in second language acquisition. *Working Papers in TESOL & Applied Linguistics*, 4(3), [Online <http://journals.tc-library.org/index.php/tesol/article/viewFile/69/75> Stand 29.08.2008]
- Tenorth, H.-E. (2001). Englisch: Ein Kerncurriculum, seine Notwendigkeit und seine Gestalt. In H.-E. Tenorth (Hrsg.), *Kerncurriculum Oberstufe. Mathematik – Deutsch – Englisch. Expertise im Auftrag der Ständigen Konferenz der Kultusminister* (S. 156-161). Weinheim: Beltz.
- Thomé, G. (1998). *Orthographieerwerb*. Frankfurt am Main: Lang.

- Thomé, G. (2003). Entwicklung der basalen Rechtschreibkenntnisse In U. Bredel, H. Günther, P. Klotz, J. Ossner & G. Siebert-Ott (Hrsg.), *Didaktik der deutschen Sprache – Ein Handbuch*. Paderborn: Schöningh.
- Thomé, G. & Gomolka, J. (2007). Rechtschreiben. In B. Beck & E. Klieme (Hrsg.), *Sprachliche Kompetenzen. Konzepte und Messung. DESI-Studie (Deutsch-Englisch-Schülerleistungen-International)* (S. 140-146). Weinheim: Beltz.
- Thomé, G. & Thomé, D. (2004). *Oldenburger Fehleranalyse OLFA. Instrument und Handbuch zur Ermittlung der orthographischen Kompetenz aus freien Texten ab Klasse 3 und zur Qualitätssicherung von Fördermaßnahmen*. Oldenburg: Igel-Verlag Wissenschaft.
- Thurstone, L. L. (1945). *Multiple factor analysis*. Chicago: University of Chicago Press.
- Timmermann, W. (2000). Transfer: ein altbekanntes Konzept im Kontext neuerer kognitiver Sprach(erwerbs)theorie. In C. Riemer (Hrsg.), *Kognitive Aspekte des Lehrens und Lernens von Fremdsprache* (S. 171-185). Tübingen: Narr.
- Turkeltaub, P. E., Gareau, L., Flowers, D. L., Zeffiro, T. A. & Eden, G. F. (2003). Development of neural mechanisms for reading. *Nature Neuroscience*, 6(6), 767-773.
- Ullman, M. T. (2001). A neurocognitive perspective on language: The declarative / procedural model. *Neuroscience*, 2, 717-727.
- Urban, K. K. (1977). *Verstehen gesprochener Sprache*. Düsseldorf: Schwann.
- Van der Linden, W. J., Veldkamp, B. P. & Carlson, J., E. (2004). Optimizing balanced incomplete block designs for educational assessments. *Applied Psychological Measurement*, 28(5), 317–331.
- Van Patten, B. (Hrsg.). (2004). *Processing instruction: theory, research, and commentary*. Mahwah, NJ: Erlbaum.
- Venezky, R. L. & Massaro, D. W. (1987). Orthographic structure and spelling-sound regularity in reading English words. In A. Allport, D. MacKay, W. Prinz & E. Scheerer (Hrsg.), *Language perception and production. Relationships between listening, speaking, reading, and writing* (S. 159-180). London: Academic Press.
- Verhoeven, L. (Hrsg.). (1994). *Functional literacy: Theoretical issues and educational implications* (Studies in written language and literacy. Bd. 1). Amsterdam: Benjamins.
- Vollmer, H. J., Henrici, G., Finkbeiner, C., Grotjahn, R., Schmid-Schönbein, G. & Zydariß, W. (2001). Lernen und Lehren von Fremdsprachen: Kognition, Affektion, Interaktion. Ein Forschungsüberblick. *Zeitschrift für Fremdsprachenforschung*, 12(2), 1-145.
- Wade-Wooley, L. & Siegel, L. S. (1997). The spelling performance of ESL and native speakers of English as a function of reading skill. *Reading and Writing*, 9, 387-406.
- Wagemaker, H. (Hrsg.). (1996). *Are girls better readers? Gender differences in reading literacy in 32 countries*. The Hague: International Association for the Evaluation of Educational Achievement.
- Wagner, W., Helmke, A., Schrader, F.-W., Eichler, W., Thomé, G. & Willenberg, H. (2008). Selbstkonzept und Motivation im Fach Deutsch. In DESI-Konsortium (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (S. 231-244). Weinheim: Beltz.
- Walker, D. A. (1976). *The IEA six-subject survey: An empirical study of education in twenty-one countries*. Stockholm: Almqvist & Wiksell.
- Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, 54(3), 427-450.
- Weinert, F. E. (1999). *Concepts of competence. Definition and selection of competencies*. München: Max-Planck-Institut für Psychologische Forschung.

- Wenderott, C. (2005). Zur Entwicklung des Bilingualen Lehrens und Lernens in Deutschland. *Essener Linguistische Skripte – elektronisch*, 5(1), 43-46.
- Wicherts, J. M., Dolan, C. V. & Hessen, D. J. (2005). Stereotype threat and group differences in test performance: A question of measurement invariance. *Journal of Personality & Social Psychology*, 89(5), 696-716.
- Widdowson, H. G. (1989). Knowledge of language and ability for use. *Applied Linguistics*, 10(2), 128-137.
- Willenberg, H. (2007a). Lesen. In B. Beck & E. Klieme (Hrsg.), *Sprachliche Kompetenzen. Konzepte und Messung. DESI-Studie (Deutsch-Englisch-Schülerleistungen-International)*. Weinheim: Beltz.
- Willenberg, H. (2007b). Wortschatz. In B. Beck & E. Klieme (Hrsg.), *Sprachliche Kompetenzen. Konzepte und Messung. DESI-Studie (Deutsch-Englisch-Schülerleistungen-International)* (S. 130-139). Weinheim: Beltz.
- Willenberg, H. (2008). Wortschatz Deutsch. In DESI-Konsortium (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (S. 72-80). Weinheim: Beltz.
- Willenberg, H., Gailberger, S. & Krelle, M. (2007). Argumentation. In B. Beck & E. Klieme (Hrsg.), *Sprachliche Kompetenzen. Konzepte und Messung. DESI-Studie (Deutsch-Englisch-Schülerleistungen-International)*. Weinheim: Beltz.
- Wilson, M. (2003). On choosing a model for measuring. *Methods of Psychological Research [Online]*, 8(3), 1-22.
- Wode, H., Burmeister, P., Daniel, A., Kickler, K.-U. & Knust, M. (1996). Die Erprobung von deutsch-englisch bilingualem Unterricht in Schleswig-Holstein: Ein erster Zwischenbericht. *Zeitschrift für Fremdsprachenforschung*, 7(1), 15-42.
- Wode, H., Devich-Henningsen, S., Fischer, U., Franzen, V. & Pasternak, R. (2001). *Englisch durch bilinguale Kitas und Immersionsunterricht in der Grundschule: Erfahrungen aus der Praxis und Forschungsergebnisse*. Vortrag auf dem DGFF-Kongress Fremdsprachenunterricht auf dem Prüfstand: Innovation – Qualität – Evaluation (5. Oktober 2001), Dresden.
- Wolff, D. (1997). Zur Förderung von Sprachbewußtheit und Sprachlernbewußtheit im bilingualen Sachfachunterricht. *Fremdsprachen Lehren und Lernen*, 26, 167-183.
- Wolff, D. (2001). Einige Anmerkungen zum Erwerb bereits gelernter Fremdsprachen auf den Erwerb weiterer Sprachen. In H. Schröder, P. Kumschlies & M. Gonzalez (Hrsg.), *Linguistik als Kulturwissenschaft. Festschrift für Bernd Spillner zum 60. Geburtstag* (S. 377-390). Frankfurt am Main: Lang.
- Wright, B. D. (1999). Rasch measurement models. In G. N. Masters & J. P. Keeves (Hrsg.), *Advances in measurement in educational research and assessment* (S. 85-97). Oxford: Pergamon Press.
- Wu, M. L., Adams, R. J. & Wilson, M. R. (1998). *Acer ConQuest. Generalised item response modelling software (Version II)*. Melbourne: Acer Press.
- Yuan, K.-H. & Hayashi, K. (2005). On Muthén's maximum likelihood for two-level covariance structure models. *Psychometrika*, 70(1), 147-167.
- Zydati, W. (2000). *Bilingualer Unterricht in der Grundschule: Entwurf eines Spracherwerbskonzepts für zweisprachige Immersionsprogramme*. Ismaning: Hueber.
- Zydati, W. (2002). *Leistungsentwicklung und Sprachstandserhebungen im Englischunterricht. Methoden und Ergebnisse der Evaluierung eines Schulversuchs zur Begabtenförderung: Gymnasiale Regel- und Expressklassen im Vergleich* (Kolloquium Fremdsprachenunterricht. Bd. 8). Frankfurt am Main: Lang.