

Evaluation von Fachlernen und Sprachlichkeit im Kontext bilingualer Bildung

Peter Passon
Universität Osnabrück

Diplomarbeit
Dezember 2007

Erstprüfer: Prof. Dr. T. Staufenbiel

Zweitprüfer: Prof. Dr. H. J. Vollmer

Peter Passon
Malvenweg 50
48163 Münster

ppasson@uos.de

Inhaltsverzeichnis

Zusammenfassung	IV
Danksagung	VI
Über den Autor.....	VII
1 Einleitung	1
2 Hintergrund (Theorie und Empirie).....	4
2.1 Determinanten von Schulleistung.....	4
2.2 Kognitive Fähigkeiten.....	8
2.3 Geschlecht und Alter	11
2.4 Sozialer Status	14
2.5 Klassenklima	21
2.6 Bilingualer Unterricht.....	24
2.7 Schulleistungsuntersuchungen	34
2.8 Schulleistungserfassung	36
2.9 Fragestellungen und Hypothesen	43
3 Methode	49
3.1 Untersuchungsplanung.....	49
3.2 Material	50
3.2.1 Geografietest	51
3.2.2 Schülerfragebogen.....	53
3.2.3 Kognitiver Fähigkeitstest	56
3.2.4 C-Test.....	60
3.3 Stichprobe	62
3.4 Untersuchungsdurchführung	71
3.5 Datenauswertung	73
4 Ergebnisse	76
4.1 Vorgehende Analysen	76
4.2 Mono- und bilinguale Schüler.....	81
4.3 Schulleistungen.....	83
4.4 Soziale Herkunft.....	85
4.5 Klassenklima	87
4.6 Geschlechtsunterschiede	88

4.7	Geografiekompetenz.....	89
5	Diskussion.....	92
5.1	Fachlernen und Sprachlichkeit.....	92
5.2	Schulleistungen.....	95
5.3	Soziale Herkunft und Klassenklima.....	96
5.4	Geschlechtsunterschiede.....	97
5.5	Kompetenz und Performanz.....	98
5.6	Stärken und Schwächen der Untersuchung.....	100
5.7	Ausblick.....	102
6	Literaturverzeichnis.....	105
7	Autorenverzeichnis.....	117
8	Abkürzungsverzeichnis.....	119
9	Abbildungsverzeichnis.....	120
10	Tabellenverzeichnis.....	121
11	Anhang.....	122
Anhang A	Schülerfragebogen für bilinguale Schüler.....	122
Anhang B	Geografietest für bilinguale Schüler (Auszug).....	137
Anhang C	C-Test Deutsch.....	139
Anhang D	Holistische Beurteilungsskala (Auszug).....	141
Anhang E	Tabellen zum KFT.....	143

Hinweis: Wegen der in meinen Augen besseren Lesbarkeit habe ich im Text männliche Formen der Worte verwendet. Dies soll keine Diskriminierung sein. Die Leserin oder der Leser möge sich bitte die weibliche Form dazu denken. Bei der expliziten Darstellung von Geschlechtsunterschieden verwende ich hingegen weibliche und männliche Formen.

Zusammenfassung

In der vorliegenden Arbeit sollten hauptsächlich die folgenden Fragen zu Fachlernen und Sprachlichkeit im Kontext bilingualer Bildung geklärt werden: Sind bilingual unterrichtete Schüler hinsichtlich der Fachkompetenz in Erdkunde mit ihren monolingual unterrichteten Mitschülern vergleichbar? Sind bilingual unterrichtete Schüler hinsichtlich der Sprachkompetenz in Englisch ihren monolingual unterrichteten Mitschülern überlegen? Wie verhält es sich mit der Sprachkompetenz in Deutsch beim Vergleich beider Schülergruppen?

Zur Klärung dieser Fragen wurden Daten aus Erhebungen an vier niedersächsischen Gymnasien herangezogen. Dabei wurden Schüler der zehnten Klasse untersucht, von denen ein Teil bilingualen Erdkundeunterricht in englischer Sprache und der andere Teil konventionellen Erdkundeunterricht in deutscher Sprache hatte. Bei den Datenerhebungen war das zentrale Instrument ein im Rahmen eines DFG-Projektes entwickelter Geografiekompetenztest, der acht Aufgaben mit insgesamt 17 Teilaufgaben umfasste, in Englisch und Deutsch vorlag und von den Schülern in 90 Minuten bearbeitet wurde. Weiterhin kam als Sprachkompetenztest für Englisch und Deutsch jeweils ein C-Test (Klein-Braley & Raatz, 1982) zum Einsatz. Kognitive Fähigkeiten wurden mittels des Kognitiven Fähigkeitstests (Heller & Perleth, 2000) erfasst. Zudem füllten die Schüler einen umfangreichen Schülerfragebogen aus, indem neben demografischen Daten auch Angaben zur Berufstätigkeit der Eltern, zum Klassenklima und zu weiteren Bereichen gemacht werden sollten. Aus den Daten zur Berufstätigkeit wurde die soziale Herkunft der Schüler ermittelt. Weitere Daten, wie Noten der Schüler in Deutsch, Englisch, Mathematik und Erdkunde, wurden von den Schulen zur Verfügung gestellt. Sowohl die Beurteilung der Schülerantworten im Geografiekompetenztest, als auch die eingesetzte Skala zum Klassenklima wurden anhand der erhobenen Daten positiv auf ihre Reliabilität hin überprüft.

Bei den Ergebnissen zeigte sich, dass die bilingualen Schüler entgegen ursprünglicher Annahmen keine größere Fachkompetenz in Erdkunde haben. Beide Schülergruppen sind vielmehr vergleichbar kompetent. Bezüglich der Sprachkompetenz zeigte sich, dass die bilingualen Schüler für Deutsch und Englisch besser abschnitten als die monolingualen Schüler. Der Kompetenzunterschied für Deutsch überraschte. Um die Vergleichbarkeit beider Schülergruppen sicherzustellen, wurden sie hinsichtlich der kognitiven Fähigkeiten und der sozialen Herkunft verglichen. Positiv fiel der Vergleich der kognitiven Fähigkeiten aus, für die keine Gruppenunterschiede gefunden wurden. Bei der sozialen Herkunft unterschieden

sich die Gruppen allerdings, was deren Vergleichbarkeit einschränkt. Hinsichtlich der Geografiekompetenz wird allerdings auf Grund von sehr geringen Zusammenhängen zwischen Geografiekompetenz und sozialer Herkunft keine Einschränkung vermutet.

Neben diesen Kernfragestellungen wurden weitere Hypothesen untersucht. Dabei zeigte sich, dass sowohl der Geografietest, als auch die beiden Sprachkompetenztests positiv mit den jeweiligen Schulleistungen – den Noten in Erdkunde, Deutsch und Englisch – zusammenhängen, was für die Validität der Verfahren spricht. Ebenso hingen die kognitiven Fähigkeiten mit den erfassten Schulleistungen positiv zusammen. Die kognitiven Fähigkeiten zeigten sich damit als guter Prädiktor der Schulleistungen.

Mit Blick auf die soziale Herkunft wurde angenommen, dass sich in den vier Kompetenzbereichen Geografiekompetenz, Sprachkompetenz Deutsch und Englisch, sowie den kognitiven Fähigkeiten positive Auswirkungen einer positiveren Beurteilung der sozialen Herkunft zeigen würden. Dem war nicht so. Weiterhin zeigten sich keine der angenommenen positiven Zusammenhänge zwischen dem Klassenklima und Schulleistungen sowie Geografiekompetenz.

Interessante Ergebnisse ergaben sich beim Vergleich von Schülerinnen und Schülern. Schülerinnen zeigten entgegen der Annahmen weder eine größere Sprachkompetenz in Deutsch, noch in Englisch. Bei letzterem zeigten die Schüler sogar bessere Leistungen. Keine Geschlechtsunterschiede fanden sich für die Geografiekompetenz. Bei den kognitiven Fähigkeiten schnitten wiederum die Schüler besser ab. Insgesamt leistet die vorliegende Arbeit damit einen positiven Beitrag zur empirischen Bildungsforschung für den bilingualen Unterricht.

Danksagung

Ich möchte mich an dieser Stelle bei meinen beiden Diplomarbeitsbetreuern bedanken. Sowohl Herr Prof. Dr. Staufenberg, als auch Herr Prof. Dr. Vollmer, haben sich viel Zeit für meine Anliegen und die Betreuung meiner Diplomarbeit genommen. Weiterhin danke ich meiner Mutter – Marianne Passon – dafür, dass sie mich bei der Korrektur der Arbeit auf orthographische Unzulänglichkeiten unterstützt hat.

Über den Autor

Name	Peter Passon
Anschrift	Malvenweg 50, 48163 Münster
Geburtsdatum, -ort	22.01.1982, Mönchengladbach
Staatsangehörigkeit	deutsch
Familienstand	ledig
Konfession	römisch-katholisch

Schul- und Universitätsausbildung

07/1988 – 07/1992	Katholische Grundschule Kleinenbroicherstraße, Mönchengladbach
08/1992 – 06/2001	Franz-Meyers-Gymnasium, Mönchengladbach Besuch des bilingualen Zweiges Deutsch/Englisch (Abschluss: Abitur mit Durchschnittsnote 1,9)
10/2002 – 09/2003	Psychologiestudium an der Heinrich-Heine-Universität Düsseldorf
10/2003 – 01/2008	Psychologiestudium an der Universität Osnabrück (Vordiplom: Durchschnittsnote 1,4)

Praktische Tätigkeiten

07/2000 – 08/2000	Praktikum bei Mitsubishi Electric Europe, Ratingen
07/2001 – 04/2002	Zivildienst beim Deutschen Roten Kreuz Kreisverband Rheydt e.V.
09/2004 – 01/2007	Forschungspraktikum an der Universität Osnabrück
06/2005 – 12/2007	Mitarbeit in der studentischen Beratung Kosmos e.V., Osnabrück
08/2005 – 09/2005	Praktikum an den Rheinischen Kliniken Mönchengladbach
09/2006 – 04/2007	Wissenschaftliche Hilfskraft an der Universität Osnabrück
04/2007 – 09/2007	Praktikum in der Beratung Meinsen & Steinhübel, Osnabrück

Publikationen

Wittenberg, C., Tietze, M. I., Passon, P., Kötter, S. & Hamborg, K.-C. (2006). Usability von Voice User Interfaces im Anwendungsbereich Car-Sharing. In T. Bosenick, M. Hassenzahl, M. Müller-Prove & M. Peissner (Hrsg.), *Usability Professionals 2006* (S. 52-55). Stuttgart: German Chapter der Usability Professionals Association e.V.

1 Einleitung

Seit dem Jahr 1989 hat in der Bundesrepublik Deutschland die Zahl der Schulen, die bilinguale Bildungsgänge anbieten, stark zugenommen (Thürmann, 2000). Oftmals handelt es sich dabei um bilinguale Züge Deutsch/Englisch, in deren Verlauf die Schüler Sachfachunterricht in englischer Sprache bekommen¹. Das Ziel des bilingualen Unterrichts ist letztlich, dass die Schüler sowohl in ihrer Muttersprache, als auch in einer weiteren Sprache profunde Sprachkenntnisse erwerben und zumindest in einem anderen kulturellen Kontext sicher agieren können (Bach, 2005; Christ, 1999; Vollmer, 2004). Trotz dass sich der bilinguale Unterricht – sicherlich auch vor dem Hintergrund eines zusammenwachsenden Europas und der viel diskutierten „Globalisierung“ – immer größerer Beliebtheit erfreut und an vielen Schulen mittlerweile ein bilingualer Zweig existiert, gibt es wenig Forschungsergebnisse zum bilingualen Unterricht und zum Vergleich zwischen bilinguaem und monolinguaem Unterricht.

Vollmer (2006b) schreibt, dass der bilinguale Unterricht „immer noch erstaunlich wenig empirisch untersucht worden ist, insbesondere was die fachliche Vergleichbarkeit der Lernergebnisse zwischen bilingualen und monolingualen“ (S. 205) Schülern anbelangt. Zudem gibt es wenig Erkenntnisse darüber, was es für Schüler bedeutet, fachliche Aufgabenstellungen in einer Zweitsprache bearbeiten zu müssen.

Danach ist der Forschung zum bilingualen Unterricht besonders die folgende Frage zu stellen: Sind bilingual unterrichtete Schüler hinsichtlich der Fachkompetenz in Sachfächern mit ihren monolingual unterrichteten Mitschülern vergleichbar, oder stehen sie ihnen nach? Neben dieser Frage nach dem *Fachlernen* sollte die *Sprachlichkeit* nicht vernachlässigt werden: Sind bilingual unterrichtete Schüler hinsichtlich der Sprachkompetenz in Englisch ihren monolingual unterrichteten Mitschülern überlegen? Wie verhält es sich mit der Sprachkompetenz in Deutsch beim Vergleich beider Schülergruppen?

Fachlernen und Sprachlichkeit – wie bereits im Titel dieser Diplomarbeit genannt – werden daher im Folgenden das Zentrum des Interesses bilden. Dabei soll untersucht werden, ob sich monolinguale und bilinguale Schüler bezüglich der Fachkompetenz in einem ausgewählten Sachfach unterscheiden. Diese Frage ist besonders relevant, da ohne ihre Beantwortung Kritiker immer wieder behaupten können, eine größere Sprachkompetenz,

¹ Bereits an dieser Stelle sei darauf hingewiesen, dass der Autor der vorliegenden Arbeit während der Gymnasialzeit in einem bilingualen Zweig Deutsch/Englisch unterrichtet worden ist.

beispielsweise in Englisch, würde in bilingualen Bildungsgängen durch eine geringere Fachkompetenz in den Sachfächern erkaufte. Als Sachfach wird in der vorliegenden Arbeit Erdkunde untersucht, dass an vielen Schulen als eines der ersten bilingualen Sachfächer gelehrt wird. „Erdkunde ist aus didaktischen Gründen oft erstes Sachfach, da Anschaulichkeit und deskriptive Sprachhandlungen fachimmanent sind“ (Sekretariat der ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [SKMK], 2006, S. 17). Darüber hinaus soll die Sprachkompetenz der Schüler in Deutsch und in Englisch verglichen werden. Zusätzlich werden verschiedene andere Einflussgrößen auf Fachlernen und Sprachlichkeit untersucht, wie beispielsweise die kognitiven Fähigkeiten oder die soziale Herkunft.

Die vorliegende Arbeit ist eingebettet in das von der Deutschen Forschungsgemeinschaft (DFG) geförderte Projekt *Fachlernen und Fremdsprachlichkeit im bilingualen Unterricht: Aufgabenbasierte Kognition, Kommunikation, Kooperation*, welches unter der Leitung von Helmut Johannes Vollmer an der Universität Osnabrück durchgeführt wurde (Vollmer, 2006b). Dieses Projekt – für das im Folgenden die Kurzbezeichnung „DFG-Projekt“ verwendet wird – beschäftigt sich mit der Untersuchung von bilinguaem Sachfachunterricht. Im Rahmen des DFG-Projektes wurden Daten an Schülern der zehnten Klasse an Gymnasien in Niedersachsen erhoben und dabei neben vielen Angaben zum persönlichen Hintergrund der Schüler auch die Fach- und Sprachkompetenz, sowie die kognitiven Fähigkeiten erfasst. Neben dem Einsatz von Tests und Fragebögen wurde zusätzlich in Teilstudien mit lautem Denken, Überarbeiten von schriftlichen Aufgabenlösungen und Videografie gearbeitet. Je nach Untersuchungsgruppe mussten die teilnehmenden Schüler Aufgaben einzeln oder kooperativ im Paar lösen. Näheres zu diesen Teilstudien berichtet Vollmer (2006a).

Sowohl die vorliegende Arbeit, als auch das DFG-Projekt, orientieren sich bei der Datenerhebung am Vorgehen bei PISA, dem Programme for International Student Assessment der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD). Einen Überblick über dieses Programm gibt beispielsweise Helmke (2005). Neben PISA wurden bereits viele nationale und internationale Untersuchungen im Schulbereich durchgeführt (Baumert, 2001; Bos & Postlethwaite, 2002), jedoch wurde PISA wie kaum eine andere so intensiv in der deutschen Öffentlichkeit diskutiert. Dabei kann im Rahmen des DFG-Projektes keine vergleichbar umfassende Analyse mit entsprechend generalisierbaren Ergebnissen wie bei PISA geleistet werden. Einerseits schon auf Grund der im Vergleich zu PISA geringen Stichprobe, andererseits wegen der deutlich geringeren Ressourcen.

In jedem Fall zeigen Untersuchungen wie PISA und die anschließende Diskussion der Ergebnisse die Notwendigkeit von Evaluation im Bildungsbereich, um Verbesserungen an der Ausbildung der nachfolgenden Generationen vornehmen zu können. Dies ist auch eine Zielsetzung des DFG-Projektes. Die Unterschiede in den fachlichen und sprachlichen Kompetenzen von bilingualen- und monolingualen Schülern sind einer genaueren Betrachtung würdig und können eine Grundlage bilden, den bilingualen und auch den monolingualen Unterricht zu verbessern.

Nach dieser Einleitung wird im zweiten Kapitel der vorliegenden Arbeit der nötige theoretische und empirische Hintergrund zur Untersuchung der skizzierten Fragestellungen dargelegt werden. Im sich anschließenden dritten Kapitel wird die Untersuchungsmethodik behandelt. Im Anschluss daran werden im vierten Kapitel die Ergebnisse vorgestellt, die im fünften Kapitel diskutiert werden.

2 Hintergrund (Theorie und Empirie)

In diesem Kapitel soll der theoretische und empirische Hintergrund beleuchtet werden. Dazu soll in den folgenden Abschnitten zuerst ein Überblick über Vorstellungen zu verschiedenen Determinanten von Schulleistung gegeben werden. Daran anschließend werden verschiedene ausgewählte Determinanten wie beispielsweise kognitive Fähigkeiten und sozialer Status genauer dargestellt. Gegen Ende des Kapitels wird auf bilingualen Unterricht sowie die Erfassung von Schulleistungen eingegangen. Das Kapitel schließt mit der Darstellung der Hypothesen der vorliegenden Arbeit.

2.1 Determinanten von Schulleistung

Bei Untersuchungen im schulischen Kontext ist es wichtig, vorab klarzustellen, was unter Schulleistung zu verstehen ist und welche Faktoren sie beeinflussen. Unter dem Begriff *Schulleistung* lässt sich sowohl die Leistung der Schule als Institution, als auch die Leistung der Schüler dieser Institution verstehen. In der vorliegenden Arbeit werden unter Schulleistungen die Leistungen von Schülern verstanden. Andere Autoren, wie beispielsweise Helmke und Weinert (1997), sprechen in diesem Zusammenhang auch von *Schülerleistungen*.

Es kann davon ausgegangen werden, dass die Leistung der Schüler zu einem großen Teil davon abhängt, in welchem Ausmaß die Schule ihre Funktion als Bildungseinrichtung erfüllt, in welchem Ausmaß es ihr gelingt, ihren Schülern nützliche Fertigkeiten und notwendiges Wissen zu vermitteln. Dies zeigt sich nach Weinert (2001a) besonders in Ländern, die nicht über ein ausreichend entwickeltes Bildungssystem verfügen: Für diese Länder finden sich deutlichere Auswirkungen der Dauer des Schulbesuchs auf die Berufschancen, als in Ländern mit entwickeltem Bildungssystem. Auch wenn es trivial scheint, dass mit der Dauer des Schulbesuchs die Berufschancen steigen, so wird an diesem Beispiel klar, dass die Wirkung von Schule und damit ebenso die Leistung der Schüler immer vom gesellschaftlichen Kontext beeinflusst werden.

Ein Ansatz, der versucht Schulleistungen zu erklären, ist die *theory of school learning* von Bloom (1976). Zu Beginn seiner Monografie von 1976 gibt Bloom einen Überblick, in dem er drei verschiedene ihm bekannte Sichtweisen bezüglich Schulleistungen kurz und knapp anspricht:

1. There are good learners and there are poor learners
 2. There are faster learners and there are slower learners
 3. Most students become very similar with regard to learning ability, rate of learning, and motivation for further learning—when provided with favorable learning conditions
- (Bloom, 1976, S. ix-x.)

Folgt man dem ersten Punkt, so entscheiden relativ unveränderbare Merkmale in der Person des Lerners darüber, ob er gute oder schlechte Schulleistungen zeigt. Der zweite Punkt bezieht sich auf ein Modell, das im Kern davon ausgeht, dass die Schulleistungen direkt vom Verhältnis der aufgewendeten zur benötigten Lernzeit abhängen. Auch hier wird implizit eine relativ unveränderliche Komponente in der Person des Lerners angenommen, die ausmacht, dass verschiedene Lerner unterschiedlich viel Zeit benötigen, um die gleiche Stoffmenge zu lernen. Der dritte Punkt schließlich bezieht sich auf Blooms Modell von 1976, das auf die Lernbedingungen abhebt.

Bloom (1976) sieht drei Bereiche als relevant an. Der erste Bereich sind Schülermerkmale. Unter dieser Überschrift unterscheidet er zwischen kognitivem Verhalten und affektiven Eigenschaften. Darunter versteht er beispielsweise kognitive Fähigkeiten des Lerners, dessen Lerngeschichte und dessen Motivation, etwas Neues zu lernen. Der zweite Bereich betrifft den Unterricht. Zum einen zählt hierzu die Qualität des Unterrichts und zum anderen die Lernaufgabe selbst, die das zentrale Element seines Modells darstellt. Der dritte Bereich ist die Ergebnisseite, mit dem erreichten Leistungsniveau, der Lerngeschwindigkeit und den affektiven Ergebnissen beim Lerner. Anders ausgedrückt stellen in Blooms Modell von 1976 damit kognitive und affektive Merkmale des Lerners die Faktoren dar, auf die Unterrichtsqualität und Lernaufgabe abgestimmt werden müssen, damit im Endeffekt gleiche Schulleistungen resultieren. Zur Veranschaulichung ist das Modell in Abbildung 1 in Originalsprache wiedergegeben.

Nach Weinert (2001a) ist dieses Modell im Zusammenhang der *pädagogischen Egalisierungstheorie* anzusiedeln, nach welcher die Zielsetzung von Schule unter anderem ist, intellektuelle Unterschiede zwischen den Schüler abzubauen und so Chancengleichheit zu schaffen. Theoretisch steht dahinter das *aptitude-treatment-interaction-model*, nach dem Schüler mit unterschiedlichen Lernvoraussetzungen von unterschiedlichen Lehrmethoden unterschiedlich profitieren. Danach hängt die Schulleistung besonders von der Passung zwischen den individuellen Lernvoraussetzungen und der Lehrmethode ab. Laut Weinert (2001a) konnten diese theoretischen Erwartungen allerdings nicht bestätigt werden.

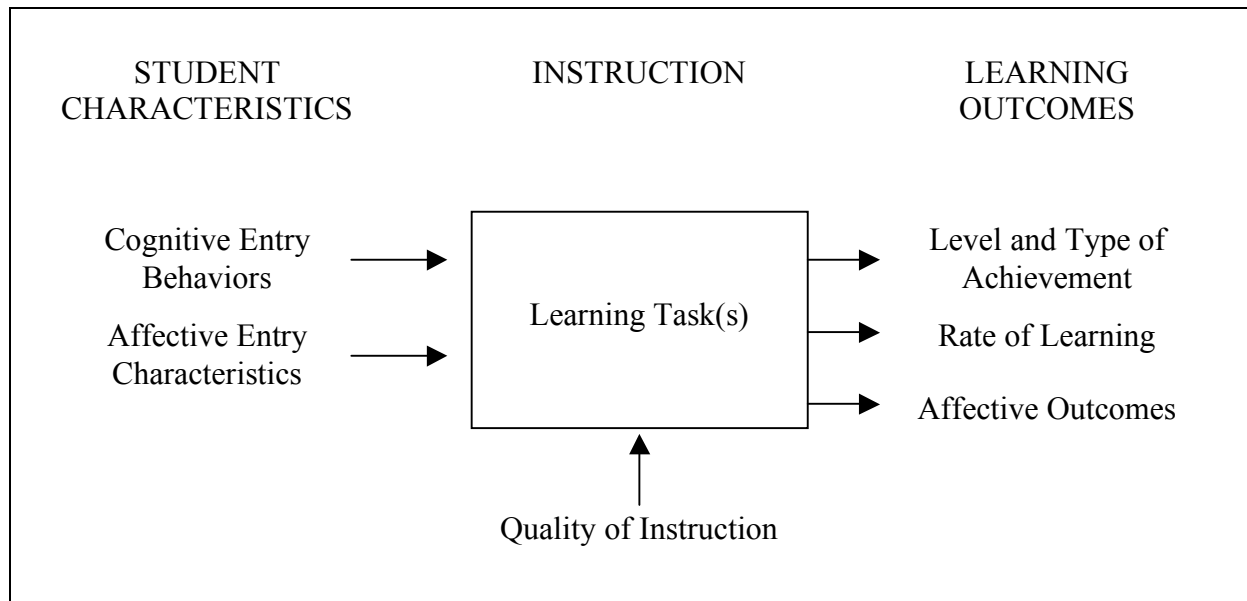


Abbildung 1 Wichtige Variablen im Modell schulischen Lernens von Bloom (1976, S. 11).

In der Literatur finden sich neben dem dargestellten Modell von Bloom (1976) noch weitere Ansätze, die beispielsweise bei Helmke und Weinert (1997) genauer erläutert werden. Folgt man Helmke und Schrader (2001) oder Helmke (2005), so gibt es keine Modelle, die versuchen, möglichst viele verschiedene Einflussfaktoren in ihrer Wirkung auf Schulleistung darzustellen. Die benannten Autoren betonen zurecht, dass es bei der Betrachtung von Einflussfaktoren der Schulleistung nicht um einzelne, isoliert wirkende Kausalfaktoren gehen kann. Statt dessen ist davon auszugehen, dass vielfältige Einflussfaktoren an der Entwicklung und Ausprägung der Schulleistungen mitwirken. Zudem müssen reziproke Effekte zwischen den einzelnen Einflussfaktoren berücksichtigt werden. Helmke und Schrader (2001) sprechen in diesem Zusammenhang von *multipler Determiniertheit* von Schulleistungen.

In ihrer umfassenden Arbeit zu Bedingungen von Schulleistung nennen Helmke und Weinert (1997) größtenteils Einflussfaktoren ohne Bindung an ein Modell oder eine Theorie. Sie unterscheiden dabei zwischen äußeren Faktoren, wie den gesellschaftlichen und wirtschaftlichen Rahmenbedingungen, Einflüssen der Schulorganisation, Einflüssen der Schulklasse und zwischen inneren Faktoren, wie motivationalen, kognitiven und emotionalen Lernvoraussetzungen. Weiterhin werden Einflüsse der Eltern betrachtet, die durch genetische Faktoren, sozialen Status oder das Erziehungsverhalten wirken können. Schließlich wird noch der Einfluss durch den Lehrer und den Unterricht behandelt. Dabei wird davon ausgegangen, dass verschiedene Faktoren additiv zusammen wirken und so ihren Einfluss wechselseitig

kompensieren können. Denkbar ist allerdings auch eine multiplikative Verknüpfung vieler verschiedener Faktoren und damit nur eine eingeschränkte wechselseitige Kompensierbarkeit. Zudem kann zwischen Einflussfaktoren *Koppelung* vorliegen, wonach Mindestausprägungen bei bestimmten Einflussfaktoren nötig sind, um bestimmte Wirkungen zu erzielen (Helmke & Schrader, 2001).

Um die in der vorliegenden Arbeit relevanten Einflussfaktoren von Schulleistung systematisch behandeln zu können, soll auf die Arbeit von Kühn (1983) zurückgegriffen werden. Da auch zum damaligen Zeitpunkt keine umfassende Theorie zu Einflussfaktoren existierte, stellte Kühn (1983) aus der Literatur die als wichtig geltenden Einflussfaktoren zusammen und wählte nach seinem subjektiven Ermessen einige für die genauere Betrachtung aus. In Abbildung 2 sind einige Einflussfaktoren in Anlehnung an Kühn (1983) aufgeführt.

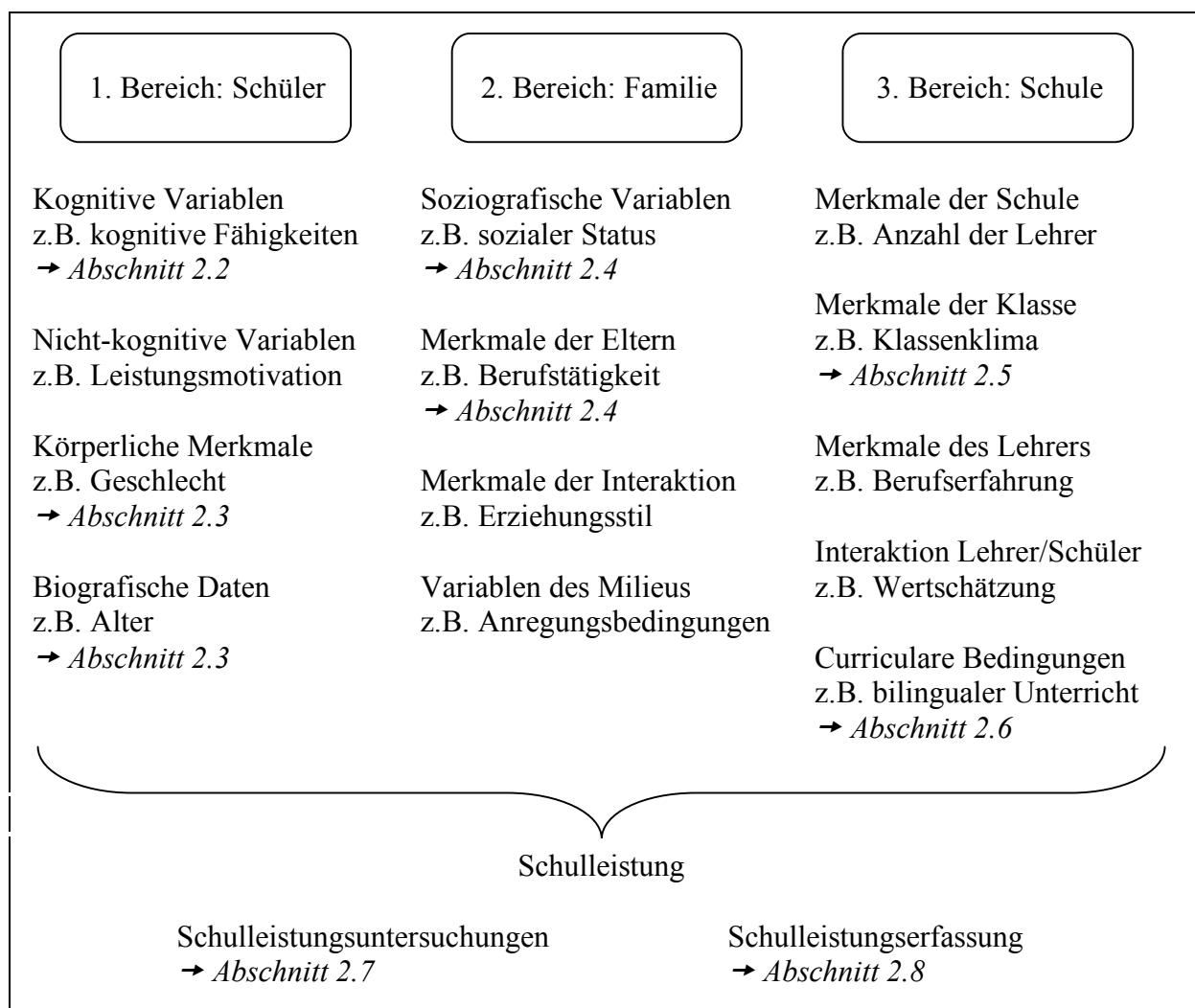


Abbildung 2

Übersicht zu den Einflussfaktoren von Schulleistung (in Anlehnung an Kühn, 1983, S. 16f).

Die einzelnen Einflussfaktoren sind dabei einem von drei Bereichen zugeordnet, je nach dem, ob sie die Person der Schüler selbst, die Herkunftsfamilie, oder die Schule betreffen. Innerhalb dieser drei Bereiche sind die Einflussfaktoren zu inhaltlichen Gruppen zusammengefasst, wobei zu jeder Gruppe meist nur ein Beispiel angegeben ist. Weitere Beispiele gibt Kühn (1983). Auch in der vorliegenden Arbeit wurden nur ausgewählte Einflussfaktoren behandelt, die dem Autor im aktuellen Kontext relevant schienen. Sie sind in Abbildung 2 durch den Verweis auf einen Abschnitt gekennzeichnet, in dem sie genauer behandelt werden. Im folgenden Abschnitt soll – wie in Abbildung 2 ersichtlich – auf die kognitiven Fähigkeiten eingegangen werden.

2.2 Kognitive Fähigkeiten

Ein erster wichtiger Einflussfaktor von Schulleistungen sind die kognitiven Fähigkeiten der Schüler. Als die zentrale kognitive Fähigkeit wird die Intelligenz angesehen. In der Literatur finden sich viele Belege dafür, dass Intelligenzmaße sehr gute Einzelprädiktoren zur Vorhersage der Schulleistung sind (Helmke & Weinert, 1997; Tücke, 2005). Bereits Sander (1931) berichtet Zusammenhänge zwischen Intelligenz und Schulleistung, ausgedrückt als Rangkorrelation in Höhe von $r_s = .75$. Zwar spricht er von Begabung; dass zur Erfassung verwendete Testverfahren scheint allerdings ein früher Intelligenztest zu sein. Auch wenn die Befunde wegen der deutlich anderen Rahmenbedingungen zum Untersuchungszeitpunkt und des verwendeten Testverfahrens für die heutige Zeit nur sehr beschränkt aussagefähig sind, so weisen sie doch das schon lange bestehende Interesse an Zusammenhängen zwischen Intelligenz und Schulleistung hin.

Intelligenzmaße versuchen die kognitiven Fähigkeiten von Personen abzubilden. Dazu werden spezielle psychologische Testverfahren verwendet, beispielsweise der Kognitive Fähigkeitstest (Heller & Perleth, 2000; vgl. Abschnitt 3.2.3), der Intelligenz-Struktur-Test (Amthauer, Brocke, Liepmann & Beauducel, 2001; Brocke, Beauducel & Liepmann, 2004) oder der Mehrfachwahl-Wortschatz-Intelligenztest (Lehrl, 1977). Die Diskussion um die Intelligenz und das Für und Wider ihrer Messung wird bereits über einen langen Zeitraum geführt. Daher soll hier nicht mehr dazu geschrieben und der Leser statt dessen auf andere Quellen verwiesen werden (z.B. Zimbardo & Gerrig, 1999; Tücke, 2005; Klauer, 2001a). Um den Einfluss der kognitiven Fähigkeiten auf Schulleistungen in seiner Wirkweise zu erklären, sind zwei Punkte von besonderer Bedeutung (Helmke & Weinert, 1997; Helmke & Schrader,

2001). Zum einen kann davon ausgegangen werden, dass sich Personen mit höherer Intelligenz besser auf neue Aufgaben einstellen und besser lösungsrelevante Aspekte von Probleme erfassen können. Zum anderen haben Personen mit höherer Intelligenz in der Vergangenheit vermutlich effektiver Wissen angesammelt und damit eine bessere Grundlage für zukünftige Lernprozesse erworben.

In einer älteren Arbeit untersuchten Beer, Kutalek und Schnell (1968) den Zusammenhang zwischen Intelligenz und Schulleistung. Dabei berechneten sie einen mittleren Korrelationskoeffizienten von $r = .44$ und ein 99%-Konfidenzintervall für die Population von .45 bis .46. Die Ergebnisse entsprechen einer Varianzaufklärung von etwa 20%. Dies ist ein Hinweis darauf, dass die Schulleistung von weiteren Faktoren beeinflusst wird und eine Vorhersage der Schulleistung allein über die Intelligenz – trotz der genannten Stärke von Intelligenzmaßen als Einzelprädiktoren – unzureichend bleiben muss.

Kolem (1981) hat in einer Untersuchung an Schülern der vierten Klasse aus Bremen und Ritterhude ($N = 626$) die Zusammenhänge von Sprache, Schulleistung und Intelligenz untersucht. Dabei geht sie von der Hypothese aus, dass Noten im Wesentlichen von der Sprachfertigkeit der Schüler abhängen und weniger von den Schulleistungen und der Intelligenz. Die Sprachfertigkeit wurde durch einen Wortverständnis- und einen Leseverständnistest operationalisiert, während die Intelligenz im Wesentlichen durch einen kultur- und sprachfreien Intelligenztest erfasst wurde. Die Schulleistung wurde an den Noten der Schüler festgemacht. Sie berichtet von Zusammenhängen zwischen Intelligenz, Wortverständnis und Leseverständnis: Mit der Deutschnote zeigten sie Korrelationen von $r = .54$ mit Wortverständnis und $r = .59$ mit Leseverständnis, sowie von $r = .44$ mit der Intelligenz. Mit der Mathematiknote berichtet sie Korrelationen von $r = .52$ mit Wortverständnis und $r = .59$ mit Leseverständnis, sowie von $r = .54$ mit der Intelligenz. Für den Bereich des Sachunterrichts berichtet sie Korrelationen von $r = .59$ mit Leseverständnis und Wortverständnis sowie von $r = .54$ mit Intelligenz.

Alle berichteten Korrelationen fielen signifikant aus. Der eingesetzte sprachfreie Intelligenztest erwies sich tatsächlich als wenig von der Sprachfertigkeit abhängig; zumindest zeigte der Vergleich der Intelligenzwerte des sprachfreien Tests geringe und nicht signifikante Korrelationen mit einem anderen, nicht sprachfreien Intelligenztest. Schulleistungen zeigten in dieser Untersuchung substantielle Zusammenhänge mit der Intelligenz, wobei die Zusammenhänge zwischen Schulleistungen und Sprachfertigkeit tendenziell höher ausfielen.

Auch Kühn (1983) hat unter anderem den Zusammenhang von Intelligenz und Schulleistung erforscht. Dazu untersuchte er 510 Schüler neben anderen Instrumenten mit dem Intelligenztest *Aufgaben zum Nachdenken* (AzN 4+). Dieser Test besteht aus fünf Subtests, wie sie sich von der Bezeichnung her bei vielen Intelligenztests wieder finden: „Rechnen“, „Analogien“, „Zahlenreihen“, „Satzergänzung“ und „Instruktionsverständnis“. Die Schulleistung wurde wiederum an Noten festgemacht. Die Ergebnisse sind in Tabelle 1 wiedergegeben. Dabei zeigten sich besonders hohe Korrelationen zwischen den Noten in Deutsch, Mathematik und Sachkunde und den Intelligenztestwerten. Für die eher künstlerisch und praktisch orientierten Nebenfächer fallen die Korrelationen geringer aus. Nochmals geringer sind die Zusammenhänge zwischen den Intelligenztestwerten und den Noten im Unterrichtsfach Leibeserziehung. In diesem Unterrichtsfach haben kognitive Fähigkeiten vermutlich einen geringeren Stellenwert im Vergleich zu körperlichen Fähigkeiten.

Tabelle 1

Die Tabelle zeigt Produkt-Moment-Korrelationskoeffizienten zwischen den Ergebnissen im Intelligenztest *Aufgaben zum Nachdenken* (AzN 4+) und Noten in Unterrichtsfächern. Die Noten wurden dabei transformiert, damit positive Korrelationskoeffizienten resultierten. (erstellt in Anlehnung an Kühn, 1983, Tabelle 6, S. 149)

		Unterrichtsfächer						
		Deutsch	Mathematik	Sachkunde/ Heimatkunde	Kunsterziehung	Musik	Leibeserziehung	Werken/ Nadellarbeit
Subtests AzN 4+	Rechnen	.46	.60	.50	.25	.30	.08*	.19
	Analogien	.45	.48	.42	.27	.32	.13	.19
	Zahlenreihen	.47	.57	.49	.28	.35	.12	.26
	Satzergänzung	.53	.52	.51	.28	.32	.16	.22
	Instruktionsverständnis	.55	.55	.50	.31	.37	.12	.27
Gesamtpunktwert AzN 4+		.61	.68	.61	.35	.42	.15	.29

* *Bemerkung:* Der so markierte Korrelationskoeffizient ist nicht signifikant. Alle anderen Korrelationskoeffizienten sind bei $p < .01$ signifikant.

Die drei exemplarisch dargestellten Untersuchungen belegen, dass sich – wie bereits eingangs erwähnt – deutliche Zusammenhänge zwischen Intelligenz und Schulleistung finden lassen. Daher soll in der vorliegenden Arbeit die Intelligenz als wichtiger Einflussfaktor von Schulleistung mit berücksichtigt werden.

2.3 Geschlecht und Alter

Weitere wichtige Einflussfaktoren aus dem Bereich der Schüler sind Geschlecht und Alter. Auf diese beiden soll in der genannten Reihenfolge in diesem Abschnitt eingegangen werden. Biologisches Geschlecht und Lebensalter werden in vielen Untersuchungen standardmäßig mit erhoben und auf Effekte hin überprüft.

Kasten (2001) berichtet über Erkenntnisse in Bezug auf die Variable Geschlecht, die für die pädagogische Psychologie bedeutsam sind. Geschlecht ist in erster Linie ein biologisches Merkmal. Im Verlauf der Ontogenese spielt dann allerdings nicht nur die Interaktion von Genom und Umwelt eine Rolle, sondern auch die Sozialisation – besonders für die Entwicklung der Geschlechtsidentität. Für kognitive Leistungen finden sich widersprüchliche Befunde bezüglich möglicher Geschlechtsunterschiede. Vielfach wurde allerdings gezeigt, dass weibliche Personen im Bereich der verbalen Fähigkeiten und männliche Personen im Bereich der mathematischen und naturwissenschaftlichen Fähigkeiten und des räumlichen Vorstellungsvermögens dem jeweils anderen Geschlecht überlegen sind.

Auch die erste PISA-Untersuchung befasst sich mit Geschlechtsunterschieden. Davon berichten Stanat und Kunter (2001). Sie erwähnen ebenfalls die widersprüchliche empirische Befundlage zum Thema. Für den internationalen Vergleich berichten sie große Geschlechtsunterschiede beim Lesen: Schülerinnen sind in allen Teilnehmerstaaten signifikant besser als Schüler, wobei Deutschland im Bereich des OECD-Mittelwertes liegt. Für Mathematik finden sich Vorteile der Schüler, die jedoch deutlich kleiner als die Vorteile der Schülerinnen beim Lesen ausfallen und nicht für allen Teilnehmerstaaten signifikant sind. Für die Naturwissenschaften finden sich keine signifikanten Geschlechtsunterschiede.

Bei PISA wird zwischen zwei verschiedenen Textarten unterschieden: Bei *kontinuierlichen Texten* handelt es sich um Prosatexte, die aus Sätzen und Absätzen bestehen. Darunter fallen Erzählungen, Kommentare und Argumentationen. Bei *nicht-kontinuierlichen Texten* sind die Informationen nicht verbal in einem fortlaufenden Text dargestellt. Zu dieser Textart zählen Formulare, Tabellen, Grafiken, Diagramme und Karten. Bei genaueren Analysen der Ergebnisse für Deutschland zeigt sich, dass der Vorteil für Schülerinnen beim Lesen besonders bei kontinuierlichen Texten gegeben ist. Für nicht-kontinuierliche Texte ist er geringer. Im Umgang mit Texten zeigen Schülerinnen eine signifikant höhere Lesegeschwindigkeit als Schüler. Für die Naturwissenschaften zeigen Schülerinnen bessere Ergebnisse im

Bereich Biologie, die nicht signifikant sind. Schüler haben signifikante Stärken im Bereich Chemie und Physik. Stanat und Kunter (2001) schreiben:

Im Vergleich zu Mädchen bereitet es Jungen deutlich größere Schwierigkeiten, Texte und ihre Merkmale kritisch zu reflektieren und zu bewerten. Analysen in den Bereichen Mathematik und Naturwissenschaften weisen darauf hin, dass Mädchen insbesondere bei Aufgaben, die den Umgang mit mentalen oder mathematischen Modellen erfordern, relative Schwächen aufweisen. (S. 257)

Wichtig ist auch eine differenzierte Betrachtung nach Schulformen. Dort sind Schülerinnen und Schüler zu unterschiedlichen Anteilen vertreten: Schülerinnen sind an Gymnasien und Schüler an Haupt- und Sonderschulen überrepräsentiert. So ist es möglich, dass die geringeren Leistungen von Schülerinnen in Mathematik durch deren Überzahl an Gymnasien – wo die Leistungen meist besser sind als an den anderen Schulen – für die Gesamtgruppe der Schülerinnen etwas abgemildert werden. Bei fast allen Schulformen zeigt sich ein Vorteil für Schüler in den Bereichen Mathematik und Naturwissenschaften, sowohl für Chemie und Physik, als auch für Biologie. Die Vorteile für Schülerinnen in Biologie haben sich bei der Analyse nach Schulformen in einen Vorteil für die Schüler gewandelt. Generell verstärken sich bei diesen Analysen damit die Stärken der Schüler. Für den Bereich Lesen zeigen Schülerinnen Vorteile. Für den Bereich nicht-kontinuierlicher Texte gilt dies nicht. Dort sind die Schüler stärker, allerdings nicht signifikant.

Bei PISA wurde ebenfalls erhoben, ob die Schülerinnen und Schüler Interesse und Freude am Lesen haben. In von Stanat und Kunter (2001) durchgeführten Mediatoranalysen zeigte sich, dass Interesse einen sehr großen Einfluss auf die Leseleistung hat. Es fungiert als Mediator zwischen Geschlecht und Leseleistung und reduziert den Einfluss des Geschlechts auf die Leseleistung. Berücksichtigt man statt dem Interesse die Freude am Lesen, verschwindet der Einfluss des Geschlechts auf die Leseleistung komplett. Dies deutet „darauf hin, dass die Geschlechterdifferenzen im Bereich Lesen zumindest zum Teil durch Unterschiede in motivationalen Merkmalen vermittelt sind“ (Stanat & Kunter, 2001, S. 265). In ähnlichen Analysen konnten vergleichbare Auswirkungen für den Bereich Mathematik nicht gefunden werden.

Vergleichbare Ergebnisse für die zweite PISA-Untersuchung sind bei Zimmer, Burba und Rost (2004) berichtet. Ähnliche Ergebnisse für den naturwissenschaftlichen Bereich beim Vergleich der Bundesländer der Bundesrepublik Deutschland finden sich bei Prenzel, Carstensen, Rost und Senkbeil (2002) und bei Rost, Senkbeil, Walter, Carstensen und Prenzel

(2005). Insgesamt unterstreichen die berichteten Ergebnisse aus PISA die oben genannten Stärken und Schwächen für die Geschlechter. Darüber hinaus zeigen sie, dass die Wirklichkeit komplexer ist und Faktoren wie die Schulform mit berücksichtigt werden müssen.

Zur Erklärung von Geschlechtsunterschieden verweist Kasten (2001) auf verschiedene psychologische Theorien. Nach der *Bekräftigungstheorie*, die auf der operanten Lerntheorie beruht, werden Mädchen und Jungen vom Beginn des Lebens an für geschlechtsangemessene Verhaltensweisen verstärkt. Das würde eine differenzierte Behandlung der Kinder durch beispielsweise deren Eltern voraussetzen, was empirisch nicht belegt werden konnte. Nach der *Imitationstheorie*, die auf der sozialkognitiven Lerntheorie beruht, imitieren Kinder gleichgeschlechtliche Modelle. Nicht erklären kann die Theorie allerdings, warum beispielsweise Jungen, die in der Kindheit häufig auf erwachsene weibliche Modelle treffen, trotzdem die Verhaltensweisen gleichgeschlechtlicher Modelle übernehmen sollen. Nach der *Identifikationstheorie* identifizieren sich die Kinder mit ihren gleichgeschlechtlichen Bezugspersonen und übernehmen deren Werte und Einstellungen. Nicht erklären kann diese Theorie, wieso die Identifikation gerade mit den gleichgeschlechtlichen Bezugspersonen stattfinden soll.

Ein anderer Ansatz findet sich bei Chasiotis und Voland (1998). Sie stellen Überlegungen zur Variablen Geschlecht aus der Perspektive der *evolutionären Entwicklungspsychologie* vor. Aus diesem kurzen Abriss über theoretische Vorstellungen zur Entstehung von Geschlechtsunterschieden wird klar, dass es keinen umfassenden Erklärungsansatz gibt. Alle Theorien können sicherlich relevante Teilaspekte für verschiedene Kontexte erklären.

Lebensalter wird laut Eckensberger und Keller (1998) oftmals als unabhängige Variable angesehen. Verhalten wird in diesem Zusammenhang also als Funktion des Lebensalters erklärt. Korrekt müsste das Lebensalter lediglich als Zeitindex für die abhängige Variable verstanden werden. In dieser Arbeit wird das Lebensalter nicht als psychologisch bedeutsame Variable angesehen, sondern als Indikator für biologische Reifung und Lebenserfahrung. Wobei herausgestellt werden muss, dass sich biologische Reifung nicht unabhängig von der Umwelt vollziehen kann (Asendorpf, 1998) und damit das Lebensalter streng genommen keine validen und reliablen Rückschlüsse auf die Entwicklung zulässt.

Den Zusammenhang von Entwicklung und Zeit stellt Thomae (1959) treffend dar, wenn er schreibt, dass Entwicklung „als Reihe von miteinander zusammenhängenden Veränderungen, die bestimmten Orten des zeitlichen Kontinuums eines individuellen Lebenslaufs zuzuordnen sind“ (Thomae, 1959, S. 10) verstanden werden kann. Vorteilhaft für die vorliegende Arbeit ist, dass bei einer Datenerhebung in Schulklassen eines Jahrganges von einer

geringen Variation des Lebensalters in einem durch die Einschulung festgelegten Rahmen ausgegangen werden kann. Das Lebensalter wird damit quasi konstant gehalten und bedarf keiner weiteren Berücksichtigung. In Abschnitt 3.3 werden detaillierte Analysen dargestellt, die zeigen, dass diese Annahmen tatsächlich gerechtfertigt sind.

2.4 Sozialer Status

Nachdem bisher Einflussfaktoren aus dem Bereich der Schüler behandelt worden sind, soll im Folgenden auf den Bereich der Familie, auf die häuslichen Rahmenbedingungen eingegangen werden, die Schulleistungen beeinflussen können. Ein wichtiger Einflussfaktor aus diesem Bereich ist der soziale Status der Familie, der entscheidend von der Berufstätigkeit der Eltern abhängt. Helmke und Weinert (1997) sprechen von sozialer Schichtzugehörigkeit. Sie berichten von Befunden, wonach eine höhere soziale Schicht auch mit höheren Schulleistungen zusammenhängt. Allerdings sei direkt vorab darauf hingewiesen, dass Statusmerkmale wie Schichtzugehörigkeit oder Berufstätigkeit der Eltern keinen direkten Erklärungswert für Schulleistungen haben, sondern diese allenfalls indirekt beeinflussen können (Helmke & Schrader, 2001).

In einer älteren Arbeit von Beer et al. (1968), auf die bereits in Abschnitt 2.2 eingegangen wurde, wird nicht von häuslichen Rahmenbedingungen oder sozialem Status berichtet. Statt dessen wird der Begriff *Milieu* verwendet. Die Autoren weisen am Anfang ihrer mittlerweile fast 40 Jahre alten Arbeit darauf hin, dass die Literatur zum Thema Milieu und Schulleistung „außerordentlich umfangreich“ (Beer et al., 1968, S. 21) ist. Bis heute wird sich dieser außerordentliche Umfang entsprechend vergrößert haben. Für ihre Untersuchung zogen Beer et al. (1968) eine Stichprobe aus allen Schülern der Stadt Wien, die in der sechsten Schulstufe unterrichtet wurden ($N = 1\,397$). Zur Datenerhebung wurde als Intelligenzmaß das Begabungs-Test-System eingesetzt. Daneben kam ein Schülerfragebogen, ein Lehrerfragebogen und ein soziometrischer Test zum Einsatz.

Aus den Fragebogendaten wurden Daten zum Milieu der Schüler gewonnen, wie die Familienstruktur, die Berufsgruppe der Eltern, die Berufstätigkeit der Eltern und anderes mehr. Auf Grund der Intelligenztestwerte wurden die Schüler in vier Intelligenzgruppen unterteilt. Dabei ergaben sich mit Korrelationen nach Pearson vergleichbare korrigierte Kontingenzkoeffizienten von .17 bis .33 für den Zusammenhang zwischen den verschiedenen Milieuvvariablen und der Intelligenz. Der höchste Zusammenhang zeigte sich mit der

Berufsgruppe der Eltern, ausgedrückt über vier Kategorien zum Qualifikationsgrad der Berufe. Anhand der Noten wurden die Schüler zudem verschiedenen Schulleistungsgruppen zugeordnet. Hier fanden sich Kontingenzkoeffizienten von .10 bis .22 für die verschiedenen MilieuvARIABLEN, wobei der größte Zusammenhang mit dem Familienleben auftrat, der durch drei Kategorien zur „Gedeihlichkeit“ der Familienverhältnisse operationalisiert wurde. Für den Zusammenhang von Intelligenz und Milieu gehen die Autoren davon aus, dass die Intelligenz vom Milieu beeinflusst wird und dass beide Variablen von anderen Variablen beeinflusst werden können. Für den Zusammenhang von Schulleistung und Milieu ist nach den Ergebnissen ebenfalls davon auszugehen, dass die Schulleistung vom Milieu beeinflusst wird.

Als neuere Untersuchung, die Einflüsse der häuslichen Rahmenbedingungen berücksichtigt, sei PISA genannt. Folgt man den Ausführungen von Baumert und Schümer (2001), haben diese Faktoren bei PISA sogar einen herausragenden Stellenwert: Besonders in demokratischen Gesellschaften ist es bedeutsam, dass jeder Mensch unabhängig von häuslichen Rahmenbedingungen Gelegenheit hat, eine Grundbildung zu erreichen. Dies so weit wie möglich zu gewährleisten ist Aufgabe des Bildungssystems. Um zu prüfen, ob das Bildungssystem dazu in der Lage ist, muss zunächst die *soziale Herkunft* der Schüler bestimmt werden. Im nächsten Schritt kann dann der Bildungsgrad in Abhängigkeit von der sozialen Herkunft betrachtet werden. Baumert und Schümer (2001) schreiben dazu:

Die soziale Herkunft von Schülerinnen und Schülern wird üblicherweise mithilfe der sozioökonomischen Stellung ihrer Familien bestimmt, das heißt mithilfe von Daten zur relativen Position ihrer Eltern in einer sozialen Hierarchie, deren Ordnungsprinzipien in der Verfügung über finanzielle Mittel, Macht oder Prestige bestehen. Da Informationen über Einkommensverhältnisse, Macht und soziale Anerkennung von Individuen nicht einfach zu erhalten sind, wird die sozioökonomische Stellung in aller Regel über die Berufstätigkeit erfasst, die Hinweise auf jeden der drei Aspekte ihrer Stellung in der sozialen Hierarchie geben kann. Daten zum Beruf und zur beruflichen Stellung des Vaters oder beider Eltern waren auch in der Bildungsforschung immer die wichtigsten Kriterien, mit deren Hilfe die soziale Herkunft von Schülern bestimmt wurde. (S. 326)

Die Berufstätigkeit ist also in der Bildungsforschung eine bereits lange berücksichtigte Variable. Um dabei international vergleichbare Daten zu erhalten, dient oftmals die internationale Berufsklassifikation des *International Labor Office* als Klassifikationsgrundlage, die in der aktuellsten Fassung von 1988 vorliegt. Berufe werden dabei in ein vierstelliges Klassifikationsschema eingeordnet. Dabei werden 10 Hauptgruppen, wie *gelernte Arbeiter in Landwirtschaft und Fischerei* oder *Führungskräfte*, unterschieden, die an der ersten Stelle durch eine Ziffer von null bis neun gekennzeichnet werden. Ganzeboom und Treiman (1996) haben passend zu dieser Klassifikationsgrundlage verschiedene Maße des beruflichen Status

bestimmt. Dabei unterscheiden sie prinzipiell drei verschiedene Arten von Maßen: Berufsprestigemaße, sozioökonomische Statusmaße und kategoriale Klassenmaße.

Berufsprestigemaße sollen das Ansehen eines Berufes in der Gesellschaft wiedergeben. Ein Beispiel ist die *Standard International Occupational Prestige Scale* (SIOPS). Bei *sozioökonomischen Statusmaßen* wird eine gewichtete Summe verschiedener Kriterien wie Bildung und Einkommen berechnet. Ein solches Statusmaß ist der von Ganzeboom, De Graaf und Treiman (1992) vorgestellte *International Socio-Economic Index of occupational status* (ISEI). Bei diesen beiden Arten von Maßen handelt es sich um kontinuierliche Maße. Bei den *kategorialen Klassenmaßen* handelt es sich im Unterschied dazu – wie der Name schon sagt – um diskrete Maße, bei denen Personen in diskrete Kategorien eingeordnet werden. Trotz der Kategorisierung können bei kategorialen Klassenmaßen verschiedene relevante Informationen in die Beurteilung mit einfließen.

Das etablierteste kategoriale Klassenmaß ist das *EGP-Klassenschema* von Erikson, Goldthorpe und Portocarero (1979). Es liegt in verschiedenen, leicht veränderten Fassungen vor, wobei Ganzeboom und Treiman (1996) die in Tabelle 2 wiedergegebene Klassifikation verwenden. In Tabelle 2, die auch den Zusammenhang zwischen dem EGP-Klassenschema und dem ISEI zeigen soll, ist zu erkennen, dass sich die Mittelwerte im ISEI für die einzelnen EGP-Kategorien unterscheiden. Erikson et al. (1979) berichten von ihrem Maß im Zusammenhang mit einer Untersuchung zur Veränderung der Klassenzugehörigkeit von Menschen in England, Frankreich und Schweden über verschiedene Generationen hinweg. Gegenüber den beiden anderen Maßen hat das EGP-Klassenschema laut Baumert und Schümer (2001) den Vorteil, dass es theoretisch besser fundiert ist und keine Informationen über Bildung und Einkommen zur Bestimmung vorliegen müssen. Allerdings müssen Informationen zur beruflichen Stellung und den Weisungsbefugnissen vorhanden sein.

Ganzeboom und Treiman (1996) haben die von ihnen vorlegte Zuordnung von Berufen nach der internationalen Berufsklassifikation und den benannten Maßen des beruflichen Status anhand einer Stichprobe von Männern aus 16 Ländern durchgeführt. Diese Stichprobe wurde in insgesamt 31 Untersuchungen erhoben ($N = 73\,901$). Diese Untersuchungen waren vom Untersuchungsplan her nur auf Männer ausgelegt; es waren somit schlicht keine Daten für Frauen vorhanden. Ganzeboom und Treiman (1996) sehen darin aber keine Einschränkung der Nutzbarkeit ihrer Zuordnung. Die Zuordnung wurde zusätzlich an einer zur benannten Stichprobe unabhängigen Stichprobe aus 14 Ländern berechnet ($N = 17\,386$). Dabei zeigt

sich, dass die verschiedenen Maße des beruflichen Status über die verschiedenen Versionen der internationalen Berufsklassifikation hinweg vergleichbar sind.

Tabelle 2

Die Tabelle zeigt das EGP-Klassenschema aus Ganzeboom und Treiman (1996), wobei die EGP-Kategorie in der linken Spalte und die englische Bezeichnung in der mittleren Spalte angegeben ist. In der rechten Spalte ist zum Vergleich der Mittelwert des International Socio-Economic Index of occupational status (M_{ISEI}) für jede EGP-Kategorie angegeben. (erstellt in Anlehnung an Ganzeboom & Treiman, 1996, Tabelle 1, S. 214)

Kategorie	Beschreibung	M_{ISEI}
1	Higher service Includes mostly professionals, large enterprise employers and higher managers	68
2	Lower service Includes mostly associate professionals, lower managers, higher sales	58
3	Routine clerical/sales Includes routine clerical and sales workers	45
4	Small employers Includes small entrepreneurs	48
5	Independent Own account workers, no employees	42
6	Not used Because Farmers and Farm managers, who are originally classified in category 6, are being classified in category 11 by Ganzeboom and Treiman (1996).	-
7	Manual foremen Manual workers with supervisory status	40
8	Skilled Manual Mostly craft workers, some skilled service, and skilled machine operators	36
9	Semi-Unskilled manual Mostly machine operators, elementary laborers, elementary sales and services	31
10	Farm workers Employed farm workers, irrespective of skill level; also family farm workers	18
11	Farmers/Farm managers Self-employed and supervisory farm workers, irrespective of skill level	26

Seit einiger Zeit werden laut Baumert und Schümer (2001) neben Maßen des beruflichen Status vielfach auch Daten zum kulturellen und sozialen Kapital von Familien erhoben. Beim *kulturellen Kapital* ist die Grundannahme, dass der Bildungserfolg entscheidend vom Ausmaß der Teilhabe an der Kultur der bürgerlichen Mittelschicht abhängt. Zu kulturellem Kapital gehören Kulturgüter wie Literatur und Bildungszertifikate, aber auch verinnerlichte Denk- und Handlungsschemata.

Der Begriff des kulturellen Kapitals stammt von Bourdieu, der in seiner Arbeit von 1983 Kapital als akkumulierte Arbeit ansieht. Er grenzt seinen Kapitalbegriff von einem ökonomischen Kapitalbegriff ab, der alle gesellschaftlichen Austauschvorgänge auf den

Austausch von Waren reduziert. Er unterscheidet drei Arten von Kapital. Das ökonomische Kapital ist direkt in Geld tauschbar. Unter sozialem Kapital versteht er soziale Verpflichtungen, Beziehungen und Netzwerke, kurz „Ressourcen, die auf der *Zugehörigkeit zu einer* [sic] *Gruppe* beruhen“ (Bourdieu, 1983, S. 190f). Das kulturelle Kapital kann in drei Zustandsformen vorkommen: Zum einen verinnerlicht, wie in Form von Bildung, zum anderen objektiviert, wie in Form von Gemälden und Instrumenten und schließlich institutionalisiert, wie in Form von Titeln und Bildungszertifikaten. Die beiden letztgenannten Kapitalarten sind unter bestimmten Voraussetzungen in ökonomisches Kapital überführbar. Bei seinen Überlegungen geht er laut Kraus (1983) immer davon aus, dass eine Klassengesellschaft existiert, dass in dieser Gesellschaft die Möglichkeit zur Teilhabe an Wohlstand und Entwicklung ungleich verteilt ist und das Besitz von Kapital hauptsächlich vererbt wird.

Unter *sozialem Kapital* wird ein Netzwerk von sozialen Beziehungen verstanden, in dem Normen, Werte und Informationen ausgetauscht werden. Der Begriff des sozialen Kapitals wird zwar auch vor Bourdieu (1983) behandelt, ausführlicher geht allerdings Coleman (1988) darauf ein. Er grenzt soziales Kapital von physischem Kapital (z.B. Maschinen und Werkzeuge), Humankapital (z.B. Fähigkeiten, Fertigkeiten und Kenntnissen von Personen) und finanziellem Kapital ab. Soziales Kapital existiert in Form von Beziehungen zwischen Personen und kann die Produktivität – ebenso wie die anderen Kapitalarten – steigern. Zudem legt es die Grundlage für die Bildung der nächsten Generation und damit für deren Humankapital. In der Wirtschaft wird Bildung und Humankapital eine bedeutende Rolle für die zukünftige wirtschaftliche Entwicklung zugewiesen (Klöß & Plünnecke, 2003). Auf den letztgenannten Aspekt geht Coleman (1987) ausführlicher ein. Er betont, dass das Humankapital, gemessen an erreichten Bildungsabschlüssen, immer mehr zunimmt und das soziale Kapital, gemessen an der Zeit, die Eltern und Kinder zur Interaktion haben, immer mehr abnimmt. Diese Abnahme sollte vom Staat, beispielsweise durch die Einrichtung von Schulen, ausgeglichen werden. Die Unternehmen, welche die Abnahme durch die Bindung ihrer Mitarbeiter verursachen, sehen sich dabei scheinbar nicht in der Pflicht.

Coleman (1988) unterscheidet zwischen sozialem Kapital innerhalb von Familien und außerhalb von Familien. Innerhalb von Familien zeigt es sich in der Zeit, die Eltern mit den Kindern verbringen und ihnen ihr Humankapital zur Verfügung stellen. Demnach ist soziales Kapital über eben diese Zeit messbar oder über die Zahl der Geschwister, durch welche die begrenzte Zeit der Eltern geteilt werden muss. Außerhalb von Familien kann soziales Kapital beispielsweise in religiösen Gemeinschaften liegen. Coleman (1988) berichtet von

einer Untersuchung, die zum Ergebnis hatte, dass an katholischen Privatschulen in den Vereinigten Staaten geringere Abbrecherraten zu verzeichnen waren als an öffentlichen Schulen oder Privatschulen anderer Träger. Er führt dies nicht auf die Konfession zurück und verweist darauf, dass sich vergleichbare Ergebnisse zum Beispiel für jüdische Privatschulen zeigen. Er erklärt die Befunde damit, dass die Schüler an den Schulen der religiösen Gemeinschaften einfach mehr soziales Kapital zur Verfügung haben. Überdies weist er darauf hin, dass soziales Kapital – im Gegensatz zu den drei anderen Kapitalarten – überwiegend kein privates Gut ist. Vielmehr beeinflusst es alle am sozialen Netz Beteiligten und kann so Vorteile und Nachteile für alle bringen. Beispielsweise können Freunde und Bekannte den Wegzug einer Familie aus einem Ort wegen eines besseren Arbeitsplatzes des Vaters nur begrenzt beeinflussen. Ist der Wegzug vollzogen, können die im Ort Verbliebenen Nachteile zu tragen haben.

Wie bereits weiter oben angedeutet wird bei PISA zur Bestimmung der sozialen Herkunft der Schüler die sozioökonomische Stellung der Eltern betrachtet. Baumert und Schümer (2001) berichten genauer von der Operationalisierung: Zur Bestimmung der sozioökonomischen Stellung der Eltern wurden Informationen über den Erwerbstätigkeitsstatus und den Beruf der Eltern, sowie über deren Weisungsbefugnisse erhoben und auf dieser Grundlage ISEI-Werte und EGP-Werte bestimmt. Zudem wurden der Wohlstand der Familie über die Wohnverhältnisse und der Besitz von teuren Gebrauchsgütern wie beispielsweise Geschirrspülmaschine und Auto erfasst. Im Vorfeld wurde dazu ein Feldtest in allen an PISA beteiligten Ländern durchgeführt, um die Aussagekraft der Indikatoren zu überprüfen. Dies führte dazu, dass bei der eigentlichen Datenerhebung andere Indikatoren erhoben wurden, als ursprünglich vorgesehen. Laut Baumert und Schümer (2001) bleibt es allerdings ohne genauere Validierungsuntersuchungen fraglich, ob die verwendeten Wohlstandsindikatoren „transkulturell äquivalent sind“ (ebd., S. 332).

Zur Bestimmung des kulturellen Kapitals der Familie wurde beispielsweise das Geburtsland der Schüler, die zu Hause gesprochene Sprache und der Bildungsstand der Eltern erfasst. Überdies wurde erfragt, welche Kulturgüter wie Bilder und Gedichte in der Familie vorhanden sind, welche schulrelevanten Besitztümer wie Taschenrechner zur Verfügung stehen und wie das kulturelle Leben ausgeprägt ist. Letzteres wurde über Fragen zu Museumsbesuchen und Ähnlichem operationalisiert.

Zur Bestimmung des sozialen Kapitals wurde beispielsweise die Größe des Haushalts, die Stellung in der Geschwisterreihe, der Erziehungsstil der Eltern und deren

Erwerbstätigkeitsstatus erfasst. Letzterer ist ein Indikator für die Zeit, die Eltern mit ihren Kindern verbringen. Die Annahme ist, dass sich in dieser Zeit soziales Kapital überhaupt erst bilden kann. Bei PISA fehlten die Sozialschichtangaben von 3.1% der Schüler, wobei für schulleistungsschwächere Schüler häufiger die Angaben fehlen. Ehmke, Hohensee, Heide-meier und Prenzel (2004) stellen ein noch zu erwähnendes Maß vor, dass ökonomische, soziale und kulturelle Merkmale kombiniert, den *Index of Economic, Social and Cultural Status* (ESCS). Weiterführende Informationen zu diesem Maß finden sich bei Ramm et al. (2006).

Von den vielfältigen Befunden, die sich bei PISA im Zusammenhang mit den häuslichen Rahmenbedingungen ergeben haben, kann an dieser Stelle nur eine exemplarische Auswahl aus Baumert und Schümer (2001) wiedergegeben werden. Die Mehrheit der Schüler in Deutschland lebt in stabilen Familienverhältnissen. Mit einem Elternteil leben 16% zusammen. In über 90% der Familien ist der höchste vorhandene Bildungsabschluss der Eltern mindestens der Hauptschulabschluss mit anschließender Lehre. Von den Schülern stammen 27% aus Familien, in denen mindestens ein Elternteil einen Migrationshintergrund hat. Diese Eltern sind zum größten Teil als Arbeiter beschäftigt. Tendenziell zeigt sich, dass bei niedrigeren Kategorien im EGP-Klassenschema, ein höherer Anteil der Schüler ein Gymnasium besucht und bei höheren Kategorien wiederum ein höherer Anteil eine Hauptschule. Zur besseren Verständlichkeit ist in Tabelle 3 das bei PISA verwendete EGP-Klassenschema wiedergegeben.

Tabelle 3

Die Tabelle zeigt das bei PISA verwendete EGP-Klassenschema, wobei die Kategorie nach PISA in der ersten und die Bezeichnung in der zweiten Spalte angegeben ist. In der dritten Spalte ist die Kategorie nach Ganzeboom und Treiman (G&T, 1996) und in der vierten Spalte der auf ganze Zahlen gerundete Mittelwert im International Socio-Economic Index of occupational status (M_{ISEI}) aufgeführt. (erstellt in Anlehnung an Baumert & Schümer, 2001, Tabelle 8.3, S. 340)

Kategorie PISA	Bezeichnung	Kategorie G&T	M_{ISEI}
I	Obere Dienstklasse	1	66
II	Untere Dienstklasse	2	55
III	Routinedienstleistungen in Handel und Verwaltung	3	44
IV	Selbstständige und selbstständige Landwirte	4, 5, 11	40
V, VI	Facharbeiter und Arbeiter mit Leistungsfunktion sowie Angestellte in manuellen Berufen	7, 8	34
VII	Un- und angelernte Arbeiter sowie Landarbeiter	9, 10	28

Sprachkompetenz und Fachkompetenz in Erdkunde, wie sie in der vorliegenden Arbeit besonders interessieren, und deren Zusammenhänge mit sozialem Status werden bei PISA nicht direkt untersucht. Ein Teil von Sprachkompetenz ist sicherlich die bei PISA untersuchte Lesekompetenz. Für diese zeigen sich in Deutschland nach Baumert und Schümer (2001) geringe Unterschiede für Schüler aus Familie der oberen und unteren Dienstklasse, also der EGP-Kategorien I und II. Die Schüler aus den weiteren EGP-Kategorien III bis VII schneiden im Vergleich dazu deutlich schlechter ab. Allerdings gibt es in allen EGP-Kategorien Schüler mit geringer und mit hoher Lesekompetenz.

Weiterhin wurde bei PISA die naturwissenschaftliche Kompetenz erfasst, bei der auch geowissenschaftliche Anteile berücksichtigt sind (Prenzel, Rost, Senkbeil, Häußler & Klopp, 2001). Dabei zeigt sich für die unterschiedlichen EGP-Kategorien nach Baumert und Schümer (2001) ein der Lesekompetenz vergleichbares Muster, allerdings in einer geringeren Ausprägung. Für die gesamte internationale Erhebung „zeigt sich ein positiver Zusammenhang zwischen sozialer Herkunft und Lesekompetenz“ (Baumert & Schümer, 2001, S. 387). Dieser *soziale Gradient* der Lesekompetenz ist für deutsche Schüler im Vergleich zu den anderen Teilnehmerstaaten mit ähnlicher Sozialstruktur am steilsten. Gleichzeitig haben die deutschen Schüler im Vergleich mit diesen Staaten das geringste mittlere Leistungsniveau.

Für die vorliegende Untersuchung ist die Berücksichtigung des sozialen Status wichtig, weil laut Vollmer (2006b) möglichst nur solche Schüler zu vergleichen sind, „die von ihrem Profil und ihrer sozial-kognitiven Ausstattung her auch tatsächlich vergleichbar sind“ (Vollmer, 2006b, S. 222). So sollen die Ergebnisse gegenüber einer vermuteten Selbstselektion guter Schüler aus Familien mit hohem sozialen Status für die Teilnahme an bilingualen Bildungsgängen robuster gemacht werden. Es sei an dieser Stelle explizit darauf hingewiesen, dass im Folgenden die *soziale Herkunft der Schüler* mit dem *sozialen Status der Familien der Schüler* gleichgesetzt wird, da – wie oben dargestellt – die soziale Herkunft der Schüler besonders über die Berufstätigkeit der Eltern bestimmt wird, die den sozialen Status der Familie ausmacht.

2.5 Klassenklima

Nachdem bisher Einflussfaktoren aus dem Bereich von Schüler und Familie behandelt worden sind, soll in diesem und dem folgenden Abschnitt auf Einflussfaktoren aus dem Bereich der Schule eingegangen werden. Im Rahmen von PISA wurde unter dem Oberbegriff der

institutionellen Bedingungen schulischen Lernens unter anderem das Schulklima untersucht (Weiß & Steinert, 2001). Da im Rahmen des DFG-Projektes allerdings deutlich weniger Schüler von einzelnen Schulen untersucht wurden als bei PISA, soll der Fokus des untersuchten Klimas kleiner gewählt werden. Daher wird in diesem Abschnitt das Klassenklima behandelt.

Unter dem Klassenklima versteht man die „Wahrnehmungen und Beurteilungen von Aspekten des Unterrichts, des Lehrer-Schüler-Verhältnisses und der Lehrperson durch die Schülerinnen und Schüler einer Klasse“ (Helmke, 2005, S. 33). Diese Begriffsbestimmung deutet bereits an, dass Klassenklima ein schlecht fassbares Konstrukt ist. Befragt man die Schüler zum Klassenklima, so erhält man verschiedene subjektive Sichtweisen. Diese können von Schüler zu Schüler natürlich unterschiedlich sein. Daher kann sich empirisch eine große Streuung in der Beurteilung des Klassenklimas innerhalb von Klassen zeigen (Helmke, 2005).

In der vorliegenden Arbeit interessiert das Klassenklima aus verschiedenen Gründen. Erstens wird dem Klassenklima von Pädagogen eine große Bedeutung für erfolgreichen Unterricht und erfolgreiches Lernen beigemessen. Beispielsweise gehen Bülter und Meyer (2004) davon aus, dass ein lernförderliches Klima dem Lehrer hilft, ein Arbeitsbündnis mit den Schülern zu schließen. Darüber hinaus hat es förderliche Wirkungen auf das Selbstvertrauen der Schüler, deren Lern- und Leistungsverhalten und deren fachliche Interessenbildung. Weiterhin gehen Bülter und Meyer (2004) davon aus, dass sich durch ein lernförderliches Klima das Sozialverhalten der Schüler und deren Einstellung zu Schule und Unterricht verbessert. Zweitens ist das Klassenklima zwar laut Helmke und Weinert (1997) ein vager Begriff, besitzt aber trotzdem einen nicht zu vernachlässigenden Wert bei der Erklärung von Schulleistungen. Drittens ist das Klassenklima einer von vielen Faktoren, die beispielsweise in Niedersachsen in die Beurteilung der *Schulqualität* einfließen (Minderop, 2004). Die Untersuchung des Klassenklimas hat also auch im Bereich der Qualitätssicherung ihren Platz.

Um die vage Begrifflichkeit des Klassenklimas zu schärfen, soll an dieser Stelle auf Eder (2001) zurückgegriffen werden. Er unterscheidet verschiedene Verwendungsweisen des Begriffes „Klima“. Für die vorliegende Arbeit passt die Verwendungsweise, bei der mit dem Klassenklima die subjektiv wahrgenommene Lernumwelt beschrieben wird. Eder (2001) unterscheidet drei Klimatypen, das *individuelle Klima* (Wahrnehmung von Klima durch einzelne Personen), das *aggregierte Klima* (durchschnittliche Wahrnehmung von Klima einer Gruppe von Personen) und das *kollektive Klima* (ähnliche Umweltwahrnehmung durch Gruppen von

Personen). Das hier interessierende Klassenklima kann als aggregiertes Klima angesehen werden.

Weiterhin sieht Eder (2001) fünf verschiedene Facetten als wichtig an, um den Klimabegriff zu verstehen. Die erste Facette ist der Inhalt. Darunter versteht er Aspekte des betrachteten Bereiches einer Organisation, wie die sozialen Beziehungen der Schüler untereinander oder das Verhalten im Unterricht. In der vorliegenden Arbeit interessiert ein *Sozialklima*, da es um das Miteinander der Schüler geht. Die zweite Facette ist der Organisationsbezug, womit der untersuchte Ausschnitt aus einer Organisation angesprochen wird. Das hier interessierende Klima fokussiert auf Schulklassen und ist als *Klassenklima* zu verstehen. Die dritte Facette ist der Subjektbezug, wobei zwischen einer subjektiven Sicht der Person auf deren Umwelt (*psychologisches Klima*) und einer konsensuellen Aussage über die Umwelt in einer Organisation (*kollektives Klima*) unterschieden werden kann. Da in der vorliegenden Arbeit die subjektive Sicht der Schüler interessiert, handelt es sich um ein psychologisches Klima. Die vierte Facette ist die Aggregierungsebene, die beschreiben soll, ob Daten von Individuen (*Individualklima*) oder Mittelwerte über verschiedene Individuen (*aggregiertes Klima*) betrachtet werden. In der vorliegenden Untersuchung wird bei der Datenauswertung über verschiedene Schüler gemittelt, womit es sich um ein aggregiertes Klima handelt. Die fünfte und letzte Facette ist die Quelle und bezieht sich darauf, bei welchen Personen die Daten erhoben wurden. Im vorliegenden Fall wurden Schüler befragt; damit handelt es sich um ein *Schülerklima*.

Bei der Messung des Klimas treten nach Eder (2001) verschiedene Probleme auf. Gemessen wird das Klima meist über standardisierte schriftliche Befragung. Bei der Itemgestaltung muss einerseits die Wahrnehmungsperspektive berücksichtigt werden. Unterscheiden lassen sich dabei Selbstbeschreibungen von Einzelpersonen, Selbstbeschreibungen von Gruppen und Beschreibungen einer Gruppe über andere Gruppen. Andererseits muss der Inferenzgrad der Aussagen beachtet werden. Es gibt beschreibende und bewertende Items, sowie solche, die hochinferent und niedriginferent sind. Bei letztgenannten Items handelt es sich um objektiv nachprüfbar Aussagen, bei denen der Beantwortende keine Schlüsse ziehen muss. Eder (2001) führt verschiedene standardisierte Erhebungsverfahren zum Klima auf. Auf die dort genannten Verfahren wird im DFG-Projekt nicht zurückgegriffen. Stattdessen findet ein selbst erstelltes Verfahren Verwendung, auf das in Abschnitt 3.2.2 noch genauer eingegangen wird.

Prinzipiell sind in Schulen verschiedene Ebenen der Untersuchung denkbar, die sich auch in den oben genannten Facetten widerspiegeln. In Bezug auf die Messung des Klimas weist Eder (2001) darauf hin, dass dabei zwischen *Beobachtungsebene*, auf der die Daten erhoben werden, *Analyseebene*, auf der sie analysiert werden und der *Interpretationsebene*, auf der sie interpretiert werden, zu unterscheiden ist. Die beiden zuletzt genannten Ebenen müssen einander entsprechen. Im DFG-Projekt wurden die Daten auf Schülerebene beobachtet. In der vorliegenden Arbeit werden sie aggregiert über verschiedene Klassen analysiert und interpretiert.

Zur Problematik der Arbeit mit verschiedenen Ebenen in Untersuchungen, wie Individual-, Klassen- oder Schulebene, sei auf von Saldern (2001) hingewiesen, der Ansätze der Mehrebenenanalyse erläutert. Dabei kann es auch bei Mehrebenenanalysen dazu kommen, dass *Kontexteffekte* – damit sind in der pädagogisch-psychologischen Forschung Auswirkungen der Lehr-Lernumwelt auf das schulische Lernen gemeint – nur als statistische Artefakte auftreten. Dies konnten Lüdtke, Robitzsch und Köller (2002) sowohl mit Hilfe einer Simulation, als auch mit einer Reanalyse echter Daten zeigen. Sie sehen diese Gefahr besonders bei Variablen der Lernumwelt von Schülern, die – wie beispielsweise das Klassenklima – auf Individualebene erhoben werden und anschließend auf Klassenebene aggregiert in Berechnungen eingehen. Sie konnten zeigen, dass die Wahrscheinlichkeit statt eines echten Effektes nur ein statistisches Artefakt vorliegen zu haben, mit sinkender Reliabilität der Messungen und steigender Intraklassenkorrelation zunimmt. Die beiden älteren Arbeiten von Farkas (1974) und Hauser (1974) befassen sich ebenfalls mit dieser Problematik, wobei sich ersterer für und letzterer gegen die Untersuchung von Kontexteffekten ausspricht.

2.6 Bilingualer Unterricht

In Deutschland versteht man unter bilinguaalem Unterricht, bilingualen Bildungsgängen oder bilingualen Zügen, respektive bilingualen Zweigen, ganz allgemein Formen der schulischen Bildung, bei der im Unterricht verstärkt eine Fremdsprache zum Einsatz kommt. Im Regelfall bietet eine weiterführende Schule einen *bilingualen Zug* an, bei dem die Schüler in der fünften und sechsten Jahrgangsstufe mehr Wochenstunden Unterricht in der jeweiligen Fremdsprache bekommen als die monolingualen Schüler. Von Jahrgangsstufe sieben bis Jahrgangsstufe zehn werden dann verschiedene Sachfächer in der Fremdsprache unterrichtet. In der Sekundarstufe II wird der bilinguale Zug dann beispielsweise mit einem Leistungskurs

in der Fremdsprache und einem fremdsprachlichen Sachfach bis zum Abitur weitergeführt. Diesen Fachunterricht, der in einer anderen als der Muttersprache durchgeführt wird, bezeichnet man als *bilingualen Unterricht*. Als weitere Form *bilingualer Bildungsgänge* werden von Christ (1999) *bilinguale Module* diskutiert, bei denen phasenweise bilinguale Unterrichtseinheiten oder bilinguale Projekte durchgeführt werden.

Dem aufmerksamen Leser wird aufgefallen sein, dass hier letztlich ein-sprachiger Fachunterricht in einer Fremdsprache als bi-lingual bezeichnet wird. Von verschiedenen Autoren wird diese Namensgebung kritisiert. Niemeier (2005) betont, dass es sich beim deutschen Ansatz nicht um das gleichzeitige Benutzen zweier Sprachen handele, wie die Begrifflichkeit eigentlich impliziert. Beim *natürlichen Bilingualismus* wechseln Sprecher zwischen zwei Sprachen, teilweise in einem Satz. Dies ist beispielsweise bei Kindern von Einwanderer anzutreffen, die durch ihr Umfeld sowohl die Muttersprache der Eltern, als auch die Landessprache des Landes in dem sie leben, fließend sprechen können.

Bach (2005) schreibt, dass die deutschen bilingualen Bildungsgänge unter anderem auf Ansätzen aus Kanada beruhen. Diese Ansätze sind allerdings nicht direkt auf die deutschen Verhältnisse übertragbar, wie Vollmer (1992) herausstellt: Kanada ist ein offiziell zweisprachiges Land. Die bilingualen Bildungsgänge dort werden als *Immersionsprogramme* bezeichnet und wurden seit den 1960er Jahren von der Politik stark gefördert. Die teilnehmenden Kinder stammen mehrheitlich aus anglophonen Bevölkerungsteilen und haben vom ersten Schultag an sämtlichen Unterricht in französischer Sprache. Unterricht in englischer Sprache beginnt erst in späteren Jahrgangsstufen.

Laut Breidbach (2007) werden im Rahmen des bilingualen Unterrichts in Deutschland mit Sprachen wie Englisch oder Französisch reine Schulfremdsprachen behandelt, die keinen direkten Bezug zum außerschulischen sprachlichen Umfeld der Schüler haben. Diese Sprachen haben allenfalls Relevanz als *Lingua franca* oder für den späteren beruflichen Erfolg. Damit haben bilinguale Bildungsgänge trotz unterschiedlicher Rahmenbedingungen in Deutschland und Kanada ein gemeinsames Ziel: Die teilnehmenden Schüler sollen neben ihrer Muttersprache in einer weiteren Sprache profunde Sprachkenntnisse erwerben und Verständnis für die andere Kultur entwickeln.

Der heutige bilinguale Unterricht in deutschen Schulen reicht in seinen Ursprüngen bis ins Jahr 1963 zurück (SKMK, 2006). In diesem Jahr wurde der deutsch-französische Kooperationsvertrag geschlossen, der eine Förderung der beiden Landessprachen im jeweils anderen Partnerland vorsah. In der Folge wurde 1969 der erste bilinguale Zug an einer Schule

eingerrichtet. Eine deutliche Zunahme an bilingualen Zügen gab es laut Thürmann (2000) allerdings erst 1989. Dies hängt sicherlich damit zusammen, dass die Europäische Union Anfang der 1990er Jahre in ihrer Bedeutung für die Bürger zunahm und Mehrsprachigkeit spätestens seit diesem Zeitpunkt von Seiten der Politik als bedeutend für die Zukunft eines vereinten Europas angesehen wurde. Seitdem dominiert zudem Englisch als Sprache die bilingualen Züge, beide denen vorher meist Französisch im Fokus stand.

In allen Bundesländern der Bundesrepublik Deutschland werden heutzutage bilinguale Bildungsgänge angeboten (SKMK, 2006). Lediglich an Gymnasien gibt es dabei bilingualen Unterricht in allen Bundesländern. Für die anderen Schulformen zeigen sich je nach Bundesland deutliche Unterschiede. Besonders auffällig ist, dass es im Hauptschulbereich nur in Baden-Württemberg entsprechende Angebote gibt. Ebenfalls in allen Bundesländern wird Englisch als Sprache im bilingualen Unterricht angeboten. Alle anderen Sprachen sind nicht flächendeckend verfügbar. Französisch ist nach Englisch am stärksten vertreten. Dahinter folgen deutlich abgeschlagen Italienisch, Spanisch und Polnisch. Weitere Sprachen werden nur in einzelnen Bundesländern unterrichtet. Insgesamt bieten 868 Schulen bilingualen Unterricht in verschiedenen Organisationsformen an. Auf Unterschiede zwischen den einzelnen Bundesländern geht Jovanovic (1997) genauer ein.

Als bilinguale Fächer sind Erdkunde und Geschichte führend. Politik, Biologie, Kunst und Sport werden in der Mehrzahl der Bundesländer angeboten. Wenig angeboten werden die Fächer Mathematik, Musik, Chemie, Physik, Religion, Hauswirtschaft und Naturwissenschaft. Laut Zydati (2004) gibt es keine Fächer, die per se besonders gut oder schlecht für den bilingualen Unterricht geeignet sind. Die Kultusministerkonferenz äußert zu den verschiedenen Fächern:

Die Gründe für die Fächerwahl sind sehr unterschiedlich. Erdkunde ist aus didaktischen Gründen oft erstes Sachfach, da Anschaulichkeit und deskriptive Sprachhandlungen fachimmanent sind. Geschichte ist häufig erst das dritte Sachfach in der Reihe der nacheinander einsetzenden Sachfächer, d.h. nach Erdkunde und Politik bzw. Sozialkunde/Gemeinschaftskunde/Gesellschaftslehre, denn die historische Perspektivierung und die dazu geforderten Sprachhandlungen des Bewertens bereiten größere Schwierigkeiten. (SKMK, 2006, S. 17)

In der deutschen Schulpraxis scheint der bilinguale Unterricht oftmals als Unterricht für Eliteschüler gesehen zu werden. Dies resultiert aus verschiedenen Definitionen bilingualen Unterrichts der Bundesländer, in denen überdurchschnittliche Fremdsprachenkompetenzen als Ziel gesetzt werden. In Ländern wie den Vereinigten Staaten oder Kanada ist es dagegen Ziel,

die L2-Kompetenz in Richtung der L1-Kompetenz zu entwickeln. Mit der Bezeichnung *L1* ist die erste Sprache gemeint, die eine Person sprechen kann, eben deren *Erstsprache*. Unter *L2* versteht die Philologie analog die zweite Sprache, die eine Person sprechen kann. Jedoch ist hier zu unterscheiden, ob diese zweite Sprache innerhalb oder außerhalb eines muttersprachlichen Kontextes erlernt worden ist. Innerhalb eines solchen Kontextes erwirbt die Person eine *Zweitsprache*, außerhalb eine *Fremdsprache*. In den benannten Ländern ist also Ziel, die Kompetenzen in der Zweitsprache in Richtung der Kompetenzen in der Erstsprache zu entwickeln.

Hier kommen die unterschiedlichen Kontexte der bilingualen Bildung zum Tragen: In Deutschland sollen in erster Linie deutsche Muttersprachler mit zusätzlichen Fremdsprachenkompetenzen ausgestattet werden, um im europäischen Wettbewerb Vorteile zu besitzen. In den Vereinigten Staaten sollen dagegen beispielsweise Schüler aus Familien mit Migrationshintergrund in die Lage versetzt werden, beide Sprachen im Alltag in verschiedenen Kontexten angemessen einsetzen zu können.

In anderen europäischen Ländern wird unter der Bezeichnung *Content and Language Integrated Learning* (CLIL) in den Schulen neben der Muttersprache eine Fremdsprache als Arbeitssprache angeboten. Beim CLIL steht – wie der Name schon sagt – das gleichzeitige Lernen von Sprache und Inhalt im Zentrum der Bemühungen. In der Praxis wird dazu neben dem Fremdsprachenunterricht gezielt Sachfachunterricht in einer Fremdsprache durchgeführt, wobei sich die Umsetzung in den einzelnen europäischen Ländern, ebenso wie in den deutschen Bundesländern, unterscheidet. Auch der bilinguale Unterricht in Deutschland kann je nach Implementation als CLIL angesehen werden.

Neben der bereits aufgegriffenen Kritik an der Begrifflichkeit, gibt es durchaus weitergehende Kritik am bilingualen Unterricht. Decke-Cornill (1999) sieht durch immer mehr bilingualen Unterricht die Existenz des reinen Fremdsprachenunterrichts und den damit einhergehenden Gestaltungsspielräumen gefährdet. Ferner befürchtet sie, dass weniger leistungsstarke Schüler – besonders solche, deren Herkunftssprache nicht Deutsch ist – durch eine Sprachbarriere weniger am Lernerfolg und an der Gestaltung des Unterrichts teilhaben können. Zudem sieht sie die Gefahr, dass vorhandene Mehrsprachigkeit zu Gunsten einer Lingua franca geopfert wird. Schließlich sieht sie eine deutliche Theoriearmut des bilingualen Unterrichts. Diese sehen auch Lamsfuß-Schenk und Wolff (1999), wenn sie schreiben, dass didaktische Fragestellungen des bilingualen Sachfachunterrichts meist von Fremdsprachendidaktikern diskutiert werden. Für den bilingualen Sachfachunterricht wird von verschiedenen

Autoren eine eigene Didaktik gefordert. Auf diese Problematik geht Breidbach (2007) sehr intensiv ein.

Für Deutschland unterscheidet Bach (2005) drei verschiedene Modelle des bilingualen Unterrichts, bei denen das Verhältnis von Fremdsprachenunterricht und Sachfachunterricht jeweils ein anderes ist. Beim *linearen Modell* bildet der Fremdsprachenunterricht die sprachliche Grundlage. Hier lernen die Schüler die Fremdsprache, die sie später im Sachfachunterricht einsetzen können. Es gibt damit eine lineare Sequenz vom Fremdsprachenunterricht zum bilingualen Sachfachunterricht. Auch beim *parallelen Modell* bildet der Fremdsprachenunterricht die sprachliche Grundlage. Weiterhin werden aber kontinuierlich Fremdsprachenunterricht und Sachfachunterricht vernetzt, um Lernziele abzustimmen. Beim *integrativen Modell* schließlich handelt es sich um CLIL. Dabei sind Inhalte und Ziele im Fremdsprachenunterricht und im Sachfachunterricht aufeinander abgestimmt. Zudem wird die Fremdsprache im Sachfach als Arbeitssprache genutzt, ohne dass durch den Fremdsprachenunterricht zuerst eine gewisse Grundlage gelegt werden muss.

Auch Vollmer (2002, 2005a) unterscheidet drei Sichtweisen des bilingualen Unterrichts für Deutschland. Nach der ersten Sichtweise wird „bilingualer Unterricht als Erweiterung des fremdsprachlichen Lernens“ (Vollmer, 2005a, S. 53) gesehen. Über den Sachfachunterricht in der Fremdsprache soll die Fremdsprachenkompetenz erhöht werden. Diese Zielsetzung kommt nicht nur Eltern entgegen, die sich für ihre Kinder eine bessere sprachliche Ausbildung wünschen. Sie kommt auch den Fremdsprachenlehrern entgegen, den Verfechtern des bilingualen Unterrichts der ersten Stunde, die sich über den Sachfachunterricht in der Fremdsprache eine Steigerung der fremdsprachlichen Kompetenzen erhoffen. Diese Steigerung soll darüber erreicht werden, dass im Sachfachunterricht mehr als im Fremdsprachenunterricht auf Inhalte und natürliche Kommunikation fokussiert werden kann. Die Leistungen der Schüler im Sachfach interessieren aus dieser Sichtweise heraus weniger. Es interessiert vielmehr die Sprachkompetenz in der Fremdsprache.

Die zweite Sichtweise fokussiert auf bilingualen Unterricht als einen Fachunterricht in der Fremdsprache. Im Zentrum steht hier der Fachunterricht, der in einer anderen als der Muttersprache abgehalten wird. Das Lernen der Fremdsprache im Sachfachunterricht ist fachsprachlich und das Lernen im Fremdsprachenunterricht allgemeinsprachlich orientiert. Die Fremdsprache dient im Sachfachunterricht damit lediglich als Handwerkszeug, als Arbeitssprache. Hier wirft sich die Frage auf, ob sowohl im monolingualen, als auch im

bilingualen Unterricht das in dieser Sichtweise besonders geforderte vergleichbare Niveau von Fachkompetenz bei mono- und bilingualen Schülern erreicht werden kann.

Die dritte Sichtweise schließlich betrachtet bilingualen Unterricht wie das integrative Modell als ein Hand-in-Hand-Gehen von inhaltlichem und sprachlichem Lernen. Aus den verschiedenen Sichtweisen zieht Vollmer (2005a) das Fazit, dass Sach- und Sprachlernen nicht gegeneinander ausgespielt werden dürfen und von daher im bilingualen Unterricht beides gefördert werden muss. Dazu stellt Vollmer (2005b) fünf verschiedene Möglichkeiten vor, mit denen das Sprachlernen im Rahmen des fremdsprachlichen Fachunterrichts gefördert werden kann – beispielsweise die systematische Wortschatzarbeit.

Zur Praxis des bilingualen Unterrichts gibt es vielfältige Erfahrungsberichte (z.B. Spörl, 1998). Diese leisten zweifellos einen sinnvollen Beitrag zur Diskussion über Wirkung und Auswirkungen des bilingualen Unterrichts. Allerdings ist in diesem Bereich ebenso zweifellos empirische Forschung nötig. Im Folgenden sollen daher einige beispielhafte empirische Untersuchungen zum bilingualen Unterricht dargestellt werden. Trotz dass es nach Vollmer (2006b) wenig empirische Forschung zum bilingualen Unterricht gibt, kann an dieser Stelle nur auf einige Untersuchungen eingegangen werden, die aus Sicht des Autors besonders relevant sind. Bei Untersuchungen zum bilingualen Unterricht muss unterschieden werden zwischen solchen, bei denen Sprachkompetenz im Mittelpunkt steht und solchen, bei denen die Fachkompetenz im Sachfach das Zentrum des Interesses bildet. Wie in der Einleitung dieser Arbeit bereits angesprochen wurde, gibt es besonders für den Bereich der Fachkompetenz einen Bedarf an weiteren Untersuchungen. Besonders dazu soll das DFG-Projekt einen Beitrag leisten.

Eine relativ frühe empirische Untersuchung zum bilingualen Unterricht mit Fokus auf Sprachkompetenz stammt von Wode, Burmeister, Daniel, Kickler und Knust (1996). Sie berichten Zwischenergebnisse der Begleituntersuchungen zum bilingualen Unterrichtsprogramm in Schleswig-Holstein, die seit dem Schuljahr 1991/1992 im Gange sind. Die Ergebnisse beziehen sich dabei auf ein Unterrichtsmodell, bei dem konventioneller Fremdsprachenunterricht in Englisch erteilt wird. Zusätzlich wird in den Klassenstufen sieben bis zehn dann Geschichte oder Erdkunde auf Englisch unterrichtet und im jeweiligen Sachfach eine zusätzliche Wochenstunde Unterricht erteilt. Die Daten wurden für den ersten und den zweiten Schülerjahrgang erhoben, die im Modell unterrichtet worden sind ($N = 36$). Der Erhebungszeitpunkt lag etwa sieben Monate nach Beginn des bilingualen Unterrichts. Als Vergleichsgruppen dienten eine monolingual unterrichtete Klasse der gleichen Schule, von der auch die

bilinguale Klasse stammte und eine monolingual unterrichtete Klasse einer anderen Schule, die keinen bilingualen Unterricht anbot. So sollte untersucht werden, ob sich in einer Schule tendenziell die besseren Schüler für die Teilnahme am bilingualen Unterricht entscheiden.

Neben Fragebögen für Schüler und Eltern über deren sozialen Hintergrund, deren Einstellung zum bilingualen Unterricht und anderes kam ein eigens konstruiertes Testverfahren mit dem Namen „a difficult decision“ zum Einsatz. Damit sollte die schriftliche und die mündliche Fremdsprachenkompetenz erfasst werden. Zu Beginn der Testung sollen sich die Schüler dabei vorstellen, dass sie mit ihrer Klasse auf einem Ausflug in unwegsamem Gelände unterwegs sind. Während des Ausflugs bekommen drei Kinder Magenprobleme und ein Kind erleidet einen Knochenbruch. Neben der Situationsbeschreibung erhalten die am Test teilnehmenden Schüler eine einfache Landkarte der Umgebung. Anhand von vier Aufgaben sollen sie dann klären, wie den kranken und verletzten Mitschülern geholfen werden kann.

Der Test kann mündlich und schriftlich durchgeführt werden. Mit den einzelnen Aufgaben wird das Leseverständnis, das Beschreiben einer Landkarte, das Diskutieren oder Formulieren von Lösungsmöglichkeiten und das Abfassen eines nachträglichen Berichtes zu den Ereignissen überprüft. Durchgeführt wird der Test in Dreiergruppen. Zur Gruppenbildung teilt der Lehrer die Klasse in drei Leistungsgruppen ein. Die Dreiergruppen werden dann per Zufall mit je einem Schüler aus der überdurchschnittlichen, der durchschnittlichen und der unterdurchschnittlichen Leistungsgruppe gebildet. Die Leistungen der Schüler wurden von drei Englischlehrern in verschiedenen Kategorien global auf einer fünfstufigen Skala bewertet. Zusätzlich wurden *psycholinguistische Strukturanalysen* durchgeführt, um beispielsweise Lexikon und Syntax genauer zu bewerten.

Im Ergebnis zeigten die bilingualen Schüler laut Wode et al. (1996) einen größeren Wortschatz, der außerdem stärker differenziert war als bei den Schülern der beiden Vergleichsgruppen. Auf Grund der noch durchgeführten detaillierteren Analysen gehen die Autoren davon aus, dass die bilingualen Schüler im Gegensatz zu den Vergleichsgruppen nicht unterschiedliche Sprachlernfähigkeiten besitzen, sondern dass der bilinguale Unterricht den Sprachlernfähigkeiten der Schüler nur mehr Entfaltungsmöglichkeiten bietet. Zudem fanden sie keinen signifikanten Unterschied zwischen den beiden Vergleichsgruppen, was gegen Selektionseffekte spricht.

Eine weitere Untersuchung zur Sprachkompetenz stammt von Bredenbröcker (2002). Er untersuchte Schüler an fünf Schulen in Niedersachsen ($N = 195$). Ein Teil der Schüler hatte

in verschiedenen Sachfächern Unterricht in englischer Sprache, der andere Teil wurde im Sachfach auf Deutsch unterrichtet. Von Vorteil war, dass sowohl die monolingualen, als auch die bilingualen Schüler den gleichen Englischunterricht hatten, da nur der Sachfachunterricht als separater Kurs durchgeführt wurde. Durch dieses Konstanthalten der Variable „Englischunterricht“ konnte der Zugewinn an Sprachkompetenz in der Fremdsprache durch den englischen Sachfachunterricht besonders exakt erfasst werden. Die Schüler ($N = 195$) wurden zu drei Zeitpunkten untersucht: Zu Beginn des englischen Sachfachunterrichts in der siebten Klasse, zum Ende der siebten Klasse und zum Ende der achten Klasse nach zwei Jahren englischen Sachfachunterrichts. Die allgemeine Sprachkompetenz wurde mittels C-Tests erfasst. Mehr über dieses Testverfahren findet sich in Abschnitt 3.2.4. Ferner wurde das Leseverständnis über Mehrfachwahltests, Übersetzungen und das Zuordnen von textuellen und visuellen Stimuli geprüft und die grammatische Kompetenz durch verschiedene Vervollständigungs-, Umformungs- und Zuordnungsaufgaben getestet.

Bezüglich der allgemeinen Sprachkompetenz waren die bilingualen Schüler zu allen Erhebungszeitpunkten signifikant besser als die monolingualen Schüler. Zudem zeigten sie eine deutlichere Verbesserung der Sprachkompetenz mit der Zeit und eine geringere Streuung der Testergebnisse – also homogenere Leistungen – als die monolingualen Schüler. Auch beim Leseverständnis schnitten die bilingualen Schüler besser ab. Bezüglich der grammatischen Kompetenz waren die bilingualen Schüler ebenfalls besser als die monolingualen. Allerdings verbesserte sich der Mittelwert ersterer kaum. Es zeigt sich damit zwar ein Kompetenzvorsprung der bilingualen Schüler, aber nur ein geringer Kompetenzzuwachs über die Zeit. Ein solcher ist im Sachfachunterricht allerdings nicht in größerem Ausmaß zu erwarten, da Arbeit an der Grammatik hier nicht im Vordergrund steht.

Die Untersuchungen von Wode et al. (1996) und Bredenbröcker (2002) belegen, dass der bilinguale Unterricht förderlich für die Entwicklung der Sprachkompetenz in der Fremdsprache ist. Doch wie steht es mit der Fachkompetenz? Befunde dazu finden sich bei drei Dissertationsprojekten, die im Weiteren vorgestellt werden.

Lamsfuß-Schenk (2002a, 2002b) befasste sich mit bilingualem Geschichtsunterricht. Sie verglich zwei neunte Klassen eines Gymnasium in Nordrhein-Westfalen. Eine Klasse unterrichtete sie in Geschichte mit Deutsch, die andere mit Französisch als Unterrichtssprache. Die Fachkompetenz in Geschichte wurde über einen aus drei Aufgaben bestehenden Geschichtstest erfasst, der in beiden Klassen relativ am Anfang des zweiten Halbjahres nach Beginn des bilingualen Geschichtsunterrichts geschrieben wurde. Bei den Aufgaben handelte

es sich um eine Quellenanalyse, eine Reproduktionsaufgabe und um das Abgeben eines persönlichen Urteils. Die schriftlichen Antworten der Schüler wurden mittels eines aus der Leseforschung bekannten Verfahrens, der so genannten *qualitativen Propositionsanalyse*, ausgewertet. Es ergaben sich bei allen drei Aufgaben im Mittel deutlich höhere Punktzahlen für die bilingualen Schüler. Lamsfuß-Schenk (2002b) erklärt dies folgendermaßen:

Die bilingualen Schülerinnen und Schüler haben schon durch die Hürde der Fremdsprache eine realistischere Erwartungshaltung an die Quelle entwickelt. Sie erwarten fremde Worte und schwierig nachzuvollziehende Beziehungen zwischen diesen. Neben ihrem Vorwissen setzen sie vermehrt ihr Sprachwissen ein, um den Text zu analysieren und somit diese Quelle besser auszuschöpfen. Dies könnte als Beleg dafür verstanden werden, dass sie ein höheres Sprachwissen entwickelt haben und somit der Besonderheit einer historischen Quelle besser gerecht werden können: (S. 201)

Danach hängt die bessere sachfachliche Leistung der Schüler direkt mit den anderen sprachlichen Rahmenbedingungen im französischsprachigen Geschichtsunterricht zusammen. Diese Interpretation passt auch zu den Überlegungen von Lalla (2002), die davon ausgeht, dass *fachrelevante Arbeitsweisen*, so genannte *study skills*, eine wichtige Funktion beim bilingualen Lernen spielen. Die notwendige intensivere Behandlung der fachrelevanten Arbeitsweisen in der Fremdsprache im Rahmen des bilingualen Sachfachunterrichts – beispielsweise eine Anleitung zum methodisch sinnvollen Herangehen an ein historisches Bild als Quelle im Geschichtsunterricht – führt zu kleinschrittigerem Arbeiten bei bilingualen Schülern. Wo die monolingualen Schüler einfach drauflos schreiben, arbeiten die bilingualen Schüler systematischer. Im Verlauf des Prozesses vom Erlernen und Anwenden der fachrelevanten Arbeitsweisen kommen die bilingualen Schüler zu einer intensiveren, zu einer tieferen Verarbeitung von Sprache auf der einen und von Inhalt auf der anderen Seite und gleichzeitig zu einer Verknüpfung beider Komponenten.

Eine weitere empirische Untersuchung zur Fachkompetenz im Sachfach stammt von Koch (2005; Koch & Bündler, 2006). Sie untersuchte Schüler im naturwissenschaftlichen Anfangsunterricht an einem Gymnasium in Schleswig-Holstein. Dazu führte sie in zwei aufeinander folgenden Schuljahren selbst freiwillige Arbeitsgemeinschaften für Schüler der Jahrgangsstufe fünf durch. Sie unterrichtete dabei vier Gruppen, in jedem Schuljahr jeweils zwei: Eine Gruppe ohne Vorkenntnisse in Englisch, zwei Gruppen mit Schülern, die auf einer bilingualen Grundschule bereits erste Kenntnisse in Englisch erwerben konnten und wiederum eine Gruppe ohne Vorkenntnisse. Die drei an erster Stelle benannten Gruppen wurden auf Englisch und die zuletzt benannte auf Deutsch unterrichtet. Insgesamt nahmen im ersten

Halbjahr 48 Schüler an der Arbeitsgemeinschaft teil und im zweiten Halbjahr 34 Schüler. Vorab wurden die teilnehmenden Schüler und ebenso ihre nicht teilnehmenden Mitschüler mittels des Kognitiven Fähigkeitstests (vergleiche Abschnitt 3.2.3) und verschiedener Fragebögen bezüglich der kognitiven Fähigkeiten, der Freizeitaktivitäten und dem Interesse an Naturwissenschaften und Sprache untersucht.

Zur Erfassung der Fachkompetenz wurde zu zwei verschiedenen Unterrichtsthemen ein Test in Form von *concept mapping* durchgeführt. Dabei bekamen die Schüler etwa 20 Begriffe vorgegeben, die sie mittels verschiedener vorgeschlagener Relationen verbinden konnten. Diese Testung wurde sowohl vor, als auch nach Behandlung der jeweiligen Unterrichtsthemen durchgeführt. Beim ersten Unterrichtsthema nahmen 39 Schüler an der Testung teil und beim zweiten Unterrichtsthema 24 Schüler. Die Auswertung geschah über so genannte *Grafenmaße*. Dazu wurde beispielsweise die Anzahl der verwendeten Begriffe oder die Anzahl der verwendeten Relationen zwischen den Begriffen gezählt.

Bezüglich der Ergebnisse zum ersten Unterrichtsthema berichten Koch und Bündler (2006), dass sich die Leistung aller Schüler vom Prä- zum Posttest verbessert hat. Dabei zeigten die beiden Gruppen mit Vorkenntnissen in Englisch die deutlichsten Steigerungen, gefolgt von der in Englisch unterrichteten Gruppe ohne Vorkenntnisse. Deutlich geringer fielen dagegen die Steigerungen für die in Deutsch unterrichtete Gruppe aus. Hinsichtlich der kognitiven Fähigkeiten unterschieden sich die teilnehmenden Schüler im Mittel nicht von ihren Mitschülern in der Jahrgangsstufe und ebenso nicht von der Normstichprobe des Testverfahrens. Auch Interessenunterschiede fanden sich kaum. Lediglich die zeitliche Kollision mit anderen Freizeitaktivitäten scheint für manche interessierte Schüler ein Grund dafür gewesen zu sein, nicht an den Arbeitsgemeinschaften teilzunehmen.

Zusammenfassend folgern Koch und Bündler (2006), dass die Fachkompetenz im bilingualen Unterricht stärker zunimmt als im deutschsprachigen Unterricht und sich „keine Hinweise für eine Behinderung der Lernprozesse durch den Unterricht in der Fremdsprache finden“ (ebd., S. 74). Zur Erklärung des besseren Abschneidens der bilingual unterrichteten Schüler gehen sie davon aus, „dass durch die sprachlichen Hürden und die damit einhergehende Verlangsamung im fremdsprachlichen Fachunterricht ein stärkeres Bewusstsein und eine bessere Vertiefung beim Erlernen von fachlichen Begriffen erreicht werden“ (ebd., S. 68) kann.

Vollmer (2006b) berichtet vergleichbare Überlegungen. Er geht davon aus, dass beim Benutzen einer Fremdsprache als Arbeitssprache die Schüler schneller an die Grenzen

ihrer Ausdrucksfähigkeit stoßen. Dies führt nach seinen Beobachtungen allerdings nicht zum Aufgeben, sondern zu einer intensiveren Auseinandersetzung mit den Inhalten. Durch diese vertiefte semantische Analyse der Inhalte können diese dann besser erinnert und genutzt werden. Annahmen zu Effekten der Verarbeitungstiefe finden sich ursprünglich bei Craik und Lockhart (1972) und wurden für verschiedene Anwendungsbereiche bestätigt. Ein Beispiel ist das Erinnern der Figurenaufstellung beim Schachspiel (Lane & Robertson, 1979).

Alles in allem gehen angeführten Autoren davon aus, dass die gefundene größere Fachkompetenz bilingualer Schüler daher rührt, dass diese sich wegen der notwendigen intensiveren Beschäftigung mit der Fremdsprache automatisch intensiver mit den sachfachlichen Inhalten befassen. Als Erkenntnis aus den dargestellten Arbeiten bleibt insgesamt festzuhalten, dass bilinguale Schüler sowohl eine größere Sprachkompetenz, als auch eine größere Fachkompetenz als monolinguale Schüler entwickeln können.

2.7 Schulleistungsuntersuchungen

In den folgenden beiden Abschnitten sollen zwei weitere wichtige Bereiche angesprochen werden, die im engeren Sinne keine Einflussfaktoren von Schulleistung sind: Die wissenschaftliche Untersuchung von Schulleistungen und Möglichkeiten deren Erfassung. Im aktuellen Abschnitt geht es um Schulleistungsuntersuchungen. Dieses Thema wird aufgegriffen, weil die vorliegende Arbeit eine Schulleistungsuntersuchung darstellt, obgleich sie gemessen an den großen internationalen Vergleichen wie PISA, die regelmäßig in den Medien diskutiert werden, klein ausfällt. Nichtsdestoweniger muss sie in eben diesem Rahmen gesehen werden. Die groß angelegten Untersuchungen zur Erfassung von Schulleistungen müssen verschiedene Einflussfaktoren berücksichtigen, wie beispielsweise die in den vorangegangenen Abschnitten beschriebenen. Natürlich können nie alle erdenklichen Faktoren berücksichtigt werden, da Datenerhebung und Datenauswertung sonst nicht mehr praktikabel wären. Zudem gilt nach Weinert (2001a) für internationale Vergleichsuntersuchungen:

Schulleistungsvergleiche zwischen verschiedenen Ländern sind möglich und sinnvoll, wenn die Stichprobe von Schülern die gesamte Schülerpopulation korrekt repräsentiert, wenn die verwendeten Aufgaben den im Unterricht realisierten Lehrplänen gleichermaßen gerecht werden ("curriculare Validität") und wenn man die durchschnittlichen Leistungsunterschiede als Folge des gesamten Unterrichts bis zum jeweiligen Testzeitpunkt interpretiert. (S. 85)

Laut Bos und Postlethwaite (2002) begannen internationale Organisationen Mitte des 20. Jahrhunderts internationale Schulleistungsvergleiche durchzuführen. Die erste Untersuchung dieser Art war die *First International Mathematics Study*. An dieser Untersuchung beteiligten sich lediglich zwei deutsche Bundesländer. Bis in die neunziger Jahre des 20. Jahrhunderts beteiligte sich Deutschland kaum an weiteren internationalen Untersuchungen in schulischen Bereich. Dies liegt vielleicht daran: „Auch auf Grund der eher geisteswissenschaftlichen Tradition deutscher Pädagogik ist in Deutschland eine empirieorientierte erziehungswissenschaftliche und bildungspolitische Denkweise, die eher an einer Überprüfung von Sachverhalten als an philosophisch orientierten Normendebatten interessiert ist, nur schwach entwickelt“ (ebd., S. 253). Dabei liefern internationale Schulleistungsvergleiche wichtige Beiträge zur Bildungsplanung (z.B. Wirksamkeit von Schule, Verbesserung der Curricula oder des Einschulungsalter auf Grund von Daten aus internationalen Vergleichen), zur Grundlagenforschung (z.B. Entwicklung und Weiterentwicklung von Schulleistungstests und Forschungsmethoden) und zur Schulentwicklung (Rückmeldung der Ergebnisse an die Schulen als Grundlage für Verbesserungen).

Eine aktuelle internationale Untersuchung ist PISA. PISA ist für das DFG-Projekt von besonderer Bedeutung, da die Datenerhebung im DFG-Projekt in Anlehnung an die bei PISA konzipiert wurde. PISA ist ein Programm der OECD, dass im Jahr 2000 erstmalig durchgeführt worden ist und alle drei Jahre wiederholt wird. Ziel von PISA ist, den Teilnehmerstaaten eine Rückmeldung zur Leistungsfähigkeit ihrer Bildungssysteme zu geben und Entwicklungsmöglichkeiten aufzuzeigen. Dazu soll der erreichte Bildungsstand gegen Ende der Pflichtschulzeit erfasst werden. Untersucht werden Schüler im Alter von 15;3 bis 16;2 Jahren, da diese Altersgruppe in allen teilnehmenden Staaten noch der Schulpflicht unterliegt oder eine Schule besucht (Schümer, 2001).

In den teilnehmenden Ländern werden bei PISA repräsentative Zufallsstichproben von fünfzehnjährigen Schülern gezogen. Die Datenerhebung verläuft mittels speziell konstruierter Tests, die den Wissensstand von Schülern in den Bereichen Lesen, Mathematik und Naturwissenschaften abbilden sollen, wobei in jedem Erhebungsjahr ein Bereich den Schwerpunkt bildet. Daneben werden bereichsübergreifende Kompetenzen, so genannte *Cross-Curricular Competencies*, wie das Problemlösen oder der Umgang mit Computern, untersucht. Zusätzlich werden von Schülern und Eltern sowie den Schulleitungen mittels Fragebögen Daten gesammelt, die beispielsweise Aufschluss geben sollen über familiären

Hintergrund der Schüler oder Rahmenbedingungen in der Schule; also über potentielle Einflussfaktoren von Schulleistungen. Diese werden als *Hintergrundmerkmale* bezeichnet.

Die im Rahmen von PISA konstruierten Tests sollen eine Grundbildung der Schüler abbilden, die im Englischen als *literacy* bezeichnet wird. Allerdings ist *Grundbildung* oder auch *Literalität* dazu nur eine unzureichende Übersetzung. Beim ersten Begriff kann je nach Verständnis eine Art kulturspezifische Weltorientierung einfließen, ähnlich wie bei der Allgemeinbildung, die bei *literacy* keine Rolle spielt. Beim zweiten Begriff kann eine Beschränkung auf grundlegende sprachliche Kompetenzen mitschwingen. Die Grundbildung nach dem Verständnis von *literacy* soll eine Basis zur Teilhabe am gesellschaftlichen Leben bieten und lebenslanges weiteres Lernen ermöglichen. Von Saldern (1997) schreibt dazu:

Bildung als offener, lebenslanger und dynamischer Prozess wird zunehmend als Befähigung zur Bewältigung von Lebenssituationen im Spannungsfeld von Selbstbestimmung und Verantwortung für das Gemeinwesen verstanden.

Hierzu ist eine Grundbildung erforderlich, die die heranwachsende Generation befähigt, an der gesellschaftlichen Diskussion und an Entscheidungsprozessen über die Gestaltung von Gegenwart und Zukunft mitzuwirken. (S. 22)

Um zu gewährleisten, dass die Schüler nicht mit gänzlich unbekanntem Inhalten konfrontiert werden, sind die Testaufgaben bezüglich ihrer curricularen Validität für Deutschland beurteilt worden. Da die Überprüfung insgesamt positiv ausgefallen ist, kann davon ausgegangen werden, dass sich die bei PISA überprüfte Grundbildung in den deutschen Lehrplänen widerspiegelt. Eine detailliertere Einleitung zu PISA findet der interessierte Leser bei Baumert, Stanat und Demmrich (2001) oder bei Prenzel, Drechsel, Carstensen und Ramm (2004).

2.8 Schulleistungserfassung

Nachdem im vorangegangenen Abschnitt exemplarisch dargestellt worden ist, wie Schulleistungsuntersuchungen angelegt sind und welche Ziele sie verfolgen, soll im aktuellen Abschnitt darauf fokussiert werden, wie Schulleistungen eigentlich messbar sind. Dazu wird sowohl auf die Erfassung von Schulleistungen in der schulischen Praxis, als auch auf deren Messung im Rahmen von Forschung und Schulleistungsuntersuchungen eingegangen.

Ingenkamp und Lissmann (2005) unterscheiden verschiedene Möglichkeiten der Erfassung von Schulleistungen. In der Schulpraxis am weitesten verbreitet und allgemein bekannt sind davon die *mündlichen Prüfungen* und die *schriftlichen Prüfungsarbeiten*. Weiterhin berichten sie von verschiedenen *alternativen Beurteilungsverfahren*. Darunter

verstehen sie Diagnosebögen, Schülerbeobachtungsbögen, Lernentwicklungsberichte, Zeugnisberichte und Portfoliobeurteilungen. Beim letztgenannten Verfahren handelt es sich um das Anlegen einer Mappe, in der Schülerleistungen gesammelt werden und anhand derer sich jeder einen Überblick über die Leistungen des Schülers verschaffen kann. Da die alternativen Beurteilungsverfahren eher ein Schattendasein fristen, soll auf sie nicht näher eingegangen werden. Eine letzte Verfahrensgruppe bilden die *Schulleistungstests*. Besonders wegen der Praktikabilität und der guten messtheoretischen Eigenschaften ist deren Einsatz im Rahmen von Evaluationsuntersuchungen – wie der vorliegenden Arbeit – sinnvoll. Ingenkamp und Lissmann (2005) schreiben:

Schulleistungstests sind Verfahren der Pädagogischen Diagnostik, mit deren Hilfe Ergebnisse geplanter und an Curricula orientierter Lernvorgänge möglichst objektiv, zuverlässig und gültig gemessen und durch Lehrende (z.T. auch durch Lernende) oder Beratende ausgewertet, interpretiert und für pädagogisches Handeln nutzbar gemacht werden können (S. 156)

Dabei unterscheiden die Autoren zwei verschiedene Testarten: Bei den *bezugsgruppenorientierten Schulleistungstests* werden die Testergebnisse eines Schülers mit den Ergebnissen einer Bezugsgruppe verglichen. Bei der Konstruktion solcher Tests wird in einem ersten Schritt in der Planungsphase festgelegt, was das Untersuchungsziel ist. In einem nächsten Schritt arbeiten Experten aus den Lehrplänen die übergeordneten Lernziele heraus, welche die Grundlage für die Testaufgaben bilden. Die nächsten Schritte sind die Konstruktion der Testaufgaben und die Aufgabenanalyse. Schließlich wird der Test an der gewünschten Bezugsgruppe geeicht. Eine Bezugsgruppe kann dabei beispielsweise eine Klassenstufe sein.

Bei den *kriteriumsorientierten Schulleistungstests* werden die Testergebnisse eines Schülers mit vorher festgelegten Kriterien verglichen. Die Konstruktion dieser Tests ist in vielen Teilen mit der Konstruktion der bezugsgruppenorientierten Verfahren vergleichbar. Wesentliche Unterschiede ergeben sich hinsichtlich der Fundierung der Aufgaben und der Eichung. Zudem ist von entscheidender Bedeutung, dass die Aufgaben inhaltsvalide sind. Im Idealfall stellen die Testaufgaben dann eine repräsentative Stichprobe aus allen möglichen Aufgaben dar. Anhand der Lehrpläne ist es sicherlich schwer, alle möglichen Aufgaben abzuleiten. Dies wird wohl nur beim Fach Mathematik exakt möglich sein: Wenn der Lehrplan beispielsweise vorsieht, dass ein Schüler zu einem gewissen Zeitpunkt Additionsaufgaben mit natürlichen Zahlen von eins bis zehn beherrschen soll, so ist durch diese Vorgabe eine endliche Menge an Aufgaben festgelegt. Durch die Betonung der Inhaltsvalidität erübrigt sich die

Eichung dieser Schulleistungstest, da einzig der Vergleich mit den im Lehrplan festgelegten Lernzielen entscheidend ist. Klauer (2001b) bezeichnet die kriteriumsorientierten Schulleistungstests als *lehrzielorientierte Tests*. Er bringt die Krux bei der Konstruktion auf den Punkt: Aus einer definierten Aufgabenmenge muss eine repräsentative Aufgabenstichprobe gezogen werden.

Formelle Schulleistungstest, wie die bisher behandelten bezugsgruppen- und kriteriumsorientierten Verfahren, werden in der Schulpraxis nur selten eingesetzt. Hier kommen allenfalls *informelle Schulleistungstests* zum Einsatz. Dabei handelt es sich um von den Lehrern ad hoc konstruierte Verfahren zur Überprüfung der Schulleistung in einem ausgewählten Themenbereich. Allerdings weist Sacher (2005) zurecht darauf hin, dass ein flächendeckender Einsatz formeller Schulleistungstests, wie beispielsweise der bei PISA verwendeten Verfahren, als Grundlage der Beurteilung von Schulleistungen kritisch zu sehen ist. Ziel kann nicht sein, dass in der Schule nur das Abschneiden in standardisierten Schulleistungstests berücksichtigt wird. In der heutigen Praxis werden neben fachlichen Leistungen auch überfachliche Leistungen, Sozialverhalten, individueller Lernfortschritt und vieles mehr bewertet. Das alles kann mit Schulleistungstests nicht erfasst werden.

Tücke (1999) behandelt ganz allgemein verschiedene Formen der Leistungsbewertung. Bei der *normorientierten Leistungsbewertung* anhand einer *sozialen Bezugsnorm* – wie in der Praxis des deutschen Schulsystems meistens angewendet – werden aktuelle Leistungen eines Schülers mit der durchschnittlichen Leistung seiner Klasse verglichen. Dadurch sind die Noten zwischen Klassen schlecht vergleichbar und innerhalb der Klassen nicht abhängig von der Beherrschung des Lernstoffes durch die Schüler. Bei der *ipsativen Leistungsbewertung* nach einer *individuellen Bezugsnorm* werden die aktuellen Leistungen eines Schülers relativ zu seinen vorhergehenden Leistungen bewertet. Das bringt den Vorteil einer individuellen Rückmeldung für den Schüler. Zudem kann eine durch Anstrengung verbesserte Leistung angemessen belohnt werden. Praktikabel ist diese Form der Leistungsbewertung allerdings schon deshalb nicht, weil sehr gute und sehr schlechte Schüler nicht angemessen beurteilt werden können. Die so vergebenen Noten sind zudem interindividuell nicht vergleichbar. Der Königsweg ist somit die verbleibende *kriteriumsorientierte Leistungsbewertung* mittels einer *sachlichen Bezugsnorm*, die beispielsweise durch kriteriumsorientierte Schulleistungstests geleistet werden kann. Die Ergebnisse sind dann zwischen Schülern und über Klassen und Schulen hinweg vergleichbar.

In der schulischen Praxis wird die Leistung der Schüler vielfältig erfasst. Schüler müssen Tests, Klassenarbeiten und Klausuren schreiben, Referate halten und im Unterricht mitarbeiten. Je nach Fach können auch künstlerische oder musische Produkte der Schüler oder deren sportliche Fertigkeiten bewertet werden. Der Beurteiler ist in allen Fällen ein Lehrer und das Ergebnis ist immer eine Note.

Betrachten wir zunächst die rechtliche Seite der Bewertung von Schulleistungen. Da Bildung in Deutschland in den Kompetenzbereich der Bundesländer fällt, können sich die rechtlichen Grundlagen von Land zu Land unterscheiden. Zur Leistungsbewertung und zu den Notenstufen wird im Folgenden beispielhaft aus der *Allgemeinen Schulordnung des Landes Nordrhein-Westfalen* (Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen [SM-NRW], 1995) zitiert:

§ 21

Leistungsbewertung

- (1) Die Leistungsbewertung soll über den Stand des Lernprozesses des Schülers Aufschluß geben; sie soll auch Grundlage für die weitere Förderung des Schülers sein. Bei der Beratung über den Bildungsgang des Schülers durch die Schule soll sie eine wesentliche Hilfe sein.
- (2) Die Leistungsbewertung bezieht sich auf die im Unterricht vermittelten Kenntnisse, Fähigkeiten und Fertigkeiten.
- (3) Bei der Bewertung von Schülerleistungen ist der Eigenart der Schulstufe, der Schulform und des Unterrichtsfachs Rechnung zu tragen. Es werden der Umfang sowie die selbständige und richtige Anwendung der Kenntnisse, Fähigkeiten und Fertigkeiten sowie die Art der Darstellung bewertet.
- (4) Grundlage der Leistungsbewertung sind alle vom Schüler im Zusammenhang mit dem Unterricht erbrachten Leistungen, insbesondere schriftliche Arbeiten, mündliche Beiträge und praktische Leistungen. Die Leistungen bei der Mitarbeit im Unterricht sind für die Beurteilung eines Schülers ebenso zu berücksichtigen wie die übrigen Leistungen.
- (5) (SM-NRW, 1995, S. 25)

Im ersten Absatz des Paragraphen 21 sind die Funktionen der Bewertung von Schulleistungen aus Sicht des Gesetzgebers aufgeführt: Rückmeldung über den Stand des Lernprozesses und Grundlage für die weitere Förderung und Beratung des Schülers. Dabei bezieht sich die Leistungsbewertung nach Absatz 2 auf im Unterricht Vermitteltes, wobei nach Absatz 3 auch Rahmenbedingungen wie Schulform berücksichtigt werden müssen. Wie das genau geschehen soll, wird letztlich nicht klar angegeben. In Absatz 4 werden verschiedene konkrete Verfahren der Leistungsbewertung aufgeführt, wie mündliche und schriftliche Äußerungen der Schüler. Formelle Schulleistungstest sind an dieser Stelle nicht vorgesehen. Die Ergebnisse der Bewertung von Schulleistungen werden in Form von Noten dargestellt:

§ 25
Notenstufen

(1) Bei der Bewertung einzelner Schülerleistungen sowie in Zeugnissen werden die folgenden Notenstufen zugrunde gelegt:

1. sehr gut (1)

Die Note „sehr gut“ soll erteilt werden, wenn die Leistung den Anforderungen in besonderem Maße entspricht.

2. gut (2)

Die Note „gut“ soll erteilt werden, wenn die Leistung den Anforderungen voll entspricht.

3. befriedigend (3)

Die Note „befriedigend“ soll erteilt werden, wenn die Leistung im allgemeinen den Anforderungen entspricht.

4. ausreichend (4)

Die Note „ausreichend“ soll erteilt werden, wenn die Leistung zwar Mängel aufweist, aber im ganzen den Anforderungen noch entspricht.

5. mangelhaft (5)

Die Note „mangelhaft“ soll erteilt werden, wenn die Leistung den Anforderungen nicht entspricht, jedoch erkennen läßt, daß die notwendigen Grundkenntnisse vorhanden sind und die Mängel in absehbarer Zeit behoben werden können.

6. ungenügend (6)

Die Note „ungenügend“ soll erteilt werden, wenn die Leistung den Anforderungen nicht entspricht und selbst die Grundkenntnisse so lückenhaft sind, daß die Mängel in absehbarer Zeit nicht behoben werden können.

(2) Neben oder anstelle der Noten nach Absatz 1 kann nach Maßgabe der Ausbildungs- und Prüfungsordnung auch ein Punktsystem verwendet werden. Noten- und Punktsystem müssen untereinander übertragbar sein.

(3) Die Ausbildungs- und Prüfungsordnung kann für die Klassen 1 und 2 der Grundschule und für Sonderschulen anstelle der Noten schriftliche Aussagen über die Leistungsbewertung vorsehen. (SM-NRW, 1995, S. 27)

Die aufgeführten Notenstufen sind in allen Bundesländern in gleicher Form definiert (Avenarius, 2001). In Absatz 1 des Paragraphen 25 werden die sechs Notenstufen unter anderem mit Ziffern bezeichnet. Aus den Beschreibungen, wann eine bestimmte Note erteilt werden soll, wird eine Rangfolge der Notenstufen ersichtlich. Beginnend bei der Bestnote *sehr gut* ist die mit den folgenden Notenstufen bewertete Schulleistung immer etwas schlechter, bis zur letzten Notenstufe *ungenügend*. Neben der offensichtlichen Rangfolge ist allerdings nicht klar, wie groß die Unterschiede zwischen den einzelnen Notenstufen sind. Ist der Schritt von einer *vollen Anforderungserfüllung* zu einer *Anforderungserfüllung im Allgemeinen* genauso groß, wie der Schritt von einer *Anforderungserfüllung im Allgemeinen* zur *Anforderungserfüllung im Ganzen trotz Mängeln*? Da einerseits die Gleichabständigkeit der einzelnen Notenstufen nicht sichergestellt ist und andererseits klar eine Rangfolge vorliegt, sollte die Schulnotenskala als ordinalskaliert² betrachtet werden.

In Absatz 2 des Paragraphen 25 wird auf die Möglichkeit hingewiesen, ein Punktsystem zu verwenden. Dabei wird – im Regelfall in der Sekundarstufe II – die sechsstufige

² Mehr über die verschiedenen Skalenniveaus findet der interessierte Leser beispielsweise bei Borg und Staufenbiel (2007), Bortz (2005), Bortz und Döring (2006), sowie bei Diehl und Kohr (1977).

Schulnotenskala in eine 15 Stufen umfassende Skala überführt, wobei für alle Notenstufen außer dem *ungenügend* neben der vollen Note noch jeweils die Abstufung *plus* und *minus* berücksichtigt wird. Auch bei einer solchen Skala kann von Ordinalskalenniveau ausgegangen werden. Absatz 3 bezieht sich auf Verbalbeurteilungen, wie sie in Grund- und Sonderschulen möglich sind. Auf Vor- und Nachteile dieser Art von Beurteilung gehen beispielsweise Tarnai (2001) und Wedel (2006) ein.

Laut Avenarius (2001) hat ein Lehrer nach Sicht des Bundesverfassungsgerichtes einen gewissen *Bewertungsspielraum*, da er Noten in einem Bezugssystem von persönlichen Erfahrungen und Vorstellungen vergibt. Allerdings hat auch ein Schüler einen gewissen *Antwortspielraum*, wonach eine angemessen begründete und nachvollziehbare Lösung nicht als falsch bewertet werden darf, nur weil sie von der Sicht des Lehrers abweicht.

Aus dieser Darstellung der rechtlichen Seite der Bewertung von Schulleistungen sollte klar geworden sein, dass die offiziellen Definitionen zur Leistungsbewertung zwar einen fundierten Bewertungsrahmen suggerieren, ein solcher aber keineswegs vorhanden ist. Das beginnt schon damit, dass den Noten neben den bisher explizit genannten Funktionen weitaus mehr abverlangt wird. Dazu berichtet Tent (2001) von Funktionen, die Noten erfüllen sollen: Zum einen gibt es die *pädagogischen Funktionen*. Noten sollen Schüler zu Leistung motivieren oder einfach disziplinieren (Anreizfunktion), Rückmeldung sein und mit der kompetitiven Leistungsbewertung vertraut machen (Sozialisationsfunktion). Zum anderen haben Noten *gesellschaftliche Funktionen*. Aus Sicht der Schüler eröffnen oder verschließen sie den Zugang zu zukünftigen Entwicklungspfaden (Weichenstellfunktion). Weiterhin belegen sie einen bestimmten Bildungsgrad (Berechtigungsfunktion) und ermöglichen den Zugang zu Arbeitsplätzen (Klassifikations-, Allokations-, Selektionsfunktion). Schließlich dienen sie der Überwachung der Schulpflicht und schulischer Maßnahmen (Kontrollfunktion). An dieser Stelle fallen sofort Schwierigkeiten auf. Wie kann ein und dieselbe Note laut Allgemeiner Schulordnung (SM-NRW, 1995) über den Stand des Lernprozesses bei einem Schüler Aufschluss geben und diesem gleichzeitig als Disziplinarmaßnahme für das Stören im Unterricht einen Denkkettel verpassen? Oder anders: Einer einzelnen Note kann nicht angesehen werden, in welcher Funktion sie gerade verwendet wird.

Ein weiterer wichtiger Punkt ist, dass in der Praxis oft das arithmetische Mittel von Noten gebildet wird, welches ebenso in Prüfungsordnungen Verwendung findet. Die errechneten Durchschnittsnoten sagen allerdings streng genommen nichts aus, da die Berechnung eines Mittelwertes über verschiedene Rangzuweisungen zwar eine Zahl ergibt, diese aber

keine Rangzuweisung mehr darstellt. Auch in der Forschung werden Noten in Verfahren der Datenauswertung verwendet, die Intervallskalenniveau fordern. Die Ergebnisse dieser Datenauswertungen sagen wegen der Verletzung der Anforderungen an das Skalenniveau allerdings ebenso wenig aus, wie die Durchschnittsnoten. Obwohl manche Autoren wie Bortz und Döring (2006) annehmen, „dass die Bestätigung einer Forschungshypothese durch die Annahme eines falschen Skalenniveaus eher erschwert wird“ (ebd., S. 70), sollen in der vorliegenden Arbeit Noten streng als ordinalskalierte Daten betrachtet werden. Näheres zur Skalierung von Noten findet sich bei Lienert (1987).

In der Schulpraxis zeigen sich häufig hohe Korrelationen von Noten und Testwerten innerhalb von Klassen, aber niedrige zwischen verschiedenen Klassen (Tent, 2001). Dies ist ein Indiz dafür, dass Lehrer Schulleistungen nicht nach den Anforderungen des Lehrplan bewerten und damit statt einer sachlichen Bezugsnorm eine soziale Bezugsnorm, nämlich die Klasse, zu Grunde legen. Wenn Lehrer so verfahren sind Noten aber über verschiedenen Klassen hinweg nicht vergleichbar. Der Numerus clausus bei manchen Studienfächern macht damit keinen Sinn, da die Noten weniger über die Schulleistungen als über die Klasse des Schülers aussagen. Über diese und weitere Unzulänglichkeiten der Benotungspraxis berichtet Ingenkamp (1976). Sacher (2005) führt an, dass die hier angerissene Problematik der unterschiedlichen Benotung gleicher Schülerleistungen in der Fachwelt schon seit Jahrzehnten bekannt ist. Weitere Kritik an Noten findet sich bei Winter (2005) und Vierlinger (2005). Letzterer kritisiert schon die Vorgabe der Notenskala durch die Bürokratie und sieht darin bereits die Grundlage zur Unzulänglichkeit der Noten hinsichtlich der Leistungsbewertung.

Die Qualität von Noten lässt sich auch anhand der Testgütekriterien beurteilen. Objektivität kann für Noten in der derzeitigen schulischen Praxis nicht angenommen werden, da sie immer einem subjektiven Lehrerurteil entstammen. Steigern kann man die Objektivität von Noten, wenn beispielsweise verschiedene Lehrer die gleiche Schulleistung bewerten. Dies ist ansatzweise bei Zeugniskonferenzen der Fall. Zu den beiden anderen Gütekriterien finden sich Befunde bei Tent (2001). Bezüglich der Reliabilität berichtet er unter anderem von hoher Testwiederholungsreliabilität. Diese liegt in der Grundschule bei $r_{tt} > .8$ mit fallender Tendenz über die weitere Schullaufbahn. Für die gesamte Gymnasialzeit finden sich Werte von $r_{tt} \approx .5$ in Fremdsprachen und Sport. Zur Kriteriumsvalidität berichtet er unter anderem Korrelationen von Noten und Intelligenz um $r = .5$ und von Noten und Schulleistungstests um $r = .65$. Dies sind substantielle Zusammenhänge.

Interessante Ergebnisse zur Vorhersagevalidität von Noten finden sich in der Diskussion um die besten Prädiktoren zur Prognose von Studienerfolg. So zeigen zum Beispiel Gold und Souvignier (2005) in ihrer prospektiven Längsschnittstudie, dass sich der Studienerfolg in verschiedenen Fächern gut durch die Abiturdurchschnittsnote vorhersagen lässt. Ein besserer Prädiktor ist nur die Note der Zwischenprüfung im Studium. Allerdings sprachen diese Ergebnisse nur zum Teil für die Validität, da es sich um Korrelationen zwischen Noten handelt, denen sowohl im Bereich der Schule, als auch im Bereich der Hochschule, vergleichbare methodische Probleme anhaften. Damit ist ferner noch nichts über die Vorhersagefähigkeit bezüglich des Berufserfolgs gesagt, für den die Bildung schließlich die Grundlage bilden soll.

In der vorliegenden Arbeit sollen zur Erfassung von Schulleistungen hauptsächlich Noten verschiedener Fächer verwendet werden. Weiterhin kommt ein formeller Schulleistungstest zum Einsatz, der in Abschnitt 3.2.1 vorgestellt wird. Zudem wird sprachliche Kompetenz über spezielle sprachliche Kompetenztests erfasst, auf die in Abschnitt 3.2.4 genauer eingegangen wird.

2.9 Fragestellungen und Hypothesen

Nachdem in den vorangegangenen Abschnitten der theoretische und empirische Hintergrund der vorliegenden Arbeit dargestellt worden ist, sollen nun die relevanten Fragestellungen vorgestellt und die Hypothesen abgeleitet werden.

In erster Linie interessiert das Ausmaß der von den Schülern erworbenen Sachfachkompetenz in Erdkunde, deren Geografiekompetenz. Dies war auch das Hauptuntersuchungsanliegen im DFG-Projekt (Vollmer, 2006a, 2006b). Bezüglich der vermittelten Inhalte und bezüglich der Leistungsbewertung sind zwischen englischsprachigem und deutschsprachigem Erdkundeunterricht seitens des Schulrechts keine Unterschiede zu erwarten (Lenz, 2003). Auf Grund der in Abschnitt 2.6 berichteten Forschungsergebnisse zum bilingualen Unterricht kann angenommen werden, dass durch die sprachliche Herausforderung im englischsprachigen Erdkundeunterricht die Geografiekompetenz profitiert – unter anderem durch die intensivere Beschäftigung mit Materialien. Im Rahmen ihrer Dissertationen konnten Lamsfuß-Schenk (2002a, 2002b) und Koch (2005; Koch & Bündler, 2006) für den Unterricht in Geschichte und Naturwissenschaften Kompetenzvorteile der bilingualen Schüler zeigen. Diese Forschungsbefunde bilden die Grundlage für die erste Hypothese der vorliegenden Arbeit:

Hypothese 1

Die bilingualen Schüler haben im Vergleich zu den monolingualen Schülern eine größere Geografiekompetenz.

Im Zusammenhang mit der Fachkompetenz untersuchte Koch (2005) auch die kognitiven Fähigkeiten der Schüler und fand, dass sich die Schüler in den bilingualen Arbeitsgemeinschaften nicht von den Übrigen der Jahrgangsstufe unterscheiden. Vollmer (2006b) betont, dass die Leistungen mono- und bilingualer Schüler nur sinnvoll verglichen werden können, wenn sie sich hinsichtlich der kognitiven Fähigkeiten nicht zu sehr unterscheiden. Das führt zur zweiten Hypothese:

Hypothese 2

Die mono- und bilingualen Schüler unterscheiden sich nicht signifikant bezüglich ihrer kognitiven Fähigkeiten.

Die Untersuchungsergebnisse bei PISA belegen deutliche Effekte der sozialen Herkunft der Schüler auf die Schulleistungen (Baumert & Schümer, 2001). Damit sind auch Auswirkungen der sozialen Herkunft auf die Geografiekompetenz zu erwarten. Vollmer (2006b) weist neben der Voraussetzung vergleichbarer kognitiver Fähigkeiten bei mono- und bilingualen Schülern darauf hin, dass beide Gruppen auch bezüglich ihrer sozialen Herkunft vergleichbar sein sollten. Wenn dies gewährleistet ist, können Gruppenunterschiede nicht auf die soziale Herkunft zurückgeführt werden. Bei den in Abschnitt 2.6 genannten Untersuchungen zum bilingualen Unterricht wurde die soziale Herkunft der Schüler nicht kontrolliert; in der vorliegenden Arbeit soll dies geschehen:

Hypothese 3

Die mono- und bilingualen Schüler unterscheiden sich nicht signifikant bezüglich ihrer sozialen Herkunft.

In Untersuchungen wie denen von Wode et al. (1996) und Bredenbröcker (2002) konnte gezeigt werden, dass der bilingualer Unterricht förderlich für die Entwicklung der Sprachkompetenz in der Fremdsprache ist. Bilingual unterrichtete Schüler zeigen für diese Sprache eine größere Sprachkompetenz, als ihre jeweiligen monolingual unterrichteten Mitschüler. Die größere Sprachkompetenz entwickelt sich dabei nicht zuletzt durch die regelmäßige und

zeitlich längere Auseinandersetzung mit der Fremdsprache (Vollmer, 2005b). Daraus ergibt sich die folgende Hypothese:

Hypothese 4

Die bilingualen Schüler haben eine größere Sprachkompetenz in Englisch als die monolingualen Schüler.

Für die deutsche Sprache, die Muttersprache des Großteils der Schüler in der Stichprobe ist, werden allerdings keine Unterschiede zwischen den mono- und bilingualen Schülern vermutet. Durch den bilingualen Unterricht beschäftigen sich die Schüler vom Zeitaufwand her zwar intensiver mit der Fremdsprache, allerdings bleibt der deutschsprachige Alltagskontext für die deutschen Muttersprachler erhalten.

Hypothese 5

Die mono- und bilingualen Schüler unterscheiden sich nicht signifikant bezüglich ihrer Sprachkompetenz in Deutsch.

Die bisher benannten fünf Hypothesen beziehen sich auf die Kernfragestellungen der vorliegenden Arbeit. Die nun folgenden drei Hypothesen betreffen die Zusammenhänge zwischen der Geografiekompetenz und der Sprachkompetenz der Schüler in Deutsch und Englisch auf der einen, und den entsprechenden Fachnoten auf der anderen Seite. Der Zusammenhang zwischen Geografiekompetenz und Erdkundenote interessiert, da ein hoher Zusammenhang – unter der Annahme, dass die Erdkundenote die Fachkompetenz der Schüler in Erdkunde zumindest zu einem Teil wiedergibt – für die Validität des Geografietests spricht:

Hypothese 6

Die Schulleistungen in Erdkunde hängen positiv mit der Geografiekompetenz zusammen.

Da schon Klein-Braley und Raatz (1982) hohe Zusammenhänge zwischen der Sprachkompetenz und den Schulleistungen im sprachlichen Bereich gefunden haben, sollen diese Zusammenhänge hier auch untersucht werden:

Hypothese 7

Die Schulleistungen in Englisch hängen positiv mit der Sprachkompetenz in Englisch zusammen.

Hypothese 8

Die Schulleistungen in Deutsch hängen positiv mit der Sprachkompetenz in Deutsch zusammen.

In den nun noch folgenden Hypothesen werden weitere aus Sicht des Autors interessante Fragestellungen behandelt, darunter die Auswirkungen kognitiver Fähigkeiten und sozialer Herkunft, Auswirkungen des Klassenklimas und Geschlechtseffekte. Die kognitiven Fähigkeiten der Schüler fanden bereits Eingang in Hypothese 2. Dabei war der Hintergrund, dass sich die Schüler möglichst nicht hinsichtlich der kognitiven Fähigkeiten unterscheiden sollten, um ihre Vergleichbarkeit sicherzustellen. Das kognitive Fähigkeiten davon abgesehen positiv mit Schulleistungen zusammenhängen, zeigen Untersuchungen wie die von Beer, Kutalek und Schnell (1968), Kolem (1981) und Kühn (1983), von denen in Abschnitt 2.2 berichtet worden ist. Zur Untersuchung dieser Zusammenhänge in der vorliegenden Arbeit soll die folgende Hypothese dienen:

Hypothese 9

Die kognitiven Fähigkeiten hängen positiv mit den Schulleistungen in Deutsch (Teilhypothese A), Englisch (Teilhypothese B), Mathematik (Teilhypothese C) und Erdkunde (Teilhypothese D) zusammen.

Die soziale Herkunft der Schüler wurde schon im Zusammenhang mit Hypothese 3 angesprochen. Ebenso wie die kognitiven Fähigkeiten in Hypothese 2, soll die soziale Herkunft zwischen den untersuchten mono- und bilingualen Schülern möglichst vergleichbar sein. Andererseits sprechen Befunde wie die von Baumert und Schümer (2001) oder Ehmke, Hohensee, Heidemeier und Prenzel (2004) für starke Auswirkungen der sozialen Herkunft auf die Schulleistungen. Daher ist es von Interesse, die Bedeutung der sozialen Herkunft anhand der Stichprobe der vorliegenden Arbeit zu prüfen. Hier sollen allerdings nicht Schulleistungen, sondern stattdessen Geografiekompetenz, Sprachkompetenz und kognitive Fähigkeiten einbezogen werden, da vom Autor auch für diese Bereiche Auswirkungen der sozialen Herkunft erwartet werden:

Hypothese 10

Bei positiverer Beurteilung der sozialer Herkunft der Schüler zeigen sich eine größere Geografiekompetenz (Teilhypothese A), eine größere Sprachkompetenz in Deutsch (Teilhypothese B), eine größere Sprachkompetenz in Englisch (Teilhypothese C) und größere kognitive Fähigkeiten (Teilhypothese D).

Wie bereits in Abschnitt 2.5 dargelegt worden ist, stellt das Klassenklima eine weitere interessante Variable dar, besonders da es nach Helmke und Weinert (1997) einen nicht zu vernachlässigenden Beitrag zur Erklärung von Schulleistungen liefern kann. Daher soll geprüft werden, inwiefern sich Zusammenhänge zwischen dem hier erfassten Klassenklima und Schulleistungen sowie Geografiekompetenz finden. Dabei wird wie bei Bültner und Meyer (2004) davon ausgegangen, dass eine positive Einschätzung des Klassenklimas mit positiven Auswirkungen auf die Schulleistungen einhergeht:

Hypothese 11

Das Klassenklima hängt positiv mit den Schulleistungen in Deutsch (Teilhypothese A), Englisch (Teilhypothese B), Mathematik (Teilhypothese C) und Erdkunde (Teilhypothese D), sowie der Geografiekompetenz (Teilhypothese E) zusammen.

Die letzten beiden jetzt noch folgenden Hypothesen betreffen Geschlechtsunterschiede zwischen Schülerinnen und Schülern. Wie bereits in Abschnitt 2.3 berichtet wurde, finden sich für Schülerinnen meist bessere Leistungen im sprachlichen Bereich. Demgegenüber zeigen Schüler Stärken im mathematisch-naturwissenschaftlichen Bereich. Dies zeigte sich im Wesentlichen auch bei PISA (Stanat & Kunter, 2001). Die folgende Hypothese dient der Untersuchung der Geschlechtsunterschiede hinsichtlich der Sprachkompetenz:

Hypothese 12

Schülerinnen zeigen im Vergleich zu Schülern eine größere Sprachkompetenz in Deutsch (Teilhypothese A) und in Englisch (Teilhypothese B).

Für die Geografiekompetenz ist vorab nicht klar, ob Schülerinnen oder Schüler bessere Leistungen zeigen könnten. Dem Autor sind keine Untersuchungen bekannt, die sich mit dieser Fragestellung befassen haben. Daher wird hier angenommen, dass es in diesem Bereich keine

systematischen Unterschiede gibt. Für die kognitiven Fähigkeiten werden ebenfalls keine systematischen Unterschiede erwartet, da bei der Normierung des eingesetzten Verfahrens zur Erfassung der kognitiven Fähigkeiten keine Geschlechtsunterschiede festgestellt worden sind (vgl. Abschnitt 3.2.3). So ergibt sich die letzte Hypothese:

Hypothese 13

Schülerinnen und Schüler zeigen keine signifikanten Unterschiede bezüglich ihrer Geografiekompetenz (Teilhypothese A) und ihrer kognitiven Fähigkeiten (Teilhypothese B).

3 Methode

Dieses Kapitel trägt die Bezeichnung „Methode“. Damit ist das methodische Vorgehen bei der Datenerhebung gemeint. Im ersten Abschnitt wird dazu die vorliegende Arbeit methodisch eingeordnet und es werden die verwendeten Operationalisierungen vorgestellt. Im Anschluss daran wird im zweiten Abschnitt das bei der Datenerhebung verwendete Material – wie Fragebögen und Testverfahren – dargestellt. Danach wird im dritten Abschnitt die Stichprobe erläutert und in den beiden letzten Abschnitten über das Vorgehen bei der Datenerhebung und der anschließenden Datenauswertung berichtet.

3.1 Untersuchungsplanung

Methodisch betrachtet stellt das DFG-Projekt den Versuch einer Evaluation im Bereich des bilingualen Unterrichts dar. Genauer gesagt sollen im Kern die beiden verschiedenen Schülergruppen der monolingualen und bilingualen Schüler, die in Teilen unterschiedlichen Unterricht erhalten, hinsichtlich ihrer Geografiekompetenz und ihrer Sprachkompetenz in Deutsch und Englisch verglichen werden.

Wottawa (2006) nennt verschiedene Evaluationsmodelle, die je nach Zielsetzung der Evaluation Verwendung finden. Danach lässt sich das DFG-Projekt als *äußere Evaluation* – Wottawa (2001) spricht von *Fremdevaluation* – klassifizieren. Bei dieser Form der Evaluation wird eine Maßnahme nicht von deren Entwicklern oder den Ausführenden selbst, sondern von Außenstehenden durchgeführt. Zudem handelt es sich um eine *summative Evaluation*, welche nach Abschluss einer Maßnahme erfolgt, um diese zu bewerten. Auch wenn für die untersuchten Schüler der bilinguale oder monolinguale Unterricht nicht mit der Datenerhebung für das DFG-Projekt beendet ist, wird die Geografie- und Sprachkompetenz der Schüler erfasst, ohne sie noch gezielt durch Unterrichts Anpassungen verbessern zu wollen. Bei einer *formativen Evaluation*, welche das Gegenstück zur summativen Evaluation bildet, würde man versuchen, den bilingualen Unterricht auf Grundlage der Evaluationsergebnisse aus fortlaufenden Erhebungen zu optimieren. Dazu sollen die Ergebnisse aus dem DFG-Projekt zwar auch beitragen, dies ist allerdings ein langfristiges Ziel.

Zur Bildung der Untersuchungsgruppen konnte im Rahmen des DFG-Projektes keine randomisierte Gruppenzuteilung erfolgen, da die mono- und bilingualen Klassen an den verschiedenen Schulen vorhanden waren. Zudem war es nicht möglich, den Unterricht in den

Klassen systematisch zu variieren. Der monolinguale und der bilinguale Unterricht können gleichwohl als zwei verschiedene Behandlungen und als Ausprägungen der unabhängigen Variablen „Unterrichtstyp“ angesehen werden. Auf Kontrollgruppen wurde verzichtet, weil in erster Linie die Geografie- und Sprachkompetenz in den beiden Unterrichtstypgruppen verglichen werden sollte. Durch das Fehlen der Randomisierung und der Kontrollgruppe handelt es sich bei der vorliegenden Untersuchung nicht um ein echtes Experiment (McGuigan, 1997), sondern um ein quasi-experimentelles Vorgehen. Ferner wurde die Untersuchung nicht im Labor, sondern im Feld durchgeführt.

Um die am Ende des letzten Kapitels erarbeiteten Hypothesen untersuchen zu können, müssen die dort genannten Konstrukte operationalisiert werden. In Tabelle 4 werden die zu jedem Konstrukt gehörenden Operationalisierungen aufgeführt. Neben den dort genannten Abschnitten findet sich auch in Kapitel 2 der vorliegenden Arbeit relevanter theoretischer Hintergrund zu den einzelnen Konstrukten.

Tabelle 4

In der linken Spalte sind die in den Hypothesen (vgl. Abschnitt 2.9) verwendeten Konstrukte aufgeführt. In der mittleren Spalte sind die verwendeten Operationalisierungen genannt. In der letzten Spalte wird auf den Abschnitt verwiesen, in dem die Operationalisierung genauer behandelt wird.

Konstrukt	Operationalisierung	Abschnitt
Schulleistung	Noten	2.5
Geografiekompetenz	Geografietest	3.2.1
Soziale Herkunft	Fragen im Schülerfragebogen	3.2.2
Klassenklima	Skala im Schülerfragebogen	3.2.2
Kognitive Fähigkeiten	Kognitiver Fähigkeitstest	3.2.3
Sprachkompetenz Deutsch	C-Test Deutsch	3.2.4
Sprachkompetenz Englisch	C-Test Englisch	3.2.4

3.2 Material

In diesem Abschnitt sollen die Materialien vorgestellt werden, die bei der Datenerhebung zum Einsatz kamen. Das sind erstens der Geografietest, zweitens der Schülerfragebogen, drittes der Kognitive Fähigkeitstest und schließlich noch der C-Test.

3.2.1 Geografietest

Der Geografietest dient zur Erfassung der wichtigste abhängigen Variable, der Geografiekompetenz. Unter „Geografiekompetenz“ kann die Gesamtheit der Fähigkeiten, Fertigkeiten und Kompetenzen der Schüler im Fach Erdkunde verstanden werden. Vollmer (2006b) berichtet genauer von der Konstruktion des Geografietests. In einem erste Schritt wurde ein Modell der geografischen Kompetenz entwickelt, da ein solches noch nicht vorlag. Daran beteiligten sich Fachleute für Geografie aus Universitäten und Erdkundelehrer. In diesem Kompetenzmodell werden letztlich sieben Kompetenzbereiche unterschieden, die im Folgenden kurz vorgestellt werden.

Beim *Fachwissen* handelt es sich um das Basiswissen der Geografie. Es beinhaltet Modelle und Konzepte sowie die Fähigkeit, vorhandenes Wissen durch Verknüpfen mit neuem Wissen auszubauen und zu reorganisieren. *Methodenkompetenz* bezieht sich auf die Anwendung von Methoden aus der Geografie. Dazu gehört der Umgang mit verschiedenen Quellen und Materialien wie Karten, Klimadiagrammen, Statistiken und Bildern sowie die Lösung von Problemen. Bei der *Fachkommunikation* geht es um die „diskursive Verwendung der Sprache in fachlichen Kontexten“ (Vollmer, 2006b, S. 11). Damit ist ein angemessener fachsprachlicher Ausdruck gemeint, der es erlaubt, an der Diskussion im Fach und an Erkenntnissen des Faches teilzuhaben. Bei der *Beurteilungskompetenz* geht es um die Anwendung von grundlegenden Maßstäben der Geografie auf Sachverhalte. Sie ist abzugrenzen von der *Bewertungskompetenz*, bei welcher neben der Beurteilung auch die persönlichen Maßstäbe des Bewertenden, wie Werte und Normen, einfließen sollen. Schließlich wurden noch die Kompetenzbereiche *räumliche Orientierung* und *Handeln* definiert. Ersterer betrifft die Fähigkeit, sich im Raum orientieren und letzterer die Fähigkeit, im wirklichen Leben handeln zu können. Diese beiden Kompetenzbereiche spielen im DFG-Projekt keine Rolle, da die räumliche Orientierung als Teil des Fachwissens gesehen werden kann und das Handeln nur schwer operationalisierbar ist.

Es zeigte sich allerdings schnell, dass sich die genannten Kompetenzbereiche nicht trennscharf abbilden lassen: Die ersten drei sind bei der Bearbeitung fast jeder Aufgabe notwendig. Daher postuliert Vollmer (2006b) unterhalb der benannten Kompetenzbereiche, die er als eine Kompetenzschicht ansieht, eine weitere Kompetenzschicht. Diese bezieht sich auf *Diskursfunktionen*. Bei der Definition von Diskursfunktionen geht Zydatiś (2005) davon aus, dass eine Person bei der Aufgabebearbeitung in Abhängigkeit von der Aufgabenstellung und dem zur Lösung vorliegenden textlichen Material, auf Inhalte, auf Sprache und auf Denken

zurückgreifen muss. Die drei Aspekte Inhalte, Sprache und Denken lassen sich modellhaft als Kreise in einer Ebene darstellen, die jeweils Überschneidungsbereiche besitzen. In diesem Modell finden sich im Überschneidungsbereich von Inhalt und Sprache die fachspezifischen Begriffe und Konzepte, im Überschneidungsbereich von Inhalt und Denken die fachspezifischen Arbeitsmethoden und -techniken und im Überschneidungsbereich von Sprache und Denken die Wissensstrukturen der Person. Die Diskursfunktionen stellen nun jeweils eine Schnittmenge aus Inhalt, Sprache und Denken dar. Die einzelnen Teilaufgaben im Geografietest lassen sich anhand ihrer dominanten Diskursfunktion gut klassifizieren. Eine Übersicht dazu gibt Tabelle 5.

Tabelle 5

Übersicht über die Diskursfunktionen und ihre Zuordnung zu den einzelnen Teilaufgaben im Geografietest.

	Diskursfunktion		Teilaufgaben
	deutsche Bezeichnung	englische Bezeichnung	
Begründen/Erklären		give reasons/explain	1d, 1e, 2b
Beschreiben		describe	1a, 1b, 1c [†] , 2a [†]
Beurteilen		assess	3 [†] , 6 [†]
Bewerten		evaluate	7 [†] , 8 [†]
Interpretieren		interpret/summarise/analyse	5a, 5b, 5c
Kategorisieren/Einordnen		categorise/classify	4a, 4b, 4c

[†] *Bemerkung:* Die so gekennzeichneten Teilaufgaben wurden in der Datenauswertung verwendet. Bei den anderen wurde auf Grund der Aufgabenstellung oder wegen Zeitmangels am Ende der Bearbeitung des Geografietests nicht genügend Text produziert, um eine sinnvolle Auswertung zu ermöglichen.

Die Diskursfunktionen *Bewerten* und *Beurteilen* sind dabei analog zu den oben genannten Kompetenzbereichen zu verstehen. Damit verbleiben auf Seiten des Kompetenzmodells die drei Basisdimensionen der Geografiekompetenz, nämlich Fachwissen, Methodenkompetenz und Fachkommunikation. Die sechs relevanten Diskursfunktionen verteilen sich über alle Teilaufgaben. Jede Diskursfunktion wird bei mindestens zwei Teilaufgaben verwendet.

Bei der Konstruktion des Geografietests wurden ebenfalls Geografieexperten eingesetzt und die endgültige Form nach mehreren Pilotierungsdurchgängen erstellt. Dabei wurde angestrebt, einen allgemeingültigen Test im Sinne eines kriteriumsorientierten Schulleistungstests vorzulegen, wobei nicht Lernziele der aktuellen Jahrgangsstufe berücksichtigt werden sollten. Vielmehr sollte laut Vollmer (2006b) die grundlegende Geografiekompetenz erfasst

werden. Die beteiligten Geografieexperten hielten die verwendeten Aufgaben im Geografietest für relevant und die Aufgabenauswahl für repräsentativ.

Thematisch befassen sich die Aufgaben mit dem tropischen Regenwald, der laut Lehrplan sowohl in der Mittelstufe, als auch in der Oberstufe behandelt wird. Bei den untersuchten Schülern lag die Auseinandersetzung mit diesem Thema im Unterricht allerdings etwa zwei Jahre zurück, da der Regenwald in der achten Klasse behandelt wird und die untersuchten Schüler aus der zehnten Klasse stammen. Dies stellt kein Problem dar, denn der Geografietest soll wie beschrieben nicht die in der aktuellen Klassenstufe erarbeiteten Kenntnisse, sondern grundlegende Geografiekompetenz erfassen.

Der Geografietest liegt in zwei parallelen Formen vor; es existiert eine deutschsprachige und eine englischsprachige Form, jeweils für die mono- und die bilingualen Schüler. Insgesamt besteht der Geografietest aus acht Aufgaben. Diese setzen sich aus bis zu fünf Teilaufgaben zusammen. Bei der Hälfte der Aufgaben sind keine Teilaufgaben vorhanden. Genauer zur Verteilung der Teilaufgaben lässt sich aus Tabelle 5 entnehmen. Über alle Aufgaben ergeben sich so insgesamt 17 Teilaufgaben, zu deren Bearbeitung 90 Minuten vorgesehen sind.

Zur Veranschaulichung der Aufgaben seien an dieser Stelle zwei Beispiele aus dem Geografietest gegeben. Bei Aufgabe 1c ist die Diskursfunktion Beschreiben relevant. Im Aufgabentext heißt es “What type of climate does Kinsangani have? Describe Kinsangani’s climate in detail with the help of the chart.” Dabei haben die Schüler ein Klimadiagramm des afrikanischen Ortes Kinsangani als Material zur Verfügung, um die Aufgabe zu lösen. Bei der nächsten Aufgabe, der Aufgabe 1d, ist die Diskursfunktion Begründen/Erklären relevant. Die Aufgabenstellung lautet: “Medan is a place in Indonesia and has the following coordinates: 3° N/98° E. Is the climate in Medan different from the climate in Kinsangani? Motivate your answer.” Bei dieser Aufgabe sollen die Schüler erklären, dass das Klima bei beiden Orten vergleichbar ist, da die beiden Orte in der Demokratischen Republik Kongo und in Indonesien etwa gleich weit entfernt nördlich vom Äquator liegen. Es handelt sich um tropisches Klima. Zudem sollen die Schüler ihre Antwort begründen. In Anhang B finden sich weitere Beispielaufgaben.

3.2.2 Schülerfragebogen

Der Schülerfragebogen wurde mit dem Ziel konstruiert, alle für das Projekt relevanten Daten der Schüler zu erfassen. Er wurde in Anlehnung an den Schülerfragebogen von PISA erstellt

und greift nicht auf standardisierte Testverfahren oder überprüfte Skalen zurück. Als Fragenformate finden sich offene Fragen, Einfachwahlaufgaben, Mehrfachwahlaufgaben, Rangreihenbildungen und verschiedene Ratingskalen. Für die monolingualen und die bilingualen Schüler wurde jeweils eine eigene Fassung erstellt, wobei sich diese in etwa neun Fragen unterscheiden. Neben demografischen Daten wurden von den Schülern Angaben zur Schullaufbahn, zum Umgang mit Fremdsprachen, zur Selbsteinschätzung der sprachlichen Fähigkeiten, zum Erdkundeunterricht, zu außerschulischen Aktivitäten, zu häuslichen Lebens- und Lernbedingungen einschließlich der Berufstätigkeit der Eltern, zur Einstellung gegenüber Einzel-, Paar- und Gruppenarbeit, zur Schule und zur Klasse erhoben.

In Anhang A findet sich der Schülerfragebogen in der Fassung für bilingualen Schüler. Die Unterschiede gegenüber der Fassung für monolinguale Schüler betreffen keine Aufgaben, die für die vorliegende Arbeit relevant sind. Relevant sind zum einen Geschlecht und Alter der Schüler. Diese beiden Angaben wurden auf dem Deckblatt des Schülerfragebogens abgefragt. Beim Geschlecht konnten die Schüler ihr Geschlecht ankreuzen, beim Alter sollten sie ihr Alter in Jahren und Monaten an der dafür vorgesehenen Stelle eintragen. Mehr hervorzuheben ist die Erfassung der Berufstätigkeit der Eltern und des Klassenklimas

Die Berufstätigkeit der Eltern wurde für jedes Elternteil separat erfasst. Zuerst sollten die Schüler niederschreiben, welchen Beruf ihre Mutter ausübt und durch ankreuzen angeben, ob diese zur Zeit nicht berufstätig ist, in Teilzeit oder in Vollzeit arbeitet. Die entsprechenden Abfragen folgten dann für den Vater. Der zugehörige Teil des Schülerfragebogens ist in Abbildung 3 wiedergegeben. Über diese Informationen wurde jeweils für Mutter und Vater ein Wert im EGP-Klassenschema bestimmt. Dies geschah unter Verwendung des bei PISA eingesetzten EGP-Klassenschemas (vgl. Tabelle 3), wobei ein Beurteiler die entsprechenden Werte bestimmte. Der Niedrigste beider Werte bildet für jeden Schüler dann die Variable „Höchstes PISA-EGP“, da der niedrigste Zahlenwert den höchsten beruflichen Status innerhalb des Elternhauses widerspiegelt. An diesem EGP-Wert wird gleichzeitig der soziale Status des Elternhauses und die soziale Herkunft der Schüler festgemacht.

Das Klassenklima wurde über eine vierstufige Likert-Skala erfasst. Dabei kamen die in Abbildung 4 dargestellten sieben Items zum Einsatz. Wie bereits erwähnt, kommt damit keine standardisierte, sondern eine selbstkonstruierte Skala zum Einsatz. Dieses Vorgehen wurde aus Gründen der Einfachheit gewählt. Ob das Klassenklima über die verwendeten Items erfassbar ist, kann unter anderem im Rahmen einer Itemanalyse untersucht werden.

79. Welchen Beruf übt deine Mutter aus?

Beruf

80.

01 z.Z. nicht berufstätig

02 Teilzeit

03 Vollzeit

81. Welchen Beruf übt dein Vater aus?

Beruf

82.

01 z.Z. nicht berufstätig

02 Teilzeit

03 Vollzeit

Abbildung 3 Auszug aus dem Schülerfragebogen: Erhebung der Berufstätigkeit der Eltern.

Welche der folgenden Aussagen trifft auf dich zu? Bitte nur eine Antwort ankreuzen.		01	02	03	04
		stimme nicht zu	stimme eher nicht zu	stimme eher zu	stimme zu
129.	In unserer Klasse ist der Zusammenhalt gut.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
130.	In unserer Erdkunde klasse ist der Zusammenhalt gut.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
131.	In unserer Klasse fühle ich mich als nicht dazugehörig und als Außenseiter .	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
132.	In unserer Erdkunde klasse fühle ich mich wohl .	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
133.	Auf eine gute Klassengemeinschaft kommt es mir nicht an .	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
134.	In unserer Erdkundeklasse arbeiten die einzelnen Schüler eher gegeneinander als miteinander.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
135.	In unserer Erdkundeklasse bekommt man schnell Probleme, wenn man besonders gute (oder: relativ schlechte) Leistungen zeigt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Abbildung 4 Auszug aus dem Schülerfragebogen: Erhebung des Klassenklimas.

Abschließen sei noch darauf hingewiesen, dass bei der Datenerhebung ein Lehrerfragebogen zum Einsatz kam, der von den 18 beteiligten Lehrern ausgefüllt wurde, dessen Ergebnisse für die vorliegenden Arbeit allerdings nicht relevant sind.

3.2.3 Kognitiver Fähigkeitstest

Der Kognitive Fähigkeitstest von Heller und Perleth (2000) ist ein nach der klassischen Testtheorie konstruierter „differentieller Intelligenztest zur Ermittlung der kognitiven Ausstattung von Schülern“ (ebd., S. 8), der auch in Rahmen von PISA eingesetzt worden ist (Kunter et al., 2002). Im Folgenden soll der KFT vorgestellt und seine Konstruktion erläutert werden. Im Anschluss daran werden die Testgütekriterien Reliabilität, Validität und Normierung behandelt.

Der KFT ist die deutsche Übertragung eines ursprünglich englischsprachigen Intelligenztests, der für den Einsatz bei Schülern konzipiert ist. Er ist als Einzel- und als Gruppentest anwendbar und existiert in zwei parallelen Testformen. Da zudem zeitliche Begrenzungen für die Subtests vorliegen ist er ein „kombinierter Power-Speed-Test“ (Heller & Perleth, 2000, S. 8). Zur Bearbeitung des Tests erhalten die Schüler ein Testheft und ein der Klassenstufe entsprechendes Antwortblatt. Die Aufgaben sind im Testheft für jeden Subtest mit aufsteigender Schwierigkeit geordnet. Je Klassenstufe wird nur eine vorgegebene Auswahl an Aufgaben bearbeitet. Der KFT erfasst verbales, nonverbales und quantitatives Denken. Zu jedem dieser drei Bereiche gibt es drei Subtests, aus deren Ergebnissen sich das kognitive Gesamtleistungsniveau der Schüler ergibt.

Der verbale Teil besteht aus den Subtests *V1 Wortschatz*, *V2 Wortklassifikationen* und *V3 Wortanalogien*. Beim ersten Subtest soll zu einem gegebenen Wort aus einer Reihe von fünf Worten dasjenige heraus gesucht werden, dass zu dem vorgegebenen Wort passt. Beim Zweiten sind drei Worte gegeben, die inhaltlich zusammen gehören und durch ein dazugehöriges viertes Wort ergänzt werden sollen, dass aus fünf gegebenen Worten auszuwählen ist. Beim Dritten soll aus fünf vorgegebenen Worten dasjenige heraus gesucht werden, welches zusammen mit einem weiteren vorgegebenen Wort ein inhaltliches Paar bildet, das dem in einem Beispiel angegebenen Paar entspricht.

Der nonverbale Teil setzt sich zusammen aus den Subtests *N1 Figurenklassifikation*, *N2 Figurenanalogien* und *N3 Faltaufgaben*. Beim ersten Subtest sind Figuren gegeben, die nach einem bestimmten Merkmal klassifizierbar sind und durch eine aus fünf vorgegebenen Figuren ergänzt werden soll, die ebenfalls nach diesem Merkmal klassifizierbar ist. Beim Zweiten soll – ähnlich wie beim Subtest Wortanalogien – aus fünf vorgegebenen geometrischen Figuren diejenige herausgesucht werden, die zusammen mit einer weiteren gegebenen Figur ein Paar bildet, dass einem im Beispiel angegebenen Paar entspricht. Beim Dritten wird

in Gedanken ein Blatt Papier gefaltet und gestanzt. Danach muss entschieden werden, welches der fünf vorgegebenen Muster sich nach dem Auseinanderfalten ergeben würde.

Der quantitative Teil beinhaltet die Subtests *Q1 Mengenvergleiche*, *Q2 Zahlenreihen* und *Q3 Gleichungenbilden*. Beim ersten Subtest sind verschiedene Mengen angegeben, für die entschieden werden soll, ob zwei Mengen kleiner, größer oder gleich sind. Beim Zweiten soll aus fünf vorgegebenen Zahlen diejenige heraus gesucht werden, die eine angegebene Zahlenreihe fortsetzt, der eine bestimmte Bildungsregel zu Grunde liegt. Beim Dritten sind Zahlen und Operationszeichen gegeben, die so kombiniert werden müssen, dass sich eines der fünf angegebenen Ergebnisse ergibt, wobei immer nur eine richtige Lösung existiert.

Neben dieser soeben beschriebenen Normalform des Tests, deren Durchführung etwa 130 Minuten in Anspruch nimmt, existiert auch ein Kurzform, die in etwa 90 Minuten durchgeführt werden kann. Sie setzt sich zusammen aus den Subtests V1, V3, N1, N2, Q1 und Q2. Laut Heller und Perleth (2000) ist diese Kurzform bei Praktikern beliebt. Die Autoren geben auch Korrelationen zwischen der Normalform und der Kurzform an, die für die drei Testbereiche und für die Gesamtleistung größer oder gleich .93 sind. Bei der vorliegenden Untersuchung wurden aus zeitlichen Gründen nur die Subtests V3, Q2 und N2 der Testform A verwendet. Vollmer (2006b) rechtfertigt diese weitere Reduktion damit, dass Gruppenvergleiche und nicht Individualdiagnostik kognitiver Leistungen Ziele im Rahmen des DFG-Projektes sind. Auch Koch (2005) setzt bei ihren Datenerhebungen auf wenige ausgewählte Subtests des KFT. Aus zeitlichen Gründen verwendet sie nur die Subtests V2 und N1.

Speziell für Leistungstest, zu denen auch Intelligenztests wie der KFT zählen, wird eine Streuung der Aufgabenschwierigkeit³ von $P = 80$ bis $P = 20$ empfohlen (Lienert und Raatz, 1998). Dabei sollte die Aufgabenschwierigkeit innerhalb der genannten Grenzen von Aufgabe zu Aufgabe möglichst sinken, sprich die Aufgaben im Leistungstests sollten zum Testende hin schwerer werden. Daran orientierte man sich auch bei der Konstruktion des KFT. Zudem wurde versucht, eine korrigierter Trennschärfe von $r \geq .3$ zu erreichen. Heller und Perleth (2000) berichten für die erläuterten Koeffizienten Ergebnisse von zwei ausgewählten Subtests, die befriedigende Ergebnisse aufweisen. Sämtliche Aufgabenanalysen wurden anhand der Normierungsstichprobe gerechnet und fielen nach Angaben der Autoren

³ Mit dem Begriff „Aufgabenschwierigkeit“ ist in der vorliegenden Arbeit der statistische Koeffizient gemeint, der sich aus dem Verhältnis der Anzahl der Testteilnehmer, welche die Aufgabe richtig gelöst haben, zur Gesamtzahl der Testteilnehmer ergibt. Dabei hat eine schwer zu lösende Aufgabe eine geringe Aufgabenschwierigkeit. Mit dem Begriff „Schwierigkeit“ ist hingegen das normalsprachliche Verständnis verbunden.

für die vorliegende revidierte Fassung des KFT nicht schlechter aus als für die ältere Vorgängerfassung des Tests.

Die Reliabilität mit der Paralleltestmethode und der Testwiederholungsmethode wurde für die ältere Fassung des Tests in einer Untersuchung gemeinsam bestimmt. Dabei bekamen Schüler ($N = 447$) von Hauptschulen und Gymnasien die Testform A oder B vorgelegt. Bei einer Testung drei Wochen später bekamen die gleichen Schüler jeweils die andere Testform zur Bearbeitung. Dadurch sollte gleichzeitig die Äquivalenz der beiden Testformen und die Stabilität untersucht werden. Über die verschiedenen Subtests und Klassenstufen ergaben sich für die Testform A Ergebnisse von $r_{tt} = .04$ bis $r_{tt} = .95$ und für die Testform B von $r_{tt} = .39$ bis $r_{tt} = .95$. Die beiden Testformen können auf Grundlage der Ergebnisse als äquivalent angesehen werden. Zudem scheint der KFT zumindest über kurze Zeiträume stabile Ergebnisse zu liefern. Als Einschränkung ist zu sehen, dass diese Ergebnisse nicht mit der aktuellen Testform gewonnen worden sind. Für längere Zeiträume berichten Heller und Perleth (2000) Stabilitätskoeffizienten von $r_{tt} = .83$ bis $r_{tt} = .87$ bei einem Jahr zwischen den Testungen und von $r_{tt} = .83$ bei zwei Jahren. Diese Ergebnisse stammen von Datenerhebungen an einer Gesamtschule in München.

Die Reliabilität als innere Konsistenz wurde beim KFT mit der *Kuder-Richardson-Equation 20* (Kuder & Richardson, 1937) berechnet, die eingesetzt werden kann, wenn die Aufgabeninterkorrelationen etwa gleich hoch sind oder dieses angenommen wird. Heller und Perleth (2000) berichten für die verschiedenen Subtests und für verschiedene Klassenstufen Konsistenzkoeffizienten, die anhand der Normierungsstichprobe ermittelt wurden. Für die Testform A liegen die Ergebnisse im Bereich von $r_{tt} = .39$ bis $r_{tt} = .97$ und für die Testform B im Bereich von $r_{tt} = .33$ bis $r_{tt} = .96$. Über alle Subtests und Testformen hinweg ergibt sich ein Median der Konsistenzkoeffizienten von $r_{tt} = .79$.

Ein weiterer wichtiger Reliabilitätsaspekt ist die Profilreliabilität, die Reliabilität der Differenzen. Sie ist nach Lienert und Raatz (1998) bedeutend für die Interpretation von Testprofilen, wie sie auch beim KFT vorkommen. Sie fällt umso höher aus, je reliabler die Einzeltests sind und je niedriger ihre Interkorrelation ist. Ausreichend ist eine Profilreliabilität $r_{\text{profil}} > .5$, die der KFT für alle Klassenstufen mit Ausnahme der Klassen 11 und 12 erreicht. Insgesamt kann die Reliabilität des KFT als ausreichend angesehen werden.

Bei Heller und Perleth (2000) werden zur Übereinstimmungsvalidierung die Korrelationen der KFT-Ergebnisse und der Deutsch-, Englisch, sowie der Mathematikzensur berichtet. Für den KFT-Gesamtwert ergeben sich Korrelationen von $r = .07$ bis $r = .65$, die

besonders bei der Mathematikzensur hoch ausfallen. Zur Vorhersagevalidität berichten die Autoren wiederum Ergebnisse der älteren Testfassung, was bei den notwendigen längsschnittlichen Ansätzen nicht überrascht. Beispielsweise wird von einer Untersuchung über den Zusammenhang der KFT-Ergebnisse mit Abiturnoten berichtet. Dabei ergaben sich Korrelationen von $r = .16$ bis $r = .79$, die für eine substanzielle Kriteriumsvalidität des KFT sprechen.

Zur Konstruktvalidität berichten Heller und Perleth (2000) Korrelationen ausgewählter KFT-Subtests mit zwei anderen Intelligenztests. Die differenzierten Befunden sollen an dieser Stelle nicht genauer berichtet werden, fallen allerdings für den KFT positiv aus. Weiterhin wird von einer Untersuchung berichtet, die zeigt, dass mit Hilfe des KFT Leistungsunterschiede zwischen den Schulformen erfasst werden können. Dabei ergaben sich für die Klassenstufen fünf bis neun die folgenden T -Werte: Hauptschule $T \approx 40$ bis $T \approx 45$, Realschule $T \approx 50$ bis $T \approx 55$, Gymnasium $T \approx 60$. Insgesamt kann die Konstruktvalidität des KFT damit als ausreichend bewertet werden.

Auch zur faktoriellen Validität berichten Heller und Perleth (2000) Ergebnisse zur aktuellen Testfassung. Dabei wurden Faktorenanalysen mit Varimax-Rotation durchgeführt. Bei der Begrenzung auf drei Faktoren konnten von der gesamten Variablenvarianz 68,7% aufgeklärt werden. Bei Betrachtung der Scree-Plots zeigt sich allerdings über verschiedene Klassenstufen ein starker Abfall der Eigenwerte vom ersten auf den zweiten Faktor. Bei Letzterem gilt für den Eigenwert dann bereits in den meisten Fällen $\lambda < 1$, womit eine dreifaktorielle Lösung in Frage zu stellen ist. Eine solche Lösung würde für die drei Bereiche des KFT sprechen, da sich laut den Autoren je ein Faktor für sprachgebundenes, formallogisches und zahlengebundenes Denken finden lässt.

Die Normierungsstichprobe besteht aus 6 765 Teilnehmern. Davon sind 3 275 Schülerinnen und 3 459 Schüler. Die beiden Werte addieren sich auf Grund von fehlenden Geschlechtsangaben nicht zur Teilnehmerzahl der Normierungsstichprobe. Mit zunehmender Klassenstufe wird die Teilnehmerzahl geringer. Daher können nur die Normen für die Klassenstufen vier bis sieben als voll befriedigend angesehen werden. Für Klassenstufen und Schularten, von denen zu wenig Schüler in der Normierungsstichprobe repräsentiert sind, geben Heller und Perleth (2000) auch keine Normen an. Zur Verteilung auf die Bundesländer der Bundesrepublik Deutschland ist anzumerken, dass fast alle Datenerhebungen an bayrischen Schulen durchgeführt wurden. Lediglich 199 Schüler wurden in einem anderen Bundesland, nämlich Baden-Württemberg, erhoben. Alle Erhebungen wurden am Ende des Schuljahres 1995/1996 und in der Mitte und am Ende des Schuljahres 1996/1997 durchgeführt. Auf

geschlechtsspezifische Normen haben die Autoren verzichtet, da sich in Analysen keine Geschlechtsunterschiede ergeben hatten. Ferner fanden sich in Analysen keine deutlichen Auswirkungen der Reihenfolge der Subtests.

3.2.4 C-Test

Der C-Test ist ein Testverfahren der Sprachdiagnostik, mit dem die allgemeine Sprachbeherrschung, die Sprachkompetenz festgestellt werden kann. Er ist aus der Kritik an seinem Vorgängerverfahren, dem Cloze-Test, hervorgegangen. Der Cloze-Test besteht aus einem beliebig langen Textstück, bei dem Worte nach einer festgelegten Regel entfernt werden, beispielsweise jedes fünfte Wort. Der Testteilnehmer muss dann die so entstandenen Lücken im Text schließen. An diesem Testverfahren wurde kritisiert, dass die entfernten Wörter keine repräsentative Stichprobe der Wortarten des Originaltextes sind, dass sich die Schwierigkeit je nach gewähltem Startpunkt der Wortentfernungen verändert, dass ein Text leicht Testteilnehmer mit vertieften Kenntnissen im behandelten Inhaltsbereich bevorzugen kann, dass die Auswertung wegen oftmals vieler möglicher richtiger Lösungen für eine Lücke nicht objektiv ist und dass selbst Muttersprachler in einem Cloze-Test praktisch nie alle Lücken richtig füllen können.

Auf der Grundlage dieser Kritik wurde der C-Test entwickelt. Eine der ersten Arbeiten, in denen er vorgestellt wird, ist der Artikel von Klein-Braley und Raatz (1982). Beim C-Test wird jedes zweite Wort eines Textes gelöscht, wobei nicht das ganze Wort, sondern nur die zweite Hälfte entfernt wird. Bei ungerader Buchstabenanzahl wird jeweils ein Buchstabe mehr entfernt. Auch bei diesem Testverfahren muss der Testteilnehmer die vorhandenen Lücken schließen. Alle richtig gefüllten Lücken, respektive gelösten Aufgaben, werden als Punkt gezählt und zu einem Gesamtwert für den Test summiert. Konstruktionsbedingt ist in fast allen Fällen nur eine Lösung pro Lücke richtig, was die Basis für eine objektive Auswertung schafft. Zudem zeigten Klein-Braley und Raatz (1982), dass Muttersprachler annähernd die höchstmögliche Punktzahl erreichen und dass die entfernten Wörter eine repräsentative Stichprobe aller Wortarten des Originaltextes darstellen. Ein weiterer Unterschied gegenüber dem Cloze-Test ist, dass beim C-Test verschiedene kürzere Texte – meist vier – anstelle eines längeren Textes verwendet werden. So wird eine Bevorteilung einzelner Testteilnehmer vermieden, die mit dem Inhalt eines Textes besonders vertraut sind. Pro Text finden sich normalerweise 20 oder 25 Lücken, was 80 bis 100 Aufgaben für einen C-Test ergibt.

Ursprünglich wurden verschiedene C-Tests mit deutschen und englischen Texten entwickelt, die für unterschiedliche Zielgruppen gedacht waren. Klein-Braley und Raatz (1982) berichten von Untersuchungen mit Mutter-, und Fremd-, und Zweitsprachlern, darunter englische Offiziere, deutsche Anglistikstudenten, deutsche Schüler, englische Schüler und in Deutschland lebende Ausländer. In vielen verschieden gestalteten Untersuchungen konnten sie zeigen, dass sich in den meisten Fällen für einen C-Test Texte mit abnehmender Aufgabenschwierigkeit finden lassen. Zudem liegt die Trennschärfe etwa bei $.60 < r_{jt} < .80$. Es fanden sich substantielle Korrelationen mit anderen Sprachtestverfahren und eine hohe Korrelation mit der Deutschnote ($r_{tc} = .8$). Die gefundenen Reliabilitätskennwerte liegen um $r_{tt} = .8$ und es zeigten sich in Faktorenanalysen hohe Ladungen auf einen einzigen Faktor, der mit „globale Sprachbeherrschung“ beschrieben werden kann. Weitere Daten, auch zur testtheoretischen Fundierung, finden sich bei Faust (1985).

Bei Raatz und Klein-Braley (1983) finden sich zusätzliche Befunde – unter anderem zur Legastheniediagnose – welche die Güte des C-Tests belegen. Dort gehen die Autoren genauer auf den theoretischen Hintergrund des C-Test ein, wobei zwei wesentliche Punkte hier herausgestellt werden sollen. Zum einen werden im Bereich der Sprachdiagnostik die Begriffe Kompetenz und Performanz anders als in anderen Kontexten verstanden. *Kompetenz* bezieht sich dabei auf potentiell Erzeugen von Sprachstrukturen, *Performanz* auf das wirkliche Erzeugen von Texten. Mit dieser Sichtweise bietet ein Mehrfachwahltest die Möglichkeit Kompetenz zu erfassen, indem Aufgaben verwendet werden, zu deren Lösung die Beherrschung des sprachlichen Regelsystems erforderlich ist. Zum anderen ist eine Grundannahme beim C-Test ein Kommunikationsmodell, bei dem der Hörer kontinuierlich Hypothesen darüber bildet, was der Sprecher als nächstes sagen wird und diese Hypothesen mit dem wirklich Gesprochenen vergleicht. Diese *Erwartungsgrammatik* bildet die notwendige Voraussetzung zum Lösen eines C-Tests, da ein Testteilnehmer die Lücken nur füllen kann, wenn er entsprechende Hypothesen zur Verfügung hat. Je höher die Sprachkompetenz, desto detaillierter die Hypothesen und desto besser die Lösung des C-Tests. Grotjahn (1995) formuliert es so:

Der C-Test ist wie der Cloze Test eine Operationalisierung des Konzepts des *reduced redundancy testing*. Die zugrundeliegende Annahme lautet: Je kompetenter die Lerner in einer Fremdsprache sind, desto besser können sie im Zuge einer konstruktiven und antizipatorischen Verarbeitung von Sprache Gebrauch von der natürlichen Redundanz eines Textes machen und desto besser können sie den C-Test lösen. (S. 38)

Genauer auf die mentalen Verarbeitungsprozesse beim Lösen von C-Tests geht Grotjahn (1993) ein. Er kommt zu dem Ergebnis, dass der C-Test nicht zur Erfassung von Lesekompetenz geeignet ist, da er Sprachkompetenz auf einer allgemeineren Ebene misst und höhere Verarbeitungsprozesse nur eine geringe Rolle spielen. Zudem können von der Leistung im C-Test nicht direkt Rückschlüsse auf die mündliche Kommunikationsfähigkeit gezogen werden. In einem französischen C-Test wurde systematisch die Reihenfolge von Teilsätzen verändert. Dabei zeigte sich, dass für die Lösung des C-Tests hauptsächlich der Mikrokontext bedeutsam ist, die veränderte Teilsatzreihenfolge also nur geringe Auswirkungen hat.

Grotjahn (1995) weist auf C-Tests für weitere Sprachen und die damit einhergehenden sprachspezifischen Probleme hin. Er gibt eine Auflistung verschiedener aktueller Einsatzbereiche von C-Tests. Danach werden C-Tests unter anderem verwendet zur Sprachkompetenzbestimmung bei Schülern mit Deutsch als Zweitsprache, als Einstufungsverfahren an Universitäten, bei Volkshochschulkursen, beim Bundeswettbewerb Fremdsprachen, beim psychologischen Dienst der Arbeitsagenturen und als standardisierter Schulleistungstest für die deutsche Sprache. Zudem kann der C-Test nicht nur in der klassischen Papier-und-Bleistift Form angewendet werden, sondern auch rechnergestützt.

Im DFG-Projekt wird jeweils ein deutschsprachiger und ein englischsprachiger C-Test eingesetzt. Der deutsche C-Test bestand zunächst aus acht Texten, die im Zuge der ersten Datenerhebungen an Schule 1 (vgl. Abschnitt 3.3) eingesetzt wurden. Auf Grund der so gewonnenen Erkenntnisse wurden vier Texte mit je 25 Lücken für die endgültige Form ausgewählt. Diese Form ist in Anhang C abgedruckt. Der englische C-Test lag vor, musste daher nicht vorgetestet werden und setzt sich ebenfalls aus vier Texten mit jeweils 25 Lücken zusammen. Bei beiden C-Tests konnten somit maximal 100 Punkte erreicht werden.

3.3 Stichprobe

Die Datenerhebung fand an sieben niedersächsischen Gymnasien statt. Zur Auswahl der Schulen wurde über das Schulamt der Stadt Osnabrück eine Liste aller Gymnasien im Regierungsbezirk Weser-Ems beschafft, die bilingualen Unterricht anboten. Dieser Regierungsbezirk wurde ausgewählt, weil so eine gute Erreichbarkeit von Osnabrück aus und damit eine einfachere Datenerhebung gewährleistet war. Die Liste umfasste 20 Gymnasien, von denen an einigen der bilinguale Bildungsgang erst in der Erprobung respektive im Aufbau war. Von diesen 20 Gymnasien wurden diejenigen Schulen ausgewählt, die mit Erdkunde als

bilingualem Unterrichtsfach in der siebten Klasse starteten und später ein weiteres bilinguales Unterrichtsfach hinzunehmen.

Die in Frage kommenden Schulen wurden kontaktiert. Dabei schied von vornherein eine Schule in Georgsmarienhütte aus, da diese im Rahmen der Entwicklung ihres bilingualen Bildungsganges intensiv mit dem Fachbereich Sprach- und Literaturwissenschaft der Universität Osnabrück kooperiert hatte. Zur Mitarbeit bereit waren letztendlich neun Schulen. Darunter waren auch zwei Gymnasien, an denen Geschichte als ersteinsetzendes bilinguales Sachfach gelehrt wurde. Diese beiden Schulen wurden entgegen der ursprünglichen Einschlusskriterien mit aufgenommen, um Daten über die Geografiekompetenz von Schülern zu sammeln, die nicht am bilingualen Erdkundeunterricht teilgenommen hatten. An wiederum zwei dieser neun Schulen wurden die Pilotuntersuchungen durchgeführt. Diese beiden Schulen lagen im Stadtgebiet von Osnabrück. Die eigentliche Datenerhebung fand an den restlichen sieben Schulen statt. Dabei wurden die Daten nur an Schülern der zehnten Klasse erhoben. Im Regelfall wurde nicht der komplette Jahrgang einer Schule einbezogen, da dies einfacher zu realisieren war und auch für die Schulen einen geringeren Aufwand bedeutete.

Im Folgenden wird von *Erhebungsgruppen* berichtet. Dieser Begriff wird an Stelle von *Klassen* verwendet. Das liegt daran, dass nicht in allen Fällen die Daten in Gruppen erhoben wurden, die im Schulalltag als Klasse gemeinsamen Unterricht hatten. Vielmehr waren einige Gruppen nach anderen Kriterien zusammengestellt: An den Schulen 4 und 5 wurde ein kompletter dreizügiger Jahrgang erhoben, der auf die drei Erhebungsgruppen 40-42 und 50-52 aufgeteilt wurde. Die Erhebungsgruppen 41 und 42 wurden in einer gemeinsamen Sitzung erhoben; alle anderen Erhebungsgruppen jeweils für sich allein.

Insgesamt wurde Daten an 380 (100%) Schülern erhoben, die sich auf 18 Erhebungsgruppen verteilen. Die Schüler wurden von 18 verschiedenen Sachfachlehrern unterrichtet, wobei zu den meisten Erhebungsgruppen auch genau eine Lehrperson gehörte. Von den Schülern nahmen 217 (57%) am monolingualen Unterricht und 163 (42%) am bilingualen Unterricht teil. Eine Übersicht über diese Stichprobe, die im Folgenden als *Gesamtstichprobe* bezeichnet wird, gibt Tabelle 6. Die dort angegebenen Schülercodes wurden zu Beginn der Erhebung anhand von Namenslisten der Schüler zugewiesen.

Tabelle 6

Die Tabelle gibt eine Übersicht über die Erhebungsgruppen der Gesamtstichprobe. Dargestellt sind Schule, Erhebungsgruppe (Gruppe), Schülercodes, Unterrichtstyp, Anzahl der Schüler in der Erhebungsgruppe und deren Anteil an der Hauptstichprobe. In den letzten beiden Spalten ist die Anzahl der Schüler und deren Anteil an der Erhebungsgruppe aufgeführt, die bei der Schülerfragebogenfrage „Welche Sprache sprichst du zuhause?“ auch „Deutsch“ angegeben hatten.

Schule	Gruppe	Schüler- codes	Unterrichts- typ	Anzahl Schüler	Anteil an Gesamt	Sprache Deutsch	Anteil an Gruppe
1	10	ab 1001	bilingual	28	7%	24	86%
	11	ab 1101	monolingual	30	8%	27	90%
	12	ab 1201	monolingual	19	5%	13	68%
2	20	ab 2001	bilingual	30	8%	22	73%
	21	ab 2101	monolingual	26	7%	21	81%
	22	ab 2201	monolingual	30	8%	24	80%
3	30	ab 3001	bilingual	29	8%	27	93%
	31	ab 3101	monolingual	29	8%	23	79%
4	40	ab 4001	bilingual	23	6%	14	61%
	41	ab 4101	monolingual	8	2%	7	88%
	42	ab 4201	monolingual	10	3%	8	80%
5	50	ab 5001	bilingual	19	5%	11	58%
	51	ab 5101	monolingual	13	3%	7	54%
	52	ab 5201	monolingual	22	6%	12	55%
6	60	ab 6001	bilingual	13	3%	10	77%
	61	ab 6101	monolingual	20	5%	17	85%
7	70	ab 7001	bilingual	21	6%	11	52%
	71	ab 7101	monolingual	10	3%	5	50%
Gesamt				380	100%	283	75%[†]

[†] *Bemerkung:* Dieser Wert stellt keine Spaltensumme dar, sondern den Anteil der Schüler an der Gesamtstichprobe, die „Deutsch“ angegeben haben.

Bei der Betrachtung von Tabelle 6 zeigt sich, dass in den meisten Erhebungsgruppe 20 bis 30 Schüler vorhanden sind. Bemerkenswert ist, dass in den Erhebungsgruppen 12 und 40 der Anteil der Schüler, die angaben, zu Hause deutsch zu sprechen, unter 70% liegt. Noch deutlicher ist dies bei den Schulen 5 und 7, bei denen der Anteil zwischen 60% und 50% liegt. In Tabelle 7 findet sich eine Übersicht über das zahlenmäßige Verhältnis von Schülerinnen und Schülern.

Tabelle 7

Die Tabelle zeigt die Verteilung der Geschlechter in den Erhebungsgruppen der Gesamtstichprobe. Dargestellt ist die Schule, die Erhebungsgruppe (Gruppe) und die Anzahl der Schülerinnen und Schüler mit ihrem jeweiligen Anteil an der Erhebungsgruppe. In der letzten Spalte ist aufgeführt, wie viele Schüler pro Erhebungsgruppe keine Angabe zu ihrem Geschlecht gemacht haben.

Schule	Gruppe	Anzahl weiblich	Anteil weiblich	Anzahl männlich	Anteil männlich	ohne Geschlechtsangabe
1	10	17	61%	11	39%	0
	11	11	37%	19	63%	0
	12	9	47%	10	53%	0
2	20	25	83%	5	17%	0
	21	15	58%	11	42%	0
	22	23	77%	7	23%	0
3	30	14	48%	15	52%	0
	31	21	72%	7	24%	1
4	40	6	26%	17	74%	0
	41	2	25%	6	75%	0
	42	2	20%	8	50%	0
5	50	14	74%	5	26%	0
	51	9	69%	4	31%	0
	52	11	50%	10	46%	1
6	60	7	54%	6	46%	0
	61	5	25%	13	65%	2
7	70	14	67%	4	19%	3
	71	2	20%	8	80%	0
Gesamt		207	55% [†]	166	44% [†]	7
Gesamt NDS 1 [*]		130 863	55%	108 350	45%	
Gesamt NDS 2 [×]		139 851	55%	115 622	45%	
Gesamt NDS 3 [◇]		40 137	51%	39 146	49%	

[†] Diese Werte stellen keine Spaltensummen dar, sondern den Anteil der weiblichen und männlichen Schüler an der Gesamtstichprobe.

^{*} Die Werte in dieser Zeile sind Daten für Schüler an Gymnasien im Bundesland Niedersachsen (NDS) von September 2005 (Niedersächsisches Landesamt für Statistik, 2006).

[×] Die Werte in dieser Zeile sind Daten für Schüler an Gymnasien im Bundesland Niedersachsen von 2005/2006 (Statistisches Bundesamt, 2007a).

[◇] Die Werte in dieser Zeile sind Daten für Schüler der Jahrgangsstufe 10 im Bundesland Niedersachsen von 2005/2006 (Statistisches Bundesamt, 2007b).

Beim Vergleich des Anteils weiblicher und männliche Schüler an der Gesamtstichprobe mit Daten aus verschiedenen Quellen, die sich auf die Anteile weiblicher und männlicher Schüler in ganz Niedersachsen beziehen, zeigt sich in Tabelle 7 eine hohe Übereinstimmung. Hinsichtlich der Geschlechterverteilung ist die Gesamtstichprobe damit repräsentativ für die Schüler an Gymnasien in Niedersachsen. In Tabelle 8 finden sich Angaben zum Alter der Schüler in den Erhebungsgruppen.

Tabelle 8

Die Tabelle gibt eine Übersicht über das Alter der Schüler in den Erhebungsgruppen der Gesamtstichprobe. Dargestellt ist die Schule und die Erhebungsgruppe (Gruppe). In den folgenden vier Spalten sind Kennwerte zum Alter der Schüler in Jahren gegeben. In den letzten vier Spalten sind wiederum Kennwerte zum Alter der Schüler gegeben, welches jedoch über das Alter in Monaten und anschließende Division durch 12 errechnet wurde. Die berichteten Kennwerte sind Minimum (Min.), Maximum (Max.), Mittelwert (M) und Standardabweichung (SD).

Schule	Gruppe	Alter in Jahren				Alter in Monaten / 12			
		Min.	Max.	M	SD	Min.	Max.	M	SD
1	10	15	17	15.85	0.46	15.25	17.75	16.40	0.48
	11	16	17	16.10	0.31	16.00	17.83	16.58	0.44
	12	16	16	16.00	0.00	16.00	16.92	16.48	0.27
2	20	13	17	15.87	0.68	13.92	17.58	16.38	0.61
	21	15	17	16.04	0.64	15.92	17.83	16.55	0.63
	22	15	17	16.04	0.43	15.83	17.83	16.55	0.47
3	30	15	16	15.97	0.19	15.92	16.83	16.50	0.31
	31	15	18	16.14	0.65	15.58	18.17	16.60	0.59
4	40	15	17	15.95	0.50	15.50	17.50	16.50	0.52
	41	16	16	16.00	0.00	16.17	16.92	16.55	0.23
	42	16	17	16.25	0.46	16.00	17.83	16.69	0.69
5	50	15	17	15.94	0.56	15.92	17.67	16.41	0.50
	51	15	17	16.00	0.41	15.83	17.50	16.51	0.44
	52	16	18	16.10	0.45	16.00	18.17	16.43	0.47
6	60	15	16	15.90	0.32	15.92	16.92	16.54	0.30
	61	16	17	16.07	0.27	16.00	17.50	16.54	0.43
7	70	16	17	16.24	0.44	16.00	17.92	16.69	0.46
	71	15	16	15.90	0.32	15.92	16.92	16.50	0.35
Gesamt [†]		13	18	16.01	0.46	13.92	18.17	16.51	0.48

[†] *Bemerkung:* Die Werte in dieser Zeile sind keine Spaltensummen, sondern die entsprechenden Kennwerte bezogen auf alle 351 Schüler, für die Altersangaben vorhanden waren.

In Tabelle 8 werden Kennwerte für zwei verschiedene Arten der Altersbestimmung angegeben. In den Schülerfragebögen sollten die Schüler in den meisten Fällen ihr Alter in Jahren und Monaten angeben. In wenigen Schülerfragebögen wurde hingegen nach dem Geburtsjahr und dem Geburtsmonat gefragt. Aus diesen Werten wurde ausgehend vom Erhebungszeitpunkt das Alter in Jahren und Monaten berechnet. Bei Schülern, die keine Monate angaben, wurde von null Monaten ausgegangen. Beide Arten der Altersbestimmung liefern vergleichbare Ergebnisse, wie in Tabelle 8 ersichtlich ist. Auffällig sind nur zwei relativ kleine Erhebungsgruppen, bei denen das Alter je nach Art der Altersbestimmung streut beziehungsweise nicht streut. Mit Blick auf die aktuelle Einschulungspraxis in Niedersachsen – Kinder, die bis zum 30.06. eines Jahres das sechste Lebensjahr vollendet haben, werden im gleichen Jahr nach den Sommerferien eingeschult – ist ein mittleres Alter im Bereich von 16 bis 17 Jahren

zu erwarten, da die Erhebung gegen Ende des Schuljahres durchgeführt wurde. Bei etwa gleicher Streuung weichen die Mittelwerte für die beiden Arten der Altersberechnung um .50 voneinander ab. Beide Mittelwerte liegen damit im erwarteten Bereich.

Für allen weiteren Berechnungen, über die in der vorliegenden Arbeit berichtet wird, liegt – wenn nicht anders angegeben – nicht die Gesamtstichprobe von 380 Schülern zu Grunde. Diese Gesamtzahl reduziert sich um 206 Schüler auf 174 Schüler, und zwar aus den im Folgenden genannten Gründen. Zum einen werden zwei Gastschülerinnen nicht mit in die Auswertungen einbezogen, da sie nicht wie die anderen Schüler mehrere Jahre in Schule 3 unterrichtet wurden. In Schule 4 werden vier Schüler nicht mit in die Auswertung genommen, weil sie statt Erdkunde als bilinguales Fach Geschichte hatten. Damit können die Ergebnisse dieser Schüler nichts über den bilingualen Erdkundeunterricht aussagen. Schule 1 wird gar nicht berücksichtigt, da die Erhebung dort unter besonderen Bedingungen stattgefunden hat: Aus den Erhebungsgruppen 10 und 11 bearbeiteten 45 Schüler den Geografietest entweder in Partnerarbeit oder unter der Zusatzinstruktion des lauten Denkens. Dadurch sind die Ergebnisse im Geografietest für diese Schüler nicht direkt mit denen vergleichbar, die den Geografietest in der Erhebungsgruppe wie eine Klassenarbeit bearbeitet haben. Nur 13 Schüler aus diesen beiden Erhebungsgruppen bearbeiteten den Geografietest in der Gruppensituation in Einzelarbeit. Allerdings herrschten auch dabei andere Arbeitsbedingungen, weil die restlichen Schüler während der Zeit der Testbearbeitung anderweitig beschäftigt werden mussten. Ebenso ausgenommen wurden die Schulen 6 und 7, da bei diesen wiederum Geschichte als bilinguales Fach unterrichtet wurde.

Weiterhin reduziert sich die Gesamtzahl um vier Schüler (Schülerkodes 3025, 3109, 4003, 4207), deren Antworten im Geografietest zur Ausbildung der Beurteiler eingesetzt wurden. Es handelt sich dabei um zwei monolinguale und zwei bilinguale Schüler. Für weitere Ausbildungszwecke wurden die Antworten von drei weiteren Schülern (Schülerkodes 1101, 2008, 5219) verwendet, die allerdings in der Stichprobe verblieben. Die Antworten aller dieser Schüler wurden zufällig aus den Schülerantworten ausgewählt. Die Erhebungsgruppe 21 wurde komplett herausgenommen, da die Schüler während der Erhebung nicht diszipliniert gearbeitet hatten und der Lehrer die Kontrolle über die Klasse verlor. Bis hier hin beläuft sich die Reduktion der Stichprobe auf 176 Schüler. Schließlich wurden noch die Ergebnisse von 30 Schülern nicht mit berücksichtigt, die zwei oder mehr Aufgaben im Geografietest nicht bearbeitet hatten. Die fehlenden Aufgaben können besonders auf Zeitmangel bei der Testbearbeitung gegen Ende der Schulstunde zurückzuführen sein. Zudem mussten beispielsweise in

Erhebungsgruppe 40 und 42 mehrere Schüler wegen des Busses früher gehen. In den Erhebungsgruppen 40 und 52 musste jeweils ein Schüler zum Zahnarzt.

Im Folgenden soll genauer auf die Stichprobe eingegangen werden, die sich durch die oben genannten Reduktionen ergibt. Diese Stichprobe wird – gegenüber der oben behandelten Gesamtstichprobe, an der keine Reduktionen vorgenommen wurden – mit *Hauptstichprobe* bezeichnet. Die Hauptstichprobe besteht aus 174 (100%) Schülern, von den 90 (52%) monolingual und 84 (48%) bilingual unterrichtet wurden. Damit ergibt sich im Unterschied zur Gesamtstichprobe ein ausgewogeneres Verhältnis von monolingualen zu bilingualen Schülern. Eine Übersicht über die Hauptstichprobe gibt Tabelle 9.

Tabelle 9

Die Tabelle gibt eine Übersicht über die Erhebungsgruppen der Hauptstichprobe. Dargestellt sind Schule, Erhebungsgruppe (Gruppe), Schülercodes, Unterrichtstyp, Anzahl der Schüler in der Erhebungsgruppe und deren Anteil an der Hauptstichprobe. In den letzten beiden Spalten ist die Anzahl der Schüler und deren Anteil an der Erhebungsgruppe aufgeführt, die bei der Schülerfragebogenfrage „Welche Sprache sprichst du zuhause?“ auch „Deutsch“ angegeben hatten.

Schule	Gruppe	Schülercodes	Unterrichtstyp	Anzahl Schüler	Anteil an Gesamt	Sprache Deutsch	Anteil an Gruppe
2	20	ab 2001	bilingual	27	16%	20	74%
	22	ab 2201	monolingual	27	16%	23	85%
3	30	ab 3001	bilingual	28	16%	26	93%
	31	ab 3101	monolingual	25	14%	21	84%
4	40	ab 4001	bilingual	13	8%	9	69%
	41	ab 4101	monolingual	8	5%	7	88%
	42	ab 4201	monolingual	7	4%	6	86%
5	50	ab 5001	bilingual	16	9%	11	69%
	51	ab 5101	monolingual	6	3%	2	33%
	52	ab 5201	monolingual	17	10%	10	59%
Gesamt				174	100%	135	78%[†]

[†] *Bemerkung:* Dieser Wert stellt keine Spaltensumme dar, sondern den Anteil der Schüler an der Hauptstichprobe, die „Deutsch“ angegeben haben.

In 40% der Erhebungsgruppen aus Tabelle 9 finden sich 25 oder mehr Schüler, in 60% der Erhebungsgruppen 17 oder weniger. Beim Vergleich von Gesamt- und Hauptstichprobe (Tabelle 6 und Tabelle 9) fällt auf, dass der Anteil der Schüler, die angaben zu Hause deutsch zu sprechen, in der Hauptstichprobe um 3% höher liegt. Für Erhebungsgruppe 40 hat sich der Anteil erhöht, bleibt aber wiederum unter 70%. Unter dieser Marke bleibt der Anteil ebenso

für alle Erhebungsgruppen aus Schule 5. In Tabelle 10 findet sich eine Übersicht über das zahlenmäßige Verhältnis von Schülerinnen und Schülern.

Tabelle 10

Die Tabelle zeigt die Verteilung der Geschlechter in den Erhebungsgruppen der Hauptstichprobe. Dargestellt ist die Schule, die Erhebungsgruppe (Gruppe) und die Anzahl der Schülerinnen und Schüler mit ihrem jeweiligen Anteil an der Erhebungsgruppe. In der letzten Spalte ist aufgeführt, wie viele Schüler pro Erhebungsgruppe keine Angabe zu ihrem Geschlecht gemacht haben.

Schule	Gruppe	Anzahl weiblich	Anteil weiblich	Anzahl männlich	Anteil männlich	ohne Geschlechtsangabe
2	20	22	82%	5	19%	0
	22	20	84%	7	26%	0
3	30	13	46%	15	54%	0
	31	18	72%	7	28%	0
4	40	2	15%	11	85%	0
	41	2	25%	6	75%	0
	42	2	29%	5	71%	0
5	50	11	69%	5	31%	0
	51	6	100%	0	0%	0
	52	10	59%	7	41%	0
Gesamt		106	61% [†]	68	39% [†]	0
Gesamt NDS 1 [*]		130 863	55%	108 350	45%	
Gesamt NDS 2 [*]		139 851	55%	115 622	45%	
Gesamt NDS 3 [◇]		40 137	51%	39 146	49%	

[†] Diese Werte stellen keine Spaltensummen dar, sondern den Anteil der weiblichen und männlichen Schüler an der Hauptstichprobe.

^{*} Die Werte in dieser Zeile sind Daten für Schüler an Gymnasien im Bundesland Niedersachsen (NDS) von September 2005 (Niedersächsisches Landesamt für Statistik, 2006).

^{*} Die Werte in dieser Zeile sind Daten für Schüler an Gymnasien im Bundesland Niedersachsen von 2005/2006 (Statistisches Bundesamt, 2007a).

[◇] Die Werte in dieser Zeile sind Daten für Schüler der Jahrgangsstufe 10 im Bundesland Niedersachsen von 2005/2006 (Statistisches Bundesamt, 2007b).

Im Gegensatz zur Gesamtstichprobe (Tabelle 7) findet sich beim Vergleich des Anteils weiblicher und männlicher Schüler an der Hauptstichprobe (Tabelle 10) mit Daten aus verschiedenen Quellen, die sich auf die Anteile weiblicher und männlicher Schüler in ganz Niedersachsen beziehen, ein zu Gunsten der weiblichen Schüler verschobenes Geschlechterverhältnis. In der Hauptstichprobe finden sich 6% mehr weibliche Schüler und 5% weniger männliche Schüler. In Tabelle 11 finden sich Angaben zum Alter der Schüler in den Erhebungsgruppen.

Beim Vergleich von Gesamt- und Hauptstichprobe (Tabelle 8 und Tabelle 11) fällt auf, dass die Mittelwerte für beide Arten der Altersbestimmung jeweils identisch sind. Trotz

der oben genannten Reduktionen bleibt das mittlere Alter beider Stichproben gleich. Auch wenn für dieses Merkmal damit keine bedeutenden Veränderungen entstanden sind, zeigen sich wie berichtet hinsichtlich der häuslichen Sprache oder des Geschlechterverhältnisses Veränderungen. Ob die Reduktionen für andere Merkmale stichprobenneutral sind, ist fraglich.

Tabelle 11

Die Tabelle gibt eine Übersicht über das Alter der Schüler in den Erhebungsgruppen der Gesamtstichprobe. Dargestellt ist die Schule und die Erhebungsgruppe (Gruppe). In den folgenden vier Spalten sind Kennwerte zum Alter der Schüler in Jahren gegeben. In den letzten vier Spalten sind wiederum Kennwerte zum Alter der Schüler gegeben, welches jedoch über das Alter in Monaten und anschließende Division durch 12 errechnet wurde. Die berichteten Kennwerte sind Minimum (Min.), Maximum (Max.), Mittelwert (M) und Standardabweichung (SD).

Schule	Gruppe	Alter in Jahren				Alter in Monaten / 12			
		Min.	Max.	M	SD	Min.	Max.	M	SD
2	20	13	17	15.89	0.70	13.92	17.58	16.40	0.63
	22	15	17	16.04	0.44	15.83	17.83	16.56	0.48
3	30	15	16	15.96	0.19	15.92	16.83	16.50	0.31
	31	15	18	16.13	0.61	15.58	18.17	16.60	0.57
4	40	15	16	15.92	0.29	15.50	16.92	16.52	0.39
	41	16	16	16.00	0.00	16.17	16.92	16.55	0.23
	42	16	17	16.40	0.55	16.00	17.83	16.90	0.79
5	50	15	17	15.86	0.54	15.92	17.67	16.37	0.46
	51	16	16	16.00	0.00	16.25	16.92	16.50	0.31
	52	16	18	16.13	0.52	16.00	18.17	16.46	0.53
Gesamt [†]		13	18	16.01	0.49	13.92	18.17	16.51	0.49

[†] *Bemerkung:* Die Werte in dieser Zeile sind keine Spaltensummen, sondern die entsprechenden Kennwerte bezogen auf alle 165 Schüler, für die Altersangaben vorhanden waren.

Für die theoretisch bedeutsame Variable der sozialen Herkunft der Schüler wurde mittels eines Mann-Whitney-U-Tests geprüft, ob sich die Elternhäuser der Schüler der Hauptstichprobe von den Elternhäusern der Schüler unterscheiden, die wegen der Reduktionen nicht in die Hauptstichprobe übernommen wurden. Als Maß der sozialen Herkunft wurde dafür der höchste EGP-Wert der Eltern verwendet, welcher nach dem bei PISA verwendeten EGP-Klassenschema (vgl. Tabelle 3) bestimmt wurde. Dieser EGP-Wert, für den deskriptive Kennwerte in Tabelle 12 angegeben sind, war für 346 Schüler vorhanden. Wie ersichtlich lag der mittlere Rang der Schüler, die nicht in die Hauptstichprobe übernommen wurden unter dem mittleren Rang derer, die in die Hauptstichprobe übernommen wurden. Dies deutet auf einen niedrigeren durchschnittlichen EGP-Wert und damit auf einen höheren sozialen Status in

dieser Gruppe hin. Bei einem a priori Alpha-Niveau von .10 ergibt der Test allerdings kein signifikantes Ergebnis ($U = 13\ 896.50$, $p = .26$). Zumindest bezüglich der sozialen Herkunft haben die Reduktionen damit nicht zu signifikanten Gruppenunterschieden geführt.

Tabelle 12

Die Tabelle zeigt deskriptive Kennwerte zum EGP-Wert für in die Hauptstichprobe übernommene und nicht übernommene Schüler. Angegeben ist neben der Gruppenzugehörigkeit die Anzahl der Schüler, Minimum (Min.), Maximum (Max.), Median und der mittlere Rang (Rang_M) pro Gruppe.

Gruppe	Anzahl	Min.	Max.	Median	Rang_M
In Hauptstichprobe übernommene Schüler	163	1	7	3.00	179.75
Nicht in Hauptstichprobe übernommene Schüler	183	1	7	3.00	167.94

Die *Zielpopulation*, über die Aussagen gemacht werden sollen, sind die Schüler an niedersächsischen Gymnasien. Die *Erhebungsgesamtheit*, auf deren Grundlage die Stichprobe gezogen wurde, sind die Schüler der Schulen, die sich zu einer Zusammenarbeit bereit erklärt haben. Die Erhebungsgesamtheit ist damit kleiner als die Zielpopulation; es liegt *undercoverage* vor. Zudem handelt es sich nicht um eine *Zufallsstichprobe*, sondern um eine *Gelegenheitsstichprobe*, da an allen Schulen Daten erhoben wurden, an denen es möglich war.

3.4 Untersuchungsdurchführung

Die Datenerhebung fand im Mai und Juni 2005 statt. Die notwendigen Pilotuntersuchungen zum Test der verwendeten Materialien wurden im März und April 2005 durchgeführt. Lediglich die Pilotierung des C-Test Deutsch wurde im Zuge der ersten Datenerhebungen an Schule 1 vorgenommen (vgl. Abschnitt 3.2.4). Für alle Schüler der Hauptstichprobe gab es drei Erhebungszeitpunkte. Die ersten beiden umfassten 90 Minuten und der dritte 45 Minuten, was einer Doppelstunde Unterricht respektive einer einzelnen Unterrichtsstunde entspricht. Der zweite Erhebungszeitpunkt konnte auf zweimal 45 Minuten aufgeteilt werden. Alle Daten wurden in den Räumen der Schulen erhoben.

Vor der Datenerhebung zum ersten Erhebungszeitpunkt wurden die Schüler über die geplante Untersuchung informiert und um eine schriftliche Einverständniserklärung der Erziehungsberechtigten gebeten. Zum ersten Erhebungszeitpunkt bearbeiteten die Schüler den deutschen und englischen C-Test, sowie den KFT. Von diesem Test kamen nur die Subskalen

V3, Q2 und N2 der Testform A zum Einsatz (vgl. Abschnitt 3.2.3). Zudem füllten die Schüler den Schülerfragebogen aus. Die mono- und bilingualen Schüler bekamen dabei jeweils einen für sie zugeschnittenen Schülerfragebogen. In der gleichen Zeit wurde den Lehrern der Lehrerfragebogen vorgelegt. Zum zweiten Erhebungszeitpunkt, der etwa ein bis zwei Wochen nach dem ersten lag, bearbeiteten die Schüler den Geografietest. Die monolingualen Schüler bekamen die deutschsprachige und die bilingualen die englischsprachige Fassung. Zum dritten und letzten Erhebungszeitpunkt hatten die Schüler Gelegenheit, ihre eigene Lösung des Geografietests zu überarbeiten. Damit sollten metakognitive Fähigkeiten der Schüler untersucht werden, worauf in der vorliegenden Arbeit allerdings nicht näher eingegangen wird. Eine Übersicht zur Datenerhebung gibt Abbildung 5.

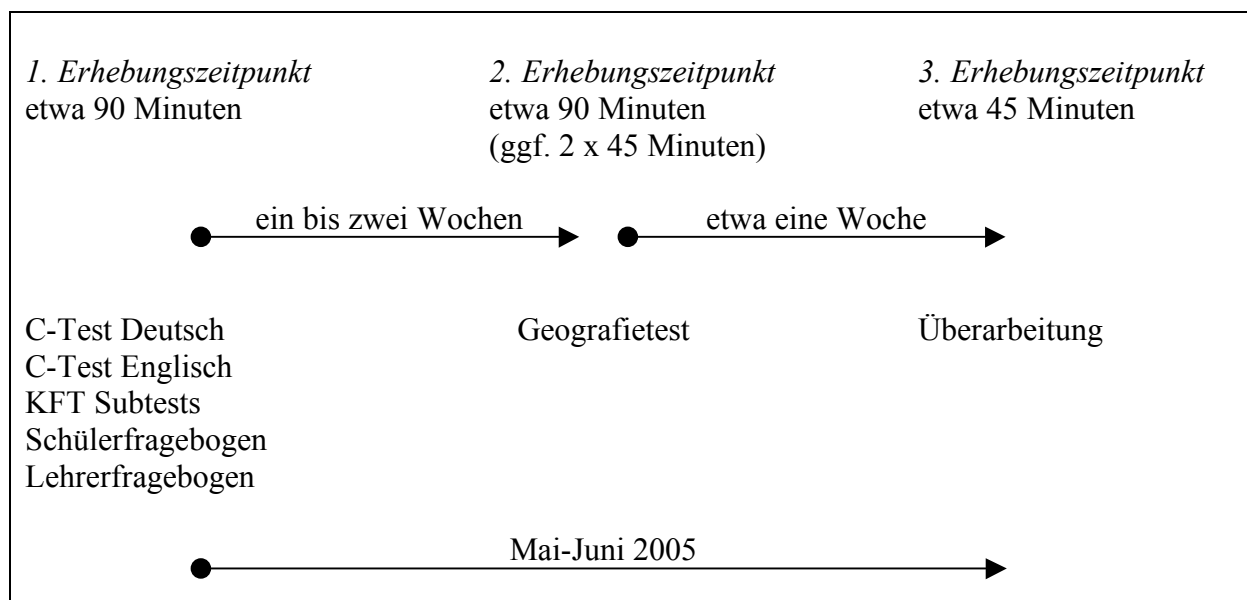


Abbildung 5 Übersichtsschema zur Durchführung der Datenerhebung.

Für die Datenerhebung der Hauptstichprobe sind noch einige Besonderheiten zu berichten. In Erhebungsgruppe 30 wurde der KFT in zwei Fassungen durchgeführt, einmal ohne und einmal mit Verbalteil V3, da dieser in der ersten Fassung der Schülermaterialien versehentlich fehlte. In Erhebungsgruppe 22 kam wiederum versehentlich ein fehlerhafter englischer C-Test zum Einsatz. Zudem sei nochmals darauf hingewiesen, dass in die Datenauswertung nur sechs der 17 Teilaufgaben des Geografietests eingehen, da bei den anderen Teilaufgaben von den

Schülern nicht genügend Text produziert wurde, um das Ergebnis sinnvoll beurteilen zu können. Die Noten der Schüler wurden direkt über die jeweilige Schule erhoben.

3.5 Datenauswertung

In diesem letzten Abschnitt des Methodenteils soll einerseits auf die Auswertung des Geografietests eingegangen werden, da dieser das zentrale Erhebungsinstrument des DFG-Projektes darstellt. Andererseits sollen kurz die Variablen vorgestellt werden, die in dieser Arbeit – besonders im Ergebnisteil – verwendet werden.

Als Arbeitsergebnisse der Schüler liegen deren schriftliche Antworten auf die Aufgaben im Geografietest vor. In den Sprachwissenschaften gibt es Vorgehensempfehlungen zur Auswertung solcher *Verschriftlichungen* (z.B. Weigle, 2002; Alderson, Clapham & Wall, 1995). Dabei werden eine oder mehrere Beurteilungsskalen konstruiert, die normalerweise vier bis neun Stufen umfassen und geschulten Beurteilern als Grundlage zur Beurteilung der Verschriftlichungen dienen. Bei PISA wurde in ähnlicher Weise vorgegangen (z.B. Artelt, Stanat, Schneider & Schiefele, 2001). Nach Weigle (2002) werden unter anderem holistische und analytische Beurteilungsskalen unterschieden.

Für die Auswertung des Geografietests sind in erster Linie die *holistischen Beurteilungsskalen* von Bedeutung. Dabei wird jeder Verschriftlichung in Abhängigkeit vom Gesamteindruck des Beurteilers ein einzelner Wert der Skala zugewiesen. Wichtig ist eine vorhandene Beurteilungsanweisung, eine Verankerung der Skala. Unter Umständen gibt es Beispieltex te für die Stufen der Skala. Vorteil der holistischen Beurteilung ist die Schnelligkeit der Durchführung. Zudem besteht die Möglichkeit, spezielle Aspekte effizient zu fokussieren, indem die Beurteilungsanweisung auf die gewünschten Aspekte ausgerichtet wird. Ein Nachteil ist die unter Umständen schwierige Interpretation der Skalenwerte, da die Beurteiler jeweils andere Gesichtspunkte fokussieren und trotzdem zu identischen Skalenwerten gelangen können.

Im Vorfeld der eigentlichen Beurteilungen ist es erforderlich, die Beurteiler angemessen zu schulen, damit sie die Beurteilungsskalen zuverlässig anwenden können. Über Mehrfachbeurteilungen der gleichen Verschriftlichungen durch verschiedene Beurteiler und die anschließende Überprüfung der Beurteilerübereinstimmung, kann sowohl die Güte der Beurteilung, als auch die Güte der Beurteilungsskala im Schulungsprozess und nach Abschluss der finalen Beurteilung überprüft werden.

Mathematische Verfahren zu Bestimmung dieser als Beurteilerübereinstimmung bezeichneten Größe gibt es verschiedene, je nach Untersuchungsplan und Skalenniveau. Übersichten finden sich beispielsweise bei Orwin (1994), sowie bei Asendorpf und Wallbott (1979). Die gängigsten Verfahren, die auch in der vorliegenden Arbeit verwendet werden, sind Kappa (κ ; Cohen, 1960; Conger, 1980; Fleiss & Cohen, 1973; Hubert, 1977) und die Intraklassenkorrelation (Rustenbach, 2003; Shrout & Fleiss, 1979).

In Anhang D ist ein Auszug aus der für die Auswertung des Geografietests verwendeten holistische Beurteilungsskala wiedergegeben. Sie wurde in englischer Sprache erstellt und sowohl auf die Verschriftlichungen der monolingualen, als auch auf die der bilingualen Schüler angewendet. Bei der holistischen Beurteilungsskala handelt es sich um eine siebenstufige verankerte Skala, wobei sechs die bestmögliche und null die schlechteste Beurteilung darstellt. Beurteilt wird der Grad der Aufgabenerfüllung oder genauer die Angemessenheit und Klarheit der Diskursfunktion.

Nach dieser Ausführung zur Auswertung des Geografietests soll nun noch auf die verwendeten Variablen eingegangen werden. In Tabelle 13 findet sich eine Übersicht über die Variablenbezeichnungen und die möglichen Ausprägungen der Variablen. Zudem ist aufgeführt, aus welcher Quelle die Daten stammen, welches Skalenniveau angenommen wird und welche Rolle die Variablen im Untersuchungsplan und in der Datenauswertung einnehmen.

Bei der Variablen „Beurteilung Geografietest“ handelt es sich um den Mittelwert der Beurteilungen über alle sechs Teilaufgaben, wobei die Beurteilung wie erwähnt auf Grundlage der holistischen Beurteilungsskala durchgeführt wurde. Bei der Variablen „Klassenklima“ handelt es sich um den Mittelwert über Items aus dem Schülerfragebogen, deren Auswahl über eine Itemanalyse (Abschnitt 4.1) festgelegt wird. Bei der Variablen „Note Erdkunde“ handelt es sich um den Median der Noten im Fach Erdkunde aus den letzten fünf Halbjahren. Bei „Note Deutsch“, „Note Englisch“ und „Note Mathematik“ handelt es sich um die Noten aus dem letzten Zeugnis. Bei den Variablen, die den KFT betreffen, liegen die Testrohwerte der einzelnen Schüler zu Grunde.

Tabelle 13

Die Tabelle gibt eine Übersicht über die in der vorliegenden Arbeit verwendeten Variablen. Angegeben sind die Variablenbezeichnungen, die möglichen Ausprägungen, die jeweilige Datenquelle und das angenommene Skalenniveau. In der letzten Spalte ist die Rolle der Variablen aufgeführt.

Variablenbezeichnung	mögliche Ausprägung	Quelle	Skalenniveau	Rolle [†]
Beurteilung Geografietest	0 bis 6	Beurteilung des Geografietest	intervall	AV, KR
Ergebnis C-Test Deutsch	0 bis 100	C-Test Deutsch	intervall	AV, KR
Ergebnis C-Test Englisch	0 bis 100	C-Test Englisch	intervall	AV, KR
Ergebnis KFT	0 bis 65	Kognitiver Fähigkeitstest	intervall	AV, KR
Ergebnis KFT (verbal)	0 bis 20	Subtest V3 des KFT	intervall	KR
Ergebnis KFT (quantitativ)	0 bis 20	Subtest Q2 des KFT	intervall	KR
Ergebnis KFT (nonverbal)	0 bis 25	Subtest N2 des KFT	intervall	KR
Klassenklima	1 bis 4	Skala im Schülerfragebogen	intervall	KR
Höchstes PISA-EGP	1 bis 7	Angaben im Schülerfragebogen	ordinal	AV, UV
Note Deutsch	1 bis 6	Information der Schulen	ordinal	KR
Note Englisch	1 bis 6	Information der Schulen	ordinal	KR
Note Mathematik	1 bis 6	Information der Schulen	ordinal	KR
Note Erdkunde	1 bis 6	Information der Schulen	ordinal	KR
Geschlecht	weiblich/ männlich	Frage im Schülerfragebogen	nominal	UV
Unterrichtstyp	mono-/ bilingual	Information der Schulen	nominal	UV

[†] *Bemerkung:* Bei der Rolle einer Variablen im Untersuchungsplan und in der Datenauswertung wird an dieser Stelle für Mittelwertsvergleiche zwischen unabhängiger Variablen (UV) und abhängiger Variablen (AV) unterschieden. Zudem sind Variablen gekennzeichnet, die als Prädiktor oder Kriterium bei Korrelation und Regression (KR) verwendet werden.

4 Ergebnisse

Im Folgenden sollen die Ergebnisse der durchgeführten Datenanalysen dargestellt werden. Im ersten Abschnitt wird über vorangehende Analysen berichtet. Dazu gehören die Ergebnisse zur Beurteilerübereinstimmung beim Geografietest und die Analysen der Skala zur Erfassung des Klassenklimas. In den weiteren Abschnitten werden dann sukzessive die Ergebnisse zu den Hypothesen aus Abschnitt 2.7 vorgestellt. Alle Analysen wurden mit dem Statistikprogramm SPSS 12.0.2G durchgeführt und in einer SPSS Syntaxdatei dokumentiert.

Wenn nicht anders berichtet, wurden zweiseitige statistische Tests angewendet. Das a priori Alpha-Niveau bei statistischen Hypothesentests wurde in Abhängigkeit von der jeweiligen Hypothese festgelegt. Bei Voraussetzungsprüfungen und anderen Hypothesen, welche die inhaltlich relevante Aussage in der Nullhypothese haben, beträgt es .10, da so das Risiko eines Beta-Fehlers – beispielsweise die Entscheidung für die Nullhypothese der Varianzhomogenität obwohl keine Varianzhomogenität vorliegt – geringer gehalten wird. In allen anderen Fällen ist das a priori Alpha-Niveau auf .05 festgelegt. Bei nicht signifikanten Ergebnissen und Darstellung der Ergebnisse in Tabellenform wird der genaue p -Wert angegeben. Die verwendeten Analyseverfahren gelten unter vielen Bedingungen als robust gegen Verletzungen (Bortz, 1999). Daher wurden nur gängige Voraussetzungsprüfungen vorgenommen. Detaillierte Analysen zur Verteilung der Daten wurden nur für den KFT durchgeführt. Die Ergebnisse dazu sind in Anhang E dargestellt.

4.1 Vorangehende Analysen

In diesem Abschnitt sollen die Ergebnisse von Analysen dargestellt werden, die vor der eigentlichen Datenanalyse durchgeführt werden mussten. Dazu zählen die Beurteilerübereinstimmung beim Geografietest und die Analysen der Skala zur Erfassung des Klassenklimas. Ohne dies vorab zu klären kann die Güte der Daten zur Geografiekompetenz und zum Klassenklima nicht eingeschätzt werden.

In Tabelle 14 finden sich die Ergebnisse zur Beurteilerübereinstimmung beim Geografietest. Neben den Werten von Cohens Kappa sind noch weitere Statistiken angegeben: Die Intraklassenkorrelation wird aufgeführt, da Cohens Kappa nicht in allen Fällen berechnet werden konnte. Weiterhin sind zum Vergleich Werte für zwei Korrelationskoeffizienten angegeben. Zu beachten ist, dass auf Grund des hohen Aufwandes und begrenzter personeller

Ressourcen leider nur etwa 20% der Verschriftlichungen der Schüler bei jeder Teilaufgabe von zwei Beurteilern bearbeitet werden konnten. Dieser Anteil erscheint allerdings ausreichend, da so immer etwa 40 Schülerantworten doppelt beurteilt wurden.

Tabelle 14

In der Tabelle finden sich die Ergebnisse zur Analyse der Beurteilerübereinstimmung für die holistische Beurteilungsskala des Geografietests. Neben Anzahl und Anteil der Doppelurteile sind Cohens Kappa (κ), die Intraklassenkorrelation (ICC), sowie die Korrelation nach Spearman (r_s) und nach Pearson (r_p) aufgeführt.

Teilaufgabe	Anzahl der Doppelurteile	Anteil an Hauptstichprobe	κ	ICC*	r_s	r_p
1c	39	22%	- [†]	.93	.94	.94
2a	40	23%	- [†]	.93	.93	.93
3	40	23%	- [†]	.93	.95	.94
6	41	24%	.93	.97	.97	.97
7	40	23%	.71	.88	.87	.88
8	40	23%	.89	.96	.95	.96

[†] Die höchste Beurteilungskategorie wurde von einem Beurteiler nicht ausgenutzt. Somit konnte keine symmetrische Datentabelle gebildet und der Wert für Cohens Kappa nicht berechnet werden.

* Bei der berichteten Intraklassenkorrelation handelt es sich um den Koeffizienten absoluter Übereinstimmung im gemischten Zwei-Wege Modell (absolute agreement, two-way mixed model).

Zur Interpretation der Ergebnisse nach der Berechnung von Kappa und der Intraklassenkorrelation werden in der Literatur Anhaltspunkte gegeben. Rustenbach (2003) verweist darauf, dass die Beurteilerübereinstimmung „stets nur als eine obere Grenze der Güte einer Übereinstimmung gelten“ (ebd., S. 51f) kann. Orwin (1994) berichtet Daumenregeln, wonach Werte unter .40 als schlecht, von .40 bis .59 als brauchbar, von .60 bis .74 als gut und größer als .75 als exzellent angesehen werden können. Er weist gleichzeitig darauf hin, dass diese Daumenregeln aus dem Kontext der Diagnostik von Psychopathologie stammen. Es ist unklar, ob sie in anderen Kontexten verwendet werden sollten. Ungeachtet dieser letzten Punkte sprechen die Ergebnisse in Tabelle 14 für eine ausreichend hohe Beurteilerübereinstimmung. Da beide Korrelationskoeffizienten etwa gleiche Ausprägungen erreichen, kann weiterhin davon ausgegangen werden, dass bei den Beurteilungen kaum Ausreißer vorhanden sind. Die Beurteilungsergebnisse des Geografietests können somit in der Datenanalyse verwendet werden.

Da das im Rahmen des DFG-Projektes verwendete Verfahren zur Erfassung des Klassenklimas eine selbst konstruierte Skala darstellt, die vorher noch nie im Einsatz war, ist eine Itemanalyse angezeigt. Die Skala, die sich aus sieben Items zusammensetzt, ist in Anhang A als Teil des Schülerfragebogens abgedruckt. Die einzelnen Items verfügen über eine

vierstufige Likertskala von eins bis vier, welche die Verankerungen „stimme nicht zu“, „stimme eher nicht zu“, „stimme eher zu“ und „stimme zu“ aufweist. Im ersten Schritt wurden alle sieben Items einer Itemanalyse unterzogen. Dabei ergaben sich die in Tabelle 15 dargestellten Ergebnisse.

Tabelle 15

In dieser Tabelle finden sich Itemnummern, Itemstamm und wichtige deskriptive Statistiken zu den sieben Items der Skala zum Klassenklima. Die Abkürzung „SFB“ steht für „Schülerfragebogen“. In die Berechnungen gingen die Daten von 161 Schülern ein.

Item	Item im SFB	Itemstamm	<i>M</i>	<i>SD</i>	korrigierte Trennschärfe
1	129	In unserer Klasse ist der Zusammenhalt gut.	3.03	0.93	.66
2	130	In unserer Erdkunde Klasse ist der Zusammenhalt gut.	3.03	0.86	.68
3 [†]	131	In unserer Klasse fühle ich mich als nicht dazugehörig und als Außenseiter.	3.60	0.74	.27
4	132	In unserer Erdkunde Klasse fühle ich mich wohl.	3.28	0.70	.51
5 [†]	133	Auf eine gute Klassengemeinschaft kommt es mir nicht an.	3.48	0.82	.19
6 [†]	134	In unserer Erdkunde Klasse arbeiten die einzelnen Schüler eher gegeneinander als miteinander.	3.50	0.68	.32
7 [†]	135	In unserer Erdkunde Klasse bekommt man schnell Probleme, wenn man besonders gute (oder: relativ schlechte) Leistungen zeigt.	3.63	0.64	.20

[†] *Bemerkung:* Die Rohwerte dieser Items wurden jeweils durch Subtraktion der Ziffer fünf invertiert.

Im zweiten Schritt wurden die Items entfernt, die eine korrigierte Trennschärfe kleiner oder gleich .20 aufwiesen, da dieser Wert als zu gering angesehen wurde. Dies betraf zwei Items. Die wichtigsten Daten zu den verbleibenden Items sind in Tabelle 16 dargestellt.

Für die nun aus fünf Items bestehende Skala ergibt sich eine Reliabilität nach Cronbach in Höhe von $\alpha = .74$, bei einem Skalenmittelwert von $M = 16.46$ ($SD = 2.75$). Dieses Ergebnis ist nicht optimal, zeigt aber eine ausreichende Reliabilität für die Verwendung der Ergebnisse zum Klassenklima in der Datenanalyse. Wie in Tabelle 17 dargestellt, würde sich die

Reliabilität der Skala durch die Elimination weiterer Items nicht bedeutend verbessern. Daher wurden keine weiteren Items entfernt.

Tabelle 16

In dieser Tabelle finden sich Itemnummern, Itemstamm und wichtige deskriptive Statistiken zu den fünf verbleibenden Items der Skala zum Klassenklima. Die Abkürzung „SFB“ steht für „Schülerfragebogen“. In die Berechnungen gingen die Daten von 163 Schülern ein.

Item	Item im SFB	Itemstamm	<i>M</i>	<i>SD</i>	korrigierte Trennschärfe
1	129	In unserer Klasse ist der Zusammenhalt gut.	3.04	0.93	.71
2	130	In unserer Erdkunde Klasse ist der Zusammenhalt gut.	3.04	0.86	.74
3 [†]	131	In unserer Klasse fühle ich mich als nicht dazugehörig und als Außenseiter.	3.60	0.73	.24
4	132	In unserer Erdkunde Klasse fühle ich mich wohl.	3.28	0.71	.53
5 [†]	134	In unserer Erdkunde Klasse arbeiten die einzelnen Schüler eher gegeneinander als miteinander.	3.50	0.68	.31

[†] *Bemerkung:* Die Rohwerte dieser Items wurden jeweils durch Subtraktion der Ziffer fünf invertiert.

Tabelle 17

In dieser Tabelle finden sich Mittelwert, Standardabweichung und Cronbachs Alpha der Skala für den Fall, dass das jeweilige Item entfernt würde.

Item	<i>M</i>	<i>SD</i>	Cronbachs α
1	20.52	2.53	.59
2	20.52	2.57	.59
3 [†]	19.95	2.94	.78
4	20.27	2.80	.68
5 [†]	20.06	2.93	.75

[†] *Bemerkung:* Die Rohwerte dieser Items wurden jeweils durch Subtraktion der Ziffer fünf invertiert.

Zur genaueren Untersuchung der Skala zum Klassenklima wurde im letzten Schritt eine *Faktorenanalyse im engeren Sinne* (common factor analysis; SPSS: Hauptachsen-Faktorenanalyse) durchgeführt. Der anfängliche Verlauf der Eigenwerte ist in Abbildung 6 dargestellt. Nach dem Scree-Test sollte ein Faktor extrahiert werden. Nach dem Kaiser-Kriterium

könnten zwei Faktoren extrahiert werden, wobei der zweite Faktor einen Eigenwert nur wenig über 1.00 aufweisen würde.

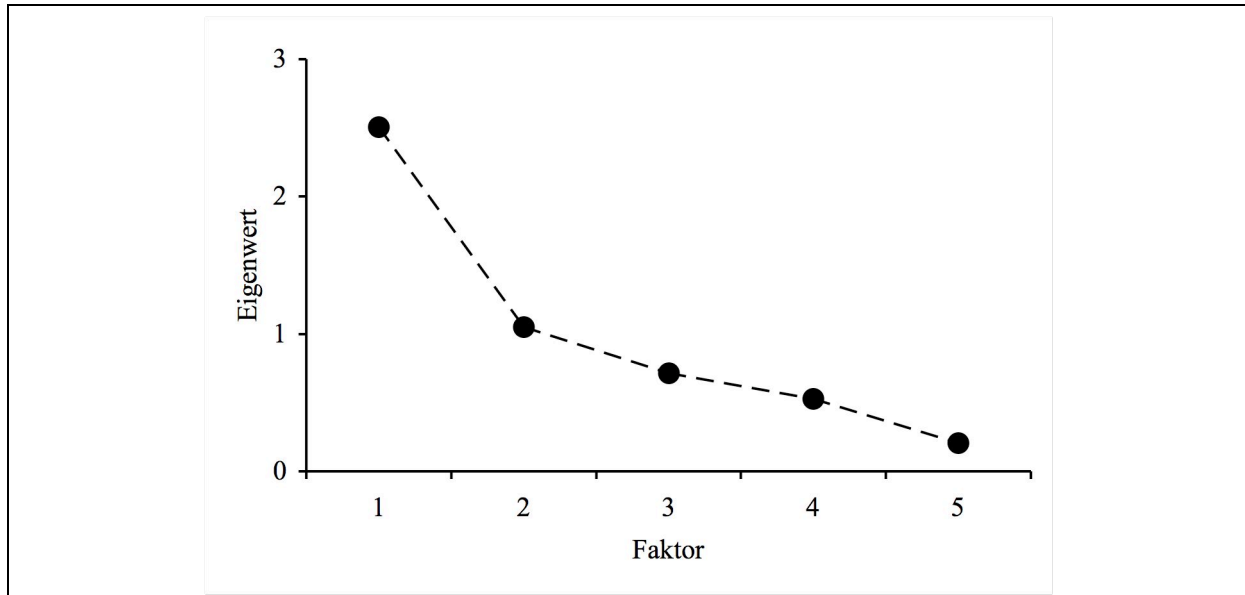


Abbildung 6 Anfänglicher Verlauf der Eigenwerte (Eigenwertediagramm, Screeplot).

Bei der Faktorenanalyse wurden zwei Faktoren extrahiert. Die Ergebnisse dazu sind in Tabelle 18 wiedergegeben. Dort zeigt sich, dass alle Items auf den ersten Faktor hoch laden. Lediglich bei Item 5 tritt eine negative und geringfügig höhere Ladung auf dem zweiten Faktor auf. Dieser besitzt einen sehr geringen Eigenwert und klärt im Vergleich zum ersten Faktor wenig Varianz auf. Insgesamt deuten die Ergebnisse der Faktorenanalyse damit auf einen Faktor hin, der den Items zu Grunde liegt und mit *Klassenklima* bezeichnet werden könnte. Das spricht – wie schon die Ergebnisse der Itemanalyse – für den Einsatz der Skala zum Klassenklima im Rahmen der Datenanalyse. Dabei wird der Mittelwert über die fünf Items der Skala unter der Variablenbezeichnung „Klassenklima“ verwendet.

Tabelle 18

Die Tabelle zeigt die Ergebnisse der mit den Items der Skala zum Klassenklima durchgeführten Faktorenanalyse. Neben den Ladungen der Items auf die Faktoren sind die Kommunalitäten der Items, die Eigenwerte der Faktoren und die erklärte Varianz angegeben.

Variable	Faktor 1	Faktor 2	Kommunalität h^2
Item 1	.90	-.28	.88
Item 2	.87	.04	.76
Item 3	.63	.31	.18
Item 4	.37	.30	.49
Item 5	.29	-.31	.23
Eigenwert λ	2.18	0.36	
Erklärte Varianz	43.61%	7.16%	
Erklärte Varianz (kumuliert)	43.61%	50.77%	

4.2 Mono- und bilinguale Schüler

Zuerst sollen die Ergebnisse zu den ersten fünf Hypothesen berichtet werden, da diese die zentralen Fragestellungen zu mono- und bilingualen Schülern betreffen. Die Variable „Unterrichtstyp“ dient dabei als unabhängige Variable. Die erste Hypothese lautet: „Die bilingualen Schüler haben im Vergleich zu den monolingualen Schülern eine größere Geografiekompetenz“. Für die Gruppe der monolingualen Schüler zeigte sich ein höherer Mittelwert in der Variablen „Beurteilung Geografietest“ ($M = 2.59$, $SD = 0.56$, $N = 90$), als für die Gruppe der bilingualen Schüler ($M = 2.35$, $SD = 0.71$, $N = 84$). Der Welch-Test ergab mit $t(157) = 2.47$, $p < .05$ ein signifikantes Ergebnis. Die Effektgröße lag bei $d = 0.38$ ($SD_{\text{gepooit}} = 0.64$) und spricht für einen kleinen Effekt. Ein praktisch bedeutsamer Effekt würde bestehen, wenn sich die Mittelwerte in „Beurteilung Geografietest“ um eine Stufe der Beurteilungsskala unterschieden. Unter Annahme der gleichen Stichprobenstreuung wäre dies bei einer Effektgröße von $d = 1.56$ der Fall. Damit muss die Hypothese verworfen werden. Entgegen der Annahme, dass bilinguale Schüler eine größere Geografiekompetenz aufweisen als monolinguale Schüler, muss vielmehr davon ausgegangen werden, dass sich beide Gruppen nicht bedeutsam hinsichtlich ihrer Geografiekompetenz unterscheiden.

Die zweite Hypothese lautet: „Die mono- und bilingualen Schüler unterscheiden sich nicht signifikant bezüglich ihrer kognitiven Fähigkeiten“. Die monolingualen Schüler zeigten in der Variablen „Ergebnis KFT“ einen geringeren Mittelwert ($M = 44.36$, $SD = 7.98$, $N = 86$) als die bilingualen Schüler ($M = 46.51$, $SD = 7.75$, $N = 79$). Der durchgeführte Welch-Test ergab $t(162) = -1.75$, $p < .10$ und damit ein signifikantes Ergebnis bei einer Effektgröße von

$d = 0.27$ ($SD_{\text{gepoolt}} = 7.87$). Dieser kleine Effekt zeigt, dass sich die beiden Gruppen hinsichtlich „Ergebnis KFT“ weniger als eine halbe Standardabweichung unterscheiden. Daher wird trotz des signifikanten Testergebnisses nicht von bedeutsamen Gruppenunterschieden ausgegangen. Die Hypothese, dass sich die mono- und bilingualen Schüler hinsichtlich ihrer kognitiven Fähigkeiten nicht unterscheiden, kann somit beibehalten werden.

Die dritte Hypothese – „die mono- und bilingualen Schüler unterscheiden sich nicht signifikant bezüglich ihrer sozialen Herkunft“ – wurde mittels eines Mann-Whitney-U-Tests untersucht. In der Variablen „Höchstes PISA-EGP“ zeigte sich für die monolingualen Schüler ein größerer mittlerer Rang (89.62, $N = 84$) als für die bilingualen Schüler (73.90, $N = 79$). Der höhere mittlere Rang deutet dabei auf eine tendenziell höhere Stufe im EGP-Klassenschema für monolinguale Schüler hin. Beim Test ergab sich $U = 2\,678.00$, $p < .10$ und damit ein signifikanter Effekt. Die Hypothese muss verworfen werden. Es scheint so zu sein, dass sich die mono- und bilingualen Schüler entgegen der Annahme hinsichtlich ihrer sozialen Herkunft unterscheiden, wobei die bilingualen Schüler „aus besserem Hause“ stammen.

Da die beiden Gruppen unter diesen Bedingungen nur eingeschränkt hinsichtlich der Geografiekompetenz verglichen werden können, wurde zur Abschätzung des Zusammenhangs zwischen „Beurteilung Geografietest“ und „Höchstes PISA-EGP“ explorativ der Korrelationskoeffizient nach Spearman bestimmt. Dabei ergab sich mit $r_s = -.06$ inferenzstatistisch eine insignifikante Korrelation ($p = .45$, $N = 163$). Aufgeteilt für beide Gruppen der Variablen „Unterrichtstyp“ ergab sich für die monolingualen Schüler $r_s = -.13$ ($p = .25$, $N = 84$) und für die bilingualen Schüler $r_s = -.06$ ($p = .61$, $N = 79$). Auch wenn beide Korrelationen insignifikant ausfielen, zeigte sich für die monolingualen Schüler ein geringfügig größerer Zusammenhang. Obwohl sich die mono- und bilingualen Schüler in ihrer sozialen Herkunft unterscheiden, kann auf Grund der berichteten explorativen Analysen davon ausgegangen werden, dass der Einfluss der sozialen Herkunft auf die Geografiekompetenz gering ist.

Die vierte Hypothese – „die bilingualen Schüler haben eine größere Sprachkompetenz in Englisch als die monolingualen Schüler“ – wurde wiederum mittels eines Welch-Tests untersucht. In der Variablen „Ergebnis C-Test Englisch“ zeigten die monolingualen Schüler einen deutlich geringeren Mittelwert ($M = 46.52$, $SD = 13.70$, $N = 86$) als die bilingualen Schüler ($M = 60.69$, $SD = 12.55$, $N = 80$). Es ergab sich $t(163) = -6.95$, $p < .05$, bei einer Effektgröße von $d = 1.08$ ($SD_{\text{gepoolt}} = 13.16$). Die Hypothese wird somit nicht verworfen. Bilinguale Schüler zeigen ganz deutlich größere Sprachkompetenz in Englisch.

Die fünfte Hypothese betrifft die Sprachkompetenz Deutsch. Es wird Folgendes angenommen: „Die mono- und bilingualen Schüler unterscheiden sich nicht signifikant bezüglich ihrer Sprachkompetenz in Deutsch“. Deskriptiv zeigt die Gruppe der monolingualen Schüler einen geringeren Mittelwert ($M = 80.98$, $SD = 9.93$, $N = 89$) als die bilingualen Schüler ($M = 85.06$, $SD = 8.09$, $N = 81$). Beim Welch-Test ergab sich $t(166) = -2.95$, $p < .10$ und eine Effektgröße von $d = 0.45$ ($SD_{\text{gepooled}} = 9.10$). Auf Grund des signifikanten Testergebnisses und der mittleren Effektgröße wird die Hypothese zurückgewiesen. Entgegen der Annahme, dass sich beide Gruppen nicht unterscheiden, zeigt sich für bilinguale Schüler eine größere Sprachkompetenz in Deutsch.

4.3 Schulleistungen

Nachdem die Unterschiede zwischen mono- und bilingualen Schülern im letzten Abschnitt behandelt worden sind, sollen jetzt Zusammenhänge zwischen den Schulleistungen der Schüler und den vier erhobenen Kompetenzvariablen „Beurteilung Geografietest“, „Ergebnis C-Test Englisch“, „Ergebnis C-Test Deutsch“ und „Ergebnis KFT“ untersucht werden.

Hypothese 6 betrifft die Geografiekompetenz: „Die Schulleistungen in Erdkunde hängen positiv mit der Geografiekompetenz zusammen“. Zur Untersuchung dieses Zusammenhangs wurde der Korrelationskoeffizient nach Spearman zwischen den Variablen „Note Erdkunde“ und „Beurteilung Geografietest“ berechnet. Es ergab sich $r_s = -.24$, wobei diese Korrelation im inferenzstatistischen Test signifikant wurde ($p < .05$, $N = 174$). Damit wird die Hypothese nicht verworfen; zwischen Schulleistungen in Erdkunde und der Geografiekompetenz zeigen sich bedeutsame positive Zusammenhänge⁴.

Hypothese 7 betrifft die Sprachkompetenz Englisch: „Die Schulleistungen in Englisch hängen positiv mit der Sprachkompetenz in Englisch zusammen“. Zwischen den Variablen „Note Englisch“ und „Ergebnis C-Test Englisch“ ergab sich $r_s = -.46$, wobei der inferenzstatistische Test wiederum signifikant ausfiel ($p < .05$, $N = 166$). Die Hypothese wird auf Grundlage der Ergebnisse nicht verworfen. Vielmehr wird davon ausgegangen, dass die Schulleistungen in Englisch positiv mit der Sprachkompetenz in Englisch zusammenhängen.

⁴ Trotz des negativen Korrelationskoeffizienten besteht ein positiver Zusammenhang zwischen beiden Variablen, da mit besserer Schulleistung in Erdkunde auch ein besseres Abschneiden im Geografietest einhergeht. Das negative Vorzeichen entsteht lediglich dadurch, dass eine gute Leistung als Note durch eine kleine Zahl repräsentiert wird. Dies gilt auch für die im Folgenden berichteten Analysen mit Schulleistungen.

Hypothese 8 betrifft, wie schon Hypothese sieben, den Bereich der Sprachkompetenz. In diesem Fall geht es um die Sprachkompetenz Deutsch: „Die Schulleistungen in Deutsch hängen positiv mit der Sprachkompetenz in Deutsch zusammen“. Zwischen den Variablen „Note Deutsch“ und „Ergebnis C-Test Deutsch“ ergab sich ein korrelativer Zusammenhang von $r_s = - .41$, der wiederum signifikant ausfiel ($p < .05$, $N = 170$). Damit wird die Hypothese beibehalten und davon ausgegangen, dass die Schulleistungen in Deutsch positiv mit der Sprachkompetenz in Deutsch zusammenhängen.

Hypothese 9 betrifft den Zusammenhang zwischen den kognitiven Fähigkeiten und den Schulleistungen in verschiedenen Fächern. Für jedes der Schulfächer wurde dazu eine Teilhypothese gebildet: „Die kognitiven Fähigkeiten hängen positiv mit den Schulleistungen in Deutsch (Teilhypothese A), Englisch (Teilhypothese B), Mathematik (Teilhypothese C) und Erdkunde (Teilhypothese D) zusammen“. Zur Untersuchung der Hypothese wurden vier Korrelationen mit der Variablen „Ergebnis KFT“ berechnet. Die Ergebnisse dazu finden sich in Tabelle 19. Dort sind zusätzlich die Interkorrelationen der Notenvariablen angegeben.

Tabelle 19

Die Tabelle enthält die bivariaten Korrelationen nach Spearman zwischen der Variablen „Ergebnis KFT“ einerseits und den vier Variablen mit den Noten in Deutsch, Englisch, Mathematik und Erdkunde andererseits. Alle angegebenen Korrelationen sind im zweiseitigen Test bei $p < .05$ signifikant. In die Berechnungen zu „Ergebnis KFT“ gingen die Daten von 165 und bei den anderen Variablen die Daten von 174 Schülern ein.

Variable	Ergebnis KFT	Note Deutsch	Note Englisch	Note Mathematik	Note Erdkunde
Ergebnis KFT		-.21	-.31	-.34	-.26
Note Deutsch	-.21		.59	.42	.56
Note Englisch	-.31	.59		.49	.47
Note Mathematik	-.34	.42	.49		.46
Note Erdkunde	-.26	.56	.47	.46	

Wie die Ergebnisse in Tabelle 19 zeigen, hängt „Ergebnis KFT“ mit allen vier Notenvariablen signifikant zusammen. Hypothese 9 kann mit allen vier Teilhypothesen beibehalten werden. Die kognitiven Fähigkeiten hängen positiv mit den Schulleistungen in Deutsch, Englisch, Mathematik und Erdkunde zusammen. Auffällig ist, dass die Korrelationen zwischen den vier Notenvariablen durchweg höher ausfallen, als ihre jeweiligen Korrelationen mit „Ergebnis

KFT“. Noten untereinander haben damit mehr gemeinsam als Noten und ein externes Kriterium.

4.4 Soziale Herkunft

In diesem Abschnitt soll Hypothese 10 behandelt werden. Sie betrifft Veränderungen in den vier Kompetenzvariablen Geografiekompetenz, Sprachkompetenz Deutsch und Englisch, sowie den kognitiven Fähigkeiten, in Abhängigkeit von der sozialen Herkunft der Schüler: „Bei positiverer Beurteilung der sozialer Herkunft der Schüler zeigen sich eine größere Geografiekompetenz (Teilhypothese A), eine größere Sprachkompetenz in Deutsch (Teilhypothese B), eine größere Sprachkompetenz in Englisch (Teilhypothese C) und größere kognitive Fähigkeiten (Teilhypothese D)“.

Da die Vorhersagen für alle vier Teilhypothesen in die gleiche Richtung gehen und Zusammenhänge zwischen den vier Kompetenzvariablen nicht ausgeschlossen werden können, soll zur Untersuchung von Hypothese 10 eine multivariate Varianzanalyse durchgeführt werden. Als abhängige Variablen dienen „Beurteilung Geografietest“, „Ergebnis C-Test Deutsch“, „Ergebnis C-Test Englisch“ und „Ergebnis KFT“. Unabhängige Variable ist „Höchstes PISA-EGP“. In Tabelle 20 finden sich deskriptive Statistiken für alle abhängigen Variablen, jeweils getrennt nach den sieben Stufen der unabhängigen Variablen.

Für den Box-M-Test auf Gleichheit der Kovarianzmatrizen ergab sich ein Wert von 79.41 mit $F(60, 8\ 907.93) = 1.17, p = .17$. Damit kann von Gleichheit der Kovarianzmatrizen ausgegangen werden. Im anschließenden multivariaten Gesamtttest ergab sich $\Lambda = .83, F(24, 510.54) = 1.13, p = .30$ und eine Varianzaufklärung von $\eta^2 = 4\%$. Ein multivariater Effekt liegt bei diesem Ergebnis nicht vor. Tabelle 21 gibt eine Übersicht über die im Anschluss durchgeführten univariaten Varianzanalysen für die abhängigen Variablen und die zugehörigen Levene-Tests auf Gleichheit der Fehlervarianzen.

Die Ergebnisse zeigen, dass für keine abhängige Variable ein signifikanter Effekt vorhanden ist. Daher wurde auf eine Diskriminanzanalyse als Anschlussverfahren verzichtet. Insgesamt zeigt sich inferenzstatistisch für keine der vier Kompetenzvariablen ein bedeutender Unterschied zwischen den Stufen des EGP-Klassenschemas. Auch deskriptiv verteilen sich die höchsten Ausprägungen für die Werte der Kompetenzvariablen auf vier verschiedene Stufen im EGP-Klassenschema (vgl. die kursiven Werte in Tabelle 20). Lediglich die letzten beiden Stufen im Schema fallen durch im Vergleich zu den anderen Stufen geringere

Mittelwerte auf. Die Ergebnisse sprechen dafür, die Hypothese zu verwerfen. Entgegen der ursprünglichen Annahmen geht die Einstufung in eine höhere Kategorie des EGP-Klassenschemas nicht mit größerer Geografiekompetenz, größerer Sprachkompetenz in Deutsch und Englisch sowie größeren kognitiven Fähigkeiten einher.

Tabelle 20

Die Tabelle enthält deskriptive Statistiken für die in der multivariaten Varianzanalyse zur Untersuchung von Hypothese 10 relevanten Variablen. Angegeben sind Faktorstufen, abhängige Variablen, Anzahl der Schüler, Mittelwert und Standardabweichung. Insgesamt gingen die Daten von 156 Schülern in die Berechnungen ein. Der über alle Faktorstufen höchste Wert jeder Kompetenzvariablen ist kursiv gesetzt.

Faktorstufe	abhängige Variable	Anzahl	<i>M</i>	<i>SD</i>
Höchstes PISA-EGP = 1	Beurteilung Geografietest	36	2.47	0.73
	Ergebnis C-Test Deutsch		85.36	8.23
	Ergebnis C-Test Englisch		56.92	17.84
	Ergebnis KFT		46.44	8.67
Höchstes PISA-EGP = 2	Beurteilung Geografietest	31	2.50	0.62
	Ergebnis C-Test Deutsch		86.10	7.02
	Ergebnis C-Test Englisch		57.23	13.13
	Ergebnis KFT		46.06	8.50
Höchstes PISA-EGP = 3	Beurteilung Geografietest	42	2.53	0.55
	Ergebnis C-Test Deutsch		81.81	9.78
	Ergebnis C-Test Englisch		49.90	11.70
	Ergebnis KFT		44.50	7.63
Höchstes PISA-EGP = 4	Beurteilung Geografietest	12	2.76	0.79
	Ergebnis C-Test Deutsch		83.92	10.66
	Ergebnis C-Test Englisch		50.33	21.88
	Ergebnis KFT		45.83	8.22
Höchstes PISA-EGP = 5	Beurteilung Geografietest	12	2.51	0.52
	Ergebnis C-Test Deutsch		83.08	8.36
	Ergebnis C-Test Englisch		58.00	13.32
	Ergebnis KFT		46.08	7.89
Höchstes PISA-EGP = 6	Beurteilung Geografietest	11	1.99	0.77
	Ergebnis C-Test Deutsch		79.36	7.76
	Ergebnis C-Test Englisch		50.18	13.22
	Ergebnis KFT		43.36	6.27
Höchstes PISA-EGP = 7	Beurteilung Geografietest	12	2.18	0.69
	Ergebnis C-Test Deutsch		78.83	10.61
	Ergebnis C-Test Englisch		45.50	9.78
	Ergebnis KFT		42.83	7.59

Tabelle 21

Die Tabelle zeigt Ergebnisse der im Anschluss an die multivariate Varianzanalyse zur Untersuchung von Hypothese 10 durchgeführten univariaten Varianzanalysen und der Levene-Tests. Angegeben sind Freiheitsgrade, F -Statistik und p -Wert. Für die Varianzanalysen ist zusätzlich die Varianzaufklärung (η^2) aufgeführt.

abhängige Variable	univariate Varianzanalyse					Levene-Test			
	df_1	df_2	F	p	η^2	df_1	df_2	F	p
Beurteilung Geografietest	6	149	1.84	.10	7%	6	149	1.11	.36
Ergebnis C-Test Deutsch	6	149	1.90	.08	7%	6	149	0.63	.70
Ergebnis C-Test Englisch	6	149	2.05	.06	8%	6	149	4.15	.00
Ergebnis KFT	6	149	0.56	.76	2%	6	149	0.18	.98

4.5 Klassenklima

Bevor die erhobenen Daten zum Klassenklima den hier berichteten Analysen zugeführt werden konnten, wurde die in Abschnitt 4.1 berichtete Itemanalyse durchgeführt, um die Reliabilität der Skala zum Klassenklima abzusichern. Annahmen zum Klassenklima finden sich in Hypothese 11: „Das Klassenklima hängt positiv mit den Schulleistungen in Deutsch (Teilhypothese A), Englisch (Teilhypothese B), Mathematik (Teilhypothese C) und Erdkunde (Teilhypothese D), sowie der Geografiekompetenz (Teilhypothese E) zusammen“. Die Ergebnisse der berechneten Zusammenhangsmaße für die Teilhypothesen A bis D finden sich in Tabelle 22.

Die Korrelationen nach Spearman in Tabelle 22 zeigen, dass keine der Notenvariablen signifikant mit dem Klassenklima zusammenhängt. Auffällig ist, dass sich nur bei der Variablen „Note Mathematik“ ein negativer Korrelationskoeffizient findet. Ein solcher war für alle Notenvariablen erwartet worden. Zur Untersuchung von Teilhypothese E wurde die Korrelation nach Pearson zwischen „Klassenklima“ und „Beurteilung Geografietest“ berechnet. Dabei ergab sich $r_p = .08$, wobei die inferenzstatistische Prüfung ein insignifikantes Ergebnis erbrachte ($p = .03$, $N = 166$). Alles in allem findet sich damit kein bedeutsamer Zusammenhang zwischen der Variablen „Klassenklima“ und einer der anderen Variablen. Die Hypothese 11 wird mit allen Teilhypothesen zurückgewiesen. Entgegen der Annahmen findet sich kein Zusammenhang zwischen dem Klassenklima auf der einen und den Schulleistungen sowie der Geografiekompetenz auf der anderen Seite.

Tabelle 22

Die Tabelle enthält die bivariaten Korrelationen nach Spearman zwischen der Variablen „Klassenklima“ und den Noten in Deutsch, Englisch, Mathematik und Erdkunde. Für die Variable „Klassenklima“ gingen die Daten von 166 Schülern in die Berechnungen ein. Bei den übrigen Variablen bestand die Datenbasis aus 174 Schülern.

Variable	Klassen- klima	Note Deutsch	Note Englisch	Note Mathematik	Note Erdkunde
Klassenklima		.01	.07	-.10	.07
Note Deutsch	.01		.59*	.42*	.56*
Note Englisch	.07	.59*		.49*	.47*
Note Mathematik	-.10	.42*	.49*		.46*
Note Erdkunde	.07	.56*	.47*	.46*	

* *Bemerkung:* Die so markierten Korrelationskoeffizienten sind im zweiseitigen Test bei $p < .05$ signifikant.

4.6 Geschlechtsunterschiede

Die beiden letzten Hypothesen betreffen Geschlechtsunterschiede. Die Variable „Geschlecht“ ist daher bei den folgenden Analysen unabhängige Variable. In Hypothese 12 wird Folgendes angenommen: „Schülerinnen zeigen im Vergleich zu Schülern eine größere Sprachkompetenz in Deutsch (Teilhypothese A) und in Englisch (Teilhypothese B)“. Zur Untersuchung beider Teilhypothesen wurde jeweils der Welch-Test verwendet.

Für Teilhypothese A ergab sich in der Variablen „Ergebnis C-Test Deutsch“ ein höherer Mittelwert für die Schülerinnen ($M = 83.03$, $SD = 9.41$, $N = 106$) gegenüber den Schülern ($M = 82.75$, $SD = 9.20$, $N = 64$). Der inferenzstatistische Test ergab $t(135) = -0.19$, $p = .85$ und eine Effektgröße von $d = 0.03$ ($SD_{\text{gepoolt}} = 9.33$). Der geringe deskriptiv vorhandene Mittelwertsunterschied kann durch das insignifikante Ergebnis und die geringe Effektgröße als unbedeutsam angesehen werden. Die Teilhypothese A wird verworfen. Schülerinnen zeigen entgegen der Annahme keine größere Sprachkompetenz in Deutsch als Schüler. Vielmehr scheint die Sprachkompetenz in Deutsch bei beide Gruppen vergleichbar zu sein.

Für Teilhypothese B zeigte sich in der Variablen „Ergebnis C-Test Englisch“ ein geringere Mittelwert für die Schülerinnen ($M = 50.82$, $SD = 14.38$, $N = 101$) als für die Schüler ($M = 57.28$, $SD = 14.99$, $N = 65$). Der Test ergab $t(132) = 2.75$, $p < .05$ und eine Effektgröße von $d = 0.44$ ($SD_{\text{gepoolt}} = 14.62$). Auf Grundlage dieser Ergebnisse wird Teilhypothese B ebenfalls verworfen. Diametral zur ursprünglichen Annahme zeigen die Schüler eine deutlich

größere Sprachkompetenz in Englisch als die Schülerinnen. Hypothese 12 wird damit insgesamt zurückgewiesen.

Hypothese 13 betrifft die Geschlechtsunterschiede hinsichtlich der Geografiekompetenz und der kognitiven Fähigkeiten: „Schülerinnen und Schüler zeigen keine signifikanten Unterschiede bezüglich ihrer Geografiekompetenz (Teilhypothese A) und ihrer kognitiven Fähigkeiten (Teilhypothese B)“. Deskriptiv zeigte sich für die Schülerinnen ein geringfügig kleinerer Mittelwert in der Variablen „Beurteilung Geografietest“ ($M = 2.45$, $SD = 0.68$, $N = 106$) als für die Schüler ($M = 2.51$, $SD = 0.60$, $N = 68$). Der Welch-Test ergab $t(154) = 0.68$, $p = .50$ und eine Effektgröße von $d = 0.09$ ($SD_{\text{gepoolt}} = 0.65$). Damit kann Teilhypothese A beibehalten werden. Hinsichtlich der Geografiekompetenz finden sich keine Unterschiede zwischen Schülerinnen und Schülern.

Für Teilhypothese B ergab sich deskriptiv ein geringerer Mittelwert bei den Schülerinnen in der Variablen „Ergebnis KFT“ ($M = 43.35$, $SD = 7.93$, $N = 101$) als bei den Schülern ($M = 48.61$, $SD = 6.80$, $N = 64$). Der Welch-Test zeigte mit $t(148) = 4.54$, $p < .10$ und einer Effektgröße von $d = 0.67$ ($SD_{\text{gepoolt}} = 7.51$) ein signifikantes Ergebnis. Daher wird Teilhypothese B nicht aufrecht erhalten. Entgegen der ursprünglichen Annahme sind Schülerinnen und Schüler hinsichtlich der kognitiven Fähigkeiten nicht vergleichbar. Vielmehr zeigen Schüler deutlich größere kognitive Fähigkeiten als Schülerinnen, was durch die Ausprägung der Effektgröße unterstrichen wird. Beim Vergleich der Verteilungskennwerte der Normstichprobe des KFT mit den Kennwerten der erhobenen Daten (vgl. Anhang E) zeigt sich, dass die untersuchten Schüler besonders im quantitativen Teil und im nonverbalen Teil besser abschneiden als die vergleichbaren Schüler der Normstichprobe.

4.7 Geografiekompetenz

Abschließend sollen noch die Ergebnisse einer explorativ durchgeführten multiplen Regression berichtet werden. Dabei wurde das Kriterium „Beurteilung Geografietest“ über verschiedene Prädiktoren vorhergesagt, um die Einflüsse verschiedener Variablen auf die Geografiekompetenz zu untersuchen. Als Prädiktoren fungierten „Ergebnis C-Test Deutsch“, „Ergebnis C-Test Englisch“, „Ergebnis KFT“ und „Klassenklima“. Für „Ergebnis KFT“ ist zu beachten, dass nicht der Gesamtwert des KFT verwendet wurde, sondern jeweils die Werte der drei Subtests, die sich zum Gesamtwert addieren. Alle Prädiktoren wurden simultan in die Berechnungen aufgenommen.

In Tabelle 23 sind die Interkorrelationen der beteiligten Variablen aufgeführt. Auffällig ist, dass die Variable „Klassenklima“ mit allen anderen Variablen eine sehr geringe Korrelation aufweist, die in keine Fall Signifikanz erreicht. Bei der Untersuchung der Zusammenhänge zwischen Klassenklima und Schulleistungen hatten sich bei Hypothese 11 ähnliche niedrige Zusammenhänge ergeben. Alle anderen Variablen korrelieren hoch untereinander und mit dem Kriterium. Die drei Subtests des KFT zeigen fast identische Zusammenhänge mit „Beurteilung Geografietest“. Untereinander korrelieren sowohl die drei Subtests des KFT als auch die beiden C-Tests sehr hoch miteinander.

Tabelle 23

Die Tabelle enthält die bivariaten Korrelationen nach Pearson für alle Variablen der multiplen Regression zur Geografiekompetenz. Insgesamt gingen die Daten von 159 Schülern in die Berechnungen ein.

Variable	Beurteilung Geografietest	Ergebnis C-Test Deutsch	Ergebnis C-Test Englisch	Ergebnis KFT (verbal)	Ergebnis KFT (quantitativ)	Ergebnis KFT (nonverbal)	Klassenklima
Beurteilung Geografietest		.43*	.27*	.20*	.19*	.19*	.08
Ergebnis C-Test Deutsch	.43*		.54*	.29*	.21*	.23*	.04
Ergebnis C-Test Englisch	.27*	.54*		.49*	.31*	.40*	.03
Ergebnis KFT (verbal)	.20*	.29*	.49*		.33*	.34*	-.04
Ergebnis KFT (quantitativ)	.19*	.21*	.31*	.33*		.53*	-.01
Ergebnis KFT (nonverbal)	.19*	.23*	.40*	.34*	.53*		-.06
Klassenklima	.08	.04	.03	-.04	-.01	-.06	

* *Bemerkung:* Die so markierten Korrelationskoeffizienten sind im einseitigen Test bei $p < .05$ signifikant.

Die errechnete multiple Korrelation von $R = .45$ war statistisch signifikant mit $F(6, 152) = 6.40$, $p < .05$. Es ergab sich ein multipler Determinationskoeffizient von $R^2 = .20$ und ein korrigierter multipler Determinationskoeffizient von $R_a^2 = .17$. Da R^2 kleiner ausfiel als die Summe über die quadrierten bivariaten Korrelationen mit dem Kriterium ($r^2 = .38$) kann von Redundanz der Prädiktoren ausgegangen werden, die damit einen gemeinsamen Anteil an der Varianz der Kriteriums haben. Die Regressionsgleichung mit unstandardisierten und standardisierten Gewichten ist in Tabelle 24 ersichtlich.

Tabelle 24

Die Tabelle enthält die Ergebnisse unstandardisierten (b) und standardisierten Regressionsgewichte (β) der multiplen Regression zur Geografiekompetenz. In den beiden letzten Spalten sind die Ergebnisse der t -Tests und die p -Werte für die einzelnen Prädiktoren gegeben.

Prädiktor	b	β	t (152)	p
(Konstante)	-.64		-1.13	.26
Ergebnis C-Test Deutsch	.03	.39	4.54	.00
Ergebnis C-Test Englisch	.00	-.02	-0.24	.81
Ergebnis KFT (verbal)	.01	.06	0.75	.46
Ergebnis KFT (quantitativ)	.01	.06	0.65	.51
Ergebnis KFT (nonverbal)	.01	.06	0.69	.49
Klassenklima	.08	.07	0.94	.35

Wie in Tabelle 24 erkennbar ist, wird der inferenzstatistische Test lediglich für den Prädiktor „Ergebnis C-Test Deutsch“ signifikant. Alle anderen Prädiktoren haben Regressionsgewichte nahe Null und damit einen sehr geringen Vorhersagebeitrag.

5 Diskussion

In diesem letzten Kapitel sollen die Ergebnisse des vorherigen Kapitels diskutiert werden. Neben den mit den Hypothesen verbundenen inhaltlichen Bereichen, die nacheinander behandelt werden, sollen dabei auch einige grundlegende Punkte angesprochen werden. Zu Beginn von Abschnitt 5.1 wird die gesamte Arbeit kurz zusammengefasst. Mit einer Auseinandersetzung zu den Stärken und Schwächen der Untersuchung und einem Ausblick für die Zukunft endet das Kapitel.

5.1 Fachlernen und Sprachlichkeit

In der vorliegenden Arbeit sollten hauptsächlich die folgenden Fragen zu Fachlernen und Sprachlichkeit im Kontext bilingualer Bildung geklärt werden: Sind bilingual unterrichtete Schüler hinsichtlich der Fachkompetenz in Erdkunde mit ihren monolingual unterrichteten Mitschülern vergleichbar? Sind bilingual unterrichtete Schüler hinsichtlich der Sprachkompetenz in Englisch ihren monolingual unterrichteten Mitschülern überlegen? Wie verhält es sich mit der Sprachkompetenz in Deutsch beim Vergleich beider Schülergruppen?

Zur Klärung dieser Fragen wurden Daten aus Erhebungen an vier niedersächsischen Gymnasien herangezogen. Dabei wurden Schüler der zehnten Klasse untersucht, von denen ein Teil bilingualen Erdkundeunterricht in englischer Sprache und der andere Teil konventionellen Erdkundeunterricht in deutscher Sprache hatte. Bei den Datenerhebungen war das zentrale Instrument ein im Rahmen eines DFG-Projektes entwickelter Geografiekompetenztest, der acht Aufgaben mit insgesamt 17 Teilaufgaben umfasste, in Englisch und Deutsch vorlag und von den Schülern in 90 Minuten bearbeitet wurde. Weiterhin kam als Sprachkompetenztest für Englisch und Deutsch jeweils ein C-Test zum Einsatz. Kognitive Fähigkeiten wurden mit dem KFT erfasst. Zudem füllten die Schüler einen umfangreichen Schülerfragebogen aus, indem neben demografischen Daten auch Angaben zur Berufstätigkeit der Eltern, zum Klassenklima und zu weiteren Bereichen gemacht werden sollten. Aus den Daten zur Berufstätigkeit wurde die soziale Herkunft der Schüler ermittelt. Weitere Daten, wie Noten der Schüler in Deutsch, Englisch, Mathematik und Erdkunde, wurden von den Schulen zur Verfügung gestellt. Sowohl die Beurteilung der Schülerantworten im Geografiekompetenztest, als auch die eingesetzte Skala zum Klassenklima wurden anhand der erhobenen Daten positiv auf ihre Reliabilität hin überprüft.

Es wurde davon ausgegangen, dass die bilingualen Schüler im Vergleich zu den monolingualen Schülern eine größere Geografiekompetenz zeigen, da in anderen Untersuchungen Kompetenzvorteile für bilinguale Schüler gefunden wurden (Lamsfuß-Schenk, 2002a, 2002b; Koch, 2005; Koch & Bündler, 2006). Diese Annahme konnte an den erhobenen Daten nicht bestätigt werden. Vielmehr zeigte sich, dass die Geografiekompetenz bei beiden Schülergruppen in etwa gleich ausgeprägt ist.

Offenbar führt bilingualer Erdkundeunterricht nicht zu einer größeren Geografiekompetenz. Die Kompetenzvorteile, die Lamsfuß-Schenk (2001, 200b) und Koch (2005) fanden, bezogen sich auf französischsprachigen Geschichtsunterricht und englischsprachigen Naturwissenschaftsunterricht. Vielleicht sind die Unterschiede im Unterricht und der Didaktik zwischen Geschichte und Naturwissenschaften auf der einen, und Erdkunde auf der anderen Seite einfach zu groß, um die Erkenntnisse aus den beiden anderen Fächern auf Erdkunde zu übertragen. Ob dem so ist, könnten zukünftige Untersuchungen zeigen. Eine weitere Möglichkeit ist, dass sich deshalb keine Kompetenzvorteile für bilinguale Erdkundeschüler zeigen, weil das Thema „Regenwald“, mit dem sich der Geografietest inhaltlich befasst, in der achten Klasse behandelt wird. Für die untersuchten Schüler der zehnten Klasse ist die unterrichtliche Auseinandersetzung somit zwei Jahre her. Daher ist es vorstellbar, dass Kompetenzvorteile durch Vergessensprozesse verschwunden sind. Allerdings ist der Geografietest von Experten mit der Zielsetzung konstruiert worden, die bis zur zehnten Klasse erworbene Geografiekompetenz zu erfassen, die nicht an einen Inhaltsbereich gebunden ist.

Nicht als Ursache für ein Ausbleiben der Kompetenzvorteile kommt die Sprache und die Art des Erhebungsverfahrens in Betracht. Die Sprache, weil sich bei Koch (2005) Kompetenzvorteile eben bei englischsprachigem Unterricht zeigten. Die Art des Erhebungsverfahrens sollte auch keine Ursache sein, da bei Lamsfuß-Schenk (2001, 200b) und Koch (2005) mit schriftlichen Tests und concept mapping zwei verschiedene Verfahren verwendet wurden, die Ergebnisse also als verfahrensunabhängig angesehen werden können.

Zwei Faktoren sind noch zu berücksichtigen, bevor die Geografiekompetenz in beiden Schülergruppen als vergleichbar angesehen werden kann. Zum einen sollten sich die Schüler nicht in ihren kognitiven Fähigkeiten und zum anderen ebenfalls nicht hinsichtlich ihrer sozialen Herkunft unterscheiden (Vollmer, 2006b). Die erste Annahme konnte bestätigt werden. Damit kann davon ausgegangen werden, dass nicht nur die Schüler mit größeren kognitiven Fähigkeiten, die meist als klügere Schüler angesehen werden, den bilingualen Unterricht besuchen. Ob Unterschiede oder keine Unterschiede zwischen mono- und

bilingualen Schülern gefunden werden – auf Unterschiede in den kognitiven Fähigkeiten kann dieser nicht unmittelbar zurückgeführt werden.

Die zweite Annahme dagegen konnte nicht bestätigt werden. Zwischen den beiden Schülergruppen zeigten sich Unterschiede in der sozialen Herkunft, wobei die bilingualen Schüler eine positiver beurteilte soziale Herkunft haben. Damit kann durch die vorliegenden Daten ein Argument von Kritikern des bilingualen Unterrichts nicht widerlegt werden. Diese gehen bisweilen davon aus, dass mehrheitlich die Schüler „aus besserem Hause“ an bilingualen Bildungsgängen teilnehmen. Zusätzliche Analysen zum Zusammenhang von sozialer Herkunft und Geografiekompetenz ergaben erfreulicherweise, dass dieser Zusammenhang sehr gering ist. Der Vergleich von mono- und bilingualen Schüler scheint damit zumindest für die Geografiekompetenz problemlos durchführbar zu sein.

Die beiden nächsten Kernfragestellungen betreffen die Sprachkompetenz der Schüler. Für die bilingualen Schüler wurde eine größere Sprachkompetenz in Englisch erwartet als für die monolingualen Schüler. Dies wurde angenommen, da in verschiedenen Untersuchungen (Wode et al., 1996; Bredenbröcker, 2002) gezeigt werden konnte, dass der bilinguale Unterricht förderlich für die Entwicklung der Sprachkompetenz in der Fremdsprache ist. Diese Annahme konnte auch in der vorliegenden Arbeit bestätigt werden. Es scheint so zu sein, dass sich durch die längere Auseinandersetzung mit der Fremdsprache tatsächlich eine höhere Sprachkompetenz ergibt (Vollmer, 2005b), was sich prinzipiell auch bei den Forschungen zu kanadischen Immersionsprogrammen gezeigt hat (Vollmer, 1992). Allerdings ermöglichen die Daten der vorliegenden Arbeit keine Aussagen über die zu Grunde liegenden Spracherwerbsprozessen.

Bezüglich der Sprachkompetenz in Deutsch wurden keine Unterschiede zwischen den Schülergruppen vermutet, da ein Großteil der Schüler Deutsch als Muttersprache hat und die bilingualen Schüler außerhalb des bilingualen Unterrichts genauso viel Kontakt mit der Muttersprache haben sollten, wie die monolingualen Schüler. Allerdings konnte diese Annahme nicht bestätigt werden, da die bilingualen Schüler neben der größeren Sprachkompetenz in Englisch ebenfalls eine größere Sprachkompetenz in Deutsch aufwiesen. Dieser interessante Befund könnte daher rühren, dass unter Umständen im bilingualen Unterricht Sprachkompetenz allgemein gefördert wird oder die größere Sprachkompetenz in Englisch positiv auf die Sprachkompetenz in Deutsch wirkt. Andererseits ist es möglich, dass die Schüler, die am bilingualen Unterricht teilnahmen, einfach sprachbegabter sind.

Die drei Kernfragestellungen der vorliegenden Arbeit sind damit beantwortet. Bilingual unterrichtete Schüler sind hinsichtlich ihrer Fachkompetenz den monolingualen Schülern vergleichbar, obgleich sie keine größere Fachkompetenz zeigen, wie vermutet wurde. Bezüglich der Sprachkompetenz sind die bilingualen Schüler den monolingualen sowohl in Deutsch, als auch in Englisch überlegen.

5.2 Schulleistungen

Neben den bisher behandelten Kernfragestellungen wurden verschiedene Zusammenhänge mit Schulleistungen untersucht. Es wurde davon ausgegangen, dass die Schulleistungen im Fach Erdkunde mit der Geografiekompetenz zusammenhängen. Diese Annahme konnte bestätigt werden. Dies spricht im Sinne einer Kriteriumsvalidierung für den Geografietest.

Hinsichtlich der Sprachkompetenz haben schon Klein-Braley und Raatz (1982) hohe Zusammenhänge zwischen Ergebnissen des C-Tests und den entsprechenden Fachnoten gefunden und dies als Beleg für die Übereinstimmungsvalidität des C-Tests gewertet. Daher wurde angenommen, dass die Schulleistungen im jeweiligen Sprachfach positiv mit der jeweiligen Sprachkompetenz zusammenhängen. Auch diese Annahmen konnten bestätigt werden, wobei die Zusammenhänge geringer ausfielen als bei Klein-Braley und Raatz (1982).

Wie erwähnt handelt es sich bei den durchgeführten Analysen zu Zusammenhängen mit Noten um eine Form der Kriteriumsvalidierung, wobei Noten als Kriterium nicht unumstritten sind. Wie in Abschnitt 2.8 dargestellt wurde, ist nicht immer klar, was Noten eigentlich aussagen. Der C-Test ist ein relativ gut erprobtes Verfahren, zu dem viele Untersuchungen hinsichtlich Validität und Reliabilität durchgeführt worden sind (z.B. Faust, 1985; Grotjahn, 1993; Grotjahn, 1995; Klein-Braley & Raatz, 1982; Raatz & Klein-Braley, 1983). Der Geografietest hingegen ist hier erstmalig verwendet worden. Die berichtete Kriteriumsvalidierung kann somit als ein erster Schritt zur Überprüfung des Verfahrens betrachtet werden.

Auf der Grundlage von verschiedenen Untersuchungen (Beer, Kutalek & Schnell, 1968; Kolem, 1981; Kühn, 1983) wurde angenommen, dass die kognitiven Fähigkeiten positiv mit den Schulleistungen zusammenhängen. Auch diese Annahme konnte anhand der erhobenen Daten bestätigt werden, obwohl die hier gefundenen Zusammenhänge kleiner sind, als in den anderen Untersuchungen (ebd.). Der trotzdem quasi triviale Befund – das Alter der zitierten Arbeiten belegt die lange Forschungstradition in diesem Bereich – weist allerdings auf Grund seiner Größe darauf hin, dass die Noten nur zu einem begrenzten Anteil Aussagen

über die kognitiven Fähigkeiten machen können. Dies ändert nichts am prädiktiven Wert von kognitiven Fähigkeiten im schulischen Bereich.

5.3 Soziale Herkunft und Klassenklima

Da die soziale Herkunft erwiesenermaßen einen großen Einfluss auf die Schulleistungen hat (Baumert & Schümer, 2001; Ehmke, Hohensee, Heidemeier & Prenzel, 2004), insofern mit positiverer Beurteilung der sozialen Herkunft auch bessere Schulleistungen einhergehen, sollte mit den Daten des DFG-Projektes untersucht werden, ob sich auch für die vier Kompetenzvariablen Geografiekompetenz, Sprachkompetenz und die kognitiven Fähigkeiten derartige Auswirkungen finden lassen. Das war nicht der Fall.

Es ist nicht so, dass mit niedrigerer Stufe im EGP-Klassenschema, die eine positivere Beurteilung der sozialen Herkunft darstellt, automatisch bessere Ergebnisse in den Kompetenzvariablen einhergehen. Lediglich für die beiden letzten Stufen im EGP-Klassenschema zeigen sich relativ schlechte Werte in den Kompetenzvariablen. Es verwundert, dass sich in den Daten keine deutlicheren Auswirkungen der sozialen Herkunft zeigen. Vielleicht liegt dies daran, dass in der Stichprobe tendenziell mehr Schüler mit einer positiver bewerteten sozialen Herkunft vorhanden sind.

Wie auch bei Bültner und Meyer (2004), so wurde in der vorliegenden Arbeit davon ausgegangen, dass eine positive Einschätzung des Klassenklimas mit positiven Auswirkungen auf die Schulleistungen einhergeht. Zusätzlich wurde geprüft, ob zwischen Klassenklima und Geografiekompetenz ein Zusammenhang besteht. Die Datenanalysen brachten keine Bestätigung dieser Annahmen.

Die Ergebnisse können darauf zurückzuführen sein, dass der angenommene direkte Wirkzusammenhang zwischen Klassenklima, Schulleistung und Geografiekompetenz zu einfach ist. Vielleicht ist es aber auch der Fall, dass das Klassenklima nur einen sehr geringen Einfluss auf die Schulleistungen hat. Es ist beispielsweise denkbar, dass Klassenklima die Lernmotivation beeinflusst und nur indirekt auf die Schulleistungen wirkt. Aktuell ist Klassenklima ein sehr breites Konstrukt, das zwar oft eingesetzt wird, aber ebenso oft anders definiert wird. Weiterhin könnten die geringen Zusammenhänge daraus resultieren, dass das Klassenklima in der vorliegenden Arbeit auf Ebene der einzelnen Schüler erfasst worden ist und bei der Datenauswertung über Schüler und Klassen aggregiert wurde. Vielleicht reicht es,

in einem ersten Schritt darauf zu achten, dass vor der Aggregation über Klassen geprüft wird, ob das Klassenklima auch innerhalb der einzelnen Klassen homogener ist als außerhalb.

Ein weiterer Grund für die Ergebnisse könnte in der Operationalisierung des Klassenklimas zu finden sein. Die eingesetzte selbstkonstruierte Skala kam erstmalig zum Einsatz. An ihr fällt auf, dass invertierte Items besonders schlechte Reliabilitätskennwerte aufweisen. Zudem ist bei einem Item eine Verknüpfung durch *und* vorhanden. Diese sollte gelöst werden, indem zwei Items gebildet werden. Einige Worte in den Itemstämmen sind ferner durch Fettdruck hervorgehoben und könnten die Antworten der Schüler beeinflussen. Die Skala sollte vor dem Einsatz in weiteren Untersuchungen modifiziert und genauer geprüft werden, besonders hinsichtlich ihrer Validität. Alternativ könnte sie durch standardisierte Skalen zum Klassenklima ersetzt werden. Die durchgeführte Itemanalyse erbrachte nach der Elimination zweier Items zwar zufriedenstellende Kennwerte, ist aber kein Validitätsbeweis.

Es sei zudem erwähnt, dass in der ursprünglichen Konzeption jene fünf Items, die hier zur Skala *Klassenklima* zusammengefasst worden sind, anders verwendet werden sollten. Drei Items sollten gar nicht ausgewertet werden und die restlichen Items bildeten die Grundlage für die Skala *Soziale Atmosphäre in der Erdkunde Klasse*.

5.4 Geschlechtsunterschiede

In vielen Untersuchungen hat sich gezeigt, dass Schülerinnen Stärken im sprachlichen Bereich besitzen, wohingegen Stärken der Schüler eher im mathematisch-naturwissenschaftlichen Bereich zu finden sind (z.B. Stanat & Kunter, 2001). Darauf aufbauend wurde angenommen, dass Schülerinnen eine größere Sprachkompetenz in Deutsch und Englisch zeigen. Für Deutsch erwies sich diese Annahme als falsch. Schülerinnen wie Schüler haben hier eine vergleichbare Sprachkompetenz. Für Englisch erwies sich die Annahme ebenfalls als nicht zutreffend, im Gegenteil: Die Schüler zeigten eine größere Sprachkompetenz.

Wie ist das zu erklären? Dem Autor sind keine Untersuchungen bekannt, in denen Geschlechtsunterschiede beim Bearbeiten von C-Tests gefunden wurden. Von daher kann die Verfahrensdurchführung als Ursache ausgeschlossen werden. Natürlich könnte es der Fall sein, dass die Schülerinnen in der Stichprobe einfach eine geringere Sprachkompetenz besitzen, auch wenn Schülerinnen im Allgemeinen Stärken in diesem Bereich haben. Dass es sich bei den untersuchten Schülern um besonders sprachbegabte handelt, ist relativ unwahrscheinlich. Ohnehin haben sie einen unterdurchschnittlichen Anteil an der Stichprobe.

Schließlich wurde noch davon ausgegangen, dass sich keine Geschlechtsunterschiede bei der Geografiekompetenz und den kognitiven Fähigkeiten finden. Für die Geografiekompetenz sind dem Autor ebenfalls keine Untersuchungen bekannt, die sich mit Geschlechtsunterschieden in diesem Bereich befassen. Da im Fach Erdkunde zudem sprachliche und naturwissenschaftliche Anteile ihren Platz haben und das eingesetzte Verfahren zur Erhebung der kognitiven Fähigkeiten mangels vorhandener Unterschiede auf geschlechtsspezifische Normen verzichtet (Heller & Perleth, 2000), sprach nichts dafür, Geschlechtsunterschiede in diesem Bereich anzunehmen.

Hinsichtlich der Geografiekompetenz konnte die Hypothese bestätigt werden. Bei den kognitiven Fähigkeiten schnitten allerdings die Schüler besser ab. Ihr besseres Ergebnis könnte darin begründet liegen, dass die Schülerinnen durch die Auswahl weniger KFT-Subtests benachteiligt wurden. Gegen diese Annahme spricht, dass die Ergebnisse der Schülerinnen in der vorliegenden Arbeit vergleichbar mit denen der Schülerinnen aus der Normstichprobe des KFT ausfallen. Es ist vielmehr so, dass die Schüler im quantitativen und nonverbalen Teil des Test einfach besser abschneiden als die Schüler der Normstichprobe.

5.5 Kompetenz und Performanz

In der vorliegenden Arbeit wurde vielfach der Begriff Kompetenz verwendet: Geografiekompetenz, Sprachkompetenz und Fachkompetenz sind einige Beispiele. Da in der allgemeinen Diskussion nicht unumstritten ist, was der Kompetenzbegriff aussagt, soll darauf an dieser Stelle genauer eingegangen werden.

Von Saldern (1997) unterscheidet zwischen Kompetenz auf der einen und Performanz auf der anderen Seite. Bei *Kompetenz* handelt es sich um Fähigkeiten oder Vermögen eines Menschen. Bezogen auf den schulischen Kontext ist Kompetenz das, was eine Schüler kann. Bei *Performanz* dagegen handelt es sich um „die konkrete Realisierung, also Anwendung und Ausführung von Kompetenzen“ (ebd., S. 30). Performanz ist damit das, was der Schüler zeigt. Nun gilt für die Feststellung von Schulleistung natürlich, dass der Lehrer nur die wirklich gezeigten Leistungen der Schüler bewerten kann. An dieser Stelle soll unberücksichtigt bleiben, dass in das subjektive Lehrerurteil noch andere Wahrnehmungen einfließen können.

Das eigentliche Problem sieht von Saldern (1997) darin, dass von der Performanz ganz selbstverständlich auf die Kompetenz geschlossen wird. Dabei kann beispielsweise

Angst vor Fehlern verhindern, dass Schüler wirklich zeigen, was sie können. Performanz und Kompetenz liegen in diesem Fall angstbedingt weit auseinander, wobei nicht genau abgeschätzt werden kann, wie weit. Die gleiche Problematik gilt natürlich für die Erfassung der Geografiekompetenz durch den Geografietest und analog für Sprachkompetenz und C-Test. Mit diesen beiden Verfahren kann letztlich nur die Performanz der Schüler in der besonderen Testsituation erfasst werden. Durch die im Gegensatz zum schulischen Alltag andere Situation sind vielfältige Einflüsse auf das Verhältnis von Kompetenz und Performanz denkbar. Beispielsweise könnten die Schüler weniger motiviert sein und nur weniger von ihrer Kompetenz zeigen, da es sich um wissenschaftliche Datenerhebungen und nicht um zeugnisrelevante Leistungen handelt. Was damit aufgezeigt werden soll: In der vorliegenden Untersuchung ist genauso wenig wie in vielen anderen Untersuchungen klar, inwiefern sich die erfasste Performanz von der vorhandenen Kompetenz unterscheidet.

Unter Geografiekompetenz wird in dieser Arbeit – wie weiter oben dargestellt – die Gesamtheit der Fähigkeiten, Fertigkeiten und Kompetenzen der Schüler im Fach Erdkunde verstanden, die sich auf die Kompetenzbereiche Fachwissen, Methodenkompetenz, Fachkommunikation, Beurteilungskompetenz und Bewertungskompetenz verteilt. Grundlage dieses Verständnisses ist der im DFG-Projekt verwendete Kompetenzbegriff nach Weinert (2001b):

Daneben gibt es natürlich auch Probleme, die inhaltsunspezifisch und im engeren Sinne fächerübergreifend sind. Ihre Bewältigung hängt in erster Linie von der Verfügbarkeit allgemeiner Problemlösestrategien [*sic*] ab. Die OECD hat in diesem Zusammenhang mehrfach vorgeschlagen, den vieldeutigen Leistungsbegriff generell durch das Konzept der Kompetenz zu ersetzen Dabei versteht man unter Kompetenz die bei Individuen verfügbaren oder durch sie erlernbaren kognitiven Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten um die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können (S. 27f)

Nach einer Expertise des Bundesministeriums für Bildung und Forschung (2003) handelt es sich bei Kompetenz nach diesem Verständnis um eine Disposition, wobei darunter die potentielle Problemlöse- oder Handlungsfähigkeit verstanden wird. In der Expertise finden sich weitere Erörterungen zum Kompetenzbegriff, dessen Explikation insbesondere als notwendige Voraussetzung zur Definition von nationalen Bildungsstandards in Deutschland gesehen wird. Trotz alledem löst auch dieser Kompetenzbegriff nicht das Problem des Unterschieds von Kompetenz und Performanz. Hier sollen mögliche Lösungen dieses Problems nicht diskutiert werden; vielmehr sollten die Ausführungen dazu dienen, beim Leser den Blick für die generelle Problematik zu schärfen.

In einer explorativen Analyse wurde der Zusammenhang von Geografiekompetenz und anderen Variablen der vorliegenden Arbeit genauer untersucht. Dabei zeigte sich, dass lediglich die Sprachkompetenz Deutsch einen deutlichen Beitrag zur Vorhersage der Geografiekompetenz der Schüler leisten kann. Sowohl die Sprachkompetenz Englisch, als auch die kognitiven Fähigkeiten, als auch das Klassenklima sind zur Vorhersage der Geografiekompetenz schlecht geeignet. Wieso existiert dieser große Zusammenhang zwischen Geografiekompetenz und Sprachkompetenz Deutsch? Vielleicht sind für die erfolgreiche Bearbeitung des Geografietests gewisse Fertigkeiten erforderlich, die auch beim C-Test relevant sind. Allerdings ist dies fraglich, da beim C-Test ein Lückentext ausgefüllt werden muss, beim Geografietest hingegen relativ komplexe Aufgaben schriftlich beantwortet werden müssen. Inhaltlich hat der C-Test nichts speziell mit Erdkunde zu tun (vgl. Anhang C). Interessant wäre zu wissen, ob ein C-Test mit Texten aus dem Erdkundeunterricht sogar noch größere Zusammenhänge zum Geografietest zeigen würde. Vielleicht sagt der C-Test Deutsch andererseits bei deutschen Schülern etwas über eine sprachenübergreifende Sprachbegabung aus, die für die Bearbeitung des Geografietests auf Deutsch und Englisch vorteilhaft ist. Dann sollte sich bei Schülern mit Muttersprache Englisch und Fremdsprache Deutsch ein stärkerer Zusammenhang zwischen der Sprachkompetenz Englisch und der Geografiekompetenz zeigen.

5.6 Stärken und Schwächen der Untersuchung

Im Nachhinein sind einige Schwächen der vorliegenden Untersuchung und des Vorgehens im DFG-Projekt klar geworden. Neben diesen sollen im Folgenden allerdings die Stärken der Untersuchung nicht vernachlässigt werden.

Von herausragender Bedeutung ist der im DFG-Projekt konstruierte Geografietest. Dabei handelt es sich um ein fundiertes Verfahren, das mit viel Arbeitsaufwand von verschiedenen Experten erstellt worden ist. Hinzu kommt, dass mit dem DFG-Projekt erstmals im deutschsprachigen Raum der Versuch unternommen wurde, an einer großen Stichprobe von Schülern die Fachkompetenz in einem ausgewählten Sachfach bei mono- und bilingualen Schülern systematisch zu vergleichen. Das bei der erstmaligen Umsetzung eines so ambitionierten Vorhabens einige Schwierigkeiten auftreten – beispielsweise bei der Datenerhebung in den Schulen – sollte nachsichtig zu Kenntnis genommen werden.

Da die Konstruktion des Geografietests ohnehin viel Zeit in Anspruch genommen hat, fehlte diese für intensive Vortests des Verfahrens. Vor einer erneuten Anwendung des

Geografietests sollten Prüfungen der Validität und Reliabilität durchgeführt werden. Damit einhergehen sollte eine systematischere Beurteilerschulung zur Auswertung des Tests. Auch wenn an der Beurteilerübereinstimmung auf Grundlage der vorhandenen Ergebnisse nichts zu kritisieren ist, sollte als Datenbasis für zukünftige Untersuchungen eine Doppelbeurteilung aller Schülerantworten angestrebt werden. Ferner sollte das zu Grunde liegende Modell der Geografiekompetenz noch weiter theoretisch ausgearbeitet und genauer untersucht werden. Dies konnte in der vorliegenden Arbeit nicht geleistet werden.

Weiterhin hat sich gezeigt, dass es für die Untersuchung der Auswirkungen des Klassenklimas sinnvoller ist, selbiges mittels erprobter Verfahren zu erfassen. Die Forschung zum Klassenklima ist vermutlich zu komplex, um dieses Thema am Rande zu behandeln. Zudem wurden Daten über Schüler und Klassen zusammengefasst. Zukünftige Untersuchungen sollten sich an Ansätzen der Mehrebenenanalyse bedienen. So kann die Datenerhebung und die Datenauswertung auf verschiedenen Ebene zuverlässiger durchgeführt werden.

Zur Erfassung der kognitiven Fähigkeiten wurden ausgewählte Subtests des KFT eingesetzt. Dabei ist allerdings nicht auszuschließen, dass die Vorteile eines standardisierten und normierten Verfahrens auf diese Weise verloren gehen. Ein solches Verfahren sollte entsprechend der Durchführungsanweisungen angewendet werden. Auch für den KFT steht eine geprüfte Kurzform zur Verfügung. Um die nötige Zeit für die längere Datenerhebung zu schaffen, sollte auf die Erhebung andere Daten verzichtet werden. Leider scheint die willkürliche Auswahl von Subtests – rechtfertigt mit Zeitersparnis bei der Datenerhebung – ein häufig praktiziertes Vorgehen zu sein.

Für Folgeuntersuchungen ist weiterhin eine genauere Erfassung der sozialen Herkunft unerlässlich. Die Theorie in diesem Bereich ist weit entwickelt und bietet verschiedene standardisierte Maße an. Die Erfassung wichtiger Merkmale wie Grad der Autonomie der Eltern in ihrem Beruf fehlte im Schülerfragebogen. Auch kann der Beruf der Eltern besser erfasst werden, indem Schüler um eine kurze Beschreibung der Tätigkeiten der Eltern gebeten werden. Ein weiterer wichtiger Einflussfaktor in diesem Bereich ist die Schulbildung der Eltern. Auch diese wurde nur unzureichend erfasst. Insgesamt sollten Maße der sozialen Herkunft Verwendung finden, die über vielfältigere Ausgangsdaten gebildet werden und intervallskalierte Daten ergeben. Damit könnten dann auch Regressions- und Kovarianzanalysen durchgeführt werden. Eine untersuchungsplanerische Möglichkeit wäre der Einsatz von auf soziale Herkunft gematchten Gruppen.

Insgesamt sollte in einer Untersuchung, in der besonders Geografie- und Sprachkompetenz interessieren, davon abgesehen werden, eine Vielzahl anderer Merkmale zu erfassen. Statt dessen sollten nur wenige theoretisch oder explorativ interessierende Merkmale zusätzlich in die Untersuchung aufgenommen werden. Diese Merkmale sollten dann allerdings vorzugsweise durch erprobte Verfahren erfasst werden. Das sichert nicht nur Vergleichbarkeit zwischen verschiedenen Untersuchungen, sondern auch eine bessere Datenqualität. Von daher könnte der Umfang des Schülerfragebogens reduziert werden. Weiterhin ist die Motivation der Schüler zur Teilnahme an der Untersuchung nicht erfasst worden. Mit wenigen Items im Schülerfragebogen wäre so etwas möglich gewesen. Dies hätte man idealerweise noch mit standardisierten Protokollen zur Durchführung der Datenerhebung in den einzelnen Schulen unterstützen können.

Ein weiterer Punkt, der nicht berücksichtigt wurde, ist die *Sprachbegabung* der Schüler. Damit ist hier die Fähigkeit, Sprachen zu erwerben und mit Sprachen zu agieren gemeint. Die Sprachbegabung sollte erfasst werden, um ausschließen zu können, dass eine größere Sprachkompetenz in Englisch bei bilingualen Schüler auf eine Selbstselektion besonders sprachbegabter Schüler zurückzuführen ist.

Zwei Punkte sind hinsichtlich der Untersuchungsplanung noch zu betonen. Zum einen können mögliche Reihenfolgeeffekte der Erhebungsinstrumente nicht ausgeschlossen werden, da die Durchführungsreihenfolge bei allen Erhebungen gleich war. Bei zukünftigen Erhebungen sollte die Reihenfolge systematisch variiert werden. Zum anderen wurden die Datenerhebungen und die Beurteilungen der Schülerantworten beim Geografietests von Mitarbeitern der Projektgruppe angeleitet beziehungsweise durchgeführt. Damit können Effekte unbewusster Beeinflussung, wie beispielsweise der *Rosenthal-Effekt*, nicht ausgeschlossen werden. Am sinnvollsten ist daher ein *doppelblindes Vorgehen*, bei dem sowohl die Teilnehmer, als auch die Anleiter und Beurteiler nicht über die Hypothesen informiert werden.

5.7 Ausblick

Zukünftige Untersuchungen, die sich die hier vorliegende Arbeit oder andere Teile des DFG-Projektes zur Grundlage nehmen, sollten in jedem Fall die in Abschnitt 5.6 dargestellten Verbesserungsvorschläge aufgreifen. Weiterhin sind für solche Untersuchungen aus jetziger Sicht verschiedene Schwerpunkte denkbar. Der Geografietest könnte in anderen Bundesländern oder im deutschsprachigen Ausland angewendet werden. Zumindest die Vergleichbarkeit von

mono- und bilingualen Schüler sollte dabei repliziert werden können. Weiterhin könnte darüber nachgedacht werden, das Verfahren in englischsprachigen Ländern einzusetzen. Explorativ könnte ein C-Test mit Texten aus dem Erdkundeunterricht erstellt und angewendet werden. Der Vergleich der C-Test-Ergebnisse mit den Ergebnissen des Geografietest ließe dann weitere Aussagen über die Validität der Verfahren zu.

Der Fokus der Datenerhebung sollte bei zukünftigen Untersuchungen neben der Geografie- und Sprachkompetenz auf den kognitiven Fähigkeiten, der sozialen Herkunft, der Motivation der Schüler und deren Sprachbegabung liegen. Diese vier Aspekte sind von enormer Bedeutung für die Aussagekraft der Daten in den beiden genannten Kompetenzbereichen. Durch eine systematischere Erfassung der kognitiven Fähigkeiten und der sozialen Herkunft könnte die Vergleichbarkeit der Schülergruppen hinsichtlich dieser Größen sichergestellt werden. Die Erhebung der Motivation ist erforderlich, um abschätzen zu können, inwieweit die Schüler überhaupt engagiert sind. Die Sprachbegabung – im Sinne der Fähigkeit, Sprachen zu erwerben und mit Sprachen zu agieren – ist wichtig, um ausschließen zu können, dass die unterschiedliche Sprachkompetenz nicht auf den Unterricht, sondern eben auf die Sprachbegabung zurückgeführt werden kann.

Optimal wäre es, wenn zur Untersuchung der Fachkompetenz bei mono- und bilingualen Schülern ein experimentelles Vorgehen angewendet werden könnte. Denkbar ist beispielsweise eine randomisierte Gruppeneinteilung im Rahmen einer freiwilligen Arbeitsgemeinschaft, ähnlich wie bei Koch (2005). Die verschiedenen Gruppen könnten wahlweise auf Englisch oder Deutsch unterrichtet werden. Die relativ gut kontrollierten Bedingungen eines solchen Feldexperimentes würden dann kausale Schlüsse zulassen.

Ein weiterhin zu bearbeitender Bereich ist die Spracherwerbsforschung. Nach Vollmer (2005a) ist nur unzureichend geklärt, wie die Spracherwerbsprozesse im bilingualen Unterricht ablaufen und wie sie sich von denen unterscheiden, die im Englischunterricht stattfinden. Der Unterschied zwischen Spracherwerbsprozessen im bilingualen Fachunterricht und im englischen Sprachunterricht sollte von daher besonders untersucht werden. Zudem sollte genauer geklärt werden, wie zeitlich längerer und intensiverer Sprachunterricht zu größerer Sprachkompetenz führt. Im Sinne einer formativen Evaluation könnten dabei auch Unterrichtsprozesse betrachtet und mit Blick auf Verbesserungen beim Erwerb von Fach- und Sprachkompetenz modifiziert werden.

Ingesamt bietet die vorliegende Arbeit einen interessanten Einblick in den Vergleich mono- und bilingualer Schüler. Die Datenerhebung ist in einigen Aspekten sicherlich

kritisierbar und die berichteten Ergebnisse können nur einen kleinen Beitrag für den Bereich der schulischen Forschung leisten. Sie bietet damit allerdings viele Ansatzpunkte für Folgeuntersuchungen. Hervorzuheben sind die Ergebnisse, dass die mono- und bilingualen Schüler hinsichtlich der Fachkompetenz in Erdkunde vergleichbar sind und dass die bilingualen Schüler eine deutlich höhere Sprachkompetenz in Englisch zeigen. Beide Ergebnisse sprechen für den bilingualen Unterricht und gegen die Kritikermeinung, dass die Fachkompetenz beim bilingualen Ansatz leide. Leider erlauben die Ergebnisse wenig direkte Schlussfolgerungen für die Schulpraxis. Dazu sind genauere Untersuchungen notwendig, wobei besonders längsschnittliche Ansätze verwendet werden sollten.

6 Literaturverzeichnis

- Alderson, J. C., Clapham, C. & Wall, D. (1995). *Language test construction and evaluation*. Cambridge: Cambridge University Press.
- Amthauer, R., Brocke, B., Liepmann, D. & Beauducel, A. (2001). *I-S-T 2000 R. Intelligenz-Struktur-Test 2000 R. Manual*. Göttingen: Hogrefe.
- Artelt, C., Stanat, P., Schneider, W. & Schiefele, U. (2001). Lesekompetenz: Testkonzeption und Ergebnisse. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann & M. Weiß (Hrsg.), *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 69-137). Opladen: Leske + Budrich.
- Asendorpf, J. & Wallbott, H. G. (1979). Maße der Beobachterübereinstimmung: Ein systematischer Vergleich. *Zeitschrift für Sozialpsychologie*, 10, 243-253.
- Asendorpf, J. B. (1998). Entwicklungsgenetik. In H. Keller (Hrsg.), *Lehrbuch Entwicklungspsychologie* (S. 97-118). Bern: Huber.
- Avenarius, H. (2001). *Einführung in der Schulrecht*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Bach, G. (2005). Bilingualer Unterricht: Lernen – Lehren – Forschen. In G. Bach & S. Niemeier (Hrsg.), *Bilingualer Unterricht. Grundlagen, Methoden, Praxis, Perspektiven* (3., überarbeitete und erweiterte Aufl., S. 9-22). Frankfurt am Main: Lang.
- Baumert, J. & Schümer, G. (2001). Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann & M. Weiß (Hrsg.), *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 323-407). Opladen: Leske + Budrich.
- Baumert, J. (2001). Internationale Schulleistungsvergleiche. In D. H. Rost (Hrsg.), *Handwörterbuch pädagogische Psychologie* (2., überarbeitete und erweiterte Aufl., S. 294-303). Weinheim: Psychologie Verlags Union.
- Baumert, J., Stanat, P. & Demmrich, A. (2001). PISA 2000: Untersuchungsgegenstand, theoretische Grundlagen und Durchführung der Studie. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann & M.

- Weiß (Hrsg.), *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 15-68). Opladen: Leske + Budrich.
- Beer, F., Kutalek, N. & Schnell, H. (1968). *Der Einfluß von Intelligenz und Milieu auf die Schulleistung*. Wien: Verlag für Jugend und Volk.
- Bloom, B. S. (1976). *Human characteristics and school learning*. New York: McGraw-Hill.
- Borg, I. & Staufenbiel, T. (1997). *Theorien und Methoden des Skalierung. Eine Einführung* (3., überarbeitete Aufl.). Bern: Huber.
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation* (4. Aufl.). Heidelberg: Springer Medizin.
- Bortz, J. (1999). *Statistik für Sozialwissenschaftler* (5., vollständig überarbeitete und aktualisierte Aufl.). Heidelberg: Springer.
- Bortz, J. (2005). *Statistik für Human- und Sozialwissenschaftler* (6. Aufl.). Heidelberg: Springer Medizin.
- Bos, W. & Postlethwaite, T. N. (2002). Internationale Schulleistungsforschung. Ihre Entwicklungen und Folgen für die deutsche Bildungslandschaft. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (2., unveränderte Aufl., S. 251-267). Weinheim: Beltz.
- Bourdieu, P. (1983). Ökonomisches Kapital, kulturelles Kapital, soziales Kapital. In R. Kreckel (Hrsg.), *Soziale Ungleichheiten* (S. 183-198). Göttingen: Schwartz.
- Bredenbröcker, W. (2002). Förderung der fremdsprachlichen Kompetenz durch bilingualen Sachfachunterricht: Empirische Untersuchungen. In: S. Breidbach, G. Bach & D. Wolff (Hrsg.), *Bilingualer Sachfachunterricht. Didaktik, Lehrer-/Lernerforschung und Bildungspolitik zwischen Theorie und Empirie* (S. 141-149). Frankfurt am Main: Lang.
- Breidbach, S. (2007). *Bildung, Kultur, Wissenschaft. Reflexive Didaktik für den bilingualen Sachfachunterricht*. Münster: Waxmann.
- Brocke, B., Beauducel, A. & Liepmann, D. (2004). I-S-T 2000 R. Intelligenz-Struktur-Test 2000 R. In W. Sarges & H. Wottawa, *Handbuch wirtschaftspsychologischer Testverfahren* (2. Aufl., S. 439-445). Lengerich: Pabst.
- Bülter, H. & Meyer, H. (2004). Was ist ein lernförderliches Klima? Voraussetzungen und Wirkungen. *Pädagogik*, 56 (11), 31-36.

- Bundesministerium für Bildung und Forschung (Hrsg.). (2003). *Zur Entwicklung nationaler Bildungsstandards. Eine Expertise*. Bonn: Bundesministerium für Bildung und Forschung.
- Chasiotis, A. & Voland, E. (1998). Geschlechtliche Selektion und Individualentwicklung. In H. Keller (Hrsg.), *Lehrbuch Entwicklungspsychologie* (S. 97-118). Bern: Huber.
- Christ, I. (1999). Zur Weiterentwicklung des bilingualen Lehrens und Lernens in deutschen Schulen. Bilinguale Züge - Bilingualer Unterricht - Module „Fremdsprache als Arbeitssprachen“. *Zeitschrift für Interkulturellen Fremdsprachenunterricht*, 4 (2), 12 Seiten.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46.
- Coleman, J. S. (1987). Families and Schools. *Educational Researcher*, 16 (6), 6, 32-38.
- Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology, Organizations and Institutions Supplement*, 94, S95-S120.
- Conger, A. J. (1980). Integration and generalization of kappas for multiple raters. *Psychological Bulletin*, 88, 322-328.
- Craik, F. I. & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11, 671-684.
- Decke-Cornill, H. (1999). Einige Bedenken angesichts eines möglichen Aufbruchs des Fremdsprachenunterrichts in eine bilinguale Zukunft. *Neusprachliche Mitteilungen aus Wissenschaft und Praxis*, 52, 164-170.
- Diehl, J. M. & Kohr, H. U. (1977). *Deskriptive Statistik*. Frankfurt am Main: Fachbuchhandlung für Psychologie.
- Eckensberger, L. H. & Keller, H. (1998). Menschenbilder und Entwicklungskonzepte. In H. Keller (Hrsg.), *Lehrbuch Entwicklungspsychologie* (S. 11-56). Bern: Huber.
- Eder, F. (2001). Schul- und Klassenklima. In D. H. Rost (Hrsg.), *Handwörterbuch pädagogische Psychologie* (2., überarbeitete und erweiterte Aufl., S. 578-586). Weinheim: Psychologie Verlags Union.
- Ehmke, T., Hohensee, F., Heidemeier, H. & Prenzel, M. (2004). Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff, J. Rost & U. Schiefele (Hrsg.), *PISA 2003: Der Bildungsstand der Jugendlichen in*

- Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 225-254).
Münster: Waxmann.
- Erikson, R., Goldthorpe, J. H. & Portocarero, L. (1979). Intergenerational class mobility in three Western European societies: England, France and Sweden. *British Journal of Sociology*, 30, 415-441.
- Farkas, G. (1974). Specification, residuals and contextual effects. *Sociological Methods & Research*, 2, 333-363.
- Faust, B. (1985). *Der C-Test. Testtheoretische Überlegungen und Analysen zu einem neuen Testverfahren* (Manuskripte zur Sprachlehrforschung Nr. 25, herausgegeben vom Seminar für Fremdsprachenforschung der Ruhr-Universität Bochum). Heidelberg: Groos.
- Fleiss, J. L. & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, 33, 613-619.
- Ganzeboom, H. B. G. & Treiman, D. J. (1996). Internationally comparable measures of occupational status for the 1988 International Standard Classification of Occupations. *Social Science Research*, 25, 201-239.
- Ganzeboom, H. B. G., De Graaf, P. M. & Treiman, D. J. (1992). A standard international socio-economic index of occupational status. *Social Science Research*, 21, 1-56.
- Gold, A. & Souvignier, E. (2005). Prognose der Studierfähigkeit. Ergebnisse aus Längsschnittanalyse. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 37, 214-222.
- Grotjahn, R. (1993). Der C-Test. Grundlagentheoretische und anwendungsbezogene Untersuchungen. *Zeitschrift für Fremdsprachenforschung*, 4 (2), 77-87.
- Grotjahn, R. (1995). Der C-Test: State of the Art. *Zeitschrift für Fremdsprachenforschung*, 6 (2), 37-60.
- Hauser, R. M. (1974). Contextual analysis revisited. *Sociological Methods & Research*, 2, 365-375.
- Heller, K. A. & Perleth, C. (2000). *Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision. Manual*. Göttingen: Beltz Test.
- Helmke, A. & Schrader, F.-W. (2001). Determinanten der Schulleistung. In D. H. Rost (Hrsg.), *Handwörterbuch pädagogische Psychologie* (2., überarbeitete und erweiterte Aufl., S. 81-91). Weinheim: Psychologie Verlags Union.

- Helmke, A. & Weinert, F. E. (1997). Bedingungsfaktoren schulischer Leistungen. In F. E. Weinert (Hrsg.), *Enzyklopädie der Psychologie: Themenbereich D Praxisgebiete, Serie I Pädagogische Psychologie, Band 3 Psychologie des Unterrichts und der Schule* (S. 71-176). Göttingen: Hogrefe.
- Helmke, A. (2005). *Unterrichtsqualität – erfassen, bewerten, verbessern* (4. Aufl.). Seelze: Kallmeyersche Verlagsbuchhandlung.
- Hubert, L. (1977). Kappa Revisited. *Psychological Bulletin*, 84, 289-297.
- Ingenkamp, K. & Lissmann, U. (2005). *Lehrbuch der Pädagogischen Diagnostik* (5., völlig überarbeitete Aufl.). Weinheim: Beltz.
- Ingenkamp, K. (1976). *Die Fragwürdigkeit der Zensurengebung* (6., überarbeitete und erweiterte Aufl.). Weinheim: Beltz.
- Jovanovic, M. (1997). Bilinguale Schulprojekte in Deutschland. Stand der Entwicklung in den Bundesländern 1996. *Bildung und Erziehung*, 50, 93-112.
- Kasten, H. (2001). Geschlechtsunterschiede. In D. H. Rost (Hrsg.), *Handwörterbuch pädagogische Psychologie* (2., überarbeitete und erweiterte Aufl., S. 212-219). Weinheim: Psychologie Verlags Union.
- Klauer, K. J. (2001a). Intelligenz und Begabung. In D. H. Rost (Hrsg.), *Handwörterbuch pädagogische Psychologie* (2., überarbeitete und erweiterte Aufl., S. 280-285). Weinheim: Psychologie Verlags Union.
- Klauer, K. J. (2001b). Lehrzielorientierte Tests. In D. H. Rost (Hrsg.), *Handwörterbuch pädagogische Psychologie* (2., überarbeitete und erweiterte Aufl., S. 401-405). Weinheim: Psychologie Verlags Union.
- Klein-Braley, C. & Raatz, U. (1982). Der C-Test: ein neuer Ansatz zur Messung von allgemeiner Sprachbeherrschung. *AKS-Rundbrief*, 4, 23-37.
- Klös, H.-P. & Plünnecke, A. (2003). Human capital formation in Germany: An untapped potential. *CESifo DICE Report*, 1 (4), 39-45.
- Koch, A. & Bündler, W. (2006). Fachbezogener Wissenserwerb im bilingualen naturwissenschaftlichen Anfangsunterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, 12, 67-76.
- Koch, A. (2005). *Bilingualer naturwissenschaftlicher Anfangsunterricht. Entwicklung – Erprobung – Evaluation*. Dissertation, Christian-Albrecht-Universität Kiel. Verfügbar unter: http://e-diss.uni-kiel.de/diss_1531 [10.09.2007].

- Kolem, U. (1981). *Intelligenz, Sprache und Leistungsbewertung: Eine empirische Untersuchung der Zusammenhänge schulrelevanter Variablen in der Grundschule*. München: Minerva-Publikation.
- Krais, B. (1983). Bildung als Kapital – Neue Perspektiven für die Analyse der Sozialstruktur. In R. Kreckel (Hrsg.), *Soziale Ungleichheiten* (S. 199-220). Göttingen: Schwartz.
- Krauth, J. (1995). *Testkonstruktion und Testtheorie*. Weinheim: Psychologie Verlags Union.
- Kuder, G. F. & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, 2, 151-160.
- Kühn, R. (1983). *Bedingungen für Schulerfolg. Zusammenhänge zwischen Schülermerkmalen, häuslicher Umwelt und Schulnoten*. Göttingen: Hogrefe.
- Kunter, M., Schümer, G., Artelt, C., Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, K.-J. & Weiß, M. (2002). *PISA 2000: Dokumentation der Erhebungsinstrumente* (Materialien aus der Bildungsforschung, Nr. 72). Berlin: Max-Planck-Institut für Bildungsforschung [Elektronische Ressource]. Verfügbar unter: <http://edoc.mpg.de/get.epl?fid=3501&did=14414&ver=0> [17.12.2006].
- Lalla, C. (2002). Die Verknüpfung von Inhalts- und Fremdsprachenlernen im bilingualen Sachfachunterricht. In: S. Breidbach, G. Bach & D. Wolff (Hrsg.), *Bilingualer Sachfachunterricht. Didaktik, Lehrer-/Lernerforschung und Bildungspolitik zwischen Theorie und Empirie* (S. 223-233). Frankfurt am Main: Lang.
- Lamsfuß-Schenk, S. & Wolff, D. (1999). Bilingualer Sachfachunterricht: Fünf kritische Anmerkungen zum state of the art. *Zeitschrift für Interkulturellen Fremdsprachenunterricht*, 4 (2), 7 Seiten.
- Lamsfuß-Schenk, S. (2002a). Bilingualer Geschichtsunterricht: Die Perspektive des Sachfaches. *Neusprachliche Mitteilungen aus Wissenschaft und Praxis*, 55, 87-96.
- Lamsfuß-Schenk, S. (2002b). Geschichte und Sprache – Ist der bilinguale Geschichtsunterricht der Königsweg zum Geschichtsbewusstsein? In: S. Breidbach, G. Bach & D. Wolff (Hrsg.), *Bilingualer Sachfachunterricht. Didaktik, Lehrer-/Lernerforschung und Bildungspolitik zwischen Theorie und Empirie* (S. 191-206). Frankfurt am Main: Lang.
- Lane, D. M. & Robertson, L. (1979). The generality of the levels of processing hypothesis: An application to memory for chess positions. *Memory & Cognition*, 7, 253-256.

- Lehrl, S. (1977). Mehrfachwahl-Wortschatz-Intelligenztest MWT-B. Erlangen: Dr. med. D. Straube.
- Lenz, T. (2003). Leistungsüberprüfung und Leistungsbewertung im bilingualen Geographieunterricht. *Geographie und Schule*, 25 (143), 38-45.
- Lienert, G. A. (1987). *Schulnoten-Evaluation*. Frankfurt am Main: Athenäum.
- Lüdtke, O., Robitzsch, A. & Köller, O. (2002). Statistische Artefakte bei Kontexteffekten in der pädagogisch-psychologischen Forschung. *Zeitschrift für Pädagogische Psychologie*, 16, 217-231.
- McGuigan, F. J. (1997). *Einführung in die experimentelle Psychologie* (5. Aufl.). Frankfurt am Main: Klotz.
- Minderop, D. (2004). Unterrichtsklima: beobachtbar – messbar? *Pädagogik*, 56 (11), 26-30.
- Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen. (1995). *Allgemeine Schulordnung mit Verwaltungsvorschriften* (11., aktualisierte Aufl.). Frechen: Ritterbach.
- Niedersächsisches Landesamt für Statistik. (2006). *Allgemein bildende Schulen im September 2005* [Elektronische Ressource]. Verfügbar unter: <http://www.nls.niedersachsen.de/Tabellen/Bildung/M3001515005.html> [15.04.2007].
- Niemeier, S. (2005). Bilingualismus und „bilinguale“ Bildungsgänge aus kognitiv-linguistischer Sicht. In G. Bach & S. Niemeier (Hrsg.), *Bilingualer Unterricht. Grundlagen, Methoden, Praxis, Perspektiven* (3., überarbeitete und erweiterte Aufl., S. 23-45). Frankfurt am Main: Lang.
- Orwin, R. G. (1994). Evaluating coding decisions. In H. Copper & L. V. Hedges (Eds.), *The handbook of research synthesis*. New York: Russell Sage Foundation.
- Prenzel, M., Carstensen, C. H., Rost, J. & Senkbeil, M. (2002). Naturwissenschaftliche Grundbildung im Ländervergleich. In J. Baumert, C. Artelt, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, K.-J. Tillmann & M. Weiß (Hrsg.), *PISA 2000: Die Länder der Bundesrepublik Deutschland im Vergleich* (S. 129-158). Opladen: Leske + Budrich.
- Prenzel, M., Drechsel, B., Carstensen, C. H. & Ramm, G. (2004). PISA 2003 – Eine Einführung. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff, J. Rost & U. Schiefele (Hrsg.), *PISA 2003: Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 13-46). Münster: Waxmann.

- Prenzel, M., Rost, J., Senkbeil, M., Häußler, P. & Klopp, A. (2001). Naturwissenschaftliche Grundbildung: Testkonzeption und Ergebnisse. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann & M. Weiß (Hrsg.), *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 191-248). Opladen: Leske + Budrich.
- Raatz, U. & Klein-Braley, C. (1983). Ein neuer Ansatz zur Messung der Sprachleistung. Der C-Test: Theorie und Praxis. In R. Horn, K. Ingenkamp & R. S. Jäger, *Test und Trends 3. Jahrbuch der Pädagogischen Diagnostik* (S. 107-138). Weinheim: Beltz.
- Ramm, G., Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M., Pekrun, R., Rolff, H.-G., Rost, J. & Schiefele, U. (Hrsg.). (2006). *PISA 2003: Dokumentation der Erhebungsinstrumente*. Münster: Waxmann.
- Rost, J., Senkbeil, M., Walter, O., Carstensen, C. H. & Prenzel, M. (2005). Naturwissenschaftliche Grundbildung im Ländervergleich. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, J. Rost & U. Schiefele (Hrsg.), *PISA 2003: Der zweite Vergleich der Länder in Deutschland – Was wissen und können Jugendliche* (S.103-124)? Münster: Waxmann.
- Rustenbach, S. J. (2003). *Metaanalyse: Eine anwendungsorientierte Einführung*. Bern: Huber.
- Sacher, W. (2005). Schulleistungsdiagnose – pädagogisch oder nach dem Modell PISA? In W. Sacher (Hrsg.), *Jenseits von PISA. Wege einer neuen Prüfungskultur* (S. 6-26). Donauwörth: Auer.
- Saldern, M. von. (1997). *Schulleistung in Deutschland – ein Beitrag zur Standortdiskussion*. Münster: Waxmann.
- Saldern, M. von. (2001). Mehrebenenanalyse. In D. H. Rost (Hrsg.), *Handwörterbuch pädagogische Psychologie* (2., überarbeitete und erweiterte Aufl., S. 457-461). Weinheim: Psychologie Verlags Union.
- Sander, W. (1931). Übergang von der Grundschule zur Mittelschule. Ein Tätigkeitsbericht. *Zeitschrift für Pädagogische Psychologie und Jugendkunde*, 32, 196-201.
- Schümer, G. (2001). Institutionelle Bedingungen schulischen Lernens im internationalen Vergleich. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann & M. Weiß (Hrsg.), *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 411-427). Opladen: Leske + Budrich.

- Sekretariat der ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2006). *Konzepte für den bilingualen Unterricht – Erfahrungsbericht und Vorschläge zur Weiterentwicklung. Bericht der Schulausschusses vom 10.04.2006*. [Elektronische Ressource]. Verfügbar unter: http://www.kmk.org/doc/publ/Konzepte-bilingualer-Unterricht_10-04-2006.pdf [08.09.2007].
- Shrout, P. E. & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, 86, 420-428.
- Spörl, F. (1998). Erdkundeunterricht in Englisch – ein Erfahrungsbereich. *Geographie heute*, 19 (158), 44-45.
- Stanat, P. & Kunter, M. (2001). Geschlechterunterschiede in Basiskompetenzen. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann & M. Weiß (Hrsg.), *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 249-269). Opladen: Leske + Budrich.
- Statistisches Bundesamt. (2007a). *GENESIS-Tabelle 21111-0003.xls* [Elektronische Ressource]. Verfügbar unter: <https://www-genesis.destatis.de/genesis/online/logon> [15.04.2007].
- Statistisches Bundesamt. (2007b). *GENESIS-Tabelle 21111-0004.xls* [Elektronische Ressource]. Verfügbar unter: <https://www-genesis.destatis.de/genesis/online/logon> [15.04.2007].
- Tarnai, C. (2001). Verbale Schulleistungsbeurteilung. In D. H. Rost (Hrsg.), *Handwörterbuch pädagogische Psychologie* (2., überarbeitete und erweiterte Aufl., S. 756-760). Weinheim: Psychologie Verlags Union.
- Tent, L. (2001). Zensuren. In D. H. Rost (Hrsg.), *Handwörterbuch pädagogische Psychologie* (2., überarbeitete und erweiterte Aufl., S. 805-811). Weinheim: Psychologie Verlags Union.
- Thomae, H. (1959). Entwicklungsbegriff und Entwicklungstheorie. In H. Thomae (Hrsg.), *Entwicklungspsychologie* (S. 3-20). Göttingen: Hogrefe.
- Thürmann, E. (2000). Zwischenbilanz zur Entwicklung der bilingualen Bildungsangebote in Deutschland. In B. Helbig, K. Kleppin & F. G. Königs (Hrsg.), *Sprachlehrforschung im Wandel. Beiträge zur Erforschung des Lehrens und Lernens von Fremdsprachen*.

- Festschrift für Karl Bausch zum 60. Geburtstag* (S. 473-497). Tübingen: Stauffenberg.
- Tücke, M. (1999). *Psychologie in der Schule – Psychologie für die Schule. Eine themenzentrierte Einführung in die Pädagogische Psychologie für (zukünftige) Lehrer* (3. unveränderte Aufl.). Münster: Lit.
- Tücke, M. (2005). *Schulische Intelligenz und Hochbegabung. Basiswissen für (zukünftige) Lehrer und Eltern*. Münster: Lit.
- Vierlinger, R. (2005). Systemänderung als Antwort auf PISA. In W. Sacher (Hrsg.), *Jenseits von PISA. Wege einer neuen Prüfungskultur* (S. 52-67). Donauwörth: Auer.
- Vollmer, H. J. (1992). Immersion und alternative Ansätze des Fremdspracherwerbs in Nordamerika: Probleme des Transfers in die Bundesrepublik Deutschland. *Zeitschrift für Fremdsprachenforschung*, 3 (2), 5-38.
- Vollmer, H. J. (2002). Leistungsfeststellung und Leistungsbewertung im bilingualen Sachfachunterricht: Ein Desideratum. In: S. Breidbach, G. Bach & D. Wolff (Hrsg.), *Bilingualer Sachfachunterricht. Didaktik, Lehrer-/Lernerforschung und Bildungspolitik zwischen Theorie und Empirie* (S. 101-121). Frankfurt am Main: Lang.
- Vollmer, H. J. (2004). Einleitung: (Fremd-)sprachlicher Kompetenzerwerb im bilingualen Sachfachunterricht. In A. Bonnet & S. Breidbach (Hrsg.), *Didaktiken im Dialog. Konzepte und Wege des Lernens im bilingualen Sachfachunterricht* (S. 63-64). Frankfurt am Main: Lang.
- Vollmer, H. J. (2005a). Bilingualer Sachfachunterricht als Inhalts- und als Sprachlernen. In G. Bach & S. Niemeier (Hrsg.), *Bilingualer Unterricht. Grundlagen, Methoden, Praxis, Perspektiven* (3., überarbeitete und erweiterte Aufl., S. 47-70). Frankfurt am Main: Lang.
- Vollmer, H. J. (2005b). Förderung des Spracherwerbs im bilingualen Sachfachunterricht. In G. Bach & S. Niemeier (Hrsg.), *Bilingualer Unterricht. Grundlagen, Methoden, Praxis, Perspektiven* (3., überarbeitete und erweiterte Aufl., S. 131-150). Frankfurt am Main: Lang.
- Vollmer, H. J. (2006a). Fachlernen und Sprachlichkeit: Aufgabenbasierte Kognition, Kooperation und Kommunikation bei bilingualen und monolingualen Lernern (S. 173-187). In A. Hahn & F. Klippel (Hrsg.), *Sprachen schaffen Chancen*,

- Dokumentation des 21. Kongresses für Fremdsprachendidaktik*. München: Oldenbourg.
- Vollmer, H. J. (2006b). Fachlichkeit und Sprachlichkeit. Zwischenbilanz eines DFG-Projekts. *Zeitschrift für Fremdsprachenforschung*, 17, 201-244.
- Wedel, H. (2006). Für oder wider verbale Beurteilungen. *Praxis Fremdsprachenunterricht*, 3 (6), 3-5.
- Weigle, S. C. (2002). *Assessing Writing*. Cambridge: Cambridge University Press.
- Weinert, F. E. (2001a). Schulleistungen – Leistungen der Schule oder der Schüler? In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 73-86). Weinheim: Beltz.
- Weinert, F. E. (2001b). Vergleichende Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 17-31). Weinheim: Beltz.
- Weiß, M. & Steinert, B. (2001). Institutionelle Vorgaben und ihre aktive Ausgestaltung – Die Perspektive der deutschen Schulleitungen. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann & M. Weiß (Hrsg.), *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 427-454). Opladen: Leske + Budrich.
- Winter, F. (2005). Neue Wege der Leistungsbewertung – das Portfolio. In W. Sacher (Hrsg.), *Jenseits von PISA. Wege einer neuen Prüfungskultur* (S.68-74). Donauwörth: Auer.
- Wode, H., Burmeister, P., Daniel, A., Kickler, K.-U. & Knust, M. (1996). Die Erprobung von deutsch-englisch bilinguaem Unterricht in Schleswig-Holstein: Ein erster Zwischenbericht. *Zeitschrift für Fremdsprachenforschung*, 7, 15-42.
- Wottawa, H. (2001). Evaluation. In D. H. Rost (Hrsg.), *Handwörterbuch pädagogische Psychologie* (2., überarbeitete und erweiterte Aufl., S. 152-158). Weinheim: Psychologie Verlags Union.
- Wottawa, H. (2006). Evaluation. In A. Krapp & B. Weidenmann (Hrsg.), *Pädagogische Psychologie* (5., vollständig überarbeitete Aufl., S. 661-687). Weinheim: Beltz.
- Zimbardo, P. G. & Gerrig, R. J. (1999). *Psychologie*. Berlin: Springer.
- Zimmer, K., Burba, D. & Rost, J. (2004). Kompetenzen von Jungen und Mädchen. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff, J. Rost & U. Schiefele (Hrsg.), *PISA 2003: Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 211-216). Münster: Waxmann.

Zydati, W. (2004). Einleitung: Sachfachlicher Kompetenzerwerb im bilingualen Sachfachunterricht. In A. Bonnet & S. Breidbach (Hrsg.), *Didaktiken im Dialog. Konzepte und Wege des Lernens im bilingualen Sachfachunterricht* (S. 89-90). Frankfurt am Main: Lang.

Zydati, W. (2005). Diskursfunktionen in einem analytischen curricularen Zugriff auf Textvarietten und Aufgaben des bilingualen Sachfachunterrichts. *Fremdsprachen Lehren und Lernen*, 34, 156-173.

7 Autorenverzeichnis

- Alderson, J. C., 73, 105
Amthauer, R., 8, 105
Artelt, C., 73, 105, 110, 111
Asendorpf, J. oder Asendorpf, J. B. (vgl. Literaturverzeichnis), 13, 74, 105
Avenarius, H., 40, 41, 105
Bach, G., 1, 25, 28, 105, 106, 110, 111, 114
Baumert, J., 2, 15, 16, 17, 19, 20, 21, 36, 44, 46, 96, 105, 107, 110, 111, 112, 113, 115, 121
Beauducel, A., 8, 105, 106
Beer, F., 9, 14, 46, 95, 106
Bloom, B. S., 4, 5, 6, 106, 120
Blum, W., 107, 111, 112, 115
Bonnet, A., 114, 116
Borg, I., 40, 106
Bortz, J., 40, 42, 76, 106
Bos, W., 2, 35, 106
Bourdieu, P., 17, 18, 106
Bredenbröcker, W., 30, 31, 44, 94, 106
Breidbach, S., 25, 28, 106, 110, 114, 116
Brocke, B., 8, 105, 106
Bülter, H., 22, 47, 96, 106
Bünder, W., 32, 33, 43, 93, 109
Bundesministerium für Bildung und Forschung, 107
Burba, D., 12, 115
Burmeister, P., 29, 115
Carstensen, C. H., 12, 36, 111, 112
Chasiotis, A., 13, 107
Christ, I., 1, 25, 107
Clapham, C., 73, 105
Cohen, J., 74, 107, 108
Coleman, J. S., 18, 107
Conger, A. J., 74, 107
Craik, F. I., 34, 107
Daniel, A., 29, 115
De Graaf, P. M., 16, 108
Decke-Cornill, H., 27, 107
Demmrich, A., 36, 105
Diehl, J. M., 40, 107
Döring, N., 40, 42, 106
Drechsel, B., 36, 111
Eckensberger, L. H., 13, 107
Eder, F., 22, 23, 24, 107
Ehmke, T., 20, 46, 96, 107
Erikson, R., 16, 108, 119
Farkas, G., 24, 108
Faust, B., 61, 95, 108
Fleiss, J. L., 74, 108, 113
Ganzeboom, H. B. G., 15, 16, 17, 20, 108, 121
Gerrig, R. J., 8, 115
Gold, A., 43, 108
Goldthorpe, J. H., 16, 108, 119
Grotjahn, R., 61, 62, 95, 108
Hahn, A., 114
Hauser, R. M., 24, 108
Häußler, P., 21, 112
Heidemeier, H., 20, 46, 96, 107
Helbig, B., 113
Heller, K. A., IV, 8, 56, 57, 58, 59, 98, 108, 143, 144
Helmke, A., 2, 4, 6, 8, 14, 22, 47, 108, 109
Hohensee, F., 20, 46, 96, 107
Hubert, L., 74, 109
Ingenkamp, K., 36, 42, 109, 112
Jovanovic, M., 26, 109
Kasten, H., 11, 13, 109
Keller, H., 13, 105, 107
Kickler, K.-U., 29, 115
Klauer, K. J., 8, 38, 109
Klein-Braley, C., IV, 45, 60, 61, 95, 109, 112
Kleppin, K., 113
Klieme, E., 105, 110, 111, 112, 113, 115
Klippel, F., 114
Klopp, A., 21, 112
Klös, H.-P., 18, 109
Knust, M., 29, 115
Koch, A., 32, 33, 43, 44, 57, 93, 103, 109
Kohr, H. U., 40, 107
Kolem, U., 9, 46, 95, 110
Köller, O., 24, 111
Königs, F. G., 113
Krais, B., 18, 110
Krapp, A., 115
Krauth, J., 110
Kreckel, R., 106, 110
Kuder, G. F., 58, 110
Kühn, R., 7, 8, 10, 46, 95, 110, 121
Kunter, M., 11, 12, 47, 56, 97, 110, 113

- Kutalek, N., 9, 46, 95, 106
Lalla, C., 32, 110
Lamsfuß-Schenk, S., 27, 31, 43, 93, 110
Lane, D. M., 34, 110
Lehmann, R., 107, 111, 112, 115
Lehrl, S., 8, 111
Lenz, T., 43, 111
Leutner, D., 107, 111, 112, 115
Lienert, G. A., 42, 57, 58, 111
Lissmann, U., 36, 109
Lockhart, R. S., 34, 107
Lüdtke, O., 24, 111
McGuigan, F. J., 50, 111
Meyer, H., 22, 47, 96, 106
Minderop, D., 22, 111
Ministerium für Schule und Weiterbildung
des Landes Nordrhein-Westfalen, 39,
111, 119
Neubrand, M., 105, 107, 110, 111, 112,
113, 115
Niedersächsisches Landesamt für Statistik,
65, 69, 111
Niemeier, S., 25, 105, 111, 114
Orwin, R. G., 74, 77, 111
Pekrun, R., 107, 111, 112, 115
Perleth, C., IV, 8, 56, 57, 58, 59, 98, 108,
143, 144
Plünnecke, A., 18, 109
Portocarero, L., 16, 108, 119
Postlethwaite, T. N., 2, 35, 106
Prenzel, M., 12, 20, 21, 36, 46, 96, 105,
107, 110, 111, 112, 113, 115
Raatz, U., IV, 45, 57, 58, 60, 61, 95, 109,
112
Ramm, G., 20, 36, 111, 112
Richardson, M. W., 58, 110
Robertson, L., 34, 110
Robitzsch, A., 24, 111
Rolf, H.-G., 107, 111, 112, 115
Rost, D. H., 12, 21, 105, 107, 108, 109,
111, 112, 113, 115
Rost, J., 12, 21, 105, 107, 108, 109, 111,
112, 113, 115
Rustenbach, S. J., 74, 77, 112
Sacher, W., 38, 42, 112, 114, 115
Sander, W., 8, 112
Sarges, W., 106
Schiefele, U., 73, 105, 107, 110, 111, 112,
113, 115
Schneider, W., 73, 105, 110, 111, 112, 113,
115
Schnell, H., 9, 46, 95, 106
Schrader, F.-W., 6, 7, 8, 14, 108
Schümer, G., 15, 16, 17, 19, 20, 21, 35, 44,
46, 96, 105, 110, 112, 121
Sekretariat der ständigen Konferenz der
Kultusminister der Länder in der
Bundesrepublik Deutschland, 2, 113
Senkbeil, M., 12, 21, 111, 112
Shrout, P. E., 74, 113
Souvignier, E., 43, 108
Spörl, F., 29, 113
Stanat, P., 11, 12, 36, 47, 73, 97, 105, 110,
112, 113, 115
Statistisches Bundesamt, 65, 69, 113
Staufenbiel, T., I, VI, 40, 106
Steinert, B., 22, 115
Tarnai, C., 41, 113
Tent, L., 41, 42, 113
Thomae, H., 13, 113
Thürmann, E., 1, 26, 113
Tillmann, K.-J., 105, 110, 111, 112, 113,
115
Treiman, D. J., 15, 16, 17, 20, 108, 121
Tücke, M., 8, 38, 114
Vierlinger, R., 42, 114
Volland, E., 13, 107
Vollmer, H. J., I, VI, 1, 2, 21, 25, 28, 29,
33, 43, 44, 45, 51, 52, 57, 93, 94, 103,
114, 115
Wall, D., 73, 105
Wallbott, H. G., 74, 105
Walter, O., 12, 112
Wedel, H., 41, 115
Weidenmann, B., 115
Weigle, S. C., 73, 115
Weinert, F. E., 4, 5, 6, 8, 14, 22, 34, 47, 99,
106, 109, 115
Weiß, M., 22, 105, 106, 110, 111, 112, 113,
115
Winter, F., 42, 115
Wode, H., 29, 30, 31, 44, 94, 115
Wolff, D., 27, 106, 110, 114
Wottawa, H., 49, 106, 115
Zimbardo, P. G., 8, 115
Zimmer, K., 12, 115
Zydati, W., 26, 51, 116

8 Abkürzungsverzeichnis

CLIL	Content and Language Integrated Learning
df	degrees of freedom (Freiheitsgrade)
DFG	Deutsche Forschungsgemeinschaft
ebd.	ebenda
EGP	Class Schema of Erikson, Goldthorpe and Portocarero
ESCS	Index of Economic, Social and Cultural Status
et al.	et alii (und andere)
ggf.	gegebenenfalls
ISEI	Socio-Economic Index of occupational status
KFT	Kognitiver Fähigkeitstest
K-S-Test	Kolmogorov-Smirnov-Tests
L1	erste Sprache, Erstsprache
L2	zweite Sprache, Zweitsprache ⁵ , erste Fremdsprache ⁶
M	Mittelwert
Max.	Maximum
Min.	Minimum
NDS	Niedersachsen
OECD	Organisation for Economic Co-Operation and Development (Organisation für wirtschaftliche Zusammenarbeit und Entwicklung)
PISA	Programme for International Student Assessment (Programm zur internationalen Schülerbeurteilung)
SD	Standard deviation (Standardabweichung)
SE	Standard error (Standardfehler)
SIOPS	Standard International Occupational Prestige Scale
SKMK	Sekretariat der ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland
SM-NRW	Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen
vgl.	vergleiche
z.B.	zum Beispiel

⁵ Wenn innerhalb eines muttersprachlichen Kontextes erlernt.

⁶ Wenn außerhalb eines muttersprachlichen Kontextes erlernt.

9 Abbildungsverzeichnis

Abbildung 1	Wichtige Variablen im Modell schulischen Lernens von Bloom (1976).....	6
Abbildung 2	Übersicht zu den Einflussfaktoren von Schulleistung.....	7
Abbildung 3	Auszug aus dem Schülerfragebogen: Berufstätigkeit der Eltern.	55
Abbildung 4	Auszug aus dem Schülerfragebogen: Klassenklima.....	55
Abbildung 5	Übersichtsschema zur Durchführung der Datenerhebung.....	72
Abbildung 6	Eigenwertediagramm der Faktorenanalyse zum Klassenklima	80

10 Tabellenverzeichnis

Tabelle 1	Korrelationstabelle zu Kühn (1983)	10
Tabelle 2	EGP-Klassenschema aus Ganzeboom und Treiman (1996)	17
Tabelle 3	EGP-Klassenschema aus Baumert und Schümer (2001).....	20
Tabelle 4	Übersicht zu Konstrukten und Operationalisierungen.....	50
Tabelle 5	Übersicht zu Diskursfunktionen im Geografietest	52
Tabelle 6	Übersicht Erhebungsgruppen Gesamtstichprobe	64
Tabelle 7	Übersicht Geschlechterverteilung Gesamtstichprobe.....	65
Tabelle 8	Übersicht Altersverteilung Gesamtstichprobe	66
Tabelle 9	Übersicht Erhebungsgruppen Hauptstichprobe.....	68
Tabelle 10	Übersicht Geschlechterverteilung Hauptstichprobe	69
Tabelle 11	Übersicht Altersverteilung Hauptstichprobe.....	70
Tabelle 12	Tabelle zum U-Test bezüglich sozialer Herkunft in den Stichproben.....	71
Tabelle 13	Übersicht der verwendeten Variablen	75
Tabelle 14	Beurteilerübereinstimmung Geografietest.....	77
Tabelle 15	Itemkennwerte der Skala zum Klassenklima (sieben Items)	78
Tabelle 16	Itemkennwerte der Skala zum Klassenklima (fünf Items).....	79
Tabelle 17	Skalenkennwerte der Skala zum Klassenklima bei Itementfernung	79
Tabelle 18	Ergebnisse der Faktorenanalyse zum Klassenklima.....	81
Tabelle 19	Korrelationstabelle KFT/Schulleistungen.....	84
Tabelle 20	Deskriptive Statistiken zur sozialen Herkunft.....	86
Tabelle 21	Univariate Varianzanalysen und Levene-Tests zur sozialen Herkunft	87
Tabelle 22	Korrelationstabelle Klassenklima/Schulleistungen	88
Tabelle 23	Korrelationstabelle Geografiekompetenz	90
Tabelle 24	Ergebnisse der Regression zur Geografiekompetenz	91
Tabelle A1	Verteilungskennwerte der KFT-Subtests.....	143
Tabelle A2	Verteilungskennwerte der KFT-Subtests (Ergänzung).....	144
Tabelle A3	Verteilungskennwerte der KFT-Subtests nach Geschlecht.....	144
Tabelle A4	Verteilungskennwerte der KFT-Subtests nach Geschlecht (Ergänzung)	145

11 Anhang

Anhang A Schülerfragebogen für bilinguale Schüler

1. Schülercode

2. Schule

3. Klasse

4. Geschlecht 01 männlich
 02 weiblich

5. Alter (in Jahren, Monaten):

Heutiges Datum:.....

In diesem Fragebogen findest du Fragen über dich und deine Erfahrungen mit dem Erdkundeunterricht. Lies bitte die Fragen sorgfältig durch und beantworte sie so genau wie möglich. Die meisten Fragen sind durch Ankreuzen eines Kästchens zu beantworten, andere bieten mehrere Antwortmöglichkeiten oder können frei beantwortet werden. Deine Antworten werden vertraulich behandelt. Kein Lehrer und keine Lehrerin deiner Schule werden deine Antworten sehen.

Fülle den Fragebogen bitte sachlich und ernsthaft aus.

Vielen Dank für deine Mitarbeit.

6. In welcher Klassenstufe hattest du Erdkundeunterricht auf Englisch?

- 01 Klasse 7, 1. Halbjahr
- 02 Klasse 7, 2. Halbjahr
- 03 Klasse 8, 1. Halbjahr
- 04 Klasse 8, 2. Halbjahr
- 05 Klasse 10, 1. Halbjahr
- 06 Klasse 10, 2. Halbjahr

7. Hattest du in dieser Zeit durchgängig Erdkundeunterricht auf Englisch oder wurdest du zwischenzeitlich auf Deutsch unterrichtet?

- 01 Ich hatte durchgängig Erdkunde auf Englisch.
- 02 Ich hatte zwischenzeitlich Erdkundeunterricht auf Deutsch.

8. Wenn du zwischenzeitlich Erdkunde auf Deutsch hattest, wie lange? Bitte nur in diesem Fall und dann nur eine Möglichkeit ankreuzen.

- 01 ein Halbjahr
- 02 zwei Halbjahre
- 03 drei Halbjahre
- 04 vier Halbjahre
- 05 fünf Halbjahre
- 06 Mehr als fünf Halbjahre

9. Wirst/Wurdest du in anderen Fächern als Erdkunde auf Englisch unterrichtet? 01 nein

Wenn ja: **In welchen Fächern** wirst/wurdest du auf **Englisch** unterrichtet? Du kannst **mehrere** Antworten ankreuzen.

- 02 Geschichte
- 03 Politik
- 04 Sport
- 05 Mathematik
- 06 Kunst
- 07 Sozialkunde
- 08 Biologie
- 09 Sonstige, nämlich

10. Seit wann? und 11. wie lange wirst/wurdest du in diesen Fächern auf Englisch unterrichtet?

Seit wann?

Wie lange?

Geschichte	02	02
Politik	03	03
Sport	04	04
Mathematik	05	05
Kunst	06	06
Sozialkunde	07	07
Biologie	08	08
Sonstige	10	10

Welche der folgenden Aussagen trifft auf dich zu? Bitte jeweils nur eine Möglichkeit ankreuzen.

		01	02	03	04
		stimme nicht zu	stimme eher nicht zu	stimme eher zu	stimme zu
25.	Ich finde es wichtig , dass man im Erdkundeunterricht beigebracht bekommt, wie man erdkundespezifische Sachverhalte sprachlich angemessen ausdrückt .	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
26.	Ich finde es schwierig, Fachsprache in Erdkunde zu verstehen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
27.	In unserem Erdkundeunterricht wird gezielt auf fachsprachliche Aspekte geachtet , z.B. um uns beizubringen, wie man verschiedene Materialien auswertet, welches Vokabular man dabei verwendet oder wie man sich zu Inhalten äußert.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
28.	Ich finde es leicht, mich in Erdkunde fachsprachlich angemessen auszudrücken.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
29.	Ich verstehe manchmal nicht, was mit bestimmten Fachausdrücken in Erdkunde gemeint ist.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
30.	Ich verwende manchmal in Erdkunde Fachbegriffe , ohne sie ganz zu verstehen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
31.	Ich kann mich in allen Fächern fachsprachlich angemessen ausdrücken.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
32.	Ich verstehe manchmal nicht, was mit bestimmten Fachausdrücken in anderen Fächern gemeint ist.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
33.	Ich verwende manchmal in anderen Fächern Fachbegriffe, ohne sie ganz zu verstehen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

34. Schreib auf, welche Fächer du **am liebsten** magst.

.....

35. Bringe die für dich **wichtigsten Fächer** in eine **Rangfolge von Beliebtheit**.

- | | | | |
|----|-------|-----|-------|
| 1. | | 7. | |
| 2. | | 8. | |
| 3. | | 9. | |
| 4. | | 10. | |
| 5. | | 11. | |
| 6. | | 12. | |

36. Wie werden die **Ergebnisse** des Erdkundeunterrichts in der Regel **festgehalten**?

- 01 schriftlich
- 02 mündlich
- 03 etwa zu gleichen Teilen schriftlich und mündlich

Bitte jeweils nur eine Möglichkeit ankreuzen.

		01	02	03	04
		nie	ungefähr zweimal im Halbjahr	ungefähr zweimal im Monat	ungefähr einmal pro Woche
37.	Wie oft musst du im Erdkundeunterricht der 10. Klasse schriftliche Antworten verfassen, die länger sind als ein Satz?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
38.	Wie oft werdet Ihr im Erdkundeunterricht angehalten, größere Zusammenhänge herzustellen und diese zu versprachlichen ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
39.	Wie oft musst du deine Antwort(en) im Erdkundeunterricht der 10. Klasse begründen ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
40.	Wie oft musst du im Erdkundeunterricht der 10. Klasse ein bestimmtes Argument entwickeln und deinen Standpunkt dazu vertreten ?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
41.	Wie oft müsst ihr Texte oder Aufgabenlösungen, die ihr im Erdkundeunterricht geschrieben habt, verbessern bzw. überarbeiten?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

42. Welche **Hilfestellung** zur **Darstellung von Beziehungen** zwischen unterschiedlichen Faktoren werden im Erdkundeunterricht verwendet? Du kannst **mehrere Antworten** ankreuzen.

- 01 Keine
- 02 Pfeile zwischen zwei Begriffen
- 03 Mind map
- 04 Strukturskizze
- 05 Liste mit sprachlichen Ausdrücken für logische Beziehungen
- 06 Sonstige, nämlich

43. Welche **Formen der Überarbeitung** wendet ihr an? Du kannst **mehrere Antworten** ankreuzen.

- 01 Keine
- 02 Rechtschreibung und Zeichensetzung
- 03 Wortwahl
- 04 Satzkonstruktionen
- 05 Aufbau/Logik des Textes
- 06 Korrektur eines Faktums
- 07 Ändern inhaltlicher Aussagen
- 08 Sonstige, nämlich

44. **Wie viele Fremdsprachen** lernst du in der Schule? Kreuze bitte **nur eine** Antwort an.

- 01 Zwei Sprachen
- 02 Drei Sprachen
- 03 Vier Sprachen

45. **Welche Fremdsprachen** lernst du in der Schule?

.....

46. Mit welcher **Fremdsprache** hast du in der jeweiligen Klassenstufe **begonnen**? Fülle ggf. alle drei Felder aus.

01 5. Klasse Fremdsprache:

02 7. Klasse Fremdsprache:

03 9. Klasse Fremdsprache:

63. Hast du sonstige Kontakte mit dem Englischen?

01 nein 02 ja

64. Wenn ja, welche?

Welche Aussage trifft auf dich zu? Bitte jeweils nur eine Antwort ankreuzen.

65. Mir gefällt mein **Erdkundeunterricht** ...

01 überhaupt nicht 02 weniger gut 03 mittelmäßig 04 eher gut 05 sehr gut

66. Wie oft beschäftigst du dich **außerhalb des Erdkundeunterrichts** gezielt mit **Erdkundethemen**?
Bitte **nur eine Antwort** ankreuzen.

01 nie 02 selten 03 manchmal 04 häufig

67. Wenn ja: Mit **welchen Erdkundethemen**?

.....

Bitte die folgenden beiden Frage nur bearbeiten, wenn soeben mit „häufig“ oder „manchmal“ geantwortet wurde, sonst einfach offen lassen.

68. Wenn ja, **in welcher Form**? Du kannst **mehrere** Antworten ankreuzen.

01	<input type="checkbox"/>	über Zeitschriften
02	<input type="checkbox"/>	über das Internet
03	<input type="checkbox"/>	über Vereinstätigkeit (Umweltverbände o.ä.)
04	<input type="checkbox"/>	auf andere Art, nämlich

69. Wenn ja, auf **Englisch** oder auf **Deutsch**? Bitte **nur eine** Möglichkeit ankreuzen.

01 hauptsächlich auf Englisch
02 zu etwa gleichen Teilen auf Englisch und Deutsch
03 hauptsächlich auf Deutsch

70. Welche der folgenden Aktivitäten übst du **außerhalb der Schule** aus? Du kannst **mehrere Antworten** ankreuzen.

01 Reiseführer lesen
02 Fernsehsendungen über andere Länder und Kulturen ansehen
03 Fernsehsendungen über Naturphänomene und Naturkatastrophen ansehen
04 Städte und Länder im Atlas lokalisieren, die in den Nachrichten aktuell besprochen werden

71. Wendest du Erdkundekenntnisse **außerhalb der Schule** an? Bitte **nur eine Antwort** ankreuzen.

01 nie 02 selten 03 manchmal 04 oft

72. Besuchst du **auf Reisen** kulturelle und natürliche **Sehenswürdigkeiten** der jeweiligen Region?
Bitte **nur eine Antwort** ankreuzen.

- | | | | | | | | | | |
|----|--------------------------|----|--------------------------|----|--------------------------|----|--------------------------|----|--------------------------|
| 01 | nie oder
fast nie | 02 | selten | 03 | manchmal | 04 | oft | 05 | (fast)
immer |
| | <input type="checkbox"/> | | <input type="checkbox"/> | | <input type="checkbox"/> | | <input type="checkbox"/> | | <input type="checkbox"/> |

73. Welche der folgenden Dinge machst du manchmal? Du kannst **mehrere Antworten** ankreuzen.

- 01 Sich mit Hilfe von Karten über die Lage von Städten oder über Fahrtziele informieren
- 02 Einen Ort in einer größeren Stadt mit Hilfe eines Stadtplans auffinden
- 03 Beim Kauf eines Produkts über Herkunft und Herstellung nachdenken
- 04 Behutsam mit Ressourcen (z.B. Wasser) im Alltag umgehen
- 05 Sonstiges:.....

74. Mit wem **wohnst du zusammen**? Du kannst **mehrere Antworten** ankreuzen.

- 01 Mutter
- 02 Vater
- 03 Stiefmutter/Partnerin des Vaters
- 04 Stiefvater/Partner der Mutter
- 05 (Stief-)Bruder/Brüder
- 06 (Stief-)Schwester/n
- 07 Großmutter
- 08 Großvater
- 09 andere

Haben deine **Eltern Abitur**?

75. **Mutter** 01 nein 02 ja 03 Fachabitur
76. **Vater** 01 nein 02 ja 03 Fachabitur

Haben deine **Eltern ein Studium** abgeschlossen?

77. **Mutter** 01 nein 02 ja
78. **Vater** 01 nein 02 ja

79. Welchen **Beruf** übt deine **Mutter** aus?

Beruf

80.

- 01 z.Z. nicht berufstätig
- 02 Teilzeit
- 03 Vollzeit

81. Welchen **Beruf** übt dein **Vater** aus?

Beruf

82.

- 01 z.Z. nicht berufstätig
- 02 Teilzeit
- 03 Vollzeit

83. Gibt es bei **dir zu Hause** einen ruhigen Ort, an dem du arbeiten kannst?

01 nein 02 ja

84. Gibt es bei dir **zu Hause** ...

01 ... einen Computer **mit** Internetanschluss

02 ... Wörterbuch (Deutsch–Englisch)

03 ... Nachschlagewerke (Buch, CD-ROM)

04 ... sonstige Informationsträger/Medien?

Welche?

.....
Du kannst **mehrere** Antworten ankreuzen.

Welche Antwort trifft auf dich zu? Bitte nur eine Antwort ankreuzen.

85. Zu Hause **sprechen** wir über **aktuelle Fragen, über Bücher, Filme, Fernsehsendungen** und dergleichen ...

01	nie oder fast nie	02	etwa einmal im Monat	03	mehrmals im Monat	04	etwa einmal pro Woche	05	(fast) jeden Tag
	<input type="checkbox"/>		<input type="checkbox"/>		<input type="checkbox"/>		<input type="checkbox"/>		<input type="checkbox"/>

86. Besitzt du einen **eigenen Computer**?

01 nein 02 ja

87. Schätze ungefähr, **wie viele Bücher** es bei **dir zu Hause** gibt. (Als Hilfe: Ein Regalmeter fasst etwa 40 Bücher.) Bitte **nur eine** Möglichkeit ankreuzen.

- 01 Keine Bücher
- 02 0–40 Bücher (unter 1 Regalmeter)
- 03 41–80 Bücher (zwischen 1–2 Regalmeter)
- 04 81–160 Bücher (zwischen 2–4 Regalmeter)
- 05 161–240 Bücher (zwischen 4–6 Regalmeter)
- 06 241–480 Bücher (zwischen 6– 12 Regalmeter)
- 07 Mehr als 480 Bücher (über 12 Regalmeter)

88. **Wie viele Bücher** gehören davon **dir**? Bitte **nur eine** Möglichkeit ankreuzen.

- 01 Keine Bücher
- 02 0–40 Bücher (unter 1 Regalmeter)
- 03 41–80 Bücher (zwischen 1–2 Regalmeter)
- 04 81–160 Bücher (zwischen 2–4 Regalmeter)
- 05 161–240 Bücher (zwischen 4–6 Regalmeter)
- 06 241–480 Bücher (zwischen 6– 12 Regalmeter)
- 07 Mehr als 480 Bücher (über 12 Regalmeter)

89. Wie oft nehmt ihr in deiner Familie **gemeinsame Mahlzeiten** ein?

01	nie oder fast nie	02	selten	03	nur am Wochenende	04	etwa einmal pro Woche (Mo–Fr)	05	(fast) jeden Tag
	<input type="checkbox"/>		<input type="checkbox"/>		<input type="checkbox"/>		<input type="checkbox"/>		<input type="checkbox"/>

90. Werden in eurem Haushalt **Zeitschriften** oder **Zeitungen** gelesen?

01 nein 02 ja

91. Wenn ja, **welche**? Du kannst **mehrere Antworten** ankreuzen; wenn bei euch keine Zeitungen oder Zeitschriften gelesen werden, lasse alle Kästchen frei.

01 Eine Tageszeitung

02 Der Spiegel

03 Stern

04 Bildzeitung

05 Die Zeit

06 Fokus

07 Fachzeitschriften (Ärzteblatt, Computerzeitschriften, ...)

08 Lifestyle-Zeitschriften (Freundin, Schöner Wohnen, ...)

09 andere, nämlich

92. Gibt es bei dir zuhause einen **Fernseher**?

01 nein 02 ja

93. **Wie häufig** siehst du fern?

01 Weniger als eine Stunde pro Woche

02 Etwa ein bis vier Stunden pro Woche

03 Etwa vier bis acht Stunden pro Woche

04 Etwa ein bis zwei Stunden täglich

05 Etwa zwei bis vier Stunden täglich

06 Mehr als vier Stunden täglich

94. **Welche Art Sendungen** siehst du dir vor allem an?

.....
.....

95. Wie wichtig ist es für deine Eltern, dass du **gute Noten** nach Hause bringst? (Du kannst frei antworten).

.....
.....

96. Wie wichtig ist es für deine Eltern, dass du gute Noten **in Erdkunde** hast? (Du kannst frei antworten).

.....
.....

97. Wenn du die Wahl hast, im Erdkundeunterricht **allein** an einer Aufgabe zu arbeiten **oder mit anderen zusammen**, wofür entscheidest du dich?

- 01 Einzelarbeit
- 02 Paararbeit
- 03 Gruppenarbeit (mehr als zwei Personen)

Beantworte von den folgenden drei Fragenblocks bitte nur einen, je nachdem, ob du soeben „Einzelarbeit“, „Paararbeit“ oder „Gruppenarbeit“ angekreuzt hast.

98. Wenn du dich für **Einzelarbeit** entscheidest, **warum**? Du kannst **mehrere Antworten** ankreuzen bzw. ausfüllen.

- 01 Weil man häufig die Arbeit für zwei machen muss, wenn sich der andere nicht beteiligt
- 02 Weil es länger dauert, wenn man sich einigen muss
- 03 Weil man sich auf den anderen einstellen muss, dies Energie beansprucht und nicht zur eigentlichen Lösung beiträgt
- 04 Anderer Grund, und zwar

99. Wenn du dich für **Paararbeit** entscheidest, warum? Du kannst **mehrere Antworten** ankreuzen bzw. ausfüllen.

- 01 Weil man von dem profitieren kann, was dein Arbeitspartner weiß.
- 02 Weil man zu zweit schneller zu einer Lösung kommt.
- 03 Weil man dabei mehr lernt, als wenn man allein arbeitet.
- 04 Weil man zu zweit mehr Spaß hat.
- 05 Weil man zu einer besseren Lösung kommt als allein.
- 06 Anderer Grund, und zwar

100. Wenn du dich für **Gruppenarbeit** entscheidest, warum? Du kannst **mehrere Antworten** ankreuzen bzw. ausfüllen.

- 01 Weil es in der Gruppe mehr Spaß macht.
- 02 Weil man in der Gruppe noch mehr herausfindet als in Einzel- oder Paararbeit.
- 03 Weil man sich in der Gruppe die Arbeit aufteilen kann.
- 04 Weil man in der Gruppe weniger Arbeit hat.
- 05 Anderer Grund, und zwar

Welche Aussage trifft auf dich zu? Bitte nur eine Möglichkeit ankreuzen.

		01	02	03	04
		stimme nicht zu	stimme eher nicht zu	stimme eher zu	stimme zu
101.	Wenn ich im Erdkundeunterricht auf Englisch eine Aufgabe lösen soll, arbeite ich gerne mit einem Partner zusammen .	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
102.	Wenn ich im Erdkundeunterricht auf Deutsch eine Aufgabe lösen soll, arbeite ich gerne mit einem Partner zusammen .	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

103. Wie häufig (schätzt du) **denkst du** beim Bearbeiten von Aufgaben im bilingualen Erdkundeunterricht auf **Deutsch** bzw. auf **Englisch**?

- 01 ausschließlich auf Deutsch 02 eher auf Deutsch 03 etwa zu gleichen Teilen auf Deutsch+Englisch 04 eher auf Englisch 05 ausschließlich auf Englisch

104. Wenn du im bilingualen Erdkundeunterricht Texte auf Englisch **lesen** oder Aufgaben lösen sollst, **übersetzt** du sie dann für gewöhnlich erst ins Deutsche?

- 01 nein
 02 ja
 03 Kommt darauf an, nämlich

105. Wenn im bilingualen Erdkundeunterricht einen Text auf Englisch **schreiben** sollst, **formulierst** du ihn dann direkt auf Englisch oder zuerst auf Deutsch und übersetzt ihn dann?

- 01 formuliere ihn direkt auf Englisch 02 formuliere ihn erst auf Deutsch im Kopf vor

Kommentare:

.....

.....

.....

.....

.....

.....

Bitte pro Zeile nur eine Antwort ankreuzen.

		01 viel zu gering	02 eher gering	03 gerade richtig	04 eher groß	05 viel zu groß
106.	Ich empfinde den Umfang meiner Hausaufgaben in Erdkunde als	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
107.	Ich empfinde den Schwierigkeitsgrad meiner Hausaufgaben in Erdkunde als	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
108.	Ich empfinde den Umfang meiner Hausaufgaben in der Schule insgesamt als	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
109.	Ich empfinde den Schwierigkeitsgrad meiner Hausaufgaben in der Schule insgesamt als	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
110.	Ich empfinde die Leistungsanforderungen in der Schule als	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
111.	Ich empfinde die Leistungsanforderungen in Erdkunde als	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

138. Welche der folgenden Aktivitäten kamen bzw. kommen in deinem jetzigen Erdkundeunterricht vor und

139. welche dieser Tätigkeiten oder Arbeitsformen magst du besonders?

Kreuze an, was du für zutreffend hältst. Dazu sollst du pro Zeile jeweils zwei Kreuze setzen, nämlich eines unter „kommt vor“ und eines unter „gefällt mir“.

Nr.	Aktivität oder Arbeitsform im Erdkundeunterricht	kommt vor			gefällt mir ...		
		nie/ selten	manch mal	eher häufig	eher nicht	Etwas	sehr
01	Klassenunterricht (frontal): Lehrer fragt, Schüler antworten						
02	Einzelarbeit						
03	Partnerarbeit						
04	Gruppenarbeit (ab 3 Personen)						
06	Auswerten von Diagrammen, Grafiken, Statistiken, Profilen, Querschnitten						
07	Diskussionen in der gesamten Klasse, offene Gespräche						
08	Erarbeiten schriftlicher Sachtexte aus Schulbüchern zu geografischen Themen						
09	Schüler-Kurzreferate						
10	Arbeit mit Karten und Plänen						
11	Sprachliche Übungen zum Ausdruck geografischer Sachverhalte						
12	Erstellen oder Übernehmen von Listen geografischer Fachbegriffe						
13	Tests zu geografischen Fachbegriffen						
14	Arbeit mit Bildern, Fotos, Dias (Bildquellen)						
15	Abschreiben von der Tafel						
16	Wichtige Inhalte eines Textes feststellen (z.B. durch Markieren)						
17	Schriftliche Erörterung von fachbezogenen Themen oder Problemen (Entwicklungen, Pro und Kontra)						
18	Fächerübergreifender Unterricht						
19	Freiarbeit: Schüler können selbständig Themen und Aufgaben wählen						
20	Projektarbeit: Schüler arbeiten in Gruppen zu einem Thema und präsentieren abschließend die Produkte ihrer Arbeit						
21	Unterschiedlich starke Schüler bearbeiten unterschiedlich schwere Aufgaben						
22	Schüler bringen sich gegenseitig den Lernstoff bei						

Vielen Dank für deine Mitarbeit!

Anhang B Geografietest für bilinguale Schüler (Auszug)

TASK 1. DISTRIBUTION

In this section, we would first like you to locate the tropical rain forests. Then, you are going to look at specific areas in the tropical rain forest in more detail.

Please write down your answers on the separate answering sheets.

- Have a look at the map of the world (Figure 1) showing the distribution of the tropical rain forest. Describe as precisely as possible where exactly in the world tropical rain forests can be found. Please, do not write on the map!
- Study the climate graph for Kisangani (Figure 2). Describe the exact location of the Kisangani weather station by also looking at the map (Figure 1).
- What type of climate does Kinsangani have? Describe Kisangani's climate in detail with the help of the chart.
- Medan is a place in Indonesia and has the following coordinates: 3° N/98° E. Is the climate in Medan different from the climate in Kisangani? Motivate your answer.
- Take a look at South America in Figure 1. Why are there no tropical rain forests in the west of South America?

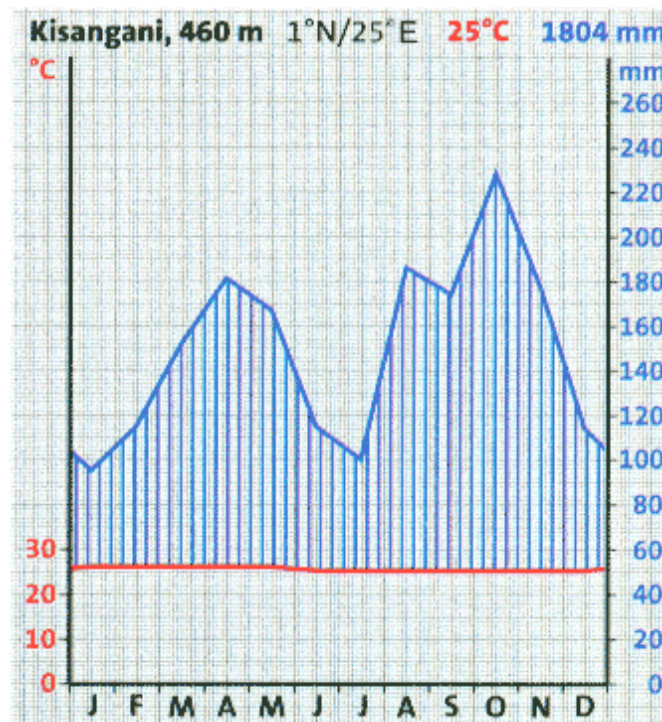



Fig. 2: Climate graph for Kisangani

TASK 6. SUSTAINABILITY

Since the world summit of Rio de Janeiro in 1992 (Agenda 21) the principle of sustainability has been globally accepted. This principle means that in terms of the use of natural resources, ecological, social and economic goals should be treated as equally important. At the same time, the rights and needs of future generations should be respected, so that they are not disadvantaged in any way through the exploitation of the resources.

Against the background of the aims of Agenda 21, how do you evaluate the following suggestion for a solution? Support your answer.



One possible way of protecting the tropical rain forest would be to turn the remaining forests into conservation areas or national parks. This would mean that nature would be left to itself again, untouched by mankind. Only a limited number of people would then be allowed to enter restricted areas in the parks along special trails and accompanied by a ranger.

Anhang C C-Test Deutsch

1. Kunstleben

Jahrzehntlang war das Kunstleben in Russland streng organisiert und strukturiert: Alle wussten genau, welche Kunst erlaubt und welche verboten war. Jetzt sind die Kulturinszenierungen zusammengebrochen und viele Künstler sind in den Westen gegangen. Daraus könnte man schließen, dass in diesem Bereich der Kunst ein Vakuum entstanden sei. Doch das Gegenteil ist der Fall. Bei einem kürzlichen Besuch in Moskau überprüfte ich von der unverminderten Vitalität der dortigen Kunstszene. Die sich in ihrer Struktur in der letzten Zeit stark verändert hat, steht auf einem anderen Blatt.

2. Krieg

Dass der Krieg der Vater aller Dinge sei, sagt man heute nicht mehr. Wir erfahren jedoch täglich, wie die historischen Katastrophen als Katalysator ungeahnter und ungewollter Veränderungen, im Großen wie im Kleinen, wirken. Diesen Weg muss man nicht nur als Unglück, sondern kann ihn auch als Chance, schließlich so sehr als Gewinn sehen. In diesem Sinne ist sich der Titel der Lebensreise Klaus von Bismarcks zu verstehen: "Aufbruch aus Pommern".

3. Lenin-Denkmal

Das Lenin-Denkmal am Wiener Platz in Dresden, dem der Innenstadt zugekehrten Vorplatz des Hauptbahnhofs, wird jetzt abgebaut. Die Stadtver_____ haben ein_____ Mühen hin_____ sich, z_____ einem Ents_____ darüber z_____ kommen, w_____ aus d_____ Denkmal d_____ Gründers d_____ Sowjetunion wer_____ solle. Zunä_____ war e_____ einfacher Ab_____ geplant, d_____ den nachrevol_____ Stimmungen i_____ der Bevöl_____ entsprochen hä_____, zumal d_____ zu ei_____ umgestalteten Fußgän_____ erweiterte eins_____ Prachtstraße i_____ Oktober 1989 e_____ Brennpunkt der Revolution in Dresden war.

4. Demokratie

Wie modern die direkte Demokratie ist, zeigt die Entwicklung in den neuen Demokratien Osteuropas. Es wä_____ verfehlt, vorse_____ preiszugeben, wo_____ Menschen ande_____ ringen. Unbes_____ Abstriche b_____ den Volksr_____ zu mac_____, bedeutet, d_____ Gelingen d_____ europäischen Integ_____ unseres Lan_____ aufs Sp_____ zu set_____. Man wi_____ wohl n_____ zu ei_____ begründeten Verz_____ bereit se_____. Perfekti_____ und korr_____ Verhalten a_____ der interna_____ Bühne wer_____ als Argu_____ nicht genügen.

Anhang D Holistische Beurteilungsskala (Auszug)

HOLISTIC SCALES DEGREE OF TASK FULFILMENT: APPROPRIATENESS AND CLARITY OF THE DISCOURSE FUNCTION

Task 1C

Level 6

There is a well-structured, succinct, correct and *comprehensive* descriptive explanation or descriptive conclusion that precedes or follows the correct identification of the climate as a *tropical* climate.

- The descriptive explanation or descriptive conclusion includes correct information about temperature and precipitation. In terms of temperature, a correct statement is made about the average temperature by including the correct numerical value (25°C) and it is pointed out that the temperature is constant throughout the year. In terms of precipitation, it is correctly and precisely indicated that the precipitation is constantly high: namely that it is hardly under 100mm per month. It is stated, however, that the monthly amount of precipitation varies over the year and that there are two peaks in the precipitation over the year. It is further explained that the precipitation is higher than the evaporation and that it is therefore humid. *Only relevant content* is included.
- The text is *consistently* logically structured and effectively linked.
- The text *consistently* follows the lexical-grammatical conventions of a descriptive explanation or descriptive conclusion and identification, e.g. it is written in the indefinite present tense and a subject-specific register and formal style is *consistently* used. The language use shows a *very high* level of accuracy.

Level 5

There is a *generally* well-structured and clearly formulated, correct and *adequate* descriptive explanation or descriptive conclusion that precedes or follows the correct identification of the climate as a *tropical* climate.

- The descriptive explanation or descriptive conclusion includes correct information about temperature and precipitation, but *one of the sub-points under precipitation* may be *lacking*, or may *not be expressed numerically*. It is explained that the precipitation is higher than the evaporation and that it is therefore humid. A *limited amount of content* may be included that is related to the topic, but *not central* to the content requirements of the task.
- The text is *for the most part* logically structured and effectively linked.
- The text *consistently* follows the lexical-grammatical conventions of a descriptive explanation or descriptive conclusion and identification, e.g. it is written in the indefinite present tense and a subject-specific register and formal style is used *most of the time*. The language use shows a *high* level of accuracy.

Level 4

There is a descriptive explanation or descriptive conclusion that is *sufficient*, but that may show a *limited number of errors* and that precedes or follows the correct identification of the climate as a *tropical* climate. The descriptive explanation or conclusion and identification *may be* generally well structured and clearly formulated.

- The description includes information about temperature and precipitation, but a *limited number of sub-points under precipitation* may be *lacking*, or may be *incorrect*, or may *not be expressed numerically*. It *may be* pointed out that it is humid (or that it is hot and wet, or hot and damp), but an *explanation* about the relation between the amount of precipitation and degree of evaporation *may be lacking*. *Some content* may be included that is related to the topic, but *not central* to the content requirements of the task.
- The text *may not be consistently* logically structured and sufficiently linked.
- The text follows the lexical-grammatical conventions of a descriptive explanation or descriptive conclusion and identification *for the most part*, e.g. it is written in the indefinite present tense. However, there may be *some instances of inappropriateness* in terms of register and style and *some grammar and vocabulary errors*.

Level 3

The type of climate *may be* identified as *tropical*, **or** may be classified as *humid*. This is preceded or followed by a descriptive explanation or descriptive conclusion that is *still acceptable for the particular classification*, but that may show a *limited number of errors*. The descriptive explanation or conclusion and identification *may be* generally well structured and clearly formulated.

- The description includes information about temperature and precipitation, but a *limited number of sub-points under temperature or precipitation* may be *lacking*, or may be *incorrect*. Exact references *expressed numerically* are *probably not included*. It *may be* pointed out that it is humid, but an *explanation* about the relation between the amount of precipitation and degree of evaporation may be *lacking*. *Some content* may be included that is related to the topic, but *not central* to the content requirements of the task.
- The text *may not be consistently* logically structured and sufficiently linked.
- The text follows the lexical-grammatical conventions of a descriptive explanation or descriptive conclusion and identification *for the most part*, e.g. it is written in the indefinite present tense. However, there may be *some instances of inappropriateness* in terms of register and style and *some* grammar and vocabulary *errors*.

Level 2

The type of climate is not identified as tropical, it is *classified as humid*, **or** *related paraphrases* are given. This is preceded or followed by a descriptive explanation or descriptive conclusion that *may be still acceptable*, but that may show *some errors*. The descriptive explanation or conclusion and identification may show *major deficiencies* in terms of logical structuring and formulation.

- The description includes *limited information* about temperature and precipitation. *Some sub-points under temperature or precipitation* may be *lacking*, or may be *incorrect*. Exact references *expressed numerically* are *probably not included*. It *may be pointed out* that it is humid, but an *explanation* about the relation between the amount of precipitation and degree of evaporation is *lacking*. *Some content* may be included that is related to the topic, but *not central* to the content requirements of the task.
- The text *may not be* logically structured and sufficiently linked.
- The text *may not consistently* follow the lexical-grammatical conventions of a descriptive explanation or descriptive conclusion and identification, e.g. it is written in the indefinite present tense. There may be a *number of instances of inappropriateness* in terms of register and style and *frequent* grammar and vocabulary *errors*.

Level 1

There is *no identification* of a type of climate, **or** the classification is *incorrect*. There is a *limited* description of the climate that may show *some errors* and that may show *major deficiencies* in terms of logical structuring and formulation.

- The description *may focus* either on temperature *or* on precipitation. *Some sub-points under temperature or precipitation* may be *lacking*, or may be *incorrect*. Exact references *expressed numerically* are *probably not included*. It *may be pointed out* that it is humid, but an *explanation* about the relation between the amount of precipitation and degree of evaporation is *lacking*. *Some content* may be included that is only *peripherally related* to the topic.
- The text *is not* logically structured and sufficiently linked.
- The text *may not consistently* follow the lexical-grammatical conventions of a descriptive explanation or descriptive conclusion and identification, e.g. it is written in the indefinite present tense. There may be a *number of instances of inappropriateness* in terms of register and style and *frequent* grammar and vocabulary *errors*.

Level 0

None of the required discourse functions is performed. The *type of climate is not identified* and there is *no description* of the climate. The text is *severely deficient* in terms of logical structuring and formulation.

- There is *no* or *hardly any reference* related to the *climate graph*.
- The text is *unstructured* and *fragmentary*.
- The text *does not follow* the lexical-grammatical conventions of a descriptive explanation or descriptive conclusion and identification at all. The register and style is *inappropriate* and there are *serious and frequent* grammar and vocabulary *errors*.

Anhang E Tabellen zum KFT

In diesem Anhang findet der interessierte Leser Kennwerte zur Verteilung der Daten bei den eingesetzten Subtests des KFT, sowohl für die gesamte Hauptstichprobe (Tabelle A1 und Tabelle A2), als auch getrennt für Schülerinnen und Schüler (Tabelle A3 und Tabelle A4). Neben deskriptiven Statistiken finden sich die Ergebnisse des zweiseitigen Kolmogorov-Smirnov-Tests (K-S-Test) auf Normalverteilung. Wenn auf Grund des Testergebnisses keine Normalverteilung angenommen werden kann, ist die gesamte betreffende Zeile in der jeweiligen Tabelle kursiv gesetzt. Statt der Bezeichnung *Exzess* findet sich bei manchen Autoren – besonders im englischsprachigen Raum – die Bezeichnung *Kurtosis*.

Tabelle A1

Die Tabelle zeigt deskriptive Statistiken für die drei eingesetzten Subtests des KFT. Zudem sind aus dem KFT-Manual die entsprechenden Ergebnisse für die Normstichprobe und die Gymnasialnorm, jeweils für Klasse 10, abgedruckt (Heller & Perleth, 2000, S. 79f).

Subtest	<i>N</i>	<i>M</i>	<i>SD</i>	Schiefe	Exzess	<i>P</i> ₂₅	<i>P</i> ₅₀	<i>P</i> ₇₅
Hauptstichprobe								
V3	165	10.07	2.98	0.10	-0.44	8	10	12
<i>Q2</i>	<i>165</i>	<i>16.08</i>	<i>3.29</i>	<i>-1.19</i>	<i>2.44</i>	<i>14</i>	<i>17</i>	<i>19</i>
<i>N2</i>	<i>165</i>	<i>19.24</i>	<i>3.90</i>	<i>-0.73</i>	<i>0.04</i>	<i>16</i>	<i>20</i>	<i>22</i>
<i>Subtests Gesamt</i>	<i>165</i>	<i>45.39</i>	<i>7.92</i>	<i>-0.59</i>	<i>-0.32</i>	<i>40</i>	<i>47</i>	<i>51</i>
Normstichprobe Testform A								
V3	144	10.24	3.15	0.04	-0.58	8	10	13
Q2	144	15.63	3.22	-0.78	-0.13	14	17	18
N2	144	18.13	4.31	-0.67	0.03	15	19	21
Normstichprobe Testform B								
V3	139	12.88	2.98	-0.16	-0.79	11	13	15
Q2	139	11.69	3.39	0.18	-0.63	9	12	14
N2	139	19.98	3.14	-0.74	0.08	18	20	22
Gymnasialnorm Testform A								
V3	114	11.55	2.64	-0.05	-0.21	10 ⁺	12	13
Q2	114	15.90	2.97	-0.49	-0.51	14	16	18
N2	114	19.14	3.85	-1.13	1.91	17	20	22
Gymnasialnorm Testform B								
V3	115	13.76	2.79	-0.38	-0.39	12	14	16
Q2	115	12.31	3.27	0.06	-0.70	10	12	15
N2	115	20.61	3.21	-1.15	1.23	19	21	23

⁺ Bemerkung: *P*₂₅ = 9.75

Tabelle A2

Diese Tabelle zeigt – in Ergänzung zu Tabelle A1 – einige zusätzliche deskriptive Statistiken. Dazu zählen Minimum (Min.), Maximum (Max.), sowie der Standardfehler (*SE*) von Schiefe und Exzess. Zudem ist jeweils das Ergebnis des Kolmogorov-Smirnov-Tests (K-S-Test) angegeben.

Subtest	Min.	Max.	<i>SE</i> _{Schiefe}	<i>SE</i> _{Exzess}	K-S-Test <i>Z</i>	K-S-Test <i>p</i>
V3	4	18	0.19	0.38	1.07	.20
Q2	0	20	0.19	0.38	1.93	.00
N2	6	25	0.19	0.38	1.74	.01
Subtests Gesamt	21	59	0.19	0.38	1.31	.07

Tabelle A3

Die Tabelle zeigt deskriptive Statistiken für die drei eingesetzten Subtests des KFT, jeweils getrennt für Schülerinnen und Schüler. Zudem sind aus dem KFT-Manual die entsprechenden Ergebnisse für die Normstichprobe, jeweils für Klasse 10, abgedruckt (Heller & Perleth, 2000, S. 88).

Subtest	<i>N</i>	<i>M</i>	<i>SD</i>	Schiefe	Exzess	<i>P</i> ₂₅	<i>P</i> ₅₀	<i>P</i> ₇₅
Hauptstichprobe (Schülerinnen)								
V3	101	9.63	2.76	-0.01	-0.45	7	10	12
Q2	101	15.44	3.17	-0.44	-0.76	13	16	18
N2	101	18.28	4.08	-0.46	-0.29	15	19	22
Subtests Gesamt	101	43.35	7.93	-0.37	-0.48	38	44	50
Hauptstichprobe (Schüler)								
V3	64	10.75	3.21	0.06	-0.63	8	11	13
Q2	64	17.11	3.23	-2.72	11.55	16	18	19
N2	64	20.75	3.05	-1.09	1.20	19	21	23
Subtests Gesamt	64	48.61	6.80	-1.08	0.97	46	50	53
Normstichprobe Testform A (Schülerinnen)								
V3		9.59	3.27					
Q2		15.36	3.45					
N2		19.18	3.06					
Normstichprobe Testform A (Schüler)								
V3		10.58	3.03					
Q2		15.77	3.09					
N2		17.60	4.74					
Normstichprobe Testform B (Schülerinnen)								
V3		12.21	3.04					
Q2		11.82	3.61					
N2		20.27	2.95					
Normstichprobe Testform B (Schüler)								
V3		13.34	2.81					
Q2		12.09	2.98					
N2		19.67	3.39					

Tabelle A4

Diese Tabelle zeigt – in Ergänzung zu Tabelle A3 – einige zusätzliche deskriptive Statistiken. Dazu zählen Minimum (Min.), Maximum (Max.), sowie der Standardfehler (SE) von Schiefe und Exzess. Zudem ist jeweils das Ergebnis des Kolmogorov-Smirnov-Tests (K-S-Test) angegeben.

Subtest	Min.	Max.	$SE_{Schiefe}$	SE_{Exzess}	K-S-Test Z	K-S-Test p
Hauptstichprobe (Schülerinnen)						
V3	4	17	.24	.48	1.18	.13
Q2	8	20	.24	.48	1.38	.04
N2	6	25	.24	.48	1.07	.20
Subtests Gesamt	21	58	.24	.48	0.78	.58
Hauptstichprobe (Schüler)						
V3	4	18	.30	.59	0.71	.70
Q2	0	20	.30	.59	1.64	.01
N2	11	25	.30	.59	1.51	.02
Subtests Gesamt	28	59	.30	.59	1.15	.14