

9 Erhebungsinstrumente und Methode

Das folgende Kapitel widmet sich der Auswahl der Erhebungsinstrumente, die in der vorliegenden Untersuchung zum Einsatz kamen, sowie dem Ablauf der Datenerhebung. Die Dokumentation umfasst auch die Erhebungsinstrumente für Variablen, die nicht im Zentrum der Fragestellung stehen, aber als Zusatzinformation über die Kinder für die Beschreibung der Probanden oder als potenzielle Kontrollvariable miterhoben wurden. Die Erhebungs- und Auswertungsverfahren zeigen, dass für die vorliegende Studie methodisch eine Kombination quantitativer und qualitativer Zugänge angezeigt ist. Gewisse Testverfahren sind ausschliesslich für quantitative Auswertungen geeignet, während sich andere Fragen aber nur aufgrund qualitativer Analysen überhaupt sinnvoll diskutieren lassen. Letzteres betrifft die Auswertung der transkribierten Gesprächsdaten, die bewusst auf einem induktiven Analyseverfahren beruht.

9.1 Dokumentation der Erhebungsinstrumente

Die Erhebungsinstrumente, die ausgewählt wurden, um die in Kapitel 8 beschriebenen Variablen zu erheben, werden im Folgenden vorgestellt. Die Wahl der Instrumente wird mit Verweis auf einschlägige Studien begründet. Zu einzelnen Verfahren werden die Gütekriterien aufgrund der erhobenen Daten berechnet. Die verwendeten Erhebungsinstrumente sind im Anhang der Arbeit eingefügt, sofern es sich nicht um publizierte Testverfahren handelt.

9.1.1 Erhebung allgemeiner Daten sowie des sozioökonomischen Status

Personenbezogene Merkmale wie das Alter, das Geschlecht, die Muttersprache und eine allfällige Mehrsprachigkeit, das Herkunftsland, erhaltene Fördermassnahmen wie Logopädie sowie der sozioökonomische Status der Familie wurden durch einen Eltern- und einen Schülerfragebogen erfragt. Die Konzeption des Fragebogens orientiert sich an ähnlichen Fragebögen aus anderen Studien und an entsprechenden methodischen Empfehlungen zur Konzeption und Durchführung von Forschungsprojekten (Carigiet Reinhard 2012; Bortz 1993; Bühner 2006; Lienert, Raatz & Lienert-Raatz 1994; Raab-Steiner & Benesch 2010; Mummendey 2003). Einige der von den Kindern verlangten Angaben wurden der Vollständigkeit halber erhoben, gehen in die Analysen der Daten aber nicht mit ein, weil sie in Bezug auf die Untersuchungsfragen keine Rele-

vanz haben oder durch eine andere Variable bereits miterfasst werden (das ungefähre Alter etwa geht in die Variable *Klassenstufe* ein). Andere Auskünfte (etwa allfällige sprachliche Therapiemassnahmen durch einen Logopäden) könnten aufschlussreich sein für Auffälligkeiten, die sich in der Datenaufbereitung und -analyse ergeben, und wurden deshalb miterfasst. Ein dritter Bereich von Angaben wurde als Massnahme zur Erhöhung der internen Validität der Untersuchung erhoben. Das gilt für die vorliegende Studie für den sozioökonomischen Status und den Bildungshintergrund der teilnehmenden Kinder sowie für die Aufmerksamkeits- und Konzentrationsleistung.

Seit den PISA-Studien kommt dem Einfluss der Schichtzugehörigkeit auf Sprachleistungen von Kindern neue Aktualität zu (Baumert 2002; Carigiet Reinhard 2012; Steinig et al. 2009). Der sozioökonomische Status wird in Studien zur kindlichen Entwicklung systematisch miterhoben (Desgrippes & Lambelet 2016: 83). Es ist daher plausibel davon auszugehen, dass die Sprachleistungen der Kinder, wie sie in der vorliegenden Studie erhoben werden, auch von dieser Einflussgrösse abhängen. Bortz & Döring (2006) empfehlen die Erhebung von potentiellen Kontroll- und Störvariablen und die Bereinigung der abhängigen Variablen um dieselben. Da sowohl die unabhängige Variable (C-Test) als auch die abhängigen Variablen der Hauptfragestellungen Sprachleistungen und keine stabilen Personenmerkmale wie das Geschlecht oder das Alter darstellen, erscheint es nicht angezeigt, den sozioökonomischen Status als Störvariable oder als Kontrollvariable aus den abhängigen Variablen herauszupartialisieren. Es ist vielmehr davon auszugehen, dass der sozioökonomische Status und der Bildungshintergrund auf die abhängigen Variablen *und* die unabhängige Variable einwirken und so unberücksichtigt bleiben können. Die Auswertungen zum sozioökonomischen Status gehen hingegen in die Beschreibung der Stichprobe ein (Kap. 10.2.3). Als Mass für den sozioökonomischen Status wurde für die vorliegende Untersuchung in Anlehnung an andere Studien der *International Socio-Economic Index of Occupational Status* (ISEI) respektive der Wert des als höher bewerteten Elternteils (HISEI) verwendet (vgl. etwa für PISA 2000 Baumert 2002; Carigiet Reinhard 2012; Winkes 2014). In der Bildungsforschung ist es üblich, die Einkommensverhältnisse einer Familie sowie andere Aspekte ihres sozioökonomischen Status in erster Linie auf der Basis von Angaben zur Berufsausübung der Eltern zu bestimmen (Baumert 2002: 36). Der linear-skalierte sozioökonomische Index basiert auf dem Einkommen, der erworbenen Bildung und dem Beruf (Carigiet Reinhard 2012: 210). Der von Ganzeboom, de Graaf & Treiman (1992) entwickelte sozioökonomische Index hat im Gegensatz zu anderen den Vorteil, die ökonomische Stellung vom Berufsprestige zu trennen. Er wird anhand von Informationen zur Bildung und zum Einkommen der Berufstätigen bestimmt und kann Werte zwischen 16 (forstwirtschaftliche Hilfs-

Tab. 9.1: Indikatoren des sozioökonomischen Status.

Erfasster Indikator	Quelle*	Frage- nummer
höchster Schulabschluss der Mutter	Elternfragebogen	5
höchster Schulabschluss des Vaters	Elternfragebogen	5
Erwerbstätigkeitsstatus der Mutter	Elternfragebogen	6
Erwerbstätigkeitsstatus des Vaters	Elternfragebogen	6
Beruf der Mutter	Elternfragebogen	7
Beruf des Vaters	Elternfragebogen	7
berufliche Funktion der Mutter	Elternfragebogen	8
berufliche Funktion des Vaters	Elternfragebogen	8
Weisungsbefugnis der Mutter	Elternfragebogen	9
Weisungsbefugnis des Vaters	Elternfragebogen	9
Bücherbesitz der Familie	Schülerfragebogen	17

* Die Eltern- und Schülerfragebögen sind im Anhang abgedruckt.

kraft) und 90 (Richter) annehmen. Ganzeboom, de Graaf & Treiman (1992) gehen davon aus, dass Schul- und Berufsbildung über die ausgeübten Berufe in Einkommen und in Chancen zur Teilhabe an Macht umgesetzt werden. Da Berufe also bestimmte Qualifikationen voraussetzen und zu bestimmten Einkommen führen, werden sie als Vermittler zwischen Bildungsabschlüssen und Einkommenslagen betrachtet (Baumert 2002: 37). Im Eltern- und Schülerfragebogen wurden die in Tabelle 9.1 aufgeführten Indikatoren erhoben. In PISA 2000 stellte sich der Bücherbesitz der Familie als aussagekräftiger Indikator für die Bildungsnähe und die Voraussage von Leseleistungen heraus (Carigiet Reinhard 2012: 209). Aus diesem Grund wurde die Frage in den Schülerfragebogen aufgenommen. Die Angaben der Eltern zu ihren momentanen oder zu ihren zuletzt ausgeübten Berufen wurden anhand der *Internationalen Standardklassifikation der Berufe* ISCO-88 (COM) unter Berücksichtigung des höchsten Bildungsabschlusses, des Erwerbstätigkeitsstatus, der beruflichen Funktion und der Weisungsbefugnis kategorisiert und den dreistelligen Codes der Berufsuntergruppen zugeordnet. Davon ausgehend wurde mithilfe eines Umsteigeschlüssels⁴⁸ der sozioökonomische Status (ISEI) der Mutter und des Vaters ermittelt und schliesslich in den *HISEI* recodiert, indem der jeweils höhere

48 Die Umrechnung der ISCO 88-Codes in ISEI-Indexpunkte wurde anhand des Umsteigeschlüssels der Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen (GESIS) vorgenommen. Verfügbar unter: <http://www.gesis.org/missy/studie/klassifikationen/sozialwissenschaftliche-klassifikationen/isei/> (letzter Zugriff: 1. 5. 2016).

ISEI als Indikator für die sozioökonomische Stellung der Familie ausgewählt wurde.⁴⁹

9.1.2 Erhebung der globalen Sprachkompetenz (C-Test)

Im Rahmen der vorliegenden Studie wurde die globale Sprachkompetenz der Kinder erhoben, um zu prüfen, ob sie für Schreibprozess- und Textvariablen sowie Aspekte der metasprachlichen Kompetenz Vorhersagekraft besitzt. Dafür wurde ein möglichst objektives Verfahren ausgesucht, auf dessen Grundlage die Bildung von Kompetenzgruppen unter den teilnehmenden Kindern möglich ist. Es existiert insbesondere für frühe Altersstufen eine Vielzahl an standardisierten Sprachtests, die unterschiedliche Teilkompetenzen messen und abbilden. Im Zentrum standen in der vorliegenden Studie aber eine sogenannte globale Sprachkompetenz und genau nicht spezifische Teilkompetenzen, da sich erfolgreiches Textschreiben gerade durch deren Integration auszeichnet. Für die Erhebung der globalen Sprachkompetenz haben sich C-Tests als hoch reliables Messinstrument etabliert,

weil im Zuge der Bearbeitung eines derartigen omnibus-Tests auf die tiefer liegende, abstrakte Sprachkompetenz zugegriffen werden muss, die ihrerseits auf einem komplexen Zusammenspiel einer ganzen Reihe inhaltlicher und sprachlicher Komponenten beruht, wobei linguistische Elemente auf verschiedenen Ebenen der Sprachstruktur integriert werden müssen. (Vockrodt-Scholz & Zydati 2010: 38)

Neben inhaltlichen Argumenten bietet sich das Verfahren auch aus durchführungstechnischen Gründen an: Der Test ist zeitökonomisch und ohne erheblichen Materialaufwand durchführbar.

9.1.2.1 C-Tests als Instrument der Sprachdiagnose

C-Tests wurden ursprünglich insbesondere als Instrument für die Sprachstandsbestimmung erwachsener, fremdsprachiger Lernender entwickelt und

⁴⁹ Die verschiedenen Indikatoren für den sozioökonomischen Status der Familie wurden erhoben, um eine möglichst zweifelsfreie Kategorisierung zu den Berufscodes zu gewährleisten. In der Datenaufbereitung waren die verschiedenen Angaben in der Tat von Vorteil, wenn etwa die Berufsbezeichnungen nicht aussagekräftig oder nicht zu interpretieren waren. In solchen Fällen half z. B. die Angabe zum höchsten Schulabschluss für die Zuordnung. In anderen Fällen führte das Vorliegen der unterschiedlichen Angaben zu einer Konfundierung der Variablen: Teilweise stimmten der aktuelle Beruf des Elternteils mit der absolvierten Ausbildung nicht überein. In solchen Fällen wurde der ausgeübte Beruf als ausschlaggebender Indikator gewählt, da Baumert (2002: 37) Berufe als Vermittler zwischen Bildungsabschlüssen und Einkommenslagen betrachtet.

Vom 16. bis zum 19. Jh. war Großbritannien die weltweit führende See- und Kolonialmacht. Daher ko _____ es, da _____ Englisch schli _____ zur Welts _____ wurde. D _____ Industrielle Revol _____, die z _____ entscheidenden sozi _____ und polit _____ Umwälzungen füh _____, begann u _____ 1760 i _____ Großbritannien u _____ verschaffte d _____ Lande au _____ eine wirtsch _____ Führungsposition. He _____ haben vi _____ alte Indus _____ an Bede _____ verloren, denn in der modernen Gesellschaft stehen Dienstleistungsbranchen wie Banken und Versicherungen im Vordergrund. Das in der Nordsee geförderte Öl sichert die Energieversorgung des Landes.

Abb. 9.1: Ausschnitt aus einem deutschsprachigen C-Test⁵⁰. Quelle: www.C-Test.de (letzter Zugriff: 1. 7. 2016).

eingesetzt (vgl. insbesondere Grotjahn 1992, 1994, 1996, 2002a, 2006, 2010, 2014), werden heute aber ebenfalls für die Bestimmung des Sprachstandes bei Muttersprachlern verwendet (Baur & Spettmann 2009: 115, vgl. zum Einsatz bei Muttersprachlern auch Wockenfuß & Raatz 2014). Das Testverfahren setzt die Festigung grundlegender Lese- und Schreibkompetenzen voraus. Baur & Spettmann (2009: 116) empfehlen den Einsatz von C-Tests ab der vierten Jahrgangsstufe. Der Test besteht in der Regel aus vier Texten, die stufengerecht ausgewählt – die Textinhalte sollen dem Wissensstand und dem Weltwissen der Probanden entsprechen – und anhand einer bestimmten Systematik zu Lückentexten umgewandelt werden. Unter dem klassischen Tilgungsprinzip wird in der Literatur in der Regel die Tilgung der hinteren Hälfte jedes zweiten Wortes verstanden, es finden sich aber auch andere Tilgungsprinzipien; Baur & Spettmann (2009: 117) schlagen für die Klassen vier bis sieben die Tilgung der hinteren Hälfte von jedem dritten Wort vor.⁵⁰ Das Prinzip der ‚Beschädigung‘ authentischer Texte bleibt bei jeder Art der Wort(teil)tilgung bestehen und zeichnet das Testverfahren von C-Tests im Kern aus. Abbildung 9.1 bildet ein von den Entwicklern zur Verfügung gestelltes Anschauungsbeispiel eines C-Tests ab, da der CT-D 4 (Raatz & Klein-Braley 1992), der in der vorliegenden Untersuchung eingesetzte C-Test, hier nicht abgedruckt werden darf.

⁵⁰ Je nach Art der Tilgung kann laut Baur & Spettmann (2009: 117) zwischen der Kompetenzbestimmung in verschiedenen sprachlichen Teilbereichen unterschieden werden; die Stammtilgung etwa ermöglicht laut den Autoren die klare Unterscheidung zwischen Lexik und Morphologie.

⁵¹ Der abgedruckte Textausschnitt ist ein Beispiel für einen C-Test. Der Ausschnitt wurde von den Autoren der Website c-test.de zur Verfügung gestellt. Unglücklich am abgedruckten Beispiel ist die Trennung des bestehenden Wortteils und der dazugehörigen leeren Linie, die für den fehlenden Teil vorgesehen ist, auf verschiedene Zeilen. Das sollte in Tests vermieden wer-

9.1.2.2 Bottom-up- und top-down-Prozesse der Sprachverarbeitung in C-Tests

Wenn im Folgenden der Kürze halber von einer allgemeinen oder globalen Sprachkompetenz gesprochen wird, dann ist zu präzisieren, dass damit primär eine allgemeine schriftliche oder schriftbasierte Kompetenz gemeint ist. Die mündliche Sprachverständigung ist integraler Bestandteil einer globalen Sprachkompetenz, wird mit C-Tests aber nicht getestet.

Zwischen der tiefer liegenden, latenten Sprachkompetenz (die der C-Test als eine Form des indirekten Testens messen will) und dem kommunikativen Umgang mit Sprache können zum Teil beträchtliche Unterschiede liegen, weil ein ‚guter Sprecher‘ kein ‚guter Grammatiker‘ und ein ‚guter Hörer‘ kein ‚guter Schreiber‘ sein muss. (Vockrodt-Scholz & Zydatiŕ 2010: 5)

Der Vorteil von C-Tests im Gegensatz zu anderen Instrumenten der Sprachstandsbestimmung besteht darin, dass sie eine global-ganzheitliche, also eine integrative Sprachstandsmessung erlauben und damit eine punktuelle Momentaufnahme der allgemeinen Sprachkompetenz abgeben (Vockrodt-Scholz & Zydatiŕ 2010: 4). Leistungen im C-Test weisen erstaunlich hohe Korrelationen mit verschiedenen Aussenkriterien wie Schulnoten, Lehrerurteilen über den Sprachstand der Kinder oder Ergebnissen aus anderen Sprachtests auf (Grotjahn 2002b: 215). In der Praxis erprobte C-Tests weisen eine hohe Reliabilität auf. Dissens besteht in der Forschung jedoch darüber, was die Konstruktvalidität des Tests ausmacht, also darüber, was der Test eigentlich genau misst. Diskutiert wird insbesondere die Frage, welche mental-kognitiven Prozesse beim Bearbeiten des Tests aktiviert werden; ob es insbesondere mikrostrukturelle Verarbeitungsprozesse sind, auf die sich die Probanden konzentrieren oder ob makrostrukturelle Verarbeitungsprozesse ebenfalls eingesetzt werden. Vertreten werden hier zwei gegenteilige Positionen: Die Analyse introspektiver Aussagen von Probanden weisen darauf hin, dass *bottom up*-Prozesse die übergeordnete Rolle spielen, wohingegen Fehleranalysen Verarbeitungszusammenhänge zwischen sprachlichen Elementen nachweisen, die im Text weite Abstände haben (Vockrodt-Scholz & Zydatiŕ 2010: 5–6). Sowohl Klein-Braley (1996) als auch Hastings (2002) betonen beide die Relevanz makrostruktureller Verarbeitungsprozesse: Sie zeigen, dass mikrostrukturelle Verarbeitungsprozesse nur bei Probanden mit schlechten Testergebnissen dominieren, wohingegen fortgeschrittene Lerner viel stärker den Makrokontext berücksichtigen. Die erfolgreiche Bearbeitung von C-Tests setzt die wechselseitige Synthese von *bot-*

den, da davon auszugehen ist, dass die Aufgabe für gewisse Probanden dadurch zusätzlich erschwert wird.

tom up- und *top down*-Prozessen der Sprachverarbeitung voraus. Es müssen orthographische, lexikalische, morphologische, syntaktische, semantische und kontextuelle Informationen integriert werden. Vockrodt-Scholz & Zydati (2010: 6) vermuten daher, dass die ablaufenden Prozesse mit den Vorgngen vergleichbar sind, die fr komplexe Sprachleistungen im Kontext natrlicher Sprachverwendung bentigt werden, und warnen davor, das Testkonstrukt des C-Test zu sehr in die Nhe der rezeptiven Leseverstehenskompetenz zu rcken, zu deren Messung er in seiner ursprnglichen Form entwickelt worden war.

9.1.2.3 Zur Rolle sprachproduktiver Fhigkeiten fr die Bearbeitung von C-Tests

Ein strittiger Aspekt in der Diskussion ber C-Tests ist die Frage, inwiefern rezeptive und/oder produktive Sprachkompetenzen getestet werden. Vockrodt-Scholz & Zydati (2010: 6) untersuchten in ihrer Studie, ob eine Zuordnung von untergeordneten Teilkompetenzen zu einem bergeordneten globalen Kompetenzniveau mglich ist und gehen dabei von der Annahme aus, dass sprachproduktive Faktoren fr die Testbearbeitung einen sehr viel hheren Stellenwert haben, als in der einschlgigen Literatur eingerumt wird. Sie weisen tatschlich erstaunlich deutliche Zusammenhnge zwischen einzelnen Teilkompetenzen und dem ‚Gesamt‘ der fremdsprachlichen Fhigkeiten der Probanden nach, was die Robustheit des C-Tests als Instrument zur Erfassung der tiefer liegenden (latenten) Sprachkompetenz von Fremdsprachlernern unterstreicht (Vockrodt-Scholz & Zydati 2010: 36). Die Leistungen in den zwei von den Autoren eingesetzten Schreibaufgaben – dabei handelt es sich zum einen um eine schriftliche Zusammenfassung eines gesprochenen Monologs und um das Verfassen einer Bildergeschichte – weisen statistisch einen hohen Zusammenhang mit den Leistungen im C-Test auf. Als besonders aufschlussreich stellte sich fr die von Vockrodt-Scholz & Zydati untersuchten Probanden der Fehlerquotient als qualitatives Merkmal textsortengebundenen Schreibens heraus. Er erreichte vergleichsweise hohe Signifikanz- und Zuordnungswerte. Die formalsprachliche Richtigkeit (*accuracy*) scheint bei fortgeschrittenen fremdsprachlichen Lernenden ein zuverlssiger Indikator fr deren Niveau zu sein (Vockrodt-Scholz & Zydati 2010: 37).

Dem Zusammenhang zwischen den C-Test-Leistungen und verschiedenen textqualitativen Merkmalen geht die vorliegende Studie nach, indem die Kindertexte anhand orthographischer, grammatischer, lexikalischer und syntaktischer Variablen ausgewertet wurden. Was hingegen unbercksichtigt bleibt, ist die inhaltliche Qualitt und Struktur der Texte. Das *Raten* der inhaltlichen Qualitt von Texten ist immer mit subjektiven Normvorstellungen verbunden (Egli Cuenat 2016: 63). Die Analyse der Kindertexte beruht im vorliegenden

Kontext auf Variablen, die kein holistisches *Raten* der Texte notwendig machen.

9.1.2.4 Der CT-D 4 und seine Testgütekriterien

Reliabilität

Die Analyse der Testgütekriterien erfolgt üblicherweise anhand einer Reliabilitätsprüfung der inneren Konsistenz des Tests, der Berechnung der Trennschärfekoeffizienten und der Schwierigkeit der einzelnen Items. Es werden im Folgenden sowohl die von den Testentwicklern angegebenen Gütekriterien als auch die aufgrund der selber erhobenen Daten berechneten Werte angegeben.

C-Tests bestehen aus Texten und beschränken sich explizit nicht auf die Arbeit an einzelnen Sätzen. Das heisst für die Auswertung von C-Tests gleichzeitig, dass die einzelne Lücke nicht als einzelnes Item im Sinne der klassischen Testtheorie betrachtet werden kann, deren je 25 Lücken voneinander stochastisch nicht unabhängig sind. Die kleinsten Einheiten eines C-Tests sind darum die einzelnen Texte, die Raatz & Klein-Braley (1992) als sogenannte ‚Superitems‘ verstehen (vgl. dazu auch Eckes 2011: 416). In die Berechnungen geht der einzelne Text als einzelnes Superitem ein, wodurch der Gesamttest vier Superitems aufweist. Im Sinne der klassischen Testtheorie ist die Schwierigkeit eines solchen Superitems die auf den Maximalwert bezogene durchschnittliche Punktzahl.

Der CT-D 4 von Raatz & Klein-Braley (1992) liegt in zwei Parallelversionen vor. Die Reliabilität liegt laut den Autoren bei deutschen Schülern und Schülerinnen für die Version A nach Cronbachs Alpha bei $\alpha = .88$ ($N = 731$) und für die Version B bei $\alpha = .86$ ($N = 698$). Insgesamt kann die Reliabilität damit als zufriedenstellend gelten. Als Retest-Zuverlässigkeit geben Raatz & Klein-Braley (1992) für dieselbe Testform im Mittel $r_{tt} = .69$, für verschiedene Testformen $r_{tt} = .73$ an.

Die Reliabilitätsprüfung (innere Konsistenz) der einzelnen Superitems ergibt für die hier erhobene 4. Klasse ein Cronbachs Alpha von $\alpha = .891$ (Testversion A) respektive $\alpha = .887$ (Testversion B) und für die 6. Klasse ein Cronbachs Alpha von $\alpha = .894$ (Testversion A) respektive $\alpha = .877$ (Testversion B). Nach Fisseni (1997: 124) sind Reliabilitäten mit Alphas $< .80$ als niedrig, Alphas zwischen $.80$ und $.90$ als mittel und Reliabilitäten $> .90$ als hoch zu bewerten. Die Alphas für die vier Teilgruppen liegen also im mittleren Bereich. Für alle 115 Kinder zusammen erreicht der Test ein Cronbachs Alpha von $\alpha = .915$, das damit im hohen Bereich liegt. Die Alphas verbessern sich durch Weglassen eines Superitems nicht, was dafür spricht, die Summe aller vier Texte als Ankervariable für die Studie zu wählen und keinen der Texte auszuschliessen.

Tab. 9.2: Trennschärfen und Schwierigkeit des CT-D 4.

	Klassenstufe 4		Klassenstufe 6	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
CT-D 4 Text 1	.76	58.0	.73	78.3
CT-D 4 Text 2	.77	53.2	.78	71.2
CT-D 4 Text 3	.78	53.9	.77	67.3
CT-D 4 Text 4	.76	47.4	.71	68.6
<i>Gesamt</i>	.77	53.1	.75	71.4

Anm.: *r* = Trennschärfekoeffizient, *p* = Schwierigkeitsindex

Da der Test für die 4. Klassenstufe entwickelt, aber auch für die zwei Jahre älteren Kinder eingesetzt wurde, gilt besonderes Augenmerk der Frage, ob der Test für die 6. Klasse zu leicht war. Die Graphiken zu den Lagemassen (Abb. 9.2, Tab. 9.5: Die Kinder der 6. Klassenstufe erzielten insgesamt bessere Ergebnisse (zwischen 41 und 96 bei maximalen 100 Punkten, $SD = 13.8$) als die jüngeren Kinder (zwischen 24 und 85 bei maximalen 100 Punkten, $SD = 16.5$), es gibt aber keinen Deckeneffekt, im Gegenteil: Es erreichte keines der älteren Kinder den Maximalwert. Der Test scheint also für die älteren Kinder nicht zu leicht zu sein. Präziser lässt sich die Aufgabenschwierigkeit (*p*, in Tab. 9.2 zu lesen als Prozente) statistisch ermitteln: Sie entspricht dem prozentualen Anteil der Probanden, die eine Aufgabe korrekt gelöst haben (Lienert, Raatz & Lienert-Raatz 1994: 73). Der Schwierigkeitsgrad sollte zwischen 25 % und 75 % richtiger Lösungen liegen. Da die einzelnen Lücken im C-Test stochastisch nicht unabhängig sind (ein Text mit je 25 Lücken stellt eine Aufgabe dar), errechnet sich der Schwierigkeitsgrad aus dem erreichten Punktmittelwert relativ zur möglichen Lückenzahl (Vockrodt-Scholz & Zydati 2010: 9). Die Aufgabenschwierigkeit liegt bei nur einem Superitem für die 6. Klassenstufe bei knapp über 75 % richtiger Lösungen (Text 1 für die 6. Klassenstufe). Die Schwierigkeit des Tests ist also für die älteren Kinder keinesfalls zu gering; im Gegenteil erweist sich der CT-D 4 für die 4. Klassenstufe als recht anspruchsvoll. Wäre davon auszugehen, dass viele Kinder unter Zeitdruck einen Text gar nicht bearbeiten konnten, wäre eine Schwierigkeitsanalyse unter Inangriffnahmekorrektur angezeigt. Da von allen Probanden nur ein einziges Kind der 4. Klasse in einem Text 0 Punkte erzielte, ist diese Korrektur nicht nötig. Das entsprechende Mädchen hat sich am Ausfüllen der Lücken auch in den Texten 3 und 4 versucht, wie das Testheft zeigt; die fehlenden Punkte sind also nicht darauf zurückzuführen, dass für die Bearbeitung gar keine Zeit geblieben ist.

Die Trennschärfe respektive der Trennschärfekoeffizient eines Items (r) entspricht der korrigierten Korrelation einer Aufgabe mit einer Skala. Die Trennschärfe drückt also aus, wie gut ein Item eine Skala widerspiegelt. Nur Items mit hoher Trennschärfe unterscheiden zwischen ‚guten‘ und ‚schlechten‘ Probanden (Brosius 2013: 828–829; Carigiet Reinhard 2012: 205). Fisseni (1997: 124) bezeichnet Trennschärfen von $< .30$ als niedrig, Werte zwischen $.30$ und $.50$ als mittel und Werte $> .50$ als hoch. Die Trennschärfekoeffizienten liegen also alle im sehr hohen Bereich, wie Tabelle 9.2 zu entnehmen ist.

Validität

Die Validität des Tests bestimmen Raatz & Klein-Braley (1992) anhand des Vergleichs mit den Zeugnisnoten im Fach Deutsch. Insgesamt gesehen ergibt sich eine zwischen Deutschnote und Testergebnis eine mittlere Korrelation von $r = .72$ (zwischen $r = .20$ und $r = .92$) (Raatz & Klein-Braley 1992: 13–14). Obwohl der Zusammenhang mit der Deutschnote damit hoch ausfällt, ist dieses Vorgehen zur Prüfung der Validität unbefriedigend. Aus diesem Grund wird die Validität für die vorliegende Studie aufgrund der Sprachstandsmessung KEKS überprüft.

Wie aus Tabelle 9.3 hervorgeht, korrelieren die globale Sprachkompetenz und die Kompetenz in den einzelnen Teilbereichen sehr signifikant. Mit Werten von $r = .777$ (Grammatik), $r = .663$ (Wortschatz) und $r = .712$ (Rechtschreibung) liegen die Korrelationen nach Pearson im hohen Bereich.⁵²

Insgesamt bestätigen die Ergebnisse damit die Annahme, dass ein bedeutender Zusammenhang besteht zwischen Leistungen im C-Test und Leistungen in den Teilbereichen Grammatik, Lexik und Orthographie. Die Korrelationskoeffizienten liegen zudem alle recht nahe beieinander, es fällt also keiner der drei sprachlichen Teilbereiche deutlich ab, wenn deren Zusammenhang mit der globalen Sprachkompetenz im Zentrum steht. Mithilfe einer multiplen Regressionsanalyse wurde weiter überprüft, welche der Variablen (z -transformierte Werte des Untertests *Grammatik*, des Untertests *Wortschatz* und des Untertests *Rechtschreibung*) signifikanten Aufklärungswert für die Leistungen im C-Test (z -transformiert), also für die globale schriftliche Sprachkompetenz, haben. Damit sollte gleichzeitig geklärt werden, wie sich die drei Faktoren (Untertest Grammatik, Untertest Wortschatz und Untertest Rechtschreibung) zuei-

⁵² Nach Cohen (1992: 157) gelten Werte über $r = .30$ als mittlere und Werte über $r = .50$ als grosse Effektstärke. Werden die Korrelationen getrennt nach Altersgruppen berechnet, ergeben sich für die 4. Klasse bei $N = 44$ Werte von $r = .776$ (Grammatik), von $r = .549$ (Wortschatz) und von $r = .636$ (Rechtschreibung) und für die 6. Klasse bei $N = 71$ Werte von $r = .724$ (Grammatik), von $r = .560$ (Wortschatz) und von $r = .628$ (Rechtschreibung).

Tab. 9.3: Korrelationen der globalen Sprachkompetenz und der Kompetenz in einzelnen sprachlichen Teilbereichen und Interkorrelationen der Merkmale.

	1	2	3	4
1 C-Test ^b	1.00			
2 Untertest Grammatik	.78**	1.00		
3 Untertest Wortschatz	.66**	.63**	1.00	
4 Untertest Rechtschreibung ^a	.71**	.65**	.52**	1.00

* $p < .05$ (2-seitig), ** $p < .01$ (2-seitig)

Anm.: $N = 115$, a. Total der Untertests zur Rechtschreibung (bestehend aus den Subtests zur Schreibung deutscher Wörter und zu Pseudowortschreibungen); b. Version B gemäss Auswertungshinweisen um 4 Punkte korrigiert

inander verhalten, wenn sie alle drei als erklärende Variablen für die C-Test-Leistungen angenommen werden. Die Bedingungen für eine lineare Regression sind erfüllt: Es besteht ein linearer Zusammenhang zwischen der abhängigen und den unabhängigen Variablen. Es liegt zudem Homoskedastizität (eine Gleichverteilung der Residuen) vor, es besteht keine Autokorrelation der Residuen (der Durbin-Watson-Wert liegt bei 2.091⁵³) und keine Multikollinearität (die Toleranzwerte liegen zwischen .463 und .587 und die Werte des Varianzinflationsfaktors⁵⁴ zwischen 1.703 und 2.162).

Durch ein schrittweises Vorgehen (Methode *schrittweise* in SPSS) wurden die unabhängigen Variablen nacheinander in das Modell aufgenommen. Prädiktoren werden bei diesem Vorgehen einzeln integriert, aber auch ausgeschlossen, wenn ein Prädiktor nach der Integration eines anderen Prädiktors nicht mehr signifikant ist. Dadurch zeigt sich, welches Modell (ob mit einer, zwei oder allen drei unabhängigen Variablen) am meisten Varianz aufklärt. Die Anpassungsgüte des Modells wird durch das Bestimmtheitsmass R^2 angegeben, das durch die Multiplikation mit 100 ausweist, wie viel Prozent der Varianz der abhängigen Variable durch das entsprechende Modell aufgeklärt wird. Da R^2 durch Hinzufügen zusätzlicher erklärender Variablen grösser wird oder gleich gross bleibt, sich der Fit der Gleichung aber niemals verkleinern

53 Ein Durbin-Watson-Wert von 0 deutet auf eine vollständig positive Autokorrelation, ein Wert von 2 auf keine Autokorrelation und ein Wert von 4 auf eine vollständige negative Autokorrelation hin (vgl. zum Durbin-Watson-Koeffizient Brosius 2013: 577–580).

54 SPSS überprüft das Vorhandensein resp. das Fehlen von Multikollinearität anhand des Toleranzwerts und des Variationsinflationsfaktors. Der Toleranzwert sollte nicht kleiner als 0.250 und der Varianzinflationsfaktor nicht grösser als 5.000 sein. Sind die Werte kleiner bzw. grösser liegt keine Multikollinearität vor (vgl. zur Kollinearitätsprüfung u. a. Brosius 2013: 580).

Tab. 9.4: Ergebnisse der multiplen Regressionsanalyse.

Modell	Prädiktoren	β	R^2	<i>korrigiertes</i> R^2	F
1	Grammatik	.777***	.604	.601	172.72***
2	Grammatik Rechtschreibung	.545*** .357***	.678	.672	117.81***
3	Grammatik Rechtschreibung Wortschatz	.429*** .312*** .233**	.710	.702	90.40***

* $p < .05$; ** $p < .01$; *** $p < .001$

Anm.: β = standardisierte Regressionskoeffizienten mit ermitteltem Signifikanzniveau, R^2 / *korrigiertes* R^2 = Bestimmtheitsmass / Fit der Gleichung mit ermitteltem Signifikanzniveau⁵⁴, F = F-Wert mit ermitteltem Signifikanzniveau

kann (und dadurch durch die Integration einer Vielzahl von Variablen theoretisch künstlich erhöht werden kann), wird neben dem R^2 häufig das *korrigierte* R^2 betrachtet, bei dessen Berechnung zusätzlich die Anzahl der bei der Schätzung verwendeten erklärenden Variablen berücksichtigt wird (Brosius 2013: 541–567).

Tabelle 9.4 zeigt die Ergebnisse der Regressionsanalyse. Die von den Kindern erzielten Leistungen in den Bereichen *Grammatik*, *Rechtschreibung* und *Wortschatz* erweisen sich alle als (hoch) signifikante Prädiktoren für die sogenannte globale Sprachkompetenz, wie sie im C-Test gemessen wird. Die drei Variablen zusammen (Modell 3) klären rund 70 % der Varianz (*korrigiertes* R^2) auf und sagen die globale Sprachkompetenz damit bemerkenswert gut voraus. Aufschlussreich ist ebenfalls, dass die Regressionskoeffizienten für alle drei Prädiktoren hoch sind. Das Resultat bestätigt die These, dass in C-Tests nicht ausschliesslich Lesekompetenzen gemessen werden, sondern insbesondere auch sprachproduktive Fähigkeiten in der Grammatik, der Orthographie und im Wortschatz, die des Weiteren für die kompetente Beherrschung des schriftlichen Ausdrucks von besonderer Relevanz sind.

⁵⁵ Beim *korrigierten* R^2 wird – anders als bei R^2 – mitberücksichtigt, wie viele erklärende Variablen in das Modell aufgenommen werden. Da der Fit der Gleichung durch das Hinzufügen weiterer erklärender Variablen niemals schlechter werden kann, führt die Aufnahme vieler erklärender Variablen in das Regressionsmodell zu einem immer höheren R^2 , aber nur zu einer vermeintlich guten Regressionsschätzung. Das wird durch das *korrigierte* R^2 ausgeschlossen.

9.1.2.5 Der CT-D 4 im Vergleich der Altersgruppen

Der CT-D 4 wurde als Sprachleistungstest für 4. Klassen konzipiert und normiert, in der vorliegenden Studie aber auch für die 6. Klassen eingesetzt. Die deskriptiven Masszahlen für die einzelnen Klassenstufen sowie die Paralleltestversionen A und B für die Texte 1 bis 4, aus denen der Test insgesamt besteht, sind Tabelle 9.5 zu entnehmen. Pro Text konnten maximal 25 Punkte erreicht werden. Die Streuung der Rohwerte bewegt sich bei den vierten Klassen zwischen 3.20 und 6.09 Standardabweichungen und bei den 6. Klassen zwischen 2.51 und 4.63 Standardabweichungen um den Mittelwert.

Tab. 9.5: Ermittelte Kennwerte für die Texte („Superitems“) 1–4 des CT-D 4.

	<i>N</i>	<i>Min.</i>	<i>Max.</i>	<i>M</i>	<i>SD</i>
4. Klasse Serie A					
CT-D 4 Text 1	24	3	23	15.58	4.45
CT-D 4 Text 2	24	6	21	14.96	4.17
CT-D 4 Text 3	24	7	21	15.42	3.20
CT-D 4 Text 4	24	4	23	12.38	4.46
4. Klasse Serie B^a					
CT-D 4 Text 1	20	5	22	13.20	4.6
CT-D 4 Text 2	20	2	22	11.30	4.79
CT-D 4 Text 3	20	0	22	11.15	6.09
CT-D 4 Text 4	20	2	20	11.20	5.56
6. Klasse Serie A					
CT-D 4 Text 1	37	12	25	21.05	2.51
CT-D 4 Text 2	37	10	25	19.76	3.50
CT-D 4 Text 3	37	7	23	17.89	3.77
CT-D 4 Text 4	37	6	25	17.43	4.57
6. Klasse Serie B^a					
CT-D 4 Text 1	34	8	24	17.97	4.17
CT-D 4 Text 2	34	9	23	15.65	4.18
CT-D 4 Text 3	34	8	25	15.65	3.93
CT-D 4 Text 4	34	7	24	16.85	4.63

Anm.: a. Die Parallelversionen A und B des CT-D 4 können als hinreichend parallel angesehen werden (Interkorrelation von $\alpha = .73$ bei $N = 46$, vgl. Raatz & Klein-Bräley 1992: 13), Testform B muss aber um 4 Punkte korrigiert werden. Hier aufgeführt sind die effektiv erreichten Punkte in den einzelnen Texten, die 4 Punkte sind noch nicht dazugerechnet. *N* = Anzahl Probanden, *Min.* = minimal erreichter Wert, *Max.* = maximal erreichter Wert, *M* = Mittelwert, *SD* = Standardabweichung

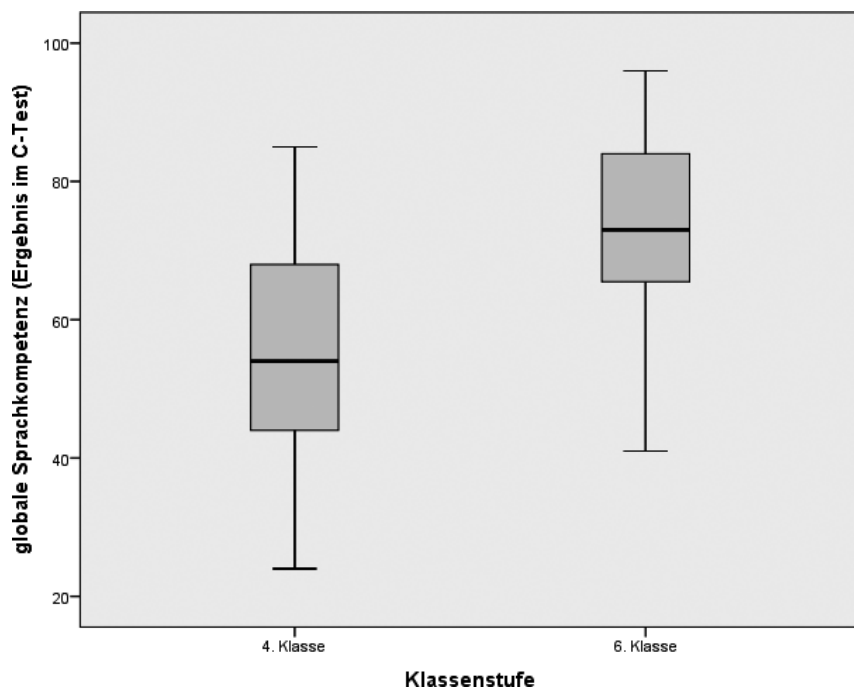


Abb. 9.2: Boxplot zu den Ergebnissen der KEKS-Untertests getrennt nach Klassenstufe.

Empfohlen wird von Raatz & Klein-Bräley (1992: 13) aufgrund eines etwas höheren Schwierigkeitsgrades der Version B eine Korrektur der Ergebnisse um vier Punkte. Ein Mittelwertvergleich des arithmetischen Mittels der C-Test-Resultate (Version B korrigiert um 4 Punkte) der jüngeren mit jenem der älteren Kinder mittels T-Test ergibt einen hoch signifikanten Unterschied ($t(113) = 6.44$, $p < .001$, $d = 1.21$). Die Kinder der 6. Klasse schneiden im C-Test also im Schnitt signifikant besser ab als die jüngeren. Tabelle 9.5 und Abbildung 9.2 zeigen für beide Altersstufen die errechnete Spannbreite zwischen dem minimal und dem maximal erreichten Testwert.

9.1.3 Erhebung der Kompetenz in der Lexik, der Grammatik und der Orthographie

Die Kompetenz in den Bereichen Lexik, Grammatik und Orthographie wurde anhand einzelner Untertests aus der Testbatterie *KEKS: Kompetenzerfassung in Kindergarten und Schule*, einem diagnostischen Konzept zur Erfassung von Kompetenzen und zur Beschreibung von Lernentwicklungen in den Bereichen

Tab. 9.6: Deskriptive Kennwerte zu den Ergebnissen in den KEKS-Untertests.

	<i>N</i>	<i>Min.</i>	<i>Max.</i>	<i>M</i>	<i>SD</i>
4. Klasse					
Grammatik	44	5	15	10.52	3.19
Wortschatz	44	2	15	8.89	3.01
Rechtschreibung ^a	44	14	38	23.68	5.47
6. Klasse					
Grammatik	71	3	15	12.56	2.53
Wortschatz	71	3	15	11.82	2.46
Rechtschreibung ^a	71	7	39	29.65	5.91

Anm.: a. Total der Untertests zur Rechtschreibung (bestehend aus den Subtests zur Schreibung deutscher Wörter und zu Pseudowortschreibungen)

N = Anzahl Probanden, *Min.* = minimal erreichter Wert, *Max.* = maximal erreichter Wert, *M* = Mittelwert, *SD* = Standardabweichung

Deutsch, Mathematik und Englisch aus Deutschland (May & Bennöhr 2013a: 5; May 2014), erhoben. Die Testbatterie wurde für die Klassenstufen bis Ende des vierten Schuljahres (KEKS 4Ü) konzipiert. Die Version KEKS 4Ü, die für das Ende des 4. resp. den Beginn des 5. Schuljahres vorgesehen ist, wurde in der vorliegenden Studie sowohl für die 4. als auch für die 6. Klasse eingesetzt. Da vergleichbare Sprachtests für die 6. Klasse nicht vorliegen und kein anderer Sprachtest für die 4. und die 6. Klasse existiert, schien es nach Rücksprache mit den Testentwicklern angemessen, den KEKS 4Ü auch für die ältere Schülergruppe zu verwenden.

Da die KEKS-Testserie auch für die Kinder der 6. Klasse eingesetzt wurde, gilt es, eine entsprechende Überprüfung der Kennwerte und der Testschwierigkeit vorzunehmen. Die in den Untertests aus der Testserie erzielten Ergebnisse im Klassenvergleich sind in Tabelle 9.6 aufgeführt.

Die Graphik zu den Lagemassen (Abb. 9.3) und die deskriptiven Kennwerte (Tab. 9.6) zeigen, dass die Testergebnisse bei den Kindern der 6. Klasse durchschnittlich besser sind. Es liegt eine breite Streuung vor, auch wenn relativ viele Kinder der 6. Jahrgangsstufe im Test sehr hohe Werte erreicht haben. Insbesondere beim Untertest zur grammatischen Kompetenz (arithmetisches Mittel von 12.56 bei einem maximal erreichbaren Wert von 15) könnte ein Deckeneffekt vermutet werden, so dass die Kinder in den höheren Punktbereichen nicht hinreichend differenziert würden. Den testtheoretischen Maximalwert von 15 Punkten erreichten 25 Kinder, was einem prozentualen Anteil von knapp 28 % entspricht. Werden Kinder, die 14 oder 15 erreicht haben, mitge-

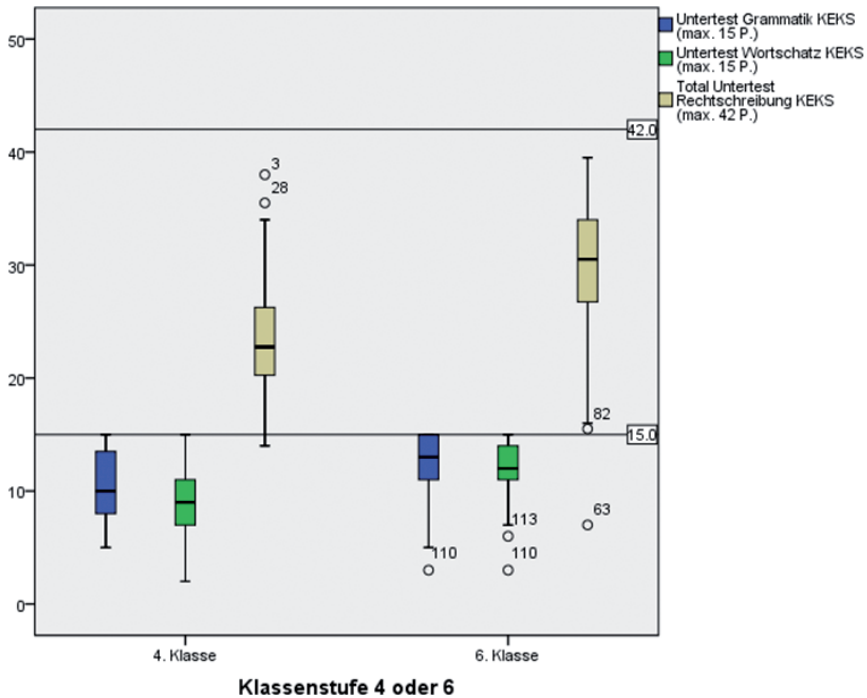


Abb. 9.3: Boxplot zu den Ergebnissen der KEKS-Untertests getrennt nach Klassenstufe.

zählt, ergibt sich auch hier ein prozentualer Anteil von etwas über 46 % – auch dieser Wert weist insgesamt noch auf einen mittleren Schwierigkeitsgrad hin. Der Test unterscheidet also ausreichend in grammatisch schwache, mittlere und hohe Kompetenzen. Die interne Konsistenz für die drei Untertests ist zufriedenstellend: Die Reliabilitätsstatistik ergibt für die 4. Klassenstufe ein Cronbachs Alpha von $\alpha = .751$ und für die 6. Klassenstufe ein Cronbachs Alpha von $\alpha = .775$.

Insgesamt stellt sich die KEKS-Batterie damit als taugliches Instrument heraus, um bei den Kindern der 4. sowie der 6. Klasse Kenntnisse und Kompetenzen in der Orthographie, der Grammatik und im Wortschatz zu erheben. Zudem zeigt sich, dass Kinder, die in einem der Bereiche gut abschnitten, tendenziell auch in den anderen entsprechend gute Werte erzielen.

Die Untertests aus der Sprachstandsmessung KEKS (vgl. Kap. 9.1.3) erheben die grammatische, die orthographische und die lexikalische Kompetenz. Hier wird argumentiert, dass sie durch das Erhebungssetting ebenfalls explizites Sprachwissen in den drei Teilbereichen miterfassen. Sie zielen im Rahmen einer klas-

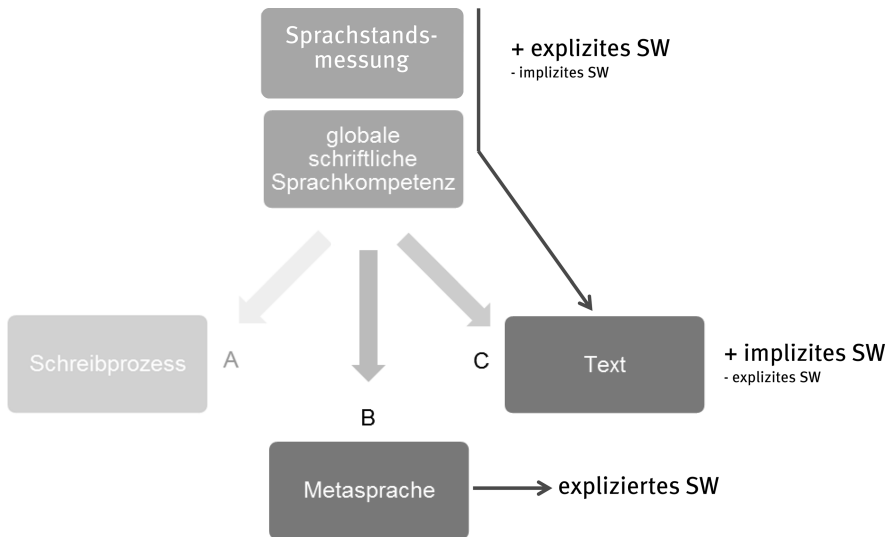


Abb. 9.4: Erhebung verschiedener Sprachwissenstypen.

sischen Testsituation (den Kindern werden insbesondere Sätze zur Vervollständigung vorgelegt) zwar auf eine Kompetenzmessung ab. Das Setting führt aber dazu, dass von den Kindern tendenziell stärker explizites Wissen (in Abb. 7.3 und Abb. 9.4 als *+explizites Wissen* aufgeführt) abgerufen wird. Im freien Schreiben hingegen spielt der Rückgriff auf implizites Sprachwissen (in Abb. 9.4 als *+implizites Wissen* aufgeführt) eine grössere Rolle: Das Auswählen eines korrekt flektierten Verbs für einen Satz mit Lücke (Testsituation) dürfte viel eher Bezüge zu explizitem Wissen über das Sprachsystem fordern als das Verfassen eines Satzes oder eines Textes mit einer bestimmten kommunikativen Absicht (freies Schreiben). Während die Sprachstandsmessung Sprachwissen also in einem klar konturierten Teilbereich abfragt, liegt der Fokus im freien Schreiben nicht auf einem sprachlichen Teilbereich, sondern verlangt im Gegenteil ein kompetentes Kombinieren verschiedener Fertigkeiten.

Zwischen diesen beiden beschriebenen Typen von Aufgaben und Anforderungen an die Kinder bildet der C-Test, das Instrument zur Messung der globalen Sprachkompetenz, den Übergang, indem er den Einsatz expliziten Wissens und den Rückgriff auf implizites Wissen gleichermaßen begünstigt. Die in der Studie erhobenen Daten ermöglichen daher die empirisch gestützte Diskussion darüber, wie *+implizites*, *+explizites* und *expliziertes* Sprachwissen zusammenhängen. Das explizierte Wissen wird in den Gesprächen mit den Kindern erhoben, wie im folgenden Kapitel geschildert wird. Abbildung 9.4 führt die Wissenstypen und die Korpora, anhand derer sie untersucht werden, auf.

9.1.4 Erhebung des Textes

Schreiben ist als Mittel zur Kommunikation immer kontextabhängig und zielgerichtet. Das gilt ebenso, wenn auch in eingeschränktem Mass, für schulisches Schreiben. Es ist eingegliedert in didaktische Settings, die ihrerseits entsprechenden Einfluss haben auf den Schreibprozess, das intendierte Ziel und den Text; „das didaktische Setting formt den Text der Lerner“ (Bachmann & Becker-Mrotzek 2010: 195). Der kommunikative Zweck ist in vielen Fällen des Schreibunterrichts nicht unmittelbar gegeben respektive entspricht nicht einer authentischen Alltagssituation. Bachmann & Becker-Mrotzek (2010: 195) empfehlen, Schreibaufgaben und Schreibenlässe darauf hin auszurichten, dass sie sinnvolle und lernförderliche Lernprozesse anregen und den Schreibauftrag so situieren, dass die Lernenden darauf mit einer sinnvollen Textform reagieren können. Sie fassen das Konstrukt eines lernförderlichen Schreibunterrichts unter dem Terminus *Aufgaben mit Profil* zusammen (Bachmann & Becker-Mrotzek 2010: 194). Zu den Bedingungen zählen sie folgende vier Merkmale (Bachmann & Becker-Mrotzek 2010: 195):

- Der zu schreibende Text muss für die Schüler/innen eine identifizierbare Funktion erfüllen; sie müssen erkennen können, welches kommunikative Problem damit bearbeitet werden soll. Denn nur wenn sie das Ziel und die Adressaten ihres Textes kennen, können sie sinnvolle Entscheidungen über den Aufbau, den propositionalen Gehalt und die Auswahl der sprachlichen Muster und Mittel treffen.
- Des Weiteren müssen die Schüler/innen Gelegenheit haben, sich das erforderliche Weltwissen, aber auch von Fall zu Fall einschlägiges sprachliches Wissen, anzueignen, damit sie wissen, worüber sie schreiben können und sollen.
- Die Schüler/innen müssen die Gelegenheit bekommen, ihren Text in einem Kontext sozialer Interaktion zu verfassen. Eingebettet in eine soziale Situation können vor allem junge Lerner die Zerdehnung der Sprechsituation leichter überwinden.
- Und schliesslich müssen sie Gelegenheit haben, die Wirkung ihres Textes auf die Leser zu überprüfen, so wie sie es aus der mündlichen Kommunikation gewöhnt sind (einigermassen etabliert haben sich in diesem Zusammenhang sogenannte ‚Schreibkonferenzen‘).

Sind diese Bedingungen erfüllt, wird von gut profilierten Aufgaben gesprochen. Das ist bei der Schreibaufgabe, die in der vorliegenden Studie eingesetzt wurde, weitgehend der Fall (vgl. Abb. 9.5): Die Kinder waren dazu angehalten, einen Brief an die zukünftige Schulleiterin zu verfassen, die aus Deutschland stammt und sich zum Zweck der Standardsprachförderung zum Ziel gesetzt

Deine Schule bekommt eine neue Schulleiterin aus Deutschland. Sie möchte, dass an der Schule kein Dialekt mehr gesprochen werden darf. Im Unterricht und auch in den Pausen müssen alle Kinder Hochdeutsch sprechen, damit sie gut Hochdeutsch lernen.

Schreib der Schulleiterin einen Brief und beschreib, wann und warum du Dialekt oder Hochdeutsch sprichst. Schreib deine Meinung zu ihrem Vorschlag und begründe sie.

Abb. 9.5: Schreibaufgabe.

hat, dass Schüler und Schülerinnen auch in der Pause und zwischen Schullektionen zum Hochdeutschsprechen angehalten werden. Als *identifizierbare Funktion* kann die Rückmeldung an die neue Schulleiterin und deren Beeinflussung gelten. Diese stellt eine zwar unbekannte, aber durch bestimmte Merkmale (sie stammt aus Hamburg und ist neu in der Schweiz) konturierte Person mit einer spezifischen Funktion im schulischen Kontext dar. Allerdings ist einzuräumen, dass in den entsprechenden Schulen kein tatsächlicher Wechsel der Schulleitung anstand. Obwohl Rückfragen zur Authentizität des Schreibanlasses durch die Probanden selten waren, muss doch davon ausgegangen werden, dass sie sich des fiktiven Charakters der beschriebenen Situation bewusst waren.

Die Schreibaufgabe war zudem so ausgerichtet, dass jedes Kind auf das eigene *Weltwissen* zurückgreifen konnte: Dem Nebeneinander von Dialekt und Standard begegnen die Schüler der Deutschschweiz spätestens mit der Einschulung und sie ist damit Teil der individuellen Spracherfahrung. Die Reflexion über die Schweizer Diglossie ist in diesem Alter im Lehrplan empfohlen: Das Lehrmittel *Die Sprachstarken* (Lötscher et al. 2007), das kantonal vorgeschrieben und in allen befragten Klassen verwendet wird, integriert das Thema im vierten Schuljahr (*Sprechweisen – Wo sprichst du wie?* Lötscher et al. 2007, S. 24 f.) und nimmt es später verschiedentlich wieder auf. Die Kinder kennen die Verwendung zweier Varietäten damit einerseits aus der eigenen Spracherfahrung und andererseits aus dem Unterricht, wobei die Bearbeitung des entsprechenden Kapitels nach Aussagen der Lehrpersonen für die ältere Klassenstufe ca. zwei Jahre und für die jüngere mindestens einige Monate zurücklag. Diese vorangehende Auseinandersetzung mit dem Thema ermöglichte den Kindern die Aneignung von Weltwissen, teilweise – wenn auch in eingeschränktem Mass – auch von sprachlichem Wissen. Mit der Schreibaufgabe wurde also bewusst an ein alltagsrelevantes Thema angeknüpft, gleichzeitig aber eine Dekontextualisierung elizitiert, indem darauf hingewiesen wurde, dass die neue Schulleiterin aus Deutschland stamme. Die Einbettung der Schreibaufgabe in einen *Kontext sozialer Interaktion* war ebenso wie die *Überprüfung der Leserwir-*

kung bedingt gegeben: Die Untersuchungsleiterin war in der Schreibstunde anwesend und wurde als Person vorgestellt, die sich für die Meinung der Kinder interessiert, was die Zerdehnung der Kommunikationssituation insofern zu überbrücken half, als der imaginierten Schulleiterin eine präsente Person als Überbringerin der Nachrichten an die tatsächliche Adressatin zugeordnet wurde. Das anschliessende Gespräch mit der Untersuchungsleiterin, in dem eingangs auch die Meinung des Kindes, die im Brief vertreten wurde, Gegenstand war, ermöglichte zwar eine gewisse Überprüfung der Textwirkung, diese war aber sehr eingeschränkt. Die Einschränkung bestand insbesondere durch den Umstand der asymmetrischen Hierarchie: Schreibkonferenzen, die Bachmann & Becker-Mrotzek (2010: 195) in Zusammenhang mit diesem Aspekt der Aufgabenprofilierung als Beispiel anführen, ermöglichen den Austausch unter *peers* und gerade nicht zwischen Lehrperson (oder Untersuchungsleiterin) und Kind. Damit darf die Aufgabe insgesamt als teilweise profiliert bezeichnet werden.

Durch die Schreibaufgabe wurden die Kinder angehalten, sich einer Textsorte zu bedienen, die sie in dieser Form aller Wahrscheinlichkeit nach nicht aus Übung beherrschen, da das ausgebaute erörternde Schreiben und das Verfassen formal und kommunikativ adäquater Briefe in der Grundschule nicht Gegenstand der Vermittlung ist. Damit wird vermieden, dass die Kinder auf im Unterricht erworbene Textschablonen oder Musterargumentationen zurückgreifen (vgl. dazu auch Augst et al. 2007: 200, die das gleiche Vorgehen gewählt haben).

Für die Schreibaufgabe standen den Kindern maximal 45 Minuten zur Verfügung. Die volle Zeitspanne wurde nur von einem Kind ausgenutzt. Zeitdruck war also keine relevante Einflussgrösse für die Texte. Erprobt wurde die Schreibaufgabe von einer Studentin im Rahmen einer Masterarbeit (Rotzetter 2012) sowie in der Vorerhebung für die Datengewinnung der vorliegenden Studie. Zudem wurde sie zusammen mit der Lehrperson, in deren Klasse die Vorerhebungen stattgefunden hatten, besprochen und angepasst. In Bezug auf das Anspruchsniveau, die Verständlichkeit und das Thema hat sie sich daher in mehreren Anwendungen als angemessen herausgestellt.

9.1.5 Erhebung des Schreibprozesses durch *Keystroke Logging*

Zur Erhebung der Textprodukte und -prozesse wurde auf ein Erhebungsinstrument zurückgegriffen, das im deutschsprachigen Raum, insbesondere in der Schweiz, bislang noch selten zum Einsatz kam: *Keystroke Logging*. Das verwendete Programm *inputlog* (Leijten & van Waes 2014) wurde zur Erhebung von Prozessdaten zu Schreibvorgängen entwickelt. Eine computerbasierte Erhe-

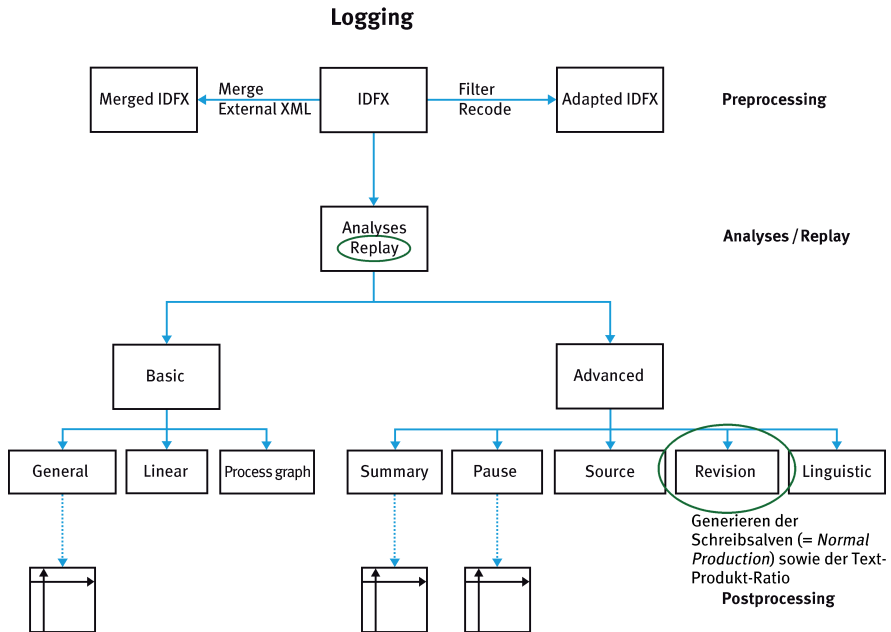


Abb. 9.6: Detaillierte Übersicht über den Flow von *inputlog* (Vgl. Leijten & van Waes 2014).

bung mit *Keystroke Logging* liefert drei Datenbereiche: das Schreibprodukt als statisches Endprodukt des Schreibens, ein Log-File mit dem zeitlichen Verlauf der Textproduktion inklusive Schreib-, Lösch- und Revisionsoperationen und den entsprechenden Zeitangaben sowie einen Abspielmodus, der die Textproduktion visualisierbar macht (vgl. hierzu auch Weder 2010: 88). Die in *inputlog* möglichen Analyseschritte und die dadurch generierten Files sind in Abbildung 9.6 ersichtlich. Für eine Beschreibung der Inhalte der einzelnen Analysen sei auf Leijten & van Waes (2014) verwiesen. Die in der vorliegenden Studie ausgewerteten Variablen (*Text-Prozess-Ratio* und *Salvenumfang*) und Auswertungsgrundlagen (als Grundlage dienten die aufgezeichneten Filme sowie die Darstellung in S-Notation und der Prozess-Graph) wurden mit Hilfe der Software generiert und teilweise manuell korrigiert und angepasst.

Ob von Hand oder von Computer geschrieben wird, verändert den Schreibprozess und die Untersuchungsmöglichkeit desselben. Empirische Studien (vgl. beispielsweise Morphy & Graham 2012) belegen, dass sich das Schreiben am Computer insbesondere bei schwachen Schreibern und Schreiberinnen positiv auf die Textqualität, auf die Textstruktur, auf die grammatische und die orthographische Korrektheit und insbesondere auf die Schreibmotivation aus-

wirkt. Zu einem ähnlichen Befund kommt die Schweizer Interventionsstudie von Schneider et al. (2012): Sie konnte einen stärkeren Leistungszuwachs für Kinder in ihren narrativen (nur bedingt aber in den instruktionalen) Texten nachweisen, die statt auf Papier auf einer Internetplattform schrieben. Dies gilt für die 4. und 5. Primarklasse, nicht aber für die 1. und 2. Klasse (Schneider et al. 2013: 52). Diese Befunde sprechen dafür, dass sich computergestützte Erhebungsverfahren für Kinder im Alter der hier untersuchten Probanden (4. und 6. Klassenstufe) eignen, um prozessbezogene Aspekte des Schreibens zu untersuchen.

Was auch die computergestützte Erfassung von Schreibprozessen nicht erfassen kann, sind kognitive Vorgänge und Denkprozesse. Daher „münden in der Schreibprozess-Forschung alle Überlegungen zur Methodenwahl in der Erkenntnis, dass *Keystroke-Logging* mit direkten Beobachtungsmethoden zu kombinieren sei, z. B. mit Verbalisierung, um objektiv gewonnene Daten unter Beizug subjektiver Daten zu interpretieren“ (Weder 2010: 88–89). In der vorliegenden Studie wurde zur Erhebung metasprachlicher Äusserungen und des Sprachwissens der Kinder die Text- und Produkterhebung ergänzt durch eine Konfrontation der Kinder mit ihrem Schreibprozess.

9.1.6 Erhebung und Aufbereitung der Daten zum explizierten Sprachwissen

Studien zu metasprachlichen Kompetenzen und Sprachbewusstheit erheben ihre Daten häufig, indem die Probanden Akzeptabilitätsurteile zu ihnen vorgelegten sprachlichen Strukturen abgeben sollen. Alternativ sind sie angehalten, abstrakte Kategorien zu erklären oder zu beschreiben, so etwa die Kategorie *Wort*. Diese Äusserungen werden dann als Indiz für eine vorliegende oder fehlende Sprachbewusstheit gewertet. In der vorliegenden Studie wurde ein anderes Vorgehen gewählt: Anschliessend an die Schreibstunde wurde den Kindern ihr Textentstehungsprozess vorgespielt mit der Bitte, diesen und Beobachtungen aller Art zu kommentieren. Methodisch lehnt sich das Vorgehen an die Erhebung von retrospektiven *Think-aloud*-Protokollen an. Die Methode des lauten Denkens wurde in der Sprachforschung angewendet, um Hinweise auf Prozesse zu erhalten, die sich auf der kognitiven Ebene abspielen: Die Probanden sind angehalten, während des Schreibens oder während des Lösen bestimmter Aufgaben laut zu artikulieren, was ihnen dabei durch den Kopf geht, sie sollen also laut denken (vgl. für den Einsatz in der Fremdsprachforschung etwa Heine 2005; Heine & Schramm 2007). Als erheblicher Nachteil des Vorgehens ist (gerade bei Kindern) die zusätzliche kognitive Belastung durch eine weitere Aktivität, das Aussprechen dessen, was gedacht wird, zu nennen. Retrospektive Kommentierungen andererseits können nur das betreffen, was erinnert

wird, sie unterliegen also dadurch bereits Restriktionen. Im Unterschied zu diesem Verfahren werden die Kommentare der Kinder in der vorliegenden Studie nicht synchron, sondern zeitlich versetzt zum Schreibvorgang, also asynchron respektive *off-line*, anhand retrospektiver Gespräche (*stimulated recall*) erhoben (vgl. zur Methode Gass & Mackey 2000, Hinweis in Weder 2010: 88). Die Methode hat das Ziel, „to stimulate writers’ retrospective accounts of their composing strategies and to raise their awareness of them“ (Abdel Latif 2008: 42). Innere Vorgänge werden nach dem Schreiben aufgrund eines Stimulus – im vorliegenden Fall der Textentstehungsprozess – verbalisiert. Die Verbalisierungen lassen auf Wissensstrukturen, kognitive Prozesse und Problemlösestrategien schliessen (Weder 2010: 88), die der direkten Beobachtung nicht zugänglich gemacht werden können:

In vielen Fällen jedoch kann das, was sich äußerlich als Ringen ums richtige Wort darstellt, auch andere Ursachen [als Formulierungsprobleme] haben. Die Materialisierung eines Gedankens [sic] umfasst nicht nur das Abrufen und Aneinanderreihen einzelner Wörter, sondern auch viele Schritte der Konkretisierung, die sich zum Teil abwechselnd auf der mentalen, averbalen Ebene und der verbalen schriftlichen Oberfläche abspielen, teilweise auch zwischen den beiden Ebenen auf dem Weg vom Gedanken zum Wort. Dem Bewusstsein ist das Wenigste davon zugänglich, doch gibt es in der Schreib- und Sprachproduktionsforschung Ansätze zur Diagnose und Erklärung der kognitiven Prozesse, die hier ablaufen. (Molitor-Lübbert 2002: 35)

Recalls sind in Kombination mit *Keystroke-Logging* ein geeignetes Verfahren, um zu beschreiben, wie sich Kinder über Sprachliches äussern, wie sie sich erklären, was sie kommentieren und welche eigene „Grammatik“ sie dabei ansetzen. Dies ungeachtet dessen, ob darauf sprachhandelnd in der konkreten Situation, die den Anlass für das Gespräch bildete, tatsächlich zurückgegriffen wurde, was methodisch auch mit lautem Denken oder anderen Zugängen in keinem Fall zweifelsfrei zu entscheiden ist. Zudem ist als weiterer Unterschied zu *Think-aloud*-Protokollen anzumerken, dass die Kommentierungen des Kindes in ein Gespräch mit der Untersuchungsleiterin eingebettet sind. Die Äusserungen kamen im Rahmen eines Gesprächs zustande und nicht im Rahmen eines Monologs zu einem vorgegebenen Input. Die Gespräche stellen damit eine Art ‚sprachbetrachtende Diskussion‘ mit dem Schreibprozess als Input dar. Dieser wurde den Kindern als Film, der beschleunigt oder verlangsamt sowie jederzeit gestoppt werden konnte, vorgespielt. Das Vorgehen hat den Nachteil der erschwerten Vergleichbarkeit von Äusserungen, weil sich der Input von Proband zu Proband unterscheidet. Ausserdem spielen das Beisein und die Gesprächsteilnahme der Untersuchungsleiterin eine Rolle dabei, wie die Äusserungen zustande kommen, zumal teilweise Rückfragen gestellt wurden. Dieser Einfluss ist allerdings zu relativieren, da sich die Nachfragen der

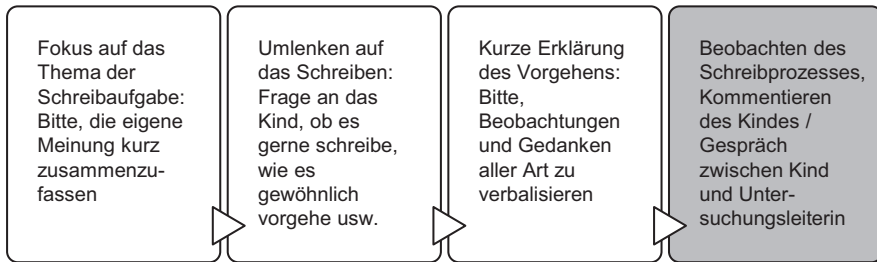


Abb. 9.7: Ablauf und inhaltliche Struktur der *recalls*.

Untersuchungsleiterin weitgehend darauf beschränkten, nach einer Begründung einer getätigten Revision zu fragen, die das Kind nicht von sich aus kommentierte. Die Einbettung in ein Gespräch war notwendig, um für die Kinder die Hemmschwelle, die die Erhebungssituation hervorrufen konnte, möglichst tief zu halten. Als entscheidenden Vorteil des Settings ist hervorzuheben, dass die Kinder unmittelbar auf eine individuelle Sprachproduktionserfahrung zurückgreifen konnten, da ihnen nicht irgendein Schreibprozess, sondern der eigene vorgelegt wurde.

Der Ablauf der Erhebung gestaltete sich auch hier in allen Klassen gleich: Nach dem Verfassen des Textes gingen die Kinder zurück in das Klassenzimmer und folgten dem regulären Unterricht. Jeweils ungefähr 30 Minuten nach dem Ende der Schreibstunde wurde das erste Kind zur mündlichen Befragung aus der Klasse geholt. Nach dem Zufallsprinzip wurde bis zur Mittagspause ein Kind nach dem anderen befragt, danach wurde die erste Erhebungsphase – unabhängig davon, wie viele *recalls* erhoben werden konnten – abgeschlossen. Die Zeit zwischen Schreibvorgang und Kommentierung desselben unterscheidet sich von Kind zu Kind, beträgt aber für alle mindestens 30 und maximal 110 Minuten. Inhaltlich bestehen die Gespräche aus vier Teilen (vgl. Abb. 9.7). In die spätere Auswertung miteinbezogen wurde nur der Hauptteil des Gesprächs (in der Abbildung blau hervorgehoben).

Die Gespräche mit den Kindern wurden mit der Software *f4* (Dresing & Pehl 2013) vollständig transkribiert (zu den verwendeten Transkriptionsregeln vgl. die Ausführungen im Anhang) und mit Hilfe von *MaxQda*, einer Software zur qualitativen Datenauswertung, annotiert und ausgewertet. Mit Bezug auf die Studie zur Struktur und Funktion metasprachlicher Kommentare von Vorschulkindern von Stude (2012) wurde dabei von Referenzbereichen metasprachlicher Kommunikation und Reflexion ausgegangen. Diese können jeweils etwa metasyntaktische, metasemantische oder metapragmatische Aspekte gleichermaßen umfassen.

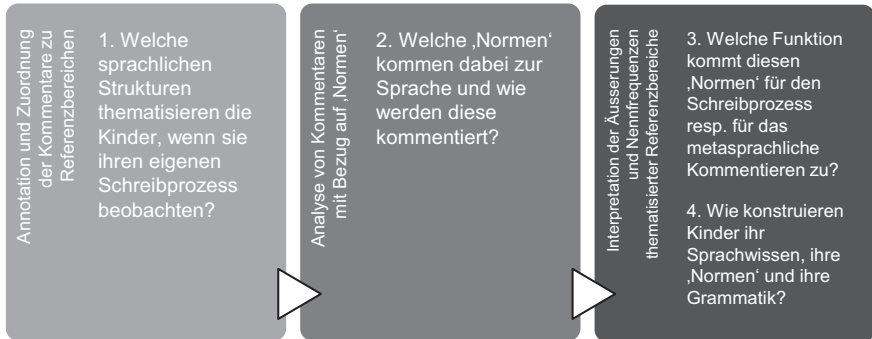


Abb. 9.8: Fragen und Vorgehen bei der Auswertung der Kindergespräche.

Da während des Gesprächs mit den Kindern durch den Input automatisch Revisionen in den Fokus der Aufmerksamkeit gerückt wurden, bildet die Zuordnung der Kinderkommentare zu verschiedenen Referenzbereichen zu einem grossen Teil ebenfalls ab, auf welchen sprachlichen Ebenen die Kinder im Schreibprozess ihre Texte überarbeitet haben. Die Metasprachprofile sind damit teilweise auch Überarbeitungsprofile. In der Auswertung wird der Frage nachgegangen, ob und inwiefern sich diese Profile zwischen sprachstarken und sprachschwächeren Kindern unterscheiden. Die Gegenüberstellung liesse Rückschlüsse darauf zu, welche Arten und Bereiche von ‚Normen‘ bei den Kindern präsent sind und welchen Funktionen sie diesen zuordnen.

Abbildung 9.8 visualisiert in drei vereinfachten Schritten das Vorgehen von der Annotation der Gespräche bis zu deren Interpretation in Bezug auf die Normvorstellungen.

9.1.7 Erhebung des Selbstkonzepts der Kinder

In Kapitel 3.3.4 wurde die Rolle der Überzeugungen in Bezug auf die eigenen Kompetenzen diskutiert. In der psychologischen Forschung werden diese Überzeugungen unter dem Terminus *Selbstkonzept* (Bandura 1986) subsumiert. Das Selbstkonzept stellte sich in verschiedenen Studien zu unterschiedlichen Fähigkeitsbereichen als zentraler Prädiktor für die Leistungen eines Individuums heraus (vgl. Kap. 3.3.4). Im Rahmen der Datenerhebung für diese Studie wurde das Selbstkonzept der Kinder in Bezug auf das Fach Deutsch, das Schreiben, den Umgang mit dem PC und die Tastaturkompetenz erhoben. Die Daten dienen der Beschreibung der Probanden (vgl. Kap. 10.2.4) sowie als Kontrollvariablen: So ermöglicht das Selbstkonzept der Kinder in Bezug auf

das Tastaturschreiben einzuschätzen, ob das Textschreiben von den Kindern der beiden Altersgruppen aufgrund der Tastaturbeherrschung unterschiedlich ausfiel (vgl. Kap. 11.1.1).

Die Erhebung des Selbstkonzepts gestaltete sich folgendermassen: Im Schülerfragebogen wurde den Schülern und Schülerinnen eine Reihe von Skalen vorgelegt, die Aussagen zum Selbstkonzept in Bezug auf das Fach Deutsch, in Bezug auf das Schreiben, in Bezug auf die PC- und in Bezug auf die Tastaturbeherrschung messen. Bandura (1995, 1997, vgl. Pajares 2003: 142) hat klare Richtlinien für die Operationalisierung und Messung des Selbstkonzepts erstellt: Das Selbstkonzept in Bezug auf einen bestimmten Bereich wird anhand einer oder mehrerer Skalen bestehend aus verschiedenen Items erhoben. Letztere sind zu formulieren als *kann*-Sätze, die eine Beurteilung der eigenen Kapazität abgeben (*Ich kann gut Briefe schreiben.*). Bandura (1995, 1997) empfiehlt, den Probanden die Fragen zum Selbstkonzept in zeitlich möglichst kleinem Abstand vor der Erhebung dessen, was mit ihm verglichen werden soll, vorzulegen. Letzteres war im vorliegenden Fall nicht möglich. Der Zeitraum, der zwischen den beiden Erhebungsdaten lag, war hingegen nicht allzu gross (4–6 Wochen).

Die Konzeption der Skalen und Ausformulierungsarten der Items lehnt sich in der vorliegenden Studie an eine der aktuell umfangreichsten Erhebungen, an die PISA-Studie (Baumert 2001; Moser 2001), an. In Bezug auf die Formulierungsmuster wurde dort von Banduras (1995) Empfehlung abgewichen: Den Kindern wurden keine *kann*-Sätze vorgelegt, sondern Behauptungen in Bezug auf die eigenen Vorlieben (*Ich schreibe Aufsätze lieber auf dem Computer als von Hand.*) oder die eigene Kompetenz (*Ich schreibe ohne grosse Probleme auf der Tastatur.*). Der Unterschied zu den *kann*-Sätzen entspricht der mittlerweile etablierten Differenzierung in Selbstkonzept (Einschätzung in Bezug auf die eigenen Fähigkeiten) und Selbstwirksamkeitskonzept (Einschätzung in Bezug darauf, was mit den eigenen Fähigkeiten bewirkt werden kann) (Pajares 2003). Verwendet wurde wie in der PISA-Studie eine 4-stufige Skala, auf der 1 die negativste (*stimmt gar nicht*) und 4 die positivste Einschätzung (*stimmt genau*) darstellt.

Die Reliabilitätsanalysen ergeben für die erhobenen Skalen Werte zwischen $\alpha = .791$ und $\alpha = .861$ und Trennschärfen mit sehr guten Werten (vgl. Tab. 9.7). Aufgrund der errechneten Mittelwerte und der Streuung können Boden- und Deckeneffekte ausgeschlossen werden. Es kann also davon ausgegangen werden, dass die Skalen den entsprechenden Bereich gut erfassen und die verschiedenen Kinder durch sie differenziert werden. Für die Validität der Skalen spricht ebenfalls, dass sie – auch wenn einzelne Items in der Formulie-

Tab. 9.7: Ermittelte Kennwerte der Skalen zur Messung des Selbstkonzepts.⁵⁵

Skala	<i>n</i> _{Items}	<i>Alpha</i>	<i>r</i> _{itc}
1 Selbstkonzept in Bezug auf das Fach Deutsch	5	.86	.68
2 Selbstkonzept in Bezug auf das Schreiben	5	.81	.59
3 Selbstkonzept in Bezug auf den Umgang mit dem PC	6	.79	.55
4 Selbstkonzept in Bezug auf den Umgang mit der Tastatur	5	.80	.59

Anm.: *r*_{itc} = gemittelter Trennschärfekoeffizient, Alpha = Cronbachs Alpha (*N* = 111 für Skala 1, *N* = 112 für Skalen 2 und 3, *N* = 110 für Skala 4)

rung leicht angepasst oder gar neu hinzugefügt wurden – in anderen Studien bereits eingesetzt und evaluiert wurden.

9.1.8 Erhebung kognitiver Grundfähigkeiten

Im zweiten Erhebungsteil wurde neben der Erhebung der globalen schriftlichen Sprachkompetenz und den Sprachkompetenzen in einzelnen sprachlichen Teilbereichen auch die Aufmerksamkeits- und Konzentrationsleistung erhoben. Das Erhebungssetting für die Schreibaufgabe im Rahmen der vorliegenden Studie wurde für die Kinder massgeblich dadurch beeinflusst, dass sie ihren Text auf PC verfassten. Neben der Selbsteinschätzung der Kinder in Bezug auf ihre Tastaturschreibkompetenz und die Übung, die sie im Umgang mit dem PC haben, könnte, so die Vermutung, die Aufmerksamkeits- und Konzentrationsfähigkeit Einfluss darauf nehmen, wie intensiv und wie konzentriert die Kinder auf dem PC an ihrem Text arbeiten. Konzentration erforderte zudem die Bearbeitung des C-Tests: Die Kinder mussten die vier Lückentexte unter Zeitdruck möglichst sorgfältig bearbeiten. Um einen möglichen Einfluss der Konzentrationsleistung zu kontrollieren oder sie als erklärende Grösse hinzuziehen zu können, wurde die Aufmerksamkeits- und Konzentrationsleistung erhoben.

Als Testverfahren wurde der *Aufmerksamkeits- und Konzentrationstest d2-Revision* (d2-R) (Brickenkamp, Schmidt-Atzert & Liepmann 2010) gewählt. Der Test wurde vielfach erprobt und liegt in einer revidierten Fassung vor. Er erfüllt die Testgütekriterien mehr als zufriedenstellend und ist in der Durchführung zeitökonomisch und unkompliziert. Die Probanden sind im d2-R vor die Aufgabe gestellt, auf Zeilen mit ähnlichen Graphemen jeweils ein mit zwei Strichen

⁵⁶ Da sich hinsichtlich der beiden Gruppen keine grossen Unterschiede ergaben (vgl. Kap. 10.2.4), werden die Kennwerte für die Probanden beider Altersgruppen zusammen aufgeführt.

markiertes *d* zu finden und durchzustreichen, was dem Testverfahren seinen Namen gibt (Brickenkamp, Schmidt-Atzert & Liepmann 2010).

Die Auswertung wurde gemäss Anleitung vorgenommen, woraus sich für jeden Probanden verschiedene Kennzahlen (Konzentrationsleistung, bearbeitete Zielobjekte, Fehlerprozente, Konzentrationsverlauf und Arbeitsstil) ergaben. Der hier interessierende Wert, die Konzentrationsleistung (KL), wurde anhand der Normtabellen in einen Standardwert (SW) transformiert. Ausschlaggebend dafür war das Alter des Probanden. Der Standardwert gibt an, wie die Leistung des Kindes im Vergleich zu der Leistung gleichaltriger Kinder zu beurteilen ist (Brickenkamp, Schmidt-Atzert & Liepmann 2010: 36).

Aus der Schreibforschung ist bekannt – und das geht auch aus den visuellen Modellierungen des Schreibprozesses hervor (vgl. etwa die Modelle von Hayes 1996 oder Berninger & Winn 2008, beschrieben in Kap. 4.2) –, dass das Arbeitsgedächtnis einen zentralen Einfluss auf die Schreibfähigkeit hat. Es wäre daher wünschenswert, in Erhebungen der schriftsprachlichen Fähigkeiten die Arbeitsgedächtnisleistung zu integrieren, um sie als mögliche Einflussgrösse bei der Datenauswertung mit zu berücksichtigen, wie das in der vorliegenden Studie etwa mit dem sozioökonomischen Hintergrund gemacht wurde. Die Verfahren zur Erhebung des Arbeitsgedächtnisses sind allerdings zeitaufwändig und hätten die zur Verfügung stehende Erhebungszeit mit den Kindern überstrapaziert. Aus diesem Grund wurde darauf verzichtet.

9.2 Beschreibung der Datenerhebung

9.2.1 Rekrutierung der teilnehmenden Kinder und Klassen

Nach Einholen der benötigten Erlaubnis der Erziehungsdirektion des Kantons Freiburg CH wurden verschiedene Schulleitungen für eine Teilnahme an der Studie angefragt. Die Auswahl erfolgte weitgehend zufällig anhand einer Liste von Schulen aus dem deutschsprachigen Teil des Kantons respektive richtete sich nach der Bereitschaft und der Verfügbarkeit der Lehrpersonen und Klassen. Ausgeschlossen wurden innerhalb des Kantons Schulen, die zweisprachig sind oder die zwei Klassenstufen in einer Klasse unterrichten. Aus erhebungsökonomischen Gründen wurden zudem Schulen bevorzugt, die jeweils eine oder zwei vierte respektive sechste Klassen zur Verfügung stellen konnten.

9.2.2 Ablauf der einzelnen Erhebungen

Die Datenerhebung wurde zwischen Mitte März und Ende April 2013 in insgesamt sieben Klassen, davon drei 4. und vier 6. Klassen, an vier Primarschulen des Kantons Freiburg durchgeführt. Jede Klasse wurde zweimal besucht, wobei in einer ersten Phase die Texte und *recalls* und in einer zweiten die standardisierten Tests erhoben wurden. Der Erhebung voraus ging ein Informationsschreiben an die Eltern und eine Erhebung ausgewählter Sozialvariablen durch dieselben.

Die erste Erhebungsphase (vgl. Abb. 9.9) wurde im zeitlichen Ablauf in allen Klassen gleich gestaltet: Nach einer vorgängig abgesprochenen Einführung in die Thematik durch die Lehrperson zu Beginn der ersten Vormittagslektion schrieben die Kinder in einem eigens mit Laptops eingerichteten Raum den Text zur vorgegebenen Schreibaufgabe (vgl. Schreibaufgabe im Anhang). Die Untersuchungsleiterin war jeweils sowohl während des Einstiegs als auch während der Schreibstunde anwesend. Die Einführung in die Thematik war kurz gehalten (ca. 5 Minuten) und verfolgte zwei Zwecke: Anhand einiger Stichworte wurde das Thema der Schreibaufgabe (Dialekt – Standard) eingeführt, wobei ebenfalls Synonyme (Mundart/Dialekt, Standard/Schriftsprache/Hochdeutsch) geklärt wurden (vgl. Information z. H. der Lehrperson im Anhang). Die Reflexion der diglossischen Sprachsituation wird als Unterrichtsgegenstand im Lehrplan für die 4. Klasse empfohlen und in den folgenden Schuljahren wieder aufgenommen. Die Wissenssynchronisierung im Vorfeld der Erhebung sollte alle Kinder in Bezug auf den Faktor *Vorwissen* auf einen gleichen oder ähnlichen Stand bringen. Nach der Einführung wurde die Schreibaufgabe gemeinsam gelesen und allfällige Verständnisfragen konnten geklärt werden.

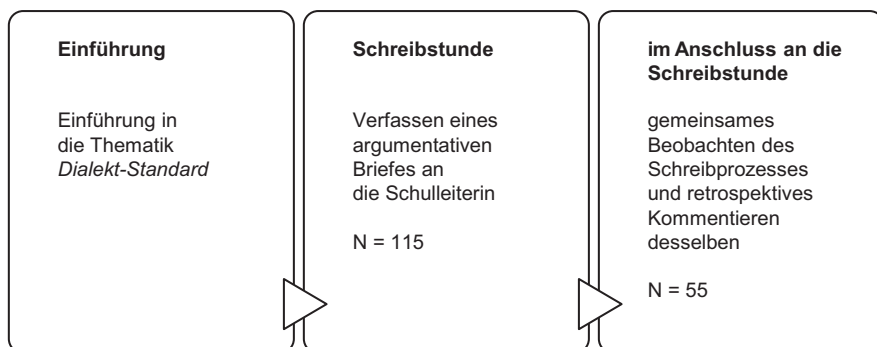


Abb. 9.9: Ablauf der ersten Erhebungsphase.

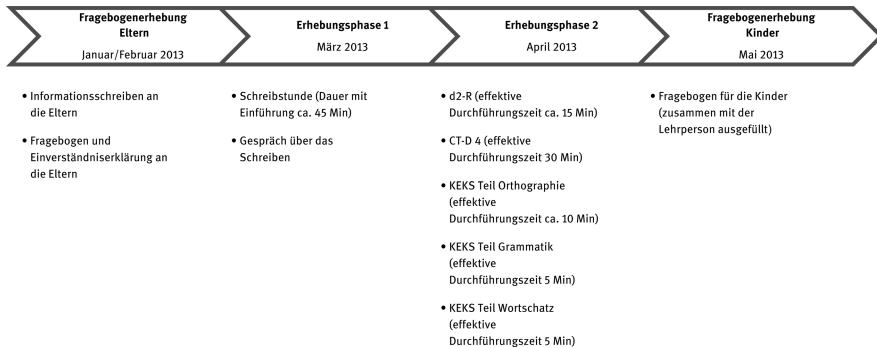


Abb. 9.10: Ablauf der gesamten Erhebung.

Die zweite Erhebungsphase war ca. 4–6 Wochen später. Sie beinhaltete einen standardisierten Test zur Erhebung der Aufmerksamkeit (d2-R), ein Verfahren zur Messung der globalen Sprachkompetenz (CT-D 4) sowie drei Untertests aus einer Sprachstandsmessung (KEKS). Die Erhebungen dauerten insgesamt rund 90 Minuten (zwei Schullektionen). Die Auflistung, wie sie in Abbildung 9.10 aufgeführt ist, entspricht der Reihenfolge der Testdurchführung.

9.3 Methoden der Datenanalyse und untersuchte Gruppen

Anhand der vorangehend präsentierten Erhebungsinstrumente werden Daten erhoben, die sich entweder ausschliesslich für die quantitative Auswertung, nur für eine qualitative Analyse oder für eine Kombination der beiden Methoden eignen. Die Kombination methodisch unterschiedlicher Zugänge als „Optimierungsstrategie“ (Ecarius & Miethe 2011: 7) von Forschungsergebnissen ist nicht neu und nicht unproblematisch (vgl. bspw. Flick 2011 oder Berthele 2012). Auf eine Grundsatzdiskussion wird hier insbesondere darum verzichtet, weil sich die unterschiedlichen Methoden jeweils auf unterschiedliche Arten von Daten beziehen.⁵⁷

⁵⁷ Als Ausnahme davon können die Frequenzen der kommentierten Referenzbereiche in den Kapiteln 11.2.2 und 11.2.3 betrachtet werden, die auf eine Abhängigkeit von den Altersgruppen und/oder den Kompetenzgruppen hin untersucht werden: Hier werden Kommentare Referenzbereichen zugeordnet und ausgezählt. Die Gegenüberstellung und Interpretationen von Häufigkeiten ist aber auch in der qualitativen Forschung durchaus üblich.

9.3.1 Form und Auswertung der Daten

Tabelle 9.8 führt die in der vorliegenden Studie verwendeten Verfahren auf und gibt über die Form der erhobenen Daten sowie über die Auswertungsmöglichkeiten Auskunft.

Tab. 9.8: Übersicht über das Korpus: Form der Daten und Auswertungsmöglichkeit.

	Form der Daten	Auswertung
Text	Texte (Word-Datei)	Qualitativ / quantitativ
Schreibprozess	u. a. S-Notation, Prozess-Graph, Film mit Schreibprozess	Qualitativ / quantitativ
<i>recall</i>	kommentierter Schreibprozess / Gespräch	qualitativ
C-Test	Punktwert	quantitativ
Elternfragebogen	geschlossene Fragen	quantitativ
Schülerfragebogen	geschlossene Fragen / (einzelne offene Fragen)	quantitativ
KEKS Untertests	Punktwerte <i>Orthographie, Grammatik</i> und <i>Wortschatz</i>	quantitativ
d2-R	Punktwert	quantitativ

Für einen Grossteil der Auswertungen wird auf statistische Verfahren zurückgegriffen. Diese werden hier vorgestellt, während methodische Aspekte zu den qualitativen Analysen in Kapitel 9.1.6 ausgeführt werden.

9.3.2 Überblick über die untersuchten Probandengruppen

Um die Lesbarkeit der Auswertungskapitel 10 und 11 zu erleichtern, werden hier die verwendeten statistischen Verfahren sowie die untersuchten Kindergruppen im Überblick dargestellt. Das hat ein Vorgeifen auf spätere Ausführungen zur Folge: Die untersuchten Kompetenzgruppen werden datengeleitet aufgrund der C-Test-Ergebnisse gebildet. Diese Gruppenbildung wird erst in Kapitel 10.3.2 vorgenommen und begründet. Dennoch werden die Kompetenzgruppen hier in einer Übersicht bereits dargestellt, wobei sich die Ausführungen auf eine kurze Beschreibung und insbesondere auf das Aufzählen der Gruppen beschränken. Für präzisere Angaben und entsprechende Kennwerte sei auf Kapitel 10.3 verwiesen.

In die meisten Analysen, die auf diese Kompetenzgruppen zurückgreifen, gehen zwei Kompetenzgruppen ein, die als Gruppe BC und DE bezeichnet wer-

Tab. 9.9: Kompetenzgruppen, Anzahl Probanden und durchschnittlicher Wert im C-Test.

	<i>N</i>	<i>M</i>	<i>SD</i>
Gruppe A	5	28.40	3.36
Gruppe B	15	41.47	3.93
Gruppe C	28	57.36	3.96
<i>Gruppen BC</i>	<i>43</i>	<i>51.81</i>	<i>8.59</i>
Gruppe D	28	69.93	2.43
Gruppe E	32	82.22	4.23
<i>Gruppen DE</i>	<i>60</i>	<i>76.48</i>	<i>7.10</i>
Gruppe F	7	94.14	1.77

Anm.: *N* = Anzahl Probanden; *M* = Mittelwert; *SD* = Standardabweichung

den. In Bezug auf einzelne Fragen werden die vier Gruppen B, C, D und E einzeln ausgewertet (vgl. etwa die Analysen zum Salvenumfang in Kap. 11.1.3). In einem Fall, für die Auswertungen zu den Referenzbereichen (vgl. Kap. 11.2.3), ist eine Neugruppierung nötig. Da das nur eine einzelne Auswertung betrifft, wird diese im entsprechenden Kapitel begründet. Tabelle 9.9 führt die Gruppen mit der jeweiligen Anzahl Probanden, dem Mittelwert im C-Test und der Standardabweichung auf.

9.3.3 Übersicht über die statistischen Verfahren

Für die Datenaufbereitung und -auswertung wurden *inputlog*, das Transkriptionssystem *f4*, die Statistiksoftware *SPSS 21/22* und *MaxQda*, ein Programm zur qualitativen Datenanalyse, verwendet. Die Effektgrößen wurden mit dem Programm *G-Power* berechnet. In der vorliegenden Studie werden Text, Schreibprozess und Sprachwissen in Bezug auf zwei Dimensionen untersucht: Eine davon ist die globale Sprachkompetenz. Sie steht im Zentrum des Interesses. Die zweite Dimension ist die Altersdimension (vgl. dazu die Konkretisierung der Forschungsfragen in Kap. 7.1, speziell Abb. 7.4 in Unterkap. 7.2.4). Die beiden Dimensionen unterscheiden sich im Skalenniveau: Bei der Altersgruppe handelt es sich um ordinale Daten, während die Ergebnisse zur globalen Sprachkompetenz intervallskaliert sind. Aus diesem Grund wird für die statistischen Analysen auf unterschiedliche Verfahren zurückgegriffen: Für die Vergleiche zwischen den Altersgruppen werden T-Tests durchgeführt. Es handelt sich bei allen durchgeführten T-Tests um *T-Tests für unabhängige Stichproben*.

Die notwendigen Eigenschaften für die Durchführung eines T-Tests sind ein Intervallskalenniveau und eine Normalverteilung. Bortz (1993: 91) geht für

praktische Zwecke allerdings davon aus, dass die Verteilung der Mittelwerte für beliebige Verteilungsformen des Merkmals in der Population schon bei einem Stichprobenumfang von $N \geq 30$ hinreichend normalverteilt ist. Zudem wird die Normalverteilungsannahme für den T-Test häufig als nicht allzu kritisch angesehen, da der Test relativ robust auf Verletzungen der Annahme reagiert (Brosius 2013: 479). Ein Levene-Test auf Homogenität der Varianzen geht der Durchführung eines T-Tests jeweils voraus. Dieser Schritt wird im Untersuchungsbericht für die einzelnen Durchführungen des Tests nicht eigens vermerkt. Angegeben wird jeweils die Effektgrösse d . Sie wird durch das Programm *G-Power* berechnet. Cohen (1988) bezeichnet bei T-Tests für unabhängige Stichproben Effektgrössen von $d = .20$ als klein, Effektgrössen von $d = .50$ als mittel und Effektgrössen von $d = .80$ als gross.

T-Tests ermöglichen in SPSS keine Mitberücksichtigung einer Kontrollvariable. Für die Berechnungen, die das voraussetzen, wird deswegen auf ein allgemeines lineares Modell (univariat) zurückgegriffen. Es ermöglicht das Kontrollieren einer Variable, indem diese als Kovariate aufgenommen wird. Wie für die ANOVAS werden pro Kennwert (F) zwei Freiheitsgrade, das Signifikanzniveau und die Effektstärke f angegeben. Nach Cohen (1988) wird bei Werten von $f = .10$ von einem schwachen Effekt, bei $f = .25$ von einem mittleren Effekt und bei $f = .40$ von einem starken Effekt gesprochen.

Wenn zwei intervallskalierte Variablen auf ihren Zusammenhang hin untersucht werden, wird jeweils der Korrelationskoeffizient r nach Pearson berechnet. Der Korrelationskoeffizient nach Pearson gibt gleichzeitig die Effektgrösse an. Nach Cohen (1992: 157) gelten Werte über $r = .30$ als mittlere und Werte über $r = .50$ als grosse Effektstärke. Teilweise wird partialkorreliert, d. h. dass die Korrelation zweier Variablen berechnet wird, während eine dritte Variable statistisch kontrolliert wird. Mit Bortz & Döring (2006: 740) wird von folgenden Signifikanzniveaus ausgegangen: p -Wert $\leq 5\%$: signifikant, p -Wert $\leq 1\%$: sehr signifikant, p -Wert $\leq 0.1\%$: hoch signifikant.

Für die Berechnung der Korrelation zwischen einer intervallskalierten und einer ordinalskalierten Variable (beispielsweise dem Selbstkonzept) werden Rangkorrelationen berechnet. Die Daten werden dafür in Rängen ausgedrückt und mit *Spearman's Rangkorrelationskoeffizient* (*Spearman's Rho*) angegeben. Koeffizienten ab .1 werden als kleiner, ab .3 als mittlerer und ab .5 als grosser Effekt aufgefasst (Cohen 1988).

Für alle Mittelwertsvergleiche über mehr als zwei Gruppen hinweg werden einfaktorielle Varianzanalysen (ANOVA) durchgeführt. Die Effektstärke f wird mit dem Programm *G-Power* berechnet. Ein signifikantes Ergebnis durch die ANOVA sagt noch nichts darüber aus, welche der untersuchten Gruppen

sich voneinander unterscheiden. Dafür wurden Post-Hoc-Mehrfachvergleiche (Bonferroni⁵⁸) eingesetzt.

Die datengestützte Validitätsanalyse des C-Tests aufgrund der standardisierten Sprachstandsmessung in den Teilbereichen *Wortschatz*, *Grammatik* und *Orthographie* greift auf eine multiple hierarchische Regressionsanalyse zurück. Durch das Verfahren werden diejenigen unabhängigen Variablen (Prädiktoren) extrahiert, die einen statistisch bedeutsamen Aufklärungswert für die Varianz der abhängigen Variable haben und diese am besten vorhersagen können. Solche Regressionen werden auch durchgeführt, wenn sich für eine Variable sowohl eine Abhängigkeit vom Alter als auch von der globalen Sprachkompetenz zeigt. Die Regression zeigt in solchen Fällen, wie das Verhältnis der Prädiktoren zueinander ist respektive ob und wieviel Varianz sie jeweils aufklären, wenn sie gemeinsam betrachtet werden. Für alle Regressionen wird die Methode *schrittweise* in SPSS ausgewählt, die die unabhängigen Variablen nacheinander in das Modell aufnimmt. Prädiktoren werden bei diesem Verfahren einzeln integriert oder ausgeschlossen, wenn ein Prädiktor nach der Integration eines anderen Prädiktors nicht mehr signifikant ist. So kann gezeigt werden, welches Modell am meisten Varianz aufklärt. Die Anpassungsgüte des Modells wird durch das Bestimmtheitsmass R^2 angegeben, das durch die Multiplikation mit 100 ausweist, wie viel Prozent der Varianz der abhängigen Variable durch das entsprechende Modell aufgeklärt wird. Das R^2 (auch Fit genannt) kann durch Hinzufügen zusätzlicher erklärender Variablen grösser werden oder gleich gross bleiben, allerdings kann es sich niemals verkleinern. So kann der Fit der Gleichung durch die Integration einer Vielzahl von Variablen theoretisch künstlich erhöht werden. Aus diesem Grund wird neben dem R^2 häufig das *korrigierte R^2* betrachtet. Es berücksichtigt zusätzlich die Anzahl der bei der Schätzung verwendeten erklärenden Variablen (Brosius 2013: 541–567).

⁵⁸ Für die Mehrfachvergleiche stehen verschiedene Verfahren zur Verfügung, die von der Gleichheit/Ungleichheit der Gruppengrößen und der Varianzen abhängen. Bonferroni ist ein konservativer Test, eignet sich aber für alle Varianten (Field 2013: 459–460).