UNIVERSITÄT DUISBURG ESSEN

Offen im Denken

Schriftsprachliche Kompetenzen in der Sekundarstufe II

Eine empirische Untersuchung

Dissertation zur Erlangung des akademischen Grades

Doktor der Philosophie (Dr. phil.)

der Fakultät für Geisteswissenschaften der Universität Duisburg-Essen

vorgelegt von

ELENA THIEDE
aus Hagen

1. Gutachter/Betreuer: Prof. Dr. Rupprecht S. Baur, Universität Duisburg-Essen
2. Gutachter: Prof. Dr. Tobias Schroedler, Universität Duisburg-Essen

Tag der Disputation: 08.08.2022

Für meine Mutter

DuEPublico
Duisburg-Essen Publications online

UNIVERSITÄT DUISBURG ESSEN
Offen im Denken

ub universitäts bibliothek

Diese Dissertation wird via DuEPublico, dem Dokumenten- und Publikationsserver der Universität Duisburg-Essen, zur Verfügung gestellt und liegt auch als Print-Version vor.

DOI: 10.17185/duepublico/76594
URN: urn:nbn:de:hbz:465-20220919-083041-8

Alle Rechte vorbehalten.

Inhaltsverzeichnis

Vorwort IV

1. Einleitung 1

1.1 Problemstellung 1

1.2 Erkenntnisinteresse und methodisches Vorgehen 10

1.3 Aufbau der Arbeit 12

2. Theoretische Grundlagen 15

2.1 Zur Definition von Sprachkompetenz – eine begriffliche Eingrenzung 15

2.1.1 (Schrift-)Sprachliche Kompetenzen in SkaLa® 17

2.1.2 Sprachkompetenz im C-Test 20

2.1.3 Sprachkompetenz in der gymnasialen Oberstufe 24

2.2 Operationalisierung von Sprachkompetenz in der Oberstufe 34

2.2.1 Operationalisierung von Sprachkompetenz durch Noten 34

2.2.2 Standardisierte Sprachtests in der Oberstufe 36

2.3 Zur Erfassung von Sprachkompetenz mit dem C-Test 38

2.3.1 Vom Cloze-Test zum C-Test 38

2.3.2 Redundanz als Zentralbegriff für das Funktionsprinzip von C-Tests 45

2.3.3 Zum Zusammenhang zwischen Erwartungsgrammatik, dem Prinzip der reduzierten Redundanz und der Messung von Sprachkompetenz mit dem C-Test 50

2.3.4 Modifizierte C-Tests für fortgeschrittene Erst-, Zweit- und Fremdsprachensprecher*innen 55

2.3.5 Rechtsseitige vs. Linksseitige Tilgung 63

2.4 Gütekriterien 76

3. Hypothesen 86

3.1 Zur Validität des C-Tests 86

3.2 Zu den Hintergrundvariablen 88

4. Methode 91

4.1 Die Testinstrumente 91

4.1.1 Der LGVT 6-12 91

4.1.2 Entwicklung des C-Tests 94

4.1.2.1 Voruntersuchungen 99

4.1.2.2 Zusammenstellung des finalen C-Tests 103

4.1.3 Fragebogenkonstruktion zur Erfassung (sprachbiographischer) Hintergrundvariablen 104

4.1.3.1 Anmerkungen: Migrationshintergrund – Muttersprache – Erstsprache – Familiensprache – Mehrsprachigkeit – Bildungsgrad der Eltern 106
4.2 Stichprobenkonstruktion und Versuchsplanung 113
4.3 Datenerhebung 119
4.4 Auswertungsmethode 124
4.4.1 Operationalisierung der Auswertung 124
4.4.2 Klassische Testtheorie vs. Item-Response-Theory 126
4.4.3 Normierung und Interpretation 130
4.4.4 Verfahren zur Überprüfung der Hypothesen 134
5. Ergebnisse 137
5.1 Stichprobenbeschreibung 137
5.1.1 Schulen, Deutschnoten, Alter und Geschlecht 137
5.1.2 Migrationshintergrund und Sprache/Sprachverwendung 139
5.1.3 Bildungsgrad der Eltern und Studienwunsch 146
5.2 Empirische Überprüfung des eingesetzten C-Test-Sets (Schwierigkeit und Verteilung) 148
5.3 Normierung und Reliabilität 151
5.4 Zur Validität des C-Tests 154
5.4.1 C-Test und LGVT 6-12 155
5.4.2 C-Test und Deutschnote 157
5.5 Deskriptive und bivariate Ergebnisse 164
5.5.1 C-Test und Phase in der Oberstufe 165
5.5.2 C-Test und Geschlecht 168
5.5.3 C-Test und Sprache 172
5.5.4 C-Test und Bildungsniveau 185
5.5.5 C-Test und Schulstandorttyp 191
5.5.6 C-Test und Schulform 196
5.6 Multivariate Ergebnisse 200
6. Interpretation und Diskussion 206
7. Fazit und abschließender Ausblick 216
Literaturverzeichnis 225
Abbildungsverzeichnis 243
Tabellenverzeichnis 244
Anhang VI
1. C-Tests VI
2. Berechnung C-Test und Deutschnote IX
3. Berechnungen C-Test und Sprachgruppen XIII

4. Berechnungen C-Test und Bildungsgrad XV
5. Berechnungen C-Test und Schulstandorttyp XVIII
6. Regression XIX

Vorwort

Die vorliegende Dissertation wurde im Rahmen des *Sprachkompetenzprojekts Diagnose und Förderung* im Zentrum für Lehrerbildung der Universität Duisburg-Essen angefertigt.

An dieser Stelle möchte ich allen beteiligten Personen danken, die mich bei der Anfertigung meiner Dissertation unterstützt haben.

Mein besonderer Dank gilt Herrn Prof. Dr. Rupprecht S. Baur, der mich bereits während meines Studiums mit dem C-Test vertraut gemacht und mir die Umsetzung der gesamten Arbeit durch die fortwährende Unterstützung und die vielen förderlichen Ratschläge ermöglicht hat. Ebenso gilt mein Dank Frau Prof. Dr. Heike Roll.

Frau Dr. Anja Pitton und Herrn Dr. Dirk Scholten-Akoun möchte ich von Herzen danken für die enorme Unterstützung und die unzähligen produktiven Gespräche, die meine Gedanken bewegt und meine Bemühungen maßgeblich vorangebracht haben. Mein spezieller Dank gilt auch für das stets offene Ohr in allen Belangen, ihr Vertrauen und ihre Geduld.

Überaus dankbar bin ich Dr. Sebastian Weirich und Tom Heilmann, die meine unendlichen Fragen zu Statistik immer fachkompetent und geduldig beantwortet und mich besonders bei der analytischen Aufbereitung und Auswertung der Daten beraten haben.

Auch möchte ich Christoph Chlosta meinen Dank aussprechen, der mich auf meinem Weg mit hilfreichen Anregungen, konstruktiver Kritik und seinem realistischen Blick auf die Dinge stets aufgemuntert und begleitet hat.

Für die vielen anregenden und hilfreichen Gespräche, die Unterstützung bei der Datenerhebung, die lieben Worte und die effektiven Kaffee- und Mittagspausen bin ich meinen (ehemaligen) Arbeitskolleg*innen Anna Mashkovskaya, Maurice Lorenz und Andreas Klemmer als auch allen Hilfskräften des Ressorts zu Dank verpflichtet. Ohne ihre Unterstützung in Wort und Tat wäre diese Arbeit nicht möglich gewesen.

Ganz besonders möchte ich mich bei meiner Familie, meinem Partner und bei meinen Freunden für ihre Geduld, Rücksichtnahme und für die unaufhörliche Unterstützung in

jeglicher Hinsicht bedanken. Dieser Beistand war mir immer sehr wichtig, vor allem im Verlauf dieser Dissertation.

Zu guter letzte möchte ich mich bei allen Lehrkräften und Schüler*innen bedanken, die sich bereit erklärt haben, an der hier vorgestellten Studie teilzunehmen. Ohne ihren Einsatz hätte diese Arbeit nicht entstehen können.

1. Einleitung

1.1 Problemstellung

Sprachliches Handeln gehört zu den zentralen Kernkompetenzen in vielen Bereichen des Lebens: Über ausgebaute Lese- und Schreibfertigkeiten zu verfügen befähigt einerseits zur aktiven gesellschaftlichen Teilhabe, andererseits gelten sie als notwendige Voraussetzung für die erfolgreiche Bewältigung von Schule, Berufsausbildung und Studium. Damit kommt der Sprachkompetenz besonders in schulischen Lehr-Lernprozessen eine Schlüsselfunktion zu und auch in der Sprachdiagnostik nimmt der Begriff einen zentralen Stellenwert ein: Da sprachliche Kompetenzen einen wichtigen Bestandteil von Bildung ausmachen, erfahren sie in empirischen Studien zu Kompetenzen und Leistungen von Schüler*innen starke Beachtung (Jude, 2008, S. 7) – allerdings mit Einschränkungen.

Ein Blick auf die Forschungslandschaft zur Sprachkompetenzdiagnostik im schulischen Bereich zeigt, dass Untersuchungen von sprachlichen Fähigkeiten und Fertigkeiten von Kindern und Jugendlichen mit dem Ziel der Förderung bzw. Aufdeckung von Förderbedarfen vor allem im Bereich der vorschulischen Erziehungsphase (z. B. Delfin 4 oder HAVAS 5), in den ersten Jahren der schulischen Erziehung (z. B. ELFE 1-6, PIRLS, IGLU, VERA 3) und im Bereich der Sekundarstufe I (z. B. VERA 8, PISA) mit einer Vielzahl von Instrumenten vorgenommen werden (vgl. zu einem Überblick Schnieders & Komor, 2005). In dieser Entwicklungsphase hält man offenbar eine Förderung auf der Grundlage von Test-Diagnosen für sinnvoll und erwartet Auswirkungen auf die gesamte schulische Entwicklung. Doch spätestens mit Abschluss der sprachlichen Instruktion am Ende der Sekundarstufe I wird anscheinend davon ausgegangen, dass die basalen schriftsprachlichen Kompetenzen durchgängig erworben sind (Feilke, 2010, 2013; Feilke et al., 2013, S. 18; Petersen, 2014b; Sitta, 2005, S. 157): Schriftsprachliche Kompetenzen sowie der sprachliche Lernstand der Schüler*innen werden „im stark literaturlastigen Deutschunterricht“ (Willenberg, 2001, S. 88) in der Sekundarstufe II in der Regel nur noch im Rahmen der regulären schriftlichen Aufgaben und Klassenarbeiten im Fach

Deutsch oder auch in anderen Fächern als Teilkompetenz der fachlich-inhaltlichen Darstellungen erfasst und nicht mehr in spezifischen Tests systematisch erhoben (Becker-Mrotzek, 2017, S. 6). Dabei wird die Bedeutung der sprachlichen Richtigkeit für die Leistungsbewertung zunehmend geringer (vgl. z. B. EPA 2002 für Abiturvorschläge und Kapitel 2.2).

Ein Blick in die aktuelle Forschungsliteratur zeigt, dass weder die systematische Überprüfung noch die Förderung von (basalen schrift-)sprachlichen Kompetenzen im Laufe der gymnasialen Oberstufe eine bedeutende Rolle spielen (Petersen, 2014c, S. 1). So sind Ergebnisse, wie beispielsweise die von Riebling (2013), die durch eine Befragung von Hamburger Lehrkräften zu dem Schluss kommt, dass sich Maßnahmen und Angebote zur Sprachbildung hauptsächlich auf die Erprobungsstufe fokussieren, mit den fortschreitenden Klassenstufen kontinuierlich abnehmen und in der Sekundarstufe II fast gar nicht mehr vorkommen, kaum verwunderlich oder gar überraschend (Petersen, 2014c, S. 3; Riebling, 2013, 136 f.).

Die Annahme, dass der Erwerb und die Entwicklung der basalen schriftsprachlichen Kompetenz mit Eintritt in die Oberstufe abgeschlossen ist und dass es in der gymnasialen Oberstufe primär um die Vermittlung eines fundierten fachlichen Orientierungswissens zur Erschließung und Reflexion von Texten und Medien „als Voraussetzung für das Erreichen der fachspezifischen wissenschaftspropädeutischen Grundanforderungen“ (MSB NRW, 2014, S. 10) gehe, dokumentieren auch die einschlägigen Untersuchungen zur Vorbereitungsqualität in der Oberstufe. Grundprobleme beim Formulieren, Beherrschung der Regeln zum sprachrichtigen schriftlichen Ausdruck usw. lassen sich in der Regel nur andeutungsweise aus den publizierten Untersuchungsergebnissen herauslesen. In einer Längsschnittstudie von Trautwein et al. (2006) zur schulischen Vorbereitungsqualität auf das Studium beispielsweise geben die Abiturient*innen am Ende der 13. Klassenstufe an, sich insgesamt nur unzureichend durch die gymnasiale Oberstufe auf die Anforderungen im Studium vorbereitet zu fühlen (Trautwein et al., 2006, S. 63). Dabei wird auch die Vorbereitung in Hinblick auf die Bewältigung sprachlicher Aufgaben, wie „Fragestellungen wissenschaftlich bearbeiten“, „Texte systematisch analysieren“ und „in Diskussionen schlüssig argumentieren“ seitens der Oberstufenschüler*innen bemängelt (Trautwein et al., 2006, S. 57). Zwei Jahre nach

dem ersten Befragungszeitpunkt, d. h. nach Aufnahme des Studiums und ersten Erfahrungen an der Universität, fällt die Beurteilung der Vorbereitungsqualität durch die Oberstufe sogar noch ungünstiger aus als zum ersten Zeitpunkt (ebd.). Die Ergebnisse relativieren Trautwein et al. (2006) und weisen darauf hin, dass sich ihre Befunde auf subjektive Einschätzungen von Schüler*innen bzw. Studierenden stützen. Tatsächlich lässt sich auf der Grundlage einer Befragung nicht klären, ob die als defizitär wahrgenommene Vorbereitung in der Oberstufe tatsächlich unzureichend ist, wie sich die angedeuteten sprachlichen Unzulänglichkeiten konkreter beschreiben ließen oder ob die Anforderungen zu Studienbeginn als (unerwartet) hoch empfunden werden und auch tatsächlich mit den schulisch erworbenen Kompetenzen nicht problemlos bewältigt werden können (ebd., S. 63). Einen Impuls zur Aufklärung auf dem Feld der basalen sprachlichen Defizite formulieren Trautwein et al. allerdings nicht. Die subjektiv empfundene Diskrepanz in den laut Kernlehrplan vermittelten „wissenschaftspropädeutischen Grundanforderungen“ (MSB NRW, 2014) zwischen der gymnasialen Oberstufe und den Anforderungen der Hochschule deuten allenfalls ganz allgemein auf eine „mangelnde curriculare Verzahnung der beiden Bildungsinstitutionen“[1] (Petersen, 2014b, 2014c, S. 2; Trautwein et al., 2006).

Den Befund einer eher mäßigen Vorbereitungsqualität der gymnasialen Ausbildung auf ein Studium aus Sicht der Schüler*innen bzw. Studierenden sehen die Autor*innen auch in Ergebnissen aus ähnlichen themenbezogenen Studien bestätigt (Trautwein et al., 2006, S. 63), die aber ebenfalls die basalen sprachlichen Kompetenzen nicht oder nur marginal in den Blick nehmen. So zeigen Untersuchungen des Hochschul-Informations-Systems (HIS) bereits zu Beginn der 1990er Jahre, dass 34 % der befragten Studierenden durch die Schule „nur in ausreichendem Maße mit dem im Studium geforderten Arbeitstechniken – richtiges Zitieren, Protokoll einer Diskussion anfertigen, Experimente planen etc. – vertraut gemacht“ worden sind (Durrer & Heine, 1995, S. 11) und auch Huber et al.; Huber (2009) kommen zu dem Schluss, dass es den Studierenden den

[1] Welche expliziten „wissenschaftspropädeutischen Grundanforderungen“ (Kernlehrplan Deutsch Sek II, 2013, S. 10) und zu welchem Ausmaß diese im Rahmen der Oberstufe überhaupt vermittelt und erreicht werden sollen (sollten), wird im Kernlehrplan für die Oberstufe (KMK GoST für das Fach Deutsch NRW) nicht weiter erläutert. Damit bleibt auch die Frage, ob nun also die gymnasiale Oberstufe oder die Universitäten selbst nicht hinreichend auf das wissenschaftliche Arbeiten/Schreiben im Studium vorbereitet an dieser Stelle offen.

eigenen Einschätzungen zufolge mehr an wissenschaftlichen Arbeitstechniken (u. a. Verfassen eines wissenschaftlichen Textes, Argumentation wissenschaftlicher Ergebnisse) fehlt als an Stoffkenntnissen (Huber et al., 2009, S. 117).

Ähnliche Ergebnisse zeigt auch der Gesamtbericht zur Exmatrikuliertenbefragung von 2015 der Universität Konstanz, in der jährlich (ehemalige)[2] Bachelorstudierende nach Gelingens- und Bedingungsfaktoren zum Studienerfolg sowie zu Gründen zum Studienabbruch befragt werden. Neben der wahrgenommenen Doppelbelastung von Studium und Berufstätigkeit während des Studiums sind es vor allem fachliche und sprachliche Faktoren und Anforderungen, die von den Studierenden als erschwerend bzw. belastend eingestuft werden: 31 % der Absolvent*innen und 23 % der Abbrecher*innen geben an, dass sie nicht ausreichend mit Präsentationstechniken und Rhetorik vertraut sind und mehr als der Hälfte der Absolvent*innen (52 %) und 45 % der Abbrecher*innen fehlen nach eigenen Angaben Kenntnisse und Techniken zum Verfassen schriftlicher, akademischer Arbeiten (*Exmatrikuliertenbefragung 2010 Bis 2014*, 2015, 22 ff.; Werner, 2019). Auf welcher Ebene aber diese Kenntnisse und Techniken zum Verfassen schriftlicher, akademischer Arbeiten als unzureichend wahrgenommen werden, bleibt jedoch unklar. Fix und Dittmann (2008) kommen in ihrer Studie zu Exzerpierkenntnissen von Oberstufenschüler*innen zu dem Schluss, dass die Fähigkeiten und Kenntnisse der untersuchten Schüler*innen in diesem Bereich für die Anforderungen der Universität nicht ausreichend sind (Fix & Dittmann, 2008, S. 57) und deshalb den Studienerfolg insgesamt bedrohen.

Die Zusammenstellung dieser Ergebnisse verdeutlicht nun vor allem, dass die Oberstufenschüler*innen und Studierende der ersten Semester die Propädeutik hinsichtlich wissenschaftlicher Arbeitstechniken und wissenschaftssprachlicher Anforderungen (Rhetorik und Argumentationstechnik) für unzureichend halten. Die

[2] In den Jahren der Untersuchung (2010-2014) brechen den Befragungsdaten der Universität Konstanz zufolge 17 % der Bachelorstudierenden ihr Studium ab. Die Autoren weisen in diesem Zusammenhang jedoch auf eine Untersuchung von Heubelein (2014) hin, in der die Studierendenabbruchquote auf der Grundlage von statistischen Daten ermittelt wurde und bei 28 % für Bachelorstudierenden liegt. Den Unterschied zwischen den Daten von Heubelein (2014) und den an der Universität Konstanz (2015) erhobenen erklären die Autoren mit einer möglichen Verzerrung beim Rücklauf der Befragung (vgl. Heubelein, Ulrich u. a.: Die Entwicklung der Studienabbruchquoten an den deutschen Hochschulen. Statistische Berechnungen auf der Basis des Absolventenjahrgangs 2012; Stabsstelle Qualitätsmanagement: Gesamtbericht Exmatrikulationsbefragung 2010 bis 2014 (2015), S. 9 ff.).

Untersuchungen bestätigen aber andererseits auch durch Schweigen oder nur implizites, unscharfes Erfassen die oben von u. a. Feilke kolportierte Annahme, dass die basalen schriftsprachlichen Kompetenzen zu Beginn der Sekundarstufe II als bereits vollständig ausgebaute Fähigkeiten und Fertigkeiten vorlägen (Feilke, 2010; Feilke et al., 2013) und dass auf diesen aufbauend in der Oberstufe nun lückenlos mit der Vermittlung von Kenntnissen und Techniken zum Verfassen schriftlicher, wissenschaftlicher Arbeiten angeknüpft werden könne. In den Befragungen wird eben fast ausschließlich von der Vorbereitungsgüte in Bezug auf das wissenschaftlich argumentierende Formulieren oder das Verfassen von akademischen Arbeiten berichtet, nicht aber von etwa basalen Mängeln in der schriftsprachlichen Kompetenz (z. B. in den Bereichen, Grammatik, Orthografie, Interpunktion, sinnentnehmendes Lesen, usw.). Die Gründe für die auftretenden Probleme bei der Bewältigung der ersten Hürden in der akademischen Ausbildung werden so – wie in den oben genannten Studien gezeigt – beim Verfassen wissenschaftlicher Arbeiten verortet, nicht aber auf der gleichsam darunter liegenden Ebene, d. h. im defizitären Ausbau basaler sprachlicher Kompetenzen, vermutet. Dabei gibt es durchaus Indizien dafür, dass die Annahme, auf die basalen schriftsprachlichen Kompetenzen könne in der Oberstufe fraglos zurückgegriffen werden, weshalb auf standardisierte oder schul- und stufenübergreifende Sprachstandsfeststellungen in der Sekundarstufe II an den Schulen folglich verzichtet werden könne, eine kritische Hinterfragung lohnte. So meldete beispielsweise das Bundeskriminalamt im Jahr 2016, dass es von 120 Stellen nur 62 besetzen konnte, da zu viele Bewerber*innen trotz Abiturs im Deutschtest, der grundlegende Kenntnisse in den Bereichen Rechtschreibung, Grammatik, Wortschatz abfragt, durchgefallen seien (Flade, 2016).

Und auch vor dem Hintergrund empirischer Untersuchungen (z. B. DESI 2003/2004) bleibt fraglich, ob das, was zumindest formal in den Kernlehrplänen für das Ende der Sekundarstufe I und die Oberstufe vorausgesetzt wird – nämlich, dass die Schüler*innen am Ende der Klasse 9 „strukturiert, verständlich, sprachlich variabel und stilistisch stimmig“ (MSB NRW, 2007, S. 16) schreiben sowie die „Grundregeln der Rechtschreibung und Zeichensetzung sicher beherrschen und häufig vorkommende Wörter, Fachbegriffe und Fremdwörter richtig schreiben“ (MSB NRW, 2007, S. 20) und

am Ende der Sekundarstufe II „selbstständig die sprachliche Darstellung in Texten mithilfe von Kriterien (u. a. stilistische Angemessenheit, Verständlichkeit, syntaktische und semantische Variationsbreite) beurteilen und überarbeiten“ sowie „die normgerechte Verwendung der Sprache (Rechtschreibung, Grammatik und Zeichensetzung) in Texten prüfen und diese überarbeiten“ (MSB NRW, 2014, 33 f.) können – überhaupt vorausgesetzt werden kann. So zeigen beispielsweise Ergebnisse aus der DESI-Studie (2003/2004), in welcher Daten zu den schriftsprachlichen Kompetenzen im Deutschen und Englischen von über 10.500 Schüler*innen der Jahrgangsstufe 9 aller Schulformen erhoben und ausgewertet wurden, dass ein nicht geringer Teil der Jugendlichen in der Klasse 9 über nicht hinreichend ausgebaute basale schriftsprachliche Kompetenzen verfügt: Im Rahmen der Untersuchung wurden dabei u. a. zu Beginn und zum Ende der Jahrgangsstufe 9 die Texte der teilnehmenden Schüler*innen in Bezug auf semantisch-pragmatische Teilkompetenzen – d. h. hinsichtlich Wortwahl, Textkonstruktion und Sprachstil – und sprach-systematischen Teilkompetenzen – d. h. bezogen auf Satzkonstruktion, Orthografie und Interpunktion – untersucht und anhand normierter Kompetenzskalen bewertet (Beck & Klieme, 2007; DESI-Konsortium, 2006b, 2008; Hartig et al., 2008, S. 41) mit dem Ergebnis, dass knapp 30 % der Schüler*innen am Ende der Klasse 9 im semantisch-pragmatischen Bereich nicht über einen „nach curricularen Erwartungen unzureichenden Kompetenzstand“ hinaus kommen, was bedeutet, dass die Texte dieser Schüler*innen sich vor allem an umgangssprachlichen Sprachnormen orientieren, logisch fehlerhaft aufgebaut sind und wesentliche, inhaltliche Textelemente nicht beinhalten (Klieme, 2006, S. 7). In Bezug auf die Sprachsystematik weist dabei mehr als ein Drittel der Schüler*innen (35 %) erhebliche Mängel in den Bereichen Orthographie, Interpunktion und Syntax auf, die eine zielgerichtete, kommunikative Übermittlung von Informationen so gut wie unmöglich machen (DESI-Konsortium, 2006b, S. 7; Neumann & Lehmann, Rainer, H., 2008, 89 f.). Diese Befunde deuten nun also eine völlig andere Ausgangslage zu Beginn des Unterrichts in der Sekundarstufe II an.

Vor dem Hintergrund dieser Annahme, dass die basalen schriftsprachlichen Kompetenzen in der Oberstufe zumindest in Teilen eben nicht hinreichend ausgebaut sind, stellt sich vor dem Hintergrund weiterer Forschungsergebnisse darüber hinaus eine

etwas anders gelagerte Frage, nämlich, ob möglicherweise bestimmte Sprachgruppen (beispielsweise monolinguale oder mehrsprachige Schüler*innen) eher von defizitär ausgebauten basalen schriftsprachlichen Kompetenzen betroffen sind. So kommt Neumann (2014; 2010) beispielsweise in einer Re-Analyse der DESI-Daten zu dem Schluss, dass sich besonders Leistungsunterschiede zwischen „deutschsprachig allochthonen und nicht deutschsprachigen" Jugendlichen zeigen, die „im Bereich der pragmatischen Teilkompetenzen vor allem die Wortwahl und den Textaufbau, weniger die inhaltlichen Aspekte und die stilistischen Ausführungen" und im Bereich der sprach-systematischen Teilkompetenz am häufigsten die Wortgrammatik, gefolgt von der Satzkonstruktion und der Orthographie, betreffen (Neumann, 2014, S. 182; Neumann & Domenech, 2010, S. 18). Dabei schneiden die nicht-deutschsprachigen Schüler*innen besonders in den Anforderungsbereichen, in denen basale lexikalisch-semantische, syntaktische und orthographische schriftsprachliche Fähigkeiten und Fertigkeiten gefragt sind, schlechter ab als ihre ausschließlich deutschsprachigen Mitschüler*innen. Interessanterweise verändern sich diese Leistungsunterschiede im Verlauf der Untersuchung dahingehend, dass sie im sprach-pragmatischen Bereich weniger, im sprach-systematischen dagegen größer werden, ohne dass die Differenzen jedoch überwunden werden können (Neumann, 2014, S. 182). Damit bestätigen sich auch generelle Ergebnisse aus der DESI-Studie und beispielsweise aus den Untersuchungen von Cantone-Altıntaş und Haberzettl (2009, 43 ff.), in denen u. a. gezeigt wurde, dass Schüler*innen mit Deutsch als Zweitsprache in der Sekundarstufe I vor allem im Bereich des Wortschatzes größeren Förderbedarf aufweisen (Klieme, 2006, S. 4).

In ihrer Untersuchung zur Schreibfähigkeit von Schüler*innen mit Deutsch als Zweitsprache in der Sekundarstufe II kommt auch Petersen (Petersen, 2014a, 2014b; 2014c) zu dem Schluss, dass mehrsprachige Oberstufenschüler*innen zwar nicht weniger komplexe syntaktische Strukturen[3] in ihren Texten produzieren, dass sich aber insgesamt vor allem bei den Texten der mehrsprachigen Elftklässler*innen „viele Belege für Unsicherheiten bei der Wortbildung, beim Gebrauch fester Wendungen, bei der Wahl eines angemessenen Registers und bei der Konstruktion komplexer Äußerungen

[3] Zur Bestimmung der syntaktischen Komplexität der Texte wurden die Textlänge, die Satzlänge, der Gebrauch von Nebensätzen und komplexer Nominalphrasen erfasst (Petersen (2014c, S. 6).

finden“ (Petersen, 2014c, S. 11). Auch hier zeichnen sich also Defizite im Bereich basaler schriftsprachlicher Kompetenzen – vor allem in Bezug auf lexikalisch-semantische Fähigkeiten und Fertigkeiten – ab, die in Form von Formulierungsschwierigkeiten sichtbar werden[4] (Petersen, 2014a, S. 249). Texte, die nach Einschätzung der Autorin einen sprachlichen Förderbedarf im Bereich der komplexen Syntax nahelegen, sind dabei sowohl in der Gruppe der monolingual deutschsprachigen als auch in der Gruppe der mehrsprachigen Schüler*innen zu finden, wobei offen bleibt, wie viele der untersuchten Texte in diese Kategorie fallen[5] (Petersen, 2014a, S. 253).

Dass auch Studierende (zumindest in Teilen) nicht über vollständig ausgebaute basale sprachliche Kompetenzen verfügen, dass also am vollständigen Erfolg der sprachlichen Basisausbildung durch die Schule und die Sekundarstufe II gezweifelt werden muss, darauf deuten die regelmäßigen Beschwerden von Hochschuldozent*innen über eklatante sprachliche Schwächen der Studierenden bei der Formulierung von Texten hin, besonders in den Bereichen Orthografie, Interpunktion, Lexik und Grammatik (Bethke, 2014; Scholten-Akoun, 2016a, 24 f.). So kommen Bremerich-Vos und Scholten-Akoun (2016) durch eine Befragung von Dozierenden an der Universität Duisburg-Essen zu dem Ergebnis, dass die Lehrenden die schriftsprachlichen Kompetenzen ihrer Studierenden tendenziell negativ einschätzen (Scholten-Akoun, 2016a, S. 24), was kürzlich in einem Interview mit Dozierenden[6] der Universität Duisburg-Essen bestätigt werden konnte: Man sei alarmiert über die Tatsache, dass Texte von Studierenden[7] eine Vielzahl von Zeichensetzungsfehlern sowie lexikalischen und grammatikalischen Fehlern aufweisen und Studierende die sprachlichen Standards, die am Ende der Sekundarstufe I – und eigentlich schon nach der 6. und 7. – Klasse vorhanden sein müssten, offensichtlich nicht erfüllen (Frohn, 2018). Damit wird die in der einschlägigen Literatur bereits zu Beginn des neuen Jahrtausends erhobene Klage über große Probleme der

[4] Petersen weist an dieser Stelle explizit darauf hin, dass die Frage, inwieweit im Bereich der Lexik tatsächlich ein Förderbedarf für mehrsprachige Schüler*innen besteht, auf Grundlage einer größeren Datenbasis differenzierter untersucht werden muss (Petersen 2014c, S. 250)

[5] In der Studie von Petersen geht es nicht um die Aufdeckung von sprachlichen Förderbedarfen im Allgemeinen, sondern um den Vergleich von Schüler*innengruppen (bezgl. Sprachverwendung und Alter) in Bezug auf syntaktische Schreibfähigkeiten und schriftliche Argumentationskompetenz.

[6] Interviewt wurden Dr. Patrik Voßkamp und Dr. Ulrike Behrens, Dozierende an der Universität Duisburg-Essen am Institut für Germanistik in den Bereichen Linguistik und Sprachdidaktik.

[7] Gemeint sind hier Studierende im Lehramts-Masterstudiengang Germanistik

Studierenden beim Schreiben einmal mehr wiederholt (Ehlich et al., 2000; Furchner et al., 1999, 61 ff.).

Diesen subjektiven Eindrücken verleiht eine Untersuchung an der Universität Duisburg-Essen (UDE) und zwei weiteren Universitäten in NRW eine gefestigtere empirische Evidenz: Im Rahmen der Studie wurde im Zentrum für Lehrerbildung der UDE ein Diagnoseinstrument zur Messung der Schreib- und Lesekompetenz entwickelt und zur Erfassung der sprachlichen Kompetenzen einer beinahe vollständigen Kohorte aller Studienanfänger*innen in den Lehramtsstudiengängen eingesetzt. Als Diagnoseinstrument wird ein zweistufiges Assessment eingesetzt, dem sich zu stellen seit 2015 die Einschreibeordnung der UDE verbindlich von allen Lehramtsstudienbewerber*innen verlangt: Der erste Testteil ist ein linksseitig getilgter C-Test[8], der zweite eine kombinierte Lese-Schreibaufgabe, die von den Studierenden das Lesen und Zusammenfassen eines Zeitungsartikels zu einem pädagogischen Thema und eine Stellungnahme zu einer These des Leseimpulses verlangt. Die von den Studierenden produzierten Freitexte im Umfang von zwischen 350 und knapp 1.000 Wörtern werden anhand von zwischen 40 und 50 (inhaltlichen, strukturellen, formalen und sprachlichen) Variablen von geschulten Codierer*innen ausgewertet (Martin et al., 2016, 123 ff.). In ihrer Untersuchung kommen Bremerich-Vos und Scholten-Akoun (2016) zu dem Ergebnis, dass die Mehrheit der Lehramtsstudierenden in der Studieneingangsphase wichtige Aspekte (zentrale Fragestellung, zentrale These) eines Textes nicht oder nicht korrekt wiedergeben kann und/oder erhebliche Schwierigkeiten im sprachlichen Bereich (elementare Grammatik- und Rechtschreibfehler usw.) hat (Bremerich-Vos & Weirich, 2016, 165 ff.; Mashkovskaya & Baur, 2016a, 213 ff.). Ein großer Teil der Studierenden verfügt zudem lediglich über einen limitierten Wortschatz. Insgesamt sind bei etwa einem Drittel der Studierenden, die ein Lehramtsstudium beginnen, die basalen schriftsprachlichen Fertigkeiten[9] defizitär, wobei die Gruppe der Studierenden mit Migrationshintergrund i. d. R. erkennbar schlechtere Leistungen zeigt, was gerade für diese Gruppe den Studienbeginn erschwert (Baur & Scholten-Akoun,

[8] Ausführliche Informationen zum (linksseitig getilgten) C-Test in Kapitel 2.3.
[9] Zur genaueren Definition der basalen schriftsprachlichen Kompetenzen ausführlich in Kapitel 2.

2010; Bremerich-Vos & Scholten-Akoun, 2016; Scholten-Akoun et al., 2013; Scholten-Akoun et al., 2014).

Obwohl gerade angehende Lehrer*innen als zukünftige „role models“ über mindestens ausreichende schriftsprachliche Kompetenzen verfügen können sollten, wird auf die Bedeutung dieser Kompetenzen in Beratungen an der Universität kaum bis gar nicht eingegangen: Das digitale Unterstützungsangebot *Studiport*, in dessen Rahmen Studienbewerber*innen ihr Wissen im Bereich „Sprach- und Textverständnis“ überprüfen, in Online-Kursen erwerben und/oder vertiefen können[10], bildet hier eher die Ausnahme, ebenso wie die vom Rektorat der UDE 2015 für angehende Lehramtsstudierende in die Einschreibeordnung aufgenommene verpflichtende Teilnahme am Online-Self-Assessment *SkaLa®* (**S**prach**k**ompetenz **a**ngehender **L**ehr**a**mtsstudierender), in dem die basalen schriftsprachlichen Kompetenzen durch die oben genannten Diagnoseinstrumente (Lese-/Schreibaufgabe, C-Test) überprüft werden. Mit der Teilnahme an SkaLa® bzw. über das individuelle SkaLa®-Ergebnis sollen die Studierenden zwar zu Beginn ihres Studiums auf eventuell bestehende Defizite und die Möglichkeiten zu deren Bearbeitung mit Förderangeboten der Hochschule aufmerksam gemacht werden, dennoch beginnt der Prozess des Ausbaus der basalen schriftsprachlichen Fähigkeiten an der Universität zu spät und erschwert den Einstieg an die Hochschule.

1.2 Erkenntnisinteresse und methodisches Vorgehen

Die vorangegangen Ausführungen haben deutlich gemacht, dass die Beherrschung von basalen schriftsprachlichen Kompetenzen auch in der Sekundarstufe II nicht als selbstverständlich vorausgesetzt werden kann und „es der Schule nicht in ausreichendem Maße gelingt, allen Schüler*innen die für eine erfolgreiche Bildungslaufbahn notwendigen sprachlichen Kompetenzen in der Unterrichtssprache Deutsch zu vermitteln“ (Petersen, 2014a, S. 7; Redder et al., 2011, S. 52). Ein

[10] Das Angebot von *Studiport* sowie ausführliche Informationen zum Programm sind abrufbar unter https://www.studiport.de/.

Forschungsbedarf besteht daher hinsichtlich der Frage, bis zu welchem Grad Schüler*innen der gymnasialen Oberstufe über basale schriftsprachliche Kompetenzen verfügen, d. h. in der empirischen Erhebung des Sprachstandes von Oberstufenschüler*innen, über den mögliche Förderbedarfe sichtbar gemacht werden könnten. Für Messungen von basalen sprachlichen Fertigkeiten und Fähigkeiten – etwa wie das oben beschriebene Assessment für Studienbewerber*innen des Lehramts (SkaLa®) –, die nach dem Abitur an den Hochschulen für ein erfolgreiches Studium oder eine berufliche Ausbildung notwendig wären, liegen jedoch bisher für die Sekundarstufe II kaum Instrumente vor (s. o., vgl. auch Kapitel 2.2; vgl. auch Scholten-Akoun & Baur, 2012, 307 ff.). Etwaige Defizite im Bereich basaler schriftsprachlicher Kompetenzen sollten jedoch möglichst nicht erst an der Universität oder in der beruflichen Ausbildung erkannt und behoben werden, sondern schon in den letzten Jahren der Schulausbildung. Bedingt durch die Tatsache, dass auf Testformate zur schnellen und auch diagnostische Informationen bereitstellenden Erfassung von schriftsprachlichen Kompetenzen der anvisierten Zielgruppe für die hier vorliegende Studie nicht unmittelbar zurückgegriffen werden konnte, ist es u. a. Ziel dieser Arbeit, ein Messinstrument zur systematischen Erfassung basaler schriftsprachlicher Kompetenzen von Oberstufenschüler*innen zu entwickeln. Ein Instrument, welches sich zur Messung des Sprachstandes im schulischen und akademischen Bereich – und hier im Besonderen, wie oben erwähnt, im Rahmen von SkaLa® zur Erfassung von Sprachkompetenz fortgeschrittener Erst- und Zweitsprachsprecher*innen des Deutschen – bereits als valide, reliabel und ökonomisch erwiesen hat, ist der C-Test, der ausführlich in Kapitel 2.3 vorgestellt wird. Im Rahmen dieser Arbeit wird also zunächst ein für die hier im Forschungsinteresse stehende Kohorte modifizierter C-Test konzipiert, erprobt und evaluiert, um anschließend mit diesem den Sprachstand in der gymnasialen Oberstufe systematisch zu erheben. Die Erhebung, Auswertung und Interpretation der Daten liefert eine gesicherte empirische Evidenz und ermöglicht es so, Erkenntnisse über die Beherrschung der schriftsprachlichen Kompetenzen von Oberstufenschüler*innen zu gewinnen und mögliche Förderbedarfe aufzudecken.

In einigen der oben genannten Studien (s. o.) wurde ebenfalls deutlich, dass bestimmte Gruppen (beispielsweise in Bezug auf die Sprachverwendung) im Vergleich häufiger

weniger gute Ergebnisse in Bezug auf Sprachstandsmessung im Deutschen erzielen als andere (vgl. z. B. Baur & Scholten-Akoun, 2010; Bremerich-Vos & Scholten-Akoun, 2016; Scholten-Akoun et al., 2013; Scholten-Akoun et al., 2014; DESI-Konsortium, 2006b; Klieme 2007; Neumann, 2014). Aus diesen Ergebnissen ergibt sich des Weiteren die Forschungsfrage, ob sich auch ausweislich der im Rahmen dieser Arbeit erhobenen C-Test-Leistungen gruppenspezifische Unterschiede in der untersuchten Kohorte abzeichnen – und hier vor allem in Bezug auf Aspekte von Mehrsprachigkeit, sozioökonomischem Status und Standorttyp der Schule (ausführlich dazu Kapitel 3). Welchen Einfluss also verschiedene Hintergrundvariablen auf die hier gemessenen schriftsprachlichen Kompetenzen der Oberstufenschüler*innen haben, wird abschließend im letzten methodischen Teil dieser Arbeit durch Korrelationsstudien untersucht. Denn, um angesichts einer zunehmend heterogenen Schülerschaft in der Sekundarstufe II Chancengleichheit im Allgemeinen und einen erfolgreichen Start in der akademischen und beruflichen Ausbildung im Besonderen zu ermöglichen, sollte eine systematische Diagnose der basalen schriftsprachlichen Fähigkeiten und Fertigkeiten – ergänzend zu der allgemeinen Rückmeldung zur sprachlichen Kompetenz, die an der Schule in der Regel im Rahmen von Klausurkorrekturen erfolgt – als Voraussetzung für eine zielgenaue sprachliche Förderung erfolgen. Die so erhobenen Ergebnisse könnten weiterhin Aufschluss darüber geben, ob bzw. inwieweit beispielsweise Schulen in sozialräumlich deprivierter Lage oder bestimmte Gruppen von Schüler*innen (beispielsweise durch die Förderung des Fachs Deutsch als Zweitsprache) und/oder eine generelle systematische Sprachförderung zur Sicherung der Bildungsgerechtigkeit bedürfen.

1.3 Aufbau der Arbeit

Die Arbeit gliedert sich in folgende Abschnitte: Nach der Einleitung wird im theoretischen Teil zunächst der Begriff der Sprachkompetenz definiert bzw. für die vorliegende Arbeit begrifflich eingegrenzt. Dabei wird aus drei verschieden Perspektiven beleuchtet, was unter dem Begriff der Sprachkompetenz im Sprachkompetenzassessment für Studienbewerber*innen an der Universität Duisburg-

Essen (SkaLa®), welches die hier vorgestellte Studie maßgeblich motiviert hat (s. o.), in Bezug auf den C-Test, mit dem in dieser Arbeit schriftsprachliche Kompetenzen gemessen werden und in Bezug auf die gymnasiale Oberstufe, deren Schüler*innen die Zielkohorte bilden, zu verstehen ist. Anschließend wird erläutert, wie sprachliche Kompetenzen in diesen drei Bereichen operationalisiert werden, wobei der C-Test – als in der vorliegenden Studie zentrales Testinstrument – einen besonderen Stellenwert einnimmt: So wird in Kapitel 2.3 ausführlich dargestellt, welche theoretischen Annahmen über Sprache bzw. Sprachkompetenz dem C-Test zugrunde liegen, wie dieser als Test funktioniert, d. h. wie mit ihm Sprachkompetenz gemessen werden kann und wie bzw. warum modifizierte C-Tests für fortgeschrittene Sprecher*innen – und hier im Besonderen durch den Wechsel vom klassischen rechts- auf ein linksseitiges Tilgungsmuster, welches in den C-Tests der vorliegenden Studie eingesetzt wurde – unterschiedliche Facetten von Sprachkompetenz fokussieren. In Kapitel 2.4 wird der C-Test in Bezug auf Gütekriterien diskutiert.

Der theoretische Teil schließt in Kapitel 3 mit der zusammenfassenden Darstellung der Forschungsfragen und den aus den theoretischen Annahmen abgeleiteten Hypothesen, die im empirischen Teil der Arbeit überprüft werden.

Der empirische Teil beginnt mit der Darstellung methodischer Aspekte der durchgeführten Untersuchung (Kapitel 4). Dazu gehört zunächst die Beschreibung, Entwicklung und Erprobung der eingesetzten Testinstrumente sowie die Konzeption des Fragebogens zur Erfassung von Hintergrunddaten (4.1). Ebenso werden die Stichprobenkonstruktion (4.2), die Datenerhebung (4.3) und die Auswertung (4.4) eingehend erläutert.

In Kapitel 5 werden schließlich die Ergebnisse der Arbeit vorgestellt. Zunächst erfolgt die Beschreibung der Stichprobe (5.1). Anschließend werden die Ergebnisse zur Überprüfung des entwickelten und eingesetzten C-Tests in Bezug auf Schwierigkeit, Verteilung (5.2), Normierung und Reliabilität (5.3) dargestellt. Zur Bestimmung der Validität des entwickelten C-Tests wurde dieser mit den Deutschnoten der Schüler*innen und einem standardisierten Lesetest korreliert (5.4). Die Zusammenhänge zwischen dem C-Test und den erfassten Hintergrundvariablen (zum Geschlecht, Alter, Sprachgebrauch, Schulbesuch usw.) werden in Kapitel 5.5 zunächst

deskriptiv und anschließend durch bivariate Korrelationsverfahren ausgewertet. In Kapitel 5.6 folgen anschließend die Darstellung und die Ergebnispräsentation der multivariaten Korrelationsanalysen.

Die empirischen Befunde werden im 6. Kapitel schließlich interpretiert und diskutiert. Den Abschluss dieser Arbeit bildet das Kapitel 7 mit einem zusammenfassenden Fazit und abschließenden Ausblick.

2. Theoretische Grundlagen

Im folgenden Kapitel soll es zunächst darum gehen, den Begriff der Sprachkompetenz definitorisch für den für diese Arbeit relevanten Kontext einzugrenzen. Dabei wird Sprachkompetenz aus drei konkreten für diese Arbeit relevanten Blickwinkeln betrachtet und eine Arbeitsdefinition formuliert. Anschließend wird auf die Frage eingegangen, wie Sprachkompetenz im Bereich der gymnasialen Oberstufe – auf den sich die vorliegende Untersuchung bezieht – operationalisiert wird, bevor der C-Test als Instrument zur Erfassung von Sprachkompetenz ausführlich vorgestellt wird.

2.1 Zur Definition von Sprachkompetenz – eine begriffliche Eingrenzung

Über angemessene mündliche und schriftliche sprachliche Fertigkeiten und Fähigkeiten zu verfügen ist eine Schlüsselkompetenz von hoher gesellschaftlicher Relevanz, der im deutschen Bildungssystem eine zentrale Rolle zukommt: Sprachkompetenz bezieht sich einerseits auf Sprachkönnen in dem Sinne, dass eine bestimmte Sprache verwendet werden kann. Andererseits kann Sprachkompetenz gleichzeitig als Voraussetzung und Instrumentarium zur Aneignung von neuem Wissen und zur Bewältigung von Schulstoff angesehen werden (Bos et al., 2003, S. 14; Jude & Klieme, 2007, S. 11). Damit kommt der Sprachkompetenz besonders in schulischen Lehr-Lernprozessen eine Schlüsselfunktion zu und auch in der Sprachdiagnostik nimmt der Begriff einen zentralen Stellenwert ein: Da sprachliche Kompetenzen einen wichtigen Bestandteil von Bildung ausmachen und zur aktiven gesellschaftlichen Teilhabe befähigen, erfahren sie in empirischen Studien zu Kompetenzen und Leistungen von Schüler*innen immer stärkere Beachtung (Jude, 2008, S. 7). So konnte anhand von Ergebnissen aus PISA (2002) sowie von Befunden des *International Adult Literarcy Survey* der OECD (2000) belegt werden, dass ein hoher Zusammenhang zwischen der schriftsprachlichen rezeptiven Kompetenz und dem Schul- bzw. Bildungserfolg junger Menschen besteht (Gogolin & Neumann, 2009; Naumann et al., 2010, S. 43).

Aufgrund seiner unterschiedlichen Verwendung in diversen wissenschaftlichen Kontexten, aber auch seiner starken alltagssprachlichen Bedeutungsvielfalt, ist es

jedoch problematisch den Sprachkompetenzbegriff allgemeingültig und wissenschaftlichen Kriterien genügend zu definieren (Chudaske, 2012, S. 37; Harting, 2008, 15 ff.). So wird der Begriff der Sprachkompetenz je nach Forschungsrichtung im Spannungsfeld zwischen Linguistik, Psycholinguistik, Psychologie und sprachbezogener Lehr-Lern-Forschung oftmals in ganz divergenter Weise verwendet (Jude & Klieme, 2007, S. 10). Während der Begriff der Sprachkompetenz in der Linguistik und Psycholinguistik beispielsweise aus einer primär theoretisch-kognitiven Perspektive heraus definiert wird – was bedeutet, Sprachkompetenz als Verwendung von Sprache, unter Einbeziehung der strukturellen und grammatikalischen Regeln eines Sprachsystems sowie sprachbezogener Kognition zu betrachten – wird Sprachkompetenz im Forschungsfeld der (Entwicklungs-)Psychologie vor allem in Hinblick auf Spracherwerbsprozesse, „durch die Individuelle Wahrnehmungsentwicklung, spezifische Lernmechanismen sowie die Ausprägung sprachspezifischer, kognitiver Sprachverarbeitungsstrukturen“ bedingt sind, untersucht (Grimm, 2000; Jude & Klieme, 2007, S. 13; Szagun, 2000). Im diagnostischen Bereich der Lehr-Lern-Forschung wird Sprachkompetenz vor allem als Gesamtkompetenz verstanden, die sich aus vielen unterschiedlichen Teilkompetenzen (z. B. spezifische Lese- und Schreibkompetenzen) zusammensetzt und deren empirische Erfassung Aussagen über sprachliche Entwicklungsstände, Lernpotenziale und Lernfortschritte von Lernenden ermöglicht (Bachmann, 1990; Jude & Klieme, 2007, S. 10).

Aus diesem Grund ist es sinnvoll und notwendig, eine spezifische Arbeitsdefinition zu formulieren, die eine Erklärung des Kompetenzbegriffs in Bezug auf den konkreten Kontext und die Fragestellung dieser Arbeit enthält. Der Versuch einer Definition von Sprachkompetenz ist in diesem Zusammenhang vor allem als eine begriffliche Eingrenzung für den für diese Arbeit relevanten Kontext zu verstehen, die dabei jedoch keinerlei Anspruch erhebt, das komplexe Phänomen Sprachkompetenz auch nur annähernd vollständig abzubilden.

Da im Rahmen der vorliegenden Studie (schrift-)sprachliche Kompetenzen von Oberstufenschüler*innen untersucht und mit dem C-Test als zentrales Testinstrument erfasst werden sollen, erscheint es sinnvoll, den Begriff der Sprachkompetenz zunächst aus drei konkreten für diese Arbeit relevanten Blickwinkeln zu betrachten und sich aus

diesen heraus einer Begriffsbestimmung anzunähern. Einen ersten Bezugspunkt liefert ein Teil der in Kapitel 1 beschriebenen Ausgangslage, nämlich dass Studierende ausweislich der SkaLa®-Ergebnisse ihr Studium mit teilweise defizitär ausgebildeten sprachlichen Kompetenzen beginnen. Da die SkaLa®-Ergebnisse zumindest in Teilen die Fragestellung dieser Arbeit motiviert haben und das Testformat (so wurde u. a. ein linksseitig getilgter C-Test eingesetzt) sowie die untersuchte Kohorte eine gewisse Nähe zu der in der vorliegenden Arbeit aufweisen, soll vom SkaLa®-Test- und Untersuchungsdesign ausgehend nachvollzogen werden, welche Annahmen und Definitionen von Sprachkompetenz SkaLa® zugrunde liegen könnten. Ein zweiter Bezugspunkt für die Definition von Sprachkompetenz stellt das in dieser Arbeit zentral eingesetzte Testinstrument dar und damit die Frage, von welchem Verständnis von Sprachkompetenz ausgegangen wird, wenn diese mit dem C-Test gemessen werden soll. Den dritten Anhaltspunkt für eine konkrete Begriffsbestimmung von sprachlichen Kompetenzen stellt die zu untersuchende Zielkohorte dar, welche sich in dieser Arbeit aus Oberstufenschüler*innen zusammensetzt. Wie wird Sprachkompetenz also im Kontext der gymnasialen Oberstufe verstanden und definiert?

2.1.1 (Schrift-)Sprachliche Kompetenzen in SkaLa®

Das Online-Self-Assessment *SkaLa®* der Universität Duisburg-Essen bietet laut Homepage des Zentrums für Lehrerbildung Studienbewerbern des Lehramts die Möglichkeit, eine genaue Rückmeldung zu ihren schriftsprachlichen Kompetenzen, genauer gesagt zu ihren Lese- und Schreibfähigkeiten, zu erhalten[11]. Über eine kombinierte Lese-Schreibaufgabe sollen sowohl produktive Fertigkeiten (Schreibkompetenz) als auch rezeptive Fertigkeiten (Lesekompetenzen) erfasst werden, indem Studierende dazu aufgefordert sind, einen argumentativen (Lese-)Text unter Berücksichtigung der zentralen Frage des Textes, der zentralen These des Autors sowie der Begründung der These schriftlich zusammenzufassen (Bremerich-Vos, 2016, S. 48; Bremerich-Vos & Scholten-Akoun, 2016, S. 48). Da sich SkaLa® an

[11] Diese und weitere Informationen zu SkaLa® unter https://zlb.uni-due.de/node/352 (zuletzt abgerufen am 18.06.2018).

Studien*bewerber*innen* richtet, orientiert sich laut Testentwickler*innen der Aufgabenimpuls an den von der Kultusministerkonferenz verabschiedeten Bildungsstandards im Fach Deutsch für die Allgemeine Hochschulreife, ausweislich derer Oberstufenschüler*innen „unter Beachtung sprachlicher Normen […] argumentative Texte wiedergegeben und beurteilt werden können“ (Bremerich-Vos, 2016, S. 52; MSB NRW, 2007). In SkaLa® sollen somit nicht primär die im akademischen Kontext relevanten Kompetenzen des wissenschaftlichen Schreibens überprüft werden, sondern vielmehr einige der Voraussetzungen für die Entwicklung dieser Kompetenzen – also jene von Bremerich-Vos und Scholten-Akoun (2016) recht vage formulierten basalen sprachlichen Kompetenzen, die bereits voruniversitär erworben wurden (Scholten-Akoun, 2016a, S. 18, 2016b, S. 32). Ein Blick auf die Beurteilungskriterien[12] nach denen die Schreibaufgabe in SkaLa® ausgewertet wurde, lässt nun vermuten, dass die Testentwickler*innen unter diesen produktiven basalen sprachlichen Kompetenzen im Allgemeinen das Vermögen der Studierenden verstehen, orthografisch einwandfrei zu schreiben, die korrekte Interpunktion des eigenen Textes zu realisieren, Wörter und Kollokationen aus dem eigenen Wortschatz angemessen auszuwählen und grammatisch richtig zu verwenden und in einem angemessenen schriftsprachlichen Register die aufgegebene Wiedergabe eines Textimpulses in seiner argumentativen Struktur zu leisten (Scholten-Akoun et al., 2016, 67 ff.). Ein Teil dessen, was die Studierenden als Text produzieren wird vor allem aus einer formalsprachlichen Perspektive heraus bewertet und hier also im Sinne der vom *GER* formulierten linguistischen Kompetenz aufgefasst (Council of Europe, 2001, S. 168–171). Diese linguistischen Kompetenz setzt sich laut *GER* zusammen aus der lexikalischen Kompetenz, der grammatischen Kompetenz, der semantischen Kompetenz und der phonologischen Kompetenz, die auch in den Beurteilungsvariablen der Dimension Sprache in SkaLa® wiederzufinden sind: Bewertet werden in SkaLa sowohl lexikalische Kompetenz (wie etwa die Angemessenheit der Wortschatzbreit und die korrekte Verwendung von Kollokationen und idiomatische Wendungen), grammatische Kompetenz (wie etwa die Beherrschung grammatikalischer Regeln in den Bereichen von Numerus, Kasus, Genus und Tempus,

[12] Einen detaillierten Überblick über die Zusammenstellung der bewerteten Variablen der Schreibaufgabe in SkaLa liefert der Auszug des Codierhandbuchs in Bremerich-Vos und Scholten-Akoun (2016, S. 67–122)

Konjugation, Deklination, Rektion, Valenz, Kongruenz sowie die Anwendung von Prinzipien der Wortbildungslehre und syntaktischen Regeln), semantische Kompetenz (Fragen der Wortbedeutung und der Konnotation) und orthographische Kompetenz[13] (richtige Schreibweise der verwendeten Wörter und gängigen Abkürzungen und das Befolgen der Regeln der Zeichensetzung) (Council of Europe, 2001, S. 91; Europarat, 2001, S. 110–118; Scholten-Akoun et al., 2016, S. 67). Linguistische Kompetenz wird hier also als eine Facette von Sprachkompetenz verstanden, die definiert wird „als Kenntnis der formalen Mittel, aus denen wohlgeformte, sinnvolle Mitteilungen zusammengesetzt und formuliert werden können, und als die Fähigkeit, diese Mittel auch zu verwenden“ (Europarat, 2001, S. 110).

Die Studierenden sollen in SkaLa® jedoch nicht nur ihr produktives Sprachvermögen im linguistischen Sinne - wie oben gezeigt - unter Beweis stellen, sondern durch die Kombination des Lesetextes und dem Aufgabenimpuls der Textproduktion zeigen, dass sie die im Lesetext (an verschiedenen Stellen) enthaltenen und im Aufgabenimpuls gezielt abgefragten Informationen heraussuchen und den logisch-argumentativen Zusammenhang dieser Informationen in einem eigenen Text wiedergeben können (Bremerich-Vos, 2016, S. 40–51). Neben den produktiven Fertigkeiten, die sich in der (Schreib-)Performanz der Studierenden zeigen, werden also zusätzlich rezeptive Leistungen miterfasst. Auch wenn davon auszugehen ist, dass die Textrezeption die Textproduktion beeinflusst, sind die Dimensionen von Lesen und Schreiben in diesem Fall nicht eindeutig voneinander trennbar: Denn ob die Rezeption des Textes gelungen ist, kann durch das Aufgabendesign bei SkaLa® – anders als in dieser Untersuchung durch den Einsatz eines standardisierten Lesetests – nicht explizit miterfasst werden, da die Aufgabenbewältigung eben auch an der produktiven Textwiedergabe scheitern kann. Die Bewertung des Textverstehens wird nun bei SkaLa® auf einer inhaltlichen Ebene dahingehend überprüft, ob die zentrale Frage, die zentrale These des Autors und die Begründung der These durch Nennung der im Text vorkommenden Argumente verstanden und genannt werden. Das Abfragen dieser Fertigkeiten ließe sich im weitesten Sinne an die den PISA-Studien zugrunde liegende Definition von

[13] *Orthographische Kompetenz* wird im GER als Teil der *phonologischen Kompetenz* verstanden, da sie „die Kenntnis der Symbole, aus denen geschriebene Texte bestehen, sowie die Fertigkeit, diese wahrzunehmen und zu produzieren“ involviert (Council of Europe (2001, S. 117).

Lesekompetenz der OECD „geschriebene Texte zu verstehen, zu nutzen und über sie zu reflektieren, um eigene Ziele zu erreichen, das eigene Wissen und Potenzial weiterzuentwickeln und am gesellschaftlichen Leben teilzunehmen“ und im Einzelnen an die im Rahmen der PISA-Studien vorgeschlagenen gestuften Leseteilkompetenzen – also etwa die Fähigkeit, (leicht bis hin zu schwierig zugängliche) Informationen zu ermitteln, die Hauptgedanken des Textes in ihrem argumentativen Beziehungsgefüge (in groben Zügen bis hin zu differenziert) zu erkennen usw. anlehnen (OECD, 2004, 2016, 2019b, 2019c). Die Analyse der Textwiedergabeleistung in SkaLa beschränkt sich in diesem Fall jedoch vor allem darauf zu überprüfen, ob und wie (vollständig/richtig usw.) die Informationen aus dem Text entnommen wurden und ob und wie ein (allgemeines/differenziertes) Verständnis der Inhalte des Textimpulses sichtbar wird[14]. Die bei PISA genannten reflektierenden und bewertenden Fähigkeiten werden nicht explizit bewertet sondern vielmehr indirekt miterfasst.

Als zweites Testinstrument wird in SkaLa® ein linksseitig getilgter C-Test eingesetzt, um ein „allgemeines Bild der rezeptiven und produktiven Fertigkeiten“ (Mashkovskaya & Baur, 2016a, S. 191) der Studierenden zu gewinnen und Förderbedarfe aufzudecken, die sich auf die „diagnostischen Informationen, die der C-Test zu unterschiedlichen sprachlichen Bereichen liefert (z. B. Grammatik, Orthografie, Lexik, Semantik, Syntax, Leseversverständnis)“, stützen (Mashkovskaya & Baur, 2016b, S. 220). Da in der vorliegenden Arbeit ebenfalls der C-Test als Testinstrument zur Erfassung von schriftsprachlichen Fähigkeiten und Fertigkeiten eingesetzt wird, soll im Folgenden genauer betrachtet werden, wie einer Definition Sprachkompetenz in Bezug auf den C-Test näher gekommen werden könnte.

2.1.2 Sprachkompetenz im C-Test

Bevor es in Kapitel 2.3 ausführlicher darum gehen wird, wie der C-Test als Testkonstrukt funktioniert und wie mit ihm Sprachkompetenz gemessen werden kann, soll an dieser Stelle – zur begrifflichen Eingrenzung der Definition von Sprachkompetenz – zunächst

[14] Vgl. Ausführlich zum Aufgabenimpuls und zur Variablenübersicht der inhaltlichen Auswertung der Schreibaufgabe in SkaLa (Bremerich-Vos und Weirich (2016)

grob beleuchtet werden, was in Bezug auf den C-Test unter Sprachkompetenz verstanden werden kann.

In der (Fach-)Literatur wird der C-Test als Testinstrument beschrieben, mit dem allgemeine Sprachkompetenz (*general language proficiency*) gemessen und erfasst werden kann (Baur & Spettmann, 2007, 95 ff.; Grotjahn, 2002a, 211 ff.; Klein-Braley, 1985a, 76 ff.; Klein-Braley & Raatz, 1984, 134 ff.; Raatz & Klein-Braley, 2002, S. 75). Doch wie genau lässt sich diese allgemeine Sprachkompetenz fassen, die der C-Test zu messen vorgibt? Eine Antwort auf diese Frage könnte – zumindest in Teilen – ein Blick auf die Anforderungen bzw. die Fähigkeiten und Fertigkeiten liefern, die bei der Bearbeitung eines C-Tests eine Rolle spielen dürften. Um einen C-Test lösen, d. h. alle getilgten Worthälften im Text erfolgreich rekonstruieren zu können, ist zunächst Leseverständnis erforderlich. Das Lesen und Verstehen der C-Test-Überschrift sowie des tilgungsfreien Satzes jeweils am Textanfang und Textende kann in einem ersten Schritt hilfreich sein, da die Proband*in über das Lesen erste Informationen über die Thematik, die Erzählperspektive, die Form des Textes etc. schöpfen kann (Harsch & Schröder, 2007, 212 ff.). Mit der Aktivierung von inhaltlichem und sprachlichem Vorwissen (z. B. in Bezug auf die Textsorte, Fachwortschatz, Textaufbau, etc.) sowie einer damit verbundenen Erwartungshaltung gegenüber dem Text gehen Vorentscheidungen hinsichtlich der Bandbreite möglicher (zunächst noch abstrakter) Textrekonstruktionen einher (ebd.).

Die Rekonstruktion der einzelnen Lücken im C-Test kann nun über die Anwendung von verschiedenen Strategien auf unterschiedlichen Ebenen bzw. über den Rekurs auf unterschiedliche Facetten von Sprachkompetenz erfolgen (Mashkovskaya & Baur, 2016a, S. 195): Bei Lücken, die durch die Einbeziehung der unmittelbar vor oder nach der Tilgung stehenden Wörter oder des ganzen Satzes, also auf der Wort- und Satzebene[15] gelöst werden können, kommen besonders linguistische Fertigkeiten und Fähigkeiten zum Tragen: Um die fehlenden Buchstaben eines Wortes erfolgreich ergänzen zu können, sind neben (text-)grammatischen und syntaktischen Kompetenzen vor allem lexikalische Fähigkeiten und Fertigkeiten – d. h. präzise allgemeinsprachliche sowie fach-, domänen- und textformspezifische Wortschatzkenntnisse – erforderlich,

[15] Zur Kategorisierung von C-Test-Lücken und Lösungsstrategien auf den unterschiedlichen Ebenen siehe (Mashkovskaya (2014, 161 ff.).

ebenso wie die Fähigkeit die ergänzten Worthälften orthographisch korrekt zu ergänzen (Daller et al., 2003, S. 206). Für die in dieser Untersuchung eingesetzte C-Test-Version mit linksseitiger Tilgung spielen – neben den grammatischen, syntaktischen, kontextuellen und bildungssprachlichen Facetten von Sprachkompetenz – besonders semantische und lexikalische Fähigkeiten und Fertigkeiten eine entscheidende Rolle, da diese bei der Ergänzung von Wortanfangen stärker im Fokus stehen (Mashkovskaya & Baur, 2016a, S. 195; Scholten-Akoun et al., 2014, S. 408) (vgl. ausführlich dazu Kapitel 2.3.5). Diese mit linksseitig getilgten C-Tests vornehmlich erfassten lexikalischen-semantischen Sprachkenntnisse (dazu gehören etwa eine angemessene Wortschatzbreite, das Erkennen und die korrekte Verwendung von Kollokationen und idiomatischen Wendungen sowie Kenntnisse über Wortbedeutungen und Konnotationen) werden dabei in der Regel über den Zugriff auf das mentale Lexikon – einen Kernbereich des Sprachgebrauchs und Schnittstelle für den schulischen Spracherwerb (Singleton, 1999, 220 f.; Steinhoff, 2009) – abgerufen.

Neben dem Zugriff auf das mentale Lexikon muss – um nun auf der Textebene entscheiden zu können, welche Ergänzungsmöglichkeiten plausibel erscheinen – sowohl der Ko- als auch der Kontext des gesamten Textes einbezogen werden: Für die Rekonstruktion der Lücken in einem C-Test ist das Erkennen von Beziehungen zwischen den Sätzen und damit ein differenziertes Textverständnis erforderlich, welches verbunden ist mit der Fähigkeit, „einen Text vorwärts und rückwärts zu lesen und Leerstellen aus dem Kontext zu erschließen“ (Mashkovskaya, 2014, S. 165; Schröder & Harsch, 2003). So kann auf der Textebene überprüft werden, welche Ergänzungen semantisch sinnvoll sind und ob der Text durch die Vervollständigung der Lücken kohärent bleibt. Harsch und Schröder (2014) argumentieren in diesem Zusammenhang, dass C-Test-Ergebnisse nicht nur Rückschlüsse auf Facetten des rezeptiven Sprachkönnens sondern – beispielsweise durch die Überprüfung von Lösungen auf Plausibilität und die Vokalisierung des Textes beim stillen Lesen – auch von produktiven Fähigkeiten und Fertigkeiten zulassen.

Die erfolgreiche Bearbeitung eines C-Tests erfordert, wie in der unten stehenden Grafik von Harsch und Schröder (2007, S. 214) gezeigt, neben der Aktivierung von sprachlichen – d. h. vor allem im oben definierten Sinne linguistischen – Fähigkeiten und Fertigkeiten

den Zugriff auf verschiedene strategische Kompetenzen, wie beispielsweise den Rekurs auf Weltwissen, Vorgehensweisen zu Bedeutungserschließung oder die Anwendung von Interpolationstechniken[16] (Harsch & Schröder, 2007, S. 231). Solche Strategien sind komplexe kognitive und sprachliche Integrationsleistungen, die sowohl im Gedächtnis gespeicherte Informationen (Top-Down-Verarbeitung), textspezifischen Informationen (Bottom-Up-Verarbeitung) als auch spezifisch sprachliche Informationen beim Lösen der Lücke miteinbeziehen (Eckes, 2010, 126 f.) und damit auch „Rückschlüsse auf die tiefer liegende, abstrakt-latente Sprachkompetenz“ zulassen (Scholten-Akoun et al., 2013, S. 184; Vockrodt-Scholz & Zydatiß, 2010, S. 5).

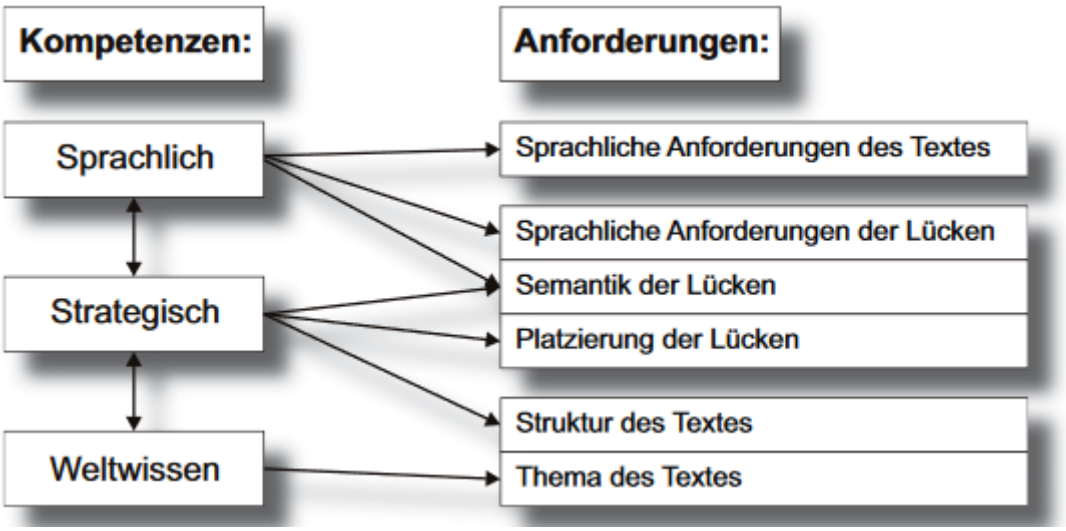


Abbildung 1: Kompetenzen und Anforderungen im C-Test nach Harsch und Schröder

Quelle: Harsch & Schröder 2007, S. 241

Ausgehend vom Funktionsprinzip und der Wirkungsweise des C-Tests[17] wird angenommen, dass die Bearbeitung eines C-Tests durch den integrativen Zugriff u. a. auf die hier genannten Fähigkeiten und Fertigkeiten gelingen kann und

[16] Unter Interpolationstechniken sind hier vor allem Vorgehensweisen gemeint, bei denen Fakten mit textuellen, sprachlichen Informationen und Wissensbeständen verknüpfen werden (Harsch, Schröder, 2007 S. 217).
[17] Ausführliche Informationen zum Funktionsprinzip des C-Tests, seiner Wirkungsweise und zur Frage, wie mit dem C-Test Sprachkompetenz gemessen werden kann in Kapitel 2.3.

Sprachkompetenz, die mit dem C-Test erfasst werden soll, als eine komplexe aber globale Gesamtkompetenz aller Sprachleistungen – sowohl rezeptiver als auch produktiver Fertigkeiten (Leseverstehen, grammatikalisches Wissen, lexikalische Kenntnisse, die Beherrschung der Rechtschreibregeln, kognitive Dispositionen, Informationsverarbeitungsprozesse, usw.) – verstanden wird (Vollmer, 1982, S. 39). Für die vorliegende Arbeit bedeutet das, dass gerade diese mit dem C-Test erfassbare, integrative sprachliche Gesamtleistungen der Oberstufenschüler*innen in den Fokus gerückt werden. Dabei stellt das Abschneiden der Oberstufenschüler*innen in dem für diesen Zweck neu entwickelten C-Test selbstverständlich kein Zeugnis über deren generelle Gestaltungsfähigkeit von Sprache aus. Aussagen dieser Reichweite verlangen eine andere erheblich erweiterte Untersuchungsbasis.

2.1.3 Sprachkompetenz in der gymnasialen Oberstufe

Ein dritter Ansatz, von dem ausgehend versucht werden soll, sich einer für diese Arbeit relevanten Definition von Sprachkompetenz anzunähern, ist die Betrachtung der geltenden Anforderungen und Kompetenzerwartungen an die untersuchte Zielkohorte von Oberstufenschüler*innen. Welche zentralen Fähigkeiten und Kompetenzen Schüler*innen in der gymnasialen Oberstufe erreichen sollen, ist seit der Verabschiedung durch die Kultusministerkonferenz (KMK) für die Allgemeine Hochschulreife in den Bildungsstandards für das Fach Deutsch nachzulesen (Becker-Mrotzek et al., 2015, 10 ff.). Basierend auf dem Kompetenzmodell für die Fachdidaktik Deutsch von (Ossner, 2006, 9 ff.) werden in den Bildungsstandards Deutsch fünf miteinander verzahnte Kompetenzbereiche beschrieben, die sich aus den prozessbezogenen Kompetenzen *Sprechen und Zuhören*, *Schreiben* und *Lesen* sowie aus den domänenspezifischen Kompetenzbereichen *Sich mit Texten und Medien auseinandersetzen* und *Sprache und Sprachgebrauch reflektieren* zusammensetzen (KMK, 2012b, S. 14):

Tabelle 1: Die Kompetenzbereiche für die gymnasiale Oberstufe für das Fach Deutsch

Domänenspezifischer Kompetenzbereich	Prozessbezogener Kompetenzbereich	Domänenspezifischer Kompetenzbereich
Sich mit Texten und Medien auseinandersetzen	Sprechen und Zuhören	Sprache und Sprachgebrauch reflektieren
	Schreiben	
	Lesen	

Der prozessbezogene Kompetenzbereich bezieht sich auf allgemeine produktive und rezeptive Fertigkeiten und Fähigkeiten in den Bereichen *Sprechen und Zuhören*, *Schreiben* und *Lesen* und damit „verstärkt auf die Auseinandersetzung mit externen Wissensbeständen, die selbstständige Wissensaneignung, die Persönlichkeitsbildung sowie die eigene mündliche und schriftliche Ausdrucksfähigkeit“ (Becker-Mrotzek et al., 2015, S. 16). In den Kompetenzanforderungen der domänenspezifischen Bereiche wird hingegen beschrieben, „an welchen zentralen fachspezifischen Inhalten diese [prozessbezogenen] Kompetenzen zu erwerben sind“ (ebd.). Schreiben, Lesen, Sprechen und Zuhören werden in der aktiven Auseinandersetzung mit konkreten domänenspezifischen Lerninhalten – also hier durch die Reflexion von Sprache und Sprachgebrauch und die Auseinandersetzung mit Texten und Medien – erworben und weiterentwickelt. Zugleich unterstützen prozessbezogene Kompetenzen den Erwerb inhaltsbezogener Fertigkeiten und Fähigkeiten. Über die einzelnen Kompetenzbereiche hinaus kann es also auch zu Überschneidungen und Vervielfachungen von (Teil-)Kompetenzen kommen, da beispielweise die Schreibkompetenz die Lesekompetenz in Anspruch nehmen kann und umgekehrt (Becker-Mrotzek & Schindler, 2007, S. 12).

Für alle fünf Bereichen sind nun insgesamt ungefähr 100 Anforderungen und Kompetenzerwartungen formuliert, die beschreiben, was Schüler*innen im Fach Deutsch in der gymnasialen Oberstufe können (sollen) (KMK, 2012b). Da es an dieser Stelle um eine begriffliche Einordnung von Sprachkompetenz geht, wird nicht auf alle 100 für das Fach Deutsch geltenden Kompetenzbeschreibungen eingegangen, sondern eben nur auf jene, die sich – in Anlehnung an die in SkaLa® erhobenen basalen Lese- und Schreibfähigkeiten (vgl. Kapitel 3.1.1) und besonders an das in dieser Untersuchung eingesetzte Testinstrument, den C-Test – auf schriftsprachliche Kompetenzen beziehen.

Da sowohl bei SkaLa® als auch bei der Bearbeitung von C-Tests Lesekompetenz im oben definierten Sinne eine wesentliche Rolle zu spielen scheint, soll an dieser Stelle zunächst ein Blick auf die an Oberstufenschüler*innen gestellten Kompetenzerwartungen im Bereich *Lesen* geworfen werden. Den Beschreibungen der Bildungsstandards für diesen Kompetenzbereich folgend (KMK, 2012b, S. 18), können die Oberstufenschüler*innen

- „den komplexen Zusammenhang zwischen Teilaspekten und dem Textganzen erschließen,
- aus anspruchsvollen Aufgabenstellungen angemessene Leseziele ableiten und diese für die Textrezeption nutzen,
- im Leseprozess ihre auf unterschiedlichen Interpretations- und Analyseverfahren beruhenden Verstehensentwürfe überprüfen,
- die Einsicht in die Vorläufigkeit ihrer Verstehensentwürfe zur kontinuierlichen Überarbeitung ihrer Hypothesen nutzen,
- Verstehensbarrieren identifizieren und sie zum Anlass eines textnahen Lesens nehmen,
- Kontextwissen heranziehen, um Verstehensbarrieren zu überwinden,
- ihr Fach- und Weltwissen flexibel einsetzen, um das Textverständnis zu vertiefen und die Relevanz des Gelesenen einzuschätzen,
- im Leseprozess ihr fachliches Wissen selbstständig zur Erschließung und Nutzung voraussetzungsreicher Texte heranziehen“ (KMK, 2012b, S. 18).

Diesen Beschreibungen der Bildungsstandards entsprechend, sollen Oberstufenschüler*innen folglich über Fähigkeiten und Fertigkeiten verfügen, die es ihnen ermöglichen, sowohl selektiv, global als auch detailliert zu lesen, Verstehensprobleme unter Einbeziehung des Ko-und Kontextes selbstständig zu lösen sowie ihr Textverständnis fortlaufend zu überprüfen. Die Kompetenzbeschreibungen für den Bereich Lesen weisen somit eine gewisse Nähe zu dem Verständnis von Lesekompetenz in Bezug auf den C-Test auf – zu nennen wären hier z. B. die Fähigkeiten einen Text von „hinten nach vorn“ zu lesen, Nutzung von Ko- und Kontext, aber auch das Überprüfen der eigenen Lösungen auf Plausibilität im Textzusammenhang (vgl. Kapitel 2.2.1). Zum anderen legen die Kompetenzerwartungen der Bildungsstandards nahe, dass Lesekompetenz ähnlich wie bei PISA als eine Art „Reading Literarcy“ verstanden

wird, die eben dort die Fähigkeit beschreibt, „geschriebene Texte zu verstehen, zu nutzen und über sie zu reflektieren, um eigene Ziele zu erreichen, das eigene Wissen und Potential weiterzuentwickeln“ (Baumert et al., 2001, S. 53). Lesen geht in diesem Sinne also weit über das einfache Dekodieren von Information hinaus, da die Leser*innen in der aktiven Auseinandersetzung mit Texten unter Einbeziehung ihres Welt-, Vor- und Sprachwissens erst die Bedeutung eines Textes konstruieren, ihn so verstehen und reflektieren können (Baumert et al., 1999, S. 2).

Schriftsprachliche Kompetenzen wie sie oben in Bezug auf SkaLa® und auf den C-Test definiert wurden, umfassen neben rezeptiven auch produktive Fähigkeiten und Fertigkeiten. Während der C-Test produktives Sprachvermögen integrativ abfragt, wird von den Studierenden bei der Bearbeitung von SkaLa® explizit verlangt, einen zusammenhängenden Text zu schreiben (vgl. Kapitel 2.1.1). In dem Kompetenzbereich *Schreiben* ist nun auch in den Bildungsstandards nachzulesen, welche schriftsprachlichen produktiven Fertigkeiten und Fähigkeiten von Oberstufenschüler*innen erwartet werden: So können die Schüler*innen „inhaltlich angemessene kohärente Texte [schreiben], die sie aufgabenadäquat, konzept-geleitet, adressaten- und zielorientiert, normgerecht, sprachlich variabel und stilistisch stimmig gestalten“ (KMK, 2012b, S. 16), wobei sie Schreibstrategien und Textmuster „für die Produktion komplexer informierender, erklärender und argumentierender sowie gestaltender Textformen selbstständig und aufgabenbezogen“ verwenden (KMK, 2012b, S. 17). Diese Beschreibungen zeigen somit wiederum eine gewisse Nähe zu der in Anlehnung an SkaLa® definierten basalen Schreibkompetenz auf, die eben – wie oben erläutert – im Wesentlichen als das Vermögen verstanden werden könnte, orthografisch einwandfrei und normgerecht zu schreiben, Wörter und Kollokationen aus dem eigenen Wortschatz angemessen auszuwählen und grammatisch richtig zu verwenden und in einem angemessenen schriftsprachlichen Register eine aufgegebene Textproduktion einer bestimmten Struktur folgend zu leisten (Scholten-Akoun et al., 2016, 67 ff.).

Obwohl nun bei der Bearbeitung eines C-Tests – anders als bei SkaLa® oder in der schriftlichen Abiturprüfung Deutsch – keine zusammenhängende Textproduktion verlangt wird, ist allen drei Aufgabenformaten gemeinsam, dass ein Teil dessen, was die Proband*innen als Text oder eben als Wortergänzung produzieren, aus einer

formalsprachlichen Perspektive heraus bewertet wird. Diese Perspektive legt wiederum ein Verständnis von Sprachkompetenz nahe, welches im Sinne der vom *GER* formulierten linguistischen Kompetenz aufgefasst werden könnte (Council of Europe, 2001). Während in den Bildungsstandards für die fortgeführte Fremdsprache (Englisch/Französisch) linguistische Fertigkeiten und Fähigkeiten explizit unter dem Punkt *Verfügen über sprachliche Mittel* (KMK, 2012a, S. 18)[18] benannt und beschrieben sind, fehlt eine solche Kategorie in den Bildungsstandards für das Fach Deutsch. Als linguistische Kompetenz kann im Fach Deutsch wohl am ehesten die Beschreibung aus dem Bereich *Schreiben* verstanden werden, die sich auf die zu erwartende Kompetenz von Oberstufenschüler*innen bezieht, „Texte orthografisch und grammatisch korrekt sowie fachsprachlich präzise, prägnant und stilistisch angemessen verfassen“ zu können (KMK, 2012b, S. 16).

Weiter konkretisiert wird diese Beschreibung in den (Beispiel-) Beurteilungsempfehlungen des Ministeriums für Schule und Weiterbildung des Landes NRW (2017) für die schriftliche Abiturprüfung für das Fach Deutsch, in der in Bezug auf die Darstellungsleistung der Oberstufenschüler*innen u. a. bewertet werden soll, ob der Prüfling „seinen Text kohärent, schlüssig, stringent und gedanklich klar“ formuliert, ob also beispielsweise einzelne Arbeitsschritte schlüssig verbunden und Sätze gedanklich nachvollziehbar verknüpft sind. Bewertet werden soll außerdem die „Verwendung von Fachtermini in sinnvollem Zusammenhang“, die „Beachtung der Tempora“, das Vorhandensein einer „adressaten- und zielgerechte Schreibweise“ und damit verbunden die Einhaltung von Schriftsprachlichkeit (MSB NRW, 2017, S. 14). Bezug nehmend auf die in den Bildungsstandards ausgeschriebene Kompetenz „orthografisch und grammatisch korrekt“ zu schreiben (s. o.) sollen zudem Punkte vergeben werden, wenn der Prüfling „lexikalisch und syntaktisch sicher, variabel und komplex (und zugleich klar)“ formuliert sowie „sprachlich richtig schreibt“ (ebd.).

Um nun auf der Grundlage dieser Beurteilungsempfehlungen ein klareres Verständnis dessen zu bekommen, was an dieser Stelle eben nicht nur als Beurteilungsmerkmal sondern, als dahinterliegende (produktive) Sprach- bzw. linguistische Kompetenz

[18] Genannt werden hier Beschreibungen der Kompetenzerwartungen der Bereiche Wortschatz, Grammatik, Aussprache, Prosodie sowie Orthografie, vgl. KMK (2012a, S. 18)

verstanden werden könnte, lohnt an dieser Stelle ein Blick auf das *Modell zu den domänenspezifischen Anforderungen des Bereichs Schreiben* von Becker-Mrotzek, welches sich auf das Kompetenzmodell für die Fachdidaktik Deutsch von Ossner (2006, 9 ff.) bezieht:

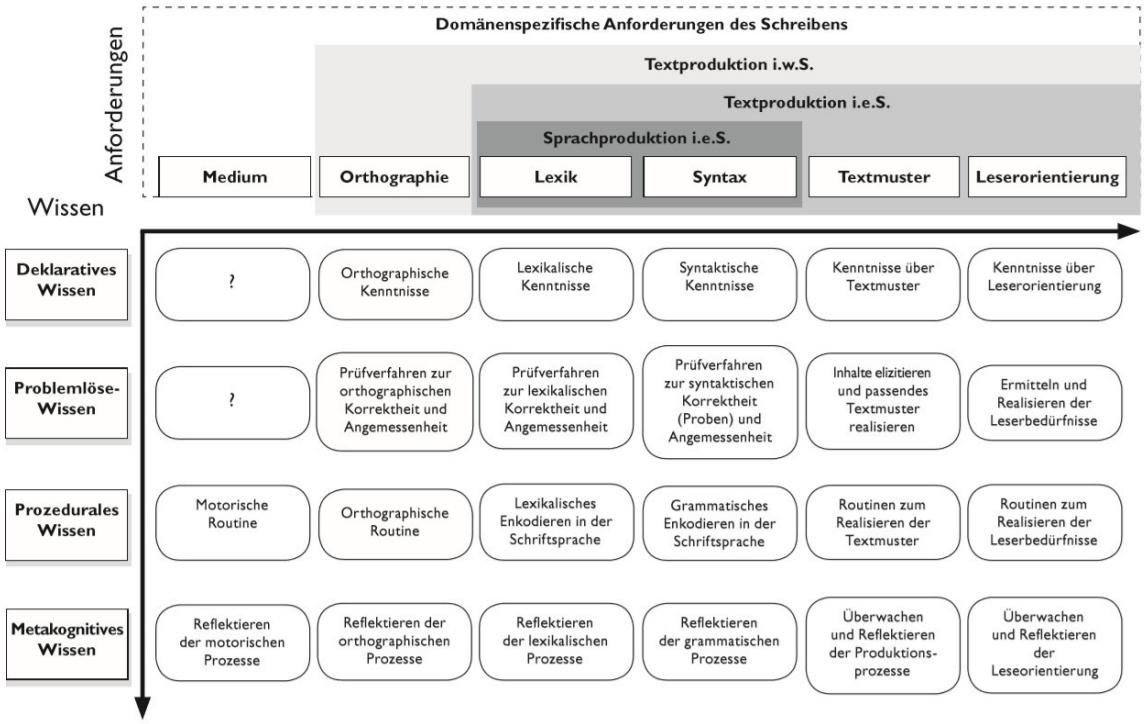


Abbildung 2: Das Schreibkompetenzmodel von Becker-Mrotzek und Schindler

Quelle: Becker-Mrotzek & Schindler 2007a, S. 24

In ihrem Schreibkompetenzmodell nehmen Becker-Mrotzek und Schindler (2007, 8 ff.) eine linguistisch orientierte und hierarchische angelegte Binnendifferenzierung vor, indem sie den Bereich *Schreiben* in Textproduktion im weiteren Sinne, Textproduktion im engeren Sinne und Sprachproduktion im engeren Sinne unterteilen (Mathiebe, 2018, S. 25). Dabei gehen die Autoren davon aus, dass eine schriftliche Textproduktion 1. motorische Prozesse einschließt (*Medium*), 2. Orthographie benötigt, die Regeln für die Verwendung von Schriftzeichen enthält (*Orthographie*), 3. die Verwendung des Lexikons (*Lexik*) sowie 4. die Verwendung von Syntax verlangt (*Syntax*), 5. das Herstellen einer bestimmten musterhaften Textstruktur beinhaltet (*Textmuster*) und sich 6. an einen

absenten Adressaten[19] richtet (*Leserorientierung*) (Becker-Mrotzek & Schindler, 2007, 12 f.). Diese inhaltlichen Anforderungen werden weiter in den ursprünglich von Mandl et al. (1986, 123 ff.) vorgeschlagenen vier Wissensbereichen *Deklaratives Wissen*, *Problemlösewissen*, *Prozedurales Wissen* und *Metakognitives Wissen* entfaltet. Diesem Modell folgend bedeutet Kompetenz also ein bestimmtes (deklaratives, problemösendes, prozedurales oder metakognitives) Wissen über eine inhaltliche Anforderung zu aktivieren (Becker-Mrotzek & Schindler, 2007, S. 10).

Anders als in den eher allgemeinformulierten Kompetenzbeschreibungen in den Bildungsstandards – orthographisch, grammatisch korrekt, präzise, prägnant und stilistisch angemessen zu schreiben (KMK, 2012b, S. 16) – formulieren Becker-Mrotzek und Schindler (2007, 10 ff.) einzelne Kompetenzbereiche, wie beispielwiese den der Orthographischen Kompetenz, der Lexikalischen Kompetenz, der Syntaktischen Kompetenz usw., denen auf verschiedenen Ebenen konkrete Können- bzw. Wissensbeschreibungen zugeordnet werden und die sich in gewisser Weise zumindest auch in den oben genannten Beurteilungskriterien für die schriftliche Abiturprüfung wiederfinden. „Lexikalisch sicher, variabel und komplex“ schreiben zu können – wie in Beurteilungsempfehlungen beschrieben (MSB NRW, 2017, S. 14) – würde dem Modell von Becker-Mrotzek und Schindler (2007) folgend somit in den Bereich der Lexikalischen Kompetenz fallen. Die Lexikalische Kompetenz beinhaltet in diesem Zusammenhang nun beispielsweise, dass das Verfassen eines Textes zunächst einmal lexikalische Kenntnisse voraussetzt, genauer gesagt einen allgemeinsprachlichen, fach-, themen- oder domänenspezifischen Wortschatz, mit und über den Inhalte und kausale, temporale, lokale Zusammenhänge sprachlich geäußert werden können (deklarative Wissensebene). Diese lexikalischen Kenntnisse werden nicht nur angewendet sondern im Prozess der Sprach- bzw. Textproduktion fortlaufend auf ihre Korrektheit und Angemessenheit hin überprüft (problemlösendes Wissen) (Becker-Mrotzek & Schindler, 2007, S. 13). Je sicherer und routinierter dabei auf einen breiten schriftsprachlichen und domänenspezifischen Wortschatz zurückgegriffen werden kann, desto einfacher und

[19] Da es sich bei der schriftlichen Kommunikation um eine „zerdehnte Kommunikation“ handelt (Ehlich (1983, 24 ff.), ist die Schreiber*in dazu aufgefordert, mit einer abwesenden, vielleicht sogar unbekannten Leser*in über Raum und Zeit hinweg zu kommunizieren (Becker-Mrotzek und Schindler (2007, S. 13)

automatisierter gelingt die konkrete Verschriftlichung von Gedächtnisinhalten und Gedanken (prozedurales Wissen). Durch das Reflektieren der eigenen lexikalischen Verwendungen beim Schreiben, etwa durch das Rezipieren und Vergleichen von Texten, können lexikalische Schreibprozesse auf einer metakognitiven Ebene überwacht, gesteuert und vor dem Hintergrund der Anforderungsbewältigung evaluiert werden (Becker-Mrotzek & Schindler, 2007, S. 15).

Wie hier am Beispiel verdeutlicht, setzt sich die Lexikalische Kompetenz auf ganz unterschiedlichen Ebenen aus verschiedenen Fertigkeiten und Fähigkeiten zusammen und bildet dem Modell von Becker-Mrotzek und Schindler (2007) folgend eine wichtige Teilkompetenz der hierarchisch höher angesiedelten Schreibkompetenz. Doch nicht nur im Bereich *Schreiben* spielt die Lexikalische Kompetenz eine Rolle, so ist sie beispielsweise auch ein wesentlicher Bestandteil von Lesekompetenz (für das Verstehen eines Fachtextes sind beispielsweise fach- oder themenspezifische Wortschatz-Kenntnisse notwendig). Ein solches Prinzip von Überschneidungen und Vervielfachungen von Teilkompetenzen ist jedoch kein spezifisches Merkmal der Lexikalischen Kompetenz, vielmehr lässt es sich auch auf alle anderen sprachlichen Kompetenzbereiche (z. B. auf die Orthographische Kompetenz, die Syntaktische Kompetenz, etc.) übertragen. Die übergeordneten Kompetenzen wie Schreib- und Lesekompetenz gelten damit also als eine Art Dachkompetenzen, welche teilweise auf identische sprachliche Teilkompetenzen zurückgreifen.

Die Antwort auf die Frage nach der Definition von Sprachkompetenz in Bezug auf die gymnasiale Oberstufe ist somit gewissermaßen abhängig von der Ebene von der aus der sie betrachtet wird. So ist es beispielsweise möglich, ausgehend von den Beurteilungsempfehlungen für die schriftliche Abiturprüfung (MSB NRW, 2017) Sprachkompetenz auf einer hierarchieniederen Ebene – ähnlich wie bei SkaLa® und dem C-Test – aus einer vorrangig linguistischen Perspektive (gemäß dem *GER*, s. o.) heraus zu betrachten und zu definieren. Schriftsprachliche Kompetenzen werden in diesem Zusammenhang also vor allem als lexikalische, grammatische, semantische und orthografische (Teil-)Fertigkeiten und Fähigkeiten verstanden. Dass diese basale Linguistische Kompetenz – die sich in der (Schreib-)Performanz der Oberstufenschüler*innen darin zeigt, auf einem grundlegenden Niveau normgerecht,

formalsprachlich und lexikalisch angemessen zu schreiben – nur eine Facette eines großen Sprachkompetenz-Fächers ausmacht, lässt sich auch daran erkennen, dass die Beurteilung dieser linguistischen Fähigkeiten und Fertigkeiten gerade einmal einen Anteil von 8 % der Gesamtbeurteilung ausmacht (8 von insgesamt 100 möglichen Punkten in der schriftlichen Abiturprüfung). Zur gesamten Beurteilungsgrundlage der Darstellungsperformanz auf einer hierarchiehöheren schriftsprachlich-produktiven Ebene gehören neben den linguistischen Fähigkeiten und Fertigkeiten gemäß den Beurteilungsempfehlungen auch jene schreibbezogenen (Teil-)Kompetenzen, Aufgaben angemessen zu gliedern, eigenständig und zieltextadäquat zu formulieren (kein Abschreiben des Textes), Paraphrasen oder Zitate korrekt zu kennzeichnen, etc. (MSB NRW, 2017, S. 14).

Neben der Beurteilung dieser schriftsprachlich-produktiven Fähigkeiten und Fertigkeiten – die mit 28 % zu knapp einem Drittel (28 %) in die Gesamtbewertung der schriftlichen Abiturprüfung einfließen – wird auch der Inhalt der Textproduktion als Verstehensleistung der Oberstufenschüler*innen bewertet. Damit ist die schriftliche Abiturprüfung genau wie SkaLa® eine kombinierte Lese-/Schreibaufgabe, in der Lesekompetenz zumindest indirekt miterfasst wird (vgl. Kapitel 2.1.1). Zur Lese- und Schreibkompetenz gehören damit sowohl in Bezug auf die gymnasiale Oberstufe als auch auf SkaLa® und den C-Test, jene Fähigkeiten und Fertigkeiten Strategien und Techniken anzuwenden, um Texte ganz unterschiedlicher Art selbstständig erschließen, verstehen, produzieren und reflektieren zu können (s. o.).

Allen drei Ansatzpunkten ist damit die Grundlage für das Verständnis von Sprachkompetenz als Auffassung von „Kompetenz“ als Konstituente des Begriffs „Sprachkompetenz“ gemeinsam, die sich wiederum übergreifend orientiert an der den internationalen Schulleistungsvergleichsstudien (z. B. PISA) und der Entwicklung der KMK-Bildungsstandards zugrundeliegenden, von Weinert vorgeschlagenen Definition von Kompetenz als

> „die bei Individuen verfügbaren oder durch sie erlernbaren kognitiven Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten, um die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können“ (Weinert, 2001, 27 f.).

Damit setzt sich die individuelle Kompetenz aus dem Wissen, den Fähigkeiten, dem Verstehen, Können und Handeln, der Erfahrung und der Motivation einer Person zusammen. Kompetenz wird demnach als Bindeglied zwischen Wissen und Können verstanden, welches zur Bewältigung von unterschiedlichen Situationen befähigt (Klieme, 2004, S. 12). Diese allgemeine Definition von Kompetenz spiegelt sich – übertragen auf den Begriff der *Sprach*kompetenz – auch in aktuellen Ansätzen der Sprachdiagnostik wieder (Jude, 2008, S. 15). Dabei werden sprachliche Kompetenzen in den Beschreibungen des *GER* als „Summe des (deklarativen) Wissens, der (prozeduralen) Fertigkeit und der persönlichkeitsbezogenen Kompetenzen und allgemeinen kognitiven Fähigkeiten, die es einem Menschen erlauben, Handlungen auszuführen“ und „Menschen zum Handeln mit Hilfe spezifischer sprachlicher Mittel“ zu befähigen (Council of Europe, 2001, S. 21; Jude, 2008, S. 15).

Vor diesem Hintergrund wird nun auch in der vorliegenden Arbeit der Begriff Sprachkompetenz als Teil einer individuellen Kompetenz verstanden, die wiederum als Disposition einer Person, bestimmte – möglichst konkrete – Anforderungssituationen zu bewältigen anzusehen ist, die sich in der Performanz in solchen Situationen äußert und damit an einer tatsächlichen (möglichst konkreten) Leistung messbar gemacht werden kann. Sprachkompetenz bildet somit ein komplexes Phänomen, welches eine Fülle von produktiven und rezeptiven Fähigkeiten und Fertigkeiten, die zur mündlichen und schriftlichen, rezeptiven und produktiven Verständigung und Handlung befähigen. Im Rahmen dieser Arbeit kann und soll dieses Phänomen nicht annähernd vollständig abgebildet werden, vielmehr sollen im Kontext der hier vorgestellten Untersuchung Lese- und Schreibfertigkeiten aus einer linguistischen Perspektive heraus als Facetten von Sprachkompetenz betrachtet werden. Unter Lese- und Schreibkompetenz werden in diesem Zusammenhang und in Anlehnung an die Überlegungen zum Sprachkompetenzbegriff bei SkaLa®, dem C-Test und in der Oberstufe vor allem Fertigkeiten und Fähigkeiten verstanden werden, Texte auf lexikalischer, morpho-syntaktischer, semantischer Ebene zu erschließen und (formalsprachlich korrekt) zu produzieren, selektiv, global und detailliert zu lesen, Verstehensprobleme selbstständig zu lösen sowie Lese- und Schreibprozesse zu reflektieren (s. o.).

Wie sich diese Kompetenzen nun operationalisieren, d. h. in empirisch fassbare Konstrukte und damit in konkrete Testformate und Aufgaben umsetzten lassen, soll im nächsten Abschnitt betrachtet werden.

2.2 Operationalisierung von Sprachkompetenz in der Oberstufe

2.2.1 Operationalisierung von Sprachkompetenz durch Noten

Prinzipiell bezieht sich die Leistungsbewertung im Fach Deutsch für die gymnasiale Oberstufe in ihren Grundzügen auf die Bestimmungen des Schulgesetzes für das Land NRW (§48), die Ausbildungs- und Prüfungsordnung für die Sekundarstufe II (§§ 13-17 APO-GOSt) und den Kernlehrplan Deutsch für die Oberstufe (MSB NRW, 2014). In Bezug auf das Fach Deutsch wird die Beurteilung von sprachlichen Kompetenzen und damit die Operationalisierung von Sprachkompetenz in der Regel durch die Vergabe von Punkten bzw. Noten geleistet. Über die Deutschnote wird Sprachkompetenz im weiteren Sinne (vgl. Kapitel 2.1) insofern (mit-)erfasst, als dass von Schüler*innen sowohl in mündlicher als auch in schriftlicher Form im Unterricht eingebrachte Leistungen in Form von Darbietungen und Beiträgen, die den im Kernlehrplan ausgewiesenen Kompetenzbereichen „Rezeption“ und „Produktion“ entsprechen, (MSB NRW, 2014, S. 16) bewertet werden:

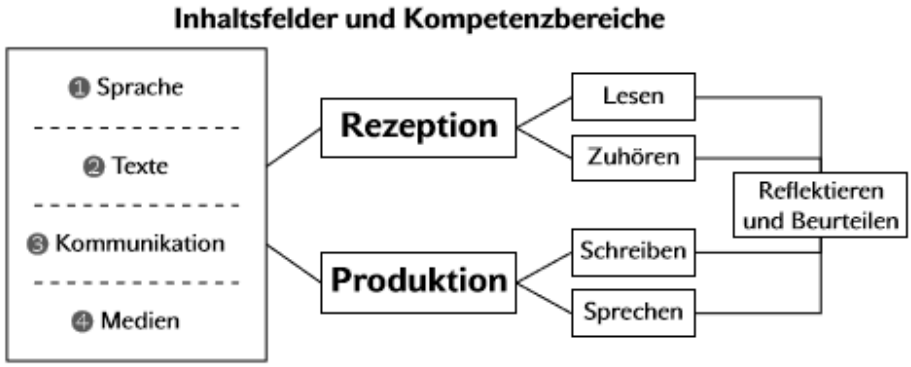


Abbildung 3: Inhaltsfelder und Kompetenzbereiche zu „Rezeption“ und „Produktion“ für das Fach Deutsch in der Sekundarstufe II

Quelle: MSB NRW, 2014, S. 16.

Die Zeugnisnote Deutsch, d. h. die Kursabschlussnote für das Fach am Ende des Schuljahres setzt sich dabei zu gleichen Teilen aus den Klausurnoten der schriftlichen Arbeiten und aus den Leistungen aus dem Beurteilungsbereich „Sonstige Mitarbeit“ zusammen. Zur Sonstigen Mitarbeit zählen gemäß des Kernlehrplanes Deutsch für die gymnasiale Oberstufe alle im Unterricht selbstständig und kooperativ erbrachten schriftlichen, mündlichen und praktischen Leistungen (ausgenommen der Klausuren und der Facharbeit), die beispielsweise in Form von Präsentationen, Referaten, Vorträgen, Protokollen, Gesprächsbeiträgen im Unterricht, Portfolios, Recherchearbeiten etc. dargeboten werden können (MSB NRW, 2014, 40 ff.).

Die Operationalisierung von (schrift-)sprachlichen Kompetenzen im engeren Sinne – um die es in dieser Arbeit vorrangig gehen soll (vgl. Kapitel 2.1) – findet also vor allem in der Beurteilung und Bewertung von Klausuren statt. Im Rahmen dieser schriftlichen Arbeiten werden Punkte auf der Grundlage von Beurteilungsrastern bzw. Erwartungshorizonten vergeben. Die Erwartungshorizonte beschreiben die der jeweiligen Klausuraufgabe entsprechenden, zu erwarteten Verstehens- und Darstellungsleistung sowie die Angaben von Merkmalen zur kriterienorientierten Bewertung der Verstehens- und Darstellungsleistung (KMK, 2012b, 23; 26). In der Verstehensleistung werden neben sachlicher Richtigkeit, der Vielfalt und Vollständigkeit inhaltlich zu nennender Aspekte und deren funktionaler Bedeutsamkeit auch die Logik und Begründung der schriftlich dargelegten Aussagen sowie ein methodisch sicherer und selbstständiger Umgang mit der Bewältigung der Aufgabe bewertet. Ein gedanklich klarer Aufbau der Textproduktion sowie eine zielgerechte und sinngebende Struktur, die Anwendung von korrekter Fachsprache und Zitation, ein angemessener allgemeinsprachlicher Ausdruck, Stil und Satzbau sowie formalsprachliche Richtigkeit sind hingegen Merkmale, die in die Beurteilung der Darstellungsleistung miteinfließen (ebd.). Zur Bewertung der gesamten Prüfungsleistung wird die Verstehensleistung mit 60 % bis 70 %, die Darstellungsleistung mit 30 % bis 40 % gewichtet (IQB, S. 2). Die Summe der erreichten Punkte (oder der erzielten Leistung in Prozent) wird nach einem Punkte-Noten-Raster den Schulnotenstufen 1-6 zugeordnet und bildet so die Klausurnote.

Die Zusammensetzung der Klausurnote und mehr noch der Zeugnis- bzw. Kursabschlussnote im Fach Deutsch ist folglich so angelegt, dass sie in komplexer Weise Bezug auf die unterschiedlichen Kompetenzbereiche des Faches nimmt und somit eine Vielzahl miteinander in Verbindung stehender Kompetenzen (s. o.) erfasst (IQB). Die Deutschnote bildet damit eine kumulative Summe aus vielen verschiedenen Fähigkeiten und Fertigkeiten, aus der Sprachkompetenz im oben definierten Sinne als einzelne Facette dieses komplexen Kompetenzfächers nicht ohne Weiteres oder nur indirekt herauspräpariert werden kann. Dies führt nun zu der Frage, wie es gelingen kann, Sprachkompetenz im engeren Sinne gezielter und expliziter als über die Deutschnote in der Oberstufe zu erfassen und zu messen.

2.2.2 Standardisierte Sprachtests in der Oberstufe

Neben der eher indirekten Beurteilung des sprachlichen Lernstandes von Schüler*innen durch das Lehrerurteil, bzw. konkreter durch die Deutschnote, können schriftsprachliche Fertigkeiten und Fähigkeiten in der Oberstufe auch durch Sprachtests messbar gemacht werden. Standardisierte Sprachtests, d. h. Tests die den klassischen Gütekriterien standhalten und für die eine Normierung vorliegt, gibt es für den Bereich der gymnasialen Oberstufe kaum bzw. nur den Lesegeschwindigkeits- und Verständnistest für die Klassenstufen 6-12 (LGVT 6-12) – bzw. die seit 2017 verfügbare, erweiterte Auflage für die Klassenstufen 5 bis 12+ (LGVT 5-12+) – zur differenzierten Ermittlung von Lesegeschwindigkeit, Lesegenauigkeit und Leseverständnis in den Klassen 6 bis 12 bzw. 5 bis 13 aller Schulformen (Schneider et al., 2017) –,auf den in Kapitel 4.1.1 ausführlicher eingegangen wird. Standardisierte Sprachtests werden in der Sekundarstufe II an den Schulen jedoch nicht systematisch eingesetzt. Eine systematische Beurteilung von Sprachkompetenz erfolgt – wie oben beschrieben – nur in indirekter Weise über die Deutschnote: Sprachkompetenz ist hier Teil einer kumulativen Summe die unterschiedlichen Kompetenzbereiche des Fachs und kann aus dieser nicht separat herausgegriffen werden (s. o.; vgl. Kapitel 2.1). Dabei scheint eine gesonderte Beurteilung der schriftsprachlichen Kompetenzen ergänzend zu den eher allgemeinen Rückmeldungen zur sprachlichen Kompetenz im Rahmen von

Klausurkorrekturen besonders sinnvoll, da über angemessene schriftsprachliche Fertigkeiten zu verfügen, eine Schlüsselkompetenz von hoher gesellschaftlicher Relevanz ist: Von besonderer Bedeutung sind diese Fähigkeiten und Fertigkeiten als Voraussetzungen für den Ausbildungs- und Studienerfolg an der Universität sowie für den Erfolg im Beruf. Mit dem Wechsel in die entsprechenden Berufe sind es auch sprachliche Fähigkeiten und Fertigkeiten, d. h. die „Beherrschung der Muttersprache in Wort und Schrift", welche in „in allen modernen Gesellschaften zum Kernbestand kultureller und sprachlicher Literalität [gehört]" (Naumann et al., 2010, S. 23), die über Karrieren entscheiden und eine Facette des Bildes der Ausbildungsgüte an Schule und Hochschule vermitteln. Im Rahmen dieser Studie soll Sprachkompetenz also als transversale Schlüsselkompetenz in den Blick genommen und mit einem der Zielkohorte dieser Arbeit geeigneten Testinstrument global gemessen und erfasst werden.

Ein standardisierter Sprachtest, welcher erstmals zu Beginn der 80er Jahre von Raatz und Klein-Braley (1981) vorgestellt wurde und bis heute als ökonomisches, objektives und reliables und Testinstrument zur globalen Erfassung von allgemeiner Sprachkompetenz überzeugt, ist der C-Test (Grotjahn, 2002b, S. 211). Ursprünglich für den Einsatz im fremdsprachlichen Bereich[20] konzipiert, wird der C-Test heutzutage in adaptierter Form auch zur Messung von sprachlichen Kompetenzen von Zweit- und Erstsprachsprecher*innen eingesetzt und gilt inzwischen als einer der am besten erforschten Sprachtests weltweit (ebd.). Im deutschsprachigen Raum konnte sich der C-Test als fester Bestandteil in der Testlandschaft etablieren: So wird er u. a. als Online-Einstufungstest für Deutsch als Fremdsprache für ausländische Studienbewerber (onSET, ehemals onDaF) (Eckes, 2010, 2014), als Test- und Förderinstrument für Deutsch als Zweitsprache für Kinder, Jugendliche und Erwachsene (Baur et al., 2006; Baur & Meder, 1994; Baur & Spettmann, 2007, 2008a, 2008b, 2009; Linnemann & Wilbert, 2014; Süßmilch, 1985), zur Erfassung von Kompetenzen in großangelegten Studien zum Bildungsmonitoring (z. B. „Deutsch English Schülerleistungen International – DESI" oder „Kompetenz und Einstellung von Schüler*innen - KESS") und zur Überprüfung von erstsprachlichen Kompetenzen eingesetzt (Bremerich-Vos & Scholten-Akoun, 2016;

[20] Vor allem im Bereich *Englisch als Fremdsprache* (vgl. Darnell (1986); Oller (1973); Oller und Conrad (1971); Taylor (1953))

Mashkovskaya, 2014; Wockenfuß, 2009; Wockenfuß & Raatz, 2006). Besonders durch die Eignung zur Messung von basalen erst- und zweitsprachlichen Fähigkeiten und Fertigkeiten und die Erfüllung der klassischen Gütekriterien[21] sowohl beim Einsatz von standardisieren C-Tests im schulischen Bereich – wie beispielsweise in den Untersuchungen von Baur et al. (Baur et al., 2012, S. 14; 2013a, S. 2) – als auch im Bereich der Studieneingangsphase – hier vor allem bei SkaLa® (vgl. Kapitel 1 und Kapitel 2.1.1, Mashkovskaya & Baur, 2016a) – ist der Einsatz des C-Tests nun auch in dieser Arbeit als Diagnoseinstrument zur systematischen und standardisierten Erfassung von Sprachkompetenz in der Oberstufe sinnvoll. Wenn also davon ausgegangen wird, dass mit dem C-Test Sprachkompetenz operationalisiert werden kann, stellt sich die Frage, welche theoretischen Grundlagen dem Konstrukt des C-Tests zugrunde liegen, nach welchem Prinzip er funktioniert und wie das nun mit der Messung von Sprachkompetenz zusammenhängt.

2.3 Zur Erfassung von Sprachkompetenz mit dem C-Test

Um also nun die Frage zu beantworten, wie es dem C-Test gelingt, Sprachkompetenz zu erfassen, soll im Folgenden zunächst ein Blick auf die Funktionsweise und das Funktionsprinzip des Tests geworfen werden, bevor anschließend auf Modifikationen bzw. unterschiedliche Versionen des C-Tests eingegangen wird.

2.3.1 Vom Cloze-Test zum C-Test

Als Vorläufer des C-Tests gilt der Cloze-Test, der erstmals 1953 von Taylor erwähnt wird. Cloze-Tests waren zur damaligen Zeit als Testformat keine neue Erfindung, da sie vom Format und Aufgabenprinzip her einfachen Lückentexten ähneln. Das Wort *cloze* im namengebenden Testformat ist eine Ableitung des englischen Wortes *closure* und kann

[21] Welche klassischen Gütekriterien ein standardisierter C-Test erfüllt/ erfüllen sollte, wird ausführlich in Kapitel 2.4 beschrieben.

in nominalisierter Form[22] mit *Schließen, Abschließen, Zusammenschluss, Beendigung* und im weitesten Sinne auch mit *Erschließen* übersetzt werden. Im wissenschaftlichen Gebrauch findet der Begriff *closure* insbesondere Verwendung in der Gestaltpsychologie und beschreibt dort die Fähigkeit der Gestaltergänzung. Anhand der unten stehenden Abbildung soll verdeutlicht werden, was damit gemeint ist: Das Dreieck in Figur A, die Säule in Figur B, der Ball in Figur C und das „schwimmende Seeungeheuer" in Figur D entstehen allein durch die menschliche Fähigkeit, unvollständige Strukturen und Gestalten zu ergänzen und diese als Sinneseindruck einer Form wahrzunehmen.

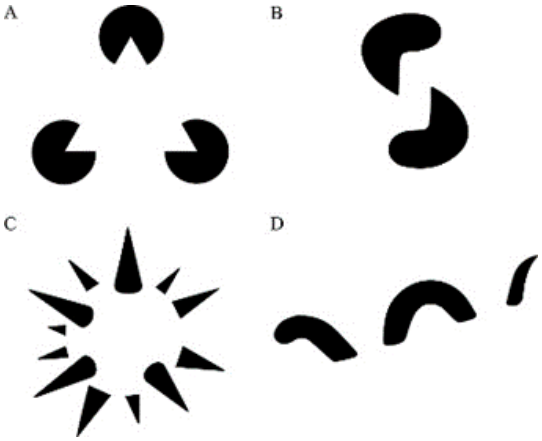


Abbildung 4: Beispiele zur Fähigkeit der Gestaltergänzung.

Quelle: Lehar, Steven 2012

Diese Fähigkeit, fehlende Informationen gedanklich zu rekonstruieren und zu einem Ganzen zu ergänzen, kann in ihrer Ausprägung von Individuum zu Individuum zu variieren (McKamey, 2006). Taylor (1953) überträgt das Prinzip der Gestaltergänzung auf den Bereich Sprache und argumentiert, dass die Fähigkeit, fehlende Wörter in einem Text zu ergänzen, indem der Text als Ganzes wahrgenommen und verstanden werden muss, bei verschiedenen Personen unterschiedlich ausgeprägt sei (McKamey, 2006, S. 151; Taylor, 1953). Dieses Vorgehen nennt Taylor „cloze procedure" und beschreibt damit den Prozess, dass „one must guess what the mutilated sentence means as a whole, then complete its pattern to fit that whole meaning" (Taylor, 1953, S. 416).

[22] Das Wort *closure* existiert in der englischen Sprache auch als Verb und wird meist im politischen Kontext verwendet mit der Bedeutung *eine Debatte beenden/schließen*. Weitaus geläufiger ist das Verb *to close* in der oben genannten Bedeutung.

Für die Konstruktion eines Tests, der nach dem Prinzip der „cloze procedure“ funktionieren soll, benutzt Taylor (1953) eine Verfahrenstechnik, bei der zufällig jedes n-te Wort in einem Text (Turnus zwischen dem 5. und 10. Wort in einem Text) gelöscht wird. Diese zufällige Auswahl von Wortlöschungen im Text soll dabei Störungen innerhalb der natürlichen Kommunikation simulieren (Klein-Braley, 1985a; Taylor, 1953, S. 419).

Taylor (1953) versucht mit dieser „cloze procedure“ zunächst die Lesbarkeit („readability“) bzw. Schwierigkeit journalistischer Texte einzuordnen, indem er Proband*innen von ihm gelöschte Wörter rekonstruieren lässt (Wockenfuß, 2009). Als richtige Lösung werden nur diejenigen Rekonstruktionen akzeptiert, die den exakten Wörtern im Originaltext entsprechen. Die Schwierigkeit eines Textes wird dann auf der Grundlage der Aufgabenauswertung interpretiert: Kann der Großteil der Proband*innen die Lücken im Text korrekt ergänzen, ist das für Taylor (1953) ein Indiz für einen leichten Text, können nur wenige Lücken rekonstruiert werden, wird der Text als schwierig eingestuft (McKamey, 2006, S. 116).

In seinem zweiten Aufsatz führt Taylor (1956) fort, dass es mit der „cloze procedure“ bzw. dem Cloze-Test nicht nur möglich sei, Aussagen über die Schwierigkeit eines Textes zu treffen, sondern, dass der Cloze-Test auch das Potential habe, das Leseverstehen („reading comprehension“) von Proband*innen in der Erstsprache einzustufen. Er argumentiert, dass „if the statement that a passage is ‘readable’ means that it is ‘understandable’, then the scores that measure readability should measure comprehension too” (McKamey, 2006, S. 116; Taylor, 1956, S. 44). Taylors Auffassung, dass sich der Cloze-Test zur Messung von Leseverstehen von Erstsprachsprecher*innen eignet (1956), wird in den folgenden Jahren von einigen Autoren und Forschern (Bormuth, 1967, 1969; Rankin & Culhane, 1969; Weaver & Kingston, 1963) anhand von kriteriumsbezogenen Validitätsstudien empirisch untersucht. Zusammenfassend können in einer Reihe dieser Studien Korrelationen von moderaten (0,5-0,7) bis guten Werten (0,8-0,95) zwischen Cloze-Tests und anderen Außenkriterien zur Überprüfung von Lesefähigkeiten nachgewiesen werden (ebd.).

Es dauert nicht lange, bis das Cloze-Test-Format auch für den Fremdsprachenbereich nutzbar gemacht wird (Darnell, 1986; Oller & Conrad, 1971; Oller & Inal, 1971). Oller

zweifelt allerdings schon früh daran, dass der ursprüngliche von Taylor (1953) vorgeschlagene Auswertungsmodus der Einschätzung der Lesekompetenz von Fremdsprachenlerner*innen gerecht werden würde (Oller, 1972, 1973, 105 ff.; Oller & Inal, 1971). Dass nur Rekonstruktionen im Originaltext vorkommender Wörter als korrekte Lösungen akzeptiert werden sollten, würde als Auswertungsregel sprachlich und inhaltlich sinnvolle Alternativlösungen übersehen, in denen sich ja ebenfalls kompetente Sprachverwendung ausdrückte. Er schlägt vor, nicht nur die exakt im Lösungstext enthaltenen Wörter als korrekte Lösung zu akzeptieren, sondern jedes inhaltlich sinnvolle, in den Textzusammenhang passende und grammatikalisch richtige Wort (Oller, 1972, 151 ff.) Dieses Verfahren sichert in der Auffassung Ollers (ebd.), dass tatsächlich Lese- und Sprachfähigkeiten und nicht – im Extremfall – die Kenntnis des Ausgangstexts über das Ergebnis des Tests entscheiden. Die Validität des Tests wird laut Oller auf diese Weise mitunter deutlich verbessert (ebd.). Empirisch gesehen stützen sich diese Aussagen auf die Resultate von kriteriumsbezogenen Validitätsstudien von Oller und Conrad (1971), die 1980 von Hinofortis ergänzt werden und belegen sollen, dass der Cloze-Test eben nicht nur Leseverstehen, sondern auch Fähigkeiten, die über die Ebene des Leseverstehens hinausgehen, miterfasst: Den Forschern gelingt es in ihren Studien nachzuweisen, dass der Cloze-Test nicht nur mit Tests zur Erfassung von Leseverstehen (0,8 bei Oller und Conrad und Hinofortis) sondern auch mit Tests zum Hörverstehen (0,71, Hinofortis), Tests zur Überprüfung von grammatischen Strukturen und grammatikalischen Wissen (0,58 bei Oller und Conrad, 0,63 bei Hinofortis), Vokabeltests (0,59 bei Oller und Conrad und Hinofortis) und mit dem Diktat (0,82 Oller und Conrad) korreliert.

Diese Ergebnisse veranlassen u. a. Oller und Conrad (1971) dazu, Cloze-Tests als „integrative“ Tests zu bezeichnen, die sich dadurch auszeichnen, dass die Proband*in verschiedene sprachliche Fähigkeiten zur gleichen Zeit abrufen muss, um die vorgelegte Aufgabe zu bewältigen. Oller betont, dass

> „integrative tests attempt to assess a learner's capacity to use many bits all at the same time, and possibly while exercising several presumed components of a grammatical system, and perhaps more than one of the traditional skills[23] or

[23] Als „traditional skills“ sind hier Fähigkeiten in den Bereichen Lesen, Hören, Schreiben und Sprachen zu verstehen.

> aspects of skills… [A]ny procedure or task that causes the learner to process sequences of elements in a language that conform to the normal contextual constraints of that language and which requires learner to relate sequences of linguistics elements via pragmatic mappings to extra linguistic contexts" (Oller, 1979, S. 37).

Er argumentiert in diesem Zusammenhang, dass der Cloze-Test als integrativer Test nicht nur einzelne, isolierte Elemente von Sprache (wie z. B. nur Leseverstehen), sondern „general language proficiency", also eine Art allgemeine Sprachfähigkeit, messe (ebd., S. 37) (vgl. Kapitel 2.3.2 und 2.3.3 zur Funktionsweise des C-Tests und zum Zusammenhang zur Messung von allgemeiner Sprachkompetenz mit dem C-Test).

Wesentlich radikaler und umfassender als Oller kritisieren vor allem Klein-Braley (1981) und Alderson (1979, 1980) das Konzept des Cloze-Tests auf der Grundlage eigener Untersuchungen im fremdsprachlichen und später auch im erstsprachlichen Bereich, wobei Klein-Braley in ihrer Dissertation nicht die theoretische Grundlage des Cloze-Prinzips in Frage stellt, wohl aber einige Aspekte des Test-Designs und der Testdurchführung (1981). Zusammenfassend werden folgende Hauptkritikpunkte von beiden Autoren genannt:

- Cloze-Tests korrelieren weitaus höher mit Teilfertigkeitstest aus den Bereichen Grammatik und Wortschatz als mit Aufgaben zur Überprüfung des Leseverstehens. Es stellt sich die Frage, ob ein einziger Cloze-Test eine Kombination aus mehreren Teilfertigkeitstests überhaupt ersetzen kann.

- Die Art der Auswertung, d. h. ob nur das im Originaltext vorkommende Wort oder alle semantisch und grammatikalisch korrekten Wörter als Lösung akzeptiert werden, hat Auswirkungen sowohl auf die Schwierigkeit, die Auswertungszeit als auch auf die Objektivität und somit auf die Reliabilität und Validität des Tests. Wenn die Tests nach der „exakten" Auswertungsmethode bewertet werden, sind sie selbst für Erstsprachsprecher*innen oft zu schwer.

- Die Tests bestehen aus nur einem Text bzw. einem Textauszug. Problematisch daran ist, dass Probanden, die mit der Thematik des Textes nicht vertraut sind einen Nachteil gegenüber Proband*innen haben könnten, die sich inhaltlich gut mit dem Thema auskennen. Wenn also das Vorwissen der Proband*innen bei der Bearbeitung des Tests eine entscheidende Rolle spielt, wird mit dem Test nicht mehr Sprachkompetenz, sondern Weltwissen überprüft, was wiederum eine ernsthafte Bedrohung für die Validität bedeuten dürfe.

- Wird nur ein Textausschnitt als Grundlage für den Tests verwendet, besteht das Risiko, dass die Tilgungen der Wörter keine repräsentative Stichprobe der Wortarten darstellen.

- Für die Berechnung von Reliabilitätswerten gilt: Je mehr homogene Items in einem Test vorhanden sind, desto höher fällt die Reliabilität aus. Um hohe Reliabilitätswerte in einem Cloze-Test zu erreichen, müssen ausreichend viele Lücken in den Text eingebaut werden, was zur Folge hat, dass der zu bearbeitende Text insgesamt sehr lang sein muss.

- Die Wahl des ersten getilgten Wortes im Text hat großen Einfluss auf die Schwierigkeit des erstellten Tests. Wird bereits im ersten Satz das erste Wort getilgt, ist der Test tendenziell schwieriger zu lösen als wenn der erste Satz tilgungsfrei bleibt und der Leser*in mehr Kontext und einen leichteren Einstieg in den Text bietet (Klein-Braley, 1981).

Raatz (1985, S. 14) und Klein-Braley (Klein-Braley, 1985a; Raatz & Klein-Braley, 1983, 1985) formulieren folgende Anforderungen an eine Überarbeitung des Testformats, um der formulierten Kritik Rechnung zu tragen:

- Der Test sollte nicht aus nur einem Textauszug sondern aus mehreren inhaltlich unterschiedlichen Texten bestehen.

- Für eine aussagekräftige Reliabilitätsberechnung sollte der Test mindestens 100 Items enthalten. Besteht der Test nicht aus einem sondern aus mehreren Texten, kann in der Summe der Texte eine ausreichend große Anzahl an Items generiert werden, wobei die Texte wesentlich kürzer ausfallen können.

- Der Anfangspunkt sowie der Rhythmus der Wortlöschungen im Text sollten im Vorhinein festgelegt sein, um persönliche Präferenzen und unterschiedliche Schwierigkeitsgrade zu vermeiden.

- Die getilgten Elemente sollten eine repräsentative Stichprobe aller Wortarten gewährleisten.

- Es sollte gebildeten, erwachsenen Erstsprachsprecher*innen gelingen, den Text ohne Beschränkung der Bearbeitungszeit weitgehend komplett zu rekonstruieren.

- Zur Erreichung einer maximalen Objektivität sollte nur eine exakte Auswertungsmethode zulässig sein.

- Der Test sollte hinreichende Werte im Bereich der Validität und Reliabilität erzielen.

- Seine Konstruktion sollte einfach und ökonomisch sein.

Auf der Grundlage dieser Kriterien wird der Cloze-Test nun von Raatz und Klein-Braley (Klein-Braley, 1985a; 1983, 1985) weiterentwickelt. In der Konstruktion des Tests beschreiten die beiden einen ganz neuen Weg: Im Gegensatz zum Cloze-Test, in dem ganze Wörter getilgt werden, schlagen sie vor, immer nur einen Teil eines Wortes zu löschen. Diese Veränderung ist namengebend für die neue Testversion, den C-Test, da von „Cloze“ lediglich das „C“ bestehen bleibt (ebd.).

Die Tilgung von nur einer Worthälfte und nicht dem Wort als Ganzes erweißt sich als Vorteil, da sie die Auswertung der einzelnen Items sehr viel präziser und objektiver macht (ebd.). Die Möglichkeit alternativer, d. h. vom Originaltext abweichender aber als korrekt akzeptierter Lösungen, ist durch die vorgegebenen Wortanfänge deutlich eingeschränkter, wenngleich alternative Lösungen vorkommen können und durch Pretests weiter einzugrenzen sind[24] (Grotjahn, 1987, 219 ff.; Raatz & Klein-Braley, 1985, S. 20). Durch das Tilgen einer Worthälfte – im Gegensatz zum ganzen Wort im Cloze-Test – ist sowohl das Lösen als auch die Korrektur von C-Tests – je nach Auswahl der Texte – für erwachsene Erstsprachsprecher*innen möglich und einfacher als bei den Cloze-Tests. Die stehenbleibenden Wortteile geben bei der Rekonstruktion zum einen Hinweise auf das zu ergänzende Wort, zum andern grenzen sie die Möglichkeiten einer orthographisch, syntaktisch und semantisch korrekten Ergänzung stark ein (Grotjahn, 1987, 219 ff.) (vgl. auch Kapitel 2.3.5).

Neu am C-Test ist auch, dass er im Gegensatz zum Cloze-Test nicht aus einem langen Text sondern aus 4-5 kürzeren Texten mit unterschiedlichen Themen (zur Ausschaltung des Weltwissen-Bias, s. o.) besteht, in denen jeweils 20 – 25 Items (Lücken) eingebaut werden[25]. So kann gewährleistet werden, dass eine C-Testbatterie, auch *Testset* oder *Testlet* genannt, insgesamt mindestens 100 Items (was als notwendige Anzahl für die Berechnung von beispielsweise Reliabilitätswerten gilt, s. o.) enthält (Grotjahn, 2019b, 585 ff.; Raatz & Klein-Braley, 1985, S. 20).

[24] Zum Verfahren zur Erfassung von Alternativlösungen in Kapitel 4.

[25] In der Regel bestehen C-Tests aus entweder 4 Texten mit jeweils 25 Lücken oder aus fünf Texten mit jeweils 20 Lücken. Dies hat den Vorteil, dass das Gesamtergebnis bei einer Maximalpunktzahl von 100 Punkten (bei 100 Items) Prozentwerten gleichgesetzt werden kann.

Raatz und Klein-Braley (1985) sowie Grotjahn (1987, 2002a) schlagen außerdem vor, den ersten und den letzten Satz eines C-Tests tilgungsfrei zu lassen (zur Erweiterung des Kontextes, s. o.) und ab dem zweiten Satz jedes 2. Wort im Text zu tilgen. Dieses Tilgungsprinzip wird *klassisch* oder *kanonisch* genannt oder auch als *2er-Tilgung* beschrieben (ebd.)[26]. Eine Tilgung in diesem Rhythmus soll die Wahrscheinlichkeit erhöhen, dass nicht nur bestimmte Wortklassen sondern theoretisch alle Wortarten bei der Tilgung berücksichtig werden (Raatz & Klein-Braley, 1985, S. 20).

Durch diese Modifikationen und die Weiterentwicklung des Cloze-Tests gelingt es Raatz und Klein-Braley (1981, 1983, 1985) den C-Test erfolgreich als einen Sprachtests vorzustellen, der integrativ auf Sprachkompetenz zuzugreifen in der Lage ist und sich aus diesem Grund im Laufe der Jahre als Verfahren zur Feststellung der „general language proficiency“ (Oller, 1983, 4 ff.), also der globalen sprachlichen Kompetenz eines Individuums (Grotjahn, 2002a, S. 211) immer stärker etabliert hat. Doch wie genau gelingt es dem C-Test nun, diese allgemeine Sprachkompetenz zu erfassen und messbar zu machen? Einen wesentlichen Teil der Antwort liefert das dem C-Test zugrunde liegende Prinzip der Redundanz, welches als Zentralbegriff für die Funktionsweise des C-Tests im folgenden Kapitel erläutert werden soll.

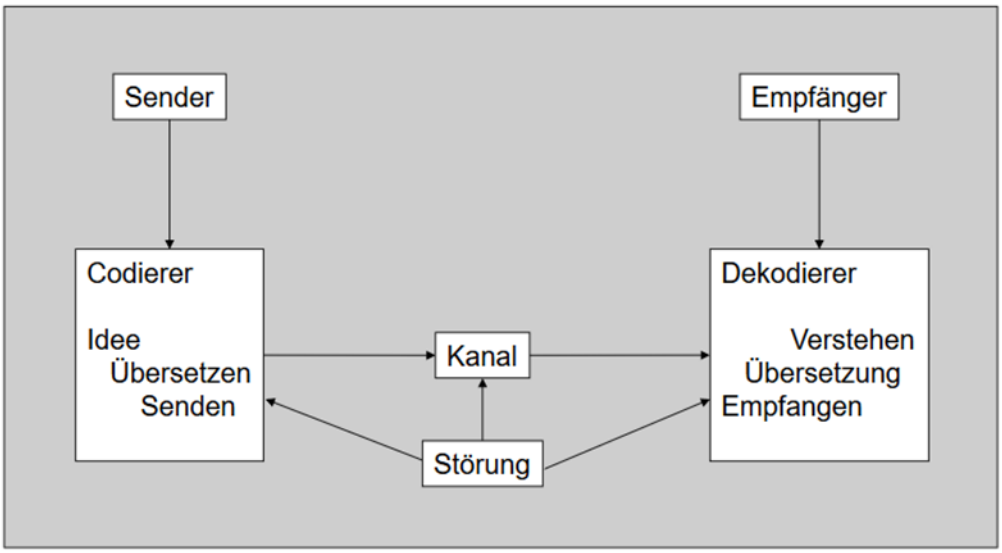
2.3.2 Redundanz als Zentralbegriff für das Funktionsprinzip von C-Tests

Der Begriff der Redundanz wird erstmals von Nyquist (2002) im Rahmen der Informationstheorie eingeführt und beschreibt die „überflüssigen“ Bestandteile in einer Nachricht, die weggelassen werden können, ohne dass Information verloren geht (Grabher, 1994, 19 ff.). Auch Nyquists Kollegen Shannon und Weaver, die Gründer der Informationstheorie, folgen dieser Definition von Redundanz:

> „It is sensibly called redundancy for this fraction of the message is in fact redundant in something close to the ordinary sense; that is to say, this fraction of the message is unnecessary […] in the sense that if it were missing the message would still be essentially complete” (Shannon & Weaver, 1949, S. 13).

[26] Es gibt mittlerweile eine Reihe von Modifikationen und Varianten des kanonischen C-Test-Formats auf die ausführlicher in Kapitel 2.3.4 eingegangen wird.

Hiermit beschreiben die beiden Kommunikationstheoretiker einen Befund, den sie bei ihren Untersuchungen für die amerikanische Telefongesellschaft Bell feststellen: Verbale menschliche Kommunikation erweist sich als erstaunlich robust gegenüber Störungen bei der Übertragung – theoretisch gesprochen: gegengenüber Störungen im Kanal.



Sender-Empfänger Modell von Shannon & Weaver (1949)

Abbildung 5: Sender- Empfänger Modell von Shannon & Weaver

Quelle: Traut-Mattausch und Dieter Frey 2006, S. 536

Bezogen auf das Sender-Empfänger-Modell von Shannon und Weaver (1949) kann bei der Übermittlung von Nachrichten Informationsverlust dann auftreten, wenn der Informationstransfer gestört wird (Traut-Mattausch & Dieter Frey, 2006, 538 f.). Eine solche Störung – auch „noise“ genannt – kann verschiedene Ursachen haben und den Informationstransfer innerhalb der Kommunikation beeinträchtigen (ebd.). Ein schlechter Handyempfang etwa kann bei einem Telefonat dazu führen, dass die vom Sender übermittelte Information nur unzureichend vom Empfänger verstanden wird. Dieser versucht dann, zunächst aus den erhaltenen Informationen die ursprüngliche Nachricht zu rekonstruieren oder – wenn das nicht funktioniert – durch Nachfragen die fehlenden Informationen zu ergänzen. Kommunikation im alltäglichen Leben wird in fast allen Bereichen von Störungen beeinflusst: dies kann eine unleserliche Handschrift sein, laute Hintergrundgeräusche, Unterbrechungen im Gespräch, etc. (Mashkovskaya,

2014). Erstaunlicher Weise müssen solche Störungen jedoch äußerst massiv sein, um den Kommunikationsprozess maßgeblich zu behindern oder gar abzubrechen. Denn oftmals reicht das Verstehen von bruchstückhaften Bestandteilen oder Äußerungen, um den Inhalt einer Nachricht weitgehend rekonstruieren zu können (Traut-Mattausch & Dieter Frey, 2006, 538 f.). Diese Stör- bzw. Fehlerresistenz in der Informationsübertragung resultiert aus dem von Shannon und Weaver (1949) erwähnten Redundanzprinzip, d. h. aus den Bestandteilen einer Nachricht, die zwar vorhanden sind, deren Wegfallen jedoch keinen Informationsverlust bewirken. Redundanz wird demzufolge nicht als Überfluss im eigentlichen Sinne verstanden, sondern als eine Art erforderliche Wiederholung oder Mehrfachnennung in der Informationsübertragung, welche in der Realität in der Regel störenden Einflüssen unterliegt. „Redundante Informationen gehen über das für die Empfänger-Verständlichkeit absolute Minimum hinaus und erhöhen damit die Sicherheit der Übertragung“ (Grabher, 1994, S. 19). Dadurch ist Informationsübertragung im Sinne von Kommunikationsprozessen enorm unempfindlich und resistent gegenüber Störeinflüssen. Doch was begründet diese erstaunlich hohe Robustheit der menschlichen Kommunikation?

Ein wesentlicher Teil der Antwort ist das Redundanzprinzip menschlicher Sprache: Jede sprachliche Nachricht kann als redundant angesehen werden, weil die in ihr vorkommenden Zeichen (Buchstaben, Silben, Wörter, etc.) aus einem begrenzten Repertoire (z. B. aus dem Alphabet) entnommen werden und sich unterschiedlich häufig wiederholen. Die Zusammensetzung der Zeichen folgt orthographischen und syntaktischen Regeln und tritt mit einer relativen Wahrscheinlichkeit in einer bestimmten seriellen Verbindung auf (Mittenecker, 2017). Innerhalb eines Wortes, eines Satzes oder eines Textes sind durch bestimmte semantische, syntaktische und pragmatische Zusammensetzungen sich auf verschiedenen Ebenen wiederholende Informationen vorhanden.

Das folgende Beispiel soll verdeutlichen, dass Sprache auf verschiedenen Ebenen redundant ist:

„€s i?t nohc ke?n Meiste.. v_m Himel gefln.“ [27]

Abbildung 6: Beispielsatz zum Redundanzprinzip

Anzunehmen ist, dass erwachsene bzw. fortgeschrittene Erstsprachsprecher*innen den oben stehenden Satz nach eingehender Betrachtung lesen können. Die Rekonstruktion des Satzes kann auf der Buchstabenebene gelingen, da der deutschen Sprachstruktur Regel- und Gesetzmäßigkeiten unterliegen, die eine gewisse Auftretenswahrscheinlichkeit mit sich bringen und eine Decodierung ermöglichen bzw. erleichtern. So kann das Wort „nohc“ im Beispielsatz vermutlich mühelos als „noch“ gelesen werden, da in der deutschen Sprache nur das „ch“ – im Gegensatz zum „hc“ – als phonologische und orthographisch korrekte Einheit auftritt. So lernen bereits Kinder unbewusst im Vorschulalter die Regeln der Phonologie. Dieses spontane Lernen erfolgt durch die eigene Sprachwahrnehmung und die eigene Sprachproduktion (Szagun, 2013). Dadurch entsteht zunächst eine phonemische Differenzierungsfähigkeit, die sich phonologische weiterentwickelt und sich mit der Lexikonentwicklung überschneidet. Durch den Erwerb der Schriftsprache werden schließlich orthographischen Regeln erlernt und eingeübt (Szagun, 2000, 2013).

Ähnlich verhält es sich mit den Fragezeichen in „i?t“ und „ke?n“. Das Fragezeichen („?“) ist ein Syngraphem der geschriebenen deutschen Sprache, welches formal als Grenzsignal funktioniert und nicht inmitten eines Wortes stehen kann. Dass das Syngraphem „Fragezeichen“ formal einen Satz beendet und in Bezug auf die Satzintention auf die als syntaktische Funktion „Frage“ hinweist, wird in der Regel zunächst beim Lesen-(Lernen) von ganzen Sätzen und Texten (z. B. wird beim Vorlesen einer Frage die Stimme am Satzende gehoben) und später beim Schreiben(-Lernen) als orthographische Regel gelernt (Gallmann, 1996, S. 1459). Da die Fragezeichen in dem hier betrachteten Beispiel in der Wortmitte ihre Funktion als Syngraphem nicht erfüllen können, ist der Leser an dieser Stelle zum Austausch und zur Sinnkonstruktion der Einheiten „i?t“ und „ke?n“ aufgefordert. Bei der Rekonstruktion von „i?t“ könnte nun

[27] Auflösung: *„ Es ist noch kein Meister vom Himmel gefallen“*.

auch die syntaktische Struktur des Satzes hilfreich sein: Ohne die Regeln jemals explizit gelernt zu haben, sind bereits 2-jährige Kinder dazu in der Lage, grammatikalisch korrekte Sätze in ihrer Erstsprache zu bilden (Szagun, 2013). Für den oben genannten Beispielsatz würde das bedeuten, dass Erstsprachsprecher*innen implizit schlussfolgern können, dass in einem deutschen Aussagesatz, der mit dem Wort „Es“ beginnt, an zweiter Stelle ein Verb steht. Soll also ein Verb mit drei Buchstaben, welches mit „i“ beginnt und mit „t“ endet, gefunden werden, liegt die Rekonstruktion von „ist“ nahe, zumal „ist“ als Verb im Deutschen häufig vorkommt. Ähnlich verhält es sich mit der Rekonstruktion des Pronomens „kein“.

Die Rekonstruktion der fehlenden Buchstaben (beispielweise das „o“ in „v_m“ oder „gefallen“ in „gfln“) wird dadurch vereinfacht, dass die Silben im Deutschen aus Konsonanten und Vokalen bestehen, wobei nie mehr als 3 Vokale und nie mehr als 5 Konsonanten in einer Silbe aufeinander folgen (Stöber, 2011, S. 313). Es reiht sich also mit einer bestimmten Wahrscheinlichkeit ein Vokal an einen Konsonanten und umgekehrt (ebd., 313). So können Wörter, in denen einzelne Buchstaben fehlen oder verfälscht sind aufgrund von Redundanzen erfolgreich decodiert werden.

Auf der phonologischen Ebene wird die Rekonstruktion dadurch erleichtert, dass gewisse Laute zum Teil in der Schriftsprache überflüssig sind. Durch die laute Aussprache der Konsonanten „gfln“ kann fast das Wort „gefallen“ [gəfalən] rekonstruiert werden, da das erste und das zweite „e“ phonologisch als Schwa [Ə] realisiert werden, wobei das zweite Schwa auch weggelassen werden könnte [gəfaln]. Das Graphem „g“ am Wortanfang wird in der deutschen Sprache in den meisten Fällen als [gə] artikuliert und macht das Graphem „e“ in diesem Fall überflüssig. Ähnlich verhält es sich mit Doppelkonsonanten, im Beispiel bei den Wörtern „Himmel“ und „gefallen“, die zwar die Länge, bzw. Kürze des nachstehenden Vokals anzeigen, vom geübten Leser jedoch sicherlich als „Rechtschreibfehler“ identifiziert und (phonologisch) korrigiert werden können.

Die Wahrscheinlichkeit, die Wörter bzw. den oben genannten Beispielsatz erfolgreich zu rekonstruieren, steigt auf der semantischen Ebene: Die Kenntnis der oder Erinnerung an die Redewendung (*Es ist noch kein Meister vom Himmel gefallen*) wird dem Leser

erleichtern, die Wörter zu vervollständigen bzw. zu korrigieren sowie den Satz inhaltlich zu verstehen.

Redundanzen sind neben der Buchstaben- und Satzebene auch auf der Wortebene und im Bereich der Grammatik zu finden. Der Satz *Sie reitet auf einem weißen Schimmel* ist gleich in mehrfacher Hinsicht redundant: Inhaltlich redundant ist das Wort „weißen" da ein Schimmel ein weißes Pferd ist, wodurch das zusätzliche Adjektiv „weiß" vor Schimmel überflüssig wird. Grammatikalisch redundant ist der unbestimmte Artikel „einem" und das Adjektiv „weißen", da beide Wörter im Kasus, Genus und Numerus an das Nomen „Schimmel" angepasst sind. Es ist davon auszugehen, dass der Satz gar ohne Präposition und Deklination „Sie reitet ein weiß Schimmel" inhaltlich gut verständlich wäre.

Das hier aufgeführte Beispiel zeigt, dass Redundanzen auf der phonologisch-morphologischen, syntaktischen, lexikalischen und auf der semantischen Ebene zu finden sind. Mithilfe dieser natürlich vorkommenden Redundanzen kann es der Empfänger*in gelingen, die beschädigte Nachricht zu korrigieren oder zu rekonstruieren, vorausgesetzt, sie kennt und beherrscht das Regelwerk, den Wortschatz und den kulturellen Hintergrund der Sprache (Grotjahn, 2002a, 2006, 2019b; Klein-Braley, 1981; Spolsky, 1973). Fehlen Teile in einem Satz oder wurden, wie im oben genannten Beispiel falsche Informationen in einen Text eingebaut, ist die Redundanz reduziert. Doch wie hängt die Fähigkeit einen beschädigten Satz oder Text zu rekonstruieren nun mit der Erfassung von Sprachkompetenz im C-Test zusammen?

2.3.3 Zum Zusammenhang zwischen Erwartungsgrammatik, dem Prinzip der reduzierten Redundanz und der Messung von Sprachkompetenz mit dem C-Test

Auf eine systematisch *reduzierte Redundanz* (Klein-Braley, 1997; Oller & Conrad, 1971; Spolsky et al., 1968) lässt sich das Konstruktionsprinzip des C-Tests verkürzen. Die Aufgabe beim Lösen eines C-Tests ist es, einen Text, in dem die natürliche Redundanz der Sprache durch das Löschen von Wortteilen reduziert ist, zu rekonstruieren. Doch wie können ebendiese Rekonstruktionsprozesse bei der Bearbeitung eines C-Tests gelingen?

Und wie kann durch das Ergänzen von Wortteilen im C-Test auf das Konstrukt einer allgemeinen Sprachkompetenz geschlossen werden?

Eine Grundlage für die Antwort auf diese Fragen liefert das von Oller beschriebene Konstrukt der sogenannten *internalisierten Erwartungsgrammatik* (*pragmatik expectancy grammar*) (1973; Oller, 1974, 1976, 165 ff., 1983, 3 ff.). Das Prinzip der Erwartungsgrammatik leitet Oller zunächst aus dem psycholinguistischen Modell zu Wahrnehmungs- und Verstehensprozessen ab, dem *Analyse-durch-Synthese-Modell* (Neisser, 1974). Das Modell der *Analyse-durch-Synthese* beschreibt den Prozess, bei dem die Rezipient*in beim Empfangen einer Nachricht die ankommenden Informationen entschlüsselt und analysiert (vgl. oben Sender-Empfänger-Modell), indem sie versucht, die eingehenden Informationen mit vergangenen Erfahrungen sowohl aus Sprachgebrauch und -handeln als auch aus dem impliziten und expliziten Regelwissen einer Sprache in Verbindung zu bringen und abzugleichen, d. h. zu „synthetisieren“ (Neisser, 1974; Wirtz, 2014, S. 142, 2017). Die Leser oder Hörer analysiert also die empfangenen Signale, „indem er sie zunächst grob verknüpft und diesen Vorentwurf dann mit den Ergebnissen der exakten Analyse vergleicht“ (Raatz & Klein-Braley, 1983, S. 110).

Oller überträgt diesen rezeptiven Verstehensprozess strukturell auch auf produktive Sprachleistungen und beschreibt ihn

> „im Prozeß des Verstehens als hypothesenbildende Instanz (in bezug auf das was ein Sprecher wohl mitteilen will, noch bevor ein Satz zu Ende formuliert ist) und im Prozeß der Sprachbildung als planbildende und die Ausführung des Plans kontrollierende Instanz“ (Vollmer, 1982, S. 41).

Dies bedeutet also, dass die Rezipient*in einer Nachricht sowohl auf der phonologischen, auf der syntaktischen, auf der semantischen als auch auf der pragmatischen Ebene versucht, „hypothesenartig die komplexen Strukturen nachzukonstruieren, die dem Plan des anderen zugrunde liegen“… „bis das, was der andere meint, nachvollzogen werden“ (Bolton, 1985, S. 48; Haug & Rammer, 1974, S. 160) oder die sprachliche Realisierung des eigenen Plans überprüft werden kann.

Diese Konstruktion von Bedeutungsstrukturen durch das fortwährende Überprüfen, Annehmen und Verwerfen von Hypothesen (Bolton, 1985, S. 47) gelingt nun laut Oller

durch die internalisierte Erwartungsgrammatik, wobei er den Begriff *Grammatik* „als ein umfassendes System von Regeln sprachlich-kommunikativen Handelns versteht“ (Bolton, 1985, S. 47). Damit beschreibt er die Annahme, dass eine Sprecher*in einer Sprache durch die Speicherung von Erfahrungen, Bedeutungen und Zusammenhängen über ein System von phonologischen, morpho-syntaktischen, semantischen und pragmatischen Erwartungsregeln verfügt. Dabei gilt, dass je ausgebildeter die Sprachkompetenz einer Sprecher*in ist, umso wahrscheinlicher und präziser sind die Hypothesen (Oller, 1973, 105 ff.).

Diese Theorie begründet nun auch den Zusammenhang zwischen dem Prinzip der reduzierten Redundanz und der Erfassung von Sprachkompetenz mit dem C-Test: Beim Lösen eines C-Tests stellt die Leser*in zunächst Hypothesen darüber auf, wie eine Lücke erfolgreich, d. h. formalsprachlich und semantisch korrekt rekonstruiert werden könnte. Durch die Einbeziehung von redundanten Informationen auf der Wort-, Satz und Textebene und den Abgleich gespeicherter phonologischer, morpho-syntaktischer, semantischer und pragmatischer Erfahrungswerte, bzw. Erwartungsregeln können die zunächst als Hypothesen entworfenen Wortergänzungen überprüft und angenommen oder als falsch verworfen werden. Diese Rekonstruktionsprozesse verlaufen (zumindest beim Lösen von rechtsseitig getilgten C-Tests[28]) bei Sprecher*innen in der Erstsprache durch die Internalisierung von sprachlichen Strukturen und Wahrscheinlich- bzw. Regelhaftigkeiten meist automatisiert und mit hoher Geschwindigkeit ab: „… the hypothesizing what will come next can be conzeptualised in terms of grammar based expectancies. It is these expectancies which enable the language user to process elements of a language at such phenomenally rapid rates” (Bolton, 1985, S. 48; Oller, 1976, S. 167). Oller beschreibt damit die Fähigkeit durch die Fülle an Erfahrungswerten sowohl im Sprachgebrauch als auch im impliziten und expliziten Regelwissen der Sprache, fehlende Informationen in einem Text mit hoher Geschwindigkeit richtig zu ergänzen (1976). In genau dieser Fähigkeit, d. h. in der internalisierten Erwartungsgrammatik, bzw. im „*knowledge of the words*“ bei (Spolsky, 1973, 165 ff.) sieht Oller einen Generalfaktor von Sprachkompetenz, dem alle sprachlichen

[28] Inwieweit Unterschiede zwischen klassischen, d.h. rechtsseitig getilgten C-Tests und linksseitig getilgten C-Tests bestehen, wird weiter in Kapitel 2.3.5 genauer erläutert.

Teilfähigkeiten, wie z. B. Lese- oder Schreibfähigkeit, zugrunde liegen. Neben Oller (1974) verstehen auch Spolsky (1973) und Vollmer (1982) Sprachkompetenz (*general language proficiency*) als eine zentrale, eindimensionale Kompetenz (*unitary competence hypothesis*), die in allen Teilfertigkeiten der Sprache (z. B. Lesen, Hören, Schreiben, Sprechen) und Sprachleistungen gleichermaßen wirksam ist (Asano, 2010, S. 5, 2014, 43 f.; Vollmer, 1982, S. 94) und die sie weder als „die Kenntnis einer bestimmten Anzahl von Vokabeln noch die Verfügung über andere isolierte Elemente dieser Sprache, sondern Kenntnis und Beherrschung ihrer Regeln“ definieren (Asano, 2010, S. 4, 2014, S. 42; Vollmer, 1982, S. 39).

Ollers Theorie, dass es entgegengesetzt der strukturalistischen Annahme einer mehrdimensionalen Sprachkompetenz, eine universelle, eindimensionale Sprachkompetenz gibt, die jeder sprachlichen Performanz sowohl in der Erst-, Zweit- als auch Fremdsprache zugrunde liegt und die als linguistische Kompetenz im Sinne einer Erwartungsgrammatik beschrieben wird, ist allerding schon früh kritisiert worden (Bolton, 1985; Vollmer, 1982, 43 ff.; Vollmer & Sang, 1980a, S. 11). So wird Ollers Auffassung, dass Rezeptions- und Produktionsleistungen den gleichen strukturellen Prozessen unterliegen, u. a. von Vollmer und Sang (1978) in Frage gestellt: Die Autoren weisen darauf hin, dass es nicht haltbar sei, Produktion und Verstehen prinzipiell als Spiegelprozesse zu sehen, da angenommen werden muss, dass die Differenzen zwischen den Prozessen des En- und Decodierens größer seien als die Gemeinsamkeiten dieser beiden Handlungsweisen (Bolton, 1985, S. 49; Vollmer, 1982, S. 42; Vollmer & Sang, 1980a). Kritisiert wird außerdem Ollers sprachliche Gleichstellung aller Sprecher*innen einer Sprache, da nicht zwangsläufig davon ausgegangen werden könne, dass die Fähigkeit eine Sprache zu produzieren und zu verstehen von unterschiedlichen Sprecher*innen jemals gleich ausgeprägt und/oder gleich angewendet wird (Bolton, 1985, S. 48). Vielmehr scheinen nach Bolton (1985), Vollmer und Sang (1980b) soziale Erfahrung und individuelle Lernprozesse das „Ausmaß der tatsächlichen Verfügungsmöglichkeiten über den Zeichen- und Regelvorrat dieses Systems“ (Vollmer & Sang, 1980b, S. 129) zu bestimmen.

Obwohl weder Oller (1976) noch Spolsky (1973) in ihren Ausführungen zwischen Erst- und Zweit- bzw. Fremdspracherwerb unterscheiden (Bolton, 1985, S. 47), ist es Spolsky

selbst, der darauf hinweist, dass „the richness of knowledge of probabilities on all levels – phonlogical, grammatical, lexical and semantic - in the language“ Fremdsprachsprecher*innen von Erstsprachsprecher*innen unterscheidet (Bolton, 1985; Spolsky, 1973, S. 174; Vollmer, 1982). Das heißt, dass nicht (nur) rein explizites Regelwissen – wie es die Begriffe Erwartungs*grammatik* bei Oller (1976) und *knowledge of the words* bei Spolsky (1973) vielleicht vermuten lassen könnten – ausschlaggebend für die erfolgreiche Rekonstruktion der Lücken ist, sondern besonders das verfügbare Ausmaß („richness“) der in Betracht kommenden Möglichkeiten zur korrekten Wortergänzung, die bei Erstsprachsprecher*innen durch die längere und natürlich erworbene Sprachanwendung viel ausgeprägter sein dürfte als bei Fremdsprachenlerner*innen. Denn im Gegensatz zu Erstsprachsprecher*innen müssen Fremdsprachenlerner*innen die Fertigkeiten und Fähigkeiten in allen Bereichen der Sprache – phonologisch, grammatikalisch, lexikalisch und semantisch – erst entwickeln und sind besonders auf den unteren Niveau-Stufen auf die redundanten Gesprächsteile im Kommunikationsprozess angewiesen, um den Inhalt einer Nachricht zu verstehen. Spolsky argumentiert in diesem Zusammenhang, dass Nicht- Erstsprachsprecher*innen für erfolgreiche Verstehensleistungen die komplette Redundanz der Sprache – und darüber hinaus manchmal sogar zusätzliche Informationen – benötigen (Spolsky, 1973, S. 170) und weist damit darauf hin, dass Fremdsprachenlerner*innen Verstehensprozesse z. B. durch verlangsamte, sehr deutliche Aussprache, Wiederholungen oder verstärkte Gestik und Mimik erleichtert werden können (Mashkovskaya, 2014, S. 82; Spolsky, 1973; Wockenfuß, 2009, S. 92).

Da der klassische C-Test, wie bereits erwähnt, ursprünglich für den fremdsprachlichen Bereich – und hier vor allem zur Messung des Sprachstandes von nicht allzu weit fortgeschrittenen Lerner*innen (Grotjahn et al., 2010, S. 298) konzipiert und eingesetzt wurde[29], kann nun an dieser Stelle zusammenfassend festgehalten werden, dass die erfolgreiche Rekonstruktion der getilgten Wortenden auf der Grundlage der sprachlichen Redundanz „als Manifest von Sprachkompetenz betrachtet, der Grad, zu dem sich diese Fähigkeit bei Nicht-Muttersprachlern manifestiert, daher als Maß für

[29] C-Test im Bereich der Erstsprache wurde vor allem zur Messung der Textschwierigkeit, nicht aber vorrangig zur Messung der Sprachkompetenz von Erstsprachsprecher*innen eingesetzt.

Sprachkompetenz“ angesehen werden kann (Sigott, 1993, S. 79). Vor diesem Hintergrund ist nachvollziehbar, warum C-Tests für Erstsprachsprecher*innen bisher vor allem zur Pilotierung von fremd- und zweitsprachlichen C-Tests und zur Erfassung der Sprachkompetenz von Kindern und Jugendlichen eingesetzt wurden (Baur & Spettmann, 2009; Grotjahn et al., 2010, S. 298).

2.3.4 Modifizierte C-Tests für fortgeschrittene Erst-, Zweit- und Fremdsprachensprecher*innen

Dass es Unterschiede im Erkennen und Nutzen von redundanten Informationen bei Rekonstruktionsprozessen zwischen erwachsenen Erstsprachsprecher*innen und Fremdsprachenlerner*innen gibt – wie oben ausgehend vom Funktionsprinzip des C-Tests theoretisch zumindest angedeutet – , zeigt sich auch empirisch in Ergebnissen aus verschiedenen Untersuchungen, in denen sich klassisch, d. h. rechtsseitig getilgte C-Tests für Erstsprachsprecher*innen oder auch sehr weit fortgeschrittene Fremd- oder Zweitsprachenlerner*innen in der Regel als zu leicht erweisen und nicht genügend zwischen den Proband*innen differenzieren (Eckes, 2006; Grotjahn et al., 2010, S. 298; Köberl & Sigott, 1994). So wurde beispielsweise in der Studie von Bremerich-Vos und Scholten-Akoun (2016) zur Sprachkompetenz von Lehramtsstudierenden (SkaLa®) festgestellt, dass sowohl C-Tests mit rechtsseitiger 3-er Tilgung[30] als auch mit klassischer 2-erTilgung[31] für die Kohorte der Studierenden in der Studieneingangsphase zu einfach sind (Bremerich-Vos & Scholten-Akoun, 2016, S. 195; Mashkovskaya, 2014, 92 ff.). Beide Testvarianten konnten im Durchschnitt zu deutlich über 90 % gelöst werden und wurden wegen dieser Deckeneffekte als ungeeignet für die Messung von Sprachkompetenz bei erwachsenen Studierenden im erst- und zweitsprachlichen Bereich eingestuft (Mashkovskaya, 2014, 92 ff.). Ähnlich verhält es sich mit einer weiteren modifizierten C-Test-Variante, in welcher nur bestimmte Wörter (z. B. Nomen und Verben), d. h.

[30] Bei der 3-er Tilgung wird jedes dritte Wort getilgt. Dieses Tilgungsverfahren hat sich im schulischen Bereich zur Messung von Erst- und Zweitsprache von Kindern in den Jahrgangsstufen 5-7 bewährt (Baur et al. (2013a, 3 ff.)

[31] Bei der 2-er Tilgung wird systematisch jedes 2. Wort getilgt.

Inhaltswörter rechtsseitig getilgt[32] wurden. Die Auswertung erwies diese Varianten wegen der ebenfalls zu beobachtenden Deckeneffekte als ungeeignet (ebd.).

Um nun einen C-Test für erwachsene Erstsprachsprecher*innen oder weit fortgeschrittene Fremd- oder Zweitsprachenlerner*innen zu konzipieren, gibt es verschiedene Möglichkeiten, die Schwierigkeit eines C-Tests der Zielkohorte anzupassen und zu erhöhen: Dies kann beispielsweise über die Reduzierung der Bearbeitungszeit, über die weitere Reduzierung der Redundanz oder über die Modifizierung des Tilgungsprinzips erfolgen.

C-Tests, in denen die Bearbeitungszeit der einzelnen Texte stark limitiert wird – sogenannte Speeded C-Tests (oder S-C-Tests)[33] – wurden bereits in einigen Studien vor allem zur Untersuchung des Zusammenhangs zwischen C-Test-Leistung, verbaler Intelligenz, logischem Denken und Informationsverarbeitungsgeschwindigkeit eingesetzt (Grotjahn, 2010b, 265 ff.; Grotjahn et al., 2010, 297 ff.; Grotjahn, 2019b, 593 f.; Wockenfuß & Raatz, 2006, 211 ff., 2014, 189 ff.; Zimmermann, 2019), mit dem Ergebnis, dass die Einführungen einer massiven Zeitlimitierung bei der Bearbeitung der einzelnen C-Testtexte im Vergleich zu C-Tests, die von den Proband*innen mit einer moderaten, d. h. einer mehr oder minder großzügigen Zeitvorgabe bearbeitet werden, zu einer Erhöhung der Testschwierigkeit führen und damit die Möglichkeit bieten, auch im oberen Kompetenzbereich, d. h. im erstsprachlichen und weit fortgeschrittenen zweit- und fremdsprachlichen Bereich, gut zu differenzieren (Grotjahn, 2010a, 265 ff.; Grotjahn et al., 2010, S. 314). In diesem Zusammenhang ist jedoch zu betonen, dass das Einbringen einer Speed-Komponente im C-Test für die erfolgreiche Rekonstruktion der Lücken „sowohl eine breite deklarative (sprachliche) Wissensbasis als auch eine hohe Prozeduralisierung und Automatisierung des Wissens und effiziente

[32] Untersuchungen zur Erhöhung der Testschwierigkeit durch die Modifikation des Tilgungsprinzips auf Inhaltswörter haben u. a. Dörnyei und Katona (1992) und Klein-Braley (1994) durchgeführt.

[33] An dieser Stelle ist darauf hinzuweisen, dass S-C-Tests nicht als reine Speed-Tests zu verstehen sind, über die durch die Vorlage von sehr einfachen C-Tests allein die Geschwindigkeit der Informationsverarbeitung gemessen wird (vgl. Grotjahn 2010, S. 268; Grotjahn, Schlak & Aguado, S. 301). Vielmehr werden S-C-Test als Niveautests eingestuft, da in ihnen Texte mit der Zielkohorte angemessener Schwierigkeit eingesetzt werden und laut Lienert und Raatz (1994, S. 35) eben diese Schwierigkeit zwischen Schnelligkeits- und Niveautests unterscheidet und nicht ob die Tests unter zeitlicher Limitierung bearbeitet werden oder nicht (ebd.).

Informationsverarbeitung“ erfordert (Grotjahn, 2010b, S. 285, 2019b). Denn im Gegensatz zu einem C-Test, dessen Bearbeitungszeit nicht massiv limitiert ist und der vor allem den Umfang des deklarativen und prozeduralen Wissens erfasst, ist ein sogenannter *S-C-Test* dahingehend konzipiert, anhand der korrekten Lösungen „über den Umfang des Wissens hinaus auch den Grad der Wissensautomatisierung und die Effizienz der Informationsverarbeitung zu messen“ (Grotjahn, 2010b; Grotjahn et al., 2010, S. 303; Grotjahn, 2019b, S. 594). Damit liegt dem S-C-Test zumindest in Teilen ein anderes Testkonstrukt zugrunde als dem konventionellen C-Test ohne massive Zeitlimitierung (ebd.).

Unabhängig von einer Limitierung der Bearbeitungszeit ist es prinzipiell auch denkbar, die Schwierigkeit eines C-Tests über die Textinhalte oder über die Textschwierigkeit der jeweiligen Teiltests zu verändern. Werden Texte als Grundlage zur Konstruktion von C-Test[34] verwendet, die beispielsweise besonders komplexe Themen und Sachverhalte, verflochtene Argumentationsstrukturen, eine hohe Dichte an Informationen, einen hohen Grad an Abstraktion der Inhalte usw. enthalten, überprüft ein auf dieser Basis entwickelter C-Test gegebenenfalls eher Fach- und/oder Weltwissen sowie möglicherweise Intelligenz als den Sprachstand der Proband*innen (Grotjahn, 2010b, S. 266; Grotjahn et al., 2010, S. 298; Raatz & Klein-Braley, 2002).

Neben inhaltlichen Komponenten kann auch die sprachliche Ausgestaltung eines Textes über seine Schwierigkeit entscheiden. Das Vorhandensein bestimmter textinterner Strukturen und formaler Merkmale, wie sie kennzeichnender Weise in fach- und/oder bildungssprachlichen Texten verwendet werden – zu nennen wären hier z. B. die vermehrte Verwendung von Fachvokabular, Fremdwörtern, Nominalisierungen, Präfix- und Suffixverben auf der lexikalisch-semantischen Ebene oder der häufige Gebrauch von komplexen Nebensatzkonstruktionen, Passivformen oder Funktionsverbgefügen auf der morphologisch-syntaktischen Ebene (Gogolin & Duarte, 2016, 489 ff.; Gogolin & Roth, 2007) – kann im Vergleich zu Texten mit einfachen textinternen Formen das Anforderungsniveau eines Textes und damit auch die Schwierigkeit, diesen lesen und

[34] Ausführliche Informationen zum Vorgehen bei der Auswahl von Grundlagentexten und zur Konstruktion eines C-Tests in Kapitel 4.1.2.

verstehen zu können, erhöhen. Die Frage nach der Beurteilung der Schwierigkeit eines Textes ist also auch in Abhängigkeit von der Vertrautheit der Leserschaft mit bestimmten Textsorten und den in diesen Textsorten gestellten Anforderungen zu beantworten.

Für eine erste Grobeinschätzung der Textschwierigkeit kann es möglicherweise hilfreich sein, den Lesbarkeitsindex (*LIX*) eines Textes zu bestimmen[35]. Dabei wird angenommen, dass je einfacher ein Test zu lesen ist, desto niedriger ist seine Schwierigkeit. Hierzu gibt es mittlerweile Online-Tools, die u. a. die Länge der Sätze und Wörter sowie die Anzahl an Silben in Wörtern in einem Text messen, um daraus Rückschlüsse auf die Lesbarkeit des Textes zu ziehen (Klein-Braley, 1985a; Lenhard & Lenhard, 2014-2017). Da der *LIX* allerdings nur einen kleinen Teil schwierigkeitsgenerierender Merkmale auf der Textoberfläche misst und nicht alle die Komplexität eines Textes beeinflussenden Faktoren adäquat abbildet, sollte er nur zu einer ersten und ungefähren Einschätzung der Textschwierigkeit genutzt werden (ebd.). Ein wesentlich elaborierteres Modell zur Einschätzung bzw. Vorhersage der Text- bzw. C-Test-Schwierigkeit schlagen Beinborn et al. (2014) vor: In der Textschwierigkeit (*paragraph difficulty*) sehen die Autor*innen einen Faktor, der sich auf die Schwierigkeit von C-Tests auswirkt und den es genauer zu betrachten gilt: Neben der durchschnittlichen Wort- und Satzlänge werden zur Bestimmung der lexikalischen Diversität eines Textes die Type-Token-Relation sowie das Vorkommen und die Häufigkeit von Verbvariationen und Pronomen gemessen. In die Berechnung zur Textschwierigkeit fließen ebenfalls syntaktische Merkmale ein, wie beispielsweise die Anzahl an unterschiedlichen Wortarten, an Bezugswörtern oder an spezifischen Satzstrukturen (z. B. Verbalphrasen, Nominalphrasen oder Nebensatzkonstruktionen) (Beinborn et al., 2014, S. 523)[36].

Eine weitere Möglichkeit zur Manipulation der Schwierigkeit im C-Test bietet die Modifikation des Tilgungsmusters: Ein C-Test für fortgeschrittene

[35] Eine absolute Messbarkeit von Textschwierigkeit gibt es nicht.
[36] Beinborn (2016) stellt unter http://spz-etest.ukp.informatik.tudarmstadt.de:9000/ctest ein kostenloses Online-Tool zur Verfügung, mit dem Texte im Englischen, Deutschen und Französischen automatisch in C-Test umgewandelt die Schwierigkeit der einzelnen Lücken anhand von drei farblich unterschiedenen Kategorien werden können. Diese Information kann dann u. a. für eine Modifikation der Schwierigkeit einzelner Lücken genutzt werden (vgl. auch Grotjahn 2019, S. 601).

Fremdsprachenlerner*innen, Zweitsprachsprecher*innen oder Erstsprachsprecher*innen könnte über eine größere Anzahl der getilgten Buchstaben, d. h. über die Reduktion der Redundanz, schwieriger zu lösen werden. Die diesem Ansatz zugrundeliegende Annahme ist: Je stärker die Redundanz eines Textes im C-Test reduziert wird, desto schwerer ist die Rekonstruktion des Textes bzw. der Lücken. Eine Reduktion der Redundanz ließe sich auch über den Wechsel der Tilgungsseite, d. h. hier den Wechsel von der rechtsseitigen zur linksseitigen Tilgung oder die Eingrenzung des Tilgungsprinzips auf Inhaltswörter realisieren (Grotjahn, 1987, S. 226, 2010a, S. 266, 2019b, 589 ff.). In Bezug auf diese letztgenannte Möglichkeit der Modifizierung zeigen Untersuchungen von Dörnyei und Katona (1992) und Klein-Braley (1994), dass über getilgte Inhaltswörter deutlich besser im oberen Kompetenzbereich differenziert werden kann als über beschädigte Strukturwörter, da Strukturwörter viel häufiger vorkommen, eine geschlossene Wortklasse bilden und aus diesem Grund potentiell einfacher zu lösen sind als Inhaltswörter (Grotjahn, 2010b, S. 266; Kaufmann, 2016, S. 124). Der Einsatz eines C-Tests mit Eingrenzung des Tilgungsprinzips auf Inhaltswörter lieferte – wie oben bereits erwähnt – in der Studie von Bremerich-Vos und Scholten-Akoun (2016, 195 ff.) allerdings keine zufriedenstellende Differenzierung und führte bei der Gruppe der Lehramtsstudierenden zu Deckeneffekten.

Die Auswirkung auf die C-Testschwierigkeit durch den Einsatz von unterschiedlichen Tilgungsvarianten untersuchen u. a.[37] auch Köberl und Sigott (1994; 1996),indem sie in ihren Studien neben einem klassischen C-Test, C-Tests mit drei weiteren Tilgungsmethoden – neben der klassischen (s.u. Variante a) – einsetzen (Köberl & Sigott, 1994, S. 179):

a. die hintere Hälfte jedes zweiten Wortes wird gelöscht (C25); zwei Drittel jedes zweiten Wortes werden gelöscht (C33);
c. jedes zweite Wort mit Ausnahme des ersten Buchstabens (CFL) oder
d. die erste Hälfte jedes zweiten Wortes wird gelöscht (X25).

[37] Den Effekt von Modifizierungen des Tilgungsprinzips untersuchten beispielsweise auch Kokkota (1988) oder Cleary (1988).

In ihrer Untersuchung von 1994 erproben Köberl und Sigott zunächst deutsche C-Tests, die von Sprecher*innen mit Deutsch als Erstsprache bearbeitet werden. Die gleichen C-Tests werden 1996 als englische Version von Proband*innen mit Englisch als Erstsprache ausgefüllt. In beiden Studien gehen Köberl und Sigott prinzipiell davon aus, dass über die Reduzierung von Redundanz die Schwierigkeit von C-Tests modifiziert werden kann, d. h. konkret, dass C-Tests umso schwieriger zu lösen sind, je mehr Informationen fehlen (Köberl & Sigott, 1994, 181 ff.). Dazu bestimmen die Autoren zunächst den Prozentsatz der gelöschten Buchstaben als Maß für die Reduzierung der Redundanz in jeder Testversion: Während im klassischen C-Test (Variante a, C25), in dem die hintere Hälfte jedes zweiten Wortes gelöscht wurde, die sprachliche Redundanz um knapp über 25 Prozentpunkte reduziert ist, steigt diese Quote um etwa 8 Prozentpunkte durch die Löschung von zwei Drittel jedes zweiten Wortes (Variante b, C33) sowohl in den deutschen als auch in den englischen C-Tests. Diese weitere Reduzierung der Redundanz um 8 Prozentpunkte im Text führt sprachunabhängig in Bezug auf die Ergebnisse der C-Tests jedoch zu einem wesentlich deutlicheren Absinken der durchschnittlichen Lösungsquote. Die Ergebnisse der Probanden liegen nach Angaben der Autoren im Durchschnitt um 20 Prozentpunkte unter dem Mittelwert, der für die klassisch getilgten C-Tests ermittelt wurde.

Für die C-Test-Version, in der nur der erste Buchstabe jedes zweiten Wortes stehenbleibt (Version c, CFL), berechnen die Autoren eine Reduzierung der Redundanz von ca. 43 Prozentpunkten für den deutschen und ca. 40 Prozentpunkten für den englischen C-Test. Die Lösungsquote dieser C-Test-Versionen liegt dabei für den englischen ca. 10 Prozentpunkte und für den deutschen C-Test etwa 14 Prozentpunkte unter der durchschnittlichen Lösungsquote der klassischen C-Tests. Das bedeutet, dass sich eine sinkende Redundanz überproportional und annähernd linear in den Lösungsquoten abbildet (ebd.).

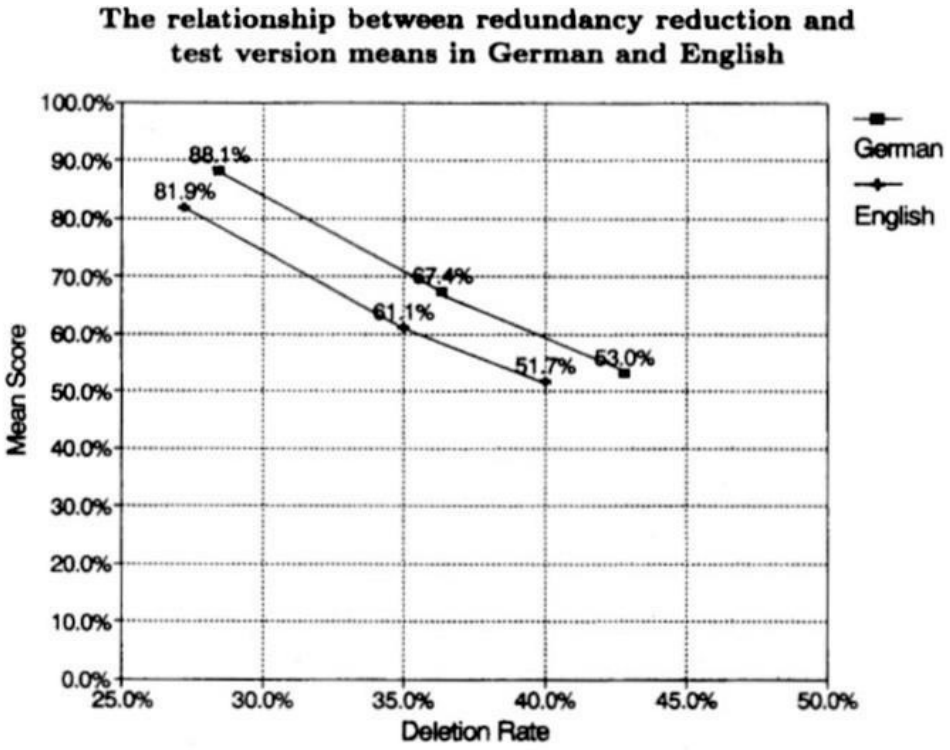


Abbildung 7: Einfluss der Reduzierung von Redundanz auf die Schwierigkeit von C-Tests nach Sigott & Köberl

Quelle: Sigott, Köberl (1996), S. 163.

Ein ganz anders Bild zeigt sich jedoch bei der Betrachtung der Ergebnisse der eingesetzten C-Tests mit linksseitiger Tilgung (Variante X25): Wenn nun die Anzahl der gelöschten Buchstaben als Maß für die Reduzierung der Redundanz bestimmt wird, so wie Köberl und Sigott (1994; 1996) es für die anderen eingesetzten Testvarianten vorschlagen, dann müsste das Maß der reduzierten Redundanz der linksseitig getilgten C-Tests, ebenso wie für die klassischen C-Tests, etwa 25 Prozentpunkten entsprechen, da in dieser Variante ebenfalls die Hälfte der Buchstaben (nur nicht am Wortende sondern am Wortanfang) gelöscht wurde. Das nach Sigott und Köberl (1996, S. 162) augenscheinlich gleiche Maß an reduzierter Redundanz führt durch den Wechsel der Tilgungsseite auf den Wortanfang jedoch zu schwierigeren C-Tests, die sich auch in Hinblick auf die jeweilige Sprache unterschiedlich verhalten: Während die durchschnittliche Lösungsquote des linksseitig getilgten, englischen C-Tests im Vergleich zu dem klassisch getilgten C-Test mit rechtsseitiger Tilgung um 15 Prozentpunkte

absinkt, liegt die Lösungsquote für den deutschen, linksseitig getilgten C-Test sogar 35 Prozentpunkte unter dem der klassisch getilgten Testvariante a).

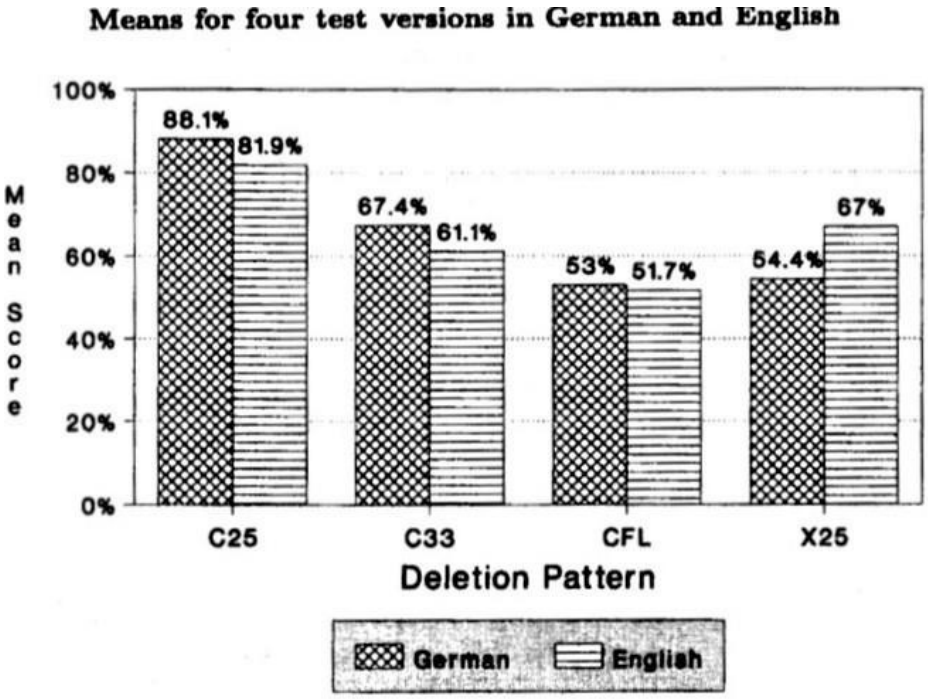


Abbildung 8: Mittelwerte der unterschiedlichen C-Test-Versionen im Englischen und Deutschen nach Sigott & Köberl

Quelle: Sigott, Köberl (1996), S. 163

Wie auch dem Diagramm zu entnehmen ist, hat die linksseitige Tilgung im deutschen C-Test (X25) einen deutlich größeren Effekt auf die Testschwierigkeit als die linksseitige Tilgung im englischen C-Test. Während die Testschwierigkeit im linkseitig getilgten, deutschen C-Test etwa der Schwierigkeit der C-Testvariante c), in welcher das getilgte Wort bis auf den ersten Buchstaben gelöscht wird (CFL), entspricht, ist der linksseitig getilgte, englische C-Test nach dem klassischen C-Test der zweitleichteste[38].

[38] Die Variante mit linksseitiger Tilgung erzielt für den deutschen C-Test außerdem sehr gute Werte im Bereich der Reliabilität und Varianz, wohingegen der englische, linksseitig getilgte C-Test nur niedrige Reliabilitätswerte aufweisen konnte (Köberl, Sigott 1994; 1996, S. 168). Auch Ergebnisse aus SkaLa®-Untersuchungen zeigen, dass deutsche C-Tests mit linksseitiger Tilgung bei der Gruppe der Studierenden eine „zufriedenstellende Varianz“ ergaben (Bremerich-Vos, Scholten-Akoun 2016, S. 195; Mashkovskaya 2014, S. 92 ff.).

Die Betrachtung dieser Ergebnisse wirft nun unweigerlich eine Reihe von Fragen auf: Wie kann es sein, dass C-Tests, die augenscheinlich das gleiche Maß an reduzierter Redundanz aufweisen, durch den Wechsel des Tilgungsmusters von rechts (d. h. vom Wortstamm oder –ende) auf links (d. h. auf den Wortanfang) schwieriger werden? Warum ist der englische C-Test mit linksseitiger Tilgung so viel einfacher als der deutsche, linksseitig getilgte C-Test? Ist die Reduzierung von Redundanz – auf dessen grundlegendem Wirkungsprinzip der C-Test funktioniert – damit möglicherweise nicht nur abhängig von der Anzahl der gelöschten Buchstaben, sondern bei der linksseitigen Tilgung auch von sprachspezifischen (morpho-syntaktischen und lexikalisch-semantischen) Strukturen?

2.3.5 Rechtsseitige vs. Linksseitige Tilgung

Die oben beschriebenen Ergebnisse aus den Untersuchungen sowohl von Köberl und Sigott (1994; 1996) als auch von Bremerich-Vos und Scholten-Akoun (2016) zeigen deutlich, dass die Schwierigkeit eines C-Tests über eine linksseitige Tilgung erhöht werden kann. Eine mögliche Erklärung dafür, dass die Rekonstruktion eines Wortes dann leichter fällt, wenn der Anfang des Wortes gegeben ist, sehen Sigott und Köberl im sogenannten *Badewanneneffekt* (Sigott & Köberl, 1996, S. 167). Der Badewanneneffekt beschreibt, dass bei der Suche nach Wörtern im mentalen Lexikon Wortanfänge und Wortenden besser zugänglich sind als der Mittelteil eines Wortes (Aitchison, 1987, 130 f.; Dittmann, 2002, 283 ff.; Juska-Bacher & Jakob, 2014, S. 52). Analog zu einer Person, die in der Badewanne liegt und deren Kopf am einen und die Füße am anderen Ende der Badewanne herausschauen, werden Wörter also über ihren Anfang (den Kopf) oder ihr Ende (die Füße) erschlossen, während die Wortmitte, die als „Körper“ im Wasser liegt, beim Zugriff auf das mentale Lexikon keine zentrale Rolle spielt (ebd.). Diesem Sinnbild folgend nimmt der Kopf beim Abrufen von Wörtern einen größeren Stellenwert ein als die Füße: So konnte in Untersuchungen gezeigt werden, dass sich Probanden besser an die Anfänge bzw. Anlaute von Wörtern erinnern konnten (zu 51 %) als an Wortenden bzw. Endlaute (zu ca. 35 %) (Aitchison, 1987, S. 131; Browman, 1978; Juska-Bacher &

Jakob, 2014, S. 52). In Bezug auf den C-Test weisen diese Ergebnisse zwar prinzipiell daraufhin, dass durch die bessere Zugänglich- bzw. Abrufbarkeit von Einträgen im mentalen Lexikon über Wortanfänge das rechtsseitige Ergänzen einer Lücke, bzw. eines Wortes, dessen Anfang bereits gegeben ist, einfacher gelingen kann als wenn der linksseitige Anfang des Wortes über das Wortende rekonstruiert werden muss. Allerdings kann mithilfe des hier beschriebenen Badewanneneffekts wohl kaum erklärt werden, warum nun die Löschung von Wortanfängen im englischen C-Test lediglich zu einer eher geringfügigen Modifikation der Schwierigkeit führt, während sie diese für den deutschen C-Test erheblich steigert (Sigott & Köberl, 1996, 167 ff.).

Diese sprachspezifischen Unterschiede in Bezug auf den C-Test sehen Sigott und Köberl (1996, 167 ff.) vielmehr in der morphologischen Struktur der jeweiligen Sprache begründet: Da die morphologische und syntaktische Struktur im Deutschen wesentlich mehr Flexionen aufweist als im Englischen, werden bei der linksseitigen Tilgung im Deutschen mehr lexikalische Morpheme gelöscht als im Englischen (ebd.). Folgendes Beispiel einer Nominalphrase soll zur Veranschaulichung dienen:

Deutsch:	***ein einfaches Beispiel***
Englisch:	***a simple example***

Abbildung 9: Beispiel Nominalphrase Deutsch und Englisch

Das Wort *einfach* wird hier im deutschen Beispiel als Adjektiv auf der morpho-syntaktischen Ebene flektiert, d. h. in Bezug auf den vorangestellten, unbestimmten Artikel und auf das nachfolgende Nomen in Genus, Kasus, Numerus und Komparation dekliniert (Neutrum, Nominativ, Singular, Positiv). Daraus ergibt sich folgende Struktur:

einfach (Wortstamm/lexikalisches Morphem) + Endung *-es* (Flexionsmorphem) = *einfaches*

Abbildung 10: Beispiel zur Flexion im Deutschen

Wenn nun in einem C-Test die erste Hälfte des Wortes *einfaches* getilgt wird, bleibt die Endung *-ches* bestehen. Da zumindest die letzten beiden Buchstaben *-es* das Flexionsmorphem bilden und damit eine grammatikalische Information erhalten, die jedes deutsche Adjektiv im Neutrum, Nominativ, Singular, Positiv mit vorangestelltem unbestimmten Artikel als mögliche Lösung in Betracht kommen lässt, geben lediglich die Buchstaben *-ch* einen Hinweis auf die lexikalische Bedeutung des zu ergänzenden Wortes, wobei die Adjektivendung *-ch* (bzw. *-lich*) im Deutschen häufig vorkommt[39]. Theoretisch könnte also an dieser Stelle eine Vielzahl von passenden Wörtern ergänzt werden (z. B. *übliches, ärmliches, ehrliches, ähnliches, deutliches* oder auch *solches, gleiches*, usw.), was im C-Test zu einer Lücke führte, die korrekt zu lösen die Einbeziehung des weiteren Ko- und Kontextes erfordern und damit ein potentiell schwieriges Item als ein auf der Wortebene[40] lösbares darstellen würde (Mashkovskaya, 2014, 161 ff.).

Die englische Übersetzung hingegen zeigt, dass das Adjektiv *simple* auf der morpho-syntaktischen Ebene nicht flektiert, sondern als Wortstamm ohne Flexionsmorphem eingesetzt wird. Wenn nun auch in diesem Fall die erste Hälfte des Wortes gelöscht wird, bleiben die Buchstaben *-ple* bestehen, die einen Teil der lexikalischen Bedeutung ausmachen und damit einen konkreten Hinweise auf das zu lösende Wort geben, bzw. die Antwortmöglichkeiten maßgeblich eingrenzen. Alternative Antwortmöglichkeiten könnten zwar beispielsweise die Adjektive *ample* (reichlich) oder *purple* (lila) sein, allerdings scheinen diese Ergänzungen in Bezug auf das nachstehende Nomen *example* eher unpassend und könnten auf der weiteren Satzebene im C-Test sicherlich schnell auf ihre Angemessenheit hin überprüft werden.

Das, was hier exemplarisch als Adjektiv-Flexion beschrieben wird, gilt auf der morpho-syntaktischen Ebene für fast alle Wörter der deutschen Sprache[41]: Flektiert werden neben Adjektiven auch Nomen, Verben, Artikel und Pronomen. Dabei realisiert sich die

[39] Ausführliche Informationen zur Wortbildung auf der morphologischen Ebene im Deutschen auf den folgenden Seiten.
[40] Vgl. für weitere Informationen zu den Prozessebenen bei der Bearbeitung von C-Tests und zur Kategorisierung von C-Test-Lücken Mashkovskaya (2014, 161 ff.)
[41] Ausgenommen sind Adverbien, Konjunktionen, Präpositionen, Interjektionen und Partikeln.

Flexion im Deutschen in der Regel – wie oben gezeigt – durch angehängte Endungen, d. h. durch Suffigierung, die ein mit der jeweiligen Wortart verbundenes grammatisches Merkmal anzeigen, wie beispielsweise das Tempus beim Verb oder der Kasus beim Nomen. Diese Suffigierung zur Markierung der grammatischen Funktion eines Wortes im Satz erfolgt dabei nach festen grammatikalischen Regeln und bildet somit eine theoretisch in sich geschlossene Liste möglicher Ergänzungen. Wird nun die vordere Hälfte eines deutschen Wortes gelöscht, bleiben am Wortende oftmals vor allem grammatikalische Marker stehen, die zwar Hinweise auf die morpho-syntaktische Funktion des zu ergänzenden Wortes geben können, aber in keinem direkten Bezug zur semantisch-lexikalischen Bedeutung des Wortes stehen müssen. Durch eine linksseitige Tilgung werden damit in einem deutschen C-Test mehr Wortstämme, bzw. lexikalische Morpheme gelöscht als in einem englischen C-Test (Sigott & Köberl, 1996, 167 ff.).

Um nun der Antwort auf die Frage, warum eine linksseitige Tilgung in einem deutschen C-Test zu einer massiven Erhöhung der Schwierigkeit gegenüber einer rechtsseitigen Tilgung führt, (noch) näher kommen zu können, soll an dieser Stelle neben der Flexion auch die Derivation, d. h. die Bildung von neuen Wörtern in der deutschen Sprache, näher betrachtet werden[42]. Was also ist aus morphologischer Sicht am Anfang eines Wortes „anders“ – und damit im linksseitig getilgten C-Test schwieriger zu ergänzen – als am Wortende?

In der deutschen Sprache werden neue Wörter (auch) durch die Derivation, d. h. durch die Kombination von Wortstamm und Derivationsmorphemen (Affixe) gebildet. Dabei können Derivationsmorpheme als Präfix vor den Wortstamm gestellt (a), als Suffix an den Wortstamm angehängt werden (b) oder als Zirkumfix den Wortstamm umgeben (c)[43] (Elsen, 2011; Meibauer et al., 2002):

a) *be-* (Präfix) + *Note* (Wortstamm) = *benote(n)*

[42] Zur Wortbildung im Deutschen gehört auch die Komposition, d. h. die Zusammensetzung von zwei Wortstämmen, auf die an dieser Stelle jedoch nicht weiter eingegangen werden soll, da im C-Test bei Komposita nur die erste bedeutungstragende Einheit getilgt wird (vgl. Grotjahn, 2002).
[43] Zur Gruppe der Derivationsaffixe gehören im Deutschen auch Interfixe und Suprafixe, die in ihrer Stellung jedoch nicht am Wortanfang oder –ende stehen.

b) Freund (Wortstamm) + *-schaft* (Suffix) = *Freundschaft*

c) ge- (Zirkumfix) + *Arbeit* (Wortstamm) *+ -tet* (Zirkumfix)= gearbeitet.

Die Derivation von Nomen als auch von Adjektiven kann durch eine große Anzahl an unterschiedlichen Suffixen realisiert werden. Durch die Suffigierung können so Nomen von Verben (z. B. *üben → Übung*), Adjektiven (z. B. *arm → Armut*) oder anderen Nomen (z. B. *Freund → Freundschaft*) abgeleitet werden, ebenso wie Adjektive von Nomen (*Sympathie → sympathisch*), Verben (*schweigen → schweigsam*), Adverbien (*gestern → gestrig*) als auch von anderen Adjektiven (*arm → ärmlich*). Mit der Suffigierung geht also vor allem ein Wechsel der Wortart einher, der die lexikalische Bedeutung des Wortstamms im Wesentlichen nicht verändert (Dürscheid et al., 2018, S. 5). Die Suffixe selbst tragen dabei keine eigenständige lexikalische Bedeutung, vielmehr fungieren sie als typische Marker für bestimmte Wortarten (die Derivation von Nomen endet meistens auf *-heit, -keit -schaft, -ung*, die der Adjektive auf *-lich, -ig, -isch, -bar, -haft*, die von Adverbien auf *-wärts, -dings, -weise*, usw.). Werden die Wörter in einem C-Test nun linksseitig getilgt, lässt das Tilgungsmuster gerade diese Suffigierung unbeschädigt und löscht dafür große Teile des Wortstamms am Anfang eines Wortes (Baur et al., 2013a, S. 15; Scholten-Akoun et al., 2014, S. 408). Dabei ist an dieser Stelle darauf hinzuweisen, dass besonders Derivationen von Nomen in bestimmten Textsorten (z. B. wissenschaftliche oder protowissenschaftliche Texte, die für die Erstellung von C-Tests für erwachsene, fortgeschrittene Sprecher oftmals angemessen scheinen (s. o.)), als Merkmal von Bildungssprache in Form von Nominalisierungen besonders häufig vorkommen (Gogolin & Duarte, 2016, 489 f.; Gogolin & Lange, 2011, S. 114).

Im Unterschied zu der großen Anzahl an Derivationsmöglichkeiten mit Suffixen sind die Möglichkeiten der Derivation durch Präfixe relativ gering (Fleischer & Barz, 2012, 255 ff.). Als Beispiel können hier Präfixe wie *un-*, *de-* oder *miss-* genannt werden, die in Kombination mit unterschiedlichen Wortstämmen neue Nomen, Verben oder Adjektive bilden (z. B. *Unmut, untätig, missachten, Missgunst, demotiviert*, etc.). Im Gegensatz zur Derivation mit Suffixen, die in der Regel einen Wortartwechsel anzeigt, führt die Präfigierung zu bedeutungsverändernden Modifizierungen.

Anders als die Derivation von Nomen, Adjektiven und Adverbien wird die Derivation von Verben im Deutschen vor allem über eine Linkserweiterung, d. h. durch das Voranstellen

von Präfixen oder Verbpartikeln vollzogen (Eichinger, 2000, S. 102). Durch die Präfigierung können Verben von fast allen Wortarten abgeleitet werden, z. B. von Nomen (z. B. *Haus → behaus(en)*), von Adjektiven (z. B. *neu → erneu(ern)*) oder auch von nicht flektierbaren Wortarten (z. B. *ja → bejah(en)*). Ebenso ist es möglich, Verben durch das Voranstellen von Präfixen oder Partikeln abzuleiten, ohne dass die Wortart *Verb* gewechselt wird (z. B. *ver + fahren → verfahren, be + raten → beraten, aus + suchen → aussuchen, usw.*) (Dürscheid et al., 2018). Was sich bei der Bildung von Präfix- und Partikelverben jedoch ändert, ist die lexikalische Bedeutung des abgeleiteten und damit neu entstandenen Lexems, wie anhand der Beispiele deutlich wird (während das Verb *fahren* hauptsächlich die Fortbewegung mit einem Fahrzeug o. ä. beschreibt, kann *verfahren* bedeuten, nach einer bestimmten Methode oder Weise vorzugehen und zu handeln oder aber vom richtigen Weg abzukommen (vgl. DUDEN), usw.). Durch die Bildung von Präfix- und Partikelverben können sich ebenso die syntaktischen Eigenschaften des Verbs verändern, so zählen Partikelverben beispielsweise zu den trennbaren Verben im Satz (Sie *suchte* die Wohnung *aus*), das Präfixverb *verfahren* erfordert die Ergänzung eines Reflexivpronomens (Er hat *sich verfahren*.), usw. (Dürscheid et al., 2018).

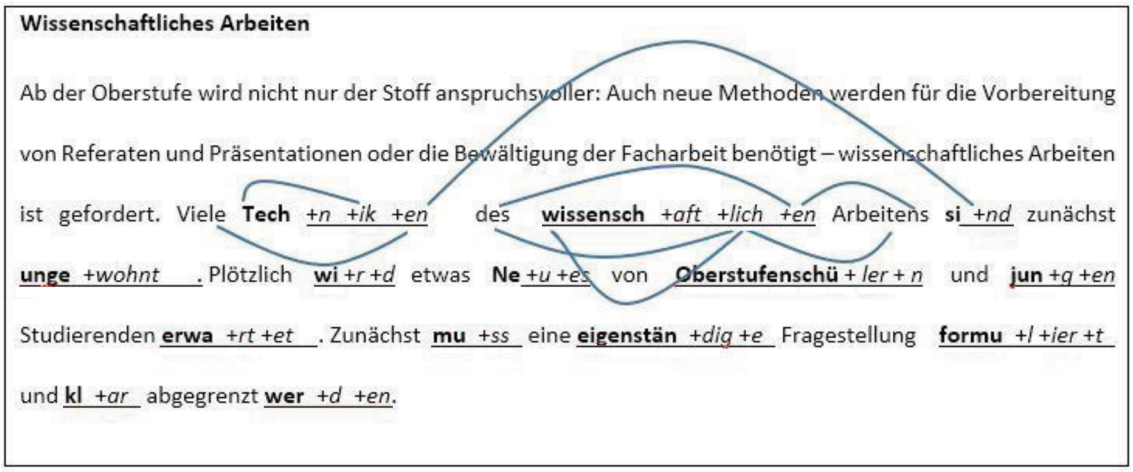
Während die Präfigierung also eine Vielzahl an Möglichkeiten zur Verbbildung darstellt, spielt die Suffigierung bei der Ableitung von Adjektiven (z. B. *aktiv → aktivier(en)*) oder Nomen (z. B. *Struktur → strukturier(en)*) eine eher untergeordnete Rolle. Die Verbbildung, die vorrangig durch Präfigierung realisiert wird, verhält sich also entgegengesetzt der Bildung von Nomen, bei der deutlich mehr Suffixe als Präfixe eingesetzt werden (Dürscheid et al., 2018; Fleischer & Barz, 2012, 9 ff.).

Ausgehend von der morphologischen Struktur der deutschen Sprache sind es also vor allem bedeutungstragende Elemente die am Anfang eines Wortes stehen, während Wortenden oftmals aus Flexionsmorphemen oder Derivationssuffixen bestehen, die hauptsächlich als grammatikalische Marker dienen, aber keine eigenständige lexikalische Information beinhalten. Auch wenn natürlich nicht alle Wörter der deutschen Sprache und die in einem C-Test vorkommen abgeleitet oder flektiert werden – Wortstämme können auch in ihrer Grundform unverändert stehen bleiben, ebenso werden beispielsweise bestimmte Wortarten (Konjunktionen, Partikel, Präpositionen,

etc.) nicht flektiert – kann angenommen werden, dass sich die Ergänzung grammatikalischer Informationen in einem rechtsseitig getilgten C-Test häufiger auf explizites oder implizites Regelwissen stützt und damit eine begrenzte Auswahl an Antwortmöglichkeiten bildet. Während sich unter Regeln also alle Fälle (mit Ausnahmen) erfassen lassen, scheint bei einem linksseitig getilgten Text der Rekurs auf Regeln nicht sonderlich hilfreich, da hier tendenziell der Rückgriff auf die Einzeleinträge des Lexikons zwingend ist. Wenn aber nun ein Tilgungswechsel vom Wortende auf den Wortanfang (Köberl & Sigott, 1994; Sigott & Köberl, 1996) zu einer erheblichen Steigerung der Schwierigkeit führt, wäre dann nicht die Annahme naheliegend, dass aufgrund der sprachlichen Struktur im Deutschen auf der morpho-syntaktischen Ebene mehr redundante Informationen zu finden sind als auf der semantisch-lexikalischen Ebene? Wäre im Falle der linksseitigen Tilgung cum grano salis davon auszugehen, dass im Gegensatz zur rechtsseitigen Tilgung gerade die Informationen getilgt werden, die nicht oder in einem geringeren Maße redundant, d. h. mehrfach vorhanden sind? Der Beantwortung dieser Fragen soll im Folgenden zumindest exemplarisch anhand eines Vergleichs einer linksseitig und einer rechtsseitig getilgten Version des gleichen C-Test[44] nachgegangen werden.

Dazu wird zunächst aus morpho-syntaktischer Sicht betrachtet, wie in einem rechtsseitig getilgten C-Test Lücken am Wortende rekonstruiert werden und welche redundanten Informationen Hinweise auf die korrekte Lösung liefern können.

[44] Der C-Test *Wissenschaftliches Arbeiten* ist in der linksseitig getilgten Version Bestandteil des in der hier vorgestellten Untersuchung eingesetzten C-Test-Sets.



Wissenschaftliches Arbeiten

Ab der Oberstufe wird nicht nur der Stoff anspruchsvoller: Auch neue Methoden werden für die Vorbereitung von Referaten und Präsentationen oder die Bewältigung der Facharbeit benötigt – wissenschaftliches Arbeiten ist gefordert. Viele **Tech** *+n +ik +en* des **wissensch** *+aft +lich +en* Arbeitens **si** *+nd* zunächst **unge** *+wohnt* . Plötzlich **wi** *+r +d* etwas **Ne** *+u +es* von **Oberstufenschü** *+ ler + n* und **jun** *+g +en* Studierenden **erwa** *+rt +et* . Zunächst **mu** *+ss* eine **eigenstän** *+dig +e* Fragestellung **formu** *+l +ier +t* und **kl** *+ar* abgegrenzt **wer** *+d +en*.

Abbildung 11: Ausschnitt aus dem C-Test *Wissenschaftliches Arbeiten* mit rechtsseitiger Tilgung

Das Lösen der ersten Lücke erfordert die Ergänzung des Wortanfangs *Tech* mit *n (Wortstamm) + ik (Derivationsmorphem) + en (Flexionsmorphem)* zum Wort *Technik*. Um diese erste Lücke lösen zu können, gibt es auf der morpho-syntaktischen Ebene mehrere Informationen, die dies ermöglichen. Wird der Wortanfang *Tech* betrachtet, zeigt die Großschreibung des ersten Buchstabens an, dass es sich bei dem zu ergänzenden Wort um ein Nomen handelt, welches durch die Suffixendung *-ik* zu *Technik* ergänzt werden kann. Das unbestimmte Zahlwort *Viele,* welches ohne Artikel vor der Lücke steht, weist auf der morpho-syntaktischen Ebene darauf hin, dass das nachstehende Nomen im Plural steht und verweist durch die Endung *-e* (*viel + e = viele*) auf den Nominativ oder Akkusativ des Nomens. Ob es sich um bei der zu ergänzenden Lücke um eine Kasusergänzung im Nominativ oder Akkusativ handelt, kann in Bezug auf die Satzstruktur beantwortet werden: Da *Viele Techniken* vor dem Prädikat steht und im Satz als Subjekt fungiert, steht *Techniken* im Nominativ, wobei diese Information für das Lösen der Lücke vernachlässigt werden könnte, da das Wort *Technik* im Plural in jedem Kasus über das Morphem *-en* flektiert wird.

Die zweite Lücke wird durch die Ergänzung des Wortanfangs *wissensch* mit *-aft (Teil des Derivationsmorphen) + lich (Derivationsmorphem) + en (Flexionsmorphem)* gelöst. Der Wortanfang *wissensch-* verweist durch das *-sch* auf das erste Derivationsmorphem hin: *wissen + schaft → Wissenschaft.* Die Position des zu ergänzenden Wortes zwischen einem bestimmten Artikel und einem Nomen sowie der kleingeschriebene

Anfangsbuchstabe weisen auf die Wortart Adjektiv hin, die durch das Suffixendung *-lich* abgeleitet werden kann. Da sich das attributiv verwendete Adjektiv *wissenschaftlich* innerhalb der Nominalphrase in seiner Form dem Nomen, auf das es sich bezieht anpasst, muss eine grammatische Kongruenz in Genus, Numerus und Kasus mit dem Nomen bestehen. Da *Arbeitens* im Neutrum, Singular und Genitiv steht, muss das Flexionsmorphem *-en* (schwache Flexion mit bestimmten Artikel im Neutrum, Singular, Genitiv) an *wissenschaftlich* (*wissenschaftlich + en* = *wissenschaftlichen*) angehängt werden. Durch die Flexion des bestimmten Artikels von *das* (nicht-flektiert *das* Arbeiten) zu *des* trägt dieser innerhalb der Nominalphrase ebenfalls die gleichen grammatikalischen Informationen.

Insgesamt gibt es in dem hier gezeigten C-Test-Ausschnitt 14 Lücken. Bei vier Lücken erfordert ein Teil der Ergänzung des Wortendes das Einfügen eines oder mehrerer Derivationsmorpheme: -ik (*Techn**ik**en*), -schaft (*wissen**schaft**lichen*), -lich (*wissenschaft**lich**en*), -ler (*Oberstufenschü**ler**n*) und -ig (*eigenständ**ig***). Flektiert werden alle hier zu ergänzenden Wörter bis auf Adjektiv *klar*. Das bedeutet, dass etwa 93 % der Lücken über eine Suffigierung am Wortende in ihren grammatikalischen Merkmalen in Bezug auf andere Bezugswörter bzw. in ihrer grammatischen Funktion im Satz angepasst werden: Die Nomen *Techniken* und *Oberstufenschülern*, das substantivierte Adjektiv *Neues* sowie die drei Adjektive *wissenschaftlichen*, *jungen* und *eigenständige* werden dekliniert, d. h. sie kongruieren durch das Anhängen von Suffixen in Numerus, Genus und Kasus mit ihren zugehörigen Wörtern und Satzteilen. Dabei stimmen die vor und bei Adjektiven auch nach der Lücke stehenden Bezugswörter innerhalb einer Nominalphrase in ihren grammatikalischen Merkmalen überein. So bezieht sich beispielsweise das Indefinitpronomen *etwas*, das stellvertretend für ein Artikelwort steht, auf das substantivierte Adjektiv *Neues*, welches als im Nominativ, Neutrum, Singular durch die Endung *–es* flektiert werden muss, da Genus, Kasus und Numerus im Artikelwort *etwas* nicht angegeben sind. Im weiteren Satzverlauf erfordert die Präposition *von*, die ohne Artikel vor *Oberstufenschülern* steht, für die korrekte Lösung der nachfolgenden Lücke eine Suffigierung zur Bestimmung von Dativ und Plural (*Oberstufenschüler +n*) ebenso für das pränominale Adjektivattribut (*jung +en*) vor *Studierenden*, das pränominale Adjektivattribut *eigenständige* (*eigenständig +e*)

zwischen dem unbestimmten Artikel *eine* und dem Nomen *Fragestellung* wird durch das Flexionsmorphem *-e* in Nominativ, Singular, Femininum angepasst, usw. Auch die flektierten Prädikate weisen eine Numeruskongruenz zum Subjekt des Satzes auf (*Viele Techniken* = Nominativ, Plural → *sind* = 3. Person Plural).

Wenn nun in einem C-Test nicht das Wortende sondern der Anfang jedes zweiten Wortes getilgt wird, bleiben am Ende der zu ergänzenden Lücke gerade diese grammatikalischen Informationen bestehen (Scholten-Akoun et al., 2014, 408 ff.). Muss also der Wortanfang eines Wortes im C-Test ergänzt werden, können die nicht getilgten grammatikalischen Informationen, d. h. vor allem die Derivations- und Flexionsmorpheme am Ende der Lücken sowie die zur Wortgruppe oder zum Satzglied gehörigen Bezugswörter vor und nach der Lücke zwar Hinweise auf die Wortart des zu ergänzenden Wortes geben (in der ersten Lücke zeigt das Derivationsmorphem *-ik* ein feminines Nomen an, das Flexionsmorphem *-en* sowie das Bezugswort *Viele* den Plural, usw.), allerdings kann innerhalb des Textes die inhaltliche Bedeutung der zu ergänzenden Wortstämme bzw. lexikalischen Morpheme am Wortanfang oftmals nur über die vorherige Nennung einer Informationen oder indirekt über eine ähnliche Information erschlossen werden.

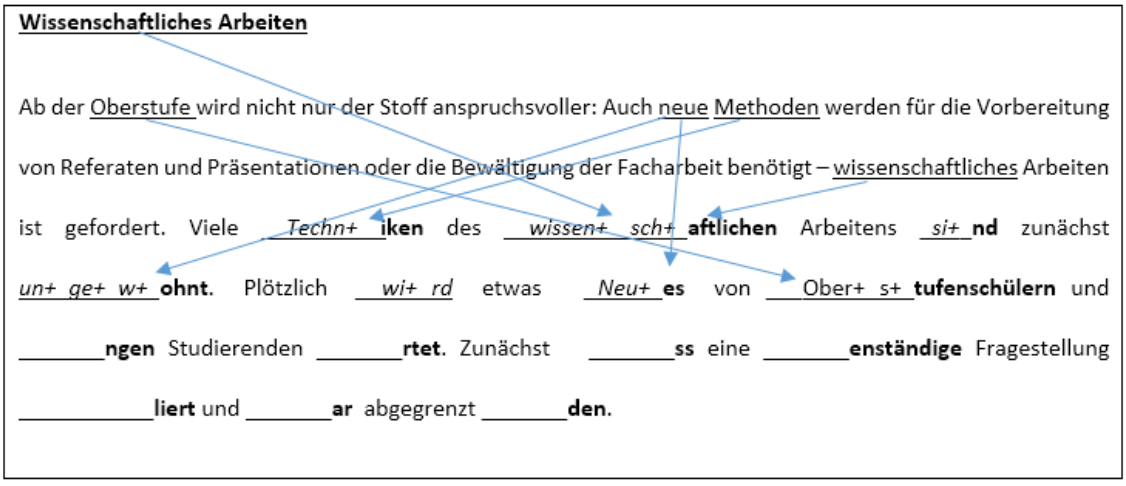


Abbildung 12: Ausschnitt aus dem C-Test *Wissenschaftliches Arbeiten* mit linksseitiger Tilgung

Das Lösen der ersten Lücke im hier gezeigten Ausschnitt eines linksseitig getilgten C-Tests erfordert die Ergänzung der Anfangsbuchstaben *Tech-* zu *Techn**iken***. Zwar gibt es – wie oben gezeigt – durch das Bezugswort *Viele*, das nicht getilgte Ende des Wortes sowie seiner Position am Satzanfang Hinweise auf die Wortart (Nomen) und die syntaktische Funktion (Subjekt) des gesuchten, bzw. zu ergänzenden Wortes, allerdings muss für die inhaltliche Erschließung der Lücke der weitere Kontext miteinbezogen werden. So kann beispielsweise über das Nomen *Methoden* im ersten tilgungsfreien Satz des C-Test-Ausschnitts auf das zu ergänzende Wort *Techniken* geschlossen werden, da beide Nomen in ihrer Bedeutung synonym verwendet werden können und die im ersten Satz beschriebenen *neue(n) Methoden* (des wissenschaftlichen Arbeitens) sowohl inhaltlich als Thema als auch syntaktisch (sowohl *Methoden* als auch *Techniken* stehen im Nominativ, Plural und bilden zusammen mit ihren Bezugswörtern die jeweilige Nominalphrase in der Funktion des Subjekts im Satz) in dem Wort *Techniken* wieder aufgegriffen, bzw. weitergeführt werden. Ähnlich verhält es sich mit der nächsten Lücke, die durch *wissen +sch-* zu *wissensch**aftlichen*** ergänzt werden muss. Grammatikalisch korrekt könnte diese Lücke mit einer Vielzahl von abgeleiteten Adjektiven (von Nomen, die auf *-schaft* enden) mit der Endung *-lich(en)* (z. B. *freundschaftlich(en), leidenschaftlich(en), gesellschaftliche(en), gemeinschaftlich(en)*, usw.) ergänzt werden, um die Lücke jedoch auch semantisch richtig lösen zu können, muss die bedeutungstragende Einheit am Anfang des Wortes durch die Einbeziehung des weiteren Kontextes und redundanter Informationen erschlossen werden. Der Ausdruck *wissenschaftliches Arbeiten* könnte in diesem Fall als redundante Information bezeichnet werden, da dieser sowohl in der Überschrift als auch im Einschub des ersten tilgungsfreien Satzes des C-Tests genannt wird. Im weiteren Verlauf des hier gezeigten, linksseitig getilgten C-Tests verschiebt sich der Akzent der Redundanz also weg von einer grammatischen Kategorie hin zu einer ausschließlich semantischen, wodurch ebenso von *ab der Oberstufe* auf die zu ergänzende Lücke *Obers+ tufenschülern*, von *neue Methoden* auf *etwas Neu +es* und *un +ge +wohnt* geschlossen werden kann. Das im adjektivischen Gebrauch stehende Partizip II *ungewohnt* ist in diesem Fall auch ein Beispiel für eine Bedeutungsänderung durch Präfigierung: Sowohl das Präfix *-un* als auch das Präfix *-ge* bestimmen maßgeblich die Bedeutung des Wortes (***ge**wohnt* hat beispielsweise eine ganz andere inhaltliche Bedeutung als ***be**wohnt*, ***un**gewohnt* verhält

sich antonym zu ***ge**wohnt*), die semantisch auf der Textebene über die Beschreibung *neu(e)* zumindest indirekt impliziert werden könnte (etwas Neues ist eher ungewohnt als gewohnt).

Es gibt aber auch Lücken, bei deren Rekonstruktion nicht direkt auf im Text vorkommende redundante Informationen oder die inhaltliche Einbeziehung des weiteren Kontextes zurückgegriffen werden kann. Vielmehr könnten bei der Ergänzung von Verben syntaktisch-strukturelle Informationen helfen, z. B. bei den Lücken *wird* und *erwartet*: Das Adjektiv *Plötzlich* am Satzanfang erfordert eine Inversion, d. h., dass an zweiter Stelle im Satz (und vor dem Subjekt) ein Verb stehen muss, welches in diesem Fall auf *–rd* endet. Wird das Verb werden, bzw. die Lücke *wird* nun korrekt rekonstruiert, kann über die Satzkonstruktionen (das Hilfsverbs *werden* + Partizip II = Vorgangspassiv Präsens) darauf geschlossen werden, dass es sich bei der letzten Lücke des Satzes um ein passivfähiges Verb handelt, was durch die Flexion *–tet* (*erwar**tet***) bestätigt wird (gleiches gilt für das Zustandspassiv Präsens *sind…ungewohnt* im ersten Satz). Bei anderen Lücken, wie beispielsweise bei *eigenständige, formuliert* oder *klar,* gibt es jedoch keine redundanten oder direkten syntaktischen Informationen, die Hinweise auf die jeweils richtige Rekonstruktion liefern. Vielmehr müssten in diesen Fällen womöglich über die Aktivierung von weiterem Sprach-, Fach- oder Weltwissen mögliche Verbindung zu potentiellen Bezugswörtern hergestellt werden (*eine Fragestellung formulieren* und *eine Fragestellung klar abgrenzen* etwa als Funktionsverbgefüge im bildungssprachlichen, akademischen Kontext).

Anhand der hier verglichenen C-Test-Versionen konnte zumindest exemplarisch gezeigt werden, dass das Lösen der Lücken je nach Tilgungsart (rechts- oder linksseitig) in einem deutschen C-Test in Teilen durch das Einbeziehen von Redundanzen auf unterschiedlichen Ebenen gelingen kann (Stöber, 2011, 307 ff.). So sind es in einem rechtsseitig getilgten C-Test vor allem Redundanzen auf der morpho-syntaktischen Ebene, die bei der Rekonstruktion der Lücken genutzt werden, während bei der Bearbeitung von linksseitig getilgten C-Tests vorrangig Redundanzen auf der (text-) semantischen Ebene Hinweise auf die zu ergänzenden Wortanfänge liefern. Wenn es also nun um die Frage geht, wie die Reduzierung von Redundanz bei linksseitig getilgten

C-Tests sinnvoll bestimmt werden kann, dann erscheint die Anzahl der gelöschten Buchstaben allein als untaugliches Maß, da die Zahl der gelöschten Buchstaben – die bei der Tilgung einer Worthälfte unabhängig vom Tilgungsmuster gleich bleibt (vgl. oben) – nicht erklären kann, warum rechtsseitig getilgte C-Tests nun leichter sind als linksseitig getilgte C-Tests. Vielmehr könnte bzw. sollte beispielsweise berechnet werden, wie viele redundante Informationen sich auf die Lücken in einem C-Test beziehen und ob sich die redundanten Informationen vorrangig auf der Wort-, Satz- oder Textebene herstellen. Dabei kann auf der Grundlage der hier analysierten Beispiele zumindest angenommen werden, dass bei der Wortanfangstilgung, die Informationen gelöscht werden, die gar nicht oder vor allem auf der Textebene redundant sind. So könnten beispielsweise bei der Bearbeitung der zweiten Lücke[45] (*wissenschaftlichen*) im rechtsseitig getilgten C-Test allein auf der morpho-syntaktischen Ebene 5 Hinweise (Derivation *-lich* zum Adjektivattribut über die Wortstellung zwischen den zwei Bezugswörtern bestimmter Artikel (1) und Nomen (1), Kleinschreibung (1); Flexion *-en* über Kasus-Numerus-Genus-Kongruenz beider Bezugswörter (2) innerhalb der Nominalphrase) genutzt werden, um die Lücke am Wortende (vor allem grammatikalisch) korrekt zu lösen oder um Lösungsvorschläge auf ihre Richtigkeit hin zu überprüfen. In einem linksseitig getilgten C-Test gibt es für dieselbe Lücke hingegen nur eine, sich auf die zu erschließende lexikalische Bedeutung am Wortanfang beziehende, redundante Informationen, die – wie oben gezeigt – auf der Textebene zweimal genannt wird (2). Der Vergleich der linksseitigen und rechtsseitigen Tilgung in den hier gezeigten C-Test-Ausschnitten legt also die Vermutung nahe, dass das Lösen linksseitig getilgter C-Tests schwieriger ist als das von rechtsseitig getilgten C-Tests, da für die Rekonstruktion der Wortanfänge weniger Redundanzen genutzt werden können als für die Ergänzung am Wortende. Die von Sigott und Köberl (1994; 1996) empirisch fundierte Aussage, dass die Schwierigkeit eines C-Tests über das Maß der Reduzierung von Redundanz modifiziert werden kann, verliert seine Gültigkeit für einen linksseitig getilgten (deutschen) C-Test damit nicht, da - wie hier im Ansatz gezeigt wurde – über einen Tilgungswechsel vom Wortende auf den

[45] Es soll an dieser Stelle weder darum gehen, die Schwierigkeit einzelner Items zu bestimmen, da diese immer einer statistischen lokalen Abhängigkeit unterliegen noch ausgehend von einzelnen Lücken die Schwierigkeit eines C-Tests vorherzusagen Grotjahn (2002b, S. 216).

Wortanfang die Redundanz im C-Test weiter reduziert wird, aber eben in einer anderen Art und Weise, wie zunächst oben von ihnen vorgeschlagen.

2.4 Gütekriterien

Wenn der C-Test – so wie in dieser Untersuchung – als wissenschaftliches Testinstrument zur Messung von Sprachkompetenz von Oberstufenschüler*innen eingesetzt wird, sollte er einigen methodischen und praktischen Anforderungen genügen (Eckes, 2010, S. 128). Diese Anforderungen bzw. Richtlinien für die Entwicklung und den Einsatz von Tests und Testverfahren im Allgemeinen werden als *Testing Standards* bezeichnet (American Educational Research Association, 2014) und setzen sich u. a. aus den drei Hauptgütekriterien Objektivität, Reliabilität und Validität zusammen (Bühner, 2006; Eckes, 2010, S. 128; Häcker et al., 1998). Welche generellen Anforderungen und in welchem Maß die Testgütekriterien im speziellen für den C-Test gelten (sollen), wird im Folgenden genauer betrachtet[46].

Eines der drei Hauptgütekriterien, die eine wissenschaftliche Untersuchung als notwendige Bedingung erfüllen sollte, ist die Objektivität. Ein Test ist dann objektiv, wenn er von äußeren Faktoren (z. B. Raumbedingungen, Prüfungssituation, Untersuchungsleiter) unabhängig ist (Lienert & Raatz, 1998, S. 7). Das Kriterium der Objektivität spielt beim C-Test besonders in zwei Bereichen eine entscheidende Rolle: bei der Auswertung und bei der Durchführung.

Um beim Einsatz von C-Tests eine Durchführungsobjektivität zu gewährleisten, ist es wichtig, dass die Testdurchführung immer den gleichen Prinzipien folgt: Zunächst sollten die Proband*innen anhand eines Beispiels mit dem Testformat vertraut gemacht werden. Je nach Zusammensetzung der Testpersonen (Alter, Sprachstand, Auffassungsgabe etc.) kann dies über eine schriftliche Anleitung mit Tilgungsbeispielen oder durch das gemeinsame Lösen eines Beispiel-C-Tests auf der Folie am OHP, an der Tafel oder in digitaler Form stattfinden (Baur et al., 2013a, S. 6; Baur & Mashkovskaya,

[46] Inwieweit die in der hier vorliegenden Arbeit entwickelten und eingesetzten C-Tests die Testgütekriterien erfüllen, wird ausführlich in Kapitel 5.2 dargestellt.

2015, 435 ff.; Baur & Spettmann, 2008a). Alle Proband*innen sollten vor Testbeginn die Möglichkeit haben, sich mit dem Arbeitsauftrag bzw. dem Testformat vertraut zu machen und Fragen diesbezüglich stellen zu können, um eventuelle Missverständnisse und Unklarheiten, die das Testergebnis ggf. negativ beeinflussen könnten, zu vermeiden (ebd.)[47].

Ein weiteres Kriterium einer objektiven C-Test-Durchführung ist die möglichst exakte Einhaltung der Bearbeitungszeit. Was die Bestimmung der Bearbeitungszeit von C-Tests betrifft, empfiehlt Grotjahn (2002b, S. 216), diese auf der Grundlage von Erfahrungswerten aus Pre-Testungen und generell eher großzügig zu bemessen, sodass sowohl schnellen als auch langsameren Leser*innen genügend Zeit für die Bearbeitung jedes einzelnen C-Tests zur Verfügung steht (Baur & Mashkovskaya, 2015, S. 446). Da in einigen Untersuchungen nachgewiesen werden konnte, dass sich durch eine separate Zeitvorgabe und Ansage für jeden einzelnen Teiltest die Reliabilitäts- und Validitätswerte von C-Tests steigern lassen, sollten die Proband*innen innerhalb der Bearbeitungszeit nur jeweils einen Teiltest lösen (z. B. Grotjahn, 2010b, 266 ff.; Grotjahn et al., 2010, 299 ff.). Wird der C-Test im Paper-Pencil-Format durchgeführt, ist darauf zu achten, dass die einzelnen C-Tests seitens der Proband*innen beginnend mit dem ersten Teiltest eines C-Test-Sets der Reihe nach und nur nach vorheriger Ankündigung der Testleiter*in bearbeitet werden sollen. Da alle Proband*innen zur gleichen Zeit den gleichen Test lösen sollen, ist ein Vor- und Zurückblättern im Testheft zu untersagen (Baur & Spettmann, 2007, 95 ff., 2008b, S. 433). Wird der C-Test online durchgeführt, gibt es oftmals die Möglichkeit – so wie in der online C-Test-Version in der hier vorliegenden Arbeit – eine Zeitbegrenzung für die Bearbeitung einer Aufgabe bzw. eines C-Tests festzulegen, sodass der jeweilige C-Test nach Ablauf der Zeit nicht weiter bearbeitet werden kann und automatisch abgegeben wird (vgl. Kapitel 4.3 zur Beschreibung der Versuchsdurchführung).

Um zu gewährleisten, dass bei der Testdurchführung für alle Proband*innen die gleichen Bedingungen herrschen und damit das Kriterium der Objektivität gewährleistet ist, sollte

[47] In Studien von Borgards und Raatz (2002) und Thiede (2014) wurde untersucht, inwieweit Personen, die mit dem C-Testformat vertraut sind, einen Vorteil beim Lösen eines haben, mit dem Ergebnis, dass eine Vertrautheit mit dem Testformat bzw. ein vorheriges Testtraining das C-Test-Ergebnis in der Regel nicht positiv beeinflusst.

zudem darauf geachtet werden, Täuschungsversuche wie etwa das Benutzen von Hilfsmitteln (z. B. der Duden, Handys, etc.) oder das Abschreiben von Sitznachbar*innen zu verhindern. In sehr engen Prüfungsräumen sollte möglichen Betrugsversuchen und einer damit einhergehenden Verfälschung von Testergebnissen vorgebeugt werden, indem beispielsweise Trennwände aufgestellt, zusätzliches Aufsichtspersonal oder parallele Testversionen eingesetzt werden (Mashkovskaya, 2014, S. 101).

Die Auswertung von C-Tests ist dahingehend objektiv, als dass sie in der Regel auf der Grundlage von vorgefertigten Lösungsschablonen funktioniert (Grotjahn, 2002b, S. 214; Mashkovskaya & Baur, 2016b, S. 217). In den Lösungsschablonen ist festgelegt, welche Wörter bzw. Wortergänzungen neben den im ursprünglichen Text vorkommenden Originallösungen als Alternativlösungen[48] akzeptiert werden (ebd.). Als Alternativlösungen gelten in der Regel alle zusätzlichen semantisch, orthografisch[49], lexikalisch, syntaktisch korrekten und kontextuell sinnvollen Ergänzungen, die durch Pre-Testungen ermittelt werden können (Grotjahn, 2002b, S. 217). Ein binärer Auswertungsmodus[50] und die vorherige Festlegung der zulässigen Lösungsmöglichkeiten lassen keinen Raum für abweichende oder subjektive Punktevergaben oder personenabhängige Interpretationen und gewährleisten somit eine Auswertungsobjektivität, die sich wiederum positiv auf die Reliabilität von C-Tests auswirkt (Eckes, 2010, S. 128).

Ein weiteres Merkmal für das Testgütekriterium der Objektivität, das für einige C-Teststeller*innen als zu erfüllende Bedingung für einen qualitativ hochwertigen Test angesehen wird, ist Objektivität im Sinne von Fairness. Ein Test gilt dann als fair, wenn alle Proband*innen die gleiche Chance haben, ein gutes Testergebnis zu erzielen und

[48] Untersuchungen von Arras et al. (2002) zu unterschiedlichen Auswertungsverfahren haben gezeigt, dass sich ein Auswertungsmodus, in dem neben der Originallösung auch Alternativlösungen als korrekte Lösung akzeptiert werden, positiv auf Reliabilität und Validität von C-Tests auswirkt (ebd. S. 205).
[49] Je nach Fragestellung und Untersuchungsziel der Testung, kann entschieden werden, ob orthographische Fehler als falsche Antwort gezählt werden oder nicht. Arras et al. (2002) machen jedoch in diesem Zusammenhang deutlich, dass es nur schwer möglich sein dürfte, klare Grenze zwischen „orthographisch falsch“ und „grammatisch falsch“ oder zwischen „orthographisch falsch“ und „lexikalisch falsch“ ziehen zu können Arras et al. (2002, S. 187). Vielmehr steigen sowohl die Auswertungsobjektivität als auch die Reliabilität, wenn orthographische Abweichungen konsequent als Fehler bewertet werden Eckes (2006).
[50] Ein binärer Auswertungsmodus bedeutet in diesem Zusammenhang, dass die richtige Wortergänzung einen Punkt erzielt, keine oder eine falsche Ergänzung mit null Punkten bewertet wird.

zwar unabhängig von bestimmten Gruppenzugehörigkeiten und aufgrund ihres soziokulturellen, ethnischen oder geschlechterspezifischen Hintergrundes keiner systematischen Diskriminierung ausgesetzt werden (*Der Online-Einstufungstest Deutsch Als Fremdsprache (OnDaF): Theoretische Grundlagen, Konstruktion Und Validierung;* Kubinger, 2009; Weber, 2005). Proband*innen unterschiedlicher Zugehörigkeiten sollten also bei gleicher Ausprägung des zu messenden Merkmals das gleiche Testergebnis erzielen können sollten (ebd.).

Als zusätzlichen Aspekt von Fairness nennen u. a. Eckes (2010, S. 128) und Kubinger (2009, 82 ff.) die Verrechnungsfairness (auch als Testgütekriterium *Skalierung* bekannt): Dieses Gütekriterium bezieht sich auf die Auswertung und Analyse von Testantworten und stellt den Anspruch, dass die Vorschrift für die Verrechnung von Teilnehmerantworten in Testwerte die empirischen Verhaltensrelationen der Teilnehmer*innen exakt wiederspiegelt (ebd.). Die einfachste Verrechnungsvorschrift, die üblicherweise auch bei der Auswertung von C-Tests angewendet wird, besteht darin, Testwerte als Anzahl der richtigen Antworten bzw. einen Summenscore zu bestimmen (ebd.). Dieser Verrechnungsmodus ist laut Probabilistischer Testtheorie aber nur dann fair, wenn er dem Rasch-Modell folgt, da „nur bei Gültigkeit des Rasch-Modells der Summenscore die gesamte Information ausschöpft, die ein Antwortmuster über die Ausprägung der Personenfähigkeit enthält“ (Eckes, 2010, S. 129). Welche Vorteile Berechnungen nach Probabilistischer Testtheorie (u. a. mit Rasch-Modellen) gegenüber Methoden der Klassischen Testtheorie zur Erfassung von Testwerten haben, wird im Kapitel 4.4.2 genauer erläutert.

Das zweite der drei Hauptgütekriterien ist die Reliabilität, welche den Grad der Messgenauigkeit eines Testinstrumentes angibt. Reliabel, d. h. zuverlässig, ist ein Test dann, wenn er ein bestimmtes Merkmal exakt und ohne zufällige Messfehler misst (Bühner, 2011, 141 ff.). Das Ausmaß von Reliabilität kann durch vier verschiedene Verfahrensweisen (durch *Paralleltest-Reliabilität*, *Retest-Reliabilität*, *Testhalbierungs-Reliabilität* oder *Innere Konsistenz*) geschätzt werden. Welche Verfahrensweise zur Einschätzung der Reliabilität eines Tests angewendet wird, hängt davon ab, wie dieser eingesetzt wird und welche Testdaten zur Verfügung stehen (ebd.). In Untersuchungen Ihre Reliabilität erreicht in aller Regel Werte von mindestens 0,8, häufig sogar Werte von

0,9 oder darüber zur *Inneren Konsistenz* beispielsweise erreichten C-Tests auf der Ebene der Texte oftmals Cronbachs Alpha-Werte zwischen 0,8 und 0,9 und auch in Test-Retest-Studien lagen die Reliabilitätskoeffizienten 0,8 und 0,9, weshalb die Reliabilität von C-Test als hoch eingestuft wird (Eckes, 2007; Grotjahn et al., 2002, 99 ff.; Mashkovskaya & Baur, 2016a).

Das dritte Hauptgütekriterium befasst sich mit der Validität von Tests. Vereinfacht ausgedrückt ist ein Test dann valide, wenn er das misst, was er messen soll. Analysen zur Validität eines Tests untersuchen also die Übereinstimmung des vom Test gemessenen Merkmals und dem, welches man messen will (Eckes, 2007, S. 70; Moosbrugger & Kelava, 2012, S. 13). Ob ein Testinstrument eine hinreichende Validität aufweist, hängt damit auch immer vom Kontext seiner Nutzung sowie der Verwendung und Einordnung der Testergebnisse ab (Grotjahn, 2019a, S. 602; Zimmermann, 2019, S. 64). Um die verschiedenen Ausprägungen der Validität von C-Tests zu untersuchen – zu nennen sind hier beispielswiese die kriterienbezogene bzw. Übereinstimmungsvalidität, die Konstrukt- und die Inhaltsvalidität –, wurden in zahlreichen Untersuchungen und Studien zum C-Test unterschiedliche Methoden eingesetzt (Eckes & Grotjahn, 2006; Grotjahn, 1995, 2014, 2019b, S. 602; Sigott, 2004). Besonders häufig werden dabei korrelationsstatistische Methoden – ergänzt um Faktorenanalysen – angewendet, die überprüfen, ob der eingesetzte C-Test mit anderen (bereits vorhandenen, meist standardisierten) Außenkriterien (z. B. anderen Sprachtests, Schulnoten, Lehrerurteilen, usw.) korreliert, d. h. ob und inwieweit das in beiden Tests erfasste Merkmal zusammenhängt (Grotjahn, 2019b, S. 602; Zimmermann, 2019, S. 48), bzw., ob der C-Test mit dem ihm zugrunde gelegten theoretischen Konzept übereinstimmt (Klein-Braley, 1985b, S. 55; Zimmermann, 2019, S. 51). Insgesamt ergeben sich in den meisten Untersuchungen für klassische, d. h. rechtsseitig getilgte C-Test, dabei Korrelationen im mittleren bis hohen Bereiche zwischen den eingesetzten C-Tests und unterschiedlichen Außenkriterien, ebenso wie „hohe Ladungen der C-Tests auf einer als allgemeine Sprachkompetenz interpretierbaren latenten Dimension“ (Eckes & Grotjahn, 2006; Grotjahn, 2002b, 2019b, S. 602; Zimmermann, 2019).

Angesichts der Tatsache, dass der C-Test in diversen Kontexten und Bereichen eingesetzt wird, kann die Frage nach seiner Validität an dieser Stelle allerdings kaum

vollumfänglich beantwortet werden. Ebenso sind Aussagen über die Validität von C-Tests nicht als allgemeingültig zu verstehen, sondern immer in Abhängigkeit von der Zielpopulation, der Fragestellung sowie und dem Einsatzbereich zu bewerten (Asano, 2014, S. 50) (Asano, 2014, p. 50).

Da für die in dieser Arbeit eingesetzte Version der linksseitig getilgten C-Tests für die hier fokussierte Kohorte der Oberstufenschüler*innen keine vergleichbaren Studien zur Validität bzw. für linksseitig getilgte C-Tests generell nur wenige Untersuchungen vorliegen, werden im Folgenden exemplarisch einige ausgesuchte Studien zur Validität von C-Tests genannt, deren Kohorte der hier untersuchten entspricht, oder mit Studierenden bzw. Studienbewerber*innen die hier untersuchte am ehesten wiederspiegelen und/oder sich auf den Einsatz linksseitig getilgter C-Tests für Erstsprachsprecher*innen[51] beziehen.

Untersuchungen zur Validität von C-Tests u. a. in der Oberstufe liegen beispielsweise von Wockenfuß (2009; Wockenfuß & Raatz) vor, mit dem Ergebnis, dass der von ihr eingesetzte rechtsseitig getilgte *speeded-C-Test* mit der Deutschnote von Abiturient*innen im mittleren Bereich korreliert (vgl. dazu ausführlich Kapitel 6.4.2), ebenso hängt der C-Test auch u. a. in der Oberstufe mit verbaler Intelligent sowie Informationsverarbeitungsgeschwinigkeit zusammen (Wockenfuß, 2009, 180 ff.; Wockenfuß & Raatz, 2014, 189 ff.).

Studien mit rechtsseitig getilgten C-Test für deutschlernende Studierende bzw. Studienbewerber*innen, wie beispielsweise die Untersuchungen von Eckes (2014), in denen der Autor der Frage nachgeht, inwieweit sich der auf C-Tests basierende Online-Einstufungstest Deutsch als Fremdsprache (*onDaF*) zur Vorhersage für deutsche Sprach- bzw. akademische Zulassungsprüfung für Studienbewerber*innen eignet (Eckes, 2014, 138 ff.; Grotjahn, 2014, S. 17), zeigen, dass der C-Test – ähnlich wie in vorherigen

[51] Die meisten Untersuchungen zur Validität für den C-Test als Messinstrument zur Erfassung von Sprachkompetenz beziehen sich auf den fremdsprachlichen Bereich (inklusive Deutsch als Fremdsprache) und werden u. a. von Arras et al. (2002), Sigott (2004) und Eckes and Grotjahn (2006) detailliert gegenübergestellt.

Untersuchungen[52] – u. a. hoch mit allen vier Subtests von *TestDaF*[53] (Leseverstehen, Hörverstehen, Schriftlicher Ausdruck und Mündlicher Ausdruck) korreliert (Eckes, 2014, 138 ff.): Auf der Grundlage der Niveaueinstufungen nach dem GER ergeben sich zwischen den *onDaF* C-Tests und *TestDaF* Korrelationen von 0.60 für den Bereich Leseverstehen, von 0,62 für das Hörverstehen und den Schriftlichen Ausdruck und von 0,54 für den mündlichen Ausdruck[54](Eckes, 2014, 138 ff.). Mithilfe einer Hauptkomponentenanalyse konnte außerdem gezeigt werden, dass die Korrelation zwischen den Gesamtergebnissen von *TestDaF* und den *onDaF* C-Tests bei 0,71 und insgesamt noch höher liegen (Eckes, 2014, S. 158, 2016). Grotjahn sieht in dieser Studie eine (weitere) Bestätigung dafür, dass der untersuchte C-Test für die Gruppe der Studienbewerber*innen „allgemeine Sprachfähigkeit im Deutschen" misst und argumentiert, dass der von Eckes entwickelte C-Test „möglicherweise in einem beschränkten Maß auch Kommunikationsfähigkeit im Deutschen im akademischen Kontext zu erfassen erlaubt" (Grotjahn, 2014, S. 19).

Auch Asano (2014) konnte in ihrer Untersuchung sowohl mit studentischen Deutschler*innen als auch mit Erstsprachsprecher*innen substantielle, signifikante Korrelationen zwischen dem von ihr eingesetzten rechtsseitig getilgten C-Test und *TestDaF* nachweisen. Die Autorin sieht in ihren Ergebnissen u. a. ein Indiz dafür, dass der C-Test für die Gruppe der Studierenden nicht nur allgemeine Lese- und Schreibfähigkeiten misst, sondern als integratives Testinstrument alle vier klassischen Fertigkeitsbereiche von Sprache (Lesen, Schreiben, Hören, Sprechen) in gewissem Maße abbilden kann (Asano, 2014).

Für Studierende mit Deutsch als Erst- oder Zweitsprache ist Mashkovskaya (2014) im Rahmen ihrer Dissertation der Frage nachgegangen, ob und inwieweit linksseitig getilgte C-Tests Lesefertigkeiten valide messen können. Dazu ließ sie Lehramtsstudierende die von ihr entwickelten, linksseitig getilgten C-Tests und verschiedene Lesetests

[52] Vergleichbare Studien zu Korrelationen zwischen onDaF und TestDaF mit ähnlichen Ergebnissen sind nachzulesen bei Eckes und Grotjahn (2006).
[53] Der TestDaF ist eine international anerkannte, standardisierte Sprachprüfung für Deutsch als Fremdsprache auf den GER-Niveaustufen B2-C1 und dient ausländischen Studierenden als Sprachnachweis für eine deutsche Hochschulzugangsberechtigung.
[54] Auf der Grundlage der reinen *Testwerte* ergaben sich zwischen *onDaF* und *TestDaF* Korrelationen von 0.66 für das Leseverstehen und 0.69 für das Hörverstehen (Eckes (2014).

bearbeiten. Der erste Lesetest umfasst eine Kombination aus einer standardisierten Multiple-Choice- und einer Zuordnungsaufgabe, die aus dem Aufgabenpool von *TestDaF* entnommen wurden. Der zweite eingesetzte Lesetest ist ein Stolperwörtertest, der ursprünglich von Metze für die Primarstufe entwickelt und von Mashkovskaya u. a. an die Zielgruppe der erwachsenen Erstsprachsprecher*innen angepasst und kohortenspezifisch modifiziert wurde (Mashkovskaya, 2014). Durch Analysen der gewonnenen Testwerte konnte Mashkovskaya (2014) zeigen, dass der C-Test Lesefertigkeiten valide misst, da zwischen dem C-Test und den TestDaF-Lesetests mittlere Korrelationen (zwischen 0,40 und 0,58) und sogar hohe Korrelationen zwischen C-Test und Stolperwörtertest (je nach Auswertungsmodus des Stolperwörtertests zwischen 0,71 und 0,82) nachgewiesen wurden.

In ihrer Studie, die ebenfalls die Sprachkompetenz von (Lehramts-)Studierenden in der Studieneingangsphase mit linksseitig getilgten C-Tests untersucht, haben Bremerich-Vos und Scholten-Akoun (2016) neben den linksseitig getilgten C-Tests auch eine kombinierte Lese-/Schreibaufgabe eingesetzt und liefern damit Forschungsergebnisse zur Validität linksseitig getilgte C-Tests für erwachsene Erst- (und Zweit-)Sprachsprecher*innen. Die Autor*innen kommen zu dem Ergebnis, dass zwischen den meisten Globalvariablen und dem eingesetzten C-Test mittlere Korrelationen (z. B. Stilistischer Gesamteindruck -0,42, Sprachliche Kompetenz -0,38, Sprachlicher Gesamteindruck -0,35), zwischen detaillierteren Einzelvariablen und dem C-Test jedoch eher schwache Zusammenhänge bestehen (z. B. Wortgebrauch: Register -0,16, Grammatik: Richtigkeit – 0.27, Wortgebrauch: Richtigkeit -0.20).

Neben diesen ersten Ergebnissen für linksseitig getilgte C-Tests im Bereich Deutsch als Erst-/Zweitsprache und den Ergebnissen der vorliegenden Untersuchung dieser Arbeit (vgl. Kapitel 5.3) stehen weitere Untersuchungen zur Validität von linksseitig getilgten C-Tests noch aus und sind wünschenswert, um den Forschungsstand um Erkenntnisse für diese C-Test-Version zu erweitern.

Eine weitere Anforderung, deren Erfüllung einige Teststeller durchaus eine wichtige Bedeutung beimessen, ist die Testökonomie. Die Ökonomie bezieht sich auf die Wirtschaftlichkeit eines Testes und das damit verbundene Verhältnis zwischen Kosten (finanziell und zeitlich) und Nutzen (Moosbrugger, 2012c). Gemessen am

Erkenntnisgewinn nimmt der Einsatz des C-Tests relativ wenig finanzielle und zeitliche Ressourcen in Anspruch. Sowohl bei der Entwicklung und Durchführung als auch bei der Auswertung eines C-Tests entsteht ein zeitlich eher geringer Aufwand. Die Durchführungszeit beträgt nach einer kurzen Einführung und bei einer (standardisierten) Bearbeitungszeit von 5 Minuten pro Teiltest – je nach Anzahl der Teiltests im C-Test-Set – ca. 30 bis 45 Minuten. Eine manuelle Auswertung mit Lösungsschablone ist erfahrungsgemäß auch von ungeübten Codierer*innen in 1-3 Minuten pro Teiltest und ggf. abhängig von der Auswertungsmethode zu erledigen (Grotjahn, 2002b).

Was den finanziellen Aufwand betrifft, kann davon ausgegangen werden, dass bei Verwendung des „Paper-Pencil-Formats“ mit üblicherweise einem Teiltest auf einer DIN A4 Seite, die entstehenden Druckkosten überschaubar sind. Dennoch können die Beauftragung von Personen für Codier- und/oder Aufsichtstätigkeiten, anfallende Lizenzgebühren für oder etwa die Expertise von Statistiker*innen oder Psychometriker*innen zusätzliche Kostenfaktoren darstellen.

Mittlerweile gehen einige Testteller*innen dazu über, Online-C-Testversionen zu realisieren, wie z. B. Eckes (2010, 2014) im Rahmen von *onDaF* oder die Autorin bei der Datenerhebung dieser Arbeit. Aus ökonomischer Sicht hat das zwei Vorteile: Durch die Bearbeitung von C-Tests am Computer, Tablet oder Smartphone können Druck- und Materialkosten eingespart werden. Zum anderen entfällt durch eine systematische computergestützte Auswertung der Testergebnisse die Korrekturzeit. Die von den Proband*innen eingegeben Antworten werden im System erfasst und bei korrekter Lösung mit einem Punkt verrechnet. Unmittelbar nach Beendigung des Tests stehen den Proband*innen (ggf. in anonymisierter Form auch dem Prüfer) die Testergebnisse zur Verfügung (vgl. auch Kapitel 4.3 und 4.4). Eckes (2010) weist darauf hin, dass somit auch eine Komponente des Kriteriums der Zumutbarkeit aus Sicht der Testteilnehmer*innen erfüllt ist, da den Proband*innen durch die sofortige Ergebnisrückmeldung lange Wartezeiten erspart bleiben (ebd.). Unabhängig davon, ob der C-Test im Papier-und-Bleistift-Format oder als computergestützte Version zum Einsatz kommt, kann davon ausgegangen werden, dass der C-Test aufgrund seiner Aufgabenstellung und

Bearbeitungszeit die Testpersonen in zeitlicher, körperlicher und psychischer Hinsicht nicht übermäßig belastet und damit im Allgemeinen zumutbar ist (Eckes, 2010).

3. Hypothesen

Aufgrund der in Kapitel 2 besprochenen theoretischen Überlegungen zur Sprachkompetenz in der Sekundarstufe II sowie der Ergebnisse aus Studien zur Sprachkompetenz im schulischen (z. B. DESI, PISA) als auch im akademischen Eingangsbereich (z. B. SkaLa®) (vgl. Kapitel 2) sollen im Rahmen dieser Arbeit Vermutungen zum einen zum hier eingesetzten C-Test und zum anderen zur Verteilung der C-Test-Ergebnisse bzw. der Sprachkompetenz der Oberstufenschüler*innen empirisch überprüft werden. Im Folgenden werden dazu Hypothesen aufgestellt, die sich auf Zusammenhänge zwischen dem C-Test und anderen Variablen oder auf gruppenspezifische Unterschiede hinsichtlich der C-Test-Ergebnisse beziehen.

3.1 Zur Validität des C-Tests

Ausgehend vom Ziel dieser Arbeit, einen C-Test für den Einsatz in der gymnasialen Oberstufe zur Messung der Sprachkompetenz von Schüler*innen zu entwickeln, zu erproben und zu evaluieren, wurde zur Überprüfung der Validität der C-Test-Neuentwicklung ein standardisierter und normierter Lesetest eingesetzt sowie die Deutschnoten der teilnehmenden Oberstufenschüler*innen über einen Fragebogen miterfasst (vgl. Kapitel 4.1). Da aus verschiedenen Studien bereits bekannt ist, dass mit dem C-Test allgemein – und im Besonderen mit linksseitig getilgten C-Tests, wie Mashkovskaya (2014) in ihrer Untersuchung zur Sprach- und Lesekompetenz von Studierenden mit Deutsch als Erst- bzw. Zweitsprache durch mittlere bis hohe Korrelationen nachweisen konnte (Mashkovskaya, 2014, S. 241) – Facetten von Lesekompetenz und Textverständnis erfasst werden kann (Baur, Spettmann 2010, S. 95 f.), wurde vor dem Hintergrund der hier vorgestellten Untersuchung beschlossen, den LGVT 6-12, für den für die Oberstufe gesicherte Normwerte vorliegen (vgl. Kapitel 4.1.1), zur Validierung einzusetzen. Dabei wird für den entwickelten C-Test für die Oberstufe angenommen, dass Oberstufenschüler*innen, die gute Leistungen im C-Test erzielen, auch gut im LGVT 6-12 abschneiden, d. h.:

1. ***Es gibt einen statistisch signifikanten Zusammenhang zwischen dem in der Oberstufe eingesetzten, linksseitig getilgten C-Test und dem standardisierten Lesetest LGVT 6-12.***

Zur Überprüfung der Konstruktvalidität von C-Tests gibt es außerdem verschiedene Studien, in denen der Zusammenhang zwischen der C-Test-Leistung und der Note im erstsprachlichen bzw. fremdsprachlichen Bereich untersucht wurde (Wockenfuß, 2009, S. 137). Denn auch wenn das Heranziehen von Noten zur Bestimmung der Validität von C-Tests aufgrund einer möglicherweise geringen Objektivität der Noten und des klasseninternen Maßstabes prinzipiell problematisch[55] sein kann (Ingenkamp, 1995; Mashkovskaya & Baur, 2016a; Wockenfuß, 2009, S. 137), wird die Messung von schriftsprachlichen Kompetenzen im schulischen Bereich – und in Bezug auf die hier vorliegende Studie im Besonderen im Bereich der gymnasialen Oberstufe – durch die Vergabe von Punkten bzw. Noten zumindest indirekt als Teil eines komplexeren Kompetenzfächers operationalisiert (vgl. Kapitel 2.2).

Für Schüler*innen der unteren Klassenstufen beschreiben Klein-Braley und Raatz (1984) beispielsweise, dass für diese Gruppe in der Regel ein zum Teil großer Zusammenhang zwischen Deutschnoten und C-Test-Leistungen besteht (S. 145). Kesper (1995) und Gummich (1997), die den Zusammenhang zwischen Deutschnote und C-Test-Leistung bei älteren Schüler*innen systematisch untersucht haben, kommen zu dem Schluss, dass auch bei älteren Schüler*innen ein Zusammenhang zwischen Deutschnote und C-Test-Leistung besteht, dieser jedoch abnimmt, je höher die Klassenstufe ist und für die Gruppe der Studierenden wesentlich niedriger liegt als für jüngere Schüler*innen (In: Wockenfuß 2009, S. 137). Dieses Ergebnis bestätigt auch Wockenfuß (2009) in ihrer Untersuchung zur Entwicklung der allgemeinen Sprachkompetenz in der Erstsprache von Gymnasialschüler*innen und Studierenden, interessanterweise jedoch mit der Einschränkung, dass der Zusammenhang bei sich in der Abiturphase befindenden Proband*innen wieder stärker wird und im mittleren Bereich liegt (Wockenfuß, 2009,

[55] Ausführlicher zur Problematik von klasseninternen Maßstäben bei Überprüfung von Korrelationen zwischen der Deutschnote und dem C-Test-Ergebnis in Kapitel 5.4.2.

174 f., 181). Da sich die Proband*innen der hier vorgestellten Studie alle in der gymnasialen Oberstufe, d. h. in der Vorbereitung auf das Abitur bzw. in der Abiturphase befinden, soll überprüft werden, ob ein ähnlicher Befund auch für die hier untersuchte Kohorte mit dem eingesetzten C-Test bestätigt werden kann.

Ausgehend von der dieser Arbeit zugrunde liegenden Definition von Sprachkompetenz bzw. deren operationalisierte Erfassung über den C-Test als integrativen Sprachtest und die Deutschnote als kumulative Summe aus verschiedenen Fähigkeiten und Fertigkeiten, wird angenommen, dass es zwar eine Schnittmenge von schriftsprachlichen Kompetenzen gibt, die gleichermaßen mit dem C-Test gemessen als auch in der Deutschnote (mit-)erfasst werden und die sich rechnerisch in einer mittleren Korrelation zwischen C-Test und Deutschnote zeigt:

2. ***Es gibt einen mittleren, statistisch signifikanten Zusammenhang zwischen dem eingesetzten C-Test und der Deutschnote.***

3.2 Zu den Hintergrundvariablen

Eine weitere im Rahmen dieser Arbeit zu überprüfende Hypothese bezieht sich auf die in Kapitel 1 diskutierte und auf den dort dargestellten Befunden (z. B. zur Sprachkompetenz der Lehramtsstudierenden) basierende Annahme, dass die Entwicklung von Sprache bzw. basalen sprachlichen Kompetenzen mit dem Eintritt in die gymnasiale Oberstufe nicht vollständig abgeschlossen sein kann und sich die schriftsprachlichen Kompetenzen auch im Laufe der Oberstufe – nicht zuletzt durch die Vermittlung und Instruktion vertiefender rezeptiver und produktiver sprachlicher Kompetenzen – weiterentwickeln (vgl. Kapitel 1). Dass sich die globale Sprachfähigkeit von Jugendlichen und jungen Erwachsenen bis zum 25. Lebensjahr linear weiterentwickelt, konnte beispielsweise auch Wockenfuß (2009) empirisch mit dem von ihr eingesetzten C-Test bestätigen. Vor diesem Hintergrund sollte es möglich sein, mit dem hier entwickelten C-Test einen Leistungszuwachs bei den Ergebnissen der Schüler*innen der Qualifikationsphase 1 im Vergleich zu den Ergebnissen der Schüler*innen aus der Einführungsphase nachzuzeichnen. Es wird also angenommen:

*3. Schüler*innen in der Qualifikationsphase 1 erzielen signifikant höhere Leistungen im C-Test als Schüler*innen in der Einführungsphase.*

Die Ergebnisse internationaler Schulleistungsstudien wie PISA, DESI, IGLU/PIRLS, in denen u. a. sprachliche Kompetenz als zentraler Faktor für eine erfolgreiche Bildungskarriere untersucht wurde, haben gezeigt, dass der Bildungserfolg und die Bildungschancen von Kindern in Deutschland in beträchtlichem Maße vom regionalen, sozialen und familiären Kontext sowie der ethnischen Herkunft abhängen (DESI-Konsortium, 2008; Weis, Müller et al., 2019, 129 ff.). Damit gibt es im deutschen Schulsystem systematische Ungleichheiten zwischen Gruppen von Schüler*innen, die mit bestimmten Merkmalen einhergehen (Reiss et al., 2016, 377 ff.; Weis, Müller et al., 2019, 129 ff.). So zeigen beispielsweise die aktuellen PISA Ergebnisse (2015, 2018) weiterhin[56] Leistungsdisparitäten in der Kompetenzverteilung zwischen Schüler*innen aus sozial schwachen und sozial starken Familien (Müller & Ehmke, 2016, 307 ff.), ebenso wie nach wie vor bestehende Leistungsunterschiede zwischen Mädchen und Jungen (Müller & Ehmke, 2016, 247 ff.; Weis, Doroganova et al., 2019, S. 62) als auch Kompetenzrückstände bei Jugendlichen mit Zuwanderungshintergrund[57] (Reiss et al., 2016, S. 377). Zudem verdeutlichen beispielsweise die DESI-Ergebnisse (2003/2004) ebenso wie die Befunde aus PISA (2018) – vor dem Hintergrund einer mehr an der Sprachbiografie als am Migrations- bzw. Zuwanderungsstaus der Jugendlichen orientierten Unterteilung der Gruppen –, dass Schüler*innen, die neben Deutsch mindestens eine weitere Sprache in ihrem Alltag verwenden, im deutschen Schulsystem nicht die gleichen Chancen haben wie diejenigen, die nur deutschsprachig sind. Selbst wenn die mehrsprachigen Schüler*innen angeben, im Alltag überwiegend Deutsch zu sprechen, erzielen sie schlechtere Leistungen als monolingual deutschsprachige Jugendliche (Hesse et al., 2008, 208 ff.; Petersen, 2014c, S. 24).

Ausgehend von diesen Befunden soll nun im Rahmen dieser Arbeit der Frage nachgegangen werden, ob sich auch in der hier untersuchten Gruppe von Schüler*innen

[56] Hier im Vergleich zu den Ergebnissen aus PISA 200 und PISA 2009.

[57] Bei PISA wird der Zuwanderungshintergrund wie folgt eingeteilt: Ein Elternteil im Ausland, ein Elternteil in Deutschland geboren, 1. Generation (Die Jugendlichen als auch ihre Eltern sind im Ausland geboren), 2. Generation (Die Jugendlichen sind in Deutschland geboren, beide Eltern im Ausland) (Weis, Müller et al. (2019, S. 134).

aus der gymnasialen Oberstufe – einem Bereich in dem traditionellerweise eher von relativ homogenen Lerngruppen ausgegangen wird – gruppenspezifische Unterschiede in Bezug auf das Geschlecht, den Bildungsgrad der Eltern, die Sprachverwendung der Schüler*innen, den Schulstandort sowie die Schulform in der mit dem C-Test gemessenen Sprachkompetenz abzeichnen. Dazu wurden folgende Hypothesen formuliert, die es auf der Grundlage der in dieser Studie erhobenen Daten zu überprüfen gilt:

4. ***Oberstufenschülerinnen erzielen signifikant höhere C-Test-Ergebnisse als Oberstufenschüler.***

5. ***Es gibt einen statistisch signifikanten Zusammenhang zwischen den C-Test-Leistungen und dem Bildungsgrad der Eltern.***

6. ***Zur Sprachverwendung:***

    a. ***Monolingual deutschsprachige Schüler*innen erzielen eine höhere C-Test-Leistung als mehrsprachige Schüler*innen.***

    b. ***Bei einer anderen Erstsprache als Deutsch, erzielen Schüler*innen, die im familiären Kontext Deutsch als Sprache verwenden, durchschnittlich höhere C-Test-Leistungen als Schüler*innen, die überwiegend eine andere Sprache als Deutsch innerhalb der Familie sprechen.***

    c. ***Es gibt einen statistisch signifikanten Unterschied zwischen den C-Test-Ergebnissen unterschiedlicher Sprachgruppen.***

7. ***Es gibt einen statistisch signifikanten Zusammenhang zwischen dem C-Test-Ergebnis und dem Schulstandorttyp.***

8. ***Die Schüler*innen an Gymnasien erzielen signifikant höhere C-Test-Leistungen als die Schüler*innen an Gesamtschulen.***

4. Methode

Im folgenden Kapitel wird das methodische Vorgehen zur Erfassung der schriftsprachlichen Kompetenzen mit dem C-Test in der gymnasialen Oberstufe sowie zur Beantwortung der Forschungsfragen (vgl. Kapitel 1) bzw. der in Kapitel 3 formulierten Hypothesen besprochen. Zunächst werden die im Rahmen dieser Studie eingesetzten Testinstrumente vorgestellt, bevor beschrieben wird, wie die Stichprobe konstruiert und die in dieser Arbeit vorliegenden Daten erhoben wurden. Der letzte Teil dieses Kapitels befasst sich mit der Auswertung des C-Tests – unter Einbeziehung unterschiedlicher Testtheorien und verschiedener (statistischer) Verfahren zur Überprüfung der Hypothesen.

4.1 Die Testinstrumente

Zur Erfassung der Sprachkompetenz von Oberstufenschüler*innen mit dem C-Test und zur Erhebung von Hintergrunddaten für die Beantwortung der in Kapitel 3 formulierten Hypothesen, folgt diese Untersuchung dem hier skizzierten methodischen Vorgehen: Zur Überprüfung der Validität wurde in der vorliegenden Studie u. a. der LGVT 6-12 als standardisierter Lesetest eingesetzt, dem sich der erste Teil des nachfolgenden Kapitels widmet. Anschließend wird die Entwicklung des für die Zielkohorte der Oberstufenschüler*innen adressatenspezifischen C-Tests sowie die Konstruktion des Fragebogens zur Erhebung der Hintergrundvariablen vorgestellt.

4.1.1 Der LGVT 6-12

Der Name LGVT 6-12 steht für *Lesegeschwindigkeits- und Verständnistest für die Klassenstufen 6-12* und bezeichnet einen für den Sekundarbereich konzipierten, standardisierten Test zur direkten Erfassung des Leseverständnisses auf der Textebene und zur indirekten Messung des Leseverstehens über die Lesegeschwindigkeit (Schneider et al., 2007, S. 5). Der LGVT 6-12 besteht für alle Klassenstufen (6-12) aus

einem (dem gleichen) Fließtext, d. h. einer Art Kurzgeschichte (mit der Überschrift *Brot und Rosenkohl*) von insgesamt 1.727 Wörtern, bei dessen Bearbeitung an 23 Stellen im Text aus drei Wahlalternativen das korrekte – d. h. das am besten in den Textzusammenhang passende – Wort ausgewählt werden muss (Schneider et al., 2007, 5 ff.). Die Überprüfung des Leseverständnisses wird hierbei über die Anzahl der richtig ausgewählten Wörter und die Lesegeschwindigkeit über die Anzahl der gelesenen Wörter innerhalb der vorgegebenen Bearbeitungszeit (4 Minuten) gemessen (Schneider et al., 2007). Um den Rohwert für die Beurteilung der Lesegeschwindigkeit zu erfassen, werden die Schüler*innen nach Ablauf der Bearbeitungszeit dazu aufgefordert, das letzte von ihnen gelesene Wort zu markieren. Dieses Verfahren setzt dabei natürlich voraus, dass die Schüler*innen nach dem Signal zur Beendigung des Tests die Stelle im Text markieren, bis zu der sie in der vorgegebenen Zeit tatsächlich auch nur gelesen haben – eine Überprüfung der von den Proband*innen gemachten Angabe zur Anzahl der gelesenen Wörter ist allerdings – wenn der Test als Gruppentest eingesetzt wird und die Schüler*innen leise für sich lesen – kaum möglich.

Die Auswertung für das Leseverständnis erfolgt über einen Lösungsschlüssel, nach dem jede richtige Lösung, d. h. jedes korrekt ausgewählte Wort, mit 2 Punkten, keine Lösung mit 0 Punkten und eine falsche Lösung mit einem Minuspunkt gewertet wird (ebd.). Da der Text unter Einhaltung einer starken Zeitlimitierung von den Proband*innen gelesen, bzw. möglichst viele richtige Wörter ausgewählt werden sollen, ist es denkbar, dass die Proband*innen nicht den gesamten Text sondern nur jene Passagen bzw. Sätze lesen, in denen Wörter ausgewählt werden sollen. Eine solche Strategie könnte zumindest bei einigen Lücken zum Erfolg führen, die auch oder vor allem über Vor- und Weltwissen gelöst werden könnten, beispielsweise wenn in Bezug auf das Aussehen von kleinen Rosenkohlköpfen die Antwortmöglichkeiten *grün*, *rot* und *groß* vorgegeben sind (*grün* ist hier die richtige Antwort). Um flüchtiges und überspringendes Lesen seitens der Proband*innen zu vermeiden sowie den Anreiz zum Raten einer Lösung zu schmälern, scheint es ebenfalls sinnvoll, vor Beginn der Testung darauf hinzuweisen, dass falsch ausgewählte Wörter mit Minuspunkten bewertet werden.

Ungünstig ist in diesem Zusammenhang auch die Tatsache, dass der hier eingesetzte LGVT 6-12 aus nur einem Fließtext besteht und sich damit inhaltlich auf nur einen

Themenbereich – eben Brot und Rosenkohl – bezieht, was Proband*innen, die in diesem Bereich über Vor- und Weltwissen verfügen, einen Vorteil in der Bewältigung der Aufgabe verschaffen, bzw. Proband*innen, die sich weniger gut mit Brot und Rosenkohl – z. B. dem Aussehen der Kohlsorte – auskennen, benachteiligen könnte. In der aktuellen und überarbeiteten Auflage des LGVT (5-12+) haben die Testentwickler*innen den Lesetext *Brot und Rosenkohl* allerdings um zwei weitere Lesetexte mit alternativen Themen ergänzt. Der erweiterte und überarbeitete LGVT 6-12+ stand allerdings zum Zeitpunkt der Datenerhebung der hier vorgestellten Untersuchung nicht zur Verfügung.

Prinzipiell gehen die Testentwickler*innen des LGVT 6-12 davon aus, dass Lesegeschwindigkeit und Leseverständnis eng miteinander verbunden sind, da das Lesen – in Anlehnung an die in PISA und den Bildungsstandards zugrundeliegende Definition von Lesekompetenz (Jude et al., 2013, S. 201) – selbst einen aktiven Prozess der Entnahme von Informationen aus geschriebenen Texten und ihrer Verknüpfung mit Vor- und Weltwissen darstellt, welcher bei leseschwachen Personen besonders durch fehlende Automatisierung der technischen Aspekte beeinträchtigt wird (Artelt et al., 2005; Schneider et al., 2007, S. 8). Muss also nun viel Zeit und Mühe auf der Ebene des Dekodierens eines Textes verwendet werden, sind „die Möglichkeiten der aktiven Sinnentnahme“ deutlich eingeschränkt (Schneider et al., 2007, 5 ff.). Damit stellt die Lesegeschwindigkeit – so die Autor*innen weiter – „eine einfach zu messende Größe dar, um Informationen über basale Lesekompetenz zu gewinnen“ (Schneider et al., 2007).

Im Rahmen dieser Arbeit wurde der LGVT 6-12 – trotz dieser kritischen Anmerkungen – zur Überprüfung der Validität des neu entwickelten C-Tests eingesetzt. Vor dem Hintergrund einer so gut wie nicht vorhandenen Testlandschaft in der Oberstufe (vgl. Kapitel 1 und 2), läuft die Auswahl eines geeigneten Testinstruments bzw. Lesetests zur Validierung des C-Tests mehr oder weniger zwangsläufig auf den LGVT 6-12 zu, da sich dieser Test als Gruppentest in der gymnasialen Oberstufe eignet und für dieses Verfahren für den Bereich der Klassenstufen 10-12 gesicherte Normwerte[58] vorliegen.

[58] Für die Klassenstufen 6-9 liegen ebenfalls gesicherte Normwerte für Haupt-, Real-, Sekundar- und Gesamtschulen sowie Gymnasien vor, die anhand einer Stichprobe von insgesamt 2.390 Schüler*innen aus 11 Bundesländern berechnet wurden. In der erweiterten Neuauflage des LGVT 6-12+ von 2017

Ebenfalls konnten die Testentwickler*innen in Untersuchungen zur Validität empirisch zeigen, dass mit dem LGVT 6-12 Lesekompetenz effizient gemessen werden kann, was beispielsweise eine relativ hohe Korrelation von 0,59 zwischen dem Lesetest und Leseaufgaben aus PISA 2000 belegt (Schneider et al., 2007, S. 18).

4.1.2 Entwicklung des C-Tests

Da der C-Test im Rahmen der hier vorgestellten Untersuchung als zentrales Testinstrument zur Erhebung der Sprachkompetenz von Oberstufenschüler*innen eingesetzt wurde, sind bei seiner Entwicklung die von Raatz und Klein-Braley (1985, 96 ff.) und Grotjahn (2002a)[59] vorgeschlagenen grundlegenden Schritte und Anweisungen zur Testkonstruktion befolgt worden, auf die im Folgenden ausführlich eingegangen werden soll.

Vor der Erstellung eines C-Tests sollte zunächst in Abhängigkeit von der Zielgruppe bestimmt werden, welche Teststruktur für die Proband*innen geeignet erscheint. Die „klassische“ Teststruktur ist für Fremdsprachenlerner*innen konzipiert worden. Ein C-Test besteht in der Regel aus einem Testset aus 4 Texten mit jeweils 25 Lücken (oder alternativ auch aus 5 Texten mit jeweils 20 Lücken), also insgesamt aus (mindestens) 100 Lücken (Raatz & Klein-Braley, 1985, 96 ff.). Für die in dieser Studie untersuchte Kohorte der Oberstufenschüler*innen sollten ebenfalls 4 adressatenspezifische C-Tests mit jeweils 25 Lücken entwickelt werden. Zusätzlich zu diesen 4 C-Tests wurden im Rahmen dieser Untersuchung zwei weitere C-Tests aus der Untersuchung zur Sprachkompetenz von Studierenden (Bremerich-Vos & Scholten-Akoun, 2016) als Ankerung eingesetzt. Die Ankertests aus der Studieneingangsphase dienen einerseits dazu, die schulischen C-Tests zu validieren bzw. zu justieren und anderseits die Messergebnisse der Oberstufenschüler*innen direkt mit den Ergebnissen der Studierenden vergleichbar zu machen (vgl. Kapitel 5.4). Für die hier vorgestellte Untersuchung wurde also ein Testset

basieren die Normen auf einer Stichprobe von 7.142 Schüler*innen aus 13 Bundesländern (Schneider et al. 2017).

[59] Nachzulesen online auf der von Grotjahn angelegten Seite u. a. zum Aussehen, zur Konstruktion von C-Tests (http://www.c-test.de/deutsch/index.php?lang=de&content=konstruktion&section=ctest).

aus insgesamt 6 C-Tests (4 neu entwickelte C-Tests + 2 Ankertest) mit jeweils 25 Lücken generiert (vgl. unten). Für die Bearbeitung der einzelnen C-Tests wurde eine Zeitbegrenzung[60] von 5 Minuten pro C-Test festgelegt. Eine solche Zeitvorgabe (5 Minuten/Text mit 25 Lücken) wurde in den Untersuchungen von Mashkovskaya (2014) und in SkaLa® (Mashkovskaya & Baur, 2016a, S. 199) sowie in der Pre-Testung[61] dieser Studie für die Gruppe der fortgeschrittenen Erst- bzw. Zweitsprachsprecher*innen[62] als angemessen eingestuft, da die Verteilung der C-Test-Ergebnisse unter einer Bearbeitungszeit von 5 Minuten – ebenso wie in den hier erhobenen Daten aus der Pre-Testung und der Haupterhebung (vgl. Kapitel 4.2) weder Decken- noch Bodeneffekte aufweist, sondern zu einer breiten Varianz in der Untersuchungsgruppe führt.

Um einen C-Test zu erstellen, werden zunächst verschiedene, in sich geschlossene, Texte benötigt, die die Grundlage der zu entwickelnden C-Tests bilden und die dem Sprachniveau der Untersuchungsgruppe entsprechen. Um vorauszusetzen, dass die Inhalte der Texte allen Proband*innen im Wesentlichen bekannt sind und mit dem C-Test vor allem sprachliche Fertigkeiten und Fähigkeiten und nicht vorrangig Fach- oder Weltwissen getestet wird, schlagen (Baur et al., 2013b, S. 3) für die Entwicklung von C-Tests im schulischen Bereich vor, Texte bzw. Themen zu verwenden, die in der Klassenstufe unter der zu untersuchenden eingesetzt wurden. Dieser Empfehlung folgend wurden zunächst authentische Texte aus unterschiedlichen Deutschbüchern, Arbeitsheften und Zeitschriften, die sich auf die in den Kernlehrplänen Deutsch für das Ende der Sekundarstufe I an Gymnasien und an Gesamtschulen ausgeschriebenen Inhalte, Themen und Anforderungsbereiche bzw. auf die Einführungsphase der gymnasialen Oberstufe beziehen (z. B. aus *Zeit online für Schüler*, *Finale.*

[60] Wockenfuß (2009, S. 148), Grotjahn, Schlak, Aguado (2010) und Grotjahn (2010b) konnten in ihren Untersuchungen zeigen, dass eine Verkürzung der zur Verfügung gestellten Zeit pro Test die Schwierigkeit des Tests erhöht: Dabei gilt, je weniger Zeit für das Ausfüllen des Lücken vorgesehen ist, desto schwieriger wird der Test. Diese Möglichkeit der Regulation der Schwierigkeit des Tests gab der C-Test-Variante ihren Namen: Speeded-C-Tests (ebd.). Da die Schwierigkeit der in dieser Untersuchung eingesetzten C-Tests über die linksseitige Tilgung reguliert werden sollte, wurde von einer Erhöhung der Schwierigkeit über eine Speed-Komponente bei der Bearbeitung abgesehen (vgl. auch Kapitel 2.3.4). Vorschläge für den schulischen Bereich beziehen sich hauptsächlich auf die Klassenstufen 4-7. So schlagen etwa Baur, Chlosta & Goggin (2012, S. 19) und Baur, Spettman & Wrede-Jackes (2013, S. 7).

[61] Eine genaue Beschreibung der Pre-Testung folgt in Kapitel 4.1.2.

[62] Ausführliche Informationen zu modifizierten C-Tests für fortgeschrittene Erst-, Zweit- und Fremdsprachsprecher*innen siehe Kapitel 2.3.4.

Prüfungstraining 2015, *KombiKOMPAKT*, *Texte Themen und Strukturen. Deutschbuch für die Oberstufe, Der Deutschunterricht,* usw.), herausgesucht[63].

Für die Erstellung der C-Tests wurden die Ausgangstexte *Der Reiz von Fernsehserien* (Weber, 2014, S. 119) und *Medien – Ethik – Wirklichkeit* (Rohrhirsch, 2008, S. 144), die beispielsweise den Anforderungsbereich „zwischen Wirklichkeit und virtuellen Welten in Medien unterscheiden" abdecken, ebenso wie der Text *Wissenschaftliches Arbeiten* (Kelch, 2014), welcher sich auf den Anforderungsbereich „Informationsquellen gezielt nutzen" bezieht, ausgewählt (MSB NRW, 2004, 13 ff., 2007, 14 ff., 2014). Auch die Texte *Das Problem des Übersetzens* (Biermann, 1999, S. 306), *Wie reden wir denn da?* (Romberg, 2012, 143 f.), *Die Diskussion* (Biermann et al. 2004, S. 69), *Einfluss der Globalisierung* (Dahmen, 2014, S. 48) sowie *Körpersprache* (Molcho, 1999, S. 83) folgen einer inhaltlichen Anlehnung an die in der 9. bzw. 10 Klasse zu vermittelnden Phänomene wie „Wirkungen der Redeweise kennen, beachten und situations- sowie adressatengerecht anwenden" u. a. in Bezug auf Sprechtempo, Stimmführung und Körpersprache (Gestik, Mimik), „ausgewählte Erscheinungen des Sprachwandels kennen und bewerten" wie beispielsweise Veränderungen von Wortbedeutungen und fremdsprachliche Einflüsse, „Aufmerksamkeit für verbale und nonverbale Äußerungen entwickeln", „beim Sprachhandeln die Inhalts- und Beziehungsebene im Zusammenhang mit den Grundfaktoren sprachlicher Kommunikation erkennen und berücksichtigen" sowie Sprachvarietäten kennen und in ihrer Funktion unterscheiden, beispielsweise in Bezug auf Standardsprache, Umgangssprache, Gruppensprachen, Fachsprachen, gesprochene und geschriebene Sprache" (ebd.).

Der fachliche Bezug zum Unterrichtsfach Deutsch bei der Auswahl der Texte liegt darin begründet, dass das Fach Deutsch als Pflichtfach in der Sekundarstufe I von jeder Schüler*in belegt werden muss und nicht abgewählt werden kann. Dies setzt zumindest theoretisch voraus, dass die inhaltlichen (Anforderungs-)Bereiche aus der Sekundarstufe I den Schüler*innen in der Oberstufe bekannt sind. Diese Voraussetzung

[63] Prinzipiell gilt, dass die für die Entwicklung von C-Tests herangezogenen Texte möglichst authentisch und inhaltsneutral sein sollten, um mögliche themenbezogene, emotionale Aufladung zu vermeiden. Ebenso sollten die Texte der zu untersuchenden Zielkohorte bekannte Themen und weitgehend vertraute sprachliche Konventionen aufweisen (vgl. auch Kapitel 2.3.4 in dieser Arbeit zur Modifikation von Schwierigkeit im C-Test).

ist für andere Fächer nicht durchgehend gegeben. Naturwissenschaftliche oder gesellschaftswissenschaftliche Fächer werden schulabhängig in der Sekundarstufe I nicht oder nicht im gleichen Umfange unterrichtet. Deshalb sollten fachthematisch spezialisierte Texte aus diesen Disziplinen bzw. Schulfächern vermieden werden, da solche Texte gewisse Schüler*innen bevorzugen und andere benachteiligen könnten. Natürlich ist an dieser Stelle nicht von der Hand zu weisen, dass im Prinzip jede Textauswahl für die Erstellung eines C-Tests im schulischen Bereich auch unter Rückgriff auf Schulbuchtexte problematisch ist: Auch wenn davon ausgegangen werden kann, dass die ausgewählten Texte die Schüler*innen sprachlich nicht überfordern, weil sie Texte auf diesem Niveau bereits im vorangegangenen Schuljahr kennengelernt haben (sollten), könnten die unterschiedlichen Voraussetzungen der Probanden (beispielsweise Umfang und Qualität des zuvor erteilten Unterrichts, Motivation seitens der Schüler*innen, usw.) in noch zu untersuchender Art und Weise voraussagen, wie gut die Schüler*innen mit Texten aus diesem Fachgebiet zurechtkommen. Untersuchungen dieser Art wären besonders im schulischen Bereich mit C-Test-Konstruktionen leicht realisierbar und wünschenswert – auch zur Evaluation des erteilten Unterrichts.

Nach der Auswahl geeigneter Texte beginnt die Beschädigung der Wörter im zweiten Satz. Der erste sowie der letzte Satz des Textes bleiben tilgungsfrei und geben dem Leser eine kurze Einleitung in den Text und das Thema bzw. zusätzlichen Kontext. In der hier eingesetzten Variante sind alle Teiltests zudem mit Überschriften versehen, die thematisch auf den zu lesenden Text bezogen sind und dem Leser zu einer ersten Orientierung dienen sollen. Die Überschriften für die C-Tests wurden aus den Grundlagentexten übernommen – teilweise wurden die Überschriften nach Beratung in einer Expertengruppe zur besseren Verständlichkeit geringfügig verändert, beispielsweise dann, wenn sich die Überschrift auf breite Themenfelder bezog (z. B. bei dem Text *Medien – Ethik – Wirklichkeit*, wobei sich der Textabschnitt für den C-Test jedoch vor allem auf *Den Einfluss der Medien* bezog) und lauten: *Der Reiz von Fernsehserien, Das Problem des Übersetzens, Körpersprache, Sprachverfall?, Der Einfluss der Medien, Wissenschaftliches Arbeiten, Sprachwandel* und *Die Diskussion*.

Dem Rhythmus der 2-er Tilgung folgend wurde in den ausgewählten Texten nun die Hälfte jedes zweiten Wortes linksseitig gelöscht[64]. Dabei wurden die zu tilgenden Buchstaben durch einen von der Anzahl der zu tilgenden Buchstaben unabhängigen, immer gleichbleibend langen Unterstrich ersetzt. Prinzipiell sollte eine Tilgungsdarstellung der Buchstaben durch eine der Zahl der getilgten Buchstaben entsprechende Zahl der Unterstriche vermieden werden, da die Anzahl der Unterstriche die Länge des zu rekonstruierenden Wortes vorgeben würde und bei den Probanden möglicherweise ein Abzählen der Buchstaben induzieren würde, was ggf. einen negativen Effekt auf die Validität des C-Tests haben könnte (Grotjahn, 2002b, S. 212). Für die Erprobung in den angestellten Pre-Testungen wurden C-Tests im Papier-Bleistift-Format verwendet, in der Hauptuntersuchung kam das finale Testset vorwiegend als Online-Version[65] zum Einsatz. Bestand ein zu tilgendes Wort aus einer ungeraden Buchstabenzahl, so wurde – gemäß den Vorgaben zur Erstellung eines C-Tests der oben genannten Autor*innen (Baur et al., 2013b, S. 3; Grotjahn, 2002b, S. 212; Klein-Braley & Raatz, 1985) – ein Buchstabe mehr gelöscht. Eigennamen, Zahlen, Abkürzungen und geografische Bezeichnungen wurden bei der systematischen Beschädigung des Textes übersprungen und nicht getilgt (ebd.). Nach Erreichen einer Anzahl von 25 im jeweiligen Text eingebauten Tilgungen wurde der letzte Satz ohne weitere Lücken bis zum Ende ausgeschrieben.

Diesem Prinzip folgend wurden 8 neue C-Tests für den Einsatz in der gymnasialen Oberstufe entwickelt, die in einem nächsten Schritt von einer Expertengruppe (kompetente, erwachsene Erstsprachsprecher*innen aus dem akademischen Bereich) evaluiert und bearbeitet wurden. Durch die Bearbeitung wurden erste Alternativlösungen gesammelt, die zwar nicht den ursprünglichen Wörtern in den Originaltexten entsprachen, jedoch bei der Lektüre der von den Proband*innen erstellten Rekonstruktion als semantisch, orthographisch, grammatikalisch korrekte Lösung akzeptiert werden sollten (Grotjahn, 2002b, S. 217, 2019b, 585 ff.)[66], als auch

[64] Zum Unterschied zwischen rechtsseitig und linksseitig getilgten C-Tests vgl. Kapitel 2.3.5.
[65] Nähere Informationen zur (technischen) Umsetzung und Durchführung der online C-Test-Version folgen in Kapitel 4.3 zur Beschreibung der Versuchsplanung und Versuchsdurchführung.
[66] Grotjahn weist in seiner Anleitung zur Konstruktion von C-Tests darauf hin, dass eine größere Anzahl an akzeptablen Lösungsalternativen die Reliabilität des entsprechenden Tests geringfügig erhöhen kann, die Punktwerte der Lerner sich dadurch jedoch nicht wesentlich verändern (2002, S. 217). Ebenso

kleinere Modifikationen am Tilgungsmuster vorgenommen. So wurde bei einigen Lücken beispielsweise ein Buchstabe weniger getilgt, wenn die Rekonstruktion der vorgegebenen Buchstaben zu einem Wort zu schwierig oder nicht eindeutig war bzw. es zu viele Antwortmöglichkeiten gab: Bei der Lücke *von* ***_____ngen*** *Oberstufenschülern* (*von* ***jungen*** *Oberstufenschülern*) das ***–n*** von ***jungen*** nicht getilgt, da die Lücke mit einer Vielzahl von Adjektiven, Pronomen oder Zahlwörtern, die auf ***–gen*** enden (z. B. muti**gen**, lusti**gen**, mündi**gen**, eini**gen**, weni**gen**, usw.), ergänzt werden könnte) .

4.1.2.1 Voruntersuchungen

Um alle 8 neu entwickelten C-Tests zeitgleich an einer der Zielkohorte entsprechenden Stichprobe erproben zu können, wurden zwei verschiedene C-Test-Sets (Testheft A und Testheft B) mit jeweils 4 unterschiedlichen C-Tests zusammengestellt. Da die Sprachkompetenz der Oberstufenschüler*innen – wie eingangs erwähnt – im Rahmen der hier vorliegenden Untersuchung zumindest in Teilen mit der der Studierenden verglichen werden soll, wurden beide Testhefte um jeweils zwei gleiche Ankertests aus der SkaLa®-Untersuchung (s. o.) ergänzt:

Tabelle 2: C-Test-Sets

	Testheft A	**Testheft B**
1. C-Test **(Ankertest)**	Warum Schule Angst macht	Warum Schule Angst macht
2. C-Test	Das Problem des Übersetzens	Die Diskussion
3. C-Test	Körpersprache	Wissenschaftliches Arbeiten
4. C-Test **(Ankertest)**	Sind Jungen die neuen Bildungsverlierer?	Sind Jungen die neuen Bildungsverlierer?
5. C-Test	Der Reiz von Fernsehserien	Sprachverfall
6. C-Test	Sprachwandel	Der Einfluss der Medien

steigt die Auswertungsobjektivität, wenn orthographische Fehler konsequent als Fehler bewertet werden (siehe auch Eckes & Grotjahn 2006b).

Die Testhefte *A* und *B* wurden im Rahmen der Pre-Testung an zwei Gymnasien jeweils sowohl in der Einführungsphase (Klassenstufe 10) als auch in der Qualifikationsphase 1 (Klassenstufe 11)[67] an insgesamt 116 Schüler*innen erprobt. Nach der Testdurchführung erfolgte die Auswertung der C-Testhefte dichotom, d. h. jedes semantisch, orthographisch[68] und grammatikalisch korrekt ergänzte Wort ergabt einen Punkt, jede falsche oder nicht ausgefüllte Lösung keinen. Pro Test konnten so maximal 25 Punkte erreicht werden, sowie ein Gesamtscore von maximal 150 Punkten. Auf der Grundlage der Ergebnisse der Pre-Testung konnten mithilfe einer Itemanalyse erste Einschätzungen über die Schwierigkeit und Reliabilität der einzelnen C-Tests bzw. der Testhefte vorgenommen werden. Ziel der Pre-Testung war es, möglicherweise „ungeeignete" Tests zu identifizieren und gegebenenfalls aus der Testbatterie zu entfernen, bzw. aus den 8 neu entwickelten C-Tests ein finales Testheft (bestehend aus 4 neuen C-Tests und den beiden Ankertests) zu generieren (Wockenfuß, 2009, S. 156). Dazu wurden für die einzelnen C-Tests bzw. für die Testhefte mithilfe der Statistik- und Analysesoftware SPSS die Schwierigkeitsindizes, Mittelwerte, Standardabweichungen, Trennschärfe und Reliabilität berechnet.

Tabelle 3: Itemanalyse und Schwierigkeitsindizes Testheft A

Testheft A (n=65)	*P*	$\bar{x}$	*s*	$r_{i(x-i)}$
1. C-Test (Schulangst)	44,12	11,03	4,91	0,660
2. C-Test (Übersetzen)	70,08	17,52	3,24	0,618
3. C-Test (Körpersprache)	53,16	13,29	4,92	0,716
4. C-Test (Bildungsverlierer)	43,76	10,94	5,27	0,739
5. C-Test (Fernsehserien)	72,36	18,09	4,58	0,739
6. C-Test (Sprachwandel)	51,80	12,95	4,39	0,767

[67] Zur Erläuterung des Aufbaus der gymnasialen Oberstufe in NRW an Gymnasien (und Gesamtschulen) siehe Kapitel 5.2.

[68] Grotjahn (2002, 1995) empfehlen für die Auswertung von C-Tests für fortgeschrittene Lerner, „orthographische Abweichungen zur Erhöhung der Auswertungsobjektivität und -ökonomie als Fehler zu werten" (Grotjahn, 2002, S. 217). Dieses Verfahren wurde für die hier untersuchte Zielkohorte von Oberstufenschüler*innen übernommen.

C-Test Gesamt	56,11	84,17	21,91	

P (Schwierigkeit), $\bar{x}$ (Mittelwert), s (Standardabweichung), $r_{i(x-i)}$ (Trennschärfe)

Tabelle 4: Tabelle: Itemanalyse und Schwierigkeitsindizes Testheft B

Testheft B (n=51)	*P*	$\bar{x}$	*s*	$r_{i(x-i)}$
1. C-Test (Schulangst)	47,84	11,96	3,68	0,558
2. C-Test (Diskussion)	64,24	16,06	4,07	0,717
3. C-Test (Wissensch. Arbeiten)	64,56	16,14	3,81	0,684
4. C-Test (Bildungsverlierer)	42,68	10,67	4,62	0,700
5. C-Test (Sprachverfall)	60,56	15,14	4,79	0,749
6. C-Test (Medien)	63,36	15,84	4,04	0,589
C-Test Gesamt	57,16	85,75	19,28	

P (Schwierigkeit), $\bar{x}$ (Mittelwert), s (Standardabweichung), $r_{i(x-i)}$ (Trennschärfe)

In beiden Tabellen sind neben den Mittelwerten ($\bar{x}$), Standardabweichungen und Trennschärfekoeffizienten ($r_{i(x-i)}$) die Schwierigkeitsindizes (P) der einzelnen C-Tests sowie des jeweiligen Testhefte insgesamt dargestellt. Der Schwierigkeitsindex (P) charakterisiert die Lösungswahrscheinlichkeit eines C-Tests, indem er die prozentuale Häufigkeit, mit der der C-Test in der Pre-Testung gelöst werden konnte, angibt. Das bedeutet also, dass ein niedriger P-Wert (= der Test wurde prozentual seltener gelöst) einen offenkundig schwieriger zu lösenden C-Test ausweist und umgekehrt (Lienert & Raatz, 1998, S. 73; Mashkovskaya, 2014, 188 f.; Wockenfuß, 2009, S. 156).

Prinzipiell sollten die Schwierigkeitsgrade einzelner Aufgaben, die mit der Formel $P_i = \frac{\bar{x}_i}{max_{xi}} \times 100$ berechnet werden können, wobei $\bar{x}_i$ für den Mittelwert des Tests und max_{xi} für die Anzahl der Lücken im C-Test stehen, zwischen P=20 und P=80 liegen (ALTE, 2005, 16 f.; Moosbrugger & Kelava, 2012; Wockenfuß, 2009, S. 156). Die Schwierigkeitsindizes der in dieser Arbeit in der Pre-Testung eingesetzten einzelnen C-Tests liegen in beiden Testheften zwischen P=42,68 und P=72,36 und damit innerhalb

der oben angegebenen Spanne und gelten daher als weder zu leicht noch zu schwierig. Der C-Test *Der Reiz von Fernsehserien* ist im Testheft A der leichteste, in Testheft B ist es der C-Test *Wissenschaftliches Arbeiten*. Sowohl in Testheft A als auch in Testheft B bilden die Ankertests, die aus den Untersuchungen aus der Studieneingangsphase (SkaLa®) übernommen wurden, für die Stichprobe der Oberstufenschüler*innen die schwierigsten C-Tests. Beide Testhefte erreichen insgesamt einen ähnlichen Schwierigkeitsindex (P=56,11 und P=57,16) und liegen damit im Bereich der mittleren Schwierigkeit – eine günstige Ausgangslage, da davon auszugehen ist, dass Items mit mittlerer Schwierigkeit (P=50) am besten zwischen unterschiedlichen Merkmalsausprägungen, d. h. zwischen kompetenten und weniger kompetenten Probanden differenzieren (Hartig & Jude, 2004; Mashkovskaya, 2014, S. 189).

Die in der letzten Spalte dargestellten Trennschärfekoeffizienten der einzelnen C-Tests geben Auskunft darüber, inwieweit die erbrachte Leistung in den einzelnen C-Tests mit dem Gesamtergebnis des jeweiligen Testhefts zusammenhängt, d. h. dass die Trennschärfe die Korrelation zwischen den Teil-C-Tests und dem gesamten C-Test darstellt (Fisseni, 2004, 36 ff.; Wockenfuß, 2009, S. 158). Ein hoher positiver Wert eines Trennschärfekoeffizienten (max. +1) steht kennzeichnend für einen trennscharfen C-Test, der eine Vorhersage darüber erlaubt, dass eine Personen mit einem hohen Ergebnis in einem einzelnen C-Test ebenfalls einen hohen Wert im Gesamtscore des Testheftes erzielt, bzw. ein niedriges Einzel-C-Test-Ergebnis einen niedrigen Gesamtwert vorhersagt (ebd.). Je höher also der Trennschärfekoeffizient ist, desto eher vermag der C-Test zwischen leistungsschwachen und leistungsstarken Personen zu trennen, vorausgesetzt, dass alle C-Tests das Gleiche, d. h. in diesem Fall Sprachkompetenz (vgl. Kapitel 2 zur Definition von Sprachkompetenz in dieser Arbeit) messen (Mashkovskaya, 2014, S. 182; Stangl, 2019). Wie aus den oben gezeigten Tabellen ersichtlich wird, variieren die Trennschärfekoeffizienten der einzelnen C-Tests in beiden Testheften zwischen $r_{i(x-i)}$= .558 und $r_{i(x-i)}$= .767 und und liegen damit alle deutlich über der

häufig genannten Ausschluss-Untergrenze von $r_{i(x-i)}$= 0,30[69] (Fisseni, 2004, S. 36; Hartig & Jude, 2004).

Um die interne Konsistenz zu bestimmen, wurde für beide Testhefte der Reliabilitätskoeffizient (Cronbachs Alpha) bestimmt, der für das Testheft A einen Wert von α= .886 für das Testheft A und α= .855 für das Testheft B als gut eingestuft werden können (Blanz, 2015, 169 ff.)[70].

4.1.2.2 Zusammenstellung des finalen C-Tests

Aus den 8 erprobten C-Tests soll nun ein finales Testheft aus 4 neu entwickelten C-Tests und den beiden Ankertests (*Schulangst* und *Bildungsverlierer*) generiert werden (s. o.). Für die Auswahl der 4 finalen C-Tests wurde der P-Wert als entscheidendes Kriterium genutzt. Es wurden also die Tests ausgewählt, die sich in den Pretests als mittelschwer/als die schwersten/als die leichtesten herausgestellt hatten, wobei die Anordnung der C-Tests einer Art Schwierigkeitsdramaturgie folgt: Die Schüler*innen sollen zunächst den laut Pretests „einfachsten" der C-Tests innerhalb eines Testhefts lösen, was den Probanden erstens den Einstieg in den Test erleichtern und zweitens möglichst motivierend wirken soll. Die beiden Ankertests wurden in einem inversen Muster in das Testheft integriert. Lieferte der P-Wert alleine – etwa bei annähernder Wertegleichheit – keine eindeutige Entscheidungsgrundlage, wurde die Trennschärfe als zusätzliches Merkmal betrachtet.

Die Analyse der Pre-Testergebnisse unter Anwendung der beiden genannten Kriterien führte zu folgender Zusammensetzung des in der Datenerhebung eingesetzten C-Testheft von insgesamt 6 C-Tests (4 neu erprobte und zwei Ankertests):

[69] Inwieweit es sinnvoll ist oder nicht, Items aus einem Test zu entfernen, die statistische Grenzwerte unterschreiten, wird bei Bühner (2011, S. 247) diskutiert.
[70] Ausführliche Informationen zur Berechnung der Gütekriterien bzw. zu Reliabilitätsanalysen für C-Tests in Kapitel 3.4 und in Kapitel 6.3 und 6.4.

1. *Der Reiz von Fernsehserien*
2. *Wissenschaftliches Arbeiten*
3. *Sprachverfall?*
4. *Körpersprache*
5. *Sind Jungen die neuen Bildungsverlierer?*
6. *Warum Schule Angst macht*

4.1.3 Fragebogenkonstruktion zur Erfassung (sprachbiographischer) Hintergrundvariablen

Um in den C-Test-Ergebnissen Gruppenspezifika erkennen zu können, wurde den Proband*innen der hier abgedruckte Fragebogen (vgl. unten Abb. 13) zusammen mit den C-Tests vorgelegt.

Angaben zur Person:

1. **Geschlecht:** ○ männlich ○ weiblich

2. **Alter:** _____

3. **Sind Sie in Deutschland geboren?** ○ Ja ○ Nein

4. **Sind Ihre beiden Eltern in Deutschland geboren?** ○ Ja ○ Nein

5. **Was ist Ihre Muttersprache[1]?** _____

6. **Welche Sprache sprechen Sie überwiegend in Ihrer Familie – auch Deutsch kann hier genannt werden?**
a) mit Ihrer Mutter: _____
b) mit Ihrem Vater: _____
c) mit Ihren Geschwistern: _____

7. **Welche Sprache(n) verwenden Sie im Kontakt mit Ihren Freunden?**
○ nur Deutsch
○ überwiegend Deutsch
○ teils Deutsch/teils meine Muttersprache (andere als Deutsch)
○ überwiegend meine Muttersprache (andere als Deutsch)
○ nur meine Muttersprache (andere als Deutsch)
○ andere Sprache(n): _____

8. **Wurden Sie in Deutschland eingeschult?** ○ Ja ○ Nein

9. **Welche Schulform/Stufe besuchen Sie?**
Gymnasium:
○ Einführungsphase ○ Qualifikationsphase 1 ○ Qualifikationsphase 2

Gesamtschule:

o Einführungsphase o Qualifikationsphase 1 o Qualifikationsphase 2

10. In welcher Form belegen Sie (zukünftig) das Fach Deutsch?
o Grundkurs o Leistungskurs

11. Was war Ihre letzte Zeugnisnote für das Fach Deutsch? _____

12. Planen Sie nach dem Abitur ein Studium aufzunehmen?
o Ja o Nein o weiß ich noch nicht

13. Welcher ist der höchste Bildungsgrad Ihrer Eltern?

Mutter

o Hauptschule o Abitur o kein Bildungsabschluss

o Realschule o akademischer Abschluss o weiß ich nicht

Vater

o Hauptschule o Abitur o kein Bildungsabschluss

o Realschule o akademischer Abschluss o weiß ich nicht

_____

[1] Mit *Muttersprache* meinen wir die Sprache, die Sie als erste Sprache in der Kommunikation mit Ihren Bezugspersonen (also Ihren Eltern, Ihren Geschwistern usw.) verwendet haben.

Abbildung 13: Fragebogen zur Erfassung der Hintergrundvariablen

Neben Fragen zum Alter, dem Geschlecht[71], der Schulform und Kurswahl, der Deutschnote, einem eventuellen Studienwunsch und zum Bildungsgrad der Eltern sollten vor allem Informationen zur Erstsprache und zur Sprachverwendung für die Überprüfung der im Rahmen dieser Arbeit entstandenen Hypothesen (vgl. Kapitel 3) erfasst werden. Für die explizite und separate Erfassung von Sprachvariablen in sprachbezogenen Kontexten und Fragestellungen – d. h., dass, wenn man Sprachkompetenzen untersuchen will (und eigentlich auch mit „Migrationshintergrund“ in schulischen Zusammenhängen Sprachkompetenz hintergründig meint), auch Sprachdaten erheben sollte (Chlosta & Ostermann, 2005, S. 63) – plädieren auch Chlosta and Ostermann (2005, 55 ff.). So stellen die Autoren im Rahmen der SPREEG-Untersuchung zur Sprachenvielfalt an Essener Grundschulen beispielsweise fest, dass einerseits 17 % der Kinder mit Migrationshintergrund keine andere Sprache als Deutsch

[71] Im Zeitraum der Datenerhebung war die Nennung eines dritten Geschlechts bzw. einer „dritten Option“ über die Kategorie *divers* durch das BVerfG noch nicht eingeführt bzw. verpflichtend (BGBl. I Seite 2635).

in den Familien sprechen und andererseits 5,5 % der Schüler*innen ohne Migrationshintergrund mehrsprachig sind (Chlosta & Ostermann, 2005, S. 60). Über die separate Erhebung der Sprachvariablen soll also die tatsächliche Sprachverwendung erfasst werden, da es im Rahmen dieser Arbeit eben um die Überprüfung sprachlicher Kompetenzen geht sowie um die Frage, ob sich die Verwendung einer anderen Sprache möglicherweise positiv oder negativ auf die Kompetenz im Deutschen auswirkt.

4.1.3.1 Anmerkungen: Migrationshintergrund – Muttersprache – Erstsprache – Familiensprache – Mehrsprachigkeit – Bildungsgrad der Eltern

In empirischen Studien und aus der Sprach- und Bildungsforschung zeigt die Beschreibung der untersuchten Kohorte bzw. der Ergebnisse häufig eine Differenzierung nach Gruppen „mit und ohne Migrationshintergrund“[72] (Chlosta & Ostermann, 2005, S. 55). Was mit dem Merkmal *Migrationshintergrund* gemeint ist, ist weder im wissenschaftlichen Diskurs noch in amtlichen Statistiken und erhobenen Schuldaten (NRW) konsensuell festgelegt. Die ebenfalls unterschiedliche Terminologie wie z. B. *Migrationshintergrund*, *Migrationsgeschichte*, *Zuwanderungshintergrund* oder *Zuwanderungsgeschichte* lässt bereits vermuten, dass auch die zugrundeliegenden definitorischen Bestimmungen der Termini differiert. Häufig wird das Geburtsland und/oder die Staatsangehörigkeit als maßgeblicher Indikator erfasst. Ist der Befragte (also im schulischen Kontext die Schüler*in) in einem anderen Land als Deutschland geboren oder besitzt er einen Pass eines anderen Staates als Deutschland, wird ihm ein Migrationshintergrund unterstellt. Über u. a. diese Kriterien werden beispielsweise in den amtlichen Schuldaten NRW, in den Daten des Statistischen Bundesamtes, der KMK und den PISA-Studien der *Migrationshintergrund* bzw. die *Zuwanderungshintergrund* definiert. Während allerdings die KMK den Migrationshintergrund alleine über die Merkmale *Staatsangehörigkeit* und *Geburtsland* der Befragten (also in diesen Fällen der Schulkinder) bestimmt (KMK, 2020, S. 7) erfasst das Statistische Bundesamt nicht nur die Daten der befragten Personen, sondern zusätzlich u. a. auch die Staatsangehörigkeit

[72] So zeigen beispielsweise auch die PISA Ergebnisse (2000-2009), dass die Lesekompetenz von 15-jährigen in Deutschland mit Migrationshintergrund insgesamt signifikant niedriger ist als die Lesekompetenz von Jugendlichen ohne Migrationshintergrund (Stanat, Rauch et al. 2010, S. 211) (vgl. Kapitel 1 und Kapitel 3).

der Eltern der Befragten und leitet daraus einen Migrationshintergrund ab: Laut Definition des Statistischen Bundesamtes liegt ein Migrationshintergrund bei einer Person dann vor, „wenn sie selbst oder mindestens ein Elternteil die deutsche Staatsangehörigkeit nicht durch Geburt besitzt“ (Statistisches Bundesamt, 2019, S. 4).

Bei den amtlichen Schuldaten, die in NRW von den Schüler*innen erhoben werden, werden sowohl Merkmale der Schüler*innen als auch der Eltern zur Bestimmung einer *Zuwanderungsgeschichte* erfasst. Kriterium hier ist das Geburtsland (und nicht die Staatsangehörigkeit). Gefragt wird, wo eine Schüler*in und wo die Eltern der Schüler*in geboren wurden. Ist eine Schüler*in oder auch ein Elternteil nicht in Deutschland geboren worden, geht das nordrheinwestfälische Schulministerium von einer *Zuwanderungsgeschichte* aus (zur Definition vgl. unten; MSB NRW, 2019a, S. 7). In den PISA-Studien (2006-2018) wird zur Definition des „Zuwanderungsstatus“ bzw. des „Zuwanderungshintergrundes“ u. a. ebenfalls das Geburtsland der Jugendlichen und ihrer Eltern herangezogen, wobei zwischen folgende Kategorien unterschieden wird:

Ohne Zuwanderungshintergrund	Kein Elternteil im Ausland geboren.
Ein Elternteil im Ausland geboren	Ein Elternteil im Ausland, ein Elternteil in Deutschland (bzw. im jeweiligen OECD-Teilnehmerstaat) geboren.
Zweite Generation	Beide Elternteile im Ausland geboren, Jugendlicher in Deutschland (bzw. im jeweiligen OECD-Teilnehmerstaat) geboren
Erste Generation	Beide Elternteile und Jugendlicher im Ausland geboren.

Abbildung 14: Zur Definition des Zuwanderungshintergrunds bzw. des Zuwanderungsstatus nach PISA 2018

Quelle: Weis, Müller et al., 2019, S. 135

In Anlehnung an diese Vorgehensweise wurde die Festlegung des Migrationshintergrundes über die Fragebogenitems *Sind Sie in Deutschland geboren?* (Frage 3) und *Sind Ihre beiden Eltern in Deutschland geboren?* (Frage 4) operationalisiert.

Von einem *Migrationshintergrund* wird im Rahmen dieser Studie dann ausgegangen, wenn die Heranwachsende selbst im Ausland und/oder mindestens ein Elternteil im Ausland geboren wurde, d. h. also immer dann, wenn Frage 4 oder Frage 3 und Frage 4 mit einem „nein“ beantwortet wurden. Ob dabei nur ein oder beide Elternteile im Ausland bzw. in Deutschland geboren wurden, wird in dieser Untersuchung aus Gründen der Fragebogenökonomie – und vor dem Hintergrund, dass im Kontext der Fragestellung dieser Arbeit vor allem Variablen zur tatsächlichen Sprachverwendung von besonderem Interesse sind (vgl. unten) – nicht separat miterfasst. Wird im Fragebogen NUR die Frage 3 mit „nein“ beatwortet, wird in der hier vorgestellten Untersuchung nicht auf einen Migrationshintergrund geschlossen. Vielmehr wird angenommen, dass diese Personen während eines Auslandsaufenthalts der Eltern geboren wurden, z.B. während eines Auslandsstudiums oder einer Beschäftigung im Ausland.

Tatsächlich greifen die KMK und das nordrhein-westfälische Schulministerium seit geraumer Zeit zusätzlich zum Geburtsland der Schüler*innen und ihrer Eltern zur Bestimmung des Migrationshintergrundes – u. a. auch auf Anregung von Chlosta/Ostermann 2005 (vgl. oben) – die Sprachverwendung der Schüler*innen zur Bestimmung des Migrations- bzw. Zuwanderungshintergrundes mit auf. Unabhängig von Geburtsland und Staatsangehörigkeit wird so auch ausgehend von der Verwendung einer anderen Sprache als bzw. neben Deutsch innerhalb der Familie auf einen Migrationshintergrund geschlossen: So werden in den Schuldaten NRW Personen als Schüler*innen mit Zuwanderungsgeschichte gezählt, „die (a) selbst nicht in Deutschland geboren sind, (b) mindestens ein nicht in Deutschland geborenes Elternteil haben oder (c) in deren Familie eine nicht deutsche Verkehrssprache gesprochen wird“. Mindestens eines dieser Merkmale muss erfüllt sein, um als Schüler*in mit Zuwanderungsgeschichte zu gelten (MSB NRW, 2019a, S. 7). Die KMK geht – analog zum NRW-Schulministerium, allerdings ohne Berücksichtigung der Eltern – bei Schüler*innen von einem Migrationshintergrund aus, „wenn mindestens eines der folgenden Merkmale zutrifft: 1. Keine deutsche Staatsangehörigkeit, 2. Nichtdeutsches Geburtsland, 3. Nichtdeutsche Verkehrssprache in der Familie bzw. im häuslichen Umfeld (auch wenn der Schüler/die Schülerin die deutsche Sprache beherrscht)“ (KMK, 2015, S. 32).

In den PISA-Studien (2018) wird neben den Fragen zur Bestimmung des Zuwanderungshintergrund separat die *zu Hause gesprochene Sprache* von den Jugendlichen im Fragebogen über die Frage „Welche Sprache sprichst du am häufigsten zu Hause?“ erfasst (Weis, Müller et al., 2019, S. 136). Die separate Erfassung beider Variablen ermöglicht damit das Sichtbarmachen von Unterschieden in Bezug auf die zu Hause gesprochenen Sprache zwischen Jugendlichen mit und ohne Zuwanderungshintergrund bzw. ermöglicht eine differenziertere Betrachtung der unterschiedlichen Gruppen (Weis, Müller et al., 2019, S. 153).

Insbesondere Chlosta und Ostermann (2005) konnten – wie oben bereits erwähnt – zeigen, dass die unterschiedlichen Definitionen des Migrationshintergrunds bzw. des Zuwanderungshintergrundes nicht zu identischen Gruppen führen. Zwar lässt sich aus der Angabe des Geburtslands in der überwiegenden Zahl der Fälle auch auf eine Verwendung der Sprache des Geburtslandes in der Familie oder mit Freunden schließen, es bleiben aber Unsicherheiten. Fragt man beide Kriterien ab, lässt sich eine Gruppe identifizieren, deren Mitglieder sich entweder sehr schnell sprachlich assimiliert haben oder die beispielsweise auch in der dritten Generation am Sprachgebrauch ihrer Großeltern festhalten, d. h. die Sprache ihrer Familie erhalten (zur Terminologie und zu Faktoren des Spracherhalts vgl. etwa Ammon, 1991, 105 ff.; Olfert, 2019).

Da im Rahmen dieser Arbeit nun untersucht werden soll, ob eine mehrsprachige Sprachverwendung, d. h. Bilingualismus oder mindestens die Verwendung einer anderen Sprache als Deutsch im persönlichen Kontext einen Einfluss auf die Sprachkompetenz in Deutsch ausweislich der C-Test-Ergebnisse hat, wurden zur differenzierteren Erfassung der tatsächlich erworbenen und verwendeten Sprache(n) ergänzend zu den migrationsbezogenen Angaben (Geburtsland der Schüler*innen und Eltern) sprachbiografische Hintergrunddaten der Schüler*innen abgefragt. Die Sprachverwendung ist dabei die entscheidendere Größe, das Geburtsland nur eine Art Kontrollgröße bzw. ein Instrument zur Messung des Grades des Spracherhalts bzw. einer möglichen sprachlichen Assimilation.

Zur Erfassung dieser sprachbiografischen Daten wurde zunächst die Frage *Was ist Ihre Muttersprache?* (5) gestellt. Prinzipiell ist die Verwendung des Begriffs *Muttersprache* aus mehreren Gründen problematisch: Der Begriff *Muttersprache* kann zum einen „zur

Bezeichnung der Sprache verwendet [werden], die jemand als erste Sprache in der Familie erworben hat und die er ‚am besten' beherrscht" (Fürstenau, 2011, S. 31). Zum anderen kann der Begriff *Muttersprache* darauf verweisen, „dass die Sprache eine große emotionale Bedeutung für die Sprecherin oder den Sprecher hat, unabhängig vom Grad der Beherrschung der Sprache" (Fürstenau, 2011, S. 31). Aus emanzipatorischer Sicht stellt seine Verwendung ein Problem dar, da der Begriff *Muttersprache* suggeriert, dass die Erstsprache vom Kind (ausschließlich) über die Mutter erworben wird, was das Bild der Spracherwerbssituation, an der schließlich auch Väter, oder Geschwister oder andere Familienmitglieder beteiligt sind, verzerrt (Ahrenholz, 2010, S. 3). Durch die Verwendung des Begriffs *Muttersprache* kann zudem auf den „Topos der ‚Abstammung'" angespielt werden, also auf die „Vorstellung, dass die Sprache der ‚Mutter' auch ‚natürlicher Weise' die eigene sei" (Dirim & Mecheril, 2010, 99 ff.; Fürstenau, 2011, S. 31). Darüber hinaus ist der Begriff *Muttersprache* auch politisch-historisch desavouiert: So wurde der Begriff „in Abgrenzung zu staatlich verordnetem Sprachgebrauch ins Feld geführt" (Ahrenholz, 2010, S. 3), zwischen 1933 und 1945 im Zuge der versuchten Germanisierungsanstrengungen (etwa im Sudetenland, in Luxemburg oder Elsass-Lothringen, Malmedy usw.) propagandistisch missbraucht (Scholten, 2000).

Aus diesen Gründen wird im Rahmen der hier vorliegenden Arbeit der Begriff *Muttersprache* nicht verwendet. Vielmehr wird – der Definition von Ahrenholz (2010, S. 3) folgend – die Sprache, die – meist im familiären Kontext – von Geburt an gelernt wird, als *Erstsprache* bezeichnet (vgl. unten zur Kritik dieses Terminus).

Im Fragebogen allerdings wird der Begriff *Muttersprache* indes verwendet. Diese Abweichung wird mit der Annahme begründet, dass das Wort *Muttersprache* den Schüler*innen in der gymnasialen Oberstufe aus dem alltäglichen Sprachgebrauch geläufig ist und am ehesten das Konzept der Erstsprache im wissenschaftlichen Sinne repräsentiert (Petersen, 2014c, S. 19). Um dieses Verständnis zu gewährleisten, wurde die Frage nach der *Muttersprache* im Fragebogen um eine Fußnote ergänzt, in der erklärt wird, dass mit *Muttersprache* die Sprache gemeint ist, die „als erste Sprache in der Kommunikation mit Ihren Bezugspersonen (also Ihren Eltern, Ihren Geschwistern usw.) verwendet" wurde (s. o., Abb. 13). In dieser Erläuterung zeigen sich aber auch die

problematischen Aspekte des Begriffs *Erstsprache*. Die Formulierung „erste Sprache“ (bzw. *Erstsprache*) suggeriert nämlich eine Reihenfolge des Spracherwerbs von mehreren Sprachen, also einen dann auch *sukzessiven Zweitspracherwerb*[73]. Das für Schüler*innen dann problematisch sein, wenn sie von Anfang an zwei (oder mehrere Sprachen) gleichzeitig erwerben (Fürstenau, 2011, S. 31; Rothweiler & Ruberg, 2011, 2011). Der Fragebogen erzwang an dieser Stelle aber keine Wahl oder Reihenfolge (etwa durch ein Rollmenu oder eine Auswahl durch Ankreuzen o. ä.) und ließ die Möglichkeit zu, auf die Frage nach der *Muttersprache* mit mehr als nur einer Angabe zu antworten[74] – so gibt es beispielsweise Fälle, in denen „Türkisch/Deutsch“ als Antwort auf die Frage nach der *Muttersprache* gegeben wurde (vgl. dazu Kapitel 4.4 zur Auswertung dieser Daten).

Neben der Erfassung der Erstsprache(n) sollte über den Fragebogen auch die alltägliche Sprachverwendung im familiären und sozialen Kontext erhoben werden, da über die Angabe der Erstsprache allein nicht ohne Weiteres auf den tatsächlichen Sprachgebrauch geschlossen werden kann: So ist es beispielsweise möglich, dass eine Person Türkisch als Erstsprache erworben hat, innerhalb der Familie und des Freundeskreises aber ausschließlich Deutsch spricht. Ebenso kann Deutsch als Erstsprache angegeben werden. Die Befragte gibt aber parallel an, mit einem Elternteil nur Arabisch zu sprechen. Die Person wird also „trotz“ Erstsprache Deutsch als mehrsprachig erfasst. Wenn Mehrsprachigkeit also – so wie in dieser Arbeit – im Sinne einer „lebensweltlichen Mehrsprachigkeit“ nach Gogolin (2004, 55 ff.) verstanden wird, dann ist diese (lebensweltliche) Mehrsprachigkeit der Schüler*innen nicht nur durch eine andere Sprache als/neben Deutsch als Erstsprache bedingt, sondern durch die Bedeutung des alltäglichen Gebrauchs dieser Sprache(n) im familiären und sozialen Umfeld (Chlosta & Ostermann, 2005, 55 ff.). Um diesem Verständnis von Mehrsprachigkeit für eine differenzierte Betrachtung der Gruppen gerecht zu werden, wurden bei der Konstruktion des Fragebogens Items zur *Familiensprache* (*Welche Sprache sprechen Sie überwiegend mit a) Ihrer Mutter, b) Ihrem Vater, c) Ihren*

[73] Der Prozess des sukzessiven Zweitspracherwerbs zeichnet sich dadurch aus, dass der Erwerb der Zweitsprache erst nach dem Erwerb der Erstsprache (in Grundzügen) stattfindet (vgl. Rothweiler 2007, S. 106).

[74] Gogolin (1987) spricht in diesem Zusammenhang von dem Konzept „Muttersprache: Zweisprachigkeit“.

Geschwistern – auch Deutsch kann hier genannt werden? (6)) und zur Sprachverwendung im Freundeskreis (*Welche Sprache(n) verwenden Sie im Kontakt mit Ihren Freunden* (7)) aufgenommen (dieses Verfahren schlagen auch (Chlosta & Ostermann, 2005, 63 ff.) vor. Über diese Angaben können Daten darüber gewonnen werden, welche Sprache(n) überwiegend mit welchem Elternteil und den Geschwistern gesprochen werden. Im Rahmen dieser Arbeit wird folglich auf das Merkmal *mehrsprachig* geschlossen, wenn als Erstsprache und/oder als Familiensprache eine andere Sprache als/neben Deutsch gesprochen wird.

Beide Vorgehensweisen nebeneinander zu nutzen, d. h. auf einen *Migrationshintergrund* ausgehend vom Geburtsort der Eltern und der Schüler*innen zu schließen und sprachbezogene Variablen separat zu erfassen, bietet zusätzliche analytische Möglichkeiten: Wie viele Personen haben beispielsweise einen Migrationshintergrund, sind aber sprachlich vollkommen assimiliert, d. h. verwenden neben Deutsch keine andere Sprache im familiären Kontext? Die Eltern wie vieler Personen sind zwar wie sie selbst in Deutschland geboren, agieren aber lebensweltlich überwiegend mehrsprachig? Auf diese Weise lassen sich möglicherweise die Testleistungen der verschiedenen Gruppen differenzieren und Aussagen darüber formulieren, ob und wie sich der Gebrauch des Deutschen in den Familien mit Migrationshintergrund aber auch die Sprachverwendung des Deutschen im Freundeskreis, wenn diese nicht (nur) die Erstsprache ist, auf die sprachlichen Leistungen im C-Test auswirkt.

Mit der Frage *Wurden Sie in Deutschland eingeschult?* (8) wird ein Zusammenhang zwischen Dauer der schulischen Instruktion und sprachlicher Kompetenz insinuiert. Der Sprach- bzw. Schriftspracherwerb oder allgemein der Ausbau der Sprachkompetenz in der deutschen Sprache wird durch den Schulbesuch generell gefördert. Unterstellt wird darüber hinaus, dass je länger die Schule besucht wird/wurde bzw. je früher die Instruktion an einer deutschen Schule einsetzte, desto besser ausgebaute schriftsprachliche Kompetenzen im Deutschen haben die Schüler*innen erworben.

Um im Rahmen dieser Arbeit überprüfen zu können, ob neben der Sprache bzw. der Sprachverwendung auch der Bildungsgrad der Eltern einen Einfluss auf die mit dem C-Test gemessenen schriftsprachlichen Kompetenzen hat – d. h., ob beispielsweise

Schüler*innen, deren Eltern über einen niedrigen Bildungsgrad verfügen, schlechtere Test-Ergebnisse erzielen als Schüler*innen, deren Eltern beispielsweise ein Studium absolviert haben –, wird als Indikator für das Bildungsniveau der Eltern der höchste Bildungsabschluss der Mutter und des Vaters (*Was ist der höchste Bildungsgrad Ihrer Eltern (Mutter/Vater)?* (9)) erfragt.

4.2 Stichprobenkonstruktion und Versuchsplanung

In diesem Abschnitt sollen zunächst die Vorarbeiten beschrieben werden, die den Datenerhebungen an den Schulen vorausgingen. Ausgehend vom Ziel der hier vorliegenden Arbeit, mit dem C-Test schriftsprachliche Kompetenzen in der gymnasialen Oberstufe messen und die sich anschließende Frage nach gruppenspezifischen Unterschieden auf einer befriedigenden Datengrundlage aufklären zu wollen, bemühte sich die Verfasserin dieser Arbeit in Kooperation mit dem Zentrum für Lehrerbildung der Universität Duisburg-Essen, in dessen Ressort Diagnostik diese Arbeit entstand, um eine möglichst große Stichprobe von Oberstufenschüler*innen. Die Stichprobengröße und ihre zumindest angezielte Repräsentativität für (aus pragmatischen und organisatorischen Gründen) das Bundesland Nordrhein-Westfalen schafft die Grundlage für die im Folgenden näher beschriebene Auswertung der C-Test-Ergebnisse und ihrer Interpretation, die sich im Wesentlichen auf Referenznormwerte stützt, deren Festlegung einer angemessenen Stichprobengröße bedarf (vgl. detailliert u.).

Für die hier vorliegende Dissertation war der ursprüngliche Plan der Verfasserin dieser Arbeit, Daten für alle drei Stufen der gymnasialen Oberstufe[75] sowohl an Gymnasien als auch an Gesamtschulen in NRW zu erheben. Die gymnasiale Oberstufe bzw. die Sekundarstufe II umfasst in NRW an allen Gymnasien und Gesamtschulen insgesamt 3 Schuljahre, beginnend mit einer einjährigen Einführungsphase (EF), gefolgt von einer zweijährigen Qualifikationsphase (Q1 und Q2). Durch die Einführung der Verkürzung der

[75] Zugunsten der Machbarkeit dieser Studie sowie einer besseren Vergleichbarkeit der Daten, bzw. Teilgruppen (in Bezug auf die zu untersuchenden Hypothesen) wurden im Rahmen dieser Arbeit nur Daten in der gymnasialen Oberstufe an Gymnasien und Gesamtschulen erhoben, nicht aber an anderen Schulformen wie beispielsweise Abendgymnasien oder Berufskollegs.

Gymnasialschulzeit von 13 auf 12 Jahre (G8) im Jahr 2013 in NRW besuchen die Gymnasialschüler*innen die Einführungsphase in der Klasse 10, Schüler*innen an Gesamtschulen – an denen weiterhin 13 Jahre bis zum Abitur unterrichtet wird (G9) – sind in der Klassenstufe 11[76] in der Einführungsphase.

Der Zugang zu Schulen für Forschungszwecke, insbesondere im Zusammenhang mit Leistungs- oder Kompetenzmessungen, ist allerdings oftmals mit Akzeptanzproblemen seitens der Schulleitungen und/oder den involvierten Lehrkräften und einem hohen organisatorischen Aufwand verbunden, weshalb die Konstruktion der hier vorgestellten Stichprobe einigen Beschränkungen unterlag. So erwies es sich bei der Akquise von Kooperationsschulen als schwierig, dass fast alle teilnehmenden Schulen zwar mit einer Datenerhebung in der Einführungs- und Qualifikationsphase 1 einverstanden waren, einer Versuchsdurchführung in der Qualifikationsphase 2 mit Blick auf das den Schüler*innen dieser Stufe bevorstehende Abitur nicht zustimmten. Ebenso erklärten sich insgesamt mehr Gymnasien bereit an dem Untersuchungsvorhaben teilzunehmen als Gesamtschulen[77]. Die Motivation einzelner Schulen an der Studie teilzunehmen, konnte mit den individualisierten Ergebnis-Rückmeldungen bzw. deren Interpretation und Einordnung und dem Zugriff auf bereitgestellte Übungsplattform (vgl. unten) insgesamt gesteigert werden. Bedingt allerdings durch die Tatsache, dass die Auswahl der teilnehmenden Proband*innen von der Kooperationsbereitschaft der einzelnen Schulen bzw. der Schulleitung und den Lehrkräften abhängig war, ist die in dieser Studie beschriebene Stichprobe (vgl. Kapitel 5.1) am ehesten als Gelegenheitsstichprobe[78] zu verstehen (Döring & Bortz, 2016, 306 ff.).

[76] Es gibt einzelne Gymnasien, an denen nicht auf G8 umgestellt wurde, d. h. an diesen Gymnasien machen die Schüler*innen weiterhin erst nach 13 Jahren Abitur (G9). In der hier vorliegenden Stichprobe betrifft dies ein Gymnasium.

[77] Eine genaue Betrachtung des Erhebungsrahmens und der Verteilung der Stichprobe erfolgt im Kapitel 5.1.

[78] Die Gelegenheitsstichprobe zeichnet sich dadurch aus, dass Personen oder Objekte willkürlich für die Stichprobe einer Studie ausgesucht werden, die gerade zur Verfügung stehen oder leicht zugänglich sind (Döring, Bortz 2016, S. 306). Der Nachteil von Gelegenheitsstichproben besteht durch die unkontrollierte Aufnahme von Proband*innen in die Stichprobe darin, dass nicht generalisierbar auf die Grundgesamtheit übertragbar, d. h. wenig repräsentativ sind (ebd.).

Zur Vermeidung noch weitergehender Einschränkungen der Repräsentativität wurde bei der Konstruktion der Stichprobe ein möglichst breites Spektrum an unterschiedlichen Schulen hinsichtlich der Zusammensetzung der Schülerschaft angezielt. Kennwerte, die der unterschiedlichen Zusammensetzung der Schülerschaft NRW Rechnung tragen, sind die sogenannten Schulstandorttypen[79] (MSB NRW, 2011, 300 ff.). Das Standorttypenkonzept NRW zielt darauf ab, die Vergleichbarkeit von Ergebnissen aus Lernstandserhebungen zwischen Schulen zu ermöglichen. Die Zuordnung von Standorttypen wird für alle weiterführenden Schulen in NRW vom Schulministerium auf der Grundlage von Daten der amtlichen Statistik vorgenommen (ebd., S. 300). Zur Bestimmung der Standorttypen werden dabei der Anteil von Schüler*innen mit Migrationshintergrund[80] sowie das durchschnittliche Einkommen der Einwohner*innen, der Anteil der Personen mit Bezug von SGB-II-Leistungen, der Anteil der Personen, die in Haushalten in Ein-bis-Zwei-Familienhäusern leben sowie der Anteil an Personen mit Migrationshintergrund in den Wohngebieten der elterlichen Wohnungen der Schüler*innen herangezogen (QUA-LIS NRW, 2017, S. 1). Jede Schule wird so einem von insgesamt 5 Standorttypen zugeordnet, wobei an Schulen, die beispielsweise dem Standorttyp 1 zugewiesen wurden, der Anteil an Schüler*innen mit Migrationshintergrund im Durchschnitt bei 11% liegt. Ebenso gibt die Mehrzahl der Schüler*innen dieses Standorttyps an Schüler*innen an,

> „dass die elterlichen Wohnungen in Wohngebieten liegen, deren Einwohnerinnen und Einwohner tendenziell ein überdurchschnittliches Einkommen aufweisen (bezogen auf den Landesdurchschnitt), in denen der Anteil der Personen mit Bezug von SGB-II-Leistungen durchschnittlich bei ca. 5% liegt, in denen durchschnittlich 60% der Haushalte in Ein- bis Zwei-Familienhäusern leben und welche einen stark unterdurchschnittlichen Anteil an Personen mit Migrationshintergrund aufweisen (bezogen auf den Landesdurchschnitt)“ (QUA-LIS NRW, 2017, S. 1).

[79] Die *Standorttypen* wurden zu Beginn des Schuljahres 21/22 von den neuen *Sozialindexen* des MSB abgelöst, d. h. diese lagen bei der Datenerhebung derhier vorgestellten Arbeit noch nicht vor und die Schulen waren folglich ausschließlich Standorttypen zugeordnet.

[80] Die Daten zum Migrationshintergrund werden laut Schulentwicklung NRW dabei den Daten der Schulstatistik entnommen (vgl. QUA-LiS NRW), in denen unabhängig von der Staatsangehörigkeit bei einem Zuzug nach Deutschland, einer nicht deutschen Verkehrssprache oder mindestens nicht in Deutschland geborenen Elternteil von einem Migrationshintergrund ausgegangen wird (Schulstatistik NRW, 2019, S. 172 ff.).

Im Vergleich dazu liegt der Anteil von Schüler*innen mit Migrationshintergrund an Schulen des Standorttyps 5 im Durchschnitt bei 61 % (QUA-LIS NRW, 2017, S. 2), ebenso gilt für die Schüler*innen, dass

> „die elterliche Wohnung in Wohngebieten liegen, deren Einwohnerinnen und Einwohner ein stark unterdurchschnittliches Einkommen aufweisen (bezogen auf den Landesdurchschnitt), in denen der Anteil der Personen mit Bezug von SGB-II-Leistungen im Durchschnitt bei 20% liegt und die Hälfte der Wohngebiete eine SGB-II-Quote zwischen 15% und 25% aufweisen, in denen im Durchschnitt nur ein Fünftel der Haushalte in Ein- bis Zwei-Familienhäusern leben und welche einen stark überdurchschnittlichen Anteil an Personen mit Migrationshintergrund aufweisen (bezogen auf den Landesdurchschnitt)“ (QUA-LIS NRW, 2017, S. 2).

Für die Konstruktion der Stichprobe dieser Untersuchung konnten für alle 5 Standorttypen Schulen für die Teilnahme an der Studie gewonnen werden. Tendenziell war die Bereitschaft der Schulen des Standorttyps 5 höher, an der Studie teilzunehmen als die der Schulen des Standorttyps 1 (zu den Details vgl. Kapitel 5.1 zur Stichprobenbeschreibung).

Zusammenfassend lässt sich festhalten, dass weder die regionale Auswahl noch die Anzahl der in der Stichprobe vertretenen Schulformen noch die Verteilung auf die Standorttypen Anspruch auf Repräsentativität erheben kann. Allenfalls für die Schulform des Gymnasiums nähert sich die Stichprobe diesem Ideal an – es bleibt aber zu konstatieren: Im statistischen Sinne handelt es sich um eine Gelegenheitsstichprobe, die allenfalls durch den relativ großen Umfang eine solide Basis für eine quantitative Auswertungen der Daten bietet (Bortz & Döring 2006, S. 394 ff.; vgl. dazu Kapitel 5.1).

Nach der Konstruktion der Stichprobe und der Erstellung des finalen Testheftes durch die Pre-Testung (vgl. Kapitel 4.1.2.1) wurde damit begonnen, den entwickelten C-Test als Online-Version zu realisieren. Eine Online-Testung bietet den Vorteil einer automatisierten Auswertung, wodurch also die äußerst zeitaufwändige und zudem fehleranfällige manuelle Korrektur der Testhefte wegfällt. Die Auswertungsobjektivität dieser Form der Verarbeitung ist also höher: Selbst wenn C-Tests manuell mithilfe von Lösungsschablonen, d. h. festgelegten Original- und Alternativlösungen korrigiert werden, kann beispielsweise eine unleserliche Handschrift einer Proband*in die Testauswerter*in bei der Entscheidung, ob eine Lücke falsch oder korrekt ausgefüllt

worden ist, unsicher werden lassen. Einen solchen Interpretationsspielraum gibt es bei der automatisierten Online-Auswertung nicht, eine Wortergänzung wird entweder als eindeutig richtig oder als eindeutig falsch erkannt, sofern die Lösungen korrekt im System hinterlegt wurden.

Zur Realisierung der Online-Versionen der C-Tests standen zwei Systeme bzw. Plattformen zur Verfügung, nämlich *JACK* und *Moodle*. *JACK* ist ein von Prof. Dr. Goedicke an der Universität Duisburg-Essen entwickeltes, server-basiertes System für die Durchführung computergestützter Aufgaben mit automatischer Bewertung, über das auch der Aufgabentyp *Lückentexte* – wozu auch der C-Test als Aufgabenformat gehört – generiert werden kann[81]. Ebenso ist es auf der E-Learning-Plattform *Moodle* möglich, in virtuellen Kursräumen Befragungen, bzw. den oben gezeigten Fragebogen zu den Hintergrundvariablen der Schüler*innen als auch Lückentexte, bzw. C-Tests online einzustellen. Da im Zuge der Digitalisierungsoffensive an vielen Schulen auch Moodle eingeführt wurde und die Schüler*innen also zumindest zum Teil mit den Anwendungsroutinen (Zugang, Kursraumlogik usw.) dieser Plattform bereits vertraut waren, entschied sich die Autorin der Untersuchung, Moodle-Kursräume für die Durchführung der Befragung und der Tests einzurichten und dazu die Moodle-Plattform der Universität Duisburg-Essen zu nutzen.

Prinzipiell erfolgt die Registrierung für die Moodle-Plattform der Universität Duisburg-Essen über die Uni-Kennung oder einen Gastzugang. Da die zu testenden Schüler*innen nicht an der Universität Duisburg-Essen eingeschrieben sind (die Uni-Kennung wir nach erfolgreicher Einschreibung vergeben) und somit über keine Uni-Kennung verfügen, wurde zunächst eine Anmeldung über Gastzugänge in Erwägung gezogen. Der Registrierungsvorgang für die Moodle-Plattform als Gast, bei dem vorausgesetzt wird, eine gültige E-Mailadresse zu hinterlegen, an die ein Aktivierungslink geschickt wird, erwies sich in Bezug auf die zu untersuchende Zielkohorte jedoch als problematisch, da zum einen nicht einfach angenommen werden konnte, dass tatsächlich alle (auch in der Oberstufe minderjährigen) Schüler*innen über einen eigenen E-Mail-Account verfügen. Zum anderen hätte eine Registrierung über private E-Mailadressen, die beispielsweise

[81] Ausführliche Informationen zu JACK sind online unter *Specification of Software Systems* von Prof. Dr. Goedicke unter https://www.s3.uni-duisburg-essen.de/jack/ zu finden.

Klarnamen enthalten oder/oder Hinweise (Vornamen, Nachnamen, Geburtsdaten, Jahreszahlen, o. ä.) auf die über die E-Mailadresse registrierte Person liefern, gegen die hier anvisierte Test- und Fragebogenbearbeitung in anonymisierter Form und somit gegen datenschutzrechtliche Bestimmungen für die Datenerhebung in dieser Studie verstoßen. Aus diesem Grund wurde in Zusammenarbeit mit dem ZIM der Universität Duisburg-Essen beschlossen, für die Registrierung im Moodle-System pseudonymisierte Accounts über Kennungen (Anmeldedaten) aus Zahlen-Buchstaben-Kombinationen zu erstellen[82]. Über diese vorab angelegten Zugänge konnte gewährleistet werden, dass die an der Studie teilnehmenden Schüler*innen sowohl die Testung als auch den Fragebogen im Online-Kursraum über eine ihnen zufällig vergebene Kennungen für das Projektteam anonymisiert bearbeiten können[83]. Werden die Kennungen nun seitens der Schule über Listen an reale Personen vergeben, können autorisierte Lehrkräfte der jeweiligen Schule über diese interne Zuordnung nachvollziehen, welche Schüler*in unter welcher Kennung gearbeitet hat und die individuell erbrachten Ergebnisse an die Schüler*in zurückmelden und besprechen.

Ebenfalls über die Kennung ist es den an der Testung teilnehmenden Schüler*innen möglich, auf zwei E-Learning-Kurse zur Förderung sowohl rezeptiver als auch produktiver sprachlicher Fähigkeiten zuzugreifen, die im Rahmen des Projekts – in dem diese Arbeit entstanden ist – in Zusammenarbeit mit Deutschlehrer*innen zweier Kooperationsschulen entworfen wurden. Die E-Learning-Kurse beinhalten u. a. aufbereitetes Material (Erklär-Videos, Audiodateien, Lesetexte, Stichwortsammlungen, Übungsklausuren, etc.) zu abiturrelevanten Themen, Aufgaben zu Leseverstehen und Lesestrategiewissen, aber auch Übungen im formalsprachlichen Bereich (Orthografie, Zeichensetzung, Grammatik), die von den Schüler*innen eigenständig bearbeitet werden können[84].

[82] Besonderer Dank gilt an dieser Stelle besonders bei Frau Heinrich und Herrn Gollan vom ZIM der Universität Duisburg-Essen für die gute und erfolgreiche Zusammenarbeit.
[83] Weitere Informationen zum Vorgehen bei der Online-Testung folgen in Kapitel 4.3 zur Beschreibung der Versuchsdurchführung bzw. Datenerhebung.
[84] Vor dem Hintergrund der Fragestellung der hier vorgestellten Dissertation soll an dieser Stelle nicht weiter auf die beschriebenen E-Learning-Kurse eingegangen werden.

Vor jeder Datenerhebung wurden die Schulen über den Ablauf, die Zielsetzung, die technische Umsetzung und die Datenschutzrichtlinien der Untersuchung informiert.

4.3 Datenerhebung

Die Situation der Datenerhebung ist als Feldstudie zu verstehen, da sie in einer natürlichen Umgebung der teilnehmenden Proband*innen, d. h. in der Schule bzw. im Klassen- oder Kursverbund in natürlichen Gruppen stattfand. Ein genereller Nachteil von Feldstudien ist, dass mögliche Störvariablen nicht immer ausgeschaltet oder kontrolliert werden können, was die interne Validität möglicherweise beeinträchtigen und die Interpretation von kausalen Zusammenhängen erschweren kann (Döring & Bortz, 2016). Es ist davon auszugehen, dass besonders im Schulalltag unterschiedliche Störvariablen einen Einfluss auf das Verhalten der getesteten Schüler*innen haben können. So kann sich beispielsweise eine ablehnende Haltung gegenüber dem Testinstrument seitens der Lehrkraft negativ auf die Motivation der Schüler*innen auswirken, ebenso wie eine späte Uhrzeit bzw. Schulstunde, in der die Daten erhoben werden. Andererseits gehören solche Störvariablen zum natürlichen Umfeld *Schule*, wodurch die Untersuchungsbedingungen den Alltagsbedingungen ähneln und die Datenerhebung damit eine hohe externe Validität aufweist. Durch eine weitestgehende Standardisierung des Testablaufs wurde der Einfluss der Störvariablen versucht möglichst gering zu halten.

Der Zeitraum für die gesamte Datenerhebung (ausgenommen der Pre-Testung) erstreckte sich auf knapp drei Jahre (von November 2015 bis Juli 2018). Die Datenerhebung wurde von einer Versuchsleiterin, der Verfasserin dieser Arbeit, durchgeführt, die je nach Gruppengröße der zu testenden Personen bei der Testdurchführung von ein oder zwei wissenschaftlichen bzw. studentischen Hilfskräften unterstützt wurde, die ihr bei der Aufsicht und bei der Lösung technischer Probleme zur Seite standen. So traten etwa vereinzelt Schwierigkeiten bei der Registrierung im Universitätssystem oder bei der Anmeldung in Moodle auf, die dann vor Beginn der Befragungen/Tests schnell gelöst werden mussten, um die zur Verfügung stehende

Schulstunde nicht zu überziehen[85]. Die einzelnen Schulen legten vorab fest, welche Kurse wann und in welchen Räumlichkeiten getestet werden sollten: Für die Online-Testung, d. h. für die Bearbeitung des Fragebogens und des C-Tests[86], musste sichergestellt sein, dass allen an der Testung teilnehmenden Schüler*innen ein eigener Arbeitsplatz mit Zugang zum Internet zur Verfügung stand, was je nach digitaler Ausstattung der Schulen im Computerraum oder im Klassen-/Kursraum über Laptops oder Tablets gewährleistet werden konnte. Für eine Schule war eine Online-Testung aus organisatorischen und/oder technischen Gründen nicht umsetzbar, weswegen die Bearbeitung des Fragebogens und des C-Tests in diesem Fall für eine Klasse (N=22) im klassischen Papier-Beilstift-Format durchgeführt wurde.

Die Testungen wurden den Schüler*innen vorab von den Lehrkräften angekündigt[87], sodass die meisten Schüler*innen auf die Testsituation vorbereitet waren. Vor Beginn der Testdurchführung[88] stellte die Versuchsleiterin sich selbst sowie ihr Team vor und erläuterte noch einmal den Grund des Erscheinens. In diesem Zusammenhang wurde auch der weitere Testablauf erklärt sowie auf die mögliche Nutzung der oben beschriebenen E-Learning-Kurse über die zuvor vergebenen Kennungen hingewiesen. Danach wurden die Testhefte des LGVT 6-12 ausgeteilt, auf deren Deckblatt die Schüler*innen zunächst ihre Kennung eintragen sollten, bevor sie von der Testleiterin mündlich[89] als auch im Testheft auf der ersten Seite in schriftlicher Form die Testinstruktion erhielten. Wenn es nach der Instruktion und des darin enthaltenen Beispiels keine Fragen seitens der Schüler*innen gab, begann die Versuchsdurchführung. Nach Ablauf der 4 Minuten Bearbeitungszeit gab die

[85] In einigen wenigen gesundheitlich oder terminlich bedingten Ausnahmefällen, an denen die Verfasserin dieser Arbeit die Testungen nicht selbst durchführen konnte, wurde sie von erfahrenen und geschulten wissenschaftlichen/studentischen Mitarbeiter*innen vertreten. Ein besonderer Dank gilt in diesem Zusammenhang insbesondere Saniye Karaca, Andreas Klemmer, Tuğba Yalcin und Androniki Tsiurda.

[86] Der LGVT 6-12 wurde in der von den Autoren käuflich zu erwerbenden Papier-Bleistift-Version eingesetzt.

[87] Die Ankündigung durch die Lehrkräfte umfasste i. d. R. neben (schriftlichen) Informationen zur Untersuchung, die Datenschutzerklärung sowie eine Einwilligungserklärung zur Teilnahme an der Studie.

[88] Vor der Vorstellung der Testleitenden und des Projekts wurde die Anwesenheit der Schüler*innen von der jeweiligen Lehrkraft kontrolliert sowie die Einwilligungserklärung zur Datenverarbeitung und Testteilnahme eingesammelt – wenn dies nicht bereits im Vorfeld passiert war. Schüler*innen, die bzw. deren Eltern der Teilnahme an der Untersuchung nicht zugestimmt hatten, wurden in der Regel mit Aufgaben aus anderen Unterrichtskontexten oder Deutsch-Fördermaterialien versorgt.

[89] Zur Vorgabe der Testinstruktion von den Testentwicklern vgl. Schneider et al. 2007, S. 12 f.

Versuchsleiterin das Stopp-Signal, die Schüler*innen wurden gebeten, dass letzte von ihnen gelesene Wort zu markieren und die Testhefte des LGVT 6-12 wurden eingesammelt.

Anschließend wurden die Schüler*innen aufgefordert, sich mit ihrer jeweiligen Kennung[90] im Moodle-System der Universität Duisburg-Essen zu registrieren und sich in den passwortgeschützten C-Test-Kursraum einzuschreiben. Die einzelnen Anmeldeschritte wurden von der Testleiterin anhand einer Power-Point-Präsentation angeleitet und lagen den Schüler*innen zudem als Handout vor. Nach erfolgreicher Registrierung und Anmeldung erschien die Startseite des Test-Kursraumes, der folgende Kursabschnitte beinhaltete:

[90] Bei der Registrierung in Moodle werden die Schüler*innen nach der Eingabe der zuvor erhaltenen Zugangsdaten dazu aufgefordert, für ihre jeweils verwendete Kennung ein eigenes Passwort festzulegen, was für ein erneutes Einloggen über die Kennung und für den Zugriff auf die Übungskurse erforderlich ist.

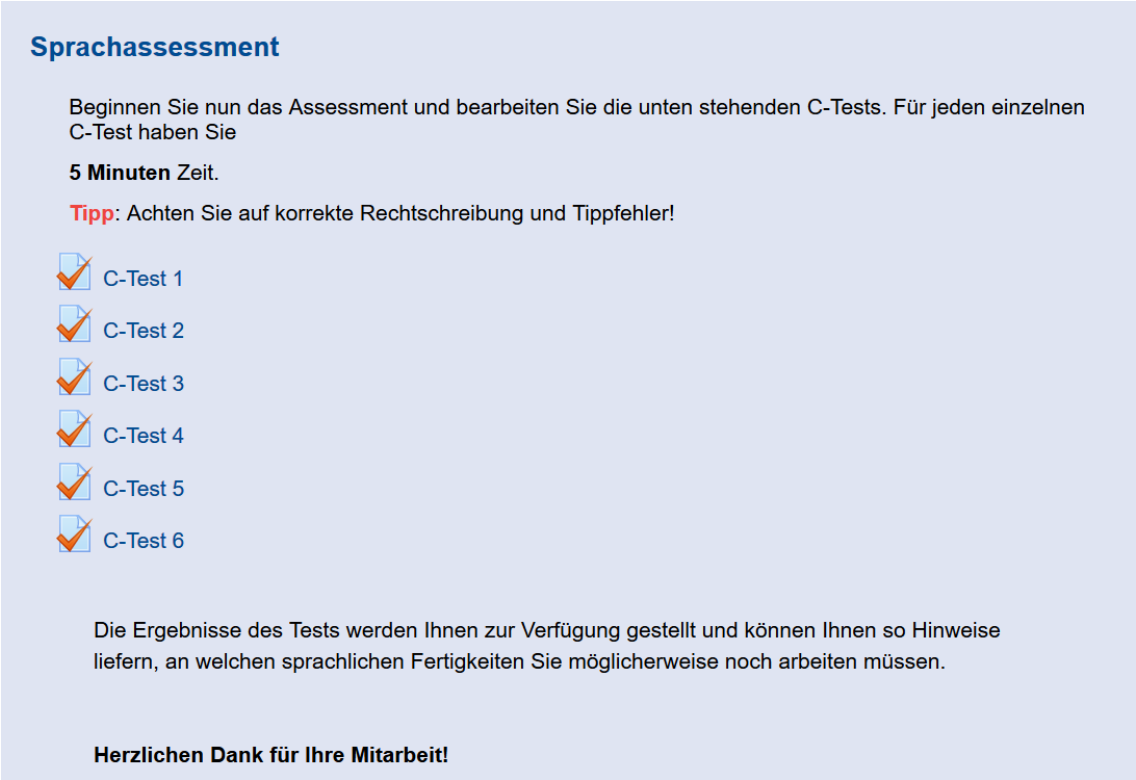


Abbildung 15: Layout des C-Test-Kursraums

Im ersten Kursabschnitt ist ein kurzer Begrüßungstext dargestellt, der noch einmal grob die Ziele der Testung bzw. des Projekts umreißt. Der zweite Kursabschnitt beinhaltet die Bearbeitung des Fragebogens, der von den Schüler*innen eigenständig bearbeitet werden konnte. Sobald alle Schüler*innen eines Testdurchgangs die Beantwortung des Fragebogens beendet hatten (in der Regel dauerte dies etwa 5-10 Minuten), wurde gemeinsam das C-Test-Beispiel im Kursabschnitt drei bearbeitet. Die Testleiterin kündigte an dieser Stelle an, dass es sich bei dem Beispiel-C-Test um eine Demo-Version handelt, deren Bearbeitung natürlich nicht mit in das Endergebnis einfließen würde. Nach Beendigung des Beispiel-C-Tests konnten mögliche Fragen zum Testformat gestellt werden. Zusätzlich wurden die Probanden noch einmal darauf hingewiesen, dass eine fehlerhafte Orthografie beim Ergänzen der Lücken in den C-Tests als Fehler gewertet werden würde und etwaige Tippfehler aus diesem Grund nach Möglichkeit vermieden bzw. korrigiert werden sollten. Zudem wurde angesagt, dass nach Beginn eines jeweiligen C-Tests die Bearbeitungszeit von 5 Minuten ablaufen würde – sichtbar durch einen Timer am linken, oberen Bildschirmrand -, der C-Test vor Ablauf der Zeit

abgegeben werden könne oder nach Ablauf der 5 Minuten automatisch abgegeben werden würde.

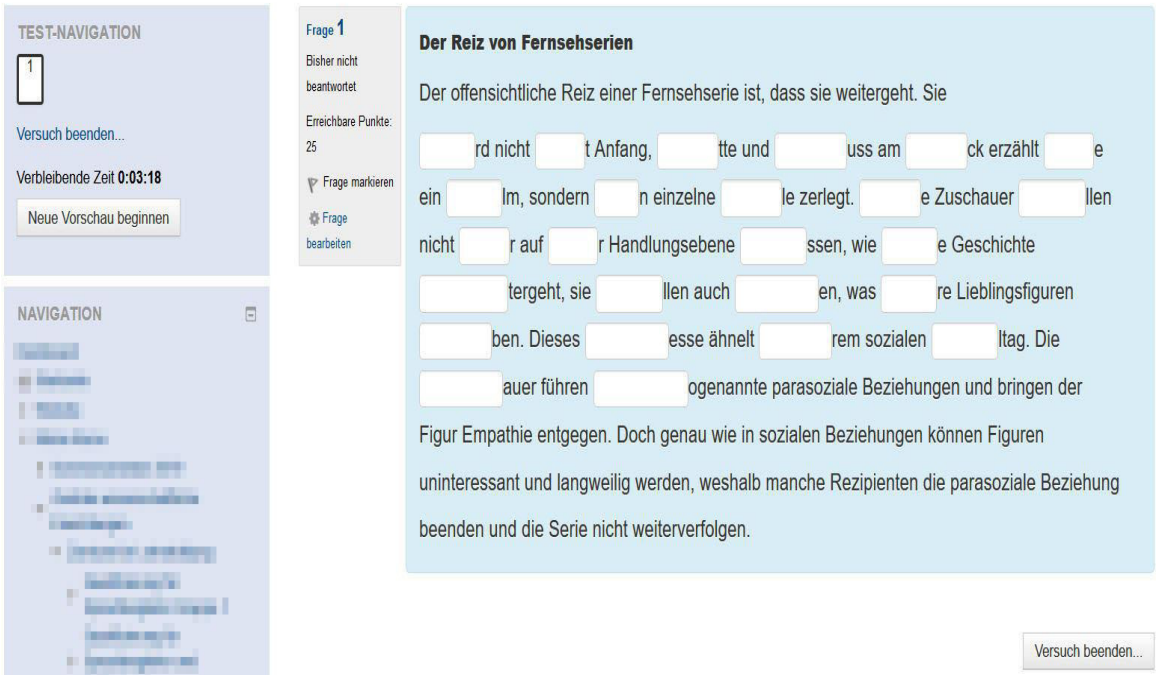


Abbildung 16: Design und Layout während der Bearbeitung der C-Tests

Die Bearbeitung war zudem auf einen Versuch pro C-Test begrenzt. Nach Abschluss der Bearbeitung aller sechs C-Tests konnten die Schüler*innen ihren Gesamtpunktwert sowie die erzielten Teil-C-Testwerte einsehen und wurden abschließend gebeten, sich aus dem Kursraum auszuloggen und von Moodle abzumelden. Der gesamte Testdurchlauf dauerte in der Regel ca. eine Zeitstunde. Die den Kennungen zugehörigen Ergebnisse bzw. deren Zuordnung in Prozentränge und Leistungsgruppen wurden den Schulen tabellarisch etwa 2-3 Wochen nach der Testdurchführung zurückgemeldet, ebenso wie eine graphischer Vergleich der durchschnittlich an der jeweiligen Schule erbrachten Leistungen zu den Durchschnittswerten aller anderen an der Studie teilnehmenden Schulen.

4.4 Auswertungsmethode

4.4.1 Operationalisierung der Auswertung

Für die Auswertung von C-Tests im schulischen Bereich, genauer gesagt für die Klassenstufen 5-7, schlagen Baur et al. (2012; 2013a, 2013b; 2007, 2009) vor, insgesamt jeweils zwei Werte für jede einzelne C-Test-Lücke zu ermitteln: Bei dem ersten Wert, dem sogenannten Richtig/Falsch-Wert (R/F-Wert), erzielt jede morphologisch-syntaktisch, semantisch und orthografisch vollständig korrekte Ergänzung einer Lücke einen Punkt (Baur et al., 2013b, S. 7). Der zweite zu bestimmende Wert ist der sogenannte Worterkennungswert (WE-Wert), welcher sich auf die semantische Rekonstruktion eines Wortes bezieht: Wird das zu ergänzende Wort einer Lücke richtig erkannt, d. h. sinngemäß korrekt ergänzt, führt dies zu einem WE-Punkt, auch wenn die Ergänzung der Lücke orthographisch oder morphologisch falsch realisiert wurde (ebd.). Die Summen aller R/F-Punkte und aller WE-Punkte ergeben dann den R/F- und WE-Wert einer Testperson für den gesamten C-Test (Baur et al., 2012; Baur et al., 2013b).

Für die Interpretation der Testergebnisse ist neben dem R/F- und dem WE-Wert nun auch die Differenz zwischen den beiden Werten interessant, da diese „Aussagen über das Verhältnis zwischen den produktiven und rezeptiven sprachlichen Fertigkeiten“ der Schüler*innen zulässt (Baur et al., 2013b, S. 9). Erzielt eine Person beispielsweise einen vergleichsweise hohen WE-Wert aber einen eher niedrigen R/F-Wert, was zu einer großen Differenz zwischen den beiden Werten führt, kann davon ausgegangen werden, dass die rezeptiven Fertigkeiten, wie Textverständnis und Wortschatzkenntnisse der Person bereits relativ gut ausgebaut sind, formalsprachliche Aspekte, beispielsweise im orthografischen und grammatikalischen Bereich, jedoch noch Schwierigkeiten bereiten (ebd.). Ein geringer Differenzwert bei insgesamt niedrigen R/F- und WE-Werten deutet wiederum darauf hin, dass die Testperson Schwierigkeiten sowohl im Bereich des Lese-

und Textverstehens als auch im formalsprachlichen Bereich hat[91] (Baur et al., 2013b, S. 10).

Diese Auswertungsmethode, d. h. die Erfassung der R/F- und WE-Werte, wurde zunächst zu Beginn der Datenerhebung mit den linksseitig getilgten C-Tests in der gymnasialen Oberstufe übernommen. Die Analysen der ersten Ergebnisse zeigten jedoch, dass es im Bereich der Sekundarstufe II kaum bzw. nur sehr geringe Differenzwerte zwischen den R/F- und den WE-Werten der Oberstufenschüler*innen gab. Auf der Grundlage dieses Befundes kann also angenommen werden, dass Schüler*innen in der Sekundarstufe II – anders als vielleicht Schüler*innen der Sekundarstufe I, aus deren Beobachtung Baur et al. ihr methodisches Vorgehen abgeleitet haben (Baur et al., 2006; Baur et al., 2012; Baur et al., 2013b) – die Lücken bzw. Wörter, die semantisch richtig erkannt werden, zum Großteil auch orthographisch[92] korrekt ergänzen können. Da über die Erfassung der WE-Werte bei der im Pretest untersuchten Kohorte keine differenzierenden diagnostischen Informationen über das Verhältnis der rezeptiven und produktiven Fertigkeiten gewonnen werden konnten, wurde für die Auswertung der in dieser Untersuchung eingesetzten C-Tests entschieden, lediglich den R/F-Wert zu erfassen.

Die Auswertung erfolgt dabei nach einem dichotomen Auswertungsmodus: Jede morphologisch-syntaktisch, semantisch und orthographisch korrekt ergänzte Lücke[93] im C-Tests wird mit einem, jede nicht oder falsch ausgefüllte Lücke mit null Punkten bewertet. Nicht bearbeitete Items, sogenannte *Missings*, werden als solche zwar im System registriert, im Auswertungsmodell jedoch der gängigen C-Test-Praxis folgend wie Falschantworten mit null Punkten bewertet (Mashkovskaya & Baur, 2016b, S. 217).

[91] Ausführliche Informationen zur Ergebnisinterpretation der R/F- und WE-Werte sowie deren Differenz finden sich bei Baur, Goggin 2007 a; 2007 b, 2008; Baur, Chlosta, Goggin 2012; Baur, Goggin, Wrede-Jackes 2013.

[92] Morphologische Strukturen spielen beim Lösen von linksseitig getilgten C-Tests eine eher untergeordnetere Rolle als beim Lösen von rechtsseitig getilgten C-Tests (vgl. Kapitel 2.3.5).

[93] Welche Alternativlösungen ebenfalls als korrekte Antwort bewertet werden sollten, wurde zuvor durch die Pre-Testung ermittelt und festgelegt (vgl. Kapitel 4.1.2.1).

Für jede Testperson werden die Itemwerte jeder einzelnen C-Test-Lücke sowie der Summenwert jedes Teil-C-Tests und der Gesamtscore für das komplette C-Test-Set (als Rohpunktewert) in der Statistik- und Analyse-Software SPSS erfasst und mit den jeweiligen Fragebogendaten über die anonymisierten Kennungen zusammengeführt. Auf der Grundlage der Itemwerte und Gesamtscores werden dann zunächst die Messqualität bzw. die Gütekriterien (Reliabilität, Itemschwierigkeit, Trennschärfe etc.) der C-Tests über eine Itemanalyse bestimmt (vgl. Kapitel 5) (Kelava & Moosbrugger, 2012, 75 ff.).

4.4.2 Klassische Testtheorie vs. Item-Response-Theory

Die Auswertung der Ergebnisse von C-Tests kann prinzipiell sowohl mit Modellen der *Klassischen* als auch der *Probabilistischen Testtheorie* (*Item-Response-Theory*) erfolgen, wobei sich die theoretischen Annahmen und damit auch die Modelle zur Berechnung unterscheiden: Die Grundannahmen der Klassischen Testtheorie bestehen im Allgemeinen darin, dass sich jeder Messwert X einer Person immer aus einem wahren Wert (*true score*) und einem Messfehler (*error*) zusammensetzt (Fisseni, 2004). Ein C-Test-Ergebnis würde so beispielsweise aus der wahren Sprachkompetenz (das zu messende Merkmal) und aus einem unsystematischen Fehlerwert, der durch unterschiedliche, unkontrollierbare Störungen (z. B. Schwankungen der Motivation oder Konzentration) zustande kommt, bestehen (Moosbrugger, 2012b, 104 ff.). Da die wahre Merkmalsausprägung laut Klassischer Testtheorie konstant ist, kann der zufällig entstehende, variierende Messfehler durch unbegrenzt viele Wiederholungen des Tests ausgeglichen bzw. herausgemittelt werden (Krampen, 2018). Würde bei einer Proband*in also unendlich oft die Sprachkompetenz mit einem C-Test gemessen werden, dann würde sich der Mittelwert aller Testungen irgendwann dem wahren Wert, bzw. der wahren Ausprägung des Merkmals *Sprachkompetenz* so weit nähern, bis er ihm quasi entspräche (ebd.). Da eine unbegrenzte Testwiederholung in der Praxis aber unmöglich ist, wird jedes einzelne Item innerhalb eines Tests, d. h. jede C-Test-Lücke, als neue Aufgabe für das zu messende Merkmal und somit als Messwiederholung angesehen. Der Testwert einer Proband*in, der somit als Schätzwert oder

Annäherungswert der wahren Merkmalsausprägung verstanden werden kann, setzt sich also nach Annahmen der Klassischen Testtheorie aus der Summe der Antworten auf die Items zusammen (Krampen, 2018). Für den C-Test bedeutet das, dass die Anzahl der richtig gelösten Items das Ergebnis für einen Teiltest und die Summe aller Teiltests den Summenscore (oder das Gesamtergebnis) bildet.

Problematisch an diesen Annahmen der Klassischen Testtheorie ist allerdings, dass sie in erster Linie, wie oben beschrieben, das formale Vorhandensein korrekter Lösungen unter Berücksichtigung von Messfehlern als Testwert fokussiert, ohne jedoch Aussagen über das Zustandekommen der Leistungen oder das Antwortverhalten der Testpersonen miteinzubeziehen (Mashkovskaya & Baur, 2016b, S. 217; Moosbrugger, 2012b, 104 ff). So erreichen beispielswiese zwei Personen formal das gleiche C-Test-Ergebnis (gleich viele Punkte), obwohl sie womöglich über ganz unterschiedliche schriftsprachliche Fähigkeiten verfügen. Denn ob von einer Person nur die leichten und von der anderen Person vor allem schwierigere Items gelöst wurden, oder ob beispielsweise die erste einfachere Hälfte eines Tests von einer Person sehr gut, die zweite Hälfte aber gar nicht mehr beantwortet wurde, bleibt bei der Auswertung nach der Klassischen Testtheorie unberücksichtigt, da allein der Testwert (Summenscore) als Ausprägung des zu messenden Merkmals zählt.

Ein weiterer Nachteil der Klassischen Testtheorie ist zudem, dass die Grundaxiome der Klassischen Testtheorie, d. h. die Annahme darüber, dass ein wahrer Testwert existiert und dass sich die Ausprägung eines Merkmals über eben diesen wahren Wert und einen zufälligen entstandenen Messfehler bestimmen lässt (Hartig & Jude, 2004, S. 5), nicht empirisch überprüfbar sind (Steyer & Eid, 2001, 205 ff.). Ebenfalls sind Berechnungen bzw. Normierungen, die auf der Grundlage der Klassischen Testtheorie erfolgen, stichprobenabhängig, da die Testpersonen bzw. deren Ergebnisse im Vergleich zur Gruppe gesehen werden, wodurch wiederum verallgemeinernde und verbindliche Aussagen über die Testergebnisse kaum möglich sind (Krampen, 2018).

Als Ergänzung der Klassischen Testtheorie wird die *Item-Response-Theorie* (*IRT*) verstanden. Im Gegensatz zu den oben beschriebenen Annahmen der Klassischen Testtheorie steht bei der Item-Response-Theorie nicht nur die Frage nach dem Test- bzw. Summenwert im Mittelpunkt, sondern vor allem *wie* eine Antwort bzw. ein

bestimmtes Antwortverhalten zustande gekommen ist (Moosbrugger, 2008, 215 ff.). Dabei wird prinzipiell angenommen, dass jedem Antwortverhalten einer Person auf ein Item eine beobachtbare Reaktion (manifeste Variable) und eine nicht beobachtbare Fähigkeit (latente Variable) zugrunde liegen (Kubinger, 2018). Die beobachtbaren Reaktionen, d. h. die Antworten auf die Items, werden als Indikatoren für latente Fähigkeiten, in diesem Fall die Sprachkompetenz, gesehen. Hierbei stehen sowohl die latente als auch die manifeste Variable in einem wahrscheinlichkeitsfunktionalen Zusammenhang (ebd.). Im Rahmen der Item-Response-Theorie geht es also um die Bestimmung dieses wahrscheinlichkeitsfunktionalen Zusammenhangs, genauer gesagt um die Wahrscheinlichkeit für das Auftreten einer bestimmten Reaktion in Abhängigkeit von der Personeneigenschaft (*Personenparameter*) und der Itemschwierigkeit (*Itemparameter*) (Moosbrugger, 2012a, 227 ff.).

Für die Bestimmung bzw. Berechnung dieser Wahrscheinlichkeitsbeziehungen zwischen den Item- und Personenparametern gibt es eine Reihe von mathematischen Modellen – das bekannteste von ihnen ist das sogenannte *Rasch-Modell* – die es ermöglichen, die Item- und Personenparameter stochastisch zu schätzen und das wahrscheinliche Antwortverhalten einer Person auf ein Item vorherzusagen (Eckes, 2007, 68 ff.). Bei Berechnungen mit dem Rasch-Modell geht es also im Wesentlichen um die Frage, wie hoch die Wahrscheinlichkeit einer Person mit einer geschätzten Personenfähigkeit ist, ein Item mit einer geschätzten Itemschwierigkeit korrekt zu lösen (Koller et al., 2012, S. 11). Dabei wird grundsätzlich angenommen, dass ein Item mit einer Wahrscheinlichkeit von mehr als 50 % gelöst wird, wenn die Itemschwierigkeit kleiner ist als die Personenfähigkeit. Andersherum wird ein Item mit einer Wahrscheinlichkeit von mehr als 50 % nicht gelöst, wenn die Schwierigkeit der Aufgabe größer ist als die Personenfähigkeit. Je größer die Differenz zwischen der individuellen Merkmalsausprägung und der Schwierigkeit des jeweiligen Items, desto größer ist damit die Wahrscheinlichkeit des Nicht-Lösens bzw. Lösens des Items (Breitenbach, 2006, 114 ff.; Moosbrugger, 2008, 215 ff.).

In IRT-Modellen sind die Item- und Teststatistiken stichprobenunabhängig. Das bedeutet, die Schätzung der Itemparameter hängt nicht von den Personenparametern ab (also davon, ob eine Stichprobe eher fähig oder eher weniger fähig ist). Umgekehrt

hängt die Schätzung der Personenparameter nicht davon ab, ob die Items eher schwer oder eher leicht sind. So erreicht beispielsweise eine Person mit einer höher geschätzten Fähigkeit auch eine höhere Lösungswahrscheinlichkeit für ein schwieriges Item als eine Person mit einer niedriger geschätzten Fähigkeit. Wird beiden Personen nun beispielsweise ein leichtes Item vorgelegt, würde die Person mit der höher geschätzten Fähigkeit nach wie vor die fähigere bleiben (Koller et al., 2012, S. 19). Der Vergleich zweier Personenfähigkeiten führt also immer zum selben Ergebnis, egal welches Item dafür verwendet wird. Gleiches gilt umgekehrt für den Vergleich von Itemparametern. Ergebnisse, die auf der Grundlage der IRT bzw. mit dem Rasch-Modell berechnet werden, sind daher – im Gegensatz zur Klassischen Testtheorie – unabhängig von der ausgewählten Stichprobe (Fischer, 1974; Koller et al., 2012, S. 19).Aufgrund der Stichprobenunabhängigkeit und empirischen Überprüfbarkeit wurde im Rahmen der hier vorliegenden Arbeit entschieden, die C-Test-Ergebnisse nicht der Klassischen Testtheorie folgend, sondern auf der Grundlage der Item-Response-Theorie (IRT) zu berechnen und zu normieren.

Eine Besonderheit bei der Berechnung gemäß der IRT bzw. bei der Anwendung des Rasch-Modells auf die C-Test-Daten ergibt sich allerdings durch das C-Test-Format. Das Rasch-Modell gilt nämlich (nur) dann, wenn die Lösungswahrscheinlichkeit einer Aufgabe, wie oben beschrieben, ausschließlich von der Fähigkeit der Person und der Itemschwierigkeit abhängt und es keine Rolle spielt, ob auch vorherige Items gelöst wurden oder nicht (Breitenbach, 2006, 114 ff., 2007; Mashkovskaya & Baur, 2016b, S. 218). Der C-Test erfüllt die Voraussetzung dieser lokalen Unabhängigkeit indes nicht. Die Lücken im C-Test stehen auf der Satzebene miteinander in Verbindung oder bauen aufeinander auf (Mashkovskaya & Baur, 2016b, S. 218). Das Lösen einer Lücke wird dann erschwert, wenn vorherige oder nachstehende Lücken nicht rekonstruiert werden konnten, da so der Kontext des Satzes fehlt (Harsch & Harting, 2010). Auch Mashkovskaya und Baur (2016a) greifen diesen Punkt auf und ergänzen, dass es kaum möglich sei, Items zu lösen, die sich auf Phraseologismen beziehen, wenn vorherige Items als Teile der Phraseologismen bereits falsch gelöst wurden (Mashkovskaya & Baur, 2016b, S. 218).

Um das Problem dieser *lokalen Abhängigkeit* der Lücken in C-Tests zu lösen, schlagen u. a. Grotjahn (2002) und Eckes (2006, 2007) vor, nicht jede einzelne Lücke als Item, sondern die Summe aller Lücken eines Teiltests als sogenanntes *Superitem* zu verstehen und statistisch zu analysieren. Demzufolge gilt jeder Teiltest als ein Superitem, das *m + 1* Itemwerte annehmen kann, wobei *m* für die Anzahl der Lücken in einem Text steht (Eckes, 2006, S. 8, 2007, S. 78). Enthält ein C-Test-Text beispielsweise 25 Lücken, so kann er als Superitem 26 verschiedene Werte annehmen, nämlich all jene, die auf einer Skala von 0 bis 25 liegen. Die Items werden also nicht mehr wie in der Klassischen Testtheorie einzeln und dichotom (richtig = 1 Punkt, falsch = 0 Punkte) betrachtet, sondern jeder einzelne C-Test als polytomes Superitem einer Skala (Mashkovskaya & Baur, 2016b, S. 218). Die Summe der sechs Superitems bildet den Gesamtscore für den C-Test.

Durch das Bilden der Superitems und des Gesamtscores können die C-Test-Ergebnisse unter Anwendung eines speziellen Rasch-Modells, dem *Ratingskalen-Modell*, analysiert und berechnet werden (Eckes, 2007, 68 ff.). Dabei wird über eine Transformation des Summenscores für jede Testperson ein geschätzter Fähigkeitswert bestimmt, ein sogenannter *WLE-Parameter* (*Weighted Likelihood Estimate*) (Eckes, 2007, S. 74; Mashkovskaya & Baur, 2016b, 218 f.). Zur Berechnungen der WLE-Parameter wurde das *Paket TAM* (Robitzsch, Kiefer & Wu, 2019) des Statistikprogramms *R* (R Core Team, 2019) verwendet.

4.4.3 Normierung und Interpretation

Zur Interpretation der C-Test-Ergebnisse ist eine Normierung des C-Tests notwendig, da über eine Normierung Vergleichswerte zur Verfügung gestellt werden, die eine Beurteilung und Ergebniseinordnung des Gesamtscores bzw. des WLE-Parameters anhand anderer Ergebnisse ermöglichen. Eine normorientierte Bewertung der C-Test-Daten bedeutet also, dass die individuellen C-Test-Ergebnisse relativ zu den durchschnittlich erbrachten Ergebnissen einer Bezugs- oder Referenzgruppe interpretiert werden.

Da im Rahmen dieser Arbeit ein standardisierter C-Tests für den Einsatz in der Sekundarstufe II generiert werden soll, bezieht sich die Referenzgruppe für die

Einordnung individueller C-Test-Ergebnisse nun also auf die in dieser Untersuchung fokussierte Kohorte der Oberstufenschüler*innen in NRW (vgl. Kapitel 4.2 zur Stichprobenkonstruktion und Kapitel 5.1 zur Stichprobenbeschreibung). Für die Normierung von C-Tests im Bereich der Erst- und Zweitsprache ist es in der C-Test-Praxis üblich, die Bezugsnorm nicht aus der gesamten Stichprobe zu bestimmen: Als Referenznormwert wird in der Regel der Median oder der (bereinigte) Mittelwert[94] allein der monolingualen Proband*innen festgelegt[95] (Baur et al., 2006, S. 394; Bremerich-Vos & Scholten-Akoun, 2016, S. 219; Scholten-Akoun & Baur, 2012). Dass die Leistungen von mehrsprachigen Personen bzw. von Personen mit Deutsch als Zweitsprache für die Ermittlung des Referenznormwertes unberücksichtigt bleiben und sich der Referenznormwert damit nicht an der gesamten Stichprobe orientiert, ruft oftmals besonders im schulischen Bereich Kritik hervor (Baur & Mashkovskaya, 2015, S. 442): So wird angenommen, dass Kinder mit Deutsch als Zweitsprache durch diese Vorgehensweise, d. h. den direkten Vergleich mit Erstsprachsprecher*innen, benachteiligt würden, da es unfair wäre, die Leistungen in einer Zweitsprache mit den Leistungen von Erstsprachsprecher*innen einer Sprache zu vergleichen (die Erstsprache wird zeitlich womöglich früher erworben und damit länger und intensiver im familiären und sozialen Umfeld verwendet) (Baur & Mashkovskaya, 2015, 453 ff.). Baur und Mashkovskaya (2015) stufen mögliche alternative Vorgehensweisen zur Bestimmung des Referenznormwertes – beispielsweise durch die Einbeziehung der Ergebnisse von Zweitsprachenlerner*innen oder das Festlegen separater Normen für bestimmte Gruppen (z. B. in Bezug auf Herkunftssprache, Ethnie oder Nationalität) – jedoch als „wenig sinnvoll und kontraproduktiv [ein], wenn in schulischen und außerschulischen Institutionen vergleichbare Bildungs- und Leistungsstandards [wie beispielsweise über das Zentralabitur] eingeführt werden sollen“ (Baur & Mashkovskaya, 2015, S. 442). Um gerade eine Benachteiligung von Schüler*innen mit sprachlichen Schwierigkeiten zu verhindern, müssten Förderbedarfe identifiziert werden, um gezielte

[94] „Bereinigter“ Mittelwert heißt in diesem Falle, dass bei der Ermittlung des Mittelwertes Extremwerte, d. h. Ausreißer sowohl im oberen als auch im unteren Bereich herausgerechnet werden (Bremerich-Vos und Scholten-Akoun (2016, S. 219)

[95] Vgl. zur statistisch-normativen Festlegung des Referenznormwertes über den Median oder den Mittelwert (Baur et al. (2006)

adressatenspezifische Förderprogramme anbieten zu können (Baur & Mashkovskaya, 2015, S. 442; Scholten-Akoun & Baur, 2012).

Die Diskussion über die Frage nach der Population, die als Basis für eine Normierung hinzugezogen werden soll, setzt jedoch oftmals implizit voraus, dass sich die Leistungen der monolingualen von den Leistungen mehrsprachigen Sprecher*innen grundsätzlich unterscheiden und sich der Referenznormwert auf der Grundlage der gesamten Stichprobe (im Vergleich zur Teilstichprobe der Monolingualen) verschieben würde. Auch wenn die Untersuchung von Bremerich-Vos & Scholten-Akoun et al. (2016) Hinweise darauf liefert, dass die monolingualen Studierenden im C-Test besser abschneiden als Studierende mit Deutsch als Zweitsprache (vgl. Mashkovskaya & Baur 2016, S. 220), bleibt zu prüfen, ob dieser Befund sich auch in der Kohorte der Oberstufenschüler*innen replizieren lässt und sich der Referenznormwert tatsächlich verschiebt, wenn nicht nur die monolingualen Schüler*innen sondern die gesamte Stichprobe als Referenzgruppe herangezogen werden würde.

Nachdem der Referenznormwert bestimmt wurde, werden für die Interpretation der C-Test-Ergebnisse zunächst Leistungsgruppen definiert, denen anschließend Förderbedarfe zugeschrieben werden (Baur et al., 2006; Baur & Spettmann, 2007, 2008a, 2008b; Bremerich-Vos & Scholten-Akoun, 2016; Mashkovskaya, 2014). Die Einteilung der Leistungsgruppen erfolgt dabei rein statistisch und unterliegt keiner inhaltich-kriterialen Begründung (Scholten-Akoun et al., 2013, S. 185). Ausgehend vom ermittelten Referenznormwert werden die Leistungsgruppen mithilfe der Standardabweichung nach folgender Vorgehensweise bestimmt[96]:

[96] Diese Angaben wurden aus Baur & Spettmanns Beschreibung zum Verfahren und zur Definition von Leistungsgruppen in Schulen (2007, S. 101 ff., 2010, s. 436 f.) sowie aus der Beschreibung der Einteilung von Leistungsgruppen bei Studierenden in der Publikation von Bremerich-Vos & Scholten-Akoun (2016, S. 220) übernommen.

Tabelle 5: Einteilung der Leitungsgruppen

Leistungsgruppe	Einordnung des Ergebnisses
I	< 2 Standardabweichungen über dem Referenznormwert
II	1 bis 2 Standardabweichung über dem Referenznormwert
III	± 1 Standardabweichung um den Referenznormwert
IV	1 bis 2 Standardabweichungen unter dem Referenznormwert
V	< 2 Standardabweichungen unter dem Referenznormwert

In den Leistungsbereich I fallen also die Test-Ergebnisse, die mehr als zwei Standardabweichung über dem Referenznormwert liegen. Ergebnisse im Leistungsbereich II sind Test-Ergebnisse zwischen einer und zwei Standardabweichungen über dem Referenznormwert. Ergebnisse, die eine Standardabweichung über oder eine Standardabweichung unter dem Referenznormwert liegen, werden hier dem Leistungsbereich III zugeordnet. Den Referenznormwert um eine bis zwei Standardabweichung unterschreitende Test-Ergebnisse liegen im Leistungsbereich IV, ihn um mehr als zwei Standardabweichungen unterschreitende im Leistungsbereich V.

Die Zuordnung zu Förderbedarfen – die sich an der zunächst rein statistischen Einteilung in die Leistungsbereiche orientiert – basiert hierbei auf dem Konzept, dass im C-Test niedrigere Werte, d. h. Werte, die unter oder weit unter den durchschnittlichen erbrachten Werten der Referenzgruppe (des Referenznormwerts) liegen, auch defizitäreren Kompetenzen entsprechen, und hier vor allem in den mit dem C-Test vorrangig erfassten sprachlichen Bereichen der Lexik, Semantik, Syntax, Leseverständnis und Orthografie (vgl. dazu Kapitel 2.1.2 und Kapitel 2.3) (Mashkovskaya & Baur, 2016a, 192 ff.) – auch wenn sich dies auf der Grundlage allein der Werte nicht eindeutig qualifizieren lässt. Testergebnisse, die in den Leistungsgruppen I und II verortet sind, signalisieren laut Mashkovskaya und Baur (2016b, 220 f.) keinen Förderbedarf in diesen Bereichen. Die Proband*innen verfügen über gut ausgebaute sprachliche Kompetenzen und können die C-Test-Lücken auf Wort-, Satz- und Textebene annähernd vollständig

rekonstruieren (Mashkovskaya, 2014, S. 118; Mashkovskaya & Baur, 2016b, 220 f.). Personen, deren Testergebnisse in der Leistungsgruppe III liegen, weisen keinen eindeutig erkennbaren Förderbedarf auf. Von einem Förderbedarf in den oben genannten sprachlichen Bereichen wird ab einer Unterschreitung des Referenznormwerts um eine bzw. mehr als eine Standardabweichung, d. h. ab dem Leistungsbereich IV, ausgegangen. Test-Ergebnisse in den Leistungsbereichen IV und V also legen nahe, dass auf sprachlicher Ebene (s. o.) teilweise erhebliche Defizite bestehen und das sprachliche Kompetenzniveau der Referenzgruppe nicht erreicht ist: Mashkovskaya und Baur (2016b, S. 221) stellen in ihrem Fazit der sprachlichen Analyse der C-Test-Bearbeitungen mit Test-Ergebnissen, die der Leistungsgruppen IV zuzuordnen sind, fest, dass die Proband*innen teilweise nicht in der Lage sind, Textzusammenhänge auf der Satz- und Textebene zu erschließen. Dieser Eindruck verschärft sich weiter bei der Betrachtung von C-Test-Bearbeitungen mit Ergebnissen der Leistungsgruppe V. Hier sind die Lücken im C-Test „teilweise übersprungen oder in einer Weise falsch vervollständigt, dass einzelne Sätze oder sogar der ganze Text keinen Sinn ergab. Zudem kamen in den Texten grammatische und orthographische Fehler vor“ (Mashkovskaya & Baur, 2016b, S. 221). Den Proband*innen mit Ergebnissen, die in der Leistungsgruppe V verortet wurden, gelingt es in der Regel nicht (oder nur bei einzelnen C-Tests), die beschädigten Texte sinnvoll zu rekonstruieren.

In welchen konkreten sprachlichen Bereichen die einzelnen Schwierigkeiten und Defizite liegen, kann mit dem C-Test nicht diagnostiziert werden. Als Screening-Instrument dienen seine Ergebnisse vor allem als „Indikatoren für generell feststellbare sprachliche Defizite“ (Baur et al., 2006; Baur et al., 2012; Baur et al., 2013b; Baur & Spettmann, 2009; Mashkovskaya & Baur, 2016b, S. 221).

4.4.4 Verfahren zur Überprüfung der Hypothesen

Zur Überprüfung der im Rahmen dieser Arbeit formulierten Hypothesen werden auf der Grundlage der erhobenen Daten sowohl Unterschiede zwischen bestimmten Gruppen als auch Zusammenhänge zwischen einzelnen Variablen berechnet (Vgl. Kapitel 5). Soll also beispielsweise bestimmt werden, ob es einen Zusammenhang zwischen der

Deutschnote und dem C-Test-Ergebnis gibt, kann über eine Korrelation die Beziehung zwischen diesen beiden Variablen berechnet werden. Korrelationen, die den Zusammenhang zwischen zwei Variablen – der Deutschnote (Variable A) und dem C-Test-Ergebnis (Variable B) wie im Beispiel – angeben, werden bivariate Korrelationen genannt (Schäfer, 2010, S. 109). Der Korrelationskoeffizient r ist dabei der statistische Wert, der das Ausmaß des linearen Zusammenhangs zwischen beiden Variablen beschreibt. Der Korrelationskoeffizient kann Werte im Bereich von -1 bis +1 annehmen (Schäfer, 2010, 114 f.). Eine Korrelation von 0 steht dabei für keinen, eine Korrelation von -1 bzw. 1 für einen perfekten negativen bzw. positiven Zusammenhang (Schäfer, 2010, S. 115). Da sich in wissenschaftlichen Untersuchungen jedoch nur äußerst selten ein perfekter Zusammenhang zwischen zwei Variablen nachweisen lässt, schlägt Cohen (1988) für die Interpretation von Korrelationskoeffizienten folgende, auf der Grundlage durchschnittlicher Korrelationswerte aus Forschungsstudien generierte, Konvention vor, nach der eine Korrelationen als gering, mittel oder hoch einzustufen ist und nach der sich die Einschätzung der Korrelationskoeffizienten dieser Untersuchung richten (Schäfer, 2010, S. 119):

r = -/+ 0,1 für eine geringe Korrelation
r = -/+ 0,3 für eine mittlere Korrelation
r = -/+ 0,5 für eine hohe Korrelation (nach Cohen, 1988).

Im ersten Teil der Datenauswertung bzw. –analyse werden also zunächst die als Hypothesen formulierten Unterschieden bzw. Zusammenhänge von jeweils zwei Variablen (z. B. Deutschnote und C-Test-Ergebnis) über bivariate Korrelationen bestimmt (vgl. Kapitel 5.4 und 5.5 Ergebnisse). Sowohl in der empirischen Bildungsforschung als auch in vielen Bereichen der (Sprach-)Lern-Lehrforschung gibt es allerdings nur selten eine Ursache für eine Wirkung. Vielmehr werden die Werte einer abhängigen Variablen, in diesem Fall die mit dem C-Test zu messende Sprachkompetenz, durch mehrere unabhängige Variablen, beispielsweise durch die Erstsprache, die Sprachverwendung, das Alter und das Geschlecht, beeinflusst. Ein Modell, mit dem diesem Umstand Rechnung getragen werden kann, ist die multiple lineare Regression (Schäfer, 2010, S. 121). Als Erweiterung der bivariaten Korrelation, können in einer

multiplen linearen Regression mehrere unabhängige Variablen gleichzeitig berücksichtig, d. h. ihr Einfluss auf die abhängige Variable geschätzt werden (Tausendpfund, 2019, 139 f.) (vgl. Kapitel 5.6). So können zum einen eventuell vorhandene Scheinkorrelation zum Vorschein gebracht werden, zum anderen können gezielte Aussagen darüber getroffen werden, welche Merkmalsausprägung in den unabhängigen Variablen im Durchschnitt zu welchem Wert der abhängigen Variablen führen. Über die lineare Regression kann ebenfalls bestimmt werden, welche im Modell enthaltenen unabhängigen Variablen einen stärkeren bzw. schwächeren Einfluss auf die abhängige Variable besitzen (Tausendpfund, 2019, 144,149). So soll im zweiten Teil der Datenauswertung der Frage nachgegangen werden, welche der über den Fragebogen erfassten Merkmale der Oberstufenschüler*innen (Geschlecht, Alter, Schulform, Erstsprache, Migrationshintergrund, Bildungsniveau der Eltern, etc.) im Vergleich einen besonders starken Einfluss auf die mit dem C-Test gemessene Sprachkompetenz haben (vgl. Kapitel 5.6).

5. Ergebnisse

5.1 Stichprobenbeschreibung

Die Datenerhebung im Erhebungszeitraum von Ende 2015 bis Ende 2018 fand an 15 Schulen in NRW (in Bergisch-Gladbach, Duisburg, Düsseldorf, Essen, Gelsenkirchen, Haan, Meerbusch und Ratingen) statt, an denen insgesamt 1585 Schüler*innen mit dem C-Test getestet wurden. In den Tabellen (vgl. u. Tab. 6 bis Tab. 9) sind die Verteilungen der getesteten Oberstufenschüler*innen in Abhängigkeit vom Schulstandorttyp, der Schulform, der Kurswahl, der Deutschnote, dem Geschlecht, dem Alter, dem Migrationsstatus, der Sprachverwendung, der Einschulung, dem Bildungsgrad der Eltern und dem Studienwunsch (vgl. Kapitel 3 und Kapitel 4.1.3 zur Erläuterung der genannten Aspekte) in Häufigkeit und Prozent angegeben.

5.1.1 Schulen, Deutschnoten, Alter und Geschlecht

Tabelle 6: Verteilung der Daten in Bezug auf Standorttyp, Deutschnote, Alter und Geschlecht

	Häufigkeit (N)[97]	**Prozent**
Standorttyp		
Standorttyp 1	128	8,1
Standorttyp 2	522	32,9
Standorttyp 3	116	7,3
Standorttyp 4	436	27,5
Standorttyp 5	383	24,2
Gesamt	**1585**	**100,0**
Schulform/Stufe		
Gymnasium EF	1013	64,2
Gymnasium Q1	402	25,5
Gesamtschule EF	162	10,3

[97] Die unterschiedlichen Gesamthäufigkeiten (N) für die jeweiligen Variablen sind im Folgenden mit der unterschiedlichen Anzahl der gültigen Fälle für das betreffende Merkmal zu erklären: So kommt es vor, dass Schüler*innen bestimmte Fragen im Fragebogen nicht beantwortet haben (z. B. keine Angabe darüber gemacht haben, welche Stufe bzw. Phase in der Oberstufe sie besuchen) und somit insgesamt weniger Daten über die abgefragte Variable vorliegen.

Gesamt	**1577**	**100,0**
Kurswahl Deutsch		
Leistungskurs	384	25,3
Grundkurs	1132	74,7
Gesamt	**1516**	**100,0**
Deutschnote		
1	131	8,6
2	523	34,3
3	635	41,6
4	223	14,6
5	13	0,9
6	1	0,1
Gesamt	**1523**	**100**
Alter		
14	36	2,3
15	537	35,1
16	601	39,2
17	274	17,9
18	67	4,4
19	15	1,0
20	1	0,1
24	1	0,1
Gesamt	**1532**	**100**
Geschlecht		
Männlich	751	48,8
Weiblich	788	51,2
Gesamt	**1539**	**100,0**

Auf der Grundlage des Standorttypenkonzepts werden in NRW alle weiterführenden Schulen einem von insgesamt 5 Standorttypen zugeordnet (QUA-LIS NRW, 2017) (vgl. Kapitel 4.2). Im Rahmen dieser Untersuchung konnten Daten an jedem der 5 Schulstandorttypen generiert werden – allerdings in stark unterschiedlichem Ausmaß pro Standorttyp. So wurde über die Hälfte der Daten an den Standorttypen 4 und 5 erfasst (51,7 %) und knapp ein Drittel am Standorttyp 2 (32,9 %); an Schulen der Standorttypen 1 und 3 jeweils deutlich unter 10 %. Die geringeren Fallzahlen können bei einem Vergleich der Standorttypen besonders in Bezug auf deskriptiv dargestellte Ausprägung der Merkmale zu Verzerrungen führen, die bei der Analyse von

Unterschieden und/oder Zusammenhängen statistisch auf Signifikanz hin zu überprüfen sind (vgl. Kapitel 5.5). Eine ähnliche Unwucht im Sample lässt sich bei der Verteilung der Daten in Bezug auf die Schulform und die Jahrgangsstufe beobachten: So wurden fast zwei Drittel der Daten in der Einführungsphase am Gymnasium (64,2 %) generiert, 25,5 % in der Qualifikationsphase 1 an Gymnasien und nur ein kleiner Teil der Gesamtstichprobe (10,3 %) in der Einführungsphase an Gesamtschulen (vgl. Kapitel 4.2 und 4.3 zur Stichprobenkonstruktion und Datenerhebung).

Der Anteil an Frauen und Männern ist in der Stichprobe in etwa gleich verteilt, allerdings zeigen sich in Bezug auf einige Variablen geschlechtsspezifische Unterschiede: Werden beispielsweise die Daten zur Deutschnote nach Geschlecht getrennt betrachtet, fällt auf, dass Schülerinnen im Vergleich zu Schülern fast doppelt so häufig eine sehr gute Note im Fach Deutsch erreichen (w/m: 10,7 %; 6,6 %). Bei den männlichen Teilnehmern hingegen kommen die Deutschnoten ausreichend (4) und mangelhaft (5) fast doppelt so häufig vor wie bei den weiblichen Teilnehmerinnen (w/m: 11,4 %; 19,4 %). Betrachtet man die Schulnoten allein, lässt sich also eine Tendenz zur Bestätigung des Stereotyps erkennen, das Mädchen eine bessere Sprachkompetenz als Jungen zuweist. Tatsächlich wählen auch deutlich mehr Schülerinnen (31,4 %) das Fach Deutsch als Leistungskurs als Schüler (19,8 %). Kompetenz und Neigung scheinen einander zu entsprechen.

5.1.2 Migrationshintergrund und Sprache/Sprachverwendung

Tabelle 7: Verteilung der Daten in Bezug auf Migrationshintergrund, Sprache und Sprachverwendung

	Häufigkeit (N)	Prozent
Migrationsstatus		
Ohne Migrationshintergrund	870	58,3
Mit Migrationshintergrund	622	41,7
1. Generation	92	15,1
2. Generation	514	84,3
Befragte Person wurde im Ausland geboren, beide Elternteile in Deutschland	4	0,7
Erstsprache		

Deutsch	1055	68,7
Türkisch	143	9,3
Russisch	54	3,5
Polnisch	37	2,4
Arabisch	36	2,3
Albanisch	33	2,1
Andere[98]	178	11,6
Gesamt	**1536**	**100**
Erstsprache Mehrsprachig	26	1,7
Sprachverwendung mit der Mutter		
Deutsch	1180	76,8
Türkisch	94	6,1
Russisch	42	2,7
Polnisch	31	2,0
Albanisch	25	1,6
Arabisch	21	1,4
Andere	143	9,3
Gesamt	**1536**	**100**
Sprachverwendung mit dem Vater		
Deutsch	1152	76,8
Türkisch	93	6,2
Russisch	37	2,5
Arabisch	27	1,8
Albanisch	25	1,7
Polnisch	23	1,5
Andere	143	9,5
Gesamt	**1500**	**100**
Sprachverwendung mit den Geschwistern		
Deutsch	1312	94
Türkisch	22	1,6
Arabisch	9	0,6
Englisch	5	0,5
Andere	48	3,4
Gesamt	**1396**	**100**
Erstsprache und Sprachverwendung innerhalb der Familie		
Monolingual deutschsprachig	995	64,6

[98] In der Kategorie „Andere“ ist Kurdisch mit 23 Sprecher*innen die nächstgrößte Sprachgruppe, gefolgt von Tamil mit 15 Sprecher*innen. Englisch, Spanisch, Griechisch, Italienisch und Vietnamesisch verzeichnen jeweils zwischen 9 und 11 Sprecher*innen, was in der hier untersuchten Stichprobe allerdings jeweils einen prozentualen Anteil von weniger als 1% ausmacht.

Mehrsprachig	546	35,4
Gesamt	**1541**	**100**
Sprachverwendung mit den Freunden		
nur Deutsch	800	52,8
überwiegend Deutsch	537	35,5
teils Deutsch/ teils andere Erstsprache(n) als Deutsch	162	10,7
überwiegend andere Erstsprache(n) als Deutsch	10	0,7
nur andere Erstsprache(n) als Deutsch	5	0,3
Gesamt	**1514**	**100**
Einschulung		
In Deutschland	1467	97,4
In einem anderen Land	39	2,6
Gesamt	**1506**	**100,0**

Einen Migrationshintergrund, der hier zunächst über den Geburtsort der Eltern bzw. Schüler*innen erfasst wurde (vgl. Kapitel 4.1.3), geben 41,7 % der befragten Schüler*innen an. Etwa 15 % zählen zur 1. Generation (d. h. sowohl sie selbst als auch ihre Eltern wurden im Ausland geboren), der deutlich überwiegende Teil der Schüler*innen mit Migrationshintergrund ist aber bereits in Deutschland geboren (sie gehören also zur 2. Generation, ihre Eltern sind im Ausland geboren, sie selbst aber in Deutschland). Nur 4 Personen geben an, dass sie selbst im Ausland, ihre Eltern aber in Deutschland geboren wurden. Angenommen wird hier, dass diese Personen während eines Auslandsaufenthalts der Eltern geboren wurden, z.B. während eines Auslandsstudiums oder einer Beschäftigung im Ausland.

Der Anteil der Schüler*innen mit Migrationshintergrund in dieser Untersuchung (41,7 %) liegt somit recht deutlich über den vom Ministerium für Schule und Bildung ermittelten Schuldaten für NRW. Die Schuldaten des Ministeriums für Schule und Bildung für NRW, das zur Definition des Migrationshintergrunds neben dem Geburtsort auch noch aus der Verwendung einer anderen Sprache als Deutsch in der Familie auf einen Migrationshintergrund schließen (vgl. ausführlicher zur Methodendifferenz Kap. 4.1.3), geben einen Anteil von 27,6 % an Schüler*innen mit Migrationshintergrund an, der im Schuljahr 17/18 die Sekundarstufe II in NRW besucht hat (MSB NRW, 2019b, 96 ff.). Würden die Kriterien des Ministeriums NRW zur Ermittlung des

Migrationshintergrunds – also *(a) selbst nicht in Deutschland geboren, (b) Eltern nicht in Deutschland geboren oder (c) innerhalb der Familie wird eine nichtdeutsche Verkehrssprache gesprochen* (MSB NRW, 2019b, S. 4) (vgl. auch Kap. 4.3.1) zugrunde gelegt, dann läge der Anteil von Schüler*innen mit Migrationshintergrund im hier erfassten Sample noch etwas höher, nämlich bei 43,9 %.

In der PISA-Studie 2018 – in der das Merkmal *Zuwanderungsgeschichte* über die gleichen Definitionskriterien (Geburtsort der Eltern/Schüler*innen) zugeschrieben wird wie in dieser Arbeit (vgl. Kapitel 4.1.3) – liegt der Anteil an Schüler*innen mit Migrationshintergrund mit 35,6 % zwar insgesamt höher als beispielsweise in den Schuldaten NRW, im Vergleich zum Anteil in dieser Arbeit (41,7 %) aber immer noch niedriger (Weis, Müller et al., 2019, S. 151). Zu vermuten ist, dass der vergleichsweise hohe Anteil in der hier untersuchten Stichprobe durch die Verzerrungen bei der Zufallsauswahl der Schulen (regionale Schwerpunktbildung *Ruhrgebiet*, vgl. Kapitel 4.2 und Kapitel 4.3) bzw. der Standorttypen begründet werden kann: Mehr als die Hälfte der getesteten Schüler*innen wird an Schulen der Standorttypen 4 und 5 (51,7 %) unterrichtet, d. h. an Schulen, an denen laut Standorttypenkonzept NRW der Anteil der Schüler*innen mit Migrationshintergrund Standorttyp 4 im Durchschnitt bei 43 % und am Standorttyp 5 durchschnittlich bei 61 % liegt[99] (QUA-LIS NRW, 2017). Mit Blick auf die Anteile der Schüler*innen mit Migrationshintergrund an den unterschiedlichen Schulstandorttypen in dieser Studie zeigt sich, dass die Anteile der Schüler*innen mit Migrationshintergrund an den Schulen des Standorttyps 5 mit durchschnittlich 59,6 % dem im Standorttypenkonzept beschrieben Durchschnittswerten in etwa entsprechen, die Schulen des Standorttyps 4 mit 50,4 % jedoch über dem berichteten Mittelwert liegen. Auch am Standorttypen 1 übertreffen die Anteile der Schüler*innen mit Migrationshintergrund die im Standorttypenkonzept ausgewiesenen Durchschnittswerte, am Standorttyp 3 hingegen werden sie unterschritten (QUA-LIS NRW, 2017). Hinzuweisen ist an dieser Stelle natürlich erneut auf das vergleichsweise kleine N für die Stichprobe zu den Standorttypen 1 und 3. Die Daten von Schüler*innen von jeweils nur zwei Schulen speisen die Stichprobe, weshalb ein Vergleich mit

[99] Beim dem Verfahren zur Bestimmung der Schulstandorttypen vgl. Kapitel 4.2.

repräsentativ erhobenen Daten wie für die PISA-Studie nur sehr eingeschränkt möglich ist.

Neben den Angaben zur Bestimmung des Migrationshintergrundes (Geburtsland der Eltern/Schüler*innen) wurden Fragen zur Erstsprache der Oberstufenschüler*innen sowie zur Sprachverwendung gestellt. Als Antwort auf die Frage nach der „Muttersprache“ (vgl. Kapitel 4.1.2) wurden insgesamt 50 Sprachen genannt: Afghanisch, Albanisch, Amharisch, Arabisch, Berberisch, Bosnisch, Bulgarisch, Chinesisch, Dari, Deutsch, Edo, Englisch, Flämisch, Französisch, Georgisch, Ghanaisch, Griechisch, Hindi, Italienisch, Koreanisch, Krio, Kroatisch, Kurdisch, Lingala, Litauisch, Mazedonisch, Nepali, Pashtu, Persisch, Pijin, Polnisch, Portugiesisch, Punjabi, Rumänisch, Russisch, Serbisch, Serbokroatisch, Spanisch, Tagalog, Tamil, Thailändisch, Tschechisch, Tunesisch, Türkisch, Turkmenisch, Ukrainisch, Ungarisch, Urdu, Vietnamesisch und Weißrussisch.

Deutsch ist mit 68,7 % die am häufigsten angegebene Erstsprache der befragten Oberstufenschüler*innen, gefolgt von Türkisch mit 9,3 %, Russisch mit 3,5 % und Polnisch mit 2,4 %, Arabisch mit 2,3 % und Albanisch mit 2,1 %. 26 Personen geben an, mindestens zwei Sprachen als Erstsprachen zu sprechen, von denen 15 Personen eine Kombination von *Deutsch + eine (oder mehrere) andere Sprache(n) als Deutsch* und 11 Personen *mindestens zwei andere Sprachen als Deutsch* als Erstsprache sprechen (vgl. Kapitel 4.1.3).

Im Kontakt zur Mutter und zum Vater wird mit jeweils 76,8 % am häufigsten Deutsch gesprochen, gefolgt von Türkisch (6,1 % und 6,2 %) und Russisch (2,7 % und 2,5 %). Während mit der Mutter am dritthäufigsten auf Polnisch (2,0 %) kommuniziert wird, gefolgt von Albanisch (1,6 %) und Arabisch (1,4 %), wird im Kontakt mit dem Vater am dritthäufigsten Arabisch (1,8 %), dann Albanisch (1,7 %) und erst am sechsthäufigsten Polnisch (1,5 %) verwendet. Darüber hinaus sprechen etwas mehr als 9 % andere als die hier genannten Sprachen.

Im Vergleich dazu wird von den befragten Personen im Kontakt zu den Geschwistern mit 94 % häufiger Deutsch gesprochen als im Kontakt zu den Eltern. Lediglich 6,1 % der

befragten Schüler*innen sprechen mit ihren Geschwistern (eine) andere Sprache(n) als Deutsch.

Ein ähnliches Bild zeigt sich auch in der Sprachverwendung im Umgang mit Freunden: Der Großteil der Proband*innen spricht ausschließlich (52,8 %) bzw. überwiegend (35 %) Deutsch im Kontakt mit Freunden. 10,7 % der Schüler*innen verwenden teils Deutsch, teils (eine) andere Erstsprache(n) als Deutsch in der Kommunikation mit Freunden und lediglich 1,1 % benutzt hier überwiegend oder gänzlich eine andere Sprache als Deutsch.

Insgesamt kann auf der Grundlage der hier vorliegenden Daten davon ausgegangen werden, dass 35,4 % der Schüler*innen mehrsprachig sind, d. h. dass sie eine andere Sprache als/neben Deutsch als Erstsprach erworben haben und/oder eine andere Sprache als Deutsch im familiären Umfeld (im Kontakt mit Mutter, Vater, Geschwistern) verwenden. Aus diesen und den oben beschrieben Zahlen lässt sich ebenfalls der Befund ableiten, dass die Schüler*innen sozusagen generationensensitiv ihre Sprachwahl treffen: Sie sind erstens beweglich und können mehrheitlich in beiden Sprachen kommunizieren. Mit der älteren Generation, d. h. mit den Eltern, wird vorzugsweise die Herkunftssprache, mit den Peers innerhalb der Familie und Freunden tendenziell eher Deutsch gesprochen.

Von den getesteten Proband*innen wurden insgesamt 39 Personen nicht in Deutschland eingeschult. Von diesen 39 Personen sind 35 mehrsprachig, 32 Schüler*innen wurden selbst nicht in Deutschland geboren. Bei diesen 35 Personen dürfte es sich mehrheitlich um sogenannte „Seiteneinsteiger“ bzw. Schüler*innen mit Fluchterfahrung handeln, die (ausgehend von der Erstsprache) aus dem arabischen, afghanischen und ukrainischen Sprachraum kommen. Sie haben das deutsche Schulsystem nicht von Klasse 1 an durchlaufen, sondern sind zu einem späteren Zeitpunkt in eine deutsche Schule bzw. die Sekundarstufe II in Deutschland eingegliedert worden.

Tabelle 8: Verteilung der Daten in Bezug auf Migrationshintergrund und Sprache

Migrationshintergrund und Sprache	Häufigkeit (N)	Prozent
Schüler*innen mit Migrationshintergrund	**622**	**41,7**
Davon mit nicht-deutscher Familiensprache/Erstsprache	505	81,2
Davon mit Familiensprache/Erstsprache Deutsch	117	18,8
Mehrsprachige Schüler*innen	**546**	**35,4**
Davon ohne Migrationshintergrund	23	4,3
Davon mit Migrationshintergrund	506	95,7

Vergleicht man nun die Angaben zum Migrationshintergrund und zur Sprachverwendung kann für die hier vorliegende Stichprobe festgehalten werden, dass ein Migrationshintergrund (definiert über den Geburtsort der Eltern/Schüler*innen, vgl. dazu Kapitel 4.1.3) ähnlich wie von Chlosta und Ostermann (2005) beobachtet, nicht in allen Fällen mit einer anderen Sprachverwendung als Deutsch einhergeht und umgekehrt: Insgesamt lässt sich für 41,7 % der befragten Personen in der Stichprobe ein Migrationshintergrund konstatieren, allerdings sind nur 35,4 % der Proband*innen mehrsprachig (d.h. sie sprechen als Erstsprache und/oder zu Hause mindestens eine andere/weitere Sprache außer Deutsch, vgl. Kapitel 4.1.3). In der hier beschriebenen Stichprobe sprechen 117 Schüler*innen mit Migrationshintergrund in ihren Elternhäusern ausschließlich Deutsch. Umgekehrt geben 23 Personen an, dass sie zwar keinen Migrationshintergrund aufweisen, neben Deutsch aber mindestens eine weitere Sprache als Erstsprache und/oder Familiensprache sprechen. Von den 564 mehrsprachigen Personen sprechen insgesamt 63 (11,5 %) Deutsch als Erstsprache verwenden, jedoch im Kontakt mit den Eltern oder Geschwistern meistens eine andere Sprache als/neben Deutsch. Umgekehrt haben 87 Personen (17,6 %) eine andere Erstsprache als Deutsch, sprechen mit der Familie aber vornehmlich Deutsch.

Die separate Erfassung der Merkmale *Migrationshintergrund* und *Sprache* bzw. *Sprachverwendung* ist also im Rahmen dieser Untersuchung insofern sinnvoll, als dass über diese Merkmale – zumindest in Teilen – unterschiedliche Gruppen erfasst werden

und Tendenzen der abnehmenden Verwendung bzw. der generationenspezifischen Verwendung der Herkunftssprachen sichtbar werden.

5.1.3 Bildungsgrad der Eltern und Studienwunsch

Tabelle 9: Verteilung der Daten in Bezug auf den Bildungsgrad der Eltern und Studienwunsch

	Häufigkeit (N)	**Prozent**
Bildungsniveau Mutter		
kein Bildungsabschluss	53	4,4
Hauptschule	91	7,5
Realschule	328	27,0
Gymnasium	374	30,8
Akademischer Abschluss	370	30,4
Gesamt	**1216**	**100,0**
Bildungsniveau Vater		
kein Bildungsabschluss	46	3,9
Hauptschule	136	11,4
Realschule	234	19,6
Gymnasium	346	29,1
Akademischer Abschluss	429	36,0
Gesamt	**1191**	**100,0**
Studienwunsch		
Ja	971	64,6
Nein	85	5,7
Unentschlossen	446	29,7
Gesamt	**1502**	**100,0**

In Bezug auf die Frage nach dem höchsten Bildungsabschluss der Eltern, lässt sich auf für die hier beschriebene Stichprobe festhalten, dass der Großteil der Eltern, d. h. 61,2 % der Mütter und fast zwei Drittel der Väter (65,1 %) der Oberstufenschüler*innen über die Hochschulreife oder einen akademischen Abschluss als höchsten Bildungsabschluss verfügen. Es zeigt sich also auch in dieser Stichprobe, dass der Schulabschluss der Eltern ein guter Prädiktor für einen analogen Schulweg der Kinder darstellt. Diesen Zusammenhang zeigen auch die Daten des Statistischen Bundesamtes: So gehen

Schüler*innen, deren Eltern über einen hohen Bildungsgrad, d. h. über die Fachhochschul- oder Hochschulreife, verfügen, besonders häufig aufs Gymnasium (64 %); wesentlich seltener besuchen Schüler*innen mit Eltern mit Abitur Hauptschulen (ca. 15 %) (Statistisches Bundesamt, 2018, S. 109). Noch deutlicher zeigt sich dieser Zusammenhang allerdings im Falle der Schüler*innen, deren Eltern keinen oder nur einen Hauptschulabschluss besitzen. Nur 8,7 % dieser Schüler*innen besuchen ein Gymnasium (Statistisches Bundesamt, 2018, S. 109). Bildungsaufstieg findet in Deutschland – wie an vielen Stellen beklagt (Weis, Müller et al., 2019, 129 ff.) zu selten statt. In der hier untersuchten Stichprobe konnten allerdings leicht überdurchschnittliche Anteile der Schüler*innen ermittelt werden, deren Eltern über keinen oder nur über einen Hauptschulabschluss verfügten. Insgesamt 11,9 % (Mütter) bzw. 15,3 % (Väter) betrug der Anteil dieser Gruppe in den untersuchten Oberstufenjahrgängen, was also deutlich über dem bundesweiten Durchschnitt liegt. Differenziert man das Ergebnis allerdings nach Schulformen (Gymnasium vs. Gesamtschule) zeigen sich deutliche Unterschiede[100]: So fällt der Anteil an Schüler*innen, deren Eltern über keinen oder einen Hauptschulabschluss verfügen an den Gesamtschulen mit 14,6 % deutlich höher aus als an den Gymnasien (5,6 %). Umgekehrt besuchen nur 46,9 % der Schüler*innen, deren Eltern eine Hochschulreife oder einen akademischen Abschluss haben, die gymnasiale Oberstufe an der Gesamtschule, wohingegen diese Gruppe am Gymnasium mit 78,8 % vertreten ist. Ein Bildungsaufstieg, so deuten es die hier erhobenen Zahlen an, ist deutlich eher an den Gesamtschulen möglich als an Gymnasien.

Den Wunsch, nach dem Abitur ein Studium aufzunehmen, äußerten fast zwei Drittel (64,6 %) aller befragten Schüler*innen unabhängig vom Bildungsgrad der Eltern. D. h., dass Schüler*innen, deren Eltern selbst studiert haben, nicht öfter ein Studium aufzunehmen beabsichtigen als Schüler*innen aus Elternhäusern mit niedrigeren Bildungsabschlüssen der Eltern. Das Interesse an einem Studium ist in den Oberstufen der Gymnasien deutlich virulenter (66,4 %) als an den Gesamtschulen (49,4 %). Etwa ein

[100] An dieser Stelle ist darauf hinzuweisen, dass die Fallzahl an den Gesamtschulen (N = 162) im Vergleich zur Datenlage an den Gymnasien sehr gering ausfallen, wodurch es zu Verzerrungen in der Verteilung kommen kann (siehe oben).

Drittel aller befragten Schüler*innen (29,7 %) sind in Bezug auf die Frage nach einem Studienwunsch noch unentschlossen.

5.2 Empirische Überprüfung des eingesetzten C-Test-Sets (Schwierigkeit und Verteilung)

Um den in dieser Arbeit entwickelten finalen C-Test (vgl. Kapitel 4.1.2) für die Oberstufe wissenschaftlich evaluieren zu können, wird der C-Test bzw. die mit dem C-Test erhobenen Daten im Folgenden zunächst hinsichtlich der Testschwierigkeit und der Datenverteilung überprüft.

Der nachfolgenden Tabelle sind die Mittelwerte der C-Test-Rohwerte ($\bar{x}$), die Standardabweichungen (SD) und die Schwierigkeitsindizes (P) der einzelnen C-Tests und des ganzen C-Test-Sets zu entnehmen. Der Schwierigkeitsindex (P) steht dabei für die Lösungswahrscheinlichkeit eines C-Tests, da er die prozentuale Häufigkeit, mit der der C-Test gelöst werden konnte, angibt (vgl. Kapitel 4.1.2). Ein niedriger P-Wert (der Test wurde prozentual seltener gelöst) weist einen C-Test somit als schwierig zu lösen aus, während ein hoher P-Wert (der Test wurde prozentual häufig gelöst) für einen eher leicht zu lösenden C-Test steht (Lienert & Raatz, 1998, S. 73; Mashkovskaya, 2014, 188 f.; Wockenfuß, 2009, S. 156). Die numerische Größe//Die Höhe des Schwierigkeitsindex P kennzeichnet damit im eigentlichen Sinne die „Leichtigkeit“ eines Items und nicht die „Schwierigkeit“ (Moosbrugger & Kelava, 2012, S. 77). Die Schwierigkeitsgrade einzelner Aufgaben können mit der Formel $P_i = \frac{\bar{x}_i}{max_{xi}} \times 100$ berechnet werden ($\bar{x}_i$ = Mittelwert des Items; max_{xi} = maximale Anzahl Items) und sollten laut Raatz und Klein-Braley (1983, 107 ff.) zwischen P=20 und P=80 variieren (Wockenfuß, 2009, S. 157). Ein P-Wert von P=50 beschreibt eine mittlere Item-Schwierigkeit (ALTE, 2005, 16 f.; Moosbrugger & Kelava, 2012; Wockenfuß, 2009, S. 156).

Tabelle 10: Tabelle: Schwierigkeitskeitsindizes P (Schwierigkeit), $\bar{x}$ (Mittelwert), SD (Standardabweichung)

C-Test (N = 1.585)	P	$\bar{x}$	SD
Der Reiz von Fernsehserien (C-Test 1)	62,47	15,62	5,534
Wissenschaftliches Arbeiten (C-Test 2)	56,77	14,19	5,336
Sprachverfall? (C-Test 3)	59,37	14,84	5,354
Körpersprache (C-Test 4)	55,42	13,86	5,087
Sind Jungen die neuen Bildungsverlierer? (C-Test 5)	35,62	8,91	4,687
Warum Schule Angst macht (C-Test 6)	49,73	12,43	5,770
C-Test Gesamt	**53,23**	**79,85**	**25,524**

Die Schwierigkeitsindizes der in dieser Arbeit eingesetzten C-Tests liegen zwischen P=35,62 und P=62,47 und befinden sich damit in einem Bereich, der für die getesteten Oberstufenschüler*innen weder als zu leicht noch als zu schwierig einzustufen ist (Raatz & Klein-Braley, 1983, 107 ff.). Die Lösungswahrscheinlichkeiten der C-Tests (1, 2, 3, 4 und 6) variieren in engen Grenzen, hier zwischen etwa 50 % und 60 %. Einzig der C-Test 5 (*Sind Jungen die neuen Bildungsverlierer?)* wurde im Durchschnitt zu nur 35,62 % gelöst und ist also für die untersuchte Kohorte der schwierigste Test (P=35,62). Der C-Test 1 (*Der Reiz der Fernsehserien)* mit einer durchschnittlichen Lösungswahrscheinlichkeit von 62,46 % zeigt sich als für die Oberstufenschüler*innen am einfachsten zu lösen. Die Lösungswahrscheinlichkeit bzw. der Schwierigkeitsgrad des gesamten C-Test-Sets liegt mit etwa P=53,23 im mittleren Bereich.

Um nun die Eignung der C-Test-Daten für bestimmte bi- und multivariate Analyseverfahren einzuschätzen zu können – wie sie in Kapitel 5.4, 5.5 und 5.6 zur Überprüfung von Zusammenhängen und Unterschieden von Variablen zum Einsatz kommen –, ist es notwendig, neben der Schwierigkeit des C-Tests auch die Verteilung der C-Test-Ergebnisse zu betrachten:

Tabelle 11: Deskriptive Statistik der C-Test-Rohwerte

	Statistik	Standardfehler
Mittelwert	79,85	0,641
Median	82,00	
Varianz	651,481	
Standardabweichung	25,524	
Minimum	6	
Maximum	141	
Schiefe	-0,267	0,061
Kurtosis	-0,477	0,123

Die Werte zur Schiefe und Kurtosis zeigen generell an, ob die C-Test-Ergebnisse symmetrisch verteilt und im Vergleich zu einer Normalverteilung eher flach oder steil gewölbt sind (Bortz & Schuster, 2010, S. 42). Bei einer perfekten Normalverteilung nehmen die Schiefe und die Kurtosis den Wert null an. Für die hier zu untersuchenden C-Test-Daten liegen die Werte für die Schiefe und Kurtosis mit -0,267 und -0,477 unter null, was bedeutet, dass die C-Test-Daten im Vergleich zu einer Normalverteilung leicht linksschief (oder rechtssteil) und eher flach, d. h. breitgipflig verteilt sind (Bortz & Schuster, 2010, S. 42; Fahrmeir et al., 2016, 44 ff.). Die leichte Linksschiefe (bzw. Rechtssteile) der C-Test-Daten ist am ehesten damit zu erklären, dass im C-Test insgesamt 150 Punkte erzielt werden können und geringe Testwerte etwas weniger häufig vorkommen als Testwerte im mittleren bis großen Zahlenbereich (um den Mittelwert von 79,85 Punkten). Ebenso liegt der Median mit 82 C-Test-Punkten über dem Mittelwert von 79,85 Punkten, was bedeutet, dass mehr C-Test-Ergebnisse über dem Mittelwert liegen als darunter. Die breitgipflige Verteilung der Daten ergibt sich daraus, dass sich die Messwerte im Vergleich zur Normalverteilung mehr um den Mittelwert massieren und weniger stark in den besonders hohen oder niedrigen Punktebereichen, d. h. den Ausläuferbereichen, verteilt sind. Diese hier beschriebenen Abweichungen von einer Normalverteilung sind allerdings gering – die Werte -0,267 und -0,477 weichen nur geringfügig von null ab – und bei der graphischen Analyse zur Verteilung der C-Test-Ergebnisse im Histogramm optisch kaum sichtbar:

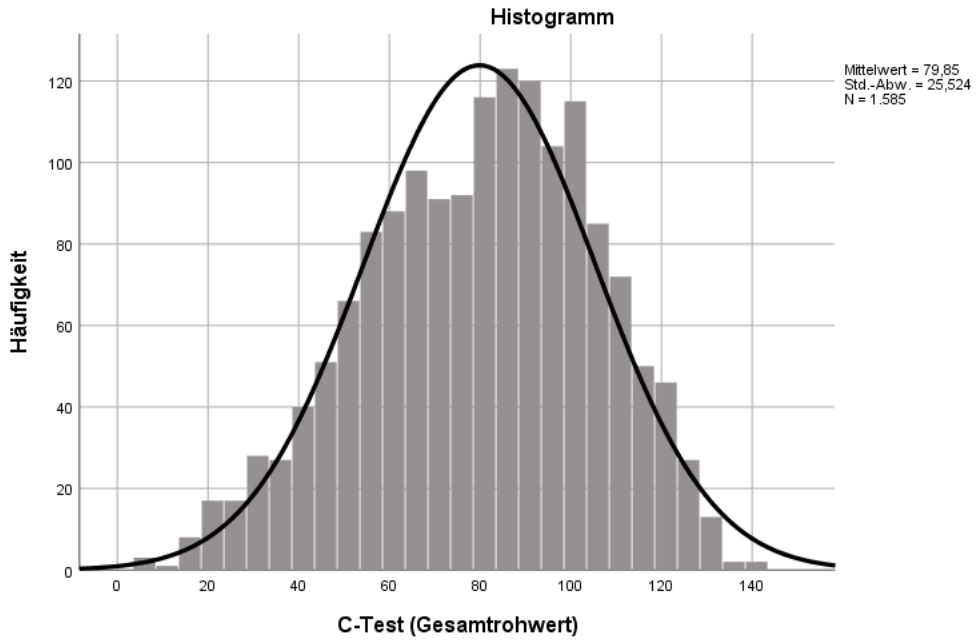


Abbildung 17: Histogramm der C-Test-Rohwerte

Die grafische Veranschaulichung der Verteilung der C-Test-Ergebnisse insinuiert damit eine annähernde, leicht nach rechts versetzte Normalverteilung. Für die hier vorliegenden C-Test-Daten (als abhängige Variable) werden für die Berechnungen von Korrelationen und Mittelwertunterschieden in Kapitel 5.4 und 5.5 Tests für parametrische Daten angewendet.

Die optische Auswertung der Verteilung der C-Test-Daten – ebenso wie die Analyse der Schwierigkeitsindizes (vgl. oben) – zeigt, dass der eingesetzte C-Test durch die nur minimale Abweichung von einer Normalverteilung in der Zielkohorte weder zu Decken- noch zu Bodeneffekten führt sondern eine breite Streuung der Ergebnisse hervorruft und damit eine ausreichende Differenzierung zwischen den Leistungen der Oberstufenschüler*innen ermöglicht.

5.3 Normierung und Reliabilität

Zur Interpretation der C-Test-Ergebnisse ist eine Normierung des C-Tests notwendig, da über eine Normierung Vergleichswerte zur Verfügung gestellt werden, die eine

Beurteilung und Ergebniseinordnung des Gesamtscores anhand anderer Ergebnisse ermöglichen. Für die Normierung der im Rahmen dieser Arbeit erhobenen Daten wurden die C-Test-Rohwerte gemäß der Item-Response-Theory rasch-skaliert (vgl. Kapitel 4.4). Dabei wird für den Gesamtscore, d. h. den C-Test-Rohwert, einer jeden Person über eine logische Transformation ein geschätzter Fähigkeitswert, ein sogenannter WLE-Parameter (Weighted Likelihood Estimation) bestimmt (vgl. Kapitel 4.4.3; Eckes 2007, S. 74; Mashkovskaya und Baur 2016b, 218 f.). Die WLE-Parameter können dabei Werte zwischen -3 und 3 annehmen, wobei negative WLE-Parameter auf eine eher geringe Personenfähigkeit schließen lassen. Für die Normierung der in dieser Studie eingesetzten C-Tests wurde dem gängigen Vorgehen in der C-Test-Praxis folgend (Baur & Spettmann, 2008a, 123 ff., 2008b, 430 ff.; Grotjahn, 2002b, 211 ff., 2019a, 585 ff.; Mashkovskaya & Baur, 2016b, 216 ff.) zunächst ein Referenznormwert bestimmt, der sich aus dem Mittelwert (der WLE-Parameter) der monolingual deutschsprachigen Proband*innen zusammensetzt (vgl. hierzu ausführlich Kapitel 4.4.3) (Baur et al., 2006, S. 394; Bremerich-Vos & Scholten-Akoun, 2016, S. 219; Scholten-Akoun & Baur, 2012). Für die untersuchte Kohorte ergaben sich ein Referenz- bzw. Mittelwert (M) von 0,0869 (das entspricht einem C-Test-Rohwert von ungefähr 87 Punkten) und eine Standardabweichung (SD) von 0,4170. Auf der Basis dieser beiden Werte der WLE-Parameter wurden dann die folgenden Leistungsbereiche definiert (vgl. Kapitel 4.4.3):

Tabelle 12: Zuordnung zu den Leistungsbereichen

Leistungs-bereiche	Normwerte	WLE-Intervalle	Rohwerte
I	> M + 2 x SD	0,9495 – 1,5806	129 - 150
II	(M+SD) < WLE < (M + 2 x SD)	0,5043 – 0,9139	111 - 128
III	(M - SD) < WLE < (M + SD)	- 0,3215 – 0,4848	62 - 110
IV	(M – 2 x SD) < WLE < (M - SD)	- 0,7346 – -0,3376	37 - 61
V	< (M – 2 x SD)	- 1,5084 - -0,7520	0 - 36

In die Leistungsgruppen I und II fallen die C-Test-Ergebnisse, die mehr als eine Standardabweichung über dem Referenznormwert liegen. Proband*innen, deren

Ergebnisse in den Leistungsbereichen I und II liegen, wird sowohl im schulischen (Baur et al., 2006; Baur et al., 2012; Baur et al., 2013b, S. 10) als auch im akademischen Bereich (Mashkovskaya, 2014; Mashkovskaya & Baur, 2016b, S. 220; Scholten-Akoun & Baur, 2012) kein sprachlicher Förderbedarf unterstellt (vgl. Kapitel 4.4.3). Ergebnisse die eine Standardabweichung über oder unter dem Referenznormwert und somit im mittleren Leistungsbereich (Leistungsgruppe III) liegen, weisen keinen eindeutig erkennbaren sprachlichen Förderbedarf aus (ebd.). Von wahrscheinlich deutlichem sprachlichen Förderbedarf wird ausgegangen, wenn das C-Test-Ergebnis den Referenznormwert um mehr als eine Standardabweichung unterschreitet (Leistungsgruppe IV), von sehr wahrscheinlich erheblichem Förderbedarf, wenn die C-Test-Leistung mehr als zwei Standardabweichungen unter dem Referenznormwert (Leistungsgruppe V) zu verorten ist (ebd.) (vgl. Kapitel 4.4.3).

Über die Einteilung der Test-Ergebnisse in Leistungsgruppen ergibt sich für die in dieser Arbeit untersuchte Kohorte von Oberstufenschüler*innen folgendes Bild:

Tabelle 13: Verteilung der Stichprobe auf die Leistungsbereiche

Leistungsbereiche	Häufigkeiten (N)	Prozent
I	17	1,1
II	162	10,2
III	1009	63,7
IV	303	19,1
V	94	5,9
Gesamt	**1585**	**100**

Von allen getesteten Oberstufenschüler*innen erzielen insgesamt 11,3 % C-Test-Ergebnisse, die in die Leistungsbereiche I und II fallen, denen also kein Förderbedarf zugeschrieben wird. Der größte Teil der Ergebnisse (63,7 %) liegt im Leistungsbereich III, d. h. innerhalb einer Standardabweichung über bzw. unter dem Referenznormwert. Ein eindeutig erkennbarer Förderbedarf lässt sich nicht erkennen. Bei etwa einem Viertel (25 %) der Schüler*innen in der gymnasialen Oberstufe ist das allerdings anders. Die Testergebnisse dieses Viertels lassen sich den Leistungsbereichen IV (19,1 %) und V

(5,9 %) zuordnen und signalisieren teils massive Förderbedarfe auch im Bereich der basalen schriftsprachlichen Fähigkeiten und Fertigkeiten (u. a. in der Lexik, Semantik, Grammatik, Orthografie und im Leseverstehen; vgl. Kapitel 4.4.3).

Die Reliabilität ist eins der zentralen Gütekriterien und gibt an, ob ein Test reliabel, d. h. zuverlässig misst (vgl. Kapitel 2.4). Für die vorliegende Stichprobe wurde die Zuverlässigkeit bzw. Messgenauigkeit der C-Tests über die Konstruktreliabilität und WLE-Reliabilität (Weighted-Likelihood-Estimates) (vgl. Kapitel 4.4.2 und 4.4.3) ermittelt. Anhand der Konstruktreliabilität wird beurteilt, wie gut ein Konstrukt durch alle ihm zugeordneten Indikatoren gemeinsam gemessen wird (Hornburg & Giering, 1996, S. 10). Die Werte beider Reliabilitäten liegen standardmäßig zwischen 0 und 1 und können im Hinblick auf die Größenordnung mit Cronbachs Alpha aus der klassischen Testtheorie verglichen werden (Rost, 2004; Zinn et al., 2017, S. 84). Prinzipiell gilt, dass je höher die einem Konstrukt zugeordneten Indikatoren miteinander korrelieren, desto valider misst der Test. Die Konstruktreliabilität (geschätzt über die EAP-Reliabilität) für die in dieser Studie eingesetzten C-Tests beträgt 0,92 und ist damit zweifellos zufriedenstellend. Die WLE-Reliabilität ist mit 0,907 nur geringfügig kleiner, aber ebenfalls hoch.

5.4 Zur Validität des C-Tests

Neben der Reliabilität und Objektivität gilt die Validität als wichtigstes Gütekriterium eines C-Tests (vgl. Kapitel 2.4). Bei der Überprüfung der Validität eines C-Tests geht es um die Frage, inwieweit die Testergebnisse das erfassen, was sie erfassen sollen (Grotjahn, 2019a, S. 602). Zur Überprüfung der Validität von C-Tests werden besonders häufig Korrelationen berechnet, die zeigen, dass C-Tests oftmals im mittleren und hohen Bereich mit verschiedenen Außenkriterien (z. B. Ergebnissen in anderen Sprachtests oder Schulnoten; vgl. dazu Kapitel 2.4) korrelieren (Eckes & Grotjahn, 2006, 209 ff.; Grotjahn, 2019a, S. 602; Wockenfuß & Raatz, 2006, 211 ff.). Die Validität der im Rahmen dieser Studie entwickelten C-Tests soll nun im Folgenden mit dem Lesetest LGVT 6-12 (vgl. Kapitel 4.1.1) sowie mit den im Fragebogen erfassten Deutschnoten der teilnehmenden Oberstufenschüler*innen (vgl. Kapitel 4.1.3) überprüft werden.

5.4.1 C-Test und LGVT 6-12

<u>Hypothese 1</u>: Es gibt einen statistisch signifikanten Zusammenhang zwischen dem in der Oberstufe eingesetzten, linksseitig getilgten C-Test und dem standardisierten Lesetest LGVT 6-12.

In den Ausführungen zur Erfassung von Sprachkompetenz mit dem C-Test in dieser Arbeit (vgl. Kapitel 2.3) wurde bereits theoretisch beschreiben, dass zum Lösen von C-Tests u. a. bestimmte Lesefähigkeiten und -fertigkeiten abgerufen werden müssen (Baur & Spettmann, 2007, 95 ff.; Mashkovskaya, 2014, S. 241). Die hier formulierte Hypothese bezieht sich auf diese Erklärung und basiert damit auf der Annahme, dass Proband*innen, die in einem Lesetest ein besonders gutes Ergebnis erzielen, im C-Test – für dessen erfolgreiche Rekonstruktion bestimmte Aspekte von Lesekompetenz also notwendig sind – eine ähnlich gute Leistung zeigen.

Mashkovskaya gelang es bereits 2014 in ihrer Untersuchung zur Sprachkompetenz von Lehramtsstudierenden, eine der Zielkohorte dieser Untersuchung ähnliche Probandengruppe, einen Zusammenhang zwischen den Leistungen in linksseitig getilgten C-Tests und Facetten von Lesekompetenz und Textverständnis von fortgeschrittenen Erst- und Zweitsprachsprecher*innen zu belegen: Die Autorin ermittelt im Rahmen ihrer Studie sowohl mittlere bis hohe Zusammenhänge zwischen den von ihr eingesetzten C-Tests und einem modifizierten Stolperwörterlesetest (r=0,82 und r=0,718) als auch zwischen den C-Tests und einem kombinierten Lesetest, bestehend aus Multiple-Choice- und Zuordnungsaufgaben (r=0,64) (Mashkovskaya, 2014, 183, 193, 241). Diese Korrelationen legen nahe, dass die C-Test-Ergebnisse als Prädiktoren für die Einschätzung der Leseleistung von fortgeschrittenen Erst- und Zweitsprachsprecher*innen angesehen werden können und der eingesetzte linksseitig getilgte C-Test damit eine mittlere bis hohe prädiktive Validität in Bezug auf die Erfassung von Lesekompetenz aufweist (Mashkovskaya, 2014, S. 196).

In der hier vorgelegten Untersuchung wird nun zur Überprüfung der Validität ein analoger Weg verfolgt, indem bestimmt werden soll, ob zwischen den im Rahmen dieser Studie entwickelten linksseitig getilgten C-Tests für die Sekundarstufe II und dem Lesetest LGVT 6-12 (vgl. Kapitel 4.1.1) eine statistisch signifikante Korrelation besteht.

Den LGVT 6-12 als Außenkriterium heranzuziehen hat den Vorteil, dass der LGVT 6-12 ein bereits erprobter und normierter Lesetest ist, der sich für den Einsatz in der gymnasialen Oberstufe eignet und der die Anforderungen eines wissenschaftlichen Tests – d. h. die Testgütekriterien – erfüllt (vgl. Kapitel 4.1.1). Eine (mittlere bis) hohe Korrelation zwischen dem C-Test und dem LGVT 6-12 würde darauf hinweisen, dass beide Testverfahren Lesekompetenz auf ähnliche Weise messen bzw. erfassen, ein schwacher Zusammenhang hingegen, könnte darauf hindeuten, dass die entwickelten C-Tests und der LGVT 6-12 unterschiedliche Facetten von Lesekompetenz überprüfen (Mashkovskaya, 2014, S. 158).

Wie in Kapitel 4.1.1 beschrieben, werden bei der Auswertung des LGVT 6-12 zwei Werte – zum einen für das Leseverstehen und zum andern für die Lesegeschwindigkeit – erfasst (Schneider et al., 2007, 12 ff.). Um nun statistisch zu überprüfen, inwieweit das C-Test-Ergebnis mit den mit dem LGVT 6-12 gemessenen Werten Leseverstehen und Lesegeschwindigkeit zusammenhängt, gilt es in Abhängigkeit des Skalenniveaus der abhängigen und unabhängigen Variable als auch in Bezug auf die Verteilung der Daten ein Verfahren zur Bestimmung des Korrelationskoeffizienten auszuwählen: Da es sich bei dem C-Test-Ergebnis als auch bei den Variablen *Leseverstehen* und *Lesegeschwindigkeit* um metrische Variablen handelt und die Daten normalverteilt sind, wird als Zusammenhangsmaß der Pearson-Korrelationskoeffizient genutzt: Beide Werte bzw. Lesetestvariablen wurden mit dem C-Test korreliert:

Tabelle 14: Korrelationsanalysen zwischen C-Test und Lesetest

		Lesegeschwindigkeit	**Leseverständnis**
Lesegeschwindigkeit	Korrelationskoeffizient	-	,548**
	Sig. (2-seitig)	-	,000
	N	-	1153
Leseverständnis	Korrelationskoeffizient	,548**	-
	Sig. (2-seitig)	,000	-
	N	1153	-

	Korrelationskoeffizient	,206**	,400**
C-Test (Gesamtrohwert)	Sig. (2-seitig)	,000	,000
	N	1153	1153

** Die Korrelation ist auf dem 0,01 Niveau signifikant (zweiseitig).

Die Korrelation zwischen dem C-Test und dem Aufgabenteil Leseverstehen des LGVT 6-12 liegt mit einem Korrelationskoeffizient von r = 0,40 im mittleren Bereich. Schwächer fällt hingegen der statistische Zusammenhang zwischen dem C-Test und der mit dem LGVT 6-12 gemessenen Lesegeschwindigkeit aus, hier zeigen die Daten eine eher geringe Korrelation von r = 0,21. Innerhalb des Textkonstrukts des LGVT 6-12 liegt die Korrelation zwischen dem Leseverstehen und der Lesegeschwindigkeit mit r = 0,55 im mittleren bis hohen Bereich und damit höher als zwischen dem C-Test und der Lesegeschwindigkeit.

Die Hypothese, dass es einen statistisch signifikanten Zusammenhang zwischen dem in der Oberstufe eingesetzten, linksseitig getilgten C-Test und dem standardisierten Lesetest LGVT 6-12 gibt, wird bestätigt.

5.4.2 C-Test und Deutschnote

<u>Hypothese 2</u>: Es gibt einen mittleren, statistisch signifikanten Zusammenhang zwischen dem eingesetzten C-Test und der Deutschnote.

Zur Überprüfung der Konstruktvalidität von C-Tests gibt es eine Reihe von Untersuchungen, in denen der Zusammenhang zwischen C-Test-Leistung und den Noten im Deutschunterricht (oder dem Unterricht in der Erstsprache) bzw. den Noten im Fremdsprachunterricht untersucht werden.

So zeigen die Untersuchungen von Klein-Braley und Raatz (1984) etwa, dass für Schüler*innen der unteren Klassenstufen in der Regel ein zum Teil großer Zusammenhang zwischen Deutschnoten und C-Test-Leistungen besteht (S. 145) und auch Baur und Spettmann (2007, 2008a) berichten von teilweise mittleren Korrelationen zwischen muttersprachlichen C-Tests und Deutschnoten in den Klassenstufen 4-6.

Kesper (1995) und Gummich (1997), die den Zusammenhang von älteren Schüler*innen zwischen Deutschnote und C-Test-Leistung systematisch untersucht haben, kommen zu dem Schluss, dass auch bei älteren Schüler*innen ein Zusammenhang zwischen Deutschnote und C-Test-Leistung besteht, dieser jedoch, je höher die Klassenstufe, abnimmt und für die Gruppe der Studierenden wesentlich niedriger liegt als für jüngere Schüler*innen (zitiert nach Wockenfuß 2009, S. 137)[101]. Die Autoren gehen davon aus, dass zu Beginn der Sekundarstufe I mit erstsprachlichen C-Tests vor allem unterrichtsrelevante sprachliche Fähigkeiten gemessen werden, die stärker in die Benotung durch die Lehrkräfte miteinfließen und die Beurteilung dieser sprachlichen Fähigkeiten im Verlauf der Sekundarstufe I immer weiter abnimmt (Wockenfuß, 2009, S. 181). Dieses Ergebnis bestätigt auch Wockenfuß (2009) in ihrer Untersuchung zur Entwicklung der allgemeinen Sprachkompetenz in der Erstsprache von Gymnasialschüler*innen und Studierenden. Bemerkenswert ist hier jedoch die Einschränkung, dass der Zusammenhang zwischen Deutschnote und C-Test-Leistung bei sich in der Abiturphase befindenden Proband*innen wieder stärker wird (r = 0,40) (Wockenfuß, 2009, 174 ff.). Wockenfuß (2009) nimmt in diesem Zusammenhang an, dass besonders in der Abiturzeit der mit dem C-Test gemessene allgemeine Sprachkompetenz in der Benotung wieder eine bedeutendere Rolle zukommt und „der benotete Leistungsbereich im Fach Deutsch breiter gefächert ist, um […] zum Schulabschluss (Abitur) einen möglichst großen Überblick der sprachlichen Fähigkeiten der Schüler zu haben bzw. zu geben“ (Wockenfuß, 2009, S. 182).

Der Erklärungsansatz von Wockenfuß (2009, S. 182), dass die Korrelation zwischen Deutschnote und C-Test-Leistung in der Oberstufe wieder zunimmt, da besonders in der Abiturphase die „allgemeine Sprachbeherrschung“ verstärkt in die Deutschnote einfließt (Wockenfuß, 2009, 181 f.), scheint jedoch zumindest mit Blick auf die Beurteilungsempfehlungen für die schriftliche Abiturprüfung im Fach Deutsch (MSB NRW, 2017, S. 15) fraglich: So sind es im Abitur im Fach Deutsch gerade die basalen Elemente der allgemeinen Sprachbeherrschung (wie etwa auf einem grundlegenden

[101] Wockenfuß (2009) verweist in ihrer Dissertationsschrift auf die Forschungsergebnisse zu Korrelationsuntersuchungen zwischen Deutschnoten und C-Testleistungen von Kesper (1995) und Gummich (1997). Bei den Studien von Kesper und Grummich handelt es sich um unveröffentlichte Diplomarbeiten.

Niveau normgerecht, formalsprachlich und lexikalisch angemessen zu schreiben, usw.), die bei der Beurteilung und Benotung der schriftlichen Abiturleistung so gut wie keine Rolle spielen (vgl. Kapitel 2.2). Genau dieser Nicht-Zusammenhang zwischen verfügbaren basalen Deutschkenntnissen und dem C-Test-Ergebnis bildet sich auch in den Befunden von Mashkovskaya ab, die in ihrer Untersuchung zur Sprachkompetenz von Lehramtsstudierenden nur eine sehr schwache, auch nicht signifikante, Korrelation von r = - 0,028 zwischen den C-Test-Leistungen und Deutschzeugnisnoten der Studierenden ermittelt (Mashkovskaya, 2014, S. 228).

Ausgehend von der dieser Arbeit zugrunde liegenden Definition von Sprachkompetenz bzw. dessen operationalisierte Erfassung über den C-Test und im Bereich der Oberstufe (vgl. Kapitel 2.2 und Kapitel 2.3), wird theoretisch angenommen, dass der C-Test als integrativer Sprachtest allgemeine Sprachkompetenz und in der hier eingesetzten Version mit linksseitiger Tilgung vor allem semantisch-lexikalische und morpho-syntaktische Fähigkeiten misst (vgl. Kapitel 2.3.5). Die Deutschnote hingegen bildet – wie in Kapitel 2.2 ausführlich dargestellt – eine eher kumulative Summe aus vielen verschiedenen Fähigkeiten und Fertigkeiten – die schriftlich, mündlich und praktisch in Form von Klausuren, Präsentationen, Referaten, Vorträgen, Protokollen, Gesprächsbeiträgen im Unterricht, Portfolios, Recherchearbeiten etc. dargeboten werden, in der aber beispielswiese Aspekte der formal- und normsprachlichen Richtigkeit eher unterbelichtet bleiben – in die die mit dem C-Test erfasste Sprachkompetenz möglicherweise als nur eine Facette eines komplexen Kompetenzfächers miteinfließt. Vor diesem theoretischen Hintergrund wird also angenommen, dass es zwar eine Schnittmenge von schriftsprachlichen Kompetenzen gibt, die mit dem C-Test gemessen und in der Deutschnote (mit-)erfasst werden, die sich rechnerisch in einer mittleren Korrelation zwischen C-Test und Deutschnote zeigt – aber eben äußerstenfalls eine mittlere.

Für die Validitätsanalyse stehen insgesamt N = 1526 gültige Daten zur Verfügung. 131 Personen geben ihre Deutschnote mit 1 (*sehr gut*) an, 523 Personen mit 2 (*gut*), 635 Oberstufenschüler*innen haben eine 3 (*befriedigend*) als Deutschnote, 223 Personen die Note 4 (*ausreichend*) und 13 Personen die Note 5 (*mangelhaft*). Nur eine Person hat die Note 6 (*ungenügend*) als schlechteste Zeugnisnote im Fach Deutsch. Der unten

abgebildete Boxplot zeigt zunächst die Verteilung der Noten in Bezug auf den C-Testscore in Punkten sowie die Streuung innerhalb der einzelnen Notengruppen.

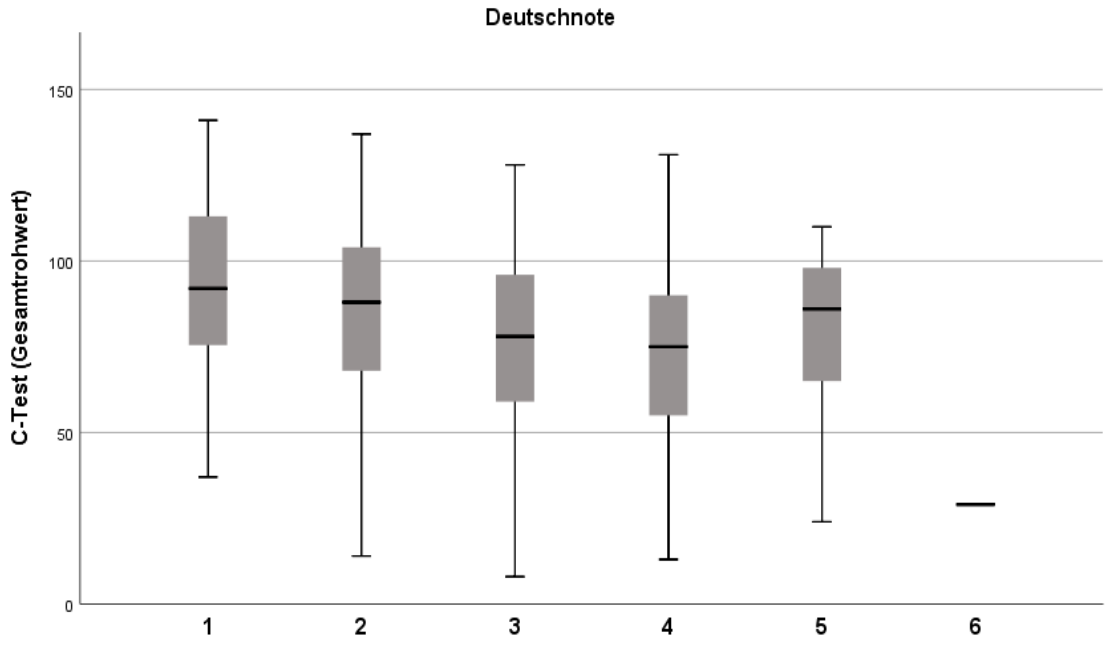


Abbildung 18: Boxplot C-Test und Deutschnote

Vergleicht man die Mediane der einzelnen Gruppen, so fällt auf, dass diese mit zunehmender, d. h. schlechter werdender Note sinken. Während der Median in der Gruppe der Oberstufenschüler*innen mit der Zeugnisnote 1 bei 93 Punkten (Mittelwert: 91) liegt, fällt er für die Notengruppe 2 auf 88 Punkte (Mittelwert: 86), für die Notengruppe 3 auf 78 Punkte (Mittelwert: 77) und in der Notengruppe 4 auf 75 Punkte (Mittelwert: 72). Für die Gruppe der Oberstufenschüler*innen, die im Fach Deutsch eine 5 haben, ist der Median bei 86 Punkten (Mittelwert: 78) zu verorten und liegt damit über den Medianen und Mittelwerten der Notengruppen 3 und 4, wobei sich dieses Leistungsbild allerdings nur auf eine sehr kleine Fallzahl (N = 13 für die Notengruppe 5) bezieht. Wie in dem Boxplot zu sehen, ist das Leistungsspektrum innerhalb der Notengruppen breit gestreut. Auffällig ist hierbei beispielsweise, dass die schwächste Leistung im C-Test mit insgesamt 8 Punkten in der Notengruppe 3 erzielt wurde und die beste C-Test-Leistung innerhalb der Notengruppe 4 (131 Punkte) über der besten C-

Test-Leistung in der Notengruppe 3 (128 Punkte) liegt. Die Maximalwerte in den Notengruppen 1 und 2 liegen dicht beieinander (141 Punkte und 137 Punkte).

Ein statistischer Vergleich der Mittelwerte der einzelnen Gruppen (über eine einfaktorielle Varianzanalyse, vgl. Anhang) zeigt, dass sich lediglich die Leistungen der Schüler*innen mit der Deutschnote 1 und 2 signifikant von den Leistungen der Schüler*innen mit den Deutschnoten 3 und 4 unterschieden. Alle anderen Leistungsunterschiede zwischen den Gruppen statistisch nicht signifikant (vgl. Anhang).

In Bezug auf die Leistungsbereiche, die den C-Test-Ergebnissen zugeordnet werden (vgl. Kapitel 5.3) sind die Deutschnoten ähnlich wie in der Darstellung der Gesamtrohwerte (siehe Boxplot) verteilt:

Tabelle 15: Verteilung der Variable *Deutschnote* auf die Leistungsbereiche

Note	LB I	LB II	LB III	LB IV	LB V	Gesamt
1	5 3,8 %	32 24,4 %	79 60,3 %	15 11,5 %	-	131 100 %
2	11 2,1 %	79 15,1 %	333 63,7 %	80 15,3 %	20 3,8 %	523 100 %
3	-	41 6,5 %	414 65,2 %	142 22,4 %	38 6,0 %	635 100 %
4	1 0,4 %	9 4,0 %	140 62,8 %	46 20,6 %	27 12,1 %	223 100 %
5	-	-	10 76,9 %	1 7,7 %	2 15,4 %	13 100 %
6	-	-	-	-	1 100 %	1 100 %
Gesamt	17 1,1 %	161 10,6 %	976 64,0 %	284 18,6 %	88 5,8 %	1526 100 %

So erzielen zwar beispielsweise die Schüler*innen mit der Deutschnote 1 und 2 am häufigsten C-Test-Ergebnisse, die den Leistungsbereichen I und II zuzuordnen sind (28,2 % und 17,2 %), allerdings gibt es auch Schüler*innen mit der Deutschnote 1, deren Ergebnisse in den Leistungsbereich IV (11,5 %) fallen. Von den Schüler*innen mit der Note 2 münden sogar mehr Personen in die Leistungsbereiche IV und V (19,1 %), denen ein erkennbarer bzw. erheblicher sprachlicher Förderbedarf zugeschrieben wird, als in die oberen Leistungsbereiche. Andersherum gibt es Schüler*innen mit der Note 4, die C-Test-Ergebnisse in den Leistungsbereichen I und II (4,4 %) erzielen, d. h. denen der C-Test keinen Förderbedarf im schriftsprachlichen Bereich attestiert. Eine große Schnittmenge bildet der Leistungsbereich III, in dem über alle Notenstufen hinweg (ausgenommen der Note 6) über 60 % der C-Test-Leistungen zu verorten sind.

Diese heterogenen, deskriptiven Ergebnisse, d. h. vor allem die breite Streuung der C-Test-Ergebnisse auf die einzelnen Notenstufen, sind wohl am ehesten damit zu erklären, dass die Notenverteilung von klasseninternen Bezugsmaßstäben abhängt (Wockenfuß, 2009, S. 173). So ist davon auszugehen, dass sich die Verteilung der (Zeugnis-)noten innerhalb einer Klasse – neben möglicherweise kriterialen Normen, wie beispielsweise über Erwartungshorizonte in Klassenarbeiten – auch am jeweiligen Leistungsspektrum der Schüler*innen innerhalb einer einzelnen Klasse, d. h. hinsichtlich der jeweiligen Bezugsgruppe orientiert. Diese klasseninterne Bezugsnorm führt dann dazu, dass die Noten innerhalb einer Klasse in der Regel annähernd normalverteilt sind, d. h. es gibt weniger häufig sehr gute bzw. sehr schlechte Noten, dafür häufiger Noten im mittleren Leistungsbereich. Wenn sich die Verteilung von Noten jedoch (auch) auf klasseninterne Maßstäbe bezieht, dann bedeutet dies auch, dass beispielsweise der Note 4 in der einen Klasse vermutlich ein ganz anderer Maßstab zugrunde liegt als einer 4 in einer anderen Klasse. Die Note selbst bildet damit also keinen objektiven Maßstab, der über alle Klassen und Schulen hinweg gilt, wodurch sich die Noten auch nicht einfach über mehrere Klassen bzw. Schulen hinweg vergleichen lassen. Sollen die Zeugnisnoten nun aber mit den C-Test-Ergebnissen korreliert werden, muss also die Problematik der klasseninternen Bezugsmaßstäbe bedacht werden. Ein Verfahren, dass dieses Problem lösen kann, ist die sogenannte *Fisher-Z-Transformation* zur Mittelung von Korrelationen: Bei der Fisher-Z-Transformation werden zunächst die Korrelationskoeffizienten für jede

Klasse einzeln ausgegeben, d. h. der Zusammenhang zwischen C-Test-Ergebnis und Deutschnote wird für jede einzelnen Klasse bzw. jeden einzelnen Deutschkurs berechnet (vgl. Anhang). Die Korrelationskoeffizienten der einzelnen Klassen werden anschließend fisher-z-transformiert und anhand der jeweiligen Klassen- oder Kursgröße bzw. Fallzahl gewichtet (Lenhard & Lenhard, 2014). Nach der Gewichtung lässt sich ein durchschnittlicher – nun klassen- und schulübergreifender – fisher-z-transformierter Korrelationskoeffizient berechnen, welcher anschließend wieder zurück transformiert wird (Wockenfuß, 2009, S. 174; Wockenfuß & Raatz, 2006, S. 226). Die so ermittelte Übereinstimmungsvalidität (bzw. Produkt-Moment-Korrelation) zwischen den C-Tests und den Deutschnoten beträgt in der hier vorliegenden Studie:

Tabelle 16: Korrelationsanalyse C-Test und Deutschnote

		Deutschnote
C-Test-Ergebnis	Korrelationskoeffizient	- 0,32**
	Sig. (2-seitig)	,000
	N	1526[102]

** Die Korrelation ist auf dem 0,01 Niveau signifikant (zweiseitig).

Diese mittlere Korrelation von r = - 0,32 ist dabei hochsignifikant. Das negative Vorzeichen der Korrelation ist in diesem Fall mit dem Notensystem zu erklären, d. h. ein gutes C-Test-Ergebnis, also ein hoher C-Testpunktwert, entspricht hier einer numerisch niedrigen Deutschnote und umgekehrt ein schlechtes C-Test-Ergebnis einer numerisch hohen Note im Fach Deutsch (Wockenfuß, 2009, S. 174).

Die Hypothese, dass es einen mittleren, statistisch signifikanten Zusammenhang zwischen dem eingesetzten C-Test und der Deutschnote gibt, wird bestätigt.

[102] Die unterschiedlichen Fallzahlen der deskriptiven und bivariaten Ergebnisse sind darauf zurückzuführen, dass einige C-Test-Ergebnisse für die fisher-z-transformation keiner Klasse bzw. keinem Kurs zugeordnet und somit nicht in der Berechnung der Korrelation berücksichtigt werden konnten.

5.5 Deskriptive und bivariate Ergebnisse

In diesem Ergebnisteil werden zur ersten Überprüfung der in Kapitel 3 formulierten Hypothesen vorerst bivariate Zusammenhänge zwischen der C-Test-Leistung und jeweils einer weiteren Variablen in den Blick genommen. Dabei erfolgt zunächst eine deskriptive Auswertung der C-Test-Leistung in Bezug auf die jeweiligen Merkmalsausprägungen der Variable (z. B. in Bezug auf Geschlecht, Bildungsniveau der Eltern, Standorttyp, Erstsprache, etc.; vgl. Kapitel 4.1.3), ehe im Anschluss – mittels unterschiedlicher Verfahren (siehe Überblick) – geprüft wird, ob sich die deskriptiv beobachteten Verteilungen statistisch signifikant voneinander unterscheiden bzw. inwieweit insgesamt ein statistisch signifikanter Zusammenhang zwischen der jeweiligen Variable und dem C-Test-Ergebnis besteht.

Tabelle 17: Überblick über die angewendeten Verfahren für Mittelwertvergleiche und bivariate Korrelationsanalysen

Abhängige Variable	Unabhängige Variable	Verfahren
C-Test-Ergebnis (metrisch)	**Geschlecht** (nominal) 2 Gruppen (unabhängig)	t-Test
	Phase in der Oberstufe (nominal) 2 Gruppen (unabhängig)	t-Test
	Bildungsniveau (ordinal) Mehr als 2 Gruppen (unabhängig)	Einfaktorielle Varianzanalyse, Spearman
	Mehrsprachigkeit (nominal) 2 Gruppen (unabhängig) **Erstsprache Deutsch** (nominal) 2 Gruppen (unabhängig) **Sprachverwendung** (nominal) 2 Gruppen (unabhängig)	t-Test
	Sprachgruppen (nominal) Mehr als 2 Gruppen (unabhängig)	Einfaktorielle Varianzanalyse, Spearman
	Standorttyp (ordinal) Mehr als 2 Gruppen (unabhängig)	Einfaktorielle Varianzanalyse, Spearman
	Schulform (nominal) 2 Gruppen (unabhängig)	t-Test

Welches Verfahren für die Überprüfung welcher Hypothese angewendet werden kann, hängt dabei in Bezug auf die Daten u. a. von ihrer Verteilung, dem Skalenniveau der

abhängigen Variablen (dem C-Test-Ergebnis), dem Skalenniveau der unabhängigen Variablen sowie der Ausprägung der Merkmalsverteilung (Anzahl, Abhängigkeit) ab und wird im Folgenden für jede Hypothese einzeln berichtet.

5.5.1 C-Test und Phase in der Oberstufe

<u>Hypothese 3</u>: Schüler*innen in der Qualifikationsphase 1 erzielen signifikant höhere Leistungen im C-Test als Schüler*innen in der Einführungsphase.

Dass die Entwicklung von Sprache bzw. basalen sprachlichen Kompetenzen mit dem Eintritt in die gymnasiale Oberstufe nicht vollständig abgeschlossen sein kann, wurde in Kapitel 1 theoretisch diskutiert. Untersuchungen wie beispielsweise die von Wockenfuß (2009) bestätigen in diesem Zusammenhang auch empirisch, dass sich die Sprachkompetenz im Laufe der Schulzeit, d. h. auch im Bereich der Oberstufe (und darüber hinaus) durchaus weiterentwickelt: So konnte Wockenfuß in ihrer Studie u. a. zeigen, dass die mit dem von ihr eingesetzten C-Test gemessene globale Sprachfähigkeit von Jugendlichen und jungen Erwachsenen bis zum 25. Lebensjahr[103] proportional ansteigt (Wockenfuß, 2009, S. 199). Vor dem Hintergrund dieser Befunde und bedingt durch die Annahme, dass schulischer Unterricht in der Sekundarstufe II einen grundlegenden Beitrag zur Kompetenzentwicklung auf dem Weg zur allgemeinen Hochschulreife leistet und im Besonderen Schüler*innen im Deutschunterricht der gymnasialen Oberstufe vertiefende rezeptive und produktive sprachliche Kompetenzen vermittelt werden (sollen) (MSB NRW, 2014, 11 ff.), wird im Folgenden überprüft, ob sich die C-Test-Leistungen der Schüler*innen in der Einführungsphase von den Leistungen der Schüler*innen in der Qualifikationsphase 1 unterscheiden. Angenommen wird dabei, dass die Schüler*innen der höheren Klassenstufe, d. h. der Qualifikationsphase 1, durchschnittlich höhere C-Test-Ergebnisse erzielen als die Schüler*innen in der darunter liegenden Stufe in der Einführungsphase.

[103] Nach dem 25. Lebensjahr nehmen die C-Test-Leistungen allerdings wieder ab (Wockenfuß 2009, S. 199).

Für die Beantwortung dieser Frage wurden Daten analysiert, die sich auf N = 1.577 gültige Fälle zur Stufe/Phase in der Oberstufe beziehen. Insgesamt gaben 1175 Testpersonen an, die Einführungsphase zu besuchen, 402 Testpersonen die Qualifikationsphase 1. Die Abb. 19 zeigt die C-Test-Leistungen und ihre Streuung für beide Phasen in der Oberstufe:

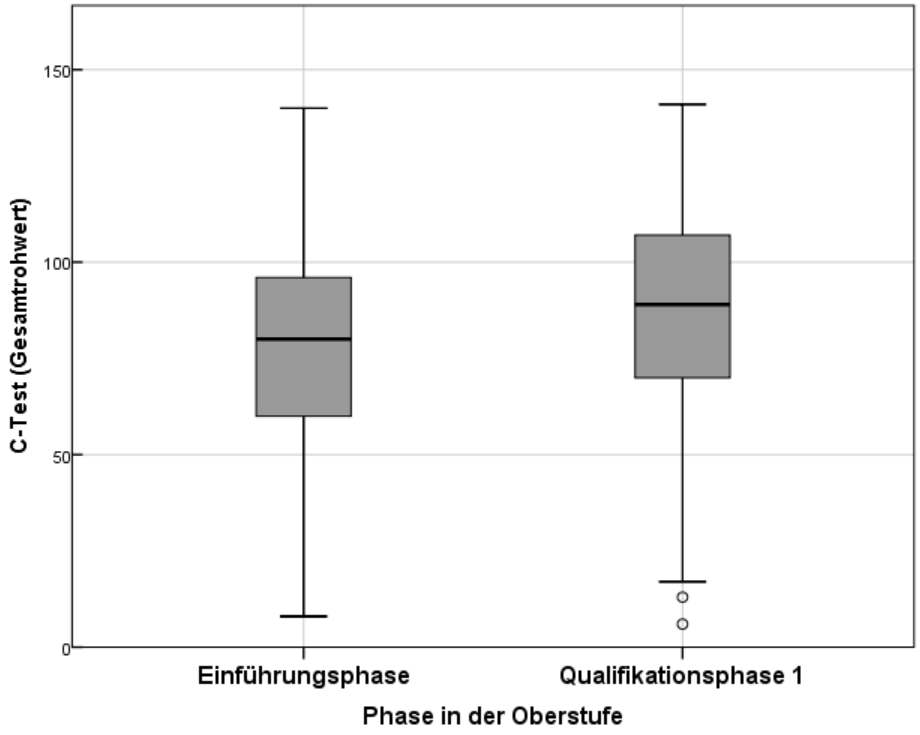


Abbildung 19: Boxplot C-Test-Ergebnis und Phase in der Oberstufe

Der Vergleich zeigt für beide Schülergruppen eine etwa gleich breite Streuung der C-Test-Ergebnisse. Betrachtet und vergleicht man die Mediane beider Gruppen miteinander, ist allerdings ein Leistungsanstieg in der Gruppe der Schüler*innen der Qualifikationsphase 1 zu erkennen: So liegt der Median in dieser Gruppe mit 89 Punkten 9 Punkte über dem Median der Gruppe der Schüler*innen in der Einführungsphase (80 Punkte).

Ein ähnliches Bild zeigt sich in der Verteilung auf die C-Test-Leistungsbereiche:

Tabelle 18: Verteilung der Variable *Phase in der Oberstufe* auf die Leistungsbereiche

Phase in der Oberstufe	LB I	LB II	LB III	LB IV	LB V	Gesamt

Einführungsphase	9	90	751	250	75	1175
	0,8 %	7,7 %	63,9 %	21,3 %	6,4 %	100 %
Qualifikationsphase 1	8	71	255	51	17	402
	2 %	17,7 %	63,4 %	12,7 %	4,2 %	100 %
Gesamt	17	161	1006	301	92	1577
	1,1 %	10,2 %	63,8 %	19,1 %	5,8 %	100 %

Betrachtet man die Verteilung der Ergebnisse, so liegen 19,7 % der Leistungen der Schüler*innen in der Qualifikationsphase 1 und nur 8,5 % der Leistungen der Schüler*innen in der Einführungsphase in den oberen Leistungsbereichen (LB I und LB II). In den Leistungsbereichen, denen Förderbedarf zugeschrieben wird (LB IV und LB V), sind mit 27,7 % fast doppelt so viele Schüler*innen der Einführungsphase vertreten im Vergleich zu den Schüler*innen der Qualifikationsphase 1 mit 16,9 %.

Um nun statistisch zu überprüfen, ob sich die zunächst deskriptiv dargestellten Leistungen der Oberstufenschüler*innen auch statistisch signifikant von den Leistungen der Oberstufenschüler unterscheiden, werden die Mittelwerte beider Gruppen mittels t-Test (vgl. Tabelle 20) miteinander verglichen:

Tabelle 19: Gruppenstatistik: C-Test-Ergebnis, Phase in der Oberstufe

Phase	N	Mittelwert	Standardabweichung
Einführungsphase	1175	77,26	24,896
Qualifizierungsphase 1	402	87,60	25,702
Gesamt	1577		

Tabelle 20: t-Test bei unabhängigen Stichproben: C-Test-Ergebnis, Phase in der Oberstufe

	t	df	Sig. (2-seitig)
C-Test (Gesamtrohwert)	-7,130	1575	,000

Die Mittelwerte der Tabelle 19 zur Gruppenstatistik dokumentieren die höhere Leistungsfähigkeit der Schüler*innen in der Qualifikationsphase gegenüber den Leistungen in der Einführungsphase. Die Signifikanzberechnung ausweislich des t-Tests weist diesen Unterschied zwischen der Einführungsphase und der Qualifikationsphase als signifikant aus. Die mit $d = \frac{m2-m1}{\sigma} = \frac{-10{,}34}{25{,}52} = 0{,}40$ berechnete Effektstärke (Lenhard & Lenhard, 2014) ergibt einen mittelgradigen Effekt, d. h. es gibt einen mittleren signifikanten Unterschied zwischen den C-Test-Leistungen den erbrachten C-Test-Leistungen in der Einführungs- und der Qualifikationsphase 1.

Ein Vergleich der C-Test-Ergebnisse der beiden Ankertests (vgl. Kapitel 4.1.2), die sowohl von den Testpersonen dieser Studie, d. h. den hier untersuchten Schüler*innen in der gymnasialen Oberstufe, als auch im Rahmen von SkaLa (vgl. Kapitel 1 und Kapitel 2.1) von Lehramtsstudienbewerber*innen bearbeitet wurden, zeigt außerdem, dass die Studienbewerber*innen durchschnittlich höhere C-Test-Leistungen erzielen als die Schüler*innen. So lösen die Studierenden in den beiden C-Tests im Durchschnitt 65,3 % der Items, die Schüler*innen in der hier untersuchten Stichprobe jedoch durchschnittlich nur 53,2 % der Items. Diese Unterschiede sind dabei auf einem 5%-Niveau statistisch signifikant.

Die Hypothese, dass Schüler*innen in der Qualifikationsphase 1 signifikant höhere Leistungen im C-Test als Schüler*innen in der Einführungsphase erzielen, wird bestätigt.

5.5.2 C-Test und Geschlecht

<u>Hypothese 4</u>: Oberstufenschülerinnen erzielen signifikant höhere C-Test-Werte als Oberstufenschüler.

Als ein zentrales Ergebnis aus den großen nationalen und internationalen Schulleistungsstudien (z. B. PISA 2000-2019, DESI 2008) der letzten Jahre ist hervorgegangen, dass es u. a. in Bezug auf schriftsprachliche Kompetenzen geschlechtsspezifische Unterschiede gibt: So zeigen beispielsweise die Ergebnisse aus

der DESI-Studie (2003/2004), dass Mädchen bei sprachbezogenen Kompetenzen bereits in der Grundschule den gleichaltrigen Jungen überlegen sind und sich dieser Leistungsvorsprung der Mädchen auch in der Klasse 9 weiter manifestiert (DESI-Konsortium, 2006b, 20 ff.). Die größten Leistungsdifferenzen zwischen den Geschlechtern sind dabei in den Bereichen der produktiven Teilkompetenzen (Textproduktion und Rechtschreibung) sowie in der Argumentation zu verzeichnen (DESI-Konsortium, 2008, S. 203). Aus den PISA-Ergebnissen 2009 geht wiederum hervor, dass Jungen deutlich geringere Kompetenzen im Lesen erzielen als Mädchen (Klieme et al., 2010, 52 ff.). Diese Unterschiede bestehen dabei nicht nur in Bezug auf die Gesamtskala der Lesekompetenz, sondern auch bezüglich aller Subskalen, unabhängig davon, ob diese nach Aspekten des Lesens oder nach Textformaten ausdifferenziert werden (Klieme et al., 2010, S. 55). Dass Mädchen im Leseverstehen insgesamt deutlich besser abschneiden als gleichaltrige Jungen, bestätigen auch die aktuellen PISA-Ergebnisse aus dem Jahr 2018 (OECD, 2019d).

Der Frage, ob diese geschlechterspezifischen Leistungsunterschiede in Bezug auf schriftsprachliche Kompetenzen auch bei höheren Altersstufen weiterhin bestehen, ist Mashkovskaya (2014) u. a. im Rahmen ihrer Untersuchung zur Sprachkompetenz von Lehramtsstudierenden nachgegangen. Zur Überprüfung der Sprachkompetenz setzt die Autorin – analog zur hier vorliegenden Studie – linkssteig getilgte C-Test ein, die von den weiblichen Studierenden signifikant besser gelöst werden als von den männlichen Studierenden (Mashkovskaya, 2014, S. 203). Während die durchschnittliche C-Test-Leistung der Studentinnen 61 Punkte (von 100) beträgt, erreichen Studenten im Schnitt nur 57 Punkte (von 100) (Mashkovskaya, 2014, S. 203). Damit liegen die C-Test-Ergebnisse der männlichen Studierenden weniger häufig auf den oberen Kompetenzstufen als die ihrer weiblichen Kommilitoninnen. Im Umkehrschluss bedeutet das, dass sie in den unteren Kompetenzstufen überrepräsentiert sind (Mashkovskaya, 2014, S. 204). Diese Ergebnisse werden auch in der Replikationsstudie, die Mashkovskaya 2016 vorgelegt hat, bestätigt: Auch hier erzielen weibliche Studierende höhere C-Test-Leistungen als männliche Studierende (Mashkovskaya & Baur, 2016b, S. 224).

Zeigt sich in den C-Test-Ergebnissen in der hier vorgelegten Untersuchung der Sprachkompetenz von Oberstufenschüler*innen nun ein ähnlicher Befunde, wie der, der für die Schüler*innen der Sek. I und für die Studienanfänger*innen berichtet wurde? Für die Beantwortung dieser Frage wurden Daten analysiert, die sich auf N = 1.539 gültige Fälle zum Geschlecht beziehen. Insgesamt gaben dabei 751 Testpersonen an, *weiblich* zu sein, 788 Testpersonen *männlich*. Die Abb. 20 zeigt geschlechtsdifferent die C-Test-Leistungen und ihre Streuung:

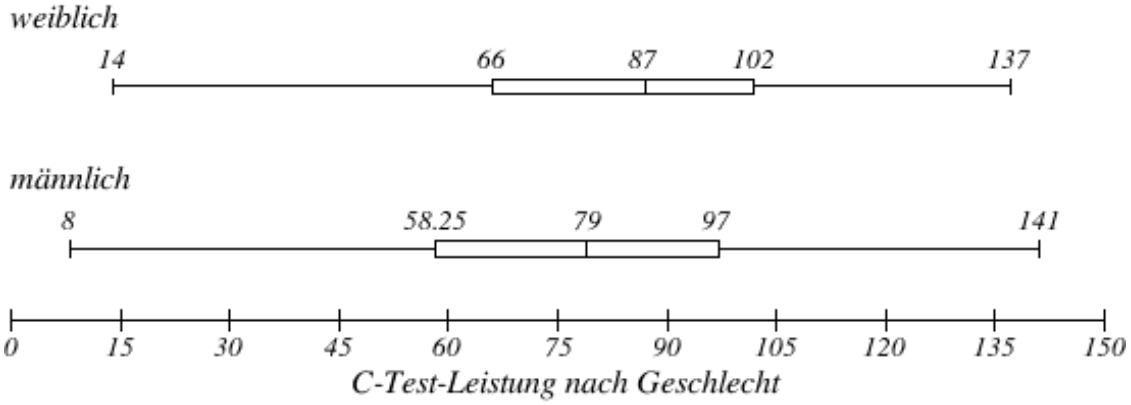


Abbildung 20: Boxplot C-Test und Geschlecht

Der Vergleich zeigt zunächst eine breitere Streuung der C-Test-Ergebnisse der männlichen Testpersonen. Sie liegen in einer Spanne zwischen 8 und 141 Punkten und streuen damit geringfügig breiter als die Ergebnisse der Schülerinnen zwischen minimal 14 und maximal 137 Punkten. Die Medianwerte beider Gruppen unterscheiden sich deutlich voneinander: Während der Median für die Schülerinnen 87 Punkte beträgt, liegt der Median für die Schüler bei 79 Punkten.

Ein ähnliches Bild zeigt sich in der Verteilung auf die C-Test-Leistungsbereiche:

Tabelle 21: Verteilung der Variable *Geschlecht* auf die Leistungsbereiche

Geschlecht	LB I	LB II	LB III	LB IV	LB V	Gesamt
männlich	7	68	493	171	49	788
	0,9 %	8,6 %	62,6 %	21,7 %	6,2 %	100 %

weiblich	10 1,3 %	93 12,4 %	490 65,2 %	118 15,7 %	40 5,3 %	751 100 %
Gesamt	17 1,1 %	161 10,5 %	983 63,9 %	289 18,8 %	89 5,8 %	1539 100 %

Betrachtet man die Verteilung der Ergebnisse, so liegen 13,7 % der weiblichen und nur 9,5 % der männlichen Schüler in den oberen Leistungsbereichen (LB I und LB II). In den Leistungsbereichen, denen Förderbedarf zugeschrieben wird (LB IV und LB V), sind hingegen Schüler öfter vertreten (27,9 %) als ihre Mitschülerinnen (21 %).

Um nun statistisch zu überprüfen, ob sich die zunächst deskriptiv dargestellten Leistungen der Oberstufenschülerinnen auch statistisch signifikant von den Leistungen der Oberstufenschüler unterscheiden, werden die Mittelwerte beider Gruppen mittels t-Test (vgl. Tabelle 23) miteinander verglichen:

Tabelle 22: Gruppenstatistik: C-Test-Ergebnis, Geschlecht

Geschlecht	N	Mittelwert	Standardabweichung
weiblich	751	83,15	25,116
männlich	788	77,14	25,517
Gesamt	1539		

Tabelle 23: t-Test bei unabhängigen Stichproben: C-Test-Ergebnis, Geschlecht

	t	df	Sig. (2-seitig)
C-Test (Gesamtrohwert)	4,443	1537	,000

Auch die Mittelwerte der Tabelle 22 zur Gruppenstatistik dokumentieren die höhere Leistungsfähigkeit der Schülerinnen. Die Signifikanzberechnung ausweislich des t-Tests weist diesen Unterschied zwischen den Geschlechtern als signifikant aus. Die Effektstärke ist indes nicht allzu groß. Berechnet man $d = \frac{m2-m1}{\sigma} = \frac{6,01}{25,52} = 0,24$ (Lenhard

& Lenhard, 2016), ergibt sich ein nur kleiner Effekt, d. h. es gibt einen signifikanten, aber eben nur kleinen Unterschied zwischen den C-Test-Leistungen der weiblichen und männlichen Teilnehmern.

Die Hypothese, dass Oberstufenschülerinnen signifikant höhere C-Test-Werte erzielen als Oberstufenschüler, wird bestätigt.

5.5.3 C-Test und Sprache

<u>Hypothesen 5</u>:

a) **Monolingual deutschsprachige Schüler*innen erzielen einen höheren C-Test-Wert als mehrsprachige Schüler*innen.**

b) **Bei einer anderen Erstsprache als Deutsch, erzielen Schüler*innen, die im familiären Kontext Deutsch als Sprache verwenden, einen höheren C-Test-Wert als Schüler*innen, die überwiegend eine andere Sprache als Deutsch innerhalb der Familie sprechen.**

c) **Es gibt einen statistisch signifikanten Unterschied zwischen den C-Test-Ergebnissen unterschiedlicher Sprachgruppen.**

Die Ergebnisse der DESI-Studie (2003/2004) zeigen – vor dem Hintergrund einer mehr an der Sprachbiografie als am Migrationsstatus der Jugendlichen orientierten Unterteilung der Gruppen (vgl. Kapitel 4.1.3 und 5.1) –, dass Schüler*innen, die neben Deutsch mindestens eine weitere Sprache in ihrem Alltag verwenden, im deutschen Schulsystem nicht die gleichen Chancen haben wie diejenigen, die nur deutschsprachig sind: Im Vergleich zu monolingual deutschsprachigen Jugendlichen erzielen mehrsprachige Schüler*innen selbst dann schlechtere sprachliche Leistungen in Deutsch, wenn sie in der alltäglichen Kommunikation hauptsächlich Deutsch verwenden (DESI-Konsortium, 2006b, S. 239; Petersen, 2014a, S. 24). Ebenso haben die nationalen

Ergebnisse von PISA 2000 und PISA 2009 gezeigt, dass neben dem Zuwanderungsstatus und der sozialen Herkunft (vgl. dazu Kapitel 4.1.3 und 5.1) auch die zu Hause gesprochene Sprache der Jugendlichen die in PISA erfasste Lesekompetenz signifikant beeinflusst (Baumert et al., 2001; Stanat et al., 2010, 200 ff.; Weis, Müller et al., 2019, 130 ff.): So erreichen sowohl in der PISA-Studie 2000 als auch 2009 Schüler*innen, die hauptsächlich eine andere Sprache als Deutsch innerhalb der Familie sprechen, niedrigere Leistungen im Bereich der Lesekompetenz als gleichaltrige Mitschüler*innen, die überwiegend Deutsch zu Hause sprechen (Weis, Müller et al., 2019, 130 f.). Und auch in der aktuellen PISA-Studie (2018) hat die Sprachverwendung im familiären Kontext einen Einfluss auf die Leseleistung: Die durchschnittliche Lesekompetenz der Jugendlichen ist signifikant niedriger, wenn im familiären Kontext hauptsächlich eine andere Sprache als Deutsch gesprochen wird (ebd. S. 155).

Ähnliche Ergebnisse beschreibt auch Neumann (2010) in ihrer Reanalyse der schriftsprachlichen Daten der DESI-Studie aus dem Jahr 2003/2004 (Beck & Klieme, 2007; DESI-Konsortium, 2008): In Bezug auf die Schreibleistungen von 9. Klässler*innen unterschiedlicher Schulformen stellt die Autorin fest, dass Schüler*innen mit nicht-deutscher Herkunftssprache am Ende der 9. Klasse schlechtere Schreibleistungen erzielen als „gemischt oder muttersprachlich deutschsprachige Jugendlich“ (Neumann & Domenech, 2010, S. 18). Über eine differenzierte Betrachtung der unterschiedlichen nicht-deutschen Sprachgruppen zeigt Neumann (2010) zudem, dass türkisch- und kurdischsprachige Jugendliche über weniger hohe Schreibkompetenzen im Deutschen (in den getesteten Bereichen Inhalt, Textaufbau, Stil) verfügen als ihre polnisch- und russischsprachigen Mitschüler*innen (Neumann & Domenech, 2010, S. 19).

Der Frage, inwieweit sich die Verwendung des Deutschen neben einer anderen Herkunftssprache als Deutsch auf die schriftsprachlichen Kompetenzen von Studierenden auswirkt, sind Mashkovskaya und Baur (2016b, S. 223) im Rahmen ihrer Arbeiten zum SkaLa®-Assessment an der Universität Duisburg-Essen nachgegangen: Dafür wurden die über SkaLa® erhobenen C-Test-Ergebnisse für Sprachkohorten mit unterschiedlichen Herkunftssprachen verglichen, u. a. mit dem Ergebnis, dass beispielswiese türkischsprachige Studierende signifikant höhere C-Test-Leistungen erzielen, wenn neben Türkisch auch Deutsch innerhalb der Familie gesprochen wird

(ebd.). Für die Kohorte der polnisch- und russischsprachigen Studierenden unterscheiden sich die C-Test-Leistungen hingegen nicht signifikant, wenn im familiären Umfeld ausschließlich die Herkunftssprache oder neben der Herkunftssprache ebenfalls Deutsch gesprochen wird (ebd.). Die Autor*innen weisen in diesem Zusammenhang außerdem darauf hin, dass Studierende, die mit beiden Elternteilen Polnisch sprechen höhere C-Test-Leistungen erzielen als Studierende, die mit ihren Eltern zu Hause nur Türkisch sprechen (ebd.).

Ausgehend von diesen Befunden soll nun im Rahmen dieser Arbeit der Frage nachgegangen werden, ob sich auch in der hier untersuchten Gruppe von Schüler*innen aus der gymnasialen Oberstufe Leistungsunterschiede hinsichtlich sprachbezogener Merkmale zeigen. Dabei sollen zunächst die C-Test-Ergebnisse von monolingual deutschsprachigen und mehrsprachigen Schüler*innen verglichen werden. Die folgenden Datenanalysen zur Sprachverwendung beziehen sich dabei auf N = 1.541 gültige Angaben, von denen 995 Schüler*innen sowohl als Erstsprache als auch innerhalb der Familie Deutsch sprechen, d. h. als monolingual deutschsprachig gelten und 546 Schüler*innen als Erstsprache und/oder als Familiensprache eine andere Sprache als/neben Deutsch sprechen und damit mehrsprachig sind. Die Abb. 21 zeigt die C-Test-Leistungen in Bezug auf eine monolingual deutsche oder eine mehrsprachige Sprachverwendung und ihre Streuungen:

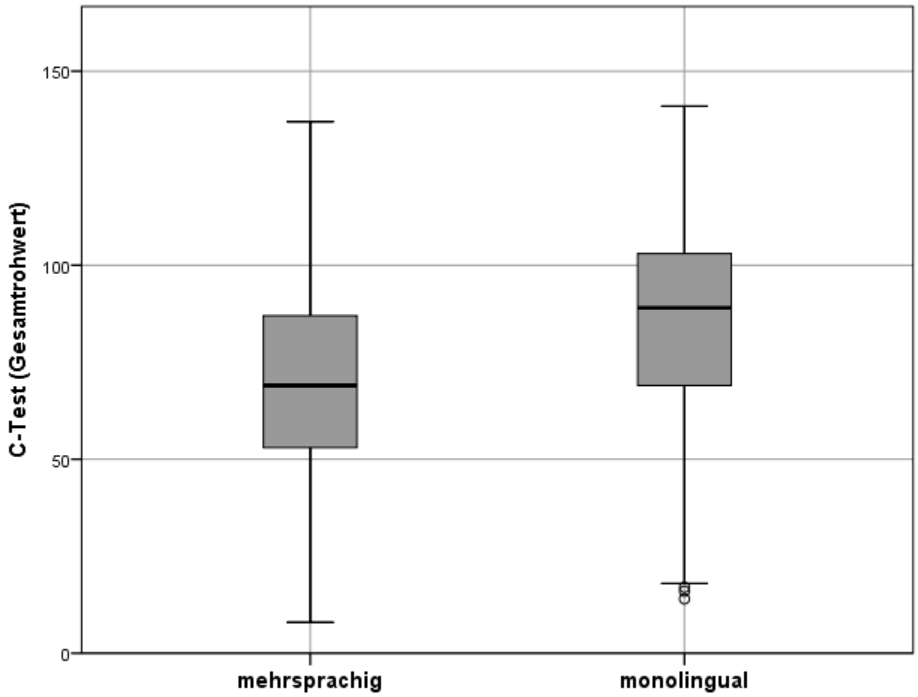


Abbildung 21: Boxplot C-Test und Sprache

Der Vergleich zeigt zunächst, dass die C-Test-Ergebnisse beider Gruppen relativ breit streuen, wobei die Ergebnisse der monolingual deutschsprachigen Schüler*innen zwischen 14 und 141 Punkten und die Ergebnisse der mehrsprachigen Schüler*innen zwischen 8 und 137 Punkten liegen. Die Medianwerte beider Gruppen unterscheiden sich jedoch recht deutlich voneinander: Während der Median für die monolingual deutschsprachigen Schüler*innen 89 Punkte beträgt, liegt der Median für die Gruppe der mehrsprachigen Schüler*innen mit 69 Punkten 20 Punkte darunter.

Ein ähnlich disparates Bild zeigt sich in der Verteilung auf die C-Test-Leistungsbereiche:

Tabelle 24: Verteilung der Variable *Sprache* auf die Leistungsbereiche

Sprache	LB I	LB II	LB III	LB IV	LB V	Gesamt
monolingual deutschsprachig	16	139	674	135	31	995
	1,6 %	14 %	67,7 %	13,6 %	3,1 %	100 %
mehrsprachig	1	22	311	154	58	546
	0,2 %	4 %	57 %	28,2 %	10,6 %	100 %

Gesamt	17	161	985	289	89	1541
	1,1 %	10,4 %	63,9 %	18,8 %	5,8 %	100 %

Betrachtet man die Verteilung der Ergebnisse, so liegen 15,6 % der C-Test-Ergebnisse der Schüler*innen, die sowohl als Erstsprache als auch zu Hause Deutsch sprechen, und nur 4,2 % der C-Test-Ergebnisse der Schüler*innen, die neben Deutsch (eine) andere Sprache(n) sprechen, in den oberen beiden Leistungsbereichen (LB I und LB II). In den Leistungsbereichen, denen Förderbedarf zugeschrieben wird (LB IV und LB V), sind mehrsprachige Schüler*innen mit ihren C-Test-Leistungen mehr als doppelt so häufig vertreten (38,8 %) wie monolingual deutschsprachige Schüler*innen (16,7 %).

Um nun zu überprüfen, ob sich die zunächst deskriptiv dargestellten Leistungen der monolingual deutschsprachigen von den Leistungen der mehrsprachigen Schüler*innen auch statistisch signifikant unterscheiden, werden die Mittelwerte beider Gruppen mittels t-Test (vgl. Tabelle 26) miteinander verglichen:

Tabelle 25: Gruppenstatistik: C-Test-Ergebnis, Sprache

Sprache	N	Mittelwert	Standardabweichung
monolingual deutschsprachig	995	86,18	24,664
mehrsprachig	546	69,18	23,895
Gesamt	1541		

Tabelle 26: t-Test bei unabhängigen Stichproben: C-Test-Ergebnis, Sprache

	t	df	Sig. (2-seitig)
C-Test (Gesamtrohwert)	-13,058	1539	,000

Auch die Mittelwerte der Tabelle 25 zur Gruppenstatistik dokumentieren die höhere Leistungsfähigkeit der monolingual deutschsprachigen Schüler*innen. Die

Signifikanzberechnung ausweislich des t-Tests weist diesen Unterschied zwischen beiden Gruppen als signifikant aus. Die Effektstärke mit $d = \frac{m2-m1}{\sigma} = \frac{17}{25{,}52} = 0{,}67$ (Lenhard & Lenhard, 2016) ergibt außerdem, dass es einen signifikanten mittelgroßen Unterschied zwischen den Leistungen der monolingual deutschsprachigen Schüler*innen und den Leistungen der mehrsprachigen Schüler*innen gibt.

Ein ähnliches Bild ergibt sich auch, wenn nicht nach dem Merkmal *Mehrsprachigkeit* – welches über die Erstsprache(n) der Schüler*innen *und* der Sprachverwendung innerhalb der Familie bestimmt wird (vgl. Kapitel 4.1.3 und 5.1) – sondern nur über die *zu Hause gesprochene Sprache* differenziert bzw. gruppiert wird: Schüler*innen, die – unabhängig von ihrer Erstsprache – überwiegend Deutsch mit ihren Eltern (und Geschwistern) sprechen, erreichen im Durschnitt 85 Punkte im C-Test, Schüler*innen, die im familiären Kontext hauptsächliche eine andere Sprache als Deutsch verwenden, erzielen im Schnitt nur 68 C-Test-Punkte. Dieser Unterschied ist signifikant und entspricht einem mittleren Effekt (d = 0,63) (Berechnungen dazu vgl. Anhang).

Wenn also die Sprachverwendung innerhalb der Familie insgesamt einen Einfluss auf das C-Test-Ergebnis hat, dann stellt sich nun die Frage, ob Schüler*innen, deren Erstsprache eine andere Sprache als Deutsch ist, möglicherweise besonders davon profitieren, wenn zu Hause überwiegend Deutsch gesprochen wird. Dazu werden im Folgenden insgesamt 481 Daten von Schüler*innen analysiert, die alle eine andere Erstsprache haben als Deutsch, von denen jedoch 73 Schüler*innen überwiegend Deutsch zu Hause sprechen und 408 Schüler*innen überwiegend (eine) andere Sprache(n) als Deutsch.

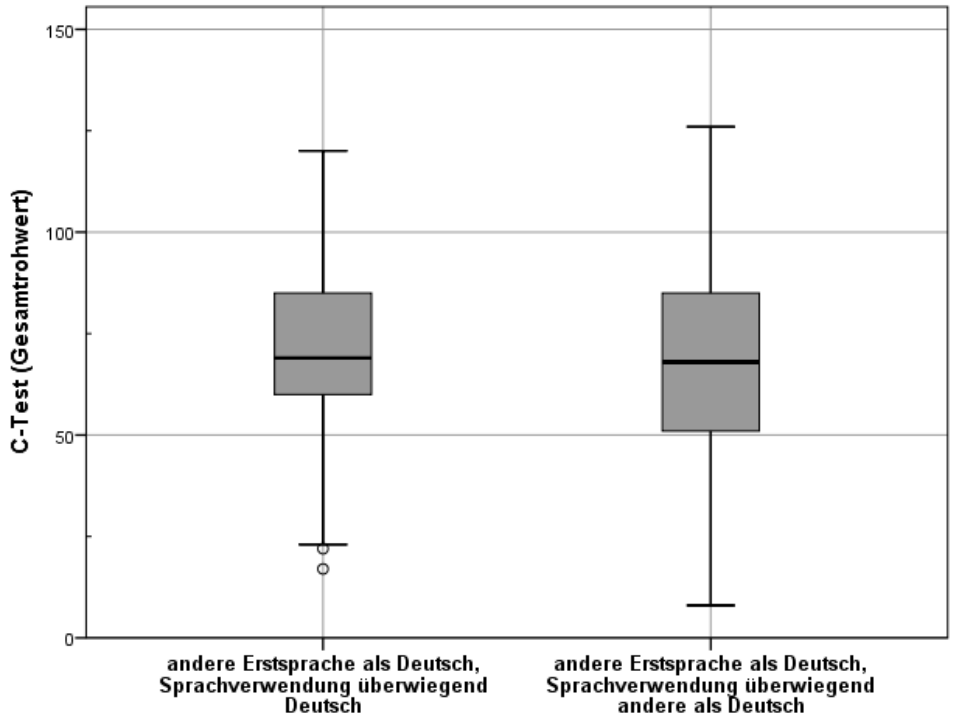


Abbildung 22: Boxplot C-Test und Erstsprache/Sprachverwendung

Der optische Vergleich der beiden Gruppen im Boxplot zeigt zunächst, dass die C-Test-Ergebnisse der mehrsprachigen Schüler*innen, die zu Hause überwiegend eine andere Sprache als Deutsch sprechen, etwas breiter streuen als die Ergebnisse der Schüler*innen, die zu Hause überwiegend Deutsch verwenden. So variieren die C-Test-Ergebnisse der erstgenannten Gruppe zwischen 8 und 126 Punkten und die der zweiten Gruppe zwischen 17 und 120. Ein Blick auf die Mediane der beiden Gruppen weist jedoch keinen direkt erkennbaren Unterschied zwischen den Gruppen auf: So liegt der Median der Schüler*innen, die zu Hause überwiegend Deutsch sprechen, bei 69 Punkten und bei der Gruppe der Schüler*innen, die im familiären Umfeld eine andere Sprache als Deutsch sprechen bei 68.

Ein ähnliches Bild zeigt sich in der Verteilung auf die C-Test-Leistungsbereiche:

Tabelle 27: Verteilung der Variable *Sprachverwendung innerhalb der Familie bei anderer Erstsprache als Deutsch* auf die Leistungsbereiche

Sprachverwendung innerhalb der Familie (Erstsprache andere als Deutsch)	**LB I**	**LB II**	**LB III**	**LB IV**	**LB V**	**Gesamt**
überwiegend Deutsch	0	1	51	15	6	74
	0 %	1,4 %	69,9 %	20,5 %	8,2 %	100 %
Überwiegend nicht Deutsch	0	16	217	125	50	408
	0 %	3,9 %	53,2 %	30,6 %	12,3 %	100 %
Gesamt	0	17	268	140	56	481
	0 %	3,5 %	55,7 %	29,1 %	11,6 %	100 %

Auffallend ist zunächst, dass keine der Schüler*innen der beiden Gruppen C-Test-Ergebnisse erzielt, die dem Leistungsbereich I zugeordnet werden können. Zwar erreichen Schüler*innen, die zu Hause überwiegend nicht Deutsch sprechen, häufiger den Leistungsbereich II, dem kein Förderbedarf zugeschreiben wird, allerdings sind sie im Vergleich zu den Schüler*innen, die mit ihren Eltern und Geschwistern meistens Deutsch sprechen, auch öfter in dem Leistungsbereich mit hohem Förderbedarf (IV und V) verortet.

Ob sich die Gruppen in Bezug auf die C-Test-Ergebnisse also überhaupt voneinander unterscheiden, soll nach der optischen Auswertung nun auch mittels t-Test statistisch überprüft werden:

Tabelle 28: Gruppenstatistik: C-Test-Ergebnis, Sprachverwendung bei nicht deutscher Erstsprache

Sprachverwendung	**N**	**Mittelwert**	**Standardabweichung**
deutsch	73	70,44	22,041
nicht deutsch	408	67,76	24,682
Gesamt	481		

Tabelle 29: t-Test bei unabhängigen Stichproben: C-Test-Ergebnis, Sprachverwendung bei nicht deutscher Erstsprache

	t	Df	Sig. (2-seitig)
C-Test (Gesamtrohwert)	,867	479	,386

Zwar zeigt die Tabelle (28), dass Schüler*innen, die zu Hause überwiegend Deutsch sprechen, im Schnitt geringfügig bessere C-Test-Ergebnisse erzielen als ihre Mitschüler*innen, die im Kontakt zu ihren Eltern (und Geschwistern) überwiegend eine andere Sprache als Deutsch sprechen. Allerdings zeigt der t-Test an, dass dieser Unterschied nicht signifikant ausfällt. Die Berechnung der Effektstärke mit $d = \frac{m2-m1}{\sigma} = \frac{2,68}{25,52} = 0,11$ (Lenhard & Lenhard, 2016) bestätigt, dass es so gut wie keinen (signifikanten) Unterschied zwischen den Leistungen von mehrsprachigen Schüler*innen gibt, die zu Hause Deutsch oder eine andere Sprache als Deutsch sprechen.

Dass sich die C-Test-Ergebnisse von mehrsprachigen Schüler*innen signifikant von denen monolingual deutschsprachiger Schüler*innen unterscheiden, konnte bereits oben gezeigt werden. Das Merkmal *Mehrsprachigkeit* wurde dabei allerdings zunächst in Abgrenzung zu einer monolingual deutschen Sprachverwendung definiert, d. h. über eine andere Erstsprache als Deutsch und/oder über eine andere Sprachverwendung als Deutsch innerhalb der Familie. Wird die Mehrsprachigkeit der Schüler*innen nun über die mit dem Fragebogen erfassten Daten differenziert nach Sprachen bzw. Sprachgruppen betrachtet, dann lässt sich für die hier untersuchte Kohorte feststellen, dass die meisten mehrsprachigen Schüler*innen neben Deutsch Türkisch sprechen, gefolgt von Russisch, Polnisch, Arabisch und Albanisch (vgl. Kapitel 5.1). In Bezug auf diese gesprochenen Sprachen soll nun überprüft werden, ob sich die C-Test-Ergebnisse bestimmter Sprachgruppen (signifikant) voneinander unterscheiden. Erzielen also beispielsweise Schüler*innen, die neben Deutsch auch Polnisch sprechen, andere C-Test-Ergebnisse als Schüler*innen, die neben Deutsch Türkisch sprechen usw.

In die folgenden Analysen sind dabei insgesamt 1.358 gültige Fälle aufgenommen worden, von denen 1.037 Fälle zur monolingual deutschsprachigen Gruppe, 152 Fälle zur Gruppe der mehrsprachig türkischen, 57 zur mehrsprachig russischen, 41 zur mehrsprachig arabischen, 38 zur mehrsprachig polnischen und 33 zur mehrsprachig albanisch Gruppe gehören. Alle Schüler*innen, auf die sich die folgenden Berechnungen zu den Sprachgruppen beziehen, gaben ebenfalls im Fragebogen an, in Deutschland eingeschult worden zu sein und damit das deutsche Schulsystem von der ersten Klasse bis zur gymnasialen Oberstufe durchlaufen zu haben (vgl. Kapitel 4.1.3).

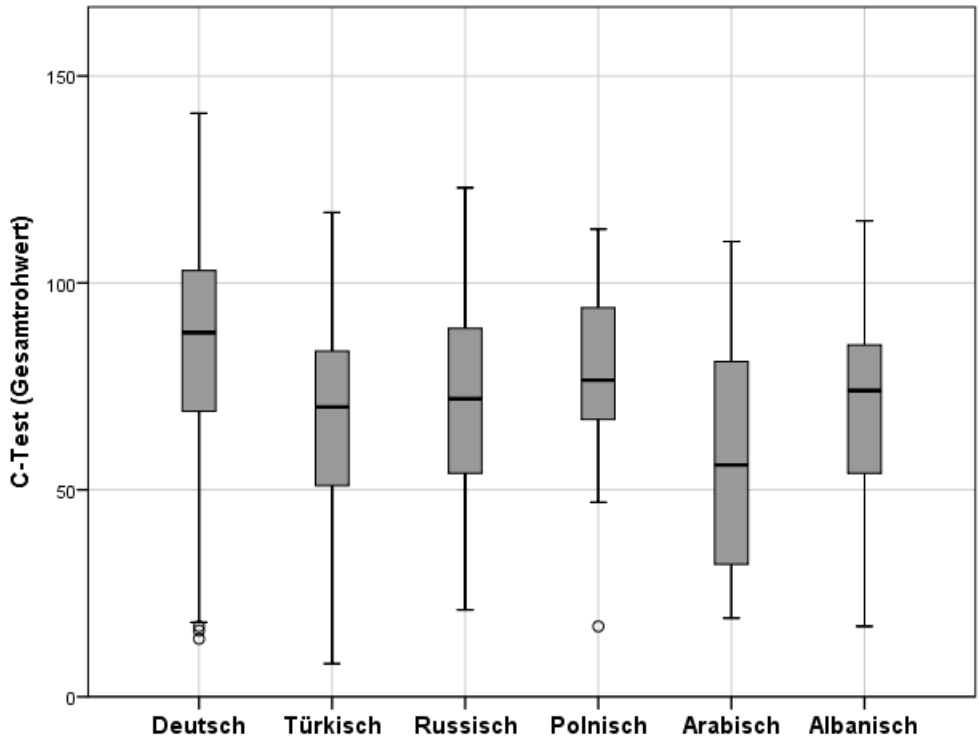


Abbildung 23: Boxplot C-Test und Sprachgruppen

Der Vergleich der hier abgebildeten größten Sprachgruppen zeigt, dass die C-Test-Ergebnisse innerhalb der Gruppen unterschiedlich breit streuen. Der insgesamt höchste C-Test-Wert (141 Punkte) wird in der Gruppe der monolingual deutschsprachigen Schüler*innen erzielt, der niedrigste Punktewert (8 Punkte) in der Gruppe der mehrsprachig türkisch Schüler*innen. Werden die Mediane der einzelnen Sprachgruppen miteinander verglichen, lässt sich feststellen, dass die Mediane der nicht-monolingual Deutschsprachigen alle unterhalb des Median der monolingualen

Sprachgruppe Deutsch (88 Punkte) liegen. An zweiter Stelle mit einem Abstand von 11,5 Punkten liegt der Medianwert der mehrsprachig polnischen Sprachgruppe (76,5 Punkte). Die Mediane der mehrsprachig türkischen, russischen, arabischen und albanischen Sprachgruppen unterscheiden sich nicht signifikant voneinander und liegen zwischen 70 und 76 Punkten. Der niedrigste Medianwert der mehrsprachig arabischen Gruppe liegt mit 56 Punkten insgesamt 32 Punkte unter dem Median der monolingual deutschsprachigen Gruppe und bildet damit die Sprachgruppe mit den im Durchschnitt schlechtesten C-Test-Ergebnissen.

Ein ähnliches Bild zeigt sich auch in Bezug auf die Verteilungen der C-Test-Ergebnisse auf die Leistungsbereiche:

Tabelle 30: Verteilung der Variable *Sprachgruppen* auf die Leistungsbereiche

Sprachgruppen	LB I	LB II	LB III	LB IV	LB V	Gesamt
Deutsch	17	144	715	147	32	1055
	1,6 %	13,6	67,8 %	13,9 %	3 %	100 %
Türkisch	0	1	86	40	16	143
	0 %	0,7 %	60,1 %	28 %	11,2 %	100 %
Russisch	0	5	30	14	5	54
	0 %	9,3 %	55,6 %	25,9 %	9,3 %	100 %
Polnisch	0	1	29	6	1	37
	0 %	2,7 %	78,4 %	16,2 %	2,7 %	100 %
Arabisch	0	0	14	10	12	36
	0 %	0 %	38,9 %	27,8 %	33,3 %	100 %
Albanisch	0	1	20	9	3	33
	0 %	3 %	60,6 %	27,3 %	9,1 %	100 %
Gesamt	17	152	894	226	69	1358
	1,3 %	11,2 %	65,8 %	16,6 %	5,1 %	100 %

Ausschließlich aus der Gruppe der monolingual Deutschsprachigen kommen Schüler*innen, deren Ergebnis im LB I liegt. 15,2 % der Ergebnisse der monolingual

deutschen Sprachgruppe liegen in den oberen Leistungsbereichen (LB I und LB II), gefolgt von den Ergebnissen in der mehrsprachig russischen (9,3 %) und in der mehrsprachig polnischen (2,7 %) Sprachgruppe. In den Leistungsbereichen, denen Förderbedarf zugeschrieben wird (LB IV und LB V), sind die Ergebnisse der mehrsprachig türkischen (39,2 %) und der mehrsprachig arabischen Sprachgruppe (61,1 %) besonders häufig verortet.

Um nun zu überprüfen, ob sich diese im Boxplot zunächst deskriptiv dargestellten Unterschiede zwischen den einzelnen Sprachgruppen auch statistisch signifikant bestätigen, werden die Mittelwerte aller Gruppen mittels einer einfaktoriellen Varianzanalyse (vgl. Tabelle 31) miteinander verglichen:

Tabelle 31: Vergleich der Sprachgruppen

Sprachvergleich		Mittelwertdifferenz	Standardfehler	Signifikanz
Türkisch	Deutsch	-18,32*	2,083	,000
	Russisch	-4,71	3,726	1,000
	Polnisch	-10,57	4,351	,228
	Arabisch	8,78	4,221	,564
	Albanisch	-3,21	4,607	1,000
Russisch	Deutsch	-13,60*	3,263	,000
	Türkisch	4,71	3,726	1,000
	Polnisch	-5,86	5,024	1,000
	Arabisch	13,50	4,912	,091
	Albanisch	1,50	5,247	1,000
Polnisch	Deutsch	-7,75	3,962	,762
	Türkisch	10,57	4,351	,228
	Russisch	5,86	5,024	1,000
	Arabisch	19,36*	5,402	,005
	Albanisch	7,36	5,708	1,000
Arabisch	Deutsch	-27,10*	3,820	,000
	Türkisch	-8,78	4,221	,564
	Russisch	-13,50	4,912	,091
	Polnisch	-19,36*	5,402	,005

	Albanisch	-11,99	5,610	,490
Albanisch	Deutsch	-15,11*	4,242	,006
	Türkisch	3,21	4,607	1,000
	Russisch	-1,50	5,247	1,000
	Polnisch	-7,36	5,708	1,000
	Arabisch	11,99	5,610	,490

*die Mittelwertdifferenz ist auf der Stufe ,05 signifikant.

Der Vergleich der Mittelwerte zeigt, dass sich die Ergebnisse der Sprachgruppen, die neben Deutsch Türkisch, Russisch, Arabisch und Albanisch sprechen, signifikant von den Ergebnissen der monolingual deutschsprachigen Gruppe unterscheiden, was bedeutet, dass die Sprecher*innen dieser mehrsprachigen Sprachgruppen signifikant niedrige C-Test-Ergebnisse erzielen im Vergleich zur monolingual deutschsprachigen Gruppe. Die Mittelwertdifferenz bzw. der Leistungsunterschied ist dabei zwischen monolingual deutschsprachigen und mehrsprachig arabischen Schüler*innen am größten. Zwischen der monolingual deutschen Sprachgruppe und der mehrsprachig polnischen Gruppe gibt es hingegen keinen signifikanten Leistungsunterschied. Im Vergleich zur mehrsprachig polnischen Sprachgruppe werden in der mehrsprachig arabischen Sprachgruppe jedoch signifikant niedrigere C-Test-Werte erzielt. Die Leistungen innerhalb der anderen mehrsprachigen Gruppen (Türkisch, Russisch, Arabisch, Albanisch) unterscheiden sich nicht signifikant voneinander.

In Bezug auf die im Rahmen dieser Arbeit untersuchten, sprachbezogenen Variablen lässt sich also nun zusammenfassend festhalten:

a) **Die Hypothese, dass monolingual deutschsprachige Schüler*innen einen höheren C-Test-Wert erzielen als mehrsprachige Schüler*innen, wird bestätigt. Die Leistungsunterschiede zugunsten der monolingual deutschsprachigen Schüler*innen sind signifikant. Ebenfalls gilt, dass Schüler*innen, die – unabhängig von ihrer Erstsprache – zu Hause Deutsch sprechen, signifikant besser abschneiden, als Schüler*innen, die innerhalb der Familie eine andere Sprache als Deutsch sprechen.**

b) **Die Hypothese, dass bei einer anderen Erstsprache als Deutsch, Schüler*innen, die im familiären Kontext Deutsch als Sprache verwenden, einen höheren C-Test-Punktwert erzielen als Schüler*innen, die überwiegend eine andere Sprache als Deutsch innerhalb der Familie sprechen, wird <u>nicht</u> bestätigt. Es gibt keinen (signifikanten) Unterschied zwischen den Leistungen von mehrsprachigen Schüler*innen, die zu Hause Deutsch oder eine andere Sprache als Deutsch sprechen.**

c) **Die Hypothese, dass es einen statistisch signifikanten Unterschied zwischen den C-Test-Ergebnissen unterschiedlicher Sprachgruppen gibt, konnte teilweise bestätigt werden: Die Leistungen der Schüler*innen, die (neben Deutsch) Türkisch, Russisch, Arabisch oder Albanisch sprechen, liegen signifikant unter den Leistungen der monolingual deutschsprachigen Schüler*innen. Zwischen den Leistungen der monolingual deutschsprachigen und (mehrsprachig) polnischen Schüler*innen bestehen hingegen keine signifikanten Unterschiede. In der polnischen Sprachgruppe werden signifikant bessere C-Test-Leistungen erzielt als in der arabischen Sprachgruppe. Alle anderen nicht-deutschen mehrsprachigen Sprachgruppen unterschieden sich in Bezug auf die C-Test-Leistungen nicht signifikant voneinander.**

5.5.4 C-Test und Bildungsniveau

<u>Hypothese 6</u>: Es gibt einen statistisch signifikanten Zusammenhang zwischen den C-Test-Ergebnissen der Oberstufenschüler*innen und dem Bildungsgrad der Eltern.

Ein stabiler Befund der internationalen Bildungsforschung ist der anhaltend starke Zusammenhang von sozialer Herkunft und Schulerfolg. Trotz der Bildungsexpansion der letzten Jahre hängt der schulische Erfolgt von Kindern nach wie vor in besonderem Maße vom Elternhaus bzw. ihrem sozioökonomischen Status ab (Hadjar et al., 2010, 224 ff.). Die aktuellen PISA-Ergebnisse (2018) bestätigen, dass Chancengerechtigkeit für das

deutsche Bildungssystem eine persistente Herausforderung bleibt (OECD, 2019b, 2019d). So hat sich u. a. die Abhängigkeit der Leistung von der sozialen Herkunft[104] im Laufe der letzten Jahre im Bereich sogar noch verstärkt (OECD, 2019a, 4 ff., 2019b; Weis, Müller et al., 2019, 129 ff.): Der Leistungsunterschied zwischen 15-Jährigen aus Familien mit günstigem sozioökonomischen Hintergrund und solchen mit ungünstigem Hintergrund ist seit 2009 im Bereich der Lesekompetenz um 9 Prozentpunkte gewachsen (OECD, 2019a, 5 f., 2019c, 251 ff.; Weis, Müller et al., 2019). Schüler*innen, die aus sozial starken Elternhäusern kommen, erreichen im Schnitt im PISA-Lesekompetenztest 113 Punkte mehr als Schüler*innen, die aus benachteiligten Familien stammen (OECD, 2019a, S. 5). Der Abstand von 24 Punkten zwischen den beiden Gruppen liegt damit über dem durchschnittlichen OECD-Abstand von 89 Punkten (OECD, 2019a, 5f., 2019c, 251 ff.).

Vor dem Hintergrund dieser Ergebnisse soll nun im Folgenden überprüft werden, ob diese Ergebnisse auch auf Schüler*innen in der gymnasialen Oberstufe übertragbar sind, d. h. ob es insgesamt einen statistischen Zusammenhang zwischen dem Bildungsniveau der Eltern, das in dieser Arbeit als ein Indikator für den sozioökonomischen Status herangezogen wird (vgl. Kapitel 4.1.3 und Kapitel 5.1), und der C-Test-Leistung der Schüler*innen gibt und ob sich die Leistungen der Oberstufenschüler*innen in Bezug auf den Bildungsgrad der Eltern im Einzelnen signifikant voneinander unterscheiden. Dabei wird in Anlehnung an die oben skizzierten Forschungsergebnisse angenommen, dass die C-Test-Ergebnisse der Schüler*innen desto höher ausfallen, je höher das Bildungsniveau der Eltern ist. Der im Fragebogen erfasste höchste Schulabschluss der Eltern (vgl. Kapitel 4.1.3) liefert dabei sowohl Hinweise auf die Dauer als auch auf den Erfolg ihrer schulischen bzw. akademischen Ausbildung. Die im Folgenden dargestellten deskriptiven Ergebnisse beziehen sich auf insgesamt N = 1.089 gültige Fälle zum Bildungsgrad der Eltern, wobei sich die Einstufung des Bildungsgrads auf den jeweils

[104] Die soziale Herkunft bzw. der sozioökonomische Status der Schüler*innen wird in PISA „anhand des PISA-Index des wirtschaftlichen, sozialen und kulturellen Status (ESCS) geschätzt, der aus mehreren Variablen des familiären Hintergrunds der Schülerinnen und Schüler abgeleitet wird: dem Bildungsniveau und dem Beruf der Eltern, der Ausstattung des Elternhauses, die als Hilfsindikator für den materiellen Wohlstand des Haushalts betrachtet werden kann, sowie der Zahl der Bücher und anderer Bildungsressourcen, die im Elternhaus verfügbar sind. Der PISA-Index des wirtschaftlichen, sozialen und kulturellen Status ist ein zusammen-gesetzter Wert, der mittels Hauptkomponentenanalyse aus diesen Indikatoren abgeleitet wurde.“(OECD (2016, 222 f.).

höchsten Schulabschluss eines (oder, wenn gleichwertig, beider) Elternteile bezieht. Fälle, in denen Angaben zu nur einem Elternteil gemacht wurden, werden für die Analysen und Berechnungen zum Bildungsniveau der Eltern herausgerechnet. Insgesamt verfügen die Eltern von insgesamt 480 Oberstufenschüler*innen über einen akademischen Abschluss, die Eltern von 348 Personen haben Abitur, in 191 Fällen haben die Eltern die Realschule erfolgreich abgeschlossen, in 52 Fällen die Hauptschule und in 18 Fällen verfügen die Eltern der Oberstufenschüler*innen über keinen Schulabschluss.

Die folgende Abbildung (Abb. 24) zeigt zunächst die Leistungsverteilung und die Streuung der C-Test-Ergebnisse in Bezug auf das Bildungsniveau der Eltern:

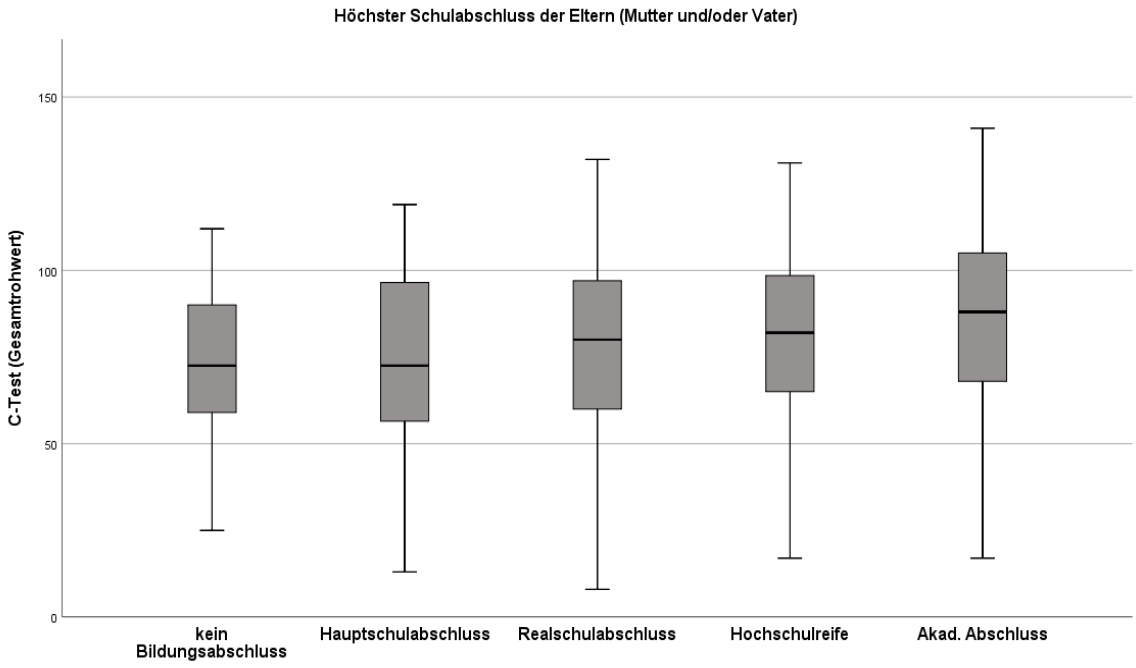


Abbildung 24: Boxplot C-Test und höchster Schulabschluss der Eltern

Wie im Boxplot zu erkennen ist, steigen die Mediane mit zunehmendem Bildungsabschluss leicht an (ausgenommen der Gruppen *kein Bildungsabschluss* und *Hauptschulabschluss*): Während der Median der Gruppen *kein Bildungsabschluss* und *Hauptschulabschluss* jeweils 72,5 Punkten beträgt, liegt er in der Gruppe der Schüler*innen mit Eltern mit *Realschulabschluss* bei 80 Punkten, in der Gruppe der Eltern mit *Abitur* bei 82 Punkten und steigt noch einmal um 3 Punkte auf insgesamt 85 Punkte in der Gruppe der Eltern mit *akademischem Abschluss* (vgl. Anhang). Insgesamt

weisen alle Gruppen eine breite Streuung der C-Test-Ergebnisse auf, wobei das Leistungsspektrum in der Gruppe der Schüler*innen, deren Eltern über einen *Realschulabschluss* verfügen, am größten ist. In dieser Gruppe wird der insgesamt niedrigste C-Test-Wert mit 8 Punkten erzielt, das beste C-Test-Ergebnis liegt bei 141 Punkten und wurde von einer Schüler*in erzielt, deren Eltern ein Studium absolviert haben. Ein Vergleich der Mittelwerte der einzelnen Gruppen (über eine einfaktorielle Varianzanalyse, vgl. Anhang) zeigt, dass sich lediglich die Leistungen der Schüler*innen, deren Eltern über einen Hauptschulabschluss oder einen Realschulabschluss verfügen, signifikant von den Leistungen der Schüler*innen mit Eltern mit akademischem Abschluss unterscheiden. Die Leistungsunterschiede zwischen den anderen Gruppen sind hingegen nicht signifikant.

In Bezug auf den Bildungsgrad der Eltern sind die C-Test-Ergebnisse wie folgt auf die Leistungsgruppen verteilt:

Tabelle 32: Verteilung der Variable *höchster Schulabschluss der Eltern* auf die Leistungsbereiche

Schulabschluss der Eltern	LB I	LB II	LB III	LB IV	LB V	Gesamt
kein Bildungsabschluss	-	1	11	5	1	18
		5,6 %	61,1 %	27,8 %	5,6 %	100 %
Hauptschulabschluss	-	2	33	11	6	52
		3,8 %	63,5 %	21,2 %	11,5 %	100 %
Realschulabschluss	1	21	116	35	18	191
	0,5 %	11,0 %	60,7 %	18,3 %	9,4 %	100 %
Hochschulreife	3	26	243	61	15	348
	0,9 %	7,5 %	69,8 %	17,5 %	4,3 %	100 %
Akademischer Abschluss	3	77	309	71	20	480
	0,6 %	16,0 %	64,4 %	14,8 %	4,2 %	100 %
Gesamt	7	127	712	183	60	1089
	0,6 %	11,7 %	65,4 %	16,8 %	5,5 %	100 %

Die Verteilung der C-Test-Ergebnisse auf die Leistungsbereiche nach dem höchsten Bildungsgrad der Eltern zeigt, dass Schüler*innen, deren Eltern über einen akademischen Abschluss verfügen, mit insgesamt 16,6 % am häufigsten C-Test-Ergebnisse erzielen, die den oberen Leistungsbereichen (I und II) zuzuordnen sind, die also keinen sprachlichen Förderbedarf haben. Schüler*innen, deren Eltern keinen Bildungsabschluss oder einen Hauptschulabschluss aufweisen, sind in der Leistungsgruppe I gar nicht und in der Leistungsgruppe II mit 5,6 % und 3,8 % deutlich seltener vertreten (bei allerdings einem insgesamt kleineren N). Von insgesamt 134 Personen erzielen nur 3 Personen (oder 2,2 %), deren Eltern einen niedrigen Bildungsgrad haben, C-Test-Ergebnisse in den obersten Leistungsbereichen. Bei sowohl ungefähr einem Drittel der Schüler*innen, deren Eltern keinen Bildungsabschluss (33,4 %) bzw. einen Hauptschulabschluss (32,7 %) erworben haben, liegt ein Förderbedarf im sprachlichen Bereich vor (sie erzielen Ergebnisse, die den LG IV und LG V zugeordnet werden können). Schüler*innen, deren Eltern Abitur oder einen akademischen Abschluss haben, attestiert der C-Test hingegen seltener (21,8 % und 19 %) Förderbedarf.

Um nun statistisch zu überprüfen, in welchem Umfang der Bildungsgrad der Eltern insgesamt mit den C-Test-Ergebnissen zusammenhängt, gilt es – in Abhängigkeit des Skalenniveaus der abhängigen und unabhängigen Variable – ein Verfahren zur Bestimmung des Zusammenhangmaßes auszuwählen: Da es sich bei dem C-Test-Ergebnis um eine metrische Variable und bei dem Bildungsgrad der Eltern um eine ordinalskalierte Variable handelt, wird als Zusammenhangsmaß der Spearman-Korrelationskoeffizient genutzt (vgl. Tabelle 33):

Tabelle 33: Korrelationsanalyse C-Test und höchster Schulabschluss der Eltern

		Höchster Schulabschluss der Eltern
C-Test-Ergebnis	Korrelationskoeffizient	0,153**
	Sig. (2-seitig)	,000
	N	1089

** Die Korrelation ist auf dem 0,01 Niveau signifikant (zweiseitig).

Der Spearman-Korrelationskoeffizient, der die Korrelation zwischen dem C-Test-Ergebnis und dem Bildungsgrad (hier: höchster Schulabschluss der Eltern bzw. mindestens eines Elternteils) beträgt r = 0,153. Der Zusammenhang ist schwach (vgl. Cohen 1988), aber hochsignifikant.

Ein weiteres interessantes Ergebnis in Bezug auf den Bildungsgrad der Eltern zeigen die Auswertungen der PISA-Ergebnisse 2003: So hat insbesondere das Bildungsniveau der Mutter einen maßgeblichen Einfluss auf die Leistungen der Schüler*innen (Kesselring & Leitner, 2007, S. 97). Die durchschnittlichen Leistungen von Schüler*innen, deren Mütter die gymnasiale Oberstufe abgeschlossen haben, liegen dabei um eine Kompetenzstufe höher als von Schüler*innen, deren Mütter nicht über einen Sekundarstufe-II-Abschluss verfügen (OECD, 2004, S. 189). Weist die Mutter ein noch höheres Bildungsniveau, d. h. einen Hochschulabschluss auf, vergrößert sich auch der Leistungsvorsprung weiter (OECD, 2004, S. 189). Auch Schymura (2013, 138 f.) konnte in ihrer Untersuchung zur Zweisprachigkeit und Schulerfolg von deutsch- und polnischsprachigen Kindern zeigen, dass die Schulleistungen der Kinder im Fach Deutsch stärker mit dem Bildungsniveau der Mutter zusammenhängen als mit dem Bildungsniveau des Vaters (Schymura, 2013, 138 f.).

Wird nun der Bildungsgrad in der vorliegenden Untersuchung nicht für beide Elternteile zusammengefasst, sondern für Mutter und Vater separat betrachtet, so ergeben sich jeweils signifikante Koeffizienten von r = 0,122 zwischen der C-Test-Leistung und dem Bildungsgrad des Vaters und r = 0,139 zwischen der C-Test-Leistung und dem Bildungsgrad der Mutter (Berechnungen siehe Anhang). Zwar ist auch die Korrelation zwischen dem C-Test-Ergebnis und dem Bildungsgrad der Mutter etwas stärker als die Korrelation zwischen dem C-Test-Ergebnis und dem Bildungsgrad des Vaters, beide Zusammenhänge bleiben jedoch schwach, wenn auch hochsignifikant. Die separat berechneten Korrelationskoeffizienten (Vater: r = 0,122; Mutter: r = 0,139) liegen zudem beide unter dem Korrelationskoeffizienten, der sich ergibt, wenn der Bildungsgrad der Eltern zusammengefasst wird (Eltern: r = 0,153).

Die Hypothese, dass es einen statistisch signifikanten Zusammenhang zwischen den C-Test-Ergebnissen der Oberstufenschüler*innen und dem Bildungsgrad der Eltern gibt, wird bestätigt.

5.5.5 C-Test und Schulstandorttyp

<u>Hypothese 7</u>: Es gibt einen statistisch signifikanten Zusammenhang zwischen den C-Test-Ergebnissen der Oberstufenschüler*innen und den Standorttypen der Schulen.

Neben den individuellen und familiären Merkmalen – wie beispielsweise dem Geschlecht, der Sprachverwendung oder dem Bildungsgrad der Eltern (vgl. oben) –, beeinflusst die sozialräumliche Lage der Schule und die sich daraus zumindest teilweise ableitende soziale Zusammensetzung der Schülerschaft deren Schulerfolg (Möller & Bellenberg, 2017, 11 ff.). So zeigen beispielsweise die Untersuchungen der PISA-Daten 2003, dass die sozialräumlichen Rahmenbedingungen der jeweiligen Schulstandorte – also vor allem die sozialwirtschaftliche Situation der Schülerfamilien und kulturellen Angebote im nachbarschaftlichen Umfeld der Schule – einen Einfluss auf die Kompetenzen der getesteten Schüler*innen haben (Baumert et al., 2005, 344 ff.; Schräpler, 2009, S. 3). In ihren Analysen der PISA-Daten (2003) konnten Baumert et al. (2005) auf der Ebene von kreisfreien Städten und Landkreisen beispielsweise nachweisen, dass sich die mittleren Leistungen bei einem Anstieg der Arbeitslosigkeit um einen Prozentpunkt, um etwa sechs Punkte auf der PISA-Leistungsskala verschlechtern (Baumert et al., 2005, 323 ff.). Möller und Bellenberg (2017, S. 23) bestätigen diesen Zusammenhang auch für die folgenden PISA-Studien und ermitteln von ihnen auf Arbeitslosigkeit zurückgeführte Leistungsdifferenzen von bis zu 30 Punkten. Die Autor*innen vermuten sogar, dass dieser Zusammenhang schulspezifisch noch sehr viel deutlicher zutage tritt, da die Analysen lediglich auf den Mittelwerten der Städte und Kreise aufruhen und soziale Ungleichheiten innerhalb der einzelnen Städte und Kreise mit diesem Vorgehen nivelliert werden (Möller & Bellenberg, 2017, S. 23).

Für den fairen Vergleich von Schulen mit ähnlichen Rahmenbedingungen bei schulübergreifenden Leistungsmessungen, wie beispielsweise den Lernstandserhebungen, wird seit 2011 der Standorttypen-Index NRW[105] verwendet (

[105] Neben dem Standorttypen-Index zum Vergleich von schulübergreifenden Leistungsmessungen gibt es in NRW den sogenannten Sozialindex, der der Steuerung von zusätzlichen Lehrerstellen an Grund- und Hauptschulen dient und seit 2015 u. a. „zur Aufteilung des für den Bereich der Lern- und Entwicklungsstörungen gebildeten Stellenbudgets auf die 53 Schulamtsbezirke herangezogen wird“ (Beierle et al. 2019, S. 28). Der Sozialindex basiert auf der Arbeitslosenquote, der Sozialhilfequote, dem

(QUA-LIS NRW, 2017). Das Konzept unterscheidet fünf Standorttypen und ist 2017 aktualisiert worden[106] (QUA-LIS NRW, 2017). Diese fünf Standorttypen werden nach dem Anteilen an Schüler*innen mit Migrationshintergrund innerhalb der Schülerschaft (unabhängig von ihrer Staatsbürgerschaft), dem durchschnittlichen Einkommen der Einwohner*innen der Wohngebiete, in dem die elterlichen Wohnungen der Schüler*innen liegen, ebenso wie dem dortigen Anteil an Personen mit Bezug von SGB-II-Leistungen, dem Anteil der Haushalte, die in Ein-bis-Zweifamilienhäusern leben und dem Anteil an Personen mit Migrationshintergrund (QUA-LIS NRW, 2017). An Schulen mit beispielswiese dem Standorttyp 1 ist der Anteil an Schüler*innen mit Migrationshintergrund vergleichsweise niedrig, ebenso gilt für die Mehrzahl der Schülerinnen und Schüler gilt,

> „dass die elterlichen Wohnungen in Wohngebieten liegen, deren Einwohnerinnen und Einwohner tendenziell ein überdurchschnittliches Einkommen aufweisen (bezogen auf den Landesdurchschnitt), in denen der Anteil der Personen mit Bezug von SGB-II-Leistungen durchschnittlich bei ca. 5% liegt, in denen durchschnittlich 60% der Haushalte in Ein- bis Zwei-Familienhäusern leben und welche einen stark unterdurchschnittlichen Anteil an Personen mit Migrationshintergrund aufweisen (bezogen auf den Landesdurchschnitt)“ (QUA-LIS NRW, 2017) (vgl. Kapitel 5.2).

An Schulen mit dem Standorttyp 5 verhält es sich andersherum, hier ist der Anteil an Schüler*innen vergleichsweise hoch (im Durchschnitt bei 60 %) und die elterlichen Wohnungen der Schüler*innen liegen in Wohngegenden, in denen das durchschnittliche Einkommen sowie der Anteil an Haushalten in Ein- bis Zwei-Familienhäusern niedrig, der Anteil an SGB-II-Empfängern und Personen mit Migrationshintergrund hingegen hoch ist (QUA-LIS NRW, 2017).

Vor dem Hintergrund der oben genannten Ergebnisse über den Zusammenhang des sozialstrukturellen Standorts bzw. der Zusammensetzung der Schülerschaft einer Schule und den erbrachten Schulleistungen soll nun auch im Rahmen dieser Arbeit überprüft

---

Anteil der Schüler*innen mit Zuwanderungsgeschichte und dem Anteil an Wohnungen in Einfamilienhäusern eines Kreisens bzw. einer kreisfreien Stadt und beschreibt damit nicht die standortspezifische Situation einer Schule (Beierle et a. 2019, S. 28).

[106] Zu Beginn des Schuljahres 2021/2022 wurden die Schulstandorte in Hinblick auf den neuen schulscharfen Sozialindex für Schulen neu berechnet und auf dieser Grundlage neu zugeordnet (vgl. Schulministerium NRW). Ein schulscharfer Schulindex war zum Zeitpunkt der Datenerhebung nicht eingeführt.

werden, ob es einen statistisch signifikanten Zusammenhang zwischen den C-Test-Leistungen und dem Schulstandorttyp gibt und ob die mit dem C-Test gemessenen schriftsprachlichen Kompetenzen an Schulen mit unterschiedlichen Standorttypen signifikant variieren. Angenommen wird hierbei, dass an Schulen mit den Standorttypen 1 und 2 im Schnitt höhere C-Test-Leistungen erzielt werden als an Schulen mit den Standorttypen 4 und 5. Die insgesamt erhobenen Testergebnisse für (N =) 1585 Schüler*innen in den gymnasialen Oberstufen aller 5 Standorttypen verteilen sich wie folgt auf die Standorttypen: Am Standorttyp 1 wurden mit 128 Schüler*innen die wenigsten und am Standorttyp 2 mit 522 die meisten Proband*innen getestet. Am Standorttyp 3 nahmen 166 Oberstufenschüler*innen an der Studie teil, am Standorttyp 4 insgesamt 436 und am Standorttyp 5 bearbeiteten 383 Schüler*innen den C-Test.
Vergleicht man nun die Aufteilung der C-Test-Leistungen an den einzelnen Schulstandorttypen miteinander, so zeigt sich folgendes Bild:

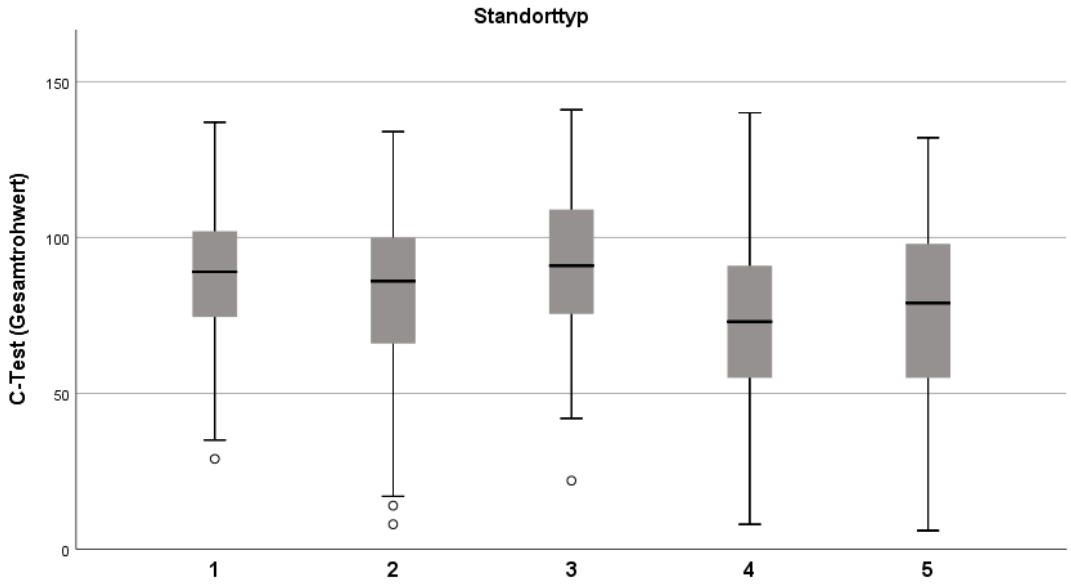


Abbildung 25: Boxplot C-Test und Schulstandorttyp

Die insgesamt höchsten C-Test-Werte werden an Schulen des Standorttyps 3 erzielt. 75 % der Schüler*innen am Standorttyp 3 erzielen C-Test-Ergebnisse, die zwischen knapp unter 50 Punkten und knapp über 110 Punkten liegen. Der Median an diesem Standorttyp ist mit 91 Punkten (Mittelwert: 91 Punkte) am höchsten. Ebenso wird in

dieser Gruppe der insgesamt höchste C-Test-Maximalwert von 141 Punkten erreicht. Die Mediane der Standorttypen 1 und 2 liegen mit 89 Punkten (Mittelwert: 88) und 86 Punkten (Mittelwert: 83) noch unter dem Median des Standorttyps 3. Die C-Test-Ergebnisse an den Standorttypen 4 und 5 sind im Vergleich zu den anderen Gruppen breiter gestreut. Der Median am Standorttyp 5 liegt mit 79 Punkten (Mittelwert: 77 Punkte) über dem Median von 73 Punkten (Mittelwert: 74 Punkte) der Schulen des Standorttyps 4. Ein Vergleich der Mittelwerte der einzelnen Gruppen (über eine einfaktorielle Varianzanalyse, vgl. Anhang) zeigt, dass sich die Leistungen der Schüler*innen an den Standorttypen 1, 2 und 3 untereinander nicht signifikant unterschieden, es jedoch zu den Leistungen, die an den Standorttypen 4 und 5 erbracht werden (zwischen denen wiederum kein signifikanter Leistungsunterschied besteht) einen signifikanten Unterschied gibt.

Wurden bisher die Rohwerte der C-Tests betrachtet, zeigt sich nun bei der differenzierten Betrachtung der C-Test-Leistungsbereiche ein mit dieser Analyse konformes Bild. Schüler*innen an Schulen an den Standorttypen 4 und 5 erzielen zu einem größeren Teil C-Test-Ergebnissen in den unteren Leistungsbereichen als an den anderen Standorttypen:

Tabelle 34: Verteilung der Variable *Standorttyp* auf die Leistungsbereiche

Standorttyp	LB I	LB II	LB III	LB IV	LB V	Gesamt
1	1	14	102	9	2	128
	0,8 %	10,9 %	79,7 %	7,0 %	1,6 %	100 %
2	8	58	349	85	22	522
	1,5 %	11,1 %	66,9 %	16,3 %	4,2 %	100 %
3	1	26	76	12	1	116
	0,9 %	22,4 %	65,5 %	10,3 %	0,9 %	100 %
4	3	25	266	111	31	436
	0,7 %	5,7 %	61 %	25,5 %	7,1 %	100 %
5	4	39	216	86	38	383
	1 %	10,2 %	56,4 %	22,5 %	9,9 %	100 %

Gesamt	17	162	1009	303	94	1585
	1,1 %	10,2 %	63,7 %	19,1 %	5,9 %	100 %

Konkret bedeutet das, dass knapp ein Drittel der Schüler*innen am Standorttyp 5 (32,4 %) mit ihren C-Test-Leistungen in den unteren Leistungsbereichen (IV und V) liegen. Die gleiche Situation lässt sich an Schulen des Standorttyps 4 beobachten (32,6 % der C-Test-Ergebnisse liegen in den Leistungsbereichen IV und V). An den Schulen des Standorttyps 3 (11,2 %), 2 (20,5 %) und 1 (8,6 %) ist dieser Anteil unterschiedlich hoch, insgesamt aber deutlich geringer. Ergebnisse, die in den oberen beiden Leistungsgruppen münden, sind am häufigsten bei den Schüler*innen am Standorttyp 3 zu verortet (mit insgesamt 23,3 %), gefolgt vom Standorttyp 2 (mit 12,6 %). An den getesteten Schulen mit Standorttyp 1 und Standorttyp 5 erzielt ein fast gleich großer Anteil von Schüler*innen (11,9 % und 11,2 %) C-Test-Leistungen, die den Leistungsbereichen I und II zuzuordnen sind und damit keinen sprachlichen Förderbedarf aufweisen.

Um nun zu überprüfen, inwieweit das C-Test-Ergebnis mit dem Standorttyp der Schulen insgesamt zusammenhängt, wird für die Berechnung des Zusammenhangsmaßes in Anhängigkeit der Variablen (siehe Tabelle 35) der Spearman-Korrelationskoeffizient genutzt:

Tabelle 35: Korrelationsanalyse C-Test und Schulstandorttyp

		Standorttyp
C-Test-Ergebnis	Korrelationskoeffizient	- 0,152**
	Sig. (2-seitig)	,000
	N	1585

** Die Korrelation ist auf dem 0,01 Niveau signifikant (zweiseitig).

Der berechnete Spearman-Korrelationskoeffizient zwischen der Variable C-Test-Ergebnis und der Variable Standorttyp beträgt r = - 0,152. Das negative Vorzeichen der

Korrelation ist in diesem Fall mit der numerischen Abstufung der Standorttypen zu erklären: Während das C-Test-Ergebnis je höher (der Testpunktwert), desto besser ist, verhält es sich bei den Standorttypen genau andersherum, d. h. je höher der numerische Wert, desto ungünstiger der Schulstandorttyp. Die Korrelationsanalyse zwischen dem C-Test-Ergebnis und der Variable *Standorttyp* zeigt einen schwachen, aber hochsignifikanten Zusammenhang.

Die Hypothese, dass es einen statistisch signifikanten Zusammenhang zwischen den C-Test-Ergebnissen der Oberstufenschüler*innen und den Standorttypen der Schulen gibt, wird bestätigt.

5.5.6 C-Test und Schulform

<u>Hypothese 8</u>: Die Schüler*innen an Gymnasien erzielen signifikant höhere C-Test-Leistungen als die Schüler*innen an Gesamtschulen.

Die PISA-Studien der letzten Jahre (2009, 2015) haben gezeigt, dass sich die Leseleistungen von Jugendlichen in Bezug auf unterschiedliche Schulformen signifikant voneinander unterscheiden (Klieme et al., 2010; Naumann et al., 2010, 23 ff.; Reiss et al., 2016; Weis, Doroganova et al., 2019, 72 f.). Für den Vergleich der Testleistungen in Abhängigkeit der Schulform werden die Daten dabei über die Gruppierungsvariablen *Gymnasium* und *nicht gymnasiale Schulform* (Hauptschule, Schule mit mehreren Bildungsgängen, integrierte Gesamtschule und Realschule) aufgeteilt (Weis, Doroganova et al., 2019, S. 68). Die aktuellen PISA-Ergebnisse (2018) bestätigen erneut – im Vergleich zu den PISA-Ergebnissen 2009 und 2015 – , dass sich die Lesekompetenz der Fünfzehnjährigen an Gymnasien im Durschnitt deutlich von den Leseleistungen der Gleichaltrigen an nicht gymnasialen Schularten unterscheidet (Weis, Doroganova et al., 2019, S. 68): Während Schüler*innen an nicht gymnasialen Schulformen im Schnitt 458 Punkte erreichen, liegt der Mittelwert der Gymnasialschüler*innen in der 9. Klasse im Bereich der Lesekompetenz bei 578 Punkten (Weis, Doroganova et al., 2019, S. 68).

Diese Leistungsunterschiede manifestieren sich dabei sowohl in der Gesamtskala zur Lesekompetenz als auch in einzelnen Teilaspekten, zu denen unterschiedliche Prozesse des Lesens, wie beispielsweise „Lokalisieren von Informationen, Textverstehen, Bewerten und Reflektieren“ gehören (Weis, Doroganova et al., 2019, S. 69). Ein Vergleich der Verteilung auf die Lesekompetenzstufen in PISA (2018) zeigt ebenfalls, dass besonders leseschwache Jugendliche selten ein Gymnasium besuchen (2 %), wohingegen diese Gruppe an nicht gymnasialen Schulformen überrepräsentiert ist (29 %). Der Anteil der besonders lesestarken Schüler*innen an nicht gymnasialen Schulformen (3 %) fällt im Vergleich zu den Gymnasien (27 %) bescheiden aus (Weis, Doroganova et al., 2019, 70 f.).

In ihrer Studie zur Sprachkompetenz von Lehramtsstudierenden konnte Mashkovskaya (2014, 217 ff.) zeigen, dass sich die Leistungen der Studierenden in dem von ihr eingesetzten linksseitig getilgten C-Test in Bezug auf die Schulform, an der der Schulabschluss erworben wurde, signifikant voneinander unterscheiden. So erzielen Studierende mit einem Gymnasialabschluss durchschnittlich höhere C-Test-Leistung als Studierende mit einem an einer integrierten Gesamtschule erworbenen Schulabschluss (Mashkovskaya, 2014, S. 219). Die Gymnasialabsolvent*innen zeigen in dieser Studie besonders häufig Leistungen, die auf keinerlei sprachlichen Förderbedarf schließen lassen (22 %). Umgekehrt weisen die Testleistungen der Studierenden mit Gesamtschulabschluss vergleichsweise sehr häufig einen Förderbedarf im sprachlichen Bereich aus (41 %) (Mashkovskaya, 2014, S. 218).

Lässt sich der oben skizzierte Befund nun auch in den hier erhobenen Daten replizieren? Lassen sich also die im Rahmen dieser Studie getesteten Schüler*innen an Gymnasien und Gesamtschulen ausweislich ihrer C-Test-Ergebnisse signifikant voneinander unterscheiden? Die deskriptive und die statistische Auswertung der Daten im Folgenden bezieht sich dabei auf die gesamte Stichprobe (N = 1585). Von den befragten Personen besuchen 1420 Schüler*innen die gymnasiale Oberstufe an einem Gymnasium und 165 Schüler*innen die Oberstufe an einer integrierten Gesamtschule.

Vergleicht man zunächst die Verteilung der C-Test-Leistungen an den beiden Schulformen, zeigt sich folgendes Bild:

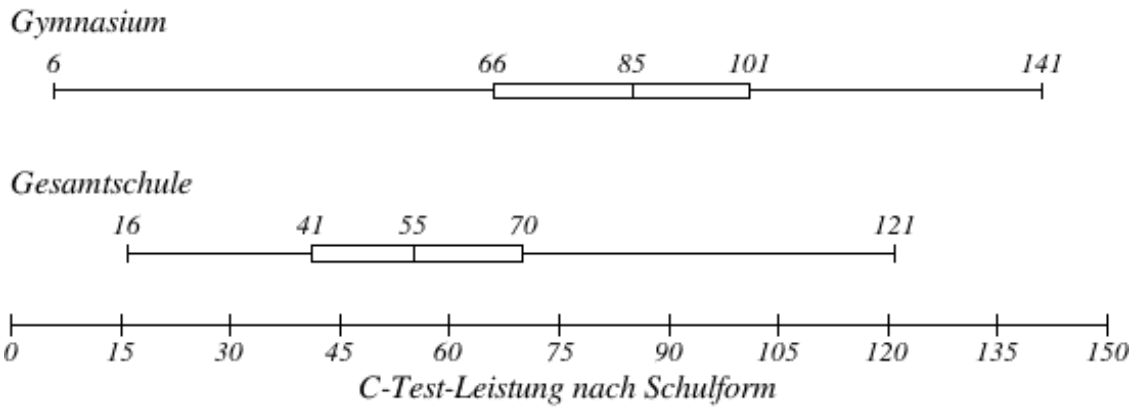


Abbildung 26: Boxplot C-Test und Schulform

Die Ergebnisse der Schüler*innen an Gymnasien sind insgesamt breiter gestreut als die der Gesamtschüler*innen und variieren zwischen 6 und 141 Punkten. Gesamtschüler*innen erzielen C-Test-Punkte, die in einer Spanne zwischen 16 und 121 Punkten zu verorten sind. Die Medianwerte beider Gruppen unterscheiden sich deutlich voneinander: Während der Median für die Gymnasialschüler*innen 85 Punkte beträgt, liegt der Median für die Gesamtschüler*innen bei lediglich 55 Punkten. Nur 25 % der Gesamtschüler*innen erzielen C-Test-Leistungen, die über 70 Punkten liegen.

Ein ähnliches Bild zeigt sich in der Verteilung auf die C-Test-Leistungsbereiche:

Tabelle 36: Verteilung der Variable *Schulform* auf die Leistungsbereiche

Schulform	LB I	LB II	LB III	LB IV	LB V	Gesamt
Gymnasium	17	160	947	231	65	1420
	1,25	11,35	66,75	16,35	4,65	100 %
Gesamtschule	-	2	62	72	29	165
		1,2 %	37,6 %	43,6 %	17,6 %	100 %
Gesamt	17	162	1009	303	94	1585
	1,1 %	10,2 %	63,7 %	19,1 %	5,9 %	100 %

Betrachtet man die Verteilung der Ergebnisse, so wird deutlich, dass 12,60 % der Ergebnisse der Gymnasialschüler*innen in den oberen Leistungsbereichen (LB I und LB

II) liegen, die Leistungen der Gesamtschüler*innen jedoch maximal in den Leistungsbereich II münden und wiederum von lediglich 1,2 % der Schüler*innen erreicht wird. In den Leistungsbereichen, denen Förderbedarf zugeschrieben wird (LB IV und LB V), sind hingegen Gesamtschüler*innen mit insgesamt 61,2 % sowohl innerhalb ihrer Gruppe als auch im Vergleich zu Schüler*innen, die ein Gymnasium besuchen (21 %), deutlich überrepräsentiert.

Um nun statistisch zu überprüfen, ob sich diese deskriptiv dargestellten Leistungsunterschiede zwischen Gymnasialschüler*innen und Gesamtschüler*innen auch statistisch (signifikant) bestätigen, werden die Mittelwerte beider Gruppen mittels t-Test (vgl. Tabelle 38) miteinander verglichen:

Tabelle 37: Gruppenstatistik: C-Test-Ergebnis, Schulform

Schulform	N	Mittelwert	Standardabweichung
Gymnasium	1420	82,45	24,773
Gesamtschule	165	57,41	20,428
Gesamt	1585		

Tabelle 38: t-Test bei unabhängigen Stichproben: C-Test-Ergebnis, Schulform

	t	df	Sig. (2-seitig)
C-Test (Gesamtrohwert)	12,499	1583	,000

Auch die Mittelwerte der Tabelle 37 zur Gruppenstatistik dokumentieren die höhere Leistungsfähigkeit der Gymnasialschüler*innen. Die Signifikanzberechnung ausweislich des t-Tests weist diesen Unterschied zwischen den Geschlechtern als signifikant aus. Für die Effektstärke, die mit $d = \frac{m2-m1}{\sigma} = \frac{25{,}04}{25{,}52} = 0{,}98$ (Lenhard & Lenhard, 2016) berechnet wird, ergibt sich ein starker Effekt, d. h. es gibt einen großen signifikanten Unterschied zwischen den C-Test-Leistungen der Gymnasialschüler*innen und denen der Gesamtschüler*innen.

Die Hypothese, dass Schüler*innen an Gymnasien signifikant höhere C-Test-Leistungen erzielen als Schüler*innen an Gesamtschulen, wird bestätigt.

5.6 Multivariate Ergebnisse

Bivariate Korrelationen, wie sie im obigen Ergebnisteil berechnet wurden, überprüfen rechnerisch den Zusammenhang zwischen zwei Variablen und sind dabei in der Regel ungerichtet. „Ungerichtet" bedeutet in diesem Kontext, dass über die einfache, bivariate Korrelation keine Informationen darüber gewonnen werden, welche Variable die andere bedingt, d. h. ob die Variable A die Variable B bedingt oder andersherum. Wenn es nun um gerichtete Zusammenhänge bzw. kausale Effekte gehen, d. h. wenn beispielsweise der Frage nachgegangen werden soll, wie sich die abhängige Variable A bei einer Änderung der unabhängigen Variablen B verändert, dann werden Regressionsanalysen angestellt (Bortz & Schuster, 2010, S. 342; Bremerich-Vos & Weirich, 2016, S. 179). Dabei gibt es verschiedene Formen der Regressionsanalysen, die sich u. a. in Bezug auf die Verfahren zur Abschätzung der Variablenbeziehung (die Regression kann als Funktion die Beziehung der Variablen nicht exakt wiedergeben, sondern sich den tatsächlichen Zusammenhängen nur annähern; vgl. unten) unterscheiden. Im Rahmen der hier eingesetzten multiplen (linearen) Regressionsanalyse (vgl. unten) soll also der Frage nachgegangen werden, wie eine unabhängige, sogenannte Prädiktorvariable (z. B. die Schulform) die abhängige Kriteriumsvariable (das C-Test-Ergebnis) beeinflusst (Bortz & Schuster, 2010, 183 ff.), z. B.: Gibt es einen signifikanten Einfluss der Schulnote auf das C-Test-Ergebnis unter Berücksichtigung der Einflüsse der anderen im Modell enthaltene unabhängigen Variablen auf das C-Test-Ergebnis?

Da die in dieser Studie zu untersuchende abhängige Variable (das C-Test-Ergebnis) metrisch skaliert ist, wird als Modell für die vorliegenden Daten eine lineare Regression gerechnet (Bortz & Schuster, 2010, S. 184). Als Erweiterung der bivariaten Korrelation, wie sie in Kapitel 5.4 und 5.5 berechnet wurden, können in einer multiplen linearen

Regression mehrere unabhängige Variablen gleichzeitig berücksichtig werden (Tausendpfund, 2019, 139 f.). So kann über die multiple, lineare Regression bestimmt werden, welche von mehreren unabhängigen Variablen (Geschlecht, Bildungsgrad der Eltern, Sprachverwendung, etc.) einen stärkeren oder welche einen schwächeren Einfluss auf die abhängige Variable (das C-Test-Ergebnis) besitzen (Tausendpfund, 2019, 144,149). Über die multiple Regression können außerdem eventuell vorhandene Scheinkorrelation (z. B. Mediations- oder Moderatoreffekte) aufgedeckt werden (Bortz & Schuster, 2010, 339 ff.).

Grundlegend wird bei der Regressionsanalyse angenommen, dass ein Zusammenhang zwischen Variablen über eine lineare Funktion abgebildet werden kann. Diese lineare Funktion bildet dabei eine Gerade, die sogenannte Regressionsgerade, mit den Regressionskoeffizienten β_0 und β_1 (bzw. β_2, β_3, β_4, usw.) (Bortz & Schuster, 2010, 184, 342). Der Regressionskoeffizient β_0 wird auch *Konstante* genannt und gibt an, an welchem Punkt die Gerade die y-Achse schneidet, d. h. welchen Wert die abhängige Variable hat, wenn die unabhängige Variable den Wert Null annimmt. Der Regressionskoeffizient β_1 beschreibt die Steigung der Regressionsgeraden und zeigt die Veränderung der abhängigen Variable (*y)* an, wenn die unabhängige Variable (*x)* um eine Einheit ansteigt (Bortz & Schuster, 2010, 185 f.).

Da die beobachteten Werte in der Regel nicht alle auf der Regressionsgeraden liegen, sondern empirisch betrachtet um die Gerade streuen, beinhaltet ein Regressionsmodell immer auch einen Fehlerterm (ε), der eben diesen Unterschied zwischen dem vorhergesagten Wert und dem tatsächlich gemessenen Wert beschreibt (Bremerich-Vos & Weirich, 2016, S. 174). Der Fehlerwert wird üblicherweise als Summe der Faktoren interpretiert, die zwar auf die abhängige Variable wirken, aber nicht auf die unabhängige Variable zurückzuführen sind (Windzio, 2013, S. 18).

Im linearen Regressionsmodell resultiert „der Wert der abhängigen Variablen *y* für das Subjekt *i* aus einer linearen Funktion einer oder mehrerer unabhängiger Variablen, deren Regressionsgewichten β, sowie einem Fehlerterm ε“ (Windzio, 2013, S. 18). Daraus ergibt sich folgendes Regressionsmodell:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{1i} \ldots + \varepsilon_i$$

y_i = abhängige Variable (C-Test-Ergebnis) des Probanden i

x_i = Wert der unabhängigen Variable(n) des Probanden i

ε_i = Fehlerterm des Probanden i

β_0 = Regressionskonstante (durchschnittlicher Wert von y, wenn alle unabhängigen Variablen auf der niedrigsten Stufe sind)

β = Regressionskoeffizient (Steigungswinkel zwischen den Ausprägungen einer Variable)

Quelle: Windzio, 2013, 18 f.

Übertragen auf die im Rahmen dieser Untersuchung erhobenen Daten bedeutet dieses Modell nun: Wenn eine unabhängige Variable (*x)* um eine Einheit steigt, so verändert sich das C-Test-Ergebnis (y) um β Punkte, gegeben alle anderen unabhängigen Variablen werden konstant gehalten. Je nach Vorzeichen von β ist diese Veränderung eine Zunahme oder eine Abnahme der C-Test-Punkte. Welche Werte die Regressionskoeffizienten dabei konkret für die hier zu untersuchenden Daten annehmen, wird im Folgenden in der Regressionsanalyse detailliert dargestellt.

Insgesamt wurden vier multivariate Regressionsanalysen berechnet, in denen die zuvor in den bivariaten Modellen untersuchten (ausschließlich) signifikanten Zusammenhänge bzw. Unterschiede einbezogen wurden (siehe Tabelle 39). Modell 1 zeigt die Koeffizienten für die Kontextvariablen *Geschlecht* und *Sprachverwendung*. In den Modellen 2 und 3 werden *Schul-* (*Stufe, Schulform, Standorttyp*) und *Leistungsvariablen* (*Deutschnote, Leseverständnis*) in die Regressionsanalysen aufgenommen. Im Modell 4 wird die Variable zum *Bildungsgrad der Eltern* in die Regression eingeschlossen. Da die Variablen *Note*, *Bildungsgrad der Eltern* und *Standorttyp* ordinalskaliert sind, müsste in der Regression prinzipiell jede einzelne Abstufung, d. h. jede einzelne Merkmalsausprägung der Variablen, im Vergleich zu einer Referenzkategorie betrachtet werden (Völkl & Korb, 2018, 19 f.). Ausgehend davon, dass jedoch vor allem der Gesamteffekt des Bildungsgrad der Eltern, der Deutschnote bzw. des Standorttyps von Interesse ist und alle drei Variablen alle mindestens fünf Merkmalsausprägungen aufweisen, können sie als „*quasi-metrische*“ Variablen behandelt werden. Aus diesem Grund werden die Variablen ohne Referenzkategorie wie eine metrisch skalierte

Variable in die Regression aufgenommen und entsprechend interpretiert (Völkl & Korb, 2018, S. 20). Für die Interpretation der Regressionstabelle (vgl. Tab. 39) gilt, dass negative Korrelationskoeffizienten auf eine Verringerung der C-Test-Punkte hinweisen, positive Werte hingegen eine Erhöhung der C-Test-Punkte anzeigen.

Tabelle 39: Regressionsanalysen

AV = C-Test-Wert	**Modell 1** Kontext-variablen	**Modell 2** Schul-variablen	**Modell 3** Leistungs-variablen	**Modell 4** SES-Variablen
Geschlecht (Ref.: weiblich)	-5,44** (1,60)	-3,93* (1,53)	-1,30 (1,43)	-1,30 (1,43)
Sprachverwendung (Ref.: mehrsprachig)	21,05*** (1,80)	17,20*** (1,79)	12,44*** (1,70)	12,43*** (1,71)
Stufe (Ref.: Einführungsphase)		10,95*** (1,71)	12,14*** (1,60)	12,14*** (1,60)
Schulform (Ref.: Gymnasium)		-11,70** (3,55)	-10,16** (3,27)	-10,13** (3,31)
Standorttyp		-3,19*** (0,67)	-3,26*** (0,62)	-3,25*** (0,62)
Deutschnote			-6,59*** (0,86)	-6,58*** (0,87)
Leseverständnis			0,95*** (0,13)	0,95*** (0,13)
Bildungsgrad der Eltern				0,06 (0,82)
Konstante	70,07*** (1,74)	79,05*** (4,98)	81,95*** (5,35)	81,63*** (6,74)
N	802	802	802	802
R^2	0,157	0,237	0,356	0,356

Standardfehler in Klammer
*** p < 0,001 ** p < 0,01 * p < 0,05

Das Modell 1, das zunächst die Merkmale *Geschlecht (männlich vs. weiblich)* und *Sprachverwendung (monolingual deutschsprachig vs. mehrsprachig)* enthält, zeigt, dass

Schüler im Vergleich zur Referenzgruppe der Schülerinnen im Durchschnitt 5,44 weniger C-Test-Punkte und damit schlechtere C-Test-Ergebnisse erzielen als Schülerinnen. Die zweite in diesem Modell aufgenommene Variable zur *Sprachverwendung* zeigt an, dass monolingual deutschsprachige Schüler*innen unter Berücksichtigung des Merkmals *Geschlecht* im C-Test durchschnittlich ein um etwa 21 C-Test-Punkte höheres Ergebnis erzielen im Vergleich zur Referenzgruppe der mehrsprachigen Schüler*innen. Der Einfluss beider Variablen ist im Modell 1 zunächst (hoch bzw. höchst-)signifikant.

In Modell 2 wird zusätzlich zu den Merkmalen *Geschlecht* und *Sprachverwendung* der Einfluss verschiedener schulbezogener Faktoren auf das C-Test-Ergebnis geschätzt. Mit der Aufnahme der drei Merkmale *Stufe (Qualifikationsphase 1 vs. Einführungsphase)*, *Schulform (Gesamtschule vs. Gymnasium)* und *Standorttyp* in das Regressionsmodell verringern sich die Regressionskoeffizienten der Merkmale *Geschlecht* und *Sprachverwendung* und das Modell hat eine höhere Varianzaufklärung (23,7 %). Während die Variablen *Stufe* und *Schulform* einen ähnlich großen Einfluss auf das C-Test-Ergebnis haben (Schüler*innen der Qualifikationsphase 1 erzielen durchschnittlich knapp 10 Punkte mehr als Schüler*innen in der Einführungsphase und Gymnasialschüler*innen knapp 11 Punkte mehr als Gesamtschüler*innen im C-Test), beeinflusst der Schulstandorttyp das C-Test-Ergebnis vergleichsweise wenig.

Das Modell 3 zeigt, dass es – werden die kontext- und schulbezogenen Variablen konstant gehalten – einen höchstsignifikanten Zusammenhang zwischen dem C-Test-Ergebnis und den aufgenommenen Leistungsvariablen gibt: Steigt die Deutschnote um eine Einheit (Note), verbessert sich das C-Test-Ergebnis im Durchschnitt um 6,59 Punkte. Im Vergleich dazu hat das mit dem LGVT 6-12 gemessene Leseverstehen zwar auch einen höchstsignifikanten – mit einem Wert von 0,95 – zugleich aber auch sehr schwachen Einfluss auf das C-Test-Ergebnis. Die Aufnahme der leistungsbezogenen Variablen in Modell 3 erhöht die Varianzaufklärung von 23,7 % (vgl. oben Modell 2) auf 35,6 %. Anzumerken ist, dass unter der Zunahme der Variablen *Deutschnote* und *Leseverständnis* der Regressionskoeffizient des Merkmals *Sprachverwendung* sinkt, der Einfluss aber signifikant bleibt im Unterschied zum Einfluss des Geschlechts auf das C-Test-Ergebnis: So besteht für die Variable *Geschlecht*, für die in der bivariaten Analyse (vgl. Kapitel 5.5.2) zunächst ein signifikanter Unterschied zwischen den Leistungen der

Schüler und Schülerinnen gezeigt wurde, in der multivariaten Regressionsanalyse, d. h. unter der Hinzunahme anderer unabhängiger Variablen, kein signifikanter Unterschied mehr.

Im Modell 4 wird schließlich noch der Bildungsgrad der Eltern mit in die Regression aufgenommen. Erstaunlicherweise hat der Bildungsgrad der Eltern keinen signifikanten Einfluss auf das C-Test-Ergebnis und wirkt sich weder merklich auf die Varianzerklärung noch auf die Regressionskoeffizienten der anderen Variablen aus. Die in den bivariaten Ergebnissen gezeigte (schwache aber signifikante) Korrelation zwischen dem Bildungsgrad der Eltern und dem C-Test-Ergebnis (vgl. Kapitel 5.5.4) kann in der linearen Regression nicht mehr bestätigt werden, d. h. wenn Kontext-, Schul- und Leistungsvariablen miteinbezogen werden.

Zusammenfassend kann festgehalten werden, dass Schüler*innen der Qualifikationsphase 1 signifikant bessere C-Test-Ergebnisse erzielen als Schüler*innen der Einführungsphase, ebenso wie Gymnasialschüler*innen signifikant höhere C-Test-Werte erreichen als Gesamtschüler*innen. Der Schulstandorttyp, die Deutschnote und das mit dem LGVT 6-12 gemessene Leseverständnis beeinflussen das C-Test-Ergebnis zwar signifikant, aber insgesamt nur geringfügig. Den größten höchstsignifikanten Einfluss auf das C-Test-Ergebnis hat die Sprachverwendung der Schüler*innen, selbst wenn Zusammenhänge zum Bildungsgrad der Eltern sowie schul- und leistungsbezogene Merkmale einbezogen werden.

6. Interpretation und Diskussion

Zur Überprüfung der im Rahmen dieser Arbeit formulierten Hypothesen (vgl. Kapitel 3) wurden in Kapitel 5 verschiedene deskriptive und (interferenz-)statistische Verfahren eingesetzt und die daraus gewonnen Ergebnisse dokumentiert. Im Folgenden Kapitel soll es nun darum gehen, die so erhaltenen Befunde zu kommentieren, in Bezug zu anderen Untersuchungen zu setzten als auch Limitationen aufzuzeigen.

Wie die Korrelationsanalysen im Ergebnisteil dieser Arbeit gezeigt haben (vgl. Kapitel 5.4.1), konnte die Hypothese, dass es einen statistisch signifikanten Zusammenhang zwischen dem Lesetest (LGVT 6-12) und dem hier eingesetzten C-Test gibt, bestätigt werden. Die mittlere Korrelation zwischen dem Lesetest und dem C-Test kann nun einerseits dahingehend gedeutet werden, dass sowohl mit dem C-Test als auch mit dem Lesetest u. a. ähnliche Facetten von Lesekompetenz erfasst werden: So kommen sowohl bei der Bearbeitung des LGVT 6-12 als auch beim C-Test z. B. Fähigkeiten und Fertigkeiten zum Tragen, einen Text sinnentnehmend zu lesen und zwischen globalem Lesen (um eine allgemeine Vorstellung von Inhalt und Aussage des Textes zu gewinnen) und selektivem Lesen (um verschiedene, aber nicht notwendigerweise alle Details des Textes, deren Zusammenhang und damit auch die Gesamtaussage zu erfassen) zu wechseln, über eine angemessene Wortschatzbreite zu verfügen, korrekte Verwendung von Kollokationen ebenso zu erkennen wie idiomatische Wendungen, Wortbedeutungen und Konnotationen usw. (vgl. Kapitel 2.1.2, 3.1 unf 5.4.1). Der mittlere Zusammenhang beider Testformate deutet also einerseits daraufhin, dass eine Rezeptionsteilfertigkeit, d. h. die Lesekompetenz, mit dem Lesetest und dem C-Test gleichermaßen gemessen wird, andererseits aber werden über den C-Test zusätzliche – und hier vor allem wie in Kapitel 2.1 beschrieben, basale (linguistische) – Facetten der Sprachkompetenz erfasst: So sind bei der Lösung von C-Tests beispielsweise auch textgrammatische und syntaktische Kompetenzen erforderlich, um die fehlenden Buchstaben eines Wortes erfolgreich rekonstruieren zu können (vgl. Kapitel 2.3.5), ebenso wie die Fähigkeit die ergänzten Worthälften orthographisch korrekt zu ergänzen, die beim Lösen des Lesetests keine Rolle spielen (vgl. Kapitel 4.1.1). Während beim Lesetest durch die Auswahl der in den Textzusammenhang passenden, richtigen Wörter

also vorrangig rezeptive Fähigkeiten und Fertigkeiten zum Tragen kommen, erfordert die Rekonstruktion der Lücken im C-Test auch den Wechsel von der Rezeption hin zur Produktion. Die mittlere Korrelation ist insofern erwartungskonform, als dass sie die (zunächst theoretisch begründete) Annahme stützt, dass mit dem C-Test vornehmlich eine zentrale, allgemeine Sprachkompetenz gemessen wird (vgl. Kapitel 2.3.), die eben nicht nur Lesekompetenz, sondern verschiedenen sprachlichen Teilfertigkeiten zugrunde liegt (vgl. Kapitel 2.1.2 und Kapitel 2.3): Denn auch, wenn nun für die in dieser Untersuchung eingesetzte C-Test-Version mit linksseitiger Tilgung besonders basale semantische und lexikalische Fähigkeiten und Fertigkeiten eine entscheidende Rolle spielen, da diese bei der Ergänzung von Wortanfängen stärker im Fokus stehen (vgl. dazu ausführlich Kapitel 2.3.), ändert sich nichts am Grundprinzip des Tests selbst, da die natürlichen sprachlichen Redundanzen auch im linksseitig getilgten C-Test weiter reduziert werden (vgl. Kapitel 2.3.4). Über das ganzheitliche Vermögen mit diesen Störungen umzugehen, sozusagen trotz „Leitungsschäden“ zu verstehen, was geschrieben oder gesagt wird, bzw. über die Fähigkeit zum sprachlichen Schließen der Lücken auf der Satz-, Wort- als auch Textebene, wird dabei sowohl bei linksseitig als auch rechtsseitig getilgten C-Tests auf eine tiefer liegende, abstrakt-latente Sprachkompetenz geschlossen (vgl. Kapitel 2.1.2 und Kapitel 2.3.4).

Die von Mashkovskaya (2014) berichteten teilweise hohen Korrelationen zwischen dem von ihr eingesetzten linksseitig getilgten C-Test und verschiedenen Lesetests für erwachsene Studierende mit Deutsch als Erst- und Zweitsprache (Mashkovskaya, 2014, 197 ff.) konnten in der hier angestellten Studie nicht repliziert werden. Die Schlussfolgerung von Mashkovskaya (2014), dass der eingesetzte C-Test mit linksseitiger Tilgung primär rezeptive Lesekompetenz erfasst kann auf der Grundlage der Studie bzw. der eben nur mittleren Korrelation zwischen dem C-Test und dem hier eingesetzten Lesetest so nicht bestätigt werden (Mashkovskaya, 2014, 197 ff.). Ob die von Mashkovskaya eingesetzten Lesetests möglicherweise Lesekompetenz auf anderen (möglicherweise hierarchiehöheren) Ebenen (vgl. Kapitel 2.3) abfragen, bzw. andere Facetten von Lesekompetenz erfassen als der hier eingesetzte Lesetest müsste in weiteren Untersuchungen (beispielsweise zur Übereinstimmungsvalidität) überprüft werden.

In Bezug auf den Zusammenhang zwischen dem mit dem LGVT 6-12 erfassten Leseverstehen und der Lesegeschwindigkeit konnte für die hier untersuchte Kohorte der Oberstufenschüler*innen hingegen ein ähnlicher Befund festgestellt werden wie für die Normierungsstichprobe des Lesetests: Die von den Testentwicklern des LGVT 6-12 berichteten mittleren bis hohen Korrelationen zwischen beiden Variablen (*Lesegeschwindigkeit* und *Leseverstehen*) zeigt sich ebenfalls für die Gruppe der hier getesteten Oberstufenschüler*innen (vgl. Kapitel 5.4.1). Die Interpretation des mittleren bis hohen Zusammenhangs bzw. die Annahme der Testentwickler des LGVT 6-12, dass Lesegeschwindigkeit und Leseverständnis eng miteinander verbunden sind, da „die Möglichkeiten der aktiven Sinnentnahme“ beim Lesen durch fehlende Automatisierung des technischen Dekodierens eines Textes bei leseschwachen Personen besonders beeinträchtigt wird bzw. die Prozesse des Dekodierens bei lesestarken Personen weitgehend automatisiert ablaufen, wodurch mehr Zeit für die aktive Entnahme von Informationen, d. h. für das sinnentnehmende Lesen, bleibt (Artelt et al., 2005; Schneider et al., 2007, S. 8), scheint auch für die hier untersuchte Stichprobe zu gelten bzw. sich zu bestätigen.

Interessanterweise ist die Korrelation zwischen der Lesegeschwindigkeit im LGVT 6-12 und dem C-Test insgesamt schwach. Aufgrund dieses Befundes kann angenommen werden, dass die Bearbeitungszeit der C-Tests von jeweils 5 Minuten für Schüler*innen in der Oberstufe ausreichend bzw. angemessen ist.

Die im Rahmen dieser Arbeit zu überprüfende Hypothese, dass es einen mittleren statistisch signifikanten Zusammenhang zwischen dem C-Test und der Deutschnote gibt, konnte ebenfalls bestätigt werden. Die mittlere Korrelation zwischen dem C-Test und der Deutschnote auf Klassenebene ist insofern erwartungskonform, als dass im hier eingesetzten linksseitig getilgten C-Test und der Deutschnote unterschiedliche Fähigkeiten und Fertigkeit erfasst werden, zugleich aber auch eine Schnittmenge (vgl. Kapitel 5.4.2). Dass unterschiedliche Fähigkeit und Fertigkeit in die Deutschnote eingehen, liegt auf der Hand: So bildet die Deutschnote eine kumulative Summe aus vielen verschiedenen Fähigkeiten und Fertigkeiten, die schriftlich, mündlich und praktisch in Form von Klassenarbeiten bzw. Klausuren, Lese- und Schreibaufgaben, Präsentationen, Referaten, Vorträgen, Protokollen, Gesprächsbeiträgen im Unterricht,

Portfolios, Recherchearbeiten etc. dargeboten werden. Der C-Test dagegen fokussiert die schriftsprachliche Basis, bzw. liefert er als integrativer Sprachtest Informationen über den globalen Sprachstand – und hier vor allem in den Bereichen Grammatik, Orthografie, Lexik, Semantik, Syntax, Leseverständnis – die allerdings u. a. auch in den verschiedenen von den Schüler*innen erstellten Texten in den Blick der beurteilenden Deutschlehrer*innen gerät. Andererseits werden über den hier eingesetzten C-Test mit linksseitiger Tilgung vor allem, wie in Kapitel 2.1 definiert, lexikalisch-semantische und morpho-syntaktische Fähigkeiten und Fertigkeiten sowie basale Aspekte der formal- und normsprachlichen Richtigkeit abgefragt, die beispielsweise bei der Beurteilung der schriftlichen Leistungen der Oberstufenschüler*innen so gut wie keine Rolle spielen (vgl. Kapitel 2.2). Vor diesem Hintergrund und der Tatsache, dass beispielsweise in der Untersuchung von Mashkovskaya, die in Bezug auf die Sprachkompetenz von Lehramtsstudierenden keine Korrelation zwischen dem von ihr eingesetzten linksseitig getilgten C-Test und den Deutschzeugnisnoten der Studierenden ermittelt wurde, ist es einerseits bemerkenswert, dass die Deutschnoten der getesteten Schüler*innen auf Klassenebene überhaupt im mittelgradigen Bereich miteinander korrelieren (Mashkovskaya, 2014, S. 228). Andererseits spiegeln die Befunde der hier vorliegenden Arbeit die Ergebnisse von Wockenfuß (2009) wider, die in ihrer Untersuchung zwischen dem von ihr eingesetzten, rechtsseitig getilgten *speeded-C-Tests* und der Deutschnote ebenfalls eine mittlere Korrelation feststellt (2009, S. 182). Die Schlussfolgerung von Wockenfuß (2009), dass der von ihr „eingesetzte muttersprachliche C-Test unterrichtsrelevante sprachliche Fähigkeiten und Fertigkeiten misst“, die eben auch von der Lehrkraft als eine Art globale „allgemeine Sprachbeherrschung“ integrativ in der Deutschnote miterfasst wird, ließe sich nun auch für den hier eingesetzten linksseitigen C-Test übertragen (Wockenfuß, 2009, 181 f.).

In Bezug auf die in dieser Arbeit gestellte Frage nach gruppenspezifischen Unterschieden bzw. Zusammenhängen hinsichtlich verschiedener Hintergrundvariablen für die Kohorte der untersuchten Oberstufenschüler*innen konnten – wie in Kapitel 5.5 und 5.6 gezeigt – die Hypothesen

- **zur Phase in der Oberstufe**
- **zur Sprachverwendung** (zumindest teilweise)

- **zum Schulstandorttyp** und
- **zur Schulform**

bestätigt werden: So konnte mit dem im Rahmen dieser Arbeit eingesetzten C-Test – ebenfalls wie in der Untersuchung von Wockenfuß (2009) – erwartungskonform ein Zuwachs der mit dem C-Test gemessenen Sprachkompetenz bei den Leistungen der Schüler*innen in der Qualifikationsphase 1 im Vergleich zu den Leistungen der Schüler*innen der Einführungsphase aufgezeigt werden. Ebenso erzielen Studierende in den Ankertests bessere C-Test-Leistungen als die getesteten Schüler*innen. Es ist also davon auszugehen, dass die Entwicklung von Sprache bzw. basalen sprachlichen Kompetenzen mit dem Eintritt in die gymnasiale Oberstufe nicht vollständig abgeschlossen ist und sich die schriftsprachlichen Kompetenzen auch im Laufe der Oberstufe und darüber hinaus weiterentwickeln.

Die Hypothese, dass es einen Leistungsunterschied zwischen mehrsprachigen und monolingual deutschsprachigen Schüler*innen gibt, konnte im Rahmen dieser Arbeit ebenfalls bestätigt werden. Anders als beispielsweise in der Untersuchung zur Schreibfähigkeit von Oberstufenschüler*innen von Petersen (2014), die zumindest für die syntaktischen Schreibfähigkeiten von Oberstufenschüler*innen keine Evidenz für Unterschiede zwischen mehrsprachigen und einsprachigen Schüler*innen gefunden hat (Petersen, 2014c, 247 ff.), divergieren die mit dem C-Test gemessenen schriftsprachlichen Kompetenzen in Hinblick auf das Merkmal der *Sprachverwendung* (mehrsprachig vs. monolingual deutschsprachig): So erreichen mehrsprachige Schüler*innen in der Oberstufe signifikant niedrigere C-Test-Ergebnisse als monolingual deutschsprachige Schüler*innen. Die Ergebnisse der Regressionsmodelle konnten ebenfalls zeigen, dass der Indikator der Sprachverwendung die insgesamt stärkste Prädiktorvariable bzw. den größten Einfluss von den in den Regressionsmodellen einbezogenen Variablen auf das C-Test-Ergebnisses darstellt (vgl. Kapitel 5.6). Dass die zu Hause gesprochene Sprache bereits in der Jahrgangsstufe 9 signifikant zur Vorhersage der Lesekompetenz beiträgt, selbst wenn Zusammenhänge zum Zuwanderungshintergrund und zur sozialen Herkunft einbezogen werden, zeigen beispielswiese die Befunde aus PISA 2018 (Weis, Müller et al., 2019, S. 155): So liegt hier

die durchschnittliche Lesekompetenz der Schüler*innen, die zu Hause hauptsächlich eine andere Sprache als Deutsch sprechen unter der Lesekompetenz der Schüler*innen, die zu Hause hauptsächlich Deutsch verwenden (ebd.). Auch wenn nun im Rahmen der hier vorliegenden Untersuchung mit dem eingesetzten C-Test unterschiedliche Facetten von schriftsprachlichen Kompetenzen – und nicht explizit Lesekompetenz – erfasst werden (vgl. oben), scheinen sich die Leistungsunterschiede zwischen vorwiegend mehrsprachigen und monolingual deutschsprachigen Schüler*innen in Bezug auf schriftsprachliche Kompetenzen auch über die Klassenstufe 9 hinweg im Bereich der gymnasialen Oberstufe weiter zu manifestieren. Der Befund der Leistungsunterschiede in Bezug auf die Sprachverwendung in der Kohorte der Oberstufenschüler*innen deckt sich damit ebenfalls mit den Ergebnissen der SkaLa®-Untersuchung, die gezeigt haben, dass Studierende, die einen Migrationshintergrund aufweisen – der über die Erstsprache der Mutter bzw. des Vaters erfasst wurde (Bremerich-Vos & Weirich, 2016, S. 166) – durchschnittlich niedrigere Leistungen im C-Test und in der Lese-Schreibaufgabe erzielen als Studierenden ohne Migrationshintergrund (Mashkovskaya & Baur, 2016b, S. 229).

Die Annahme, dass sich die Verwendung des Deutschen im familiären Kontext bei einer anderen Erstsprache als Deutsch besonders positiv auf die mit dem C-Test gemessene Sprachkompetenz auswirkt, konnte im Rahmen dieser Untersuchung allerdings nicht bestätigt werden. So zeigen die Ergebnisse keinen (signifikanten) Unterschied zwischen den Leistungen von mehrsprachigen Schüler*innen, die zu Hause Deutsch oder eine andere Sprache als Deutsch sprechen. Einschränkend ist an dieser Stelle jedoch auf die geringe Datenmenge, die der Überprüfung dieser Hypothese zugrunde lag (vgl. Kapitel 5.1 und Kapitel 5.5.3), hinzuweisen und damit von einer Generalisierbarkeit dieses Befundes abzusehen.

Auf der Grundlage der vorliegenden Daten konnte zudem – ähnlich wie beispielsweise in den Untersuchungen der PISA-Daten 2003 (Baumert et al., 2005, 344 ff.; Schräpler, 2009, S. 3) – gezeigt werden, dass die Rahmenbedingungen der jeweiligen Schulstandorte insgesamt mit den Kompetenzen der getesteten Schüler*innen zusammenhängen (vgl. Kapitel 5.5.5). In Bezug auf die einzelnen Standorttypen der Schulen unterscheiden sich die Leistungen der Oberstufenschüler*innen allerdings nur

an den Standorttypen 1-3 signifikant von den Leistungen an Schulen mit den Standorttypen 4 und 5 (vgl. Kapitel 5.5.5). Dass unterschiedliche Abschneiden an den jeweiligen Standorttypen hängt also zumindest in Teilen (auch) mit der Zusammensetzung der Schülerschaft zusammen: So bestätigt sich für die hier untersuchte Stichprobe – so wie es das Standorttypenkonzept zumindest theoretisch indiziert – dass der Anteil an Schüler*innen aus eher bildungsfernen Familien (gemessen über den Bildungsgrad der Eltern) als auch der Anteil an mehrsprachigen Schüler*innen an den Standorttypen 4 und 5 deutlich höher liegt als an den Standorttypen 1-3. Unter der Hinzunahme anderer (schul-, leistungs- und personenbezogener) Variablen hat das Merkmal *Schulstandorttyp* im Regressionsmodell allerdings einen nur schwachen Einfluss auf die gemessene Sprachkompetenz und eignet sich nur bedingt als Prädiktorvariable. Auch hier ist allerdings auf die Verzerrung hinsichtlich der Verteilung der Daten an den einzelnen Standorttypen hinzuweisen: So wurde über die Hälfte der Daten an den Standorttypen 4 und 5 erfasst und knapp ein Drittel am Standorttyp 2; die Daten zu den Standorttypen 1 und 3 wurden hingegen an jeweils nur zwei Schulen erhoben und bilden in der Gesamtstichprobe jeweils weniger als 10 % (vgl. Kapitel 5.1 und Kapitel 5.5.5). Ob und inwieweit sich signifikante Leistungsunterschiede zwischen den Schulen an den Standorttypen 1-3 bei einer Vergrößerung der Stichprobe eben dieser Standorttypen zeigen würden und ob dann möglicherweise die Variable *Standorttyp* das C-Test-Ergebnis insgesamt stärker (signifikant) beeinflussen würde, müsste in weiteren Untersuchungen überprüft werden.

Die Ergebnisse der hier vorliegenden Untersuchung zeigen zudem – ähnlich wie in den PISA-Studien der letzten Jahre (2009, 2015, 2018) –, dass sich die Leistungen der Oberstufenschüler*innen im C-Test in Bezug auf die Schulform (Gymnasium vs. Gesamtschule) signifikant voneinander unterscheiden (Klieme et al., 2010; Naumann et al., 2010, 23 ff.; Reiss et al., 2016; Weis, Doroganova et al., 2019, 72 f.). Ebenso wie in der Studie zur Sprachkompetenz von Lehramtsstudierenden, in der Mashkovskaya (2014, 217 ff.) berichtet, dass Studierende in dem von ihr eingesetzten linksseitig getilgten C-Test mit einem Gymnasialabschluss durchschnittlich höhere C-Test-Leistung als Studierende mit einem an einer integrierten Gesamtschule erworbenen Schulabschluss erzielen (Mashkovskaya, 2014, S. 219), erreichen in der hier

vorliegenden Untersuchung Oberstufenschüler*innen, die ein Gymnasium besuchen, signifikant höhere Leistungen als Oberstufenschüler*innen an Gesamtschulen. An dieser Stelle ist im Besonderen darauf hinzuweisen, dass der Leistungsunterschied möglicherweise sogar noch gravierender als ausgewiesen angenommen werden muss, da die Schüler*innen an den Gymnasien und in den Gesamtschulen jeweils in der Einführungsphase der gymnasialen Oberstufe getestet wurden, die Schüler*innen in den Gymnasien – weil die teilnehmenden Gymnasien überwiegend bereits nach acht Jahren das Abitur vergeben (G8 an allen, ausgenommen einem, teilnehmenden Gymnasien) – aber in der Einführungsphase ein Jahr jünger waren als ihre Mitschüler*innen an den Gesamtschulen, die in der Einführungsphase zum Testzeitpunkt bereits ein Jahr länger beschult worden waren (G9 an den teilnehmenden Gesamtschulen) (vgl. Kapitel 4.2).

Bei der Einordnung der Ergebnisse in Bezug auf die Leistungsunterschiede an beiden Schulformen ist allerdings darauf hinzuweisen, dass sich auch die Zusammensetzung der Schülerschaft an den Gesamtschulen von den Gymnasien unterscheidet: So liegt der prozentuale Anteil an mehrsprachigen Schüler*innen – wobei das Merkmal *Mehrsprachigkeit* die stärkste Prädiktorvariable bildet – in der hier vorliegenden Stichprobe an den Gesamtschulen mit 48,5 % deutlich höher als an den Gymnasien (33,9 %). Ebenso fällt der Anteil an Schüler*innen, deren Eltern über keinen oder einen Hauptschulabschluss verfügen an den Gesamtschulen mit 14,6 % deutlich höher aus als an den Gymnasien (5,6 %) und umgekehrt besuchen nur 46,9 % der Schüler*innen, deren Eltern eine Hochschulreife oder einen akademischen Abschluss haben, die gymnasiale Oberstufe an der Gesamtschule, wohingegen diese Gruppe am Gymnasium mit 78,8 % vertreten ist. Allerdings konnte für den hier erfassten Bildungsgrad der Eltern in der Regressionsanalyse unter Einbeziehung weiterer Variablen kein signifikanter Einfluss auf das C-Test-Ergebnis gezeigt werden.

Auch wenn sich die Leistungen im C-Test hinsichtlich der Schulform in der hier vorliegenden Studie insgesamt signifikant voneinander unterschieden, ist durch die verhältnismäßig geringe Stichprobengröße von Schüler*innen an Gesamtschulen (vor allem im Vergleich zur Kohorte der Gymnasialschüler*innen; vgl. Kapitel 5.1 und Kapitel 5.5.6) von einer Generalisierbarkeit der beschriebenen Ergebnisse eher abzusehen.

Vielmehr müssten die im Rahmen dieser Arbeit gefundenen Unterschiede in Bezug auf die mit dem C-Test gemessene Sprachkompetenz in weiteren Untersuchungen anhand einer größeren Stichprobe von Oberstufenschüler*innen an Gesamtschulen empirisch weiter überprüft werden.

Zwar konnten die Hypothesen

- **zum Bildungsgrad der Eltern sowie**
- **zum Geschlecht**

in den bivariaten Korrelationsanalysen zunächst ebenfalls erwartungskonform bestätigt werden, d. h. es gibt sowohl einen schwachen signifikanten Zusammenhang zwischen dem C-Test-Ergebnis und dem Bildungsgrad der Eltern, ebenso erzielen Schüler*innen geringfügig, aber signifikant bessere Leistungen im C-Test als Schüler. Allerdings zeigte sich in der Regressionsanalyse, dass beide Variablen unter der Zunahme anderer (schul-, sprach- und leistungsbezogener) Variablen keinen signifikanten Einfluss auf das C-Test-Ergebnis mehr haben (vgl. Kapitel 5.6). Dass das Geschlecht unter der Hinzunahme anderer Variablen keinen signifikanten Einfluss auf das C-Test-Ergebnis hat, ist beispielsweise vor dem Hintergrund der PISA-Ergebnisse (2009, 2015, 2018), die durchgängig zeigen, dass die mittlere Lesekompetenz der 15-jährigen Mädchen deutlich höher liegt als die mittlere Lesekompetenz der Jungen und der Untersuchung von Mashkovskaya (2014), in der weibliche Studierende durchschnittlich höhere Ergebnisse im eingesetzten linksseitig getilgten C-Test erzielten als männliche Studierende, bemerkenswert (Mashkovskaya, 2014, S. 205; Weis, Doroganova et al., 2019, 74 ff.).

Ebenso erstaunlich ist, dass auch für das Bildungsniveau der Eltern kein signifikanter Einfluss auf das C-Test-Ergebnis nachgewiesen konnte, da vor allem die großen Schulleistungsstudien wie PISA und DESI einen großen Zusammenhang zwischen Merkmalen der sozialen Herkunft und den Schulleistungen, z. B. der Lesekompetenz in PISA (2018), aufzeigen (Baumert et al., 2001; Weis, Müller et al., 2019, 129 ff.). Die aktuellen PISA-Ergebnisse (2018) belegen zudem, dass der Anteil der Unterschiede in der Lesekompetenz, der durch die soziale Herkunft erklärt wird im Vergleich zum OECD-Durchschnitt in Deutschland deutlich höher liegt (Weis, Müller et al., 2019, S. 158).

Das Ergebnis der hier vorliegenden Untersuchung ließe einerseits vermuten, dass es – zumindest den an der Studie teilnehmenden Schulen – gelungen ist, den Zusammenhang zwischen sozialer Herkunft und Bildungserfolg (vgl. oben) bzw. hier der Sprachkompetenz aufzubrechen. Möglicherweise sind aber eher die erhobenen Daten zur Messung des sozialen Hintergrunds (also das höchste Bildungsniveau der Eltern als Indikator für die soziale Herkunft der Schüler*innen) allein nicht ausreichend: Zwar wird beispielsweise in PISA ebenfalls das Bildungsniveau und die Bildungsdauer der Eltern abgefragt, allerdings werden diese Variablen noch mit weiteren Aspekten zur Bestimmung der sozialen Herkunft ergänzt: So werden zusätzlich zu Bildungsgrad und -dauer u. a. der berufliche Status der Eltern (die berufliche Tätigkeit der Eltern unter Berücksichtigung des Einkommens, die Art des Beschäftigungsverhältnisses sowie die Weisungsbefugnis) als auch der Besitz von Wohlstandsgütern (klassisch-kulturelle als auch lernbezogene und computerbezogene Besitztümer in der Familie und deren Verfügbarkeit im familiären Umfeld Schüler*innen) als Aspekte eines sozioökonomischen und soziokulturellen Status erfasst (Weis, Müller et al., 2019, 131 ff.). Ob also der im Rahmen dieser Studie erfasste Bildungsgrad der Eltern als Indikator für den sozioökonomischen Status der Oberstufenschüler*innen zu kurz greift, bzw. ob sich unter der Hinzunahme weiterer Variablen zur Bestimmung der sozialen Herkunft, (etwa nach dem Muster von PISA) ein signifikanter Einfluss auf die mit dem C-Test gemessene Sprachkompetenz zeigen würde, müsste in weiteren Untersuchungen überprüft werden.

7. Fazit und abschließender Ausblick

Während sich standardisierte oder schul- und stufenübergreifende Untersuchungen und Verfahren zur Feststellung von sprachlichen Fähigkeiten und Fertigkeiten vor allem auf den Primarbereich und die Sekundarstufe I des deutschen Schulsystems beziehen, werden analoge Sprachstandsfeststellungen in der Sekundarstufe II nicht mehr durchgeführt (vgl. Kapitel 1). Dass eine systematische Überprüfung von sprachlichen Kompetenzen – wie sie beispielsweise u. a. im Rahmen von PISA (zur Lesekompetenz, z. B. 2018) oder in der DESI-Studie 2003/2004 in der Jahrgangsstufe 9 erfolgt (ist) – in der gymnasialen Oberstufe nicht mehr stattfindet, dass also auch entsprechende Testinstrumente nicht entwickelt wurden, legt die Vermutung nahe, dass die Ausbildung der grundlegenden sprachlichen Kompetenzen der Schüler*innen mit Ende der Sekundarstufe I als abgeschlossen und eine weitere explizite Überprüfung als entbehrlich betrachtet werden (vgl. Kapitel 1). Vor dem Hintergrund der Ergebnisse verschiedener empirischer Untersuchungen – wie beispielsweise der DESI-Studie, die zeigen, dass die sprachlichen Kompetenzen von etwa einem Drittel der Jugendlichen in der Klasse 9 deutlich hinter den curricular als Basis für die Vermittlung der Inhalte der Stufe 10 fixierten blieben (Hartig et al., 2008, S. 41; Neumann & Lehmann, Rainer, H., 2008) (vgl. Kapitel 1) – scheint dies jedoch fragwürdig und damit auch, ob das, was zumindest formal in den Kernlehrplänen für das Ende der Sekundarstufe I vorausgesetzt wird, mit dem Eintritt in die Oberstufe tatsächlich vorausgesetzt werden kann. Denn daran, dass die Schüler*innen zumindest teilweise mit förderbedürftigen Kompetenzen die Oberstufe verlassen und ein Studium aufnehmen, besteht kein Zweifel: So zeigen neuere Untersuchungen im Rahmen des Online-Sprachassessments für Studienbewer*innen (SkaLa®) an der Universität Duisburg-Essen, dass Studierende, die ein Lehramtsstudium aufnehmen, teilweise mit erheblichen Sprachkompetenzdefiziten ihren Weg an der Universität beginnen (Bremerich-Vos & Scholten-Akoun, 2016). Etwa ein Drittel dieser Studierenden-Teilgruppe kann z. T. wichtige Aspekte eines Textes nicht oder nicht korrekt wiedergeben; etwas weniger als ein Drittel der Studierenden hat erhebliche Schwierigkeiten, grammatisch und orthographisch korrekt zu schreiben (ebd., vgl. Kapitel 1).

Die hier vorgelegte Arbeit hat versucht, einen Weg aufzuzeigen, wie die Lücke in den schulübergreifenden Sprachstandsfeststellungen zwischen der Klasse 9 und der Studieneingangsphase geschlossen werden könnte. Für die Untersuchung wurde im Rahmen der Arbeit ein Instrument entwickelt und evaluiert in dem Bestreben, die schriftsprachlichen Kompetenzen der Oberstufenschüler*innen einzuschätzen. Fokussiert wurden dabei die basalen schriftsprachlichen Fertigkeiten (zur Definition dieser Fertigkeiten vgl. Kap. 2.1).

Zur Erfassung der basalen schriftsprachlichen Kompetenzen der Oberstufenschüler*innen wurde ein linksseitig getilgter C-Test entwickelt. Rechtsseitig getilgte C-Tests in der deutschen Sprache führen beim Einsatz in Schüler*innengruppen in der Oberstufe oder bei Studierenden an deutschen Schulen bzw. Universitäten (also bei der Testung von DaM-/DaZ-Sprecher*innen mit langjähriger Sprachpraxis) unabhängig vom Tilgungsrhythmus zu Deckeneffekten (Mashkovskaya, 2014; Mashkovskaya & Baur, 2016b). Sigott und Köberl sehen eine mögliche Erklärung dafür, dass die Rekonstruktion eines Wortes dann leichter fällt, wenn der Anfang eines Wortes gegeben ist, in der Veränderung des Verfahrens der Redundanzreduzierung: So konnte in Kapitel 2.3.5 zumindest exemplarisch gezeigt werden, dass das Lösen der Lücken je nach Tilgungsart (rechts- oder linksseitig) in einem deutschen C-Test in Teilen durch das Einbeziehen von Redundanzen auf unterschiedlichen Ebenen gelingen kann: In rechtsseitig getilgten C-Tests werden Redundanzen vorzugsweise auf morpho-syntaktischer Ebene genutzt (z. B. *Er singt ein Lied* → *Er sin___ ein Li___*. bzw. *Er ____gt ein ____ed*). Die Redundanzen auf der textsemantischen Ebene sind für das Lösen der linksseitig getilgten Wörter – wie in diesem Beispiel gezeigt – notwendig, um die Lücke zu schließen, da andere Ergänzungen ebenfalls einen sinnvollen Satz ergeben können, der aber erst im Gesamtkontext auf seine Sinnhaftigkeit geprüft werden kann (*Er **bedrä**ngt ein **Mitgli**ed* o.ä. wäre beispielsweise auch möglich). Die Veränderung der Strategie der Redundanzreduzierung und damit die Steigerung der Schwierigkeit in einem linksseitig getilgten C-Test funktioniert dabei über die Anzahl der redundanten Informationen, die sich auf die Lücke in einem C-Test beziehen und darüber, ob sich die redundanten Informationen vorrangig auf der Wort-, Satz- oder Textebene herstellen. So konnte Mashkovskaya (2014) bereits nachweisen, dass linksseitig getilgte Lücken,

deren Rekonstruktion ausschließlich über die Textebene gelingt, besonders schwer zu lösen sind (S. 166 ff.). Die hier in Kapitel 2.3.5 durchgeführte exemplarische Analyse, die nun den Vergleich von linksseitigen und rechtsseitigen Tilgungen fokussiert, legt nahe, dass das Lösen von linksseitig getilgten C-Tests allgemein schwieriger ist als das von rechtsseitig getilgten C-Tests, da für die Rekonstruktion der Wortanfänge weniger und andere (übergreifendere) Redundanzen genutzt werden können als für die Ergänzung am Wortende (vgl. Kapitel 2.3.5).

Diesem Muster, d. h. der linksseitigen Tilgung folgend, wurden adressatenspezifische C-Tests für die Zielkohorte von Oberstufenschüler*innen entwickelt und erprobt. Nach ersten Pre-Testungen konnte ein finales Testheft erstellt werden (vgl. Kapitel 4.1), für welches nach dem Einsatz in der Gesamtstichprobe die empirische Analyse der erhobenen Daten zeigt, dass der entwickelte C-Test mit einer Gesamtschwierigkeit von P=53,23 im mittleren Bereich liegt, d. h. für die Oberstufenschüler*innen weder zu schwierig noch zu leicht ist (vgl. Kapitel 5.2). Die Auswertung der C-Test-Daten zeigt eine annähernde Normalverteilung. Es lassen sich weder Decken- noch Bodeneffekte beobachten. Die breite Streuung der Ergebnisse lässt eine befriedigende Differenzierung zwischen den Leistungen der Oberstufenschüler*innen vermuten (vgl. Kapitel 5.2).

Zur Überprüfung der Validität des entwickelten C-Tests wurden die C-Test Daten mit den Deutschnoten der Oberstufenschüler*innen korreliert, mit dem Ergebnis, dass es einen mittleren, dabei aber hochsignifikanten Zusammenhang von r= -0,32 zwischen den C-Test-Daten und der Deutschnote gibt (vgl. Kapitel 5.4.2). Die mittlere Korrelation lässt sich so deuten, dass mit dem eingesetzten linksseitig getilgten C-Test und der Deutschnote zwar eine Schnittmenge von ähnlichen Fähigkeiten und Fertigkeiten erfasst werden, zugleich aber die Extension der insgesamt erfassten Kompetenzen sehr unterschiedlich ist: Während die Zusammensetzung der Deutschnote so angelegt ist, dass sie in komplexer Weise Bezug auf die unterschiedlichen Kompetenzbereiche des Fachs Deutsch (Lesen, Zuhören, Schreiben, Sprechen, Beurteilen, Reflektieren usw.; vgl. Kapitel 2.1.3) nimmt und somit eine Vielzahl miteinander in Verbindung stehender Kompetenzen erfasst und in einer Note bündelt (vgl. Kapitel 2.2), fokussiert der C-Test die schriftsprachliche Basis – und hier vor allem grundlegende Kompetenzen in den Bereichen Grammatik, Orthografie, Lexik, Semantik, Syntax, Leseverständnis; vgl.

Kapitel 2.1.2 und Kapitel 2.3). Die Rückmeldung zur sprachlichen Kompetenz in Klausurkorrekturen ist dabei zwar einerseits u.U. sehr genau und differenzierter als es der C-Test leistet, die Bedeutung der mit dem C-Test gemessenen Teilkompetenz im Verhältnis zu anderen in der Gesamtnote erfassten Kompetenzen jedoch marginal (vgl. Kapitel 2.1) Dieses Ergebnis mag auf den ersten Blick zwar wenig überraschend sein, ist in seiner Aussage vor dem Hintergrund der in der Arbeit gestellten Frage, ob und zu welchem Grad die Oberstufenschüler*innen über ausgebaute schriftsprachliche Basiskompetenzen verfügen, allerdings von besonderer Relevanz, da es verdeutlicht, dass gerade diese basalen schriftsprachlichen Kompetenzen in die Deutschnote zwar als Facetten eines komplexen Kompetenzfächers indirekt miteinfließen, aber eben nicht ohne Weiteres oder nur indirekt aus dieser herauspräpariert werden können. Um nun aber die basalen schriftsprachlichen Kompetenzen – über die zu verfügen einerseits zur aktiven gesellschaftlichen Teilhabe befähigt und andererseits als notwendige Voraussetzung für die erfolgreiche Bewältigung von Schule, Berufsausbildung und Studium gelten – gezielt und explizit abbilden zu können, sollten Aussagen zum Leistungsstand im Fach Deutsch um eine explizite Rückmeldung zum Sprachstand über den standardisierten C-Test ergänzt werden. C-Test-Resultate liefern somit zusätzliche Informationen über den basalen Schriftsprachkompetenzstand der Schüler*innen in der Oberstufe und ermöglichen das systematische Aufdecken von möglichen Defiziten auch im Bereich der gymnasialen Oberstufe, denen dann mit Förderangeboten begegnet werden kann.

Zur weiteren Validitätsüberprüfung wurde der entwickelte C-Test mit dem in der Oberstufe einsetzbaren standardisierten Lesetest *LGVT 6-12* korreliert. Im Ergebnis zeigt die Korrelationsanalyse auch hier einen statistisch signifikanten mittleren Zusammenhang von r= 0,40. Bei der Interpretation der mittleren Korrelation zwischen den beiden im Rahmen dieser Untersuchung eingesetzten Tests lassen sich ähnliche Faktoren erkennen: Wie bei der Deutschnote verhält es sich mit dem im Rahmen dieser Arbeit eingesetzten Lesetest und dem C-Test so, dass einerseits ähnliche Kompetenzen in Bezug auf rezeptive Fähigkeiten und Fertigkeiten erfasst werden, mit dem C-Test andererseits jedoch basale schriftsprachliche Kompetenzen gemessen werden, die über rein rezeptive Kompetenzen hinaus gehen (vgl. Kapitel 4.1.1, Kapitel 5.4.1 und Kapitel

6). Dieses Ergebnis verdeutlicht, dass der linksseitig getilgte C-Test, wie in Kapitel 2.3.5 ausführlich erläutert, zwar verstärkt lexikalisch-semantische Fertigkeiten und Fähigkeiten fokussiert, dadurch jedoch kein reines Leseverstehen, sondern als integrativer Sprachtest nach wie vor eine allgemeine basale Sprachkompetenz, die ebenso unterschiedliche Bereiche wie beispielsweise Grammatik, Orthografie, Syntax, etc. inkludiert, erfasst.

Zurückkommend auf die in der hier vorliegenden Arbeit zentrale Frage, zu welchem Ausmaß die Schüler*innen der gymnasialen Oberstufe über basale schriftsprachliche Kompetenzen verfügen, weist auf der Grundlage der im Rahmen dieser Studie erhobenen Daten insgesamt ein Viertel der getesteten Oberstufenschüler*innen einen Förderbedarf im Bereich der mit dem C-Test gemessenen grundlegenden schriftsprachlichen Fähigkeiten und Fertigkeiten auf. Dieses Ergebnis ähnelt damit den oben genannten Befunden über das Vorliegen von Defiziten im Bereich der Sprachkompetenz von 9. Klässler*innen (DESI) (DESI-Konsortium, 2006a; Klieme, 2006) ebenso wie von Studienbewerber*innen (SkaLa®) (Bremerich-Vos & Scholten-Akoun, 2016). Damit erhalten die Zweifel an der eingangs vermuteten Annahme, die Schulsysteme betrachten den elementaren Schriftspracherwerb längstens mit dem Abschluss der Sekundarstufe I als abgeschlossen, eine empirische Stützung. Für etwa ein Viertel der Schüler*innen erweist sich die Annahme schlicht als irrig. Die schriftsprachlichen Kompetenzen nehmen zwar – eine gute Nachricht – insgesamt im Verlauf der Beschulung in der Oberstufe zu, allerdings weisen die vorliegenden Ergebnisse für zumindest einen Teil der Schüler*innen sowohl in der Einführungsphase als auch in der Qualifikationsphase 1 in der untersuchten Kohorte Förderbedarf im Bereich der basalen schriftsprachlichen Kompetenzen aus (vgl. Kapitel 5.5.1).

Studien von Baur und Scholten Akoun (2010) haben gezeigt, dass Proband*innen in Abhängigkeit beispielsweise vom Bildungshintergrund der Eltern, vom Geschlecht oder vom Migrationshintergrund unterschiedliche Leistungen in Sprachstandsmessungen erzielen (teilweise mit ähnlichen Befunden, vgl. Bremerich-Vos & Scholten-Akoun, 2016; DESI-Konsortium, 2006b; Klieme, 2006; Neumann, 2014; Scholten-Akoun et al., 2013; Scholten-Akoun et al., 2014). Ob sich dieser Befund der Gruppenspezifik auch auf der Basis der im Rahmen dieser Arbeit erhobenen C-Test-Leistungen in der hier

untersuchten Kohorte replizieren ließ, war ein die Untersuchung zusätzlich leitender Fragenfächer (vgl. Kapitel 1 und 3). Durch die Erfassung der Hintergrundvariablen (z. B. zur Sprachverwendung, zur Schulform, zum Bildungsgrad der Eltern etc.) (vgl. Kapitel 4.1.3) konnte in dieser Arbeit bestätigt werden, dass bestimmten Gruppen ausweislich der C-Test-Ergebnisse insgesamt häufiger Förderbedarfe zugeschrieben werden als anderen. Das gilt aber nicht für das Geschlecht. Die oftmals proklamierten Leistungsunterschiede zwischen Jungen und Mädchen in Bezug auf sprachliche Kompetenzen (vgl. z. B. PISA 2015, 2018) lassen sich für die gymnasiale Oberstufe auf der Basis der hier erhobenen C-Testergebnisse nicht bestätigen: Mädchen schneiden im direkten Vergleich zwar minimal besser ab als Jungen, allerdings relativiert sich dieser geschlechtsspezifische Unterschied unter Hinzunahme mehrerer Faktoren im Regressionsmodell und hat dann keinen signifikanten Einfluss mehr auf die mit dem C-Test gemessene Sprachkompetenz (vgl. Kapitel 5.5.2 und Kapitel 5.6). Ebenso ist – entgegen den Ergebnissen der großen Schulleistungsstudien wie PISA und DESI, die einen starken Zusammenhang zwischen Merkmalen der sozialen Herkunft und den Schulleistungen aufzeigen (Baumert et al., 2001; Weis, Müller et al., 2019, 129 ff.) – festzuhalten, dass auch für das Bildungsniveau der Eltern in der vorliegenden Studie kein signifikanter Einfluss auf das C-Test-Ergebnis nachgewiesen konnte (vgl. Kapitel 5.6). Gezeigt werden konnte jedoch, dass die Rahmenbedingungen der jeweiligen Schulstandorte insgesamt mit den Kompetenzen der getesteten Schüler*innen zusammenhängen (vgl. Kapitel 5.5.5): In Bezug auf die einzelnen Standorttypen der Schulen unterschieden sich die Leistungen der Oberstufenschüler*innen dabei allerdings nur an den Standorttypen 1-3 signifikant von den Leistungen an Schulen mit den Standorttypen 4 und 5 (vgl. Kapitel 5.5.5). Unter der Hinzunahme anderer (schul-, leistungs- und personenbezogener) Variablen hat das Merkmal *Schulstandorttyp* im Regressionsmodell einen nur schwachen Einfluss auf die gemessene Sprachkompetenz und eignet sich daher nur bedingt als Prädiktorvariable. An dieser Stelle ist jedoch auf die Verzerrung hinsichtlich der Verteilung der Daten an den einzelnen Standorttypen hinzuweisen, die das Ergebnis möglicherweise beeinflusst und somit nicht generalisierbar macht (vgl. Kapitel 5.5.5 und 6). Ebenso ist darauf hinzuweisen, dass mit Beginn des Schuljahres 2021/2022 die Standorttypen unter Berücksichtigung eines neu erfassten schulscharfen Sozialindex berechnet und auf dieser Grundlage neu zugeordnet

wurden[107]. Eine solche Zuordnung unter Berücksichtigung eines schulscharfen Sozialindexes lag für den Zeitraum der Datenerhebung dieser Arbeit allerdings noch nicht vor. Signifikant unterschiedlich hingegen sind die Ergebnisse in Bezug auf die Schulform: So liegen die durchschnittlich erbrachten Leistungen an den Gymnasien deutlich über den Leistungen der Gesamtschulen, an denen der Anteil von Schüler*innen mit großem Förderbedarf in der hier vorliegenden Stichprobe dreifach höher ist als an den Gymnasien (vgl. Kapitel 5.5.6). Bedingt durch die verhältnismäßig geringe Stichprobengröße von Schüler*innen an Gesamtschulen (vor allem im Vergleich zur Kohorte der Gymnasialschüler*innen; vgl. Kapitel 5.1 und Kapitel 5.5.6) müsste der beschriebene Trend in den Ergebnissen durch weitere Erhebungen stabilisiert werden (vgl. Kapitel 6).

In Bezug auf die Frage nach gruppenspezifischen Unterschieden zwischen Oberstufenschüler*innen ist abschließend festzuhalten, dass für die hier untersuchte Stichprobe ein systematischer Zusammenhang zwischen dem erfassten Merkmal *Mehrsprachigkeit* (vgl. Kapitel 4.1.3 und 5.1.2) und niedrig(er)en C-Test-Leistungen ausgemacht werden konnte: Im Vergleich zur Gruppe der monolingual deutschsprachigen Schüler*innen finden sich mehrsprachige Schüler*innen doppelt so häufig in Leistungsbereichen, für die der C-Test Förderbedarf ausweist. Von den in dieser Arbeit erfassten Hintergrundmerkmalen hat die Variable *Mehrsprachigkeit* also in der hier vorliegenden Untersuchung den insgesamt größten Einfluss auf das C-Test-Ergebnis (vgl. Kapitel 5.6).

Eingrenzend ist festzuhalten, dass das Abschneiden der Oberstufenschüler*innen in dem im Rahmen dieser Arbeit entwickelten C-Test selbstverständlich kein Zeugnis über deren generelle Gestaltungsfähigkeit von Sprache ausstellt. Aussagen dieser Reichweite müssten sich auf ein stark verändertes Untersuchungsdesign stützen können - alleine mit C-Tests sind etwa die freien sprachlichen Gestaltungsmöglichkeiten bestenfalls indirekt erfassbar. Denkbar wäre aber – etwa in Anlehnung an SkaLa® (vgl. Kapitel 2.1.1 – hier insbesondere die uneindeutige Korrelation zwischen Ergebnissen der Freitext- und der C-Test-Aufgabe) – die Ergänzung des Testinstruments um weitere Lese-

[107] Ausführliche Informationen hierzu unter https://www.schulministerium.nrw/sozialindex.

und/oder freie Schreibaufgaben. Ebenso konnten die aufgezeigten Disparitäten der mit dem C-Test gemessenen schriftsprachlichen Kompetenzen der Oberstufenschüler*innen nicht vollständig über die in den Regressionsmodellen erfassten Indikatoren erklärt werden (vgl. Kapitel 5.6). So müsste also in weiteren Studien untersucht werden, ob und inwieweit sich zusätzliche kontextuelle Variablen, wie beispielsweise die Wirksamkeit der Lehrkräfte, die Qualität des Unterrichts, die häusliche (Lern-)Umgebung, motivationale Faktoren, andere sprachliche Faktoren (wie beispielsweise ein Mangel an bildungsrelevanten familiären sprachlichen Ressourcen), Sprachfördermaßnahmen, usw. auf das C-Test-Ergebnis bzw. die damit gemessenen die Sprachkompetenz auswirken.

Die vorliegende Arbeit hat darüber hinaus Fragen auf ganz unterschiedlichen Ebenen aufgeworfen, die in weiteren Studien und Untersuchungen aufgegriffen werden könnten:

- Auf empirischer Ebene wäre dabei besonders wichtig, für ein vollständiges Bild der basalen schriftsprachlichen Fertigkeiten und Fähigkeiten in der gymnasialen Oberstufe erstens die Stichprobe an den Gesamtschulen und an den einzelnen (neu zugeordneten, vgl. oben) Standorttypen zu erweitern ebenso wie die Sprachkompetenz auch in der Qualifikationsphase 2 systematisch mit dem C-Test zu erfassen und zweitens um Daten anderer Schulformen (wie beispielsweise Abendgymnasien oder Berufskollegs, an denen die gymnasiale Oberstufe besucht und mit dem Abitur abgeschlossen werden kann) zu ergänzen.

- Auf testtheoretischer Ebene müsste durch weitere Validitätsprüfungen genauer bestimmt werden, welche Fähigkeiten und Fertigkeiten linksseitige C-Tests erfassen. Ebenso wäre es wichtig, weitere Faktoren zur Bestimmung eines sozioökonomischen Status als Moderatorvariablen einzubeziehen.

- Auf schulsystematischer Ebene wäre es weiterhin interessant Sprachdiagnostik mit dem C-Test – der bei einem einmaligen Einsatz wie in der hier vorliegenden Arbeit den Schüler*innen und Lehrkräften zunächst auf individueller Ebene Auskünfte über mögliche Defizite liefert – um Längsschnittstudien zu ergänzen, um sprachliche Entwicklungen auf individueller Ebene sichtbar zu machen.

Zudem ließe sich der C-Test zur Evaluation von Sprachfördermaßnahmen einsetzen ebenso wie zur Überprüfung der Leistungsfähigkeit der Schulsysteme, da der Einsatz eines standardisierten, normierten C-Tests Interlerngruppenvergleiche über die Schul-, Schulform- und Landesregionsgrenzen hinaus möglich macht.

Die beunruhigenden Ergebnisse zum Gesamtanteil an defizitär ausgebauten schriftsprachlichen Kompetenzen im Bereich der gymnasialen Oberstufe aber auch zum Einfluss des Migrationshintergrundes auf die sprachliche Performanz machen deutlich, dass es auch in der Sekundarstufe II einen nicht unerheblichen Bedarf an Sprachförderung gibt. Die vorliegende Untersuchung hat mitunter gezeigt, dass die Schüler*innen der gymnasialen Oberstufe im Hinblick auf ihre sprachlichen Voraussetzungen und ihren potentiellen Förderbedarf insgesamt eine heterogene Lerngruppe bilden und die Personengruppe mehrsprachiger Schüler*innen, was ihre sprachlichen Voraussetzungen betrifft, ebenso wenig eine homogene Gruppe ist. Die Konsequenz der hier vorgestellten Ergebnisse sollte also sein, Konzepte zu bedarfsgerechten, durchgängigen Sprachförderangeboten (auch) für die Sekundarstufe II (weiter-)zu entwickeln, in den Schulen zu implementieren und fortlaufend zu evaluieren, damit Bildungsgerechtigkeit nicht nur als Forschungsvorhaben existiert sondern Chancengleichheit überhaupt gewährleistet wird und folglich allen Schüler*innen einen erfolgreichen Start in der akademischen und beruflichen Ausbildung ermöglicht.

Literaturverzeichnis

Ahrenholz, B. (2010). Erstsprache - Zweitsprache - Fremdsprache. In B. Ahrenholz, I. Oomen-Welke & W. Ulrich (Hrsg.), *Deutschunterricht in Theorie und Praxis: DTP; Handbuch zur Didaktik der deutschen Sprache und Literatur in elf Bänden / hrsg. von Winfried Ulrich ; Bd. 9. Deutsch als Zweitsprache* (2. Aufl., S. 3–16). Schneider-Verl. Hohengehren.

Aitchison, J. (1987). *Words in the Mind: An Introduction to the Mental Lexikon*. Basil Blackwell.

Alderson, J. C. (1979). The Cloze Procedure and Proficiency in English as a Foreign Language. *TESOL Quarterly*, *13*(2), 219. https://doi.org/10.2307/3586211

Alderson, J. C. (1980). Native and non-native speaker performance on cloze tests. *Language Learning*, *30*(1), 59–76.

ALTE. (2005). *Handreichungen für Testautoren.*

American Educational Research Association. (2014). *Report and Recommendations for the Reauthorization of the Institute of Education Sciences*. American Educational Research Association.

Ammon, U. (1991). *Die internationale Stellung der deutschen Sprache*. de Gruyter. http://www.degruyter.com/search?f_0=isbnissn&q_0=9783110886498&searchTitles=true

Arras, U., Eckes, T. & Grotjahn, R. (2002). C-Tests im Rahmen des "Test Deutsch als Fremdsprache" (TestDaF): Erste Forschungsergebnisse. In R. Grotjahn (Hrsg.), *Fremdsprachen in Lehre und Forschung (FLF): Bd. 32. Der C-Test: Theoretische Grundlagen und praktische Anwendungen* (Bd. 4, S. 175–209). AKS-Verl.

Artelt, C., McElvany, N., Christmann, U., Richter, T., Groeben, N., Köster, J., Schneider, W., Stanat, P., Ostermeier, C., Schiefele, U., Valtin, R. & Ring, K. (2005). *Expertise - Förderung von Lesekompetenz.*

Asano, Y. (2010). *Misst ein C-Test das, was er messen soll? Konstruktdefinition als Basis für die Validierung von C-Tests*.

Asano, Y. (2014). C-Tests und 'allgemeine Sprachkompetenz': Theoretische Überlegungen und empirische Analysen. In R. Grotjahn (Hrsg.), *Language testing and evaluation: Bd. 34. Der C-Test: Aktuelle Tendenzen: The C-test: Current trends* (Bd. 34, S. 39–52). Peter Lang.

Bachmann, L. F. (1990). *Fundamental considerations in language testing*. Oxford University Press.

Baumert, J., Carstensen, C. H. & Siegle, T. (2005). Wirtschaftliche, soziale und kulturelle Lebensverhältnisse und regionale Disparitäten des Kompetenzerwerbs. In M. Prenzel, J. Baumert, W. Blum, Lehmann, Rainer, H., 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. 323–366). Waxmann.

Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, K.-J. & Weiß, M. (2001). *PISA 2000*. VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-322-83412-6

Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Tillmann, K.-J. & Weiß, M. (1999). *Internationales und nationales Rahmenkonzept für die Erfassung von Lesekompetenz in PISA.* Max-Planck-Institut für Bildungsforschung. https://www.mpib-berlin.mpg.de/Pisa/KurzFrameworkReading.pdf

Baur, R. S., Chlosta, C. & Goggin, M. (2012). *Doppel-Klick* (1. Aufl., 1. Dr). Cornelsen.

Baur, R. S., Goggin, M. & Wrede-Jackes, J. (2013a). *Der C-Test: Einsatzmöglichkieten im Bereich DaZ*. https://www.uni-due.de/imperia/md/content/prodaz/c_test_einsatzmoeglichkeiten_daz.pdf

Baur, R. S., Goggin, M. & Wrede-Jackes, J. (2013b). *Der C-Test: Einsatzmöglichkeiten im Bereich DaZ.* Universität Duisburg-Essen. https://www.uni-due.de/imperia/md/content/prodaz/c_test_einsatzmoeglichkeiten_daz.pdf

Baur, R. S., Grotjahn, R. & Spettmann, M. (2006). Der C-Test als Instrument der Sprachstandserhebung und Sprachförderung. In J. P. Timm (Hrsg.), *Fremdsprachenlernen und Fremdsprachenforschung: Kompetenzen, Standards, Lernformen, Evaluation* (S. 389–406). Narr.

Baur, R. S. & Mashkovskaya, A. (2015). C-Test-Kritik reviewed. In J. Böcker & A. Stauch (Hrsg.), *Fremdsprachen lebenslang lernen. Konzepte aus der Sprachlehrforschung - Impulse für die Praxis: Festschrift für Karin Kleppin* (S. 435–448).

Baur, R. S. & Meder, G. (1994). C-Tests zur Ermittlung der globalen Sprachfähigkeiten im Deutschen und in der Muttersprache bei ausländischen Schülern in der Bundesrepublik Deutschland. In R. Grotjahn (Hrsg.), *Der C-Test: Theoretische Grundlagen und praktische Anwendungen. Bd. 2* (Bd. 2, S. 151–178). Brockmeyer.

Baur, R. S. & Scholten-Akoun, D. (2010). *Deutsch als Zweitsprache in der Lehrerausbildung: Bedarf - Umsetzung - Perspektive.* Stiftung Mercator. www.mercator-foerderun-terricht.de/projekt/publikationen.html

Baur, R. S. & Spettmann, M. (2007). Screening - Diagnose - Förderung: Der C-Test im Bereich DaZ. In B. Ahrenholz (Hrsg.), *Deutsch als Zweitsprache: Voraussetzungen und Konzepte für die Förderung von Kindern und Jugendlichen mit Migrationshintergrund* (3. Aufl., S. 95–110). Fillibach; Fillibach-Verl.

Baur, R. S. & Spettmann, M. (2008a). Kompetenzen testen - leicht gemacht: C-Tests für die Orientierungsstufe. In C. Bainski (Hrsg.), *Handbuch Sprachförderung* (S. 123–131). Neue-Dt.-Schule-Verl.-Ges.

Baur, R. S. & Spettmann, M. (2008b). Sprachstandsmessung und Sprachförderung mit dem C-Test. In B. Ahrenholz & I. Oomen-Welke (Hrsg.), *Deutschunterricht in Theorie und Praxis: Band 9. Deutschunterricht in Theorie und Praxis: Handbuch zur Didaktik der deutschen Sprache und Literatur in elf Bänden* (S. 430–441). Schneider Hohengehren.

Baur, R. S. & Spettmann, M. (2009). Der C-Test als Instrument der Sprachdiagnose und Sprachförderung. In H.-J. Roth, H. Reich & D. Lengyel (Hrsg.), *FörMig-Edition: Bd. 5. Von der Sprachdiagnose zur Sprachförderung* (S. 115–127). Waxmann.

Beck, B. & Klieme, E. (2007). *Beltz Pädagogik*. *Sprachliche Kompetenzen: Konzepte und Messung ; DESI-Studie (Deutsch-Englisch-Schülerleistungen-International)*. Beltz.

Becker-Mrotzek, M. (2017). Das Schreiben zurückholen - Anmerkung zur Funktion des materialgeschützen Schreibens in den Bildungsstandards. *Didaktik Deutsch*, *22*(42), 4–11. https://www.pedocs.de/volltexte/2019/16875/pdf/Deutsch_Didaktik_2017_42_Becker-Mrotzek_Das_Schreiben_zurueckholen.pdf

Becker-Mrotzek, M., Kämper-van den Boogaart, M., Köster, J., Stanat, P. & Gippner, G. (Hrsg.). (2015). *Bildungsstandards aktuell: Deutsch in der Sekundarstufe II*. Schroedel.

Becker-Mrotzek, M. & Schindler, K. (2007). Schreibkompetenz modellieren. In M. Becker-Mrotzek & K. Schindler (Hrsg.), *Kölner Beiträge zur Sprachdidaktik / Reihe A: Bd. 5. Texte schreiben* (S. 7–27). Gilles & Francke.

Becker-Mrotzek, M. & Schindler, K. (2007). Schreibkompetenz modellieren. In M. Becker-Mrotzek & K. Schindler (Hrsg.), *Kölner Beiträge zur Sprachdidaktik / Reihe A: Bd. 5. Texte schreiben* (S. 7–26). Gilles & Francke.

Beinborn, L., Zesch, T., Gurevych & Iryna (2014). Predicting the Difficulty of Language Proficiency Tests. *Transactions of the Association for Computational Linguistics* (2), 517–529.

Bethke, H. (27. März 2014). Studenten können keine Rechtschreibung mehr: Sprachnotstand an der Uni. *F.A.Z*. https://www.faz.net/aktuell/beruf-chance/campus/sprachnotstand-an-der-uni-studenten-koennen-keine-rechtschreibung-mehr-12862242.html

Biermann, H. (1999). Das Problem des Übersetzens. In H. Biermann, B. Schurf, G. Brenner, D. Erlach, K. Fingerhut, M. Fingerhut & C. Grunow (Hrsg.), *Texte, Themen und Strukturen: Deutschbuch für die Oberstufe*. Cornelsen.

Blanz, M. (2015). *Forschungsmethoden und Statistik für die Soziale Arbeit: Grundlagen und Anwendungen*. W. Kohlhammer Verlag.

Bolton, S. (1985). *Die Gütebestimmung kommunikativer Tests*. G. Narr.

Borgards, S. & Raatz, U. (2002). Sind C-Tests trainierbar? In R. Grotjahn (Hrsg.), *Fremdsprachen in Lehre und Forschung (FLF): Bd. 32. Der C-Test: Theoretische Grundlagen und praktische Anwendungen* (S. 157–174). AKS-Verl.

Bormuth, J. R. (1967). Comparable Cloze and Multiple-Choice Comprehension Test Scores. *Journal of Reading* (10), 291–299. https://www.jstor.org/stable/40009351

Bormuth, J. R. (1969). Factor Validity of Cloze Tests as Measures of Reading Comprehension Ability. *Reading Research Quarterly*, *4*(3), 358. https://doi.org/10.2307/747144

Bortz, J. & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler: Mit 163 Tabellen* (7. Aufl.). *Springer-Lehrbuch*. Springer.

Bos, W., Lankes, E. M., Prenzel, M., Schwippert, K., Walther, G. & Valtin, R. (2003). *Erste Ergebnisse aus IGLU: Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich*. Waxmann.

Breitenbach, E. (2006). Einführung in die pädagogisch-psychologische Diagnostik. In S. Ellinger (Hrsg.), *Lehren und Lernen mit behinderten Menschen: Bd. 9. Grundstudium Sonderpädagogik* (2. Aufl., S. 114–141). Athena.

Breitenbach, E. (2007). *Förderdiagnostik: Theoretische Grundlagen und Konsequenzen für die Praxis* (2. Aufl.). Edition bentheim.

Bremerich-Vos, A. (2016). Das Zusammenfassen eines argumentativen Textes als Schreibaufgabe. In A. Bremerich-Vos & D. Scholten-Akoun (Hrsg.), *Schriftsprachliche Kompetenzen von Lehramtsstudierenden in der Studieneingangsphase: Eine empirische Untersuchung* (1. Aufl., S. 35–64). Schneider Hohengehren.

Bremerich-Vos, A. & Scholten-Akoun, D. (Hrsg.). (2016). *Schriftsprachliche Kompetenzen von Lehramtsstudierenden in der Studieneingangsphase: Eine empirische Untersuchung* (1. Auflage). Schneider Hohengehren.

Bremerich-Vos, A. & Weirich, S. (2016). Ergebnisse zur Lese-/ Schreibaufgabe. In A. Bremerich-Vos & D. Scholten-Akoun (Hrsg.), *Schriftsprachliche Kompetenzen von Lehramtsstudierenden in der Studieneingangsphase: Eine empirische Untersuchung* (1. Aufl., S. 165–190). Schneider Hohengehren.

Browman, C. P. (1978). Tip of the tongue and slip of the ear: Implications for language processing. *UCLA Working Papers in Phonetics* (42).

Bühner, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion* (2., aktualisierte Aufl.). Pearson Studium.

Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion* (3. Aufl.). *PS Psychologie*. Pearson Studium.

Cantone-Altıntaş, K. & Haberzettl, S. (2009). "Ich bin dagegen warum sollte man den kein Handy mitnehmen" - zur Bewertung argumentativer Texte bei Schülern mit Deutsch als Zweitsprache. In K. Schramm & C. Schroeder (Hrsg.), *Reihe Mehrsprachigkeit: Band 23. Empirische Zugänge zu Spracherwerb und Sprachförderung in Deutsch als Zweitsprache* (S. 43–65). Waxmann.

Chlosta, C. & Ostermann, T. (2005). Warum fragt man nach der Herkunft, wenn man die Sprache meint? Ein Plädoyer für eine Aufnahme sprachbezogener Fragen in demographische Untersuchungen. In *Bildungsreform Band: Bd. 14. Migrationshintergrund von Kindern und Jugendlichen: Wege zur Weiterentwicklung der amtlichen Statistik* (1. Aufl., S. 55–66). dp Moser.

Chudaske, J. (2012). *Sprache, Migration und schulfachliche Leistung: Einfluss sprachlicher Kompetenz auf Lese-, Rechtschreib- und Mathematikleistungen* (1. Auflage). VS Verlag.

Cleary, C. (1988). The C-Test in English: left-hand deletions. *RELC Journal*, *19*(2), 26–35. https://doi.org/10.1177/003368828801900203

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.

Dahmen, M. (2014). Einfluss der Globalisierung. *Finale. Prüfungstraining 2015 Zentrale Klausur am Ende der Einführungsphase NRW Deutsch*.

Daller, H., van Hout, R. & Treffers-Daller, J. (2003). Lexical richness in the spontaneous speech of bilinguals. *Applied Linguistics*, *24*, 197–222.

Darnell, D. K. (1986). *The development of an English Language proficiency test of foreign students using a clozentropy procedure.* Univerity of Colorado.

Der Online-Einstufungstest Deutsch als Fremdsprache (onDaF): Theoretische Grundlagen, Konstruktion und Validierung. https://www.testdaf.de/fileadmin/Redakteur/PDF/Forschung-Publikationen/Eckes_onDaF_2010.pdf

DESI-Konsortium (Hrsg.). (2006a). *Unterricht und Kompetenzerwerb in Deutsch und Englisch: Zentrale Befunde der Studie Deutsch Englisch Schülerleistungen International (DESI)*. https://www.dipf.de/de/forschung/aktuelle-projekte/pdf/biqua/desi-zentrale-befunde

DESI-Konsortium. (2006b). *Unterricht und Kompetenzerwerb in Deutsch und Englisch: ZusammenZentrale Befunde der Studie Deutsch-Englisch-Schülerleristungen International (DESI).* Deutsches Institut für International Pädagogische Forschung. https://www.dipf.de/de/forschung/aktuelle-projekte/pdf/biqua/desi-zentrale-befunde

DESI-Konsortium (Hrsg.). (2008). *Beltz Pädagogik*. *Unterricht und Kompetenzerwerb in Deutsch und Englisch: Ergebnisse der DESI-Studie*. Beltz Verlag.

Dirim, İ. & Mecheril, P. (2010). Die Sprache(n) der Migrationsgesellschaft. In P. Mecheril, A. Kalpaka, C. Melter, İ. Dirim & M. d. M. Castro Varela (Hrsg.), *Studium Paedagogik. Migrationspädagogik* (S. 99–120). Beltz Verlag.

Dittmann, J. (2002). Wörter im Geist. Das mentale Lexikon. In J. Dittmann & C. Schmidt (Hrsg.), *Rombach Grundkurs: Bd. 5. Über Wörter: Grundkurs Linguistik* (1. Aufl., S. 283–310). Rombach.

Döring, N. & Bortz, J. (2016). Stichprobenziehung. In Döring (Hrsg.), *Springer-Lehrbuch. Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (Bd. 46, S. 291–319). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-41089-5_9

Dörnyei, Z. & Katona, L. (1992). Validation of the C-test amongst Hungarian EFL learners. *Language Testing*, *9*(2), 187–206. https://doi.org/10.1177/026553229200900206

Durrer, F. & Heine, C. (1995). Abiturienten 90 in Studium, Berufsausbildung und Erwerbstätigkeit: Einstellungen und Bewertungen 3 1/2 Jahre nach Schulabgang. *HIS-Kurzinformation. A*, *12*.

Dürscheid, C., Elspaß, S. & Ziegler, A. (2018). Grundlagenartikel Wortbildung. In A. Ziegler & A. Thurner (Hrsg.), *Variantengrammatik des Standarddeutschen: Ein Online-Nachschlagewerk*. http://www.variantengrammatik.net/docs/Grundlagenartikel%20Wortbildung.pdf

Eckes, T. (2006). Rasch-Modelle zur C-Test-Skalierung. In R. Grotjahn (Hrsg.), *Language testing and evaluation: Bd. 6. Der C-Test: Theorie, Empirie, Anwendungen: The C-Test: Theory, empirical research, applications* (S. 1–44). P. Lang.

Eckes, T. (2007). Konstruktion und Analyse von C-Tests mit Ratingskalen-Rasch-Modellen. *Diagnostica*, *53*(2), 68–82. https://doi.org/10.1026/0012-1924.53.2.68

Eckes, T. (2010). Der Online-Einstufungstest Deutsch als Fremdsprache (onDaF): Theoretische Grundlagen, Konstruktion und Validierung. In R. Grotjahn (Hrsg.), *Language testing and evaluation: Bd. 18. Der C-Test: Beiträge aus der aktuellen Forschung* (S. 125–192). Peter Lang.

Eckes, T. (2014). Die onDaF-TestDaF-Vergleichsstudie: Wie gut sagen Ergebnisse im onDaF Erfolg oder Misserfolg beim TestDaF vorher? In R. Grotjahn (Hrsg.), *Language testing and evaluation: Bd. 34. Der C-Test: Aktuelle Tendenzen: The C-test: Current trends* (Bd. 34, S. 134–162). Peter Lang.

Eckes, T. (2016). Exploratorische und konformatorische Faktorenanalyse. In D. Caspari, F. Klippel, M. Legutke & K. Schramm (Hrsg.), *Forschungsmethoden in der Fremdsprachendidaktik: Ein Handbuch* (S. 341–355). Narr Franke Attempo.

Eckes, T. & Grotjahn, R. (2006). A closer look at the construct validity of C-tests. *Language Testing*, *23*(3), 290–325. https://doi.org/10.1191/0265532206lt330oa

Ehlich, K. (1983). Text und sprachliches Handeln. In J. Assmann, A. Assmann & C. Hardmeier (Hrsg.), *Beiträge zur Archäologie der literarischen Kommunikation. Schrift und Gedächtnis* (S. 24–43). Wilhelm Fink.

Ehlich, K., Steets, A. & Traunspurger, I. (2000). *Schreiben für die Hochschule: eine annotierte Bibliographie*. P. Lang.

Eichinger, L. M. (2000). *Deutsche Wortbildung: Eine Einführung*. *Narr Studienbücher*. Narr.

Elsen, H. (2011). *Grundzüge der Morphologie des Deutschen*. *De Gruyter Studium*. de Gruyter.

Europarat. (2001). *Gemeinsamer europäischer Referenzrahmen für Sprachen: Lernen, lehren, beurteilen*. Niveau A1, A2, B1, B2, C1, C2. Langenscheidt.

Exmatrikuliertenbefragung 2010 bis 2014: Aggregation über fünf Jahrgänge. (2015). https://www.uni-konstanz.de/typo3temp/secure_downloads/67357/0/9cfd9930b206720c5291c465601286f1e6abc01c/ExMat_5Jahre_Bericht.pdf

Fahrmeir, L., Heumann, C., Künstler, R., Pigeot, I. & Tutz, G. (2016). *Statistik: Der Weg zur Datenanalyse* (8. Aufl.). *Springer-Lehrbuch*. Springer Spektrum.

Feilke, H. (2010). Schriftliches Argumentieren zwischen Nähe und Distanz am Beispiel wissenschaftlichen Schreibens. In V. Ágel & M. Henning (Hrsg.), *Nähe und Distanz im Kontext variationslinguistischer Forschung* (S. 209–231). de Gruyter.

Feilke, H. (2013). Bildungssprache und Schulsprache - am Beispiel literal-argumentativer Kompetenzen. In M. Becker-Mrotzek, K. Schramm, E. Thürmann & H. J. Vollmer (Hrsg.), *Sprache im Fach. Sprachlichkeit und fachliches Lernen* (S. 113–130). Waxmann.

Feilke, H., Köster, J. & Steinmetz, M. (Hrsg.). (2013). *Textkompetenzen in der Sekundarstufe II* (1. Aufl.). Fillibach bei Klett.

Fischer, G. H. (1974). *Einführung in die Theorie psychologischer Tests*. Huber.

Fisseni, H.-J. (2004). *Lehrbuch der psychologischen Diagnostik: Mit Hinweisen zur Intervention* (3., überarb. u. erw. Aufl.). Hogrefe.

Fix, G. & Dittmann, J. (2008). Exzerpieren. Eine empirische Studie an Exzerpten von GymnasialschülerInnen der Oberstufe. *Linguistik Online*, *33*(1), 17–71. https://www.linguistik-online.net/33_08/fixDittmann.pdf

Flade, F. (16. Dezember 2016). BKA-Bewerber scheitern am Deutschtest: Interview mit BKA-Präseident Holger Münch. *Spiegel Online,* 2016(49).

Fleischer, W. & Barz, I. (2012). *Wortbildung der deutschen Gegenwartssprache* (3. Aufl.). *De Gruyter Studium*. de Gruyter. https://doi.org/10.1515/9783110256659

Frohn, P. (2018). Wie sollen Lehrkräfte vermitteln, was sie selbst nicht können? Zeitungsakrtikel vom 26.01.2018. *Frankfurter Allgemeine* (Blogseminar). https://blogs.faz.net/blogseminar/wie-sollen-lehrkraefte-vermitteln-was-sie-selbst-nicht-koennen/

Furchner, I., Ruhmann, G. & Tente, C. (1999). Von der Schreibberatung für Studierende zur Lehrberatung fürDozenten. In O. Kruse, E.-M. Jakobs & G. Ruhmann (Hrsg.), *Hochschulwesen. Schlüsselkompetenz Schreiben: Konzepte, Methoden, Projekte für Schreibberatung und Schreibdidaktik an der Hochschule* (S. 61–71). Luchterhand. http://www.tl.rwth-aachen.de/uploads/Publikationen/jakobs_eva-maria-kruse_otto-ruhmann_gabriela_1999_schluesselkompetenz_schreiben.pdf

Fürstenau, S. (2011). Mehrsprachigkeit als Voraussetzung und Ziel schulischer Bildung. In S. Fürstenau & M. Gomolla (Hrsg.), *Migration und schulischer Wandel: Mehrsprachigkeit* (1. Aufl., S. 25–50). VS Verlag für Sozialwissenschaften / Springer Fachmedien Wiesbaden GmbH Wiesbaden.

Gallmann, P. (1996). Interpunktion (Syngrapheme). In H. Günther & O. Ludwig (Hrsg.), *Handbücher zur Sprach- und Kommunikationswissenschaft. Schrift und Schriftlichkeit: Ein interdisziplinäres Handbuch internationaler Forschung* (S. 1456–1467). de Gruyter. http://www.personal.uni-jena.de/~x1gape/Pub/Syngraph_HSK_1996.pdf

Gogolin, I. (2004). Lebensweltliche Mehrsprachigkeit. In K.-R. Bausch, F. G. Königs & H.-J. Krumm (Hrsg.), *Giessener Beiträge zur Fremdsprachendidaktik. Mehrsprachigkeit im Fokus: Arbeitspapiere der 24. Frühjahrskonferenz zur Erforschung des Fremdsprachenunterrichts* (S. 55–61). Narr.

Gogolin, I. & Duarte, J. (2016). Bildungssprache. In J. Kilian, B. Brouër & D. Lüttenberg (Hrsg.), *Handbücher Sprachwissen: Band 21. Handbuch Sprache in der Bildung* (S. 478–499). de Gruyter.

Gogolin, I. & Lange, I. (2011). Bildungssprache und Durchgängige Sprachbildung. In S. Fürstenau & M. Gomolla (Hrsg.), *Migration und schulischer Wandel: Mehrsprachigkeit* (1. Aufl., S. 107–127). VS Verlag für Sozialwissenschaften / Springer Fachmedien Wiesbaden GmbH Wiesbaden.

Gogolin, I. & Neumann, U. (2009). *Streitfall Zweisprachigkeit: The bilingualism controversy* (1. Aufl.). VS Verlag für Sozialwissenschaften.

Gogolin, I. & Roth, H.-J. (2007). Bilinguale Grundschule: Ein Beitrag zur Förderung der Mehrsprachigkeit. *Mehrsprachigkeit bei Kindern und Erwachsenen. Erwerb, Formen, Förderung*, 31–46.

Grabher, G. (1994). *Lob der Verschwendung: Redundanz in der Regionalentwicklung : ein sozioökonomisches Plädoyer*. Ed. Sigma.

Grimm, H. (Hrsg.). (2000). *Sprachentwicklung* (Bd. 3). Hogrefe.

Grotjahn, R. (1987). How to construct and evaluate a c-test: A discussion of some problems and some statistical analyses. In R. Grotjahn, C. Klein-Braley & D. K. Stevenson (Hrsg.), *Quantitative linguistics: vol. 34. Taking their measure: The validity and validation of language tests* (S. 219–253). Brockmeyer.

Grotjahn, R. (1995). Der C-Test: State of the Art. *Zeitschrift für Fremdsprachenforschung*, *6*(2), 37–60.

Grotjahn, R. (Hrsg.). (2002a). *Fremdsprachen in Lehre und Forschung (FLF): Bd. 32*. *Der C-Test: Theoretische Grundlagen und praktische Anwendungen*. AKS-Verl.

Grotjahn, R. (2002b). Konstruktion und Einsatz von C-Tests: Ein Leitfaden für die Praxis. In R. Grotjahn (Hrsg.), *Fremdsprachen in Lehre und Forschung (FLF): Bd. 32. Der C-Test: Theoretische Grundlagen und praktische Anwendungen* (S. 211–225). AKS-Verl.

Grotjahn, R. (Hrsg.). (2006). *Language testing and evaluation: Bd. 6*. *Der C-Test: Theorie, Empirie, Anwendungen: The C-Test: Theory, empirical research, applications*. P. Lang.

Grotjahn, R. (Hrsg.). (2010a). *Language testing and evaluation: Bd. 18*. *Der C-Test: Beiträge aus der aktuellen Forschung*. Peter Lang.

Grotjahn, R. (2010b). Gesamtdarbietung, Einzeltextdarbietung, Zeitbegrenzung und Zeitdruck: Auswirkungen auf Item- und Testkennwerte und C-Test-Konstrukt. In R. Grotjahn (Hrsg.), *Language testing and evaluation: Bd. 18. Der C-Test: Beiträge aus der aktuellen Forschung* (S. 265–296). Peter Lang.

Grotjahn, R. (Hrsg.). (2014). *Language testing and evaluation: Bd. 34*. *Der C-Test: Aktuelle Tendenzen: The C-test: Current trends*. Peter Lang.

Grotjahn, R. (2019a). C-Tests. In S. Jeuk & J. Settinieri (Hrsg.), *DaZ-Handbücher. Sprachdiagnostik Deutsch als Zweitsprache: Ein Handbuch* (S. 585–610).

Grotjahn, R. (2019b). C-Tests. In S. Jeuk & J. Settinieri (Hrsg.), *DaZ-Handbücher: Bd. 2. Sprachdiagnostik Deutsch als Zweitsprache* (S. 585–610). de Gruyter Mouton.

Grotjahn, R., Klein-Braley, C. & Raatz, U. (2002). C-Tests: an overview. In J. A. Coleman, R. Grotjahn & U. Raatz (Hrsg.), *Fremdsprachen in Lehre und Forschung: Bd. 31. University language testing and the C-test* (S. 93–114). AKS-Verl.

Grotjahn, R., Schlack, T. & Aguado, K. (2010). S-C-Tests: Messung automatisierter sprachlicher Kompetenzen anhand von C-Tests mit mas-siver textspezifischer Zeitlimitierung. In R. Grotjahn (Hrsg.), *Language testing and evaluation: Bd. 18. Der C-Test: Beiträge aus der aktuellen Forschung* (S. 297–319). Peter Lang.

Häcker, H., Leutner, D., Amelang, M. & Pfannenschwarz, C. (1998). *Standards für pädagogisches und psychologisches Testen* (1. Aufl.). *Supplementum der "Diagnostica" und der "Zeitschrift für differentielle und diagnostische Psychologie": Bd. 1*. Hogrefe [u.a.].

Hadjar, A., Lupatsch, J. & Grünewald-Huber, E. (2010). Bildungsverlierer/-innen, Schulentfremdung und Schulerfolg. In G. Quenzel & K. Hurrelmann (Hrsg.), *Bildungsverlierer: Neue Ungleichheiten* (Bd. 19, S. 223–244). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-92576-9_11

Harsch, C. & Harting, J. (2010). Empirische und inhaltliche Analyse lokaler Abhängigkeiten im C-Test. In R. Grotjahn (Hrsg.), *Language testing and evaluation: Bd. 18. Der C-Test: Beiträge aus der aktuellen Forschung* (S. 193–204). Peter Lang.

Harsch, C. & Schröder, K. (2007). Textrekonstruktion: C-Test. In B. Beck & E. Klieme (Hrsg.), *Sprachliche Kompetenzen. Konzepte und Messung: DESI-Studie (Deutsch Englisch Schülerleistung International)* (S. 212–255).

Harsch, C. & Schröder, K. (2014). Der C-Test und sein schulischer Wert: Eine Lanze für mehr C-Tests in der Schule. In R. Grotjahn (Hrsg.), *Language testing and evaluation: Bd. 34. Der C-Test: Aktuelle Tendenzen: The C-test: Current trends* (S. 313–321). Peter Lang.

Hartig, J. & Jude, N. (2004). *Testtheorie und Testkonstruktion.* Universität Frankfurt, Frankfurt am Main. https://user.uni-frankfurt.de/~johartig/TK/pdf/KTT.pdf

Hartig, J., Jude, N. & Wagner, W. (2008). Methodische Grundlagen der Messung und Erklärung sprachlicher Kompetenzen. In DESI-Konsortium (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (S. 34–54). Beltz.

Harting, J. (2008). Kompetenzen als Ergebnisse von Bildungsprozessen. In N. Jude, J. Harting & E. Klieme (Hrsg.), *Kompetenzerfassung in pädagogischen Handlungsfeldern: Theorien, Konzepte und Methoden* (S. 15–25).

Haug, U. & Rammer, G. (1974). *Sprachpsychologie und Theorie der Verständigung* (1. Aufl.). *Sprache und Lernen: Bd. 39*. Pädag. Verl. Schwann.

Hesse, H.-G., Göbel, K. & Hartig, J. (2008). Sprachliche Kompetenzen von mehrsprachigen Jugendlichen und Jugendlichen nicht-deutscher Erstsprache. In DESI-Konsortium (Hrsg.), *Beltz Pädagogik. Unterricht und Kompetenzerwerb in Deutsch und Englisch: Ergebnisse der DESI-Studie* (S. 208–230). Beltz Verlag.

Hinofortis, F. B. (1980). Cloze as an alternative method of ESL placement and proficiency testing. In J. W. Oller JR. & K. Perkins (Hrsg.), *Research in Language Testing* (S. 121–128). Newbury House.

Hornburg, C. & Giering, A. (1996). Konzeptualisierung und Operationalisierung komplexer Konstrukte. Ein Leitfaden für die Marketingforschung. *Marketing ZFP*, *18*(1), 5–24. https://doi.org/10.15358/0344-1369-1996-1-5

Huber. (2009). Warum Forschendes Lernen nötig und möglich ist. In Huber, Hellmer & Schneider (Hrsg.), *Forschendes Lernen im Studium. Aktuelle Konzepte und Erfahrungen.* (S. 9–35). Universitätsverlag Webler. https://www.fh-potsdam.de/fileadmin/user_upload/forschen/material-publikation/Huber_Warum_Forschendes_Lernen_noetig_und_moeglich_ist.pdf

Huber, Hellmer & Schneider (Hrsg.). (2009). *Forschendes Lernen im Studium. Aktuelle Konzepte und Erfahrungen.* Universitätsverlag Webler.

Ingenkamp, K. (1995). *Die Fragwürdigkeit der Zensurengebung.: Text und Untersuchungsberichte*. Beltz.

IQB. *Aufgaben für das Fach Deutsch: Beschreibung der Struktur*. https://www.iqb.hu-berlin.de/abitur/dokumente/deutsch

Jude, N. (2008). *Zur Struktur von Sprachkompetenz* [Dissertation]. Johann Wolfgang Goethe-Universität, Frankfurt am Main. https://d-nb.info/995928428/34

Jude, N., Harting, J., Schipolowski, S., Böhme, K. & Stanat, P. (2013). Definition und Messung von Lesekompetenz: PISA und die Bildungsstandards. *Zeitschrift für Pädagogik*(59), 200–228.

Jude, N. & Klieme, E. (2007). Sprachliche Kompetenz aus Sicht der pädagogisch-psychologischen Diagnostik. In B. Beck & E. Klieme (Hrsg.), *Beltz Pädagogik. Sprachliche Kompetenzen: Konzepte und Messung ; DESI-Studie (Deutsch-Englisch-Schülerleistungen-International)* (S. 9–23). Beltz.

Juska-Bacher, B. & Jakob, S. (2014). Wortschatzumfang und Wortschatzqualität und ihre Bedeutung im fortgesetzen Spracherwerb. *Zeitschrift für angewandte Linguistik*, *61*(1), 49–75. https://www.phbern.ch/fileadmin/user_upload/IFE/Dokumente/Publikationen/Publik

ationen_Datenbank_nicht_loeschen/Britta_Juska-Bacher_nicht_loeschen/Wortschatzumfang_und_Wortschatzqualitaet_und_ihre_Bedeutung_im_fortgesetzten_Spracherwerb.pdf

Kaufmann, N. (2016). Die Vorhersage der Schwierigkeit deutscher C-Test-Texte: Untersuchungen am Beispiel des onDaF. *Zeitschrift für Interkulturellen Fremdsprachenunterricht*, *21*(2), 111–126. https://www.onset.de/fileadmin/Redakteure/PDF/Publikationen/829-1659-1-SM.pdf

Kelava, A. & Moosbrugger, H. (2012). Deskriptivstatistische Evaluation von Items (Itemanalyse) und Testwertverteilung. In H. Moosbrugger & A. Kelava (Hrsg.), *Springer-Lehrbuch. Testtheorie und Fragebogenkonstruktion: Mit 66 Abbildung und 41 Tabellen* (2. Aufl., S. 75–102). Springer.

Kelch, F. (2014). *Wissenschaftliches Arbeiten*. http://blog.zeit.de/schueler/2011/05/03/wissenschaftliches-arbeiten/

Kesselring, A. & Leitner, M. (2007). Soziale Herkunft und Schulerfolg. In I. Erler (Hrsg.), *Buchreihe der ÖH Uni Wien: Bd. 2. Keine Chance für Lisa Simpson? Soziale Ungleichheit im Bildungssystem* (1. Aufl., S. 90–107). Mandelbaum.

Klein-Braley, C. (1981). *Empirical investigation of cloze tests* [Dissertation]. Universität Duisburg-Essen, Duisburg.

Klein-Braley, C. (1985a). A cloze-up on the C-Test: a study in the construct validation of authentic tests. *Language Testing*, *2*(1), 76–104. https://doi.org/10.1177/026553228500200108

Klein-Braley, C. (1985b). C-Tests and Construct Validity. *Fremdsprachen und Hochschule*, *13/14*, 55–56.

Klein-Braley, C. (1994). *Language testing with the C-Test: A linguistic and statistical investigation into the strategies used by C-Te st takers, and the prediction of C-Test diffi- culty* [Habilitationsschrift]. Universität Duisburg-Essen, Duisburg.

Klein-Braley, C. (1997). C-Tests in the context of reduced redundancy testing: an appraisal. *Language Testing*(14), 47–84.

Klein-Braley, C. & Raatz, U. (1984). A survey of research on the C-Test. *Language Testing*, *1*(2), 134–146. https://doi.org/10.1177/026553228400100202

Klein-Braley, C. & Raatz, U. (Hrsg.). (1985). *Fremdsprachen und Hochschulen: Thematischer Teil: C-Tests in der Praxis*. *C-Tests in der Praxis* (13. Aufl.).

Klieme, E. (2004). Was sind Kompetenzen und wie lassen sie sich messen? *Pädagogik*(56.6), 10–13.

Klieme, E. (Hrsg.). (2006). *Zusammenfassung zentraler Ergebnisse der DESI-Studie.* Deutsches Institut für International Pädagogische Forschung. https://www.dipf.de/de/forschung/aktuelle-projekte/pdf/biqua/DESI_Ausgewaehlte_Ergebnisse.pdf

Klieme, E., Artelt, C., Hartig, J., Jude, N., Köller, O., Prenzel, M., Schneider, W. & Stanat, P. (Hrsg.). (2010). *PISA 2009: Bilanz nach einem Jahrzehnt*. Waxmann.

KMK. (2012a). *Bildungsstandards für die fortgeführte Fremdsprache (Englisch/Französisch) für die Allgemeine Hochschulreife* [Beschluss der KMK vom 18.10.2012]. https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2012/2012_10_18-Bildungsstandards-Fortgef-FS-Abi.pdf

KMK. (2012b). *Bildungsstandards im Fach Deutsch für die Allgemeine Hochschulreife* [Beschluss der KMK vom 18.10.2012]. https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2012/2012_10_18-Bildungsstandards-Deutsch-Abi.pdf

KMK. (2015). *Definitionenkatalog zur Schulstatistik 2015 (2).* KMK. https://www.kmk.org/fileadmin/Dateien/pdf/Statistik/Defkat2015_2.pdf

KMK. (2020). *Bildung in Deutschland 2020: Ein indikatorengestützter Bericht mit einer Analyse zu Bildung in einer digitalisierten Welt*. https://www.bildungsbericht.de/static_pdfs/bildungsbericht-2020.pdf

Köberl, J. & Sigott, G. (1994). Adjusting C-Test difficulty in German. In R. Grotjahn (Hrsg.), *Der C-Test: Theoretische Grundlagen und praktische Anwendungen. Bd. 2* (S. 179–192). Brockmeyer.

Kokkota, V. (1988). Letter-deletion procedure: a flexible way of reducing text redundancy. *Language Testing*, *5*(1), 115–119. https://doi.org/10.1177/026553228800500108

Koller, I., Alexandrowicz, R. & Hatzinger, R. (2012). *Das Rasch Modell in der Praxis* (1. Aufl.). *Utb-studi-e-book: Bd. 3786*.

Krampen, D. (2018). *Klassische Testtheorie.* Dorsch - Lexikon der Psychologie. https://portal.hogrefe.com/dorsch/de/startseite/stichwort-detailseite/desktop/1/keyword/klassische-testtheorie-1/

Kubinger, K. D. (2009). *Psychologische Diagnostik: Theorie und Praxis psychologischen Diagnostizierens* (2., überarb. und erw. Aufl.). Hogrefe.

Kubinger, K. D. (2018). *Item-Response-Theorie (IRT).* Dorsch - Lexikon der Psychologie. https://m.portal.hogrefe.com/dorsch/item-response-theorie-irt/

Lenhard, W. & Lenhard, A. (2014-2017). *Berechnung des Lesbarkeitsindex LIX nach Björnson*. http://www.psychometrica.de/lix.html. Bibergau: Psychometrica. DOI: 10.13140/RG.2.1.1512.3447

Lenhard, W. & Lenhard, A. (2014). *Testing the Significance of Correlations.* Psychometrica.

Lenhard, W. & Lenhard, A. (2016). *Berechnung von Effektstärken.* Psychometrica. https://www.psychometrica.de/effektstaerke.html

Lienert, G. A. & Raatz, U. (1998). *Testaufbau und Testanalyse* (6. Aufl.). *Grundlagen Psychologie*. Beltz Psychologie VerlagsUnion.

Linnemann, M. & Wilbert, J. (2014). Do C-tests measure language comprehension of learning disabled students. In R. Grotjahn (Hrsg.), *Language testing and evaluation: Bd. 34. Der C-Test: Aktuelle Tendenzen: The C-test: Current trends* (S. 223–238). Peter Lang.

Mandl, H., Friedrich, H. F. & Hron, A. (1986). Psychologie des Wissenserwerbs. In B. Weidenmann, A. Krapp, M. Hofer, G. L. Huber & H. Mandl (Hrsg.), *Pädagogische Psychologie* (S. 143–218). Urban & Schwarzenberg.

Martin, G., Bremerich-Vos, A. & Scholten-Akoun, D. (2016). Zur Kodierung der studientischen Texte. In A. Bremerich-Vos & D. Scholten-Akoun (Hrsg.), *Schriftsprachliche Kompetenzen von Lehramtsstudierenden in der Studieneingangsphase: Eine empirische Untersuchung* (1. Aufl., S. 123–138). Schneider Hohengehren.

Mashkovskaya, A. (2014). *Der C-Test als Lesetest bei Muttersprachlern* [Dissertation]. Universität Duisburg-Essen, Essen. https://duepublico.uni-duisburg-essen.de/servlets/DerivateServlet/Derivate-34531/Dissertation_2014_01_06.pdf

Mashkovskaya, A. & Baur, R. S. (2016a). C-Tests für erwachsene Erst- und Zweitsprachsprecherinnen.: Screening der (schrift-)sprachlichen Kompetenzen von Lehramtsstudierenden. In A. Bremerich-Vos & D. Scholten-Akoun (Hrsg.), *Schriftsprachliche Kompetenzen von Lehramtsstudierenden in der Studieneingangsphase: Eine empirische Untersuchung* (1. Aufl., S. 191–205). Schneider Hohengehren.

Mashkovskaya, A. & Baur, R. S. (2016b). Ergebnisse und Diskussion der C-Tests. In A. Bremerich-Vos & D. Scholten-Akoun (Hrsg.), *Schriftsprachliche Kompetenzen von*

Lehramtsstudierenden in der Studieneingangsphase: Eine empirische Untersuchung (1. Aufl., S. 213–232). Schneider Hohengehren.

Mathiebe, M. (2018). *Wortschatz und Schreibkompetenz: Bildungssprachliche Mittel in Schülertexten der Sekundarstufe I*. Waxmann Verlag GmbH.

Max-Planck-Institut für Bildungsforschung (Hrsg.). (2002). *PISA 2000: Die Studie im Überblick: Grundlagen, Methoden und Ergebnisse*. https://www.mpib-berlin.mpg.de/Pisa/PISA_im_Ueberblick.pdf

McKamey, T. (2006). Getting closure on the cloze: A validation study of the 'rational deletion' method. *Second Language Studies*, *24*(2), 114–164.

Meibauer, J., Demske, U., Geilfuß-Wolfgang, J., Pafel, J., Ramers, K. H., Rothweiler, M. & Steinbach, M. (2002). *Einführung in die germanistische Linguistik*. Metzler.

Mittenecker, E. (2017). Redundanz. In M. A. Wirtz (Hrsg.), *Dorsch - Lexikon der Psychologie* (18. Aufl.). Hogrefe.

Molcho, S. (1999). Körpersprache. In H. Biermann, B. Schurf, G. Brenner, D. Erlach, K. Fingerhut, M. Fingerhut & C. Grunow (Hrsg.), *Texte, Themen und Strukturen: Deutschbuch für die Oberstufe.* Cornelsen.

Möller, G. & Bellenberg, G. (2017). *Ungleiches ungleich behandeln*. https://www.gew-nrw.de/fileadmin/user_upload/Kampagne_Bildung-weiter-denken/GEW-NRW-Moeller-Bellenberg-Studie-Sozialindex-gesamt-Ungleiches-ungleich-behandeln.pdf

Moosbrugger, H. (2008). Item-Response-Theorie (IRT). In H. Moosbrugger & A. Kevala (Hrsg.), *Springer-Lehrbuch. Testtheorie und Fragebogenkonstruktion: Mit 43 Tabellen* (S. 215–259). Springer. https://doi.org/10.1007/978-3-540-71635-8_10

Moosbrugger, H. (2012a). Item-Response-Theorie (IRT). In H. Moosbrugger & A. Kelava (Hrsg.), *Springer-Lehrbuch. Testtheorie und Fragebogenkonstruktion: Mit 66 Abbildung und 41 Tabellen* (2. Aufl., S. 227–274). Springer.

Moosbrugger, H. (2012b). Klassische Testtheorie (KKT). In H. Moosbrugger & A. Kelava (Hrsg.), *Springer-Lehrbuch. Testtheorie und Fragebogenkonstruktion: Mit 66 Abbildung und 41 Tabellen* (2. Aufl., S. 103–117). Springer. https://link.springer.com/content/pdf/10.1007%2F978-3-642-20072-4_5.pdf

Moosbrugger, H. (2012c). *Testtheorie und Fragebogenkonstruktion* (2. Aufl.). *Springer-Lehrbuch*. Springer.

Moosbrugger, H. & Kelava, A. (Hrsg.). (2012). *Springer-Lehrbuch*. *Testtheorie und Fragebogenkonstruktion: Mit 66 Abbildung und 41 Tabellen* (2. Aufl.). Springer.

MSB NRW. (2004). *Kernlehrplan für die Gesamtschule - Sekundarstufe I in Nordrhein-Westfalen - Deutsch* (1. Aufl.). *Schriftenreihe "Schule in NRW": Bd. 3107*. Ritterbach. http://curricula-depot.gei.de/bitstream/handle/11163/15/720294398_2004.pdf?sequence=1

MSB NRW. (2007). *Kernlehrplan für den verkürzten Bildungsgang des Gymnasiums - Sekundarstufe I (G8) in Nordrhein-Westfalen. Deutsch*.

MSB NRW (2011). Neues Standorttypenkonzept: Faire Vergleiche bei Lernstandserhebungen. *Schule NRW*, *11*(6). https://www.schulentwicklung.nrw.de/e/upload/lernstand8/download/mat_2011/Amtsblatt_SchuleNRW_06_11_Isaac-Standorttypenkonzept.pdf

MSB NRW. (2014). *Kernlehrplan für die Sekundarstufe II Gymnasium/Gesamtschule in Nordrhein-Westfalen: Deutsch*. https://www.schulentwicklung.nrw.de/lehrplaene/lehrplan/11/KLP_GOSt_Deutsch.pdf

MSB NRW. (2017). *Abiturprüfung auf Basis der Kernlehrpläne - Beispielaufgabe: Deutsch, Grundkurs.* Qualitäts- und UnterstützungsAgenur - Landesinstitut für Schule. https://www.standardsicherung.schulministerium.nrw.de/cms/zentralabitur-gost/faecher/getfile.php?file=4613

MSB NRW. (2019a). *Quantia Schuljahr 2018/19: Das Schulwesen in Nordrhein-westfalen aus quantitativer Sicht* [Statistische Übersicht]. https://www.schulministerium.nrw.de/docs/bp/Ministerium/Service/Schulstatistik/Amtliche-Schuldaten/Quantita_2018.pdf

MSB NRW. (2019b). *Statistik-Telegramm 2018/19: Schuleckdaten 2018/19*. Statistische Übersicht. https://www.schulministerium.nrw.de/docs/bp/Ministerium/Service/Schulstatistik/Amtliche-Schuldaten/StatTelegramm2018.pdf

Müller, K. & Ehmke, T. (2016). Soziale Herkunft und Kompetenzerwerb. In K. Reiss, C. Sälzer, A. Schiepe-Tiska, E. Klieme & O. Köller (Hrsg.), *PISA 2015: Eine Studie zwischen Kontinuität und Innovation* (S. 286–316). Waxmann.

Naumann, J., Artelt, C., Schneider, W. & Stanat, P. (2010). Lesekompetenz von PISA 2000 bis PISA 2009. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009: Bilanz nach einem Jahrzehnt* (S. 23–72). Waxmann.

Neisser, U. (1974). *Kognitive Psychologie*. *Konzepte der Humanwissenschaften*. Klett.

Neumann, A. (2014). Jugendliche DaZ-Lerner schreiben schulische Textformen - Reanalysen der Leistungsdaten und Schülerbefragungen aus DESI und IMOSS. In B. Ahrenholz (Hrsg.), *DaZ-Forschung: v.4. Zweitspracherwerb im Jugendalter* (S. 171–192). de Gruyter Mouton.

Neumann, A. & Domenech, M. (2010). Paradoxien des Schreibens in der Bildungssprache Deutsch. *Blickpunkt Deutsch als Zweitsprache*, *2*, 11–37.

Neumann, A. & Lehmann, Rainer, H. (2008). Schreiben Deutsch. In DESI-Konsortium (Hrsg.), *Beltz Pädagogik. Unterricht und Kompetenzerwerb in Deutsch und Englisch: Ergebnisse der DESI-Studie* (S. 89–103). Beltz Verlag.

Nyquist, H. (2002). Certain Topics in Telegraph Transmission Theory. *Proceedings of the IEEE*, *90*, 617–644.

OECD. (2004). *Lernen für die Welt von morgen: Erste Ergebnisse von PISA 2003*. http://www.oecd.org/education/school/programmeforinternationalstudentassessmentpisa/34474315.pdf

OECD. (2016). *PISA 2015 Ergebnisse (Band I)*. OECD. https://doi.org/10.1787/9789264267879-de

OECD. (2019a). *Deutschland - Ländernotiz: PISA 2018 Ergebnisse.* OECD. http://www.oecd.org/berlin/themen/pisa-studie/PISA2018_CN_DEU_German.pdf

OECD. (2019b). *PISA 2018 Ergebnisse (Band I)*. OECD. https://doi.org/10.1787/1da50379-de

OECD. (2019c). *PISA 2018 Results (Volume II)*. OECD. https://www.oecd-ilibrary.org/docserver/b5fd1b8f-en.pdf?expires=1576328526&id=id&accname=guest&checksum=8EB2BD975F79CFD93FFC77528E413104 https://doi.org/10.1787/b5fd1b8f-en

OECD. (2019d, 3. Dezember). *PISA-Studie 2018* [Press release]. Paris, Berlin. http://www.oecd.org/berlin/presse/pisa-studie-2018-leistungen-in-deutschland-insgesamt-ueberdurchschnittlich-aber-leicht-ruecklaeufig-und-mit-grossem-abstand-zu-den-spitzenreitern-03122019.htm

Olfert, H. (2019). *Spracherhalt und Sprachverlust bei Jugendlichen: Eine Analyse begünstigender und hemmender Faktoren für Spracherhalt im Kontext von Migration*. *Language development: Band 40*. Narr Francke Attempto.

Oller, J. W., JR. (1972). Scoring Methods and Difficulty Levels for Cloze Tests of Proficiency in English as a Second Language. *The Modern Language Journal*, *56*(3), 151–158. https://doi.org/10.1111/j.1540-4781.1972.tb05035.x

Oller, J. W., JR. (1973). Cloze Tests of Second Language Proficiency and What They Measure. *Language Learning*, *23*(1), 105–118. https://doi.org/10.1111/j.1467-1770.1973.tb00100.x

Oller, J. W., JR. (1974). Expectancy for Successive Elements: Key Ingredient to Language Use. *Foreign Language Annals*, *7*(4), 443–452. https://doi.org/10.1111/j.1944-9720.1974.tb01503.x

Oller, J. W., JR. (1976). Evidence for General Language Proficiency Factor: An Expectancy Grammar. *Die neueren Sprachen*, *75*(2), 165–174. https://doi.org/10.1111/j.1540-4781.1976.tb06314.x

Oller, J. W., JR. (1979). *Language tests at school*. Longman.

Oller, J. W., JR. (1983). Evidence for a general language proficiency factor: An expectancy grammar. In J. W. Oller JR. (Hrsg.), *Issues in LAnguage Testing Research* (S. 3–10). Newbury House.

Oller, J. W., JR. & Conrad, C. A. (1971). The Cloze Technique and ESL Proficiency. *Language Learning* (21), 183–195.

Oller, J. W., JR. & Inal, N. (1971). A Cloze Test of English Prepositions. *TESOL Quarterly* (5), 315–326.

Organisation for Economic Co-operation and Development (Hrsg.). (2000). *Literacy in the Information Age: Final Report on the International Adult Literacy Survey*. http://www.oecd.org/education/skills-beyond-school/41529765.pdf

Ossner, J. (2006). Kompetenzen und Kompetenzmodelle im Deutschunterricht. *Didaktik Deutsch*, *12*(21), 5–19.

Petersen, I. (2014a). *Schreibfähigkeit und Mehrsprachigkeit* (1. Aufl.). *DaZ-Forschung: Bd. 5*. de Gruyter. http://www.degruyter.com/books/9783110318333

Petersen, I. (2014b). *Sprachbildung und Schreibförderung in der Sekundarstufe II.* proDaZ. https://www.uni-due.de/imperia/md/content/prodaz/petersen_sprachbildung.pdf

QUA-LIS NRW. (2017). *Deskriptive Beschreibung der Standorttypen für die weiterführenden Schulen.* Qualitäts- und UnterstützungsAgenur - Landesinstitut für Schule. https://www.schulentwicklung.nrw.de/e/upload/lernstand8/download/mat_2017/2017-02-08_Beschreibung_Standorttypen__weiterfhrende_Schulen_NEU_RUB_ang.pdf

Raatz, U. (1985). Tests of reduced redundancy: the C-test, a practical example. *Fremdsprachen und Hochschule*, 14–19.

Raatz, U. & Klein-Braley, C. (1981). The C-Test: A Modification of the Cloze Procedure. In T. Culhane, C. Klein-Braley & D. K. Stevenson (Hrsg.), *Practice and problems in the language testing* (26. Aufl.). University of Essex.

Raatz, U. & Klein-Braley, C. (1983). Ein neuer Ansatz zur Messung der Sprachlesitung. Der C-Test: Theorie und Praxis. *Tests und Trends. Jahrbuch der pädagogischen Diagnostik.*(3), 107–138.

Raatz, U. & Klein-Braley, C. (1985). How to develop a C-Test. *Fremdsprachen und Hochschule*, *13/14*, 20–22.

Raatz, U. & Klein-Braley, C. (2002). Introduction to language testing and to C-Test. In J. A. Coleman, R. Grotjahn & U. Raatz (Hrsg.), *Fremdsprachen in Lehre und Forschung: Bd. 31. University language testing and the C-test* (S. 75–91). AKS-Verl.

Rankin, E. F. & Culhane, J. W. (1969). Comparable Cloze and Multiple-choice Comprehension Test Scores. *Journal of Reading*(13), 193–198.

Redder, A., Schwippert, K., Hasselhorn, M., Forschner, S. & Fickermann, Detelv & Ehlich, Konrad. (2011). *Bilanz und Konzeptualisierung von strukturierter Forschung zu Sprachdiagnostik und Sprachförderung.* ZUSE.

Reiss, K., Sälzer, C., Schiepe-Tiska, A., Klieme, E. & Köller, O. (Hrsg.). (2016). *PISA 2015: Eine Studie zwischen Kontinuität und Innovation*. Waxmann. https://www.pisa.tum.de/fileadmin/w00bgi/www/Berichtsbaende_und_Zusammenfassungungen/Bericht_2015-openaccess.pdf https://doi.org/10.31244/9783830991007

Riebling, L. (2013). *Sprachbildung im naturwissenschaftlichen Unterricht: Eine Studie im Kontext migrationsbedingter sprachlicher Heterogenität* (1. Aufl.). Waxmann Verlag GmbH.

Rohrhirsch, F. (2008). Medien - Ethik - Wirklichkeit. In K. Dambach & A. Ramin (Hrsg.), *KombiKompakt: Deutsch in der Oberstufe* (S. 144). C.C. Buchner.

Rost, J. (2004). *Lehrbuch Testtheorie, Testkonstruktion* (2. Aufl.). *Aus dem Programm Huber, Psychologie Lehrbuch*. H. Huber.

Rothweiler, M. & Ruberg, T. (2011, 2011). *Der Erwerb des Deutschen bei Kindern mit nichtdeutscher Erstsprache: Sprachliche und außersprachliche Einflussfaktoren ; eine Expertise der Weiterbildungsinitiative Frühpädagogische Fachkräfte (WiFF) ; [Sprache]*. *WiFF-Expertisen: Bd. 12*. DJI.

Schäfer, T. (2010). *Statistik I: Deskriptive und Explorative Datenanalyse*. VS Verlag für Sozialwissenschaften / GWV Fachverlage GmbH Wiesbaden. https://doi.org/10.1007/978-3-531-92446-5

Schneider, W., Schlagmüller, M. & Ennemoser, M. (2007). *LGVT 6-12: Lesegeschwindigkeits und –verständnistest für die Klassen 6-12: Manual und Testheft*. Deutsche Schultests. Hogrefe.

Schneider, W., Schlagmüller, M. & Ennemoser, M. (2017). *LGVT 5-12+: Lesegeschwindigkeits- und Verständnistest für die Klassen 5–12+.*

Schnieders, G. & Komor, A. (2005). Eine Synopse aktueller Verfahren der Sprachstandsfeststellung. In Bundesministerium für Bildung und Forschung (Hrsg.), *Bildungsreform Band: Bd. 11. Anforderungen an Verfahren der regelmäßigen Sprachstandsfeststellung als Grundlage für die frühe und individuelle Förderung von Kindern mit und ohne Migrationshintergrund* (S. 261–342).

Scholten, D. (2000). *Sprachverbreitungspolitik des nationalsozialistischen Deutschlands*. Zugl.: Duisburg, Univ., Diss., 1999. *Duisburger Arbeiten zur Sprach- und Kulturwissenschaft: Bd. 42*. Lang.

Scholten-Akoun, D. (2016a). Eine Fragebogenerhebnung zu den sprachlichen Eingangsvoraussetzungen von Studierenden. In A. Bremerich-Vos & D. Scholten-Akoun (Hrsg.), *Schriftsprachliche Kompetenzen von Lehramtsstudierenden in der Studieneingangsphase: Eine empirische Untersuchung* (1. Aufl., S. 15–26). Schneider Hohengehren.

Scholten-Akoun, D. (2016b). Zur Genese des Projekts. In A. Bremerich-Vos & D. Scholten-Akoun (Hrsg.), *Schriftsprachliche Kompetenzen von Lehramtsstudierenden in der Studieneingangsphase: Eine empirische Untersuchung* (1. Aufl., S. 27–33). Schneider Hohengehren.

Scholten-Akoun, D. & Baur, R. S. (2012). Der C-Test als ein Instrument zur Messung der Schriftsprachkompetenzen von Lehramtsstudierenden (auch) mit Migrationshintergrund: Eine Studie. In B. Ahrenholz (Hrsg.), *Beiträge aus dem Workshop "Kinder und Jugendliche mit Migrationshintergrund". Sprachstand erheben - Spracherwerb erforschen: Beiträge aus dem 6. Workshop "Kinder und Jugendliche mit Migrationshintergrund", 2010* (S. 307–330). Ernst Klett.

Scholten-Akoun, D., Kuhnen, A. & Mashkovskaya, A. (2013). Sprachkompetenz Studierender: Design und erste Ergebnisse einer empirischen Studie. In H. Feilke, J. Köster & M. Steinmetz (Hrsg.), *Textkompetenzen in der Sekundarstufe II* (1. Aufl., S. 179–200). Fillibach bei Klett.

Scholten-Akoun, D., Mashkovskaya, A. & Tischmeyer, D. (2014). Language competencies of future teachers: Design and results of an empirical study. *Applied Linguistics Review*, *5*(2), 401–423.

Scholten-Akoun, D., Tischmeyer, D. & Martin, G. (2016). Das Kodiermanual. In A. Bremerich-Vos & D. Scholten-Akoun (Hrsg.), *Schriftsprachliche Kompetenzen von Lehramtsstudierenden in der Studieneingangsphase: Eine empirische Untersuchung* (1. Aufl., S. 65–122). Schneider Hohengehren.

Schräpler, J.-P. (2009). Verwendung von SGB II-Dichten als raumindikator für die Sozialberichterstattung am Beispiel der "soziallen Belastung" von Schulstandorten in NRW. *Statistische Analysen und Studien NRW*(57), 3–32.

Schröder, K. & Harsch, C. (2003). *Rahmenkonzeption zur Erfassung Sprachlicher Kompetenzen.* Universität Augsburg, Lehrstuhl für Didaktik des Englischen. https://www.philhist.uni-augsburg.de/lehrstuehle/anglistik/didaktik/forschung/desi/desis/DESI_C-Test_Konzeption_10_2002.html

Schymura, I. (2013). *Zweisprachigkeit, Schulerfolg und Integration von Migrantenkindern mit polnischsprachigem Hintergrund* [Dissertation]. Universität Duisburg-Essen, Essen.

Shannon, C. E. & Weaver, W. (1949). *The Mathematical Theory of Communication*. University of Illinois Press.

Sigott, G. (1993). *Zur Lernbarkeit von Englisch und Französisch für deutsche Muttersprachler: Eine exploratorische Pilotstudie*. Zugl.: Klagenfurt, Univ., Diss., 1992. */Arbeiten aus Anglistik und Amerikanistik / Buchreihe] Buchreihe zu den Arbeiten aus Anglistik und Amerikanistik: Bd. 6*. Narr.

Sigott, G. (2004). *Towards Identifying the C-Test Construct*. *Language testing and evaluation: vol. 1*. P. Lang.

Sigott, G. & Köberl, J. (1996). Deletion patterns and C-Test difficulty across languages. In R. Grotjahn (Hrsg.), *Manuskripte zur Sprachlehrforschung: 39/3. Der C-Test: Theoretische Grundlagen und praktische Anwendungen. Bd. 3* (S. 159–172). Brockmeyer.

Singleton, D. (1999). *Exploring the second language mental lexicon*. Cambridge University Press.

Sitta, H. (2005). Sachtexte im Deutschunterricht. Wozu? Welche? Ein Plädoyer für stärkere Berücksichtigung argumentativer Texte im gymnasialen Deutschunterricht. In M. Fix, R. Jost & K.-H. Fingerhut (Hrsg.), *Sachtexte im Deutschunterricht: Für Karlheiz Fingerhut zum 65. Geburtstag* (S. 150–159). Schneider Hohengehren.

Spolsky, B. (1973). What Does it Mean to Know a Language, Or How Do You Get Someone to Perform His Competence? In J. W. Oller JR. & J. C. Richards (Hrsg.), *Focus on the Learner* (S. 164–176). Newbury House.

Spolsky, B., Sigurd, B., Masahito, S., Walker, E. & Arterburn, C. (1968). Preliminary Studies in the Development of Techniques for Testing Overall Second Language Proficiency. In J.

A. Upshur & J. Fata (Hrsg.), *Language Learning Special Issue. Problems in Foreign Language Testing* (3. Aufl., S. 79–103).

Stanat, P., Rauch, D. & Segeritz, M. (2010). Schülerinnen und Schüler mit Migrationshintergrund. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009: Bilanz nach einem Jahrzehnt* (S. 200–228). Waxmann.

Stangl, W. (2019). *Online Lexikon für Psychologie und Pädagogik: Trennschärfe*. https://lexikon.stangl.eu/11255/trennschaerfe/

Statistisches Bundesamt. (2018). *Datenreport 2018: Ein Sozialbericht für die Bundesrepublik Deutschland*. bpb Bundeszentrale für politische Bildung. https://www.destatis.de/DE/Service/Statistik-Campus/Datenreport/Downloads/datenreport-2018.pdf?__blob=publicationFile

Statistisches Bundesamt. (2019). *Bevölkerung mit Migrationshintergrund: Ergebnisse des Mikrozensus 2018.* Statistisches Bundesamt. Fachserie. https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Migration-Integration/Publikationen/_publikationen-innen-migrationshintergrund.html;jsessionid=2F23D8ACBA16A79E653034FAD0B8FCF4.internet721?nn=206104

Steinhoff, T. (2009). Wortschatz - eine Schaltstelle für den schulischen Spracherwerb? *Siegener Papiere zur Aneignung sprachlicher Strukturformen (SPASS)*, *17*, 3–63.

Steyer, R. & Eid, M. (2001). *Messen und Testen: Mit Übungen und Lösungen*. *Springer-Lehrbuch*. Springer Berlin Heidelberg; Imprint; Springer.

Stöber, R. (2011). Ohne Redundanz keine Anschlusskommunikation. Zum Verhältnis von Information und Kommunikation. *M&K Medien & Kommunikationswissenschaft*, *59*(3), 307–323. https://doi.org/10.5771/1615-634x-2011-3-307

Süßmilch, E. (1985). C-Tests für ausländische Schüler: Sprachdiagnose im Unterricht Deutsch als Zweitsprache. In C. Klein-Braley & U. Raatz (Hrsg.), *Fremdsprachen und Hochschulen: Thematischer Teil: C-Tests in der Praxis. C-Tests in der Praxis* (13. Aufl.).

Szagun, G. (2000). *Sprachentwicklung beim Kind* (6. Aufl.). *Beltz-Taschenbuch: 62 : Psychologie*. Beltz.

Szagun, G. (2013). *Sprachentwicklung beim Kind: Ein Lehrbuch* (5., aktualisierte Aufl.). Beltz.

Tausendpfund, M. (Hrsg.). (2019). *Studientexte zur Soziologie*. *Quantitative Datenanalyse: Eine Einführung mit SPSS* (1. Aufl.). Springer Fachmedien Wiesbaden.

Taylor, W. (1953). Cloze prozedure: A new Tool for Measuring readibility. *Journalism Quarterly*, *30*, 415–433.

Taylor, W. (1956). Recent Developments in the Use of Cloze Procedure. *Journalism Quarterly*(33), 42–48.

Thiede, E. (2014). *Die Vertrautheit mit dem Testformat als Einflussfaktor bei der Lösung von C-Tests im schulischen Kontext* [unveröffentlichte Staatsarbeit]. Universität Duisburg-Essen, Essen.

Traut-Mattausch, E. & Dieter Frey. (2006). Kommunikationsmodelle. In H.-W. Bierhoff, D. Frey & J. Bengel (Hrsg.), *Handbuch der Psychologie / hrsg. von J. Bengel.: Bd. 3. Handbuch der Sozialpsychologie und Kommunikationspsychologie* (Bd. 3, S. 536–544). Hogrefe.

Trautwein, U., Lüdtke, O. & Husemann, N. (2006). Die Qualität der Studienvorbereitung in der gymnasialen Oberstufe: Eine Längsschnittstudie mit Absolventen von allgemeinbildenden und beruflichen Gymnasien. In A. Ittel, L. Stecher, H. Merkens & J. Zinnecker (Hrsg.), *Jahrbuch Jugendforschung: 6. Ausgabe 2006* (2006. Aufl., Bd. 31,

S. 47–67). VS Verlag fur Sozialwissenschaften. https://doi.org/10.1007/978-3-531-90537-2_3

Vockrodt-Scholz, V. & Zydatiß, W. (2010). Sprachproduktive Faktoren und die Konstruktvalidität von C-Tests: Kompetenzniveaus und Fehlerquotient in textsortengebundenen Schreibaufgaben. In R. Grotjahn (Hrsg.), *Language testing and evaluation: Bd. 18. Der C-Test: Beiträge aus der aktuellen Forschung* (S. 1–40). Peter Lang.

Völkl, K. & Korb, C. (2018). Variablen und Skalenniveaus. In K. Völkl & C. Korb (Hrsg.), *Deskriptive Statistik* (S. 7–28). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-10675-1_2

Vollmer, H. J. (1982). *Spracherwerb und Sprachbeherrschung: Untersuchungen zur Struktur von Fremdsprachenfähigkeit : ein empirischer Beitrag zu einer psycholinguistisch orientierten Sprachlehr-/-lernforschung*. *Tübinger Beiträge zur Linguistik: Bd. 187*. G. Narr.

Vollmer, H. J. & Sang, F. (1978). *Allgemeine Sprachfähigkeit und Fremdsprachenerwerb : zur Struktur von Leistungsdimensionen und linguistischer Kompetenz des Fremdsprachenlernens*. *Diskussionsbeiträge aus dem Institut für Bildungsforschung: Bd. 1*. Max-Planck-Institut für Bildungsforschung.

Vollmer, H. J. & Sang, F. (1980a). Modelle linguistischer Kompetenz und ihre empirische Fundierung. In R. Grotjahn & E. Hopkins (Hrsg.), *Empirical Research on Language Teaching and Language Acquisition* (S. 11–84). Brockmeyer.

Vollmer, H. J. & Sang, F. (1980b). Zum psycholinguistischen Konstrukt einer internalisierten Erwartungsgrammatik. *Linguistik und Didaktik* (42), 122–148.

Weaver, W. W. & Kingston, A. J. (1963). A factor analysis of the cloze procedure and other measures of reading and language ability. *Journal of Communication* (13), 252–261.

Weber, H. (2005). *Handbuch der Persönlichkeitspsychologie und differentiellen Psychologie*. *Handbuch der Psychologie: Bd. 2*. Hogrefe.

Weber, T. (2014). Der Reiz von Fernsehserien. In H. Biermann, B. Schurf, G. Brenner, D. Erlach, K. Fingerhut, M. Fingerhut & C. Grunow (Hrsg.), *Texte, Themen und Strukturen: Deutschbuch für die Oberstufe* (S. 119). Cornelsen.

Weinert, F. E. (2001). *Leistungsmessungen in Schulen* (2. Aufl.). *Beltz Pädagogik*. Beltz-Verl.

Weis, M., Doroganova, A., Hahnel, C., Becker-Mrotzek, M., Lindauer, T., Artelt, C. & Reiss, K. (2019). Lesekompetenz in PISA 2018 - Ergebnisse in einer digitalen Welt. In K. Reiss, M. Weis, E. Klieme & O. Köller (Hrsg.), *PISA 2018: Grundbildung im internationalen Vergleich* (S. 47–80). Waxmann Verlag GmbH. https://www.pisa.tum.de/fileadmin/w00bgi/www/Berichtsbaende_und_Zusammenfassungungen/PISA_2018_Berichtsband_online_29.11.pdf

Weis, M., Müller, K., Mang, J., Heine, J.-H., Mahler, N. & Reiss, K. (2019). Soziale Herkunft, Zuwanderungshintergrund und Lesekompetenz. In K. Reiss, M. Weis, E. Klieme & O. Köller (Hrsg.), *PISA 2018: Grundbildung im internationalen Vergleich* (S. 129–162). Waxmann Verlag GmbH.

Werner, R. (11. April 2019). Es ist Zeit für eine Schreiboffensive: Zu viel Mündliches an Schulen. *F.A.Z*. https://www.faz.net/aktuell/feuilleton/hoch-schule/wegen-muendlicher-leistung-schueler-koennen-nicht-richtig-schreiben-16122857.html?printPagedArticle=true#pageIndex_0

Willenberg, H. (2001). Versuch, einen Einblick in den Deutschunterricht der Sekundarstufe II zu gewinnen. In H.-E. Tenroth (Hrsg.), *Kerncurriculum Oberstufe. Mathe - Deutsch - Englisch.: Expertisen im Auftrag der KMK* (S. 82–102). Beltz.

Windzio, M. (2013). Das generalisierte lineare Regressionsmodell. In M. Windzio (Hrsg.), *Regressionsmodelle für Zustände und Ereignisse* (S. 17–37). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-531-18852-2_2

Wirtz, M. A. (Hrsg.). (2014). *Dorsch - Lexikon der Psychologie* (17., überarb. Aufl., neue Ausg). Verlag Hans Huber.

Wirtz, M. A. (Hrsg.). (2017). *Dorsch - Lexikon der Psychologie* (18., überarbeitete Auflage). Hogrefe.

Wockenfuß, V. (2009). *Diagnostik von Sprache und Intelligenz bei Jugendlichen und jungen Erwachsenen*. *Berichte aus der Psychologie*. Shaker.

Wockenfuß, V. & Raatz, U. (2006). Über den Zusammenhang zwischen Testleistung und Klassenstufe bei muttersprachlichen C-Tests. In R. Grotjahn (Hrsg.), *Language testing and evaluation: Bd. 6. Der C-Test: Theorie, Empirie, Anwendungen: The C-Test: Theory, empirical research, applications* (S. 211–242). P. Lang.

Wockenfuß, V. & Raatz, U. (2014). Zur Validität von muttersprachlichen C Tests: Bedeutung von verbaler Intelligenz und Informationsverarbeitungsgeschwindigkeit unter Berücksichtigung des Lebensalters. In R. Grotjahn (Hrsg.), *Language testing and evaluation: Bd. 34. Der C-Test: Aktuelle Tendenzen: The C-test: Current trends* (S. 189–223). Peter Lang.

Zimmermann, K. (2019). *Keine Zeit für den C-Test? Eine empirische Untersuchung zum Einfluss einer Geschwindigkeitskomponente auf das Konstrukt des C-Tests*. Universitätsverlag der TU Berlin. https://depositonce.tu-berlin.de/bitstream/11303/9202/3/zimmermann_kerstin.pdf

Zinn, B., Latzel, M. & Ariali, S. (2017). Entwicklung und Erprobung eines Instruments zur Erfassung technischen Wissens im Fach Naturwissenschaften und Technik. *Journal of Technical Education (JOTED)*, *5*(1), 76–99. https://www.journal-of-technical-education.de/index.php/joted

Abbildungsverzeichnis

Abbildung 1: Kompetenzen und Anforderungen im C-Test nach Harsch und Schröder 23
Abbildung 2: Das Schreibkompetenzmodel von Becker-Mrotzek und Schindler 29
Abbildung 3: Inhaltsfelder und Kompetenzbereiche zu „Rezeption“ und „Produktion“ für das Fach Deutsch in der Sekundarstufe II 34
Abbildung 4: Beispiele zur Fähigkeit der Gestaltergänzung. 39
Abbildung 5: Sender- Empfänger Modell von Shannon & Weaver 46
Abbildung 6: Beispielsatz zum Redundanzprinzip 48
Abbildung 7: Einfluss der Reduzierung von Redundanz auf die Schwierigkeit von C-Tests nach Sigott & Köberl 61
Abbildung 8: Mittelwerte der unterschiedlichen C-Test-Versionen im Englischen und Deutschen nach Sigott & Köberl 62
Abbildung 9: Beispiel Nominalphrase Deutsch und Englisch 64
Abbildung 10: Beispiel zur Flexion im Deutschen 64
Abbildung 11: Ausschnitt aus dem C-Test *Wissenschaftliches Arbeiten* mit rechtsseitiger Tilgung 70
Abbildung 12: Ausschnitt aus dem C-Test *Wissenschaftliches Arbeiten* mit linksseitiger Tilgung 72
Abbildung 13: Fragebogen zur Erfassung der Hintergrundvariablen 105
Abbildung 14: Zur Definition des Zuwanderungshintergrunds bzw. des Zuwanderungsstatus nach PISA 2018 107
Abbildung 15: Layout des C-Test-Kursraums 122
Abbildung 16: Design und Layout während der Bearbeitung der C-Tests 123
Abbildung 17: Histogramm der C-Test-Rohwerte 151
Abbildung 18: Boxplot C-Test und Deutschnote 160
Abbildung 19: Boxplot C-Test-Ergebnis und Phase in der Oberstufe 166
Abbildung 20: Boxplot C-Test und Geschlecht 170
Abbildung 21: Boxplot C-Test und Sprache 175
Abbildung 22: Boxplot C-Test und Erstsprache/Sprachverwendung 178
Abbildung 23: Boxplot C-Test und Sprachgruppen 181
Abbildung 24: Boxplot C-Test und höchster Schulabschluss der Eltern 187
Abbildung 25: Boxplot C-Test und Schulstandorttyp 193
Abbildung 26: Boxplot C-Test und Schulform 198

Tabellenverzeichnis

Tabelle 1: Die Kompetenzbereiche für die gymnasiale Oberstufe für das Fach Deutsch 24
Tabelle 2: C-Test-Sets 99
Tabelle 3: Itemanalyse und Schwierigkeitsindizes Testheft A 100
Tabelle 4: Tabelle: Itemanalyse und Schwierigkeitsindizes Testheft B 101
Tabelle 5: Einteilung der Leitungsgruppen 133
Tabelle 6: Verteilung der Daten in Bezug auf Standorttyp, Deutschnote, Alter und Geschlecht 137
Tabelle 7: Verteilung der Daten in Bezug auf Migrationshintergrund, Sprache und Sprachverwendung 139
Tabelle 8: Verteilung der Daten in Bezug auf Migrationshintergrund und Sprache 145
Tabelle 9: Verteilung der Daten in Bezug auf den Bildungsgrad der Eltern und Studienwunsch 146
Tabelle 10: Tabelle: Schwierigkeitskeitsindizes P (Schwierigkeit), $\overline{x}$ (Mittelwert), SD (Standardabweichung) 149
Tabelle 11: Deskriptive Statistik der C-Test-Rohwerte 150
Tabelle 12: Zuordnung zu den Leistungsbereichen 152
Tabelle 13: Verteilung der Stichprobe auf die Leistungsbereiche 153
Tabelle 14: Korrelationsanalysen zwischen C-Test und Lesetest 156
Tabelle 15: Verteilung der Variable *Deutschnote* auf die Leistungsbereiche 161
Tabelle 16: Korrelationsanalyse C-Test und Deutschnote 163
Tabelle 17: Überblick über die angewendeten Verfahren für Mittelwertvergleiche und bivariate Korrelationsanalysen 164
Tabelle 18: Verteilung der Variable *Phase in der Oberstufe* auf die Leistungsbereiche 166
Tabelle 19: Gruppenstatistik: C-Test-Ergebnis, Phase in der Oberstufe 167
Tabelle 20: t-Test bei unabhängigen Stichproben: C-Test-Ergebnis, Phase in der Oberstufe 167
Tabelle 21: Verteilung der Variable *Geschlecht* auf die Leistungsbereiche 170
Tabelle 22: Gruppenstatistik: C-Test-Ergebnis, Geschlecht 171
Tabelle 23: t-Test bei unabhängigen Stichproben: C-Test-Ergebnis, Geschlecht 171
Tabelle 24: Verteilung der Variable *Sprache* auf die Leistungsbereiche 175
Tabelle 25: Gruppenstatistik: C-Test-Ergebnis, Sprache 176
Tabelle 26: t-Test bei unabhängigen Stichproben: C-Test-Ergebnis, Sprache 176

Tabelle 27: Verteilung der Variable *Sprachverwendung innerhalb der Familie bei anderer Erstsprache als Deutsch* auf die Leistungsbereiche 179
Tabelle 28: Gruppenstatistik: C-Test-Ergebnis, Sprachverwendung bei nicht deutscher Erstsprache..... 179
Tabelle 29: t-Test bei unabhängigen Stichproben: C-Test-Ergebnis, Sprachverwendung bei nicht deutscher Erstsprache..... 180
Tabelle 30: Verteilung der Variable *Sprachgruppen* auf die Leistungsbereiche..... 182
Tabelle 31: Vergleich der Sprachgruppen..... 183
Tabelle 32: Verteilung der Variable *höchster Schulabschluss der Eltern* auf die Leistungsbereiche 188
Tabelle 33: Korrelationsanalyse C-Test und höchster Schulabschluss der Eltern..... 189
Tabelle 34: Verteilung der Variable *Standorttyp* auf die Leistungsbereiche 194
Tabelle 35: Korrelationsanalyse C-Test und Schulstandorttyp 195
Tabelle 36: Verteilung der Variable *Schulform* auf die Leistungsbereiche 198
Tabelle 37: Gruppenstatistik: C-Test-Ergebnis, Schulform 199
Tabelle 38: t-Test bei unabhängigen Stichproben: C-Test-Ergebnis, Schulform..... 199
Tabelle 39: Regressionsanalysen 203

Anhang

1. C-Tests

C-Test 1:

Der Reiz von Fernsehserien

Der offensichtliche Reiz einer Fernsehserie ist, dass sie weitergeht. Sie _____rd nicht _____t Anfang, _____tte und _____uss am _____ck erzählt _____e ein _____lm, sondern ___n einzelne _____le zerlegt. _____e Zuschauer _____llen nicht _____r auf _____r Handlungsebene _____ssen, wie _____e Geschichte _____tergeht, sie _____llen auch _____en, was _____re Lieblingsfiguren _____ben. Dieses _____esse ähnelt _____rem sozialen _____ltag. Die _____auer führen ___ogenannte parasoziale Beziehungen und bringen der Figur Empathie entgegen. Doch genau wie in sozialen Beziehungen können Figuren uninteressant und langweilig werden, weshalb manche Rezipienten die parasoziale Beziehung beenden und die Serie nicht weiterverfolgen.

C-Test 2:

Wissenschaftliches Arbeiten

Ab der Oberstufe wird nicht nur der Stoff anspruchsvoller: Auch neue Methoden werden für die Vorbereitung von Referaten und Präsentationen oder die Bewältigung der Facharbeit benötigt – wissenschaftliches Arbeiten ist gefordert. Viele _____iken des _____aftlichen Arbeitens _____nd zunächst _____ohnt. Plötzlich _____rd etwas _____es von _____tufenschülern und _____ngen Studierenden _____rtet. Zunächst _____ss eine _____enständige Fragestellung _____liert und _____ar abgegrenzt _____den. Dann _____lt sich _____e Frage: _____e recherchiert _____n geeignete _____atur? Was _____nd Quellen _____d wo _____det man _____e? Wie _____t eine _____te Facharbeit oder ein

durchdachtes Referat aufgebaut? Spätestens an der Uni wird dann von Studierenden erwartet, dass sie Hausarbeiten oder Essays schreiben, die neben einer interessanten Fragestellung und These sowie einer guten Argumentation auch noch den Kriterien wissenschaftlichen Arbeitens entsprechen.

C-Test 3:

Sprachverfall?

Sprachwandel kann schmerzhaft sein, vor allem für jene, die Sprache lieben. Denn _____de Veränderung _____llt zunächst _____en Verstoß _____en geltende _____eln dar. _____r die _____cklung von _____che über _____ere Zeit _____olgt, stellt _____och fest: _____e Fehler _____n heute _____den früher _____er später _____e Regeln _____n morgen. _____nn eine _____r Ursachen _____s Sprachwandels _____t der _____ng zur _____fachung. Er _____gte zum _____piel dafür, dass starke Verbformen wie `buk´ zu `backte´ mutierten. Wer aus Liebe zur Tradition auf alten Formen beharrt, läuft Gefahr, bald wie sein eigener Urgroßvater zu klingen.

C-Test 4:

Körpersprache

Für uns Menschen gibt es immer zwei Kommunikationsebenen. Die _____ne ist _____bal: Den _____alt einer _____mation vermitteln _____r durch _____s gesprochene _____rt. Die _____eite Ebene _____t non-verbal: _____persprache. Das _____ektive Erlebnis _____ser Gesprächssituation _____d die _____ühle und _____lungen, die _____r mit _____ser Information _____nden, artikulieren _____r selten _____ch Worte, _____och sind _____e in _____rer Gestik _____ch bestimmte _____ale erkennbar. Es ist wichtig, beide

Kommunikationsebenen richtig zu verstehen, denn oft bleibt eine wichtige zusätzliche Information unausgesprochen, oder es gibt gar einen Widerspruch zwischen dem Gesagten und dem Gemeinten, der Verwirrung und Missverständnisse stiftet.

C-Test 5:

Sind Jungen die neuen Bildungsverlierer?

Die gegenwärtig geführte Diskussion um eine mögliche Benachteiligung von Jungen in der Schule soll auf eine wissenschaftliche Basis gestellt werden. Eine _____ch Geschlecht _____nnte Analyse _____ischen Erfolgs _____ibt, dass _____gen in _____ngsbeteiligung und _____fikaterwerb gegenüber _____chen zurückgefallen _____nd, und _____ch im _____erb von _____enzen zeigen _____ch unverändert _____eile, z. B beim _____en. Jungen _____chen bei _____chen Kompetenzen _____chtere Noten _____s Mädchen _____d münden _____alb seltener _____n höhere _____ungslaufbahnen ein. _____s werden _____gungen angestellt, ob Schwierigkeiten, die Jungen möglicherweise bei der Entwicklung ihrer männlichen Identität im Umgang mit Lehrerinnen erfahren, zur Erklärung ihrer Leistungs- und Interessenentwicklung beitragen können.

C-Test 6:

Warum Schule Angst macht

Dem LBS- Kinderbarometer zufolge, einer bundesweiten Studie aus dem Jahr 2007, hat ein Drittel aller Kinder Angst davor, in der Schule zu versagen. Experten _____chen aufgrund _____r steigenden _____len von _____lern, die _____er Schulangst _____den, schon _____n einer _____echten Epidemie. _____ei liegen _____e Gründe _____r die _____ulangst in _____r Schule _____bst. Beim _____lichen Aufstehen _____t noch _____es gut, _____ch

spätestens _____im Frühstück _____nnt der _____pf zwischen _____ter und _____nd. Manche _____der reagieren _____t Bauchweh, andere reagieren mit Schlafstörungen oder Albträumen. Deshalb sollten Eltern aufmerksam werden, wenn ein Kind starke Angst vor Leistungsanforderungen hat und es vor Klassenarbeiten mehr als üblich besorgt und aufgeregt ist.

2. Berechnung C-Test und Deutschnote

Korrelation auf Klassenebene

Klasse	N	relatives N (in Prozent)	Spearman Korrelations-koeffizient	Spearman Korrelations-koeffizient: *fisher-z-transformiert*
Klasse_1	28	1,84%	-0,390	-0,412
Klasse_2	22	1,45%	-0,247	-0,252
Klasse_3	19	1,25%	-0,108	-0,109
Klasse_4	28	1,84%	-0,196	-0,199
Klasse_5	23	1,51%	0,331	0,344
Klasse_6	20	1,31%	-0,354	-0,370
Klasse_7	26	1,71%	-0,746	-0,964
Klasse_8	15	0,99%	-0,154	-0,156
Klasse_9	14	0,92%	-0,355	-0,371
Klasse_10	24	1,58%	-0,231	-0,235
Klasse_11	16	1,05%	-0,243	-0,248
Klasse_12	16	1,05%	-0,136	-0,137
Klasse_13	25	1,64%	-0,541	-0,606
Klasse_14	19	1,25%	0,418	0,445
Klasse_15	19	1,25%	-0,278	-0,286
Klasse_16	17	1,12%	-0,471	-0,511
Klasse_17	11	0,72%	-0,559	-0,631
Klasse_18	12	0,79%	-0,438	-0,470
Klasse_19	35	2,30%	-0,320	-0,332
Klasse_20	20	1,31%	-0,612	-0,712
Klasse_21	15	0,99%	-0,204	-0,207
Klasse_22	21	1,38%	-0,140	-0,141
Klasse_23	15	0,99%	-0,033	-0,033
Klasse_24	17	1,12%	-0,5	-0,549
Klasse_25	20	1,31%	-0,338	-0,352
Klasse_26	37	2,43%	-0,526	-0,585
Klasse_27	36	2,37%	-0,47	-0,510
Klasse_28	32	2,10%	-0,468	-0,508
Klasse_29	36	2,37%	-0,342	-0,356

Klasse_30	19	1,25%	0,264	0,270
Klasse_31	16	1,05%	0,247	0,252
Klasse_32	16	1,05%	-0,446	-0,480
Klasse_33	20	1,31%	-0,517	-0,572
Klasse_34	19	1,25%	-0,791	-1,074
Klasse_35	16	1,05%	-0,417	-0,444
Klasse_36	54	3,55%	0,061	0,061
Klasse_37	17	1,12%	0,206	0,209
Klasse_38	16	1,05%	-0,725	-0,918
Klasse_39	21	1,38%	-0,518	-0,574
Klasse_40	17	1,12%	-0,658	-0,789
Klasse_41	13	0,85%	-0,112	-0,112
Klasse_42	25	1,64%	0,023	0,023
Klasse_43	19	1,25%	0,469	0,509
Klasse_44	18	1,18%	-0,238	-0,243
Klasse_45	24	1,58%	0,296	0,305
Klasse_46	19	1,25%	0,214	0,217
Klasse_47	10	0,66%	0,157	0,158
Klasse_48	16	1,05%	-0,59	-0,678
Klasse_49	17	1,12%	-0,321	-0,333
Klasse_50	17	1,12%	-0,047	-0,047
Klasse_51	19	1,25%	-0,318	-0,329
Klasse_52	39	2,56%	-0,623	-0,730
Klasse_53	17	1,12%	-0,598	-0,690
Klasse_54	20	1,31%	-0,474	-0,515
Klasse_55	18	1,18%	-0,541	-0,606
Klasse_56	21	1,38%	-0,188	-0,190
Klasse_57	23	1,51%	-0,126	-0,127
Klasse_58	19	1,25%	-0,501	-0,551
Klasse_59	21	1,38%	-0,576	-0,656
Klasse_60	18	1,18%	-0,437	-0,469
Klasse_61	21	1,38%	-0,398	-0,421
Klasse_62	16	1,05%	-0,268	-0,275
Klasse_63	24	1,58%	-0,359	-0,375
Klasse_64	20	1,31%	-0,045	-0,045
Klasse_65	10	0,66%	0,486	0,531
Klasse_66	16	1,05%	-0,173	-0,174
Klasse_67	19	1,25%	-0,57	-0,648
Klasse_68	20	1,31%	0,194	0,196
Klasse_69	20	1,31%	-0,468	-0,508
Klasse_70	16	1,05%	-0,211	-0,214
Klasse_71	18	1,18%	-0,484	-0,528
Klasse_72	39	2,56%	-0,205	-0,208
Klasse_73	21	1,38%	-0,615	-0,717
Gesamt:	**1522**	**1,000**		

Vergleich der Mittelwerte: Einfaktorielle Varianzanalyse

Zwischensubjektfaktoren

		H
Was war Ihre letzte Zeugnisnote für das Fach Deutsch?	1	131
	2	523
	3	635
	4	223
	5	13

Levene-Test auf Gleichheit der Fehlervarianzen[a]

Abhängige Variable: C-Test (Gesamtrohwert)

F	df1	df2	Sig.
,250	4	1520	,910

Testet die Nullhypothese, dass die Fehlervarianz der abhängigen Variablen über Gruppen hinweg gleich ist.

a. Design: Konstanter Term + Note

Tests der Zwischensubjekteffekte

Abhängige Variable: C-Test (Gesamtrohwert)

Quelle	Typ III Quadratsumme	df	Quadratischer Mittelwert	F	Sig.	Partielles Eta hoch zwei	Dezentr. Parameter	Beobachtete Trennschärfe[b]
Korrigiertes Modell	57371,362[a]	4	14342,840	23,437	,000	,058	93,747	1,000
Konstanter Term	1769090,127	1	1769090,127	2890,758	,000	,655	2890,758	1,000
Note	57371,362	4	14342,840	23,437	,000	,058	93,747	1,000
Fehler	930211,828	1520	611,981					
Gesamtsumme	10844784,000	1525						
Korrigierter Gesamtwert	**987583,190**	**1524**						

a. R-Quadrat = ,058 (Angepasstes R-Quadrat = ,056)

b. Berechnet mit alpha = ,05

Mehrfachvergleiche

Abhängige Variable: C-Test (Gesamtrohwert)

Bonferroni

(I) Was war Ihre letzte Zeugnisnote für das Fach Deutsch?	(J) Was war Ihre letzte Zeugnisnote für das Fach Deutsch?	Mittelwertdifferenz (I-J)	Standardfehler	Sig.	95 % Konfidenzintervall	
					Untergrenze	Obergrenze
1	2	6,16	2,417	,109	-,64	12,95
	3	15,45*	2,374	,000	8,77	22,12
	4	19,75*	2,723	,000	12,09	27,40
	5	13,79	7,194	,554	-6,43	34,01
2	1	-6,16	2,417	,109	-12,95	,64
	3	9,29*	1,461	,000	5,18	13,39
	4	13,59*	1,978	,000	8,03	19,15
	5	7,63	6,946	1,000	-11,89	27,16
3	1	-15,45*	2,374	,000	-22,12	-8,77
	2	-9,29*	1,461	,000	-13,39	-5,18
	4	4,30	1,926	,256	-1,11	9,72
	5	-1,65	6,931	1,000	-21,14	17,83
4	1	-19,75*	2,723	,000	-27,40	-12,09
	2	-13,59*	1,978	,000	-19,15	-8,03
	3	-4,30	1,926	,256	-9,72	1,11
	5	-5,96	7,058	1,000	-25,80	13,89
5	1	-13,79	7,194	,554	-34,01	6,43
	2	-7,63	6,946	1,000	-27,16	11,89
	3	1,65	6,931	1,000	-17,83	21,14
	4	5,96	7,058	1,000	-13,89	25,80

Basierend auf beobachteten Mittelwerten.

Der Fehlerterm ist Mittel der Quadrate (Fehler) = 611,981.

*. die Mittelwertdifferenz ist auf der Stufe ,05 signifikant

3. Berechnungen C-Test und Sprachgruppen

Zwischensubjektfaktoren

	Wertbeschriftung	H
Vergleich Sprachgruppen	1 Deutsch	1037
	2 Türkisch	152
	3 Russisch	57
	4 Polnisch	38
	5 Arabisch	41
	6 Albanisch	33

Levene-Test auf Gleichheit der Fehlervarianzen[a]

Abhängige Variable: C-Test (Gesamtrohwert)

F	df1	df2	Sig.
,730	5	1352	,601

Testet die Nullhypothese, dass die Fehlervarianz der abhängigen Variablen über Gruppen hinweg gleich ist.

a. Design: Konstanter Term + Sprachgruppen

Tests der Zwischensubjekteffekte

Abhängige Variable: C-Test (Gesamtrohwert)

Quelle	Typ III Quadratsumme	df	Quadratischer Mittelwert	F	Sig.
Korrigiertes Modell	79494,047[a]	5	15898,809	27,629	,000
Konstanter Term	1774165,455	1	1774165,455	3083,206	,000
Sprachgruppen	79494,047	5	15898,809	27,629	,000
Fehler	777979,721	1352	575,429		
Gesamtsumme	9973584,000	1358			
Korrigierter Gesamtwert	**857473,767**	**1357**			

a. R-Quadrat = ,093 (Angepasstes R-Quadrat = ,089)

Mehrfachvergleiche

Abhängige Variable: C-Test (Gesamtrohwert)

Bonferroni

(I) Vergleich Sprachgruppen	(J) Vergleich Sprachgruppen		Standardfehler	Sig.	95 % Konfidenzintervall

		Mittelwertdifferenz (I-J)			Untergrenze	Obergrenze
Deutsch	Türkisch	18,32*	2,083	,000	12,19	24,44
	Russisch	13,60*	3,263	,000	4,01	23,20
	Polnisch	7,75	3,962	,762	-3,90	19,40
	Arabisch	27,10*	3,820	,000	15,87	38,33
	Albanisch	15,11*	4,242	,006	2,63	27,58
Türkisch	Deutsch	-18,32*	2,083	,000	-24,44	-12,19
	Russisch	-4,71	3,726	1,000	-15,67	6,24
	Polnisch	-10,57	4,351	,228	-23,37	2,22
	Arabisch	8,78	4,221	,564	-3,63	21,20
	Albanisch	-3,21	4,607	1,000	-16,76	10,34
Russisch	Deutsch	-13,60*	3,263	,000	-23,20	-4,01
	Türkisch	4,71	3,726	1,000	-6,24	15,67
	Polnisch	-5,86	5,024	1,000	-20,63	8,91
	Arabisch	13,50	4,912	,091	-,95	27,94
	Albanisch	1,50	5,247	1,000	-13,93	16,93
Polnisch	Deutsch	-7,75	3,962	,762	-19,40	3,90
	Türkisch	10,57	4,351	,228	-2,22	23,37
	Russisch	5,86	5,024	1,000	-8,91	20,63
	Arabisch	19,36*	5,402	,005	3,47	35,24
	Albanisch	7,36	5,708	1,000	-9,42	24,15
Arabisch	Deutsch	-27,10*	3,820	,000	-38,33	-15,87
	Türkisch	-8,78	4,221	,564	-21,20	3,63
	Russisch	-13,50	4,912	,091	-27,94	,95
	Polnisch	-19,36*	5,402	,005	-35,24	-3,47
	Albanisch	-11,99	5,610	,490	-28,49	4,50
Albanisch	Deutsch	-15,11*	4,242	,006	-27,58	-2,63
	Türkisch	3,21	4,607	1,000	-10,34	16,76
	Russisch	-1,50	5,247	1,000	-16,93	13,93
	Polnisch	-7,36	5,708	1,000	-24,15	9,42
	Arabisch	11,99	5,610	,490	-4,50	28,49

Basierend auf beobachteten Mittelwerten.

Der Fehlerterm ist Mittel der Quadrate (Fehler) = 575,429.

*. die Mittelwertdifferenz ist auf der Stufe 0,05 signifikant.

4. Berechnungen C-Test und Bildungsgrad

Einfaktorielle Varianzanalyse

Zwischensubjektfaktoren	Wertbeschriftung	H
Höchster Schulabschluss von Mutter ODER Vater	1 kein Bildungsabschluss	18
	2 Hauptschulabschluss	52
	3 Realschulabschluss	191
	4 Hochschulreife	348
	5 Akad. Abschluss	480

Levene-Test auf Gleichheit der Fehlervarianzen[a]

Abhängige Variable: C-Test (Gesamtrohwert)

F	df1	df2	Sig.
1,583	4	1084	,176

Testet die Nullhypothese, dass die Fehlervarianz der abhängigen Variablen über Gruppen hinweg gleich ist.

a. Design: Konstanter Term + Bildung_Eltern

Tests der Zwischensubjekteffekte

Abhängige Variable: C-Test (Gesamtrohwert)

Quelle	Typ III Quadratsumme	df	Quadratischer Mittelwert	F	Sig.	Partielles Eta hoch zwei
Korrigiertes Modell	17151,890[a]	4	4287,973	6,965	,000	,025
Konstanter Term	1795967,794	1	1795967,794	2917,372	,000	,729
Bildung_Eltern	17151,890	4	4287,973	6,965	,000	,025
Fehler	667322,839	1084	615,611			
Gesamtsumme	7972184,000	1089				
Korrigierter Gesamtwert	**684474,729**	**1088**				

a. R-Quadrat = ,025 (Angepasstes R-Quadrat = ,021)

Mehrfachvergleiche

Abhängige Variable: C-Test (Gesamtrohwert)

Bonferroni

(I) Höchster Schulabschluss von Mutter ODER Vater	(J) Höchster Schulabschluss von Mutter ODER Vater	Mittelwertdifferenz (I-J)	Standardfehler	Sig.	95 % Konfidenzintervall	
					Untergrenze	Obergrenze
kein Bildungsabschluss	Hauptschulabschluss	1,21	6,785	1,000	-17,88	20,29
	Realschulabschluss	-2,97	6,117	1,000	-20,18	14,23
	Hochschulreife	-6,49	5,997	1,000	-23,36	10,38
	Akad. Abschluss	-11,57	5,957	,523	-28,33	5,18
Hauptschulabschluss	kein Bildungsabschluss	-1,21	6,785	1,000	-20,29	17,88
	Realschulabschluss	-4,18	3,881	1,000	-15,10	6,74
	Hochschulreife	-7,70	3,689	,372	-18,07	2,68
	Akad. Abschluss	-12,78*	3,622	,004	-22,96	-2,59
Realschulabschluss	kein Bildungsabschluss	2,97	6,117	1,000	-14,23	20,18
	Hauptschulabschluss	4,18	3,881	1,000	-6,74	15,10
	Hochschulreife	-3,52	2,234	1,000	-9,80	2,77
	Akad. Abschluss	-8,60*	2,123	,001	-14,57	-2,63
Hochschulreife	kein Bildungsabschluss	6,49	5,997	1,000	-10,38	23,36
	Hauptschulabschluss	7,70	3,689	,372	-2,68	18,07
	Realschulabschluss	3,52	2,234	1,000	-2,77	9,80
	Akad. Abschluss	-5,08*	1,747	,037	-9,99	-,17
Akad. Abschluss	kein Bildungsabschluss	11,57	5,957	,523	-5,18	28,33
	Hauptschulabschluss	12,78*	3,622	,004	2,59	22,96
	Realschulabschluss	8,60*	2,123	,001	2,63	14,57
	Hochschulreife	5,08*	1,747	,037	,17	9,99

Basierend auf beobachteten Mittelwerten.

Der Fehlerterm ist Mittel der Quadrate (Fehler) = 615,611.

*. die Mittelwertdifferenz ist auf der Stufe ,05 signifikant.

Berechnung Zusammenhang Bildungsabschluss Vater und C-Test-Ergebnis

			Welcher ist der höchste Bildungsgrad Ihres Vaters?	CTest_GES
Spearman-Rho	Welcher ist der höchste Bildungsgrad Ihres Vaters?	Korrelationskoeffizient	1,000	,122**
		Sig. (2-seitig)	.	,000
		N	1191	1191
	CTest_GES	Korrelationskoeffizient	,122**	1,000
		Sig. (2-seitig)	,000	.
		N	1191	1585

**. Die Korrelation ist auf dem 0,01 Niveau signifikant (zweiseitig).

Berechnung Zusammenhang Bildungsabschluss Mutter und C-Test-Ergebnis

			Welcher ist der höchste Bildungsgrad Ihrer Mutter?	CTest_GES
Spearman-Rho	Welcher ist der höchste Bildungsgrad Ihrer Mutter?	Korrelationskoeffizient	1,000	,139**
		Sig. (2-seitig)	.	,000
		N	1216	1216
	CTest_GES	Korrelationskoeffizient	,139**	1,000
		Sig. (2-seitig)	,000	.
		N	1216	1585

**. Die Korrelation ist auf dem 0,01 Niveau signifikant (zweiseitig).

5. Berechnungen C-Test und Schulstandorttyp

Zwischensubjektfaktoren

		H
Standorttyp	1	128
	2	522
	3	116
	4	436
	5	383

Levene-Test auf Gleichheit der Fehlervarianzen[a]

Abhängige Variable: C-Test (Gesamtrohwert)

F	df1	df2	Sig.
7,477	4	1580	,000

Testet die Nullhypothese, dass die Fehlervarianz der abhängigen Variablen über Gruppen hinweg gleich ist.

a. Design: Konstanter Term + Standorttyp

Tests der Zwischensubjekteffekte

Abhängige Variable: C-Test (Gesamtrohwert)

Quelle	Typ III Quadratsumme	df	Quadratischer Mittelwert	F	Sig.
Korrigiertes Modell	49298,395[a]	4	12324,599	19,817	,000
Konstanter Term	7319408,976	1	7319408,976	11768,888	,000
Standorttyp	49298,395	4	12324,599	19,817	,000
Fehler	982647,350	1580	621,929		
Gesamtsumme	11137103,000	1585			
Korrigierter Gesamtwert	**1031945,745**	**1584**			

a. R-Quadrat = ,048 (Angepasstes R-Quadrat = ,045)

Mehrfachvergleiche

Abhängige Variable: C-Test (Gesamtrohwert)

Bonferroni

			Sig.	95 % Konfidenzintervall

(I) Standorttyp	(J) Standorttyp	Mittelwertdifferenz (I-J)	Standardfehler		Untergrenze	Obergrenze
1	2	5,38	2,460	,289	-1,54	12,29
	3	-3,15	3,197	1,000	-12,13	5,84
	4	14,52*	2,507	,000	7,47	21,57
	5	11,51*	2,546	,000	4,35	18,67
2	1	-5,38	2,460	,289	-12,29	1,54
	3	-8,52*	2,560	,009	-15,72	-1,33
	4	9,14*	1,618	,000	4,60	13,69
	5	6,13*	1,678	,003	1,42	10,85
3	1	3,15	3,197	1,000	-5,84	12,13
	2	8,52*	2,560	,009	1,33	15,72
	4	17,67*	2,605	,000	10,34	24,99
	5	14,66*	2,643	,000	7,23	22,09
4	1	-14,52*	2,507	,000	-21,57	-7,47
	2	-9,14*	1,618	,000	-13,69	-4,60
	3	-17,67*	2,605	,000	-24,99	-10,34
	5	-3,01	1,747	,849	-7,92	1,90
5	1	-11,51*	2,546	,000	-18,67	-4,35
	2	-6,13*	1,678	,003	-10,85	-1,42
	3	-14,66*	2,643	,000	-22,09	-7,23
	4	3,01	1,747	,849	-1,90	7,92

Basierend auf beobachteten Mittelwerten.

Der Fehlerterm ist Mittel der Quadrate (Fehler) = 621,929.

*. die Mittelwertdifferenz ist auf der Stufe ,05 signifikant.

6. Regression

Modellübersicht

Modell	R	R-Quadrat	Angepasstes R-Quadrat	Standardfehler der Schätzung
1	,396[a]	,157	,155	22,637
2	,487[b]	,237	,232	21,572
3	,597[c]	,356	,351	19,841
4	,597[d]	,356	,350	19,853

b. Prädiktoren: (Konstante), Sprache_Mono, Geschlecht, Phase in der Oberstufe, Schulform, Standorttyp

c. Prädiktoren: (Konstante), Sprache_Mono, Geschlecht, Phase in der Oberstufe, Schulform, Standorttyp, Was war Ihre letzte Zeugnisnote für das Fach Deutsch? , Aufgabenteil B - Leseverständnis

d. Prädiktoren: (Konstante), Sprache_Mono, Geschlecht, Phase in der Oberstufe , Schulform, Standorttyp, Was war Ihre letzte Zeugnisnote für das Fach Deutsch? , Aufgabenteil B - Leseverständnis, Höchster Schulabschluss von Mutter ODER Vater

ANOVA[a]

Modell		Quadratsumme	df	Mittel der Quadrate	F	Sig.
1	Regression	76214,362	2	38107,181	74,367	,000[b]
	Residuum	409935,728	800	512,420		
	Gesamtsumme	486150,090	802			
2	Regression	115261,940	5	23052,388	49,537	,000[c]
	Residuum	370888,149	797	465,355		
	Gesamtsumme	486150,090	802			
3	Regression	173186,922	7	24740,989	62,848	,000[d]
	Residuum	312963,168	795	393,664		
	Gesamtsumme	486150,090	802			
4	Regression	173189,346	8	21648,668	54,924	,000[e]
	Residuum	312960,744	794	394,157		
	Gesamtsumme	486150,090	802			

b. Prädiktoren: (Konstante), Sprache_Mono, Geschlecht

c. Prädiktoren: (Konstante), Sprache_Mono, Geschlecht, Phase in der Oberstufe , Schulform, Standorttyp

d. Prädiktoren: (Konstante), Sprache_Mono, Geschlecht, Phase in der Oberstufe , Schulform, Standorttyp, Was war Ihre letzte Zeugnisnote für das Fach Deutsch? , Aufgabenteil B - Leseverständnis

e. Prädiktoren: (Konstante), Sprache_Mono, Geschlecht, Phase in der Oberstufe , Schulform, Standorttyp, Was war Ihre letzte Zeugnisnote für das Fach Deutsch? , Aufgabenteil B - Leseverständnis, Höchster Schulabschluss von Mutter ODER Vater

Koeffizienten[a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	t	Sig.	Kollinearitäts-statistik	
		B	Standardfehler	Beta			Toleranz	VIF
1	(Konstante)	70,073	1,744		40,177	,000		
	Geschlecht	-5,444	1,598	-,111	-3,406	,001	1,000	1,000
	Sprache_Mono	21,051	1,799	,380	11,703	,000	1,000	1,000
2	(Konstante)	79,051	4,979		15,877	,000		
	Geschlecht	-3,925	1,533	-,080	-2,561	,011	,987	1,013
	Sprache_Mono	17,197	1,786	,310	9,631	,000	,922	1,085

	Phase in der Oberstufe	10,947	1,713	,204	6,389	,000	,936	1,069
	Schulform	-11,703	3,550	-,107	-3,296	,001	,907	1,102
	Standorttyp	-3,189	,671	-,158	-4,751	,000	,865	1,156
3	(Konstante)	81,948	5,352		15,310	,000		
	Geschlecht	-1,299	1,428	-,026	-,909	,363	,962	1,039
	Sprache_Mono	12,444	1,699	,225	7,326	,000	,862	1,161
	Phase in der Oberstufe	12,142	1,595	,227	7,613	,000	,913	1,095
	Schulform	-10,165	3,268	-,093	-3,110	,002	,906	1,104
	Standorttyp	-3,257	,618	-,161	-5,274	,000	,864	1,157
	Was war Ihre letzte Zeugnisnote für das Fach Deutsch?	-6,587	,856	-,231	-7,691	,000	,894	1,118
	Aufgabenteil B - Leseverständnis	,950	,131	,222	7,241	,000	,865	1,156
4	(Konstante)	81,627	6,740		12,110	,000		
	Geschlecht	-1,305	1,431	-,027	-,912	,362	,959	1,042
	Sprache_Mono	12,429	1,711	,224	7,264	,000	,850	1,176
	Phase in der Oberstufe	12,144	1,596	,227	7,609	,000	,913	1,095
	Schulform	-10,127	3,305	-,093	-3,064	,002	,887	1,128
	Standorttyp	-3,253	,620	-,161	-5,249	,000	,859	1,164
	Was war Ihre letzte Zeugnisnote für das Fach Deutsch?	-6,577	,866	-,231	-7,597	,000	,876	1,141
	Aufgabenteil B - Leseverständnis	,949	,132	,221	7,185	,000	,855	1,170
	Höchster Schulabschluss von Mutter ODER Vater	,064	,820	,002	,078	,938	,888	1,126

a. Abhängige Variable: C-Test (Gesamtrohwert)