

X25

English is generally acknowledged to be the world's most important language. It _____ perhaps _____ th glancing _____ fly at _____ e basis _____ r that _____ ation. There _____ e, after _____ l, thousands _____ f different _____ ages in _____ e world, _____ d each _____ ll seem _____ uely important _____ o those _____ o speak _____ t as _____ ir native _____ uage, the _____ uage they _____ ired at _____ ir mother's _____ ee.

Long before man first caught sight of his home planet from a spacecraft, astronomers, meteorologists and physicists had been pondering the question of how the Earth would appear when viewed from outer space. It _____ d already _____ en realized - _____ ng before _____ e first _____ ce flights - _____ at the _____ th could _____ t possibly _____ ok like a _____ be although _____ ch an _____ ge repeatedly _____ sted itself. _____ t should _____ t be _____ tten that _____ e Earth's _____ phere is _____ ly very _____ ely so _____ etely clear _____ d transparent _____ at an _____ rver from _____ ide might _____ nize continents _____ d oceans, _____ nds and _____ as as _____ rly as _____ n a _____ p.

Children are curious. Everything _____ ey encounter _____ s marvellous _____ d astonishing. _____ ey would _____ ke to _____ stand, and _____ s soon _____ s they _____ n talk _____ ey ask _____ ions. This _____ ing desire _____ o understand, _____ is appetite _____ r knowledge _____ nues in a _____ what more _____ cting and _____ ound form _____ ing adolescence, _____ ch is _____ fore the _____ ral age _____ r starting _____ rsity.

The passages are taken from Quirk et al. (1985, p. 3), Haber (1967, p. 9), and Belloc & Nègre (1971, p. 310).

Marion Meißner-Stiffel und Ulrich Raatz

_____ oder _____ ?

Zwei Grundlagenuntersuchungen zum C-Prinzip bei L1-Lernern*

Bei Untersuchungen von Süßmilch (1984, 1985) und Bresnihan & Ray (1994) hat sich gezeigt, daß C-Tests einen geringeren Schwierigkeitsgrad besitzen, wenn man an den Stellen, wo Wortteile getilgt sind, die meist übliche durchgezogene Linie durch so viele Striche ersetzt, wie Buchstaben fehlen. Dabei blieb jedoch die Frage offen, ob und gegebenenfalls wie diese Änderung die Reliabilität und die Validität der untersuchten C-Tests beeinflusst. Wir haben dieses Problem zunächst in einer ersten Studie bei Schülern aus 4. Klassen mit dem CT-D4 (Raatz & Klein-Braley, 1992) in der Muttersprache Deutsch untersucht. Dabei zeigte sich, daß der modifizierte C-Test erwartungsgemäß deutlich leichter wurde. Zu unserer Überraschung stiegen jedoch Reliabilität und Validität – letztere als Korrelationen mit verschiedenen Deutschnoten – stark an. In einer zweiten Untersuchung an einer größeren Stichprobe, diesmal an Schülern aus 3. Klassen, konnten wir diese Ergebnisse bestätigen.

1. Einführung

Bei einem 'klassischen' C-Test wird, beginnend mit dem zweiten Satz eines Textes, die zweite Hälfte eines jeden zweiten Wortes getilgt. Der fehlende Wortteil muß durch den Probanden ergänzt werden. Die Lücke wird in der Regel durch eine durchgezogene, immer gleich lange Linie markiert. Die Instruktion lautet in diesem Falle beispielsweise:

“Bei einigen Worten fehlt die zweite Hälfte. Ihr sollt jetzt erraten, was fehlt, und die fehlenden Wortteile in die Lücken einsetzen.” (Raatz & Klein-Braley, 1992, S. 6)

Süßmilch (1984, 1985) ist bei der Entwicklung von C-Tests für ausländische Schüler von diesem Prinzip abgewichen. Er hat die durchgezogene Linie durch so viele Striche ersetzt, wie Buchstaben fehlen, weil

“dadurch Texte insgesamt leichter werden, denn es bleibt meist nur noch eine Lösungsmöglichkeit übrig, die schneller gefunden wird.” (Süßmilch 1985, S. 77)

* Stark verkürzte Fassung der Diplomarbeit “CT-D in 3. Klassen” von M. Meißner-Stiffel zur Erlangung des Grades des Dipl. Päd.

Folgendes Textbeispiel (Instruktion des CT-D4) macht die Veränderung deutlich:

Klassische Form

Sechs Läufer standen am Start, standen gespannt und bewegungslos und dicht nebeneinander, und e_____ war s_____ still i_____ Stadion, d_____ das Knat_____ der Fah_____ im Win_____ zu hö_____ war. Al_____ warteten a_____ den Startschuß.

Modifizierte Form

Sechs Läufer standen am Start, standen gespannt und bewegungslos und dicht nebeneinander, und e_ war s_ still i_ Stadion, d_ _ das Knat_ _ _ der Fah_ _ _ im Wi_ _ zu hö_ _ _ war. Al_ _ warteten a_ _ den Startschuß.

In einer jüngst veröffentlichten Studie von Bresnihan & Ray (1994) wird dieselbe Technik verwendet, um das Prinzip des C-Tests Zweitsprachenlernern in einer multilingualen Testsituation (z.B. bei Sommerkursen an englischen Universitäten) nonverbal zu erklären. Die Autoren fanden ebenfalls heraus, daß die C-Tests durch diese Modifikation leichter werden.

Bei beiden Untersuchungen ist die Frage offen geblieben, ob sich durch die Modifikation der C-Tests Reliabilität und Validität geändert haben. So könnte man vermuten, daß durch das bessere Verständnis der Testanweisung und die größere Eindeutigkeit der Lösungen und die damit verbundene höhere Auswertungsobjektivität die Reliabilität des Tests ansteigen könnte, was Auswirkungen auf die Validität hätte. Andererseits wird von Gegnern dieser Modifikation immer wieder angeführt, daß die Probanden durch die Vorgabe der Striche für jeden fehlenden Buchstaben zum "Buchstabenzählen" angeregt würden, was letztlich nicht der Sinn des Tests sein könnte. Sie befürchten bei diesem Prinzip eine Herabsetzung der Konstruktvalidität.

Die aufgeworfenen Fragen sollen in den beiden im folgenden beschriebenen Untersuchungen geklärt werden.

2. Erste Untersuchung

2.1 Durchführung

In dieser Untersuchung sollten die Form A des CT-D4 in der Originalfassung und in der modifizierten Fassung in zwei randomisierten Stichproben von Schülerinnen und Schülern aus vierten Klassen eingesetzt und bezüglich Schwierigkeit, Reliabilität und Validität verglichen werden.

Die Studie wurde im vierten Schuljahr in drei Parallelklassen in einer Grundschule im Raum Moers (NRW) im Monat Mai, also in der zweiten Hälfte des Schuljahrs, wie im Manual des CT-D4 gefordert, durchgeführt. In allen drei Klassen erhielten die Schüler im Wechsel die beiden Testformen. Instruktion und zeitliche Vorgaben entsprachen der Originalversion des Tests.

Die Stichprobe dieser Untersuchung umfaßte 47 deutsche und 8 ausländische Schülerinnen und Schüler. Alle ausländischen Schüler beherrschten die deutsche Sprache in hinreichendem Maße, so daß ihre Ergebnisse verwendet werden konnten. Die Originalversion wurde von 9 Jungen und 18 Mädchen (Kontrollgruppe), die modifizierte Form von 14 Jungen und 14 Mädchen (Versuchsgruppe) bearbeitet.

Als Validitätskriterien wurden die letzten Schulnoten im Fach Deutsch in Lesen, Aufsatz, Rechtschreibung und mündlicher Mitarbeit herangezogen. Diese Noten wurden von den Lehrerinnen während der Testdurchführung in vorgefertigte Listen eingetragen, wobei der jeweilige Name aus Datenschutzgründen verschlüsselt wurde.

2.2 Ergebnisse

Zunächst haben wir geprüft, ob Versuchs- und Kontrollgruppe bzgl. Zusammensetzung und Ausgangsleistungen im Fach Deutsch vergleichbar sind. Die entsprechenden Daten finden sich in den Tabellen 1 und 2.

Tabelle 1
Gruppengrößen (n)

	Versuchsgruppe	Kontrollgruppe
Jungen	14	9
Mädchen	14	18
Gesamt	28	27

Der Unterschied im Anteil von Jungen und Mädchen in Versuchs- und Kontrollgruppe ist nicht signifikant ($\chi^2 = 0.96, p < .33$); auch die Mittelwerte unterscheiden sich nicht signifikant.

Tabelle 2
Mittelwerte (\bar{x}) und Standardabweichungen (s) der Noten

	Versuchsgruppe		Kontrollgruppe	
	\bar{x}	s	\bar{x}	s
Lesen	3.1	1.3	3.1	1.0
Rechtschreibung	3.2	1.4	3.2	1.1
Aufsatz	3.1	1.4	3.3	1.1
mündliche Mitarbeit	3.1	1.4	3.3	1.1

Anschließend haben wir die Mittelwerte und Standardabweichungen der Ergebnisse in den beiden C-Test-Versionen, die Reliabilitäten (Alpha-Formel nach Cronbach) und die Korrelationen zwischen C-Test und Noten als Maß für die Validität berechnet. Die Ergebnisse sind in Tab. 3 zusammengestellt. Der Mittelwertsunterschied im C-Test ist signifikant ($t = 2.07, p < .05$).

Tabelle 3
Mittelwerte (Schwierigkeiten), Standardabweichungen, Reliabilitäten und Validitäten der C-Tests

	Versuchsgruppe	Kontrollgruppe
Mittelwert	69.1	57.4
Standardabweichung	22.4	19.4
Reliabilität	.90	.75
Validitäten		
Lesen	-.79	-.48
Rechtschreiben	-.85	-.45
Aufsatz	-.87	-.33
mündliche Mitarbeit	-.84	-.51

Die Ergebnisse unserer ersten Untersuchung lassen sich wie folgt zusammenfassen:

1. Der C-Test ist durch die Modifikation deutlich leichter geworden. Der Unterschied betrug bei dieser Untersuchung ca. 12 Punkte.

2. Die modifizierte Testform hat eine sehr zufriedenstellende und bei C-Tests eigentlich übliche Reliabilität von .90. Die Reliabilität der Originalform liegt mit .75 deutlich niedriger und unterschreitet damit den im Testmanual zum CT-D4 angegebenen Wert von .90.

3. Auch bei den Validitäten, d.h. bei den Korrelationen mit den verschiedenen Noten im Fach Deutsch, ergibt sich ein deutlicher Vorsprung der modifizierten Testform. Während diese Korrelationen bei der Originalform in der Größenordnung des im Testmanual angegebenen Werts von .60 liegen, sind die Korrelationen der Versuchsform bemerkenswert hoch und erreichen das Niveau von Reliabilitätskoeffizienten. Dabei gibt es keine interpretierbaren Unterschiede zwischen den einzelnen Teilnoten.

Zusammenfassend läßt sich sagen, daß die modifizierte Testform der Originalform hinsichtlich Reliabilität und Validität wahrscheinlich deutlich überlegen ist. Allerdings ist diese Version jetzt mit einer Schwierigkeit von 69.1 für vierte Klassen voraussichtlich zu leicht. Es ist zu überlegen, ob der CT-D4 in der modifizierten Version in dritten Klassen anwendbar ist. Diese Frage soll in der zweiten, im folgenden beschriebenen Untersuchung beantwortet werden.

3. Zweite Untersuchung

3.1 Durchführung

Der CT-D4 war als globaler Deutschtest seinerzeit für dritte Klassen konzipiert worden. Bei den ersten Probedurchführungen hatte sich jedoch herausgestellt, daß dieser Test in der vorliegenden Form für dritte Klassen zu schwer war. Deshalb wurde damals der Test endgültig für vierte Klassen normiert (Raatz & Klein-Braley, 1992). Der Bedarf für einen entsprechenden Test für dritte Klassen besteht allerdings immer noch, beispielsweise im Zusammenhang mit der Diagnose einer Lese-Rechtschreib-Schwäche. Diese Lücke könnte durch die modifizierte Form des CT-D4 geschlossen werden, wie unsere erste Studie gezeigt hat.

Die Klärung dieser Frage ist Gegenstand der im folgenden beschriebenen zweiten Untersuchung. Dabei sind an einer größeren Stichprobe von Schülern aus dritten Klassen Schwierigkeitsgrad, Reliabilität und Validität der modifizierten Testform zu überprüfen. Zusätzlich soll noch der Einfluß des Geschlechts auf die Testergebnisse untersucht werden.

Die Untersuchung fand ein Jahr später wiederum gegen Ende des Schuljahrs bei 293 Schülerinnen und Schülern aus insgesamt 14 dritten Klassen aus sechs Schulen im Großraum Duisburg statt. In Tabelle 4 wird diese Stichprobe näher beschrieben.

Tabelle 4
Beschreibung der Stichprobe (n = 293)

Geschlecht	Nationalität				Gesamt
	deutsch	türkisch	polnisch	sonstige*	
Jungen	116	14	4	8	142
Mädchen	115	20	5	11	151
Gesamt	231	34	9	19	293

* darunter fallen die Nationalitäten: russisch, chinesisches, angolansich, liberianisch, spanisch, italienisch, marokkanisch, jugoslawisch, britisch und libanesich

Die Klassen, die an unserer Untersuchung teilgenommen haben, unterschieden sich in der Schülerzahl und in ihrer Zusammensetzung stark. Der Umfang schwankte zwischen 15 und 28 Schülern. Es gab fast reine Mädchenklassen und Klassen mit einem erheblichen Jungenüberschuß. Der Anteil der ausländischen Schüler lag zwischen 0% und 50%.

Als Vergleichskriterium wurden die Schulnoten im Fach Deutsch erhoben. Diese Deutschnote setzt sich im dritten Schuljahr aus drei Einzelnoten zusammen: Lesen, Rechtschreibung und mündliche Mitarbeit.

Als Testzeitpunkt wurde jeweils die 2. oder 3. Schulstunde gewählt, da die Schüler zu dieser Zeit am aufnahme- und leistungsfähigsten sind. So ist es auch im Manual des CT-D4 vorgesehen (Raatz & Klein-Braley, 1992).

Bei der Durchführung der Tests, insbesondere bei der Instruktion folgten wir weitgehend dem Manual zum CT-D4. Die Lehrer(innen) waren vor allem bei der Aufsicht behilflich, da die Form des Tests zum Abschreiben verleitete und wir die Parallelform nicht benutzten.

Während die Schüler arbeiteten, gaben die Klassenlehrer(innen) alle Zusatzinformationen (Alter, Geschlecht, Nationalität, Noten), die in vorgefertigten Listen den jeweiligen Nummern des Klassenbuchs zugeordnet wurden. Diese Nummern schrieben die Kinder auch statt ihres Namens auf die Testbögen. Damit wurde ihre Anonymität gewahrt.

Bei einigen (vor allem den ausländischen) Kindern lagen keine Noten vom letzten Halbjahreszeugnis vor, da die entsprechenden Kinder neu in der Schule waren. In diesen Fällen gaben die Lehrer aktuelle Einschätzungen der Leistungen anhand schon geschriebener Klassenarbeiten.

Mit der vorgegebenen Zeit (30 min.) kamen fast alle Schüler aus.

3.2 Ergebnisse

Tabelle 5 enthält Informationen über Mittelwert (Schwierigkeit), Standardabweichung, Reliabilität und Validitäten des modifizierten C-Tests. Diese Daten beruhen nur auf den Testergebnissen der deutschen Schüler.

Die Reliabilität wurde wieder nach Cronbach abgeschätzt. Zur Bestimmung der Validität wurden die Korrelationen zwischen Testergebnis und den drei Teilnoten im Fach Deutsch berechnet.

Tabelle 5
Testergebnisse bei den deutschen Schülern

Mittelwert (Schwierigkeit)	53.6
Standardabweichung	21.6
Reliabilität	.87
Validitäten	
Lesen	-.63
Rechtschreibung	-.65
mündliche Mitarbeit	-.53
Gesamtnote	-.70

Zwischen den einzelnen Klassen ergaben sich auch hier wieder große Unterschiede. Der Schwierigkeitsgrad des C-Tests schwankte zwischen 35.8 und 65.6. Die Validität als Korrelation zur Gesamtnote lag zwischen .58 und .88. Versuche, diese Unterschiede z.B. durch die Klassenstärke, die Zusammensetzung bzgl. des Geschlechts oder den jeweiligen Ausländeranteil zu erklären, sind wegen der geringen Anzahl der Klassen, aber auch aus methodologischen Gründen nicht sinnvoll.

Insgesamt gesehen hat der CT-D4 in der modifizierten Form für dritte Klassen die erwünschte mittlere Schwierigkeit. Die Reliabilität ist zufriedenstellend. Die Validität als Korrelation mit den Noten ist zwar nicht ganz

so hoch wie bei der ersten Untersuchung. Sie ist aber immer noch als sehr gut zu beurteilen, und sie liegt im Rahmen der bei C-Tests üblichen Werte.

Bei den ausländischen Schülern waren mit $n = 34$ die türkischen Schüler die größte Gruppe. Für diese betrug das durchschnittliche Testergebnis 36.8 Punkte. Hier lagen die Validitäten mit .08 (Lesen), .43 (Rechtschreibung), .25 (mündlich) und .35 (Gesamtnote) deutlich niedriger. Ein ähnliches Ergebnis mit .46 gegenüber einem Wert von .57 bei deutschen Schülern hatte sich auch bei der Validierung des CT-D4 gezeigt (Raatz & Klein-Braley, 1992, S. 14). Wir vermuten, daß Lehrer und Lehrerinnen bei der Beurteilung der Schulleistungen von türkischen Schülern sehr unsicher sind, insbesondere dann, wenn sie keine 'harten' Daten, wie z.B. Leistungen im Diktat, zur Verfügung haben.

Wenn man die Testergebnisse und die Noten der deutschen Schülerinnen und Schüler nach dem Geschlecht aufschlüsselt, ergeben sich die in Tabelle 6 aufgeführten Werte.

Tabelle 6
Mittelwerte der Testergebnisse und Schulnoten
bei Jungen und Mädchen

	Jungen	Mädchen
C-Test	51.1	56.1
Lesen	2.6	2.4
Rechtschreibung	2.9	2.5
mündliche Mitarbeit	2.5	2.4
Gesamtnote	2.7	2.5

Es ergibt sich bei allen Leistungskriterien tendenziell ein besseres Abschneiden der Mädchen. Signifikanz liegt allerdings nur bei der Rechtschreibnote vor ($t = 2.90$, $p < .01$). Auch im CT-D4 haben die Mädchen mit durchschnittlich 62.8 Punkten gegenüber den Jungen mit 55.5 Punkten besser abgeschnitten (Raatz & Klein-Braley, 1992, S. 15). Die höhere Leistung der Mädchen in Sprachtests entspricht der Erwartung (s. z.B. Coates, 1986; Rathenow, Laupenmühlen & Vöge, 1980). Zur Erklärung des Unterschiedes bei den Schulnoten kommt vielleicht noch hinzu, daß Mädchen meist besser beurteilt werden als ihre Klassenkameraden (Schröter, 1981, S. 244).

4. Zusammenfassung

Wir konnten in diesen beiden Untersuchungen zeigen, daß die Modifikation des C-Prinzips, nämlich bei den getilgten Wortteilen die durchgezogene Linie durch so viele Striche zu ersetzen, wie Buchstaben fehlen, den Schwierigkeitsgrad des Tests herabsetzt, dabei aber die Reliabilität und die Validität unverändert läßt, wenn nicht sogar erhöht. Damit eröffnet sich die Möglichkeit, den Anwendungsbereich von C-Tests auch für weniger fähige oder jüngere Probanden oder auch für L2-Lerner zu erweitern.

Einschränkend ist anzumerken, daß sich unsere Ergebnisse zunächst nur auf die Sprache Deutsch bei L1-Lernern beziehen. Weitere Untersuchungen müssen zeigen, ob die vorgeschlagene Modifikation auch in anderen Sprachen bzw. bei Fremdsprachenlernern eingesetzt werden kann.

Literaturverzeichnis

- Bresnihan, Brian & Ray, Stratton. (1994). C-Tests and the usefulness of non-linguistic instructions. In Rüdiger Grotjahn (Hrsg.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Bd. 2, S. 193-216). Bochum: Brockmeyer.
- Coates, Jennifer. (1986). *Woman, men and language*. London: Longman.
- Raatz, Ulrich & Klein-Braley, Christine. (1992). *Schulleistungstest Deutsch CT-D4*. Weinheim: Beltz.
- Rathenow, P., Laupenmühlen, D. & Vöge, J. (1980). *Westermann Rechtschreibtest 6 +*. Braunschweig: Westermann.
- Schröter, Gottfried. (1981). *Zensuren? Zensuren!* Baltmannsweiler: Schröter/Pädagogischer Verlag.
- Süßmilch, Edgar. (1984). Language testing with immigrant children. In Terry Culhane, Christine Klein-Braley & Douglas K. Stevenson (Hrsg.) *Practice and problems in language testing 7. Proceedings of the Seventh International Language Testing Symposium of the IUS, Colchester, October 1982* (S. 167-176). Colchester: University of Essex, Dept. of Language and Linguistics.
- Süßmilch, Edgar. (1985). C-Tests für ausländische Schüler: Sprachdiagnose im Unterricht Deutsch als Zweitsprache. In Christine Klein-Braley & Ulrich Raatz (Hrsg.), *Fremdsprachen und Hochschule 13/14: Thematischer Teil: C-Tests in der Praxis* (S. 72-82). Bochum: AKS.
- Wendeler, Jürgen. (1981). *Lernzieltests im Unterricht*. Weinheim: Beltz.