

Günther Sigott and Johann Köberl

Deletion patterns and C-Test difficulty across languages

This article examines ways of avoiding ceiling effects in German and English C-Tests when used with native speakers or very advanced FL learners. The effect of three modifications of the original C-principle on the difficulty of translationally equivalent texts for native speakers is determined and compared across languages.

1. Introduction

Among the virtues of the C-Test Klein-Braley (1994, p. 42) lists the fact that native speakers normally achieve near-perfect scores. There are, however, some problems involved here:

- It is uncertain whether native speaker ability to achieve perfect scores on a language test should be considered an indication of validity.
- Perfect or near-perfect native-speaker scores on C-Tests make the procedure unsuitable as a measure of text difficulty for native speakers.
- Experience has shown that C-Tests often prove too easy as measures of general language proficiency even for advanced foreign language learners.

Grotjahn (1987, pp. 224ff.) suggests ways of modifying the deletion pattern in order to adjust test difficulty. The effect of changes in the deletion pattern has been investigated by Kokkota (1988), Cleary (1988), Cohen, Segal & Weiss Bar-Siman-Tov (1984), and Köberl & Sigott (1994).

In Köberl & Sigott (1994) we applied three deletion patterns in addition to the one originally suggested by Raatz & Klein-Braley (1985) to four German texts, one originally in German, two translated from English, and one from French. Two of our deletion patterns involve increasing the extent of redundancy reduction by increasing the number of letters deleted: In one of them (C33) two thirds rather than one half of every second word are deleted; the other (CFL) leaves only the first letter of every second word. The third pattern (X25) does not involve increasing the proportion

of letters deleted and thus increasing redundancy reduction, but deletes the first – rather than the second – half of every other word in the text.

All three deletion patterns were found to yield tests which are more difficult than the original C-Test format (C25). When the test was administered to 60 16-to-17-year old native speakers of German, C33 lowered test facility by just above 20 percentage points, CFL by about 35 percentage points, and, interestingly, X25, with the same deletion rate as C25, by almost 34 percentage points.

This paper describes the results of an analogous study carried out with comparable native speakers of English and compares the results with those of the German study. The same texts as in Köberl & Sigott (1994) were used, but this time in their English version.

2. Method

2.1 Materials and procedures

As in Köberl & Sigott (1994), only three of the four passages were used in the statistical analyses. The first passage was considered a warming-up exercise and put at the beginning of the test in order to absorb a possible practice effect. By applying four different deletion patterns (C25, C33, CFL, X25) to the set of four texts, four test versions were constructed (for details of construction see Köberl & Sigott, 1994). Since the tests were devised with crosslinguistic comparability in mind (for a discussion see Sigott, 1992), deletion of words was allowed to run to the end of the sentence in which the 20th deletion was reached rather than stop after the 20th deletion and leave the rest of the sentence intact. This was done in order to ensure that the deletion procedure was applied to the same extent to translationally equivalent portions of text in the two languages. As a result, the number of deletions in the German version of each passage is different from the number in the English version. The extent of redundancy reduction, defined as the percentage of letters deleted, is given for each version (C25, C33, CFL, X25) in both languages in Table 1.

The four test versions were randomly assigned to the subjects (Ss). Testing was carried out during regular class periods. The tests were scored according to the "exact word" method by a native speaker of English in consultation with the authors; spelling errors were tolerated.

Table 1
Redundancy reduction in the four test versions
in German and English

Texts 2 - 4	Letters (Percentage deleted)	C25	C33	CFL	X25
German	978	700 (28.43)	622 (36.40)	559 (42.84)	700 (28.43)
English	826	601 (27.24)	537 (34.99)	496 (39.95)	601 (27.24)

2.2 Subjects

82 sixth-form pupils at a school in Bristol (UK) were tested.¹ The Ss were chosen so as to ensure maximum comparability with the Ss of the German study. All Ss were approximately 17 years old and had English as their first language.

3. Results

3.1 Findings

3.1.1 Reliability

As in the German study, internal consistency was estimated by means of Cronbach's Alpha. The coefficients for the four versions and for the two languages are given in Table 2. Since the number of deletions is not the same in the German and in the English version of the test, the coefficients for German cannot be directly compared to those for English. The coefficients can, however, be compared among the four versions within each language because the number of deletions for all four versions in a language is the same.

3.1.2 Redundancy reduction and test difficulty

In both languages the relationship between redundancy reduction and test difficulty is approximately linear, as can be seen in Figure 1. Changing the

¹ We are greatly indebted to Mr. P. Heaney and Mr. L. Hewson for administering the tests, and to the Sixth Form students of Ridings High School, Bristol, for their cooperation.

Table 2
Cronbach's alpha reliability coefficient for four versions of C-Tests in German and English

Test Version	German		English	
	N	Alpha	N	Alpha
C25	14	.16	20	.82
C33	15	.85	21	.88
CFL	17	.76	20	.92
X25	14	.86	20	.84

deletion pattern from C25 to C33, thus increasing redundancy reduction by about 8 percentage points, decreases test facility by about 20 percentage points in both languages. A further change from C33 to CFL, involving an increase in redundancy reduction by about 6 and 5 percentage points respectively, lowers test facility by about 14 percentage points in German and about 10 percentage points in English.

X25, however, behaves differently in the two languages. Without reducing redundancy in comparison with C25, it brings down test facility by ca. 34 percentage points in German, but only by 15 percentage points in English, i.e. less than either C33 or CFL.

3.1.3 Comparison of test version means

For each test version the unweighted mean was computed. Test version and language means and standard deviations are given in Table 3; means are graphically presented in Figure 1.

Table 3
Means and standard deviations (SD) for four test versions in German and English

Test version	German			English		
	N	mean	SD	N	mean	SD
C25	14	88.1	2.7	21	81.9	8.6
C33	15	67.4	9.2	20	61.1	14.7
CFL	17	53.0	7.5	20	51.7	9.0
X25	14	54.4	9.5	21	67.0	10.9

Figure 1
The relationship between redundancy reduction and test version means in German and English

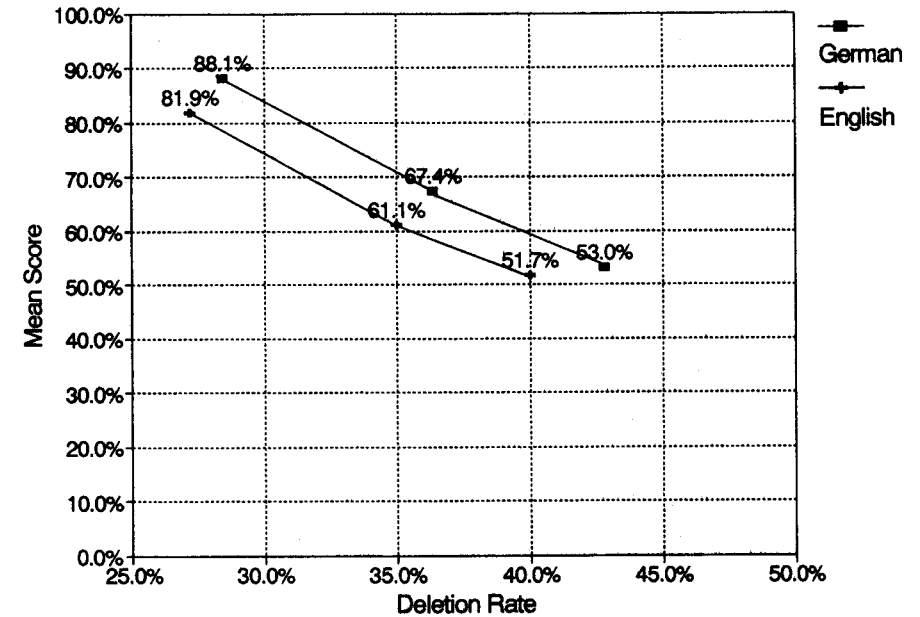
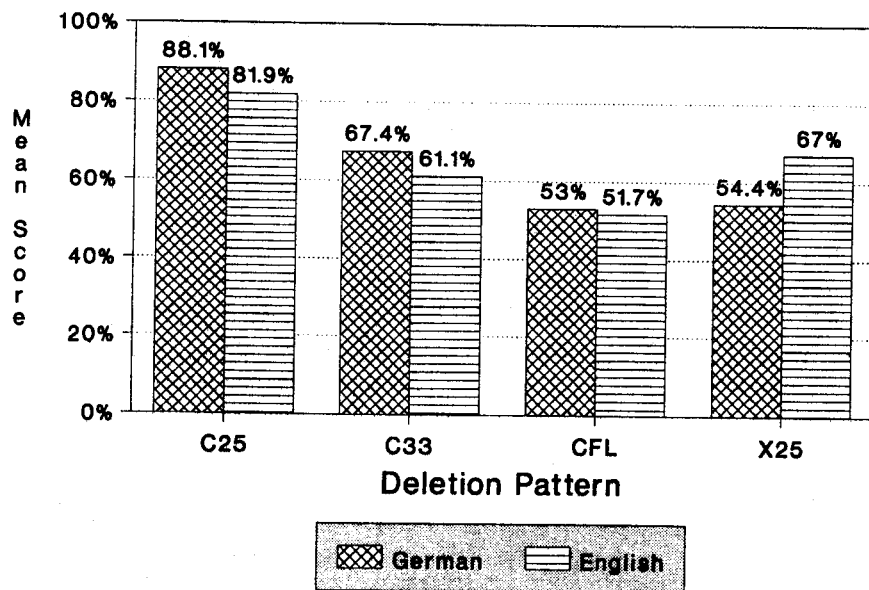


Table 4
Significant differences between pairs of test version means

	German			English		
	C25	C33	CFL	C25	C33	CFL
C33	*	—	—	*	—	—
CFL	*	*	—	*	*	—
X25	*	*		*		*

* indicates a difference significant at the .05 level as determined by Duncan procedure

Figure 2
Means for four test versions in German and English



The differences between the test version means in English were tested for statistical significance by means of analysis of variance. As in the German study, the effect of test version was found to be highly significant ($F = 27.16, p = .000$). Ten out of twelve differences between pairs of test versions are significant; in German, CFL does not differ significantly from X25, while in English the difference between C33 and X25 fails to reach significance (Duncan procedure; see Table 4).

In order to compare the test version means across languages, in a second step the factor LANGUAGE was integrated into the analysis of variance design. The results can be seen in Table 5. In addition to the effect of the factor VERSION there is now a significant interaction VERSION BY LANGUAGE, indicating that the effect of the different deletion patterns, while significant in each of the two languages, is also significantly different between the two languages.

Table 5
Analysis of Variance for VERSION BY LANGUAGE
Dependent variable = test score (unweighted means) $N = 142$

Variation	SS	DF	MS	F	Sig. of F
Within Cells	38356.56	134	286.24		
Version	59747.85	3	19915.95	69.58	.000
Language	10.49	1	10.49	0.04	.848
Version by Language	6037.52	3	2012.51	7.03	.000

In fact, the significant interaction effect is entirely due to the different behaviour of X25, which, in terms of difficulty, ranks third in German but second in English; if X25 is excluded from the analysis, the interaction effect disappears. In this case, however, a significant effect of LANGUAGE obtains: scores for C25, C33 and CFL are consistently lower in English than in German (see Table 3).

3.1.4 Comparison of passage difficulty

In order to investigate the effect of the deletion pattern on relative passage difficulty in the two languages, percentage scores for the three passages were computed (see Table 6 and Figure 3). All factors, i.e. VERSION, LANGUAGE, and TEXT were included in a multiple analysis of variance design with the three passage scores as dependent variables. The results involving the factor TEXT are summarised in Table 7. While the single effect of the factor TEXT is not significant, all three interaction effects are.

Table 6
Passage scores for three translationally equivalent passages in German and English

	Text B		Text C		Text D	
	German	English	German	English	German	English
C25	82.0	85.9	90.7	80.8	91.5	78.9
C33	60.5	68.9	72.3	54.4	69.4	60.0
CFL	45.4	65.7	57.1	41.1	56.6	48.0
X25	43.1	68.5	60.4	61.9	59.5	70.7

Figure 3
Passage scores for three translationally equivalent passages
in German and English

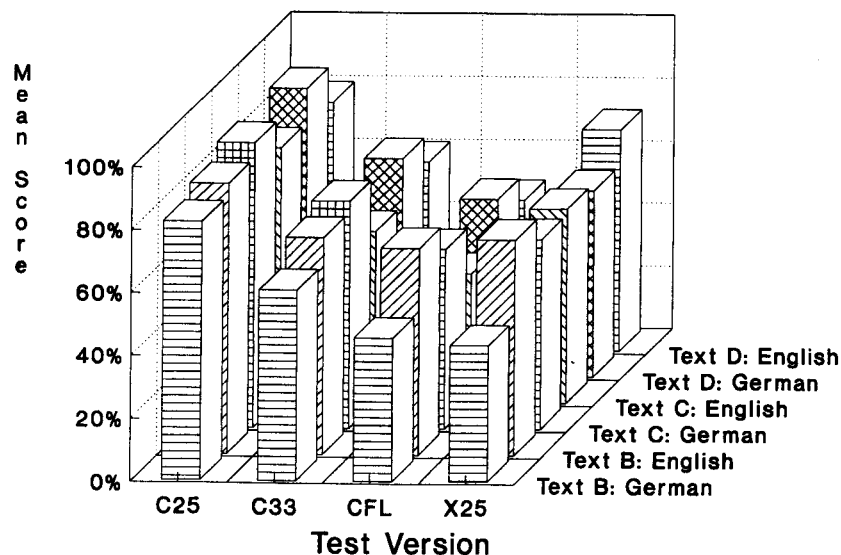


Table 7
Averaged tests of significance for trial factor TEXT
Dependent variables = passage scores for three texts (N = 142)

Variation	SS	DF	MS	F	Sig. of F
Within Cells	24789.03	268	92.50		
Texts	319.76	2	159.88	1.73	.180
Version by Texts	1944.84	6	324.14	3.50	.002
Language by Texts	11874.10	2	5937.05	64.19	.000
Version by Language by Texts	1360.62	6	226.77	2.45	.025

The interaction effect VERSION BY TEXTS indicates that the relative difficulty of the three passages is not the same in the four test versions. Moreover, the differences between the passage scores are significantly different in the two languages, as is reflected in the interaction effect LANGUAGE BY TEXTS. That is, the relative difficulty of the passages is not the same in the two languages.

Finally – and most importantly – while relative passage difficulty is different in the four test versions, these differences are not the same in the two languages either, as indicated by the significance of the three-way interaction VERSION BY LANGUAGE BY TEXTS. Thus, the difference in difficulty between translationally equivalent passages is the result of a complex interplay between text type and deletion pattern.

3.2 Discussion

The reliability of the four test versions in English is satisfactory. In contrast to German, where, presumably due to low variance, C25 fails to produce satisfactory reliability coefficients, C25 in English is more difficult and produces higher variance and hence higher reliability.

For both languages the prediction that redundancy reduction is related to test difficulty is borne out. While the causes of the increase in difficulty brought about by X25 are presumably related to those responsible for the so-called “bathtub effect” (Aitchison, 1987, pp. 119ff.) in both languages, less information is lost in English than in German by applying word-initial rather than word-final deletion. This finding may be due to differences in the morphological structure of the two languages. Blanking out the beginning of words in German, which uses considerably more inflection than English, involves damaging lexical morphemes to a greater extent than in English, where the incidence of inflections is lower.

At this stage, the following conclusions can be drawn:

- If the traditional C-Test is too easy either as a proficiency measure or as a measure of passage difficulty, increasing the extent of redundancy reduction is an appropriate way of increasing test difficulty in both German and English. If an increase by ca. 20 percentage points is needed, C33 may be used. If a further increase is required, CFL seems to be a viable solution.

- For German, X25 increases test difficulty by approximately the same extent as CFL, and, since it has proved more reliable than CFL, it may be considered as an alternative. In English, where it increases difficulty even less than C33, the only reason, if any, for preferring it to C33 is its slightly greater ease of construction, which, however, is offset by lower reliability.
- It is advisable to exercise caution when translating reduced redundancy tests in order to achieve equivalence with regard to difficulty. C25, C33 and CFL seem to become slightly, but significantly, more difficult in English than they are in German. X25 becomes considerably easier in English than it is in German. Moreover, the changes in difficulty brought about by translation may also be dependent on the text type.

Since in addition to the deletion pattern, the text type and the language of the text have been found to be related to test difficulty, further investigation of the causes of C-Test difficulty is clearly required. Research of this kind holds promise as an important contribution to C-Test validation.

References

- Aitchison, Jean. (1987). *Words in the mind*. Oxford: Blackwell.
- Belloc, Gabriel & Nègre, Gaston. (Eds.). (1971). *Expression française. Thèmes et textes* (B.E.P. 2^e année). Delagrave: s.l.
- Cleary, Christopher. (1988). The C-Test in English: left-handed deletions. *RELC Journal*, 19, 26-37.
- Cohen, Andrew D., Segal, Michal & Weiss Bar-Siman-Tov, Ronit. (1984). The C-Test in Hebrew. *Language Testing*, 1, 221-225.
- Grotjahn, Rüdiger. (1987). How to construct and evaluate a C-Test: A discussion of some problems and some statistical analyses. In Rüdiger Grotjahn, Christine Klein-Braley & Douglas K. Stevenson (Eds.), *Taking their measure: The validity and validation of language tests* (pp. 219-253). Bochum: Brockmeyer.
- Haber, Heinz. (1967). *Unser blauer Planet*. Hamburg: Rowohlt.
- Klein-Braley, Christine. (1994). Language testing with the C-Test. A linguistic and statistical investigation into the strategies used by C-Test takers, and the prediction of C-Test difficulty. Unpublished Habilitationsschrift, University of Duisburg.

- Köberl, Johann & Sigott, Günther. (1994). Adjusting C-Test difficulty in German. In Rüdiger Grotjahn (Ed.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Vol. 2, pp. 171-192). Bochum: Brockmeyer.
- Kokkora, Valmar. (1988). Letter deletion procedure: a flexible way of reducing text redundancy. *Language Testing*, 5, 115-119.
- Quirk, Randolph, Greenbaum, Sidney, Leech, Geoffrey & Svartvik, Jan. (1985). *A comprehensive grammar of the English language*. London: Longman.
- Raatz, Ulrich & Klein-Braley, Christine. (1985). How to develop a C-Test. In Christine Klein-Braley & Ulrich Raatz (Eds.), *Fremdsprachen und Hochschule 13/14: Thematischer Teil: C-Tests in der Praxis*, (pp. 20-22). Bochum: AKS.
- Sigott, Günther. (1992). Problems in the development of a proficiency measure with crosslinguistic applicability. *Arbeiten aus Anglistik und Amerikanistik*, 17(2), 151-170.

Appendix: English Test Passages

For German test passages, cf. Köberl & Sigott (1994).

C25

English is generally acknowledged to be the world's most important language. It i_____ perhaps wo_____ glancing bri_____ at t_____ basis f_____ that evalu_____. There a_____, after a_____, thousands o_____ different lang_____ in t_____ world, a_____ each wi_____ seem uniq_____ important t_____ those w_____ speak i_____ as th_____ native lang_____, the lang_____ they acqu_____ at th_____ mother's kn_____.

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 h_____ already be_____ realized - lo_____ before t_____ first sp_____ flights - th_____ the Ea_____ could n_____ possibly lo_____ like a gl_____ although su_____ an im_____ repeatedly sugg_____ itself. I_____ should n_____ be forg_____ that t_____ Earth's atmos_____ is on_____ very rar_____ so compl_____ clear a_____ transparent th_____ an obse_____ from out_____ might reco_____ continents a_____ oceans, isl_____ and se_____ as cle_____ as o_____ a m_____.

Children are curious. Everything th_____ encounter i_____ marvellous a_____ astonishing. Th_____ would li_____ to under_____, and a_____ soon a_____ they c_____ talk th_____ ask ques_____. This bur_____ desire t_____ understand, th_____ appetite f_____ knowledge cont_____ in a some_____ more refle_____ and prof_____ form dur_____ adolescence, wh_____ is ther_____ the nat_____ age f_____ starting unive_____.

C33

English is generally acknowledged to be the world's most important language. It i_____ perhaps w_____ glancing br_____ at t_____ basis f_____ that eva_____. There a_____, after a_____, thousands o_____ different lan_____ in t_____ world, a_____ each w_____ seem un_____ important t_____ those w_____ speak i_____ as t_____ native la_____, the la_____ they ac_____ at t_____ mother's k_____.

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 h_____ already b_____ realized - l_____ before t_____ first s_____ flights - t_____ the E_____ could n_____ possibly l_____ like a g_____ although s_____ an i_____ repeatedly sug_____ itself. I_____ should n_____ be for_____ that t_____ Earth's atm_____ is o_____ very ra_____ so com_____ clear a_____ transparent t_____ an ob_____ from ou_____ might rec_____ continents a_____ oceans, is_____ and s_____ as cl_____ as o_____ a m_____.

Children are curious. Everything t_____ encounter i_____ marvellous a_____ astonishing. T_____ would l_____ to und_____, and a_____ soon a_____ they c_____ talk t_____ ask que_____. This bu_____ desire t_____ understand, t_____ appetite f_____ knowledge con_____ in a so_____ more ref_____ and pr_____ form du_____ adolescence, w_____ is the_____ the na_____ age f_____ starting uni_____.

CFL

English is generally acknowledged to be the world's most important language. It i_____ perhaps w_____ glancing b_____ at t_____ basis f_____ that e_____. There a_____, after a_____, thousands o_____ different l_____ in t_____ world, a_____ each w_____ seem u_____ important t_____ those w_____ speak i_____ as t_____ native l_____, the l_____ they a_____ at t_____ mother's k_____.

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 h_____ already b_____ realized - l_____ before t_____ first s_____ flights - t_____ the E_____ could n_____ possibly l_____ like a g_____ although s_____ an i_____ repeatedly s_____ itself. I_____ should n_____ be f_____ that t_____ Earth's a_____ is o_____ very r_____ so c_____ clear a_____ transparent t_____ an o_____ from o_____ might r_____ continents a_____ oceans, i_____ and s_____ as c_____ as o_____ a m_____.

Children are curious. Everything t_____ encounter i_____ marvellous a_____ astonishing. T_____ would l_____ to u_____, and a_____ soon a_____ they c_____ talk t_____ ask q_____. This b_____ desire t_____ understand, t_____ appetite f_____ knowledge c_____ in a s_____ more r_____ and p_____ form d_____ adolescence, w_____ is t_____ the n_____ age f_____ starting u_____.

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.