

Gerhard Jakschik

Der C-Test für erwachsene Zweitsprachler als Einstufungsinstrument bei der Schul- ausbildung

Das Deutsche Erwachsenen-Bildungswerk (DEB) in Nürnberg bereitet junge Aussiedler innerhalb eines Jahres auf die Hauptschulabschlußprüfung vor. Bei Kursbeginn werden die Teilnehmer in Abhängigkeit von ihrem Sprachstand im Deutschen in Leistungsgruppen eingeteilt. Die Einstufung erwies sich in der Vergangenheit als problematisch: Anfänger mit guten oder sehr geringen Deutschkenntnissen ließen sich zwar identifizieren, eine befriedigende Differenzierung im mittleren Leistungsbereich war aber nicht möglich. Daher wird beim DEB erwogen, den "C-Test für erwachsene Zweitsprachler" für die Einstufung heranzuziehen. Eine Querschnittsuntersuchung am Ende des Schuljahres 91/92 zeigte, daß enge Zusammenhänge zwischen dem C-Test-Ergebnis und den erhobenen Deutschnoten bestehen. Vor allem aber ließ sich eine monotone Beziehung zwischen der Gruppierungsvariablen "C-Test-Gruppe" und den mittleren Deutschnoten nachweisen. Demgegenüber wies die herkömmliche Gruppeneinteilung die bekannten Unschärfen im mittleren Leistungsbereich sowie durchgängig niedrigere Zusammenhänge zu den Deutschnoten auf. In der anschließenden Längsschnittuntersuchung, in der der C-Test am Anfang und am Ende des Schuljahres 92/93 bearbeitet wurde, bestätigte sich seine Überlegenheit als Einstufungsinstrument. Darüber hinaus deutet die hohe Test-Retest-Korrelation darauf hin, daß mit dem C-Test gebildete Rangreihen eine hohe zeitliche Stabilität aufweisen. Die Verwendung des C-Tests bei der Einteilung in sprachliche Leistungsgruppen wird empfohlen.

1. Einleitung

Das Deutsche Erwachsenen-Bildungswerk (DEB) in Nürnberg ist eine Einrichtung der Deutschen Angestellten Gewerkschaft (DAG), an der junge Aussiedler den deutschen Hauptschulabschluß erlangen können. Die Kursteilnehmer werden innerhalb eines Schuljahres auf die Abschlußprüfung vorbereitet, wobei der Unterricht die Fächer Deutsch, Mathematik, Arbeitslehre, Erziehungskunde, Kunsterziehung sowie Werken/Technisches Zeichnen umfaßt. Den Kursteilnehmern, die einen Zensuredurchschnitt von 3.0 oder besser erreichen, wird der sogenannte qualifizierende Hauptschulabschluß zugesprochen (der nur in Bayern und Niedersachsen vergeben wird); die Schüler mit einem Schnitt zwischen 3.1 und 4.9 erlangen den üblichen Hauptschulabschluß.

Ein Problem, das sich den Lehrkräften des DEB zu Beginn eines jeden Schuljahres stellt, ist die Einstufung der Deutschkenntnisse der neuen Kursteilnehmer. Erfahrungsgemäß verfügen die neuen Schüler, die in erster Linie aus der ehemaligen UdSSR, aus Rumänien und aus Polen kommen, über recht heterogene Deutschkenntnisse; es erscheint daher angebracht, in Abhängigkeit vom Sprachstand im Deutschen Leistungsgruppen zu bilden, um so den Unterricht optimal an die individuellen Voraussetzungen der Kursteilnehmer anpassen zu können. Bei der Einteilung der neuen Schüler in Leistungsgruppen zeigte sich in der Vergangenheit, daß zwischen den Extrembereichen der Dimension "Sprachstand im Deutschen" relativ gut differenziert werden kann (d.h. zwischen Schülern mit minimalen oder aber sehr ansprechenden Deutsch-Vorkenntnissen), es aber schwierig ist, zwischen den Ausprägungsgraden im mittleren Bereich zuverlässig zu unterscheiden. Die Unzufriedenheit mit dieser Situation führte zu der Überlegung, ob die alljährliche Einstufungsprozedur¹ nicht durch ein reliableres und valideres Verfahren ersetzt oder doch zumindest um ein solches ergänzt werden sollte. Konkret wurde in Erwägung gezogen, den "C-Test für erwachsene Zweitsprachler" (Jakschik, 1992) einzusetzen. Der "C-Test für erwachsene Zweitsprachler" ist ein aus sechs Lückentexten bestehendes Verfahren, mit dem ein Globalindex für den Sprachstand im Deutschen ermittelt wird.² Zunächst sollte jedoch geprüft werden, inwieweit sich substantielle Zusammenhänge zwischen C-Test-Ergebnissen und den Abschlußnoten in Deutsch beim Deutschen Erwachsenen-Bildungswerk nachweisen lassen.³

¹ Die Einzelheiten des bisherigen Einstufungsverfahrens, bei dem eher qualitative Daten erhoben werden, sind dem Verfasser nicht bekannt.

² Eine Einführung in die C-Test-Methode geben z.B. Raatz & Klein-Braley (1983) und Grotjahn (1992a).

³ Bedanken möchte ich mich bei Herrn Fock, Bundesanstalt für Arbeit, für seine Anregung, den C-Test im oben beschriebenen Rahmen einzusetzen, und besonders bei Frau Fock, Deutsches Erwachsenen-Bildungswerk, für die Vorbereitung und Durchführung der Untersuchung. Mein Dank gilt auch den anderen beteiligten Lehrkräften.

2. Querschnittsuntersuchung

2.1 Beschreibung der Querschnittsuntersuchung

Im Juli 1992, kurz vor Schuljahresende, wurden sechs von insgesamt sieben Klassen des Deutschen Erwachsenen-Bildungswerkes mit dem "C-Test für erwachsene Zweitsprachler" untersucht. Die Klasse 6 wurde aus DEB-internen Gründen nicht in die Erhebung einbezogen.

Insgesamt nahmen $N = 59$ Personen an der Untersuchung teil, darunter 37 Schülerinnen und 22 Schüler (Zahl der Schüler pro Klassenstufe: 1: 11, 2: 12, 3: 7, 4: 12, 5: 9, 7: 8). Das Durchschnittsalter betrug 17.3 Jahre. Im Mittel lebten die Schüler seit knapp zwei Jahren (21.2 Monate) in Deutschland (Standardabweichung: 8.1 Monate). 30 Schüler (= 51%) wurden bereits in ihrem Herkunftsland in Deutsch unterrichtet, und zwar durchschnittlich dreieinhalb Jahre lang (41 Monate; Standardabweichung: 30.5 Monate).

Es wurden neben dem C-Test-Ergebnis folgende Leistungsdaten erhoben:

- die schriftliche Deutschnote vom Schuljahresende
- die mündliche Deutschnote vom Schuljahresende
- die Abschlußnote in Deutsch
- die Halbjahresnote in Deutsch
- die Abschlußnote in Mathematik
- die Abschlußnote in Arbeitslehre
- die Abschlußnote in Erziehungskunde
- die Abschlußnote in Kunsterziehung
- die Abschlußnote in Werken/Technisches Zeichnen.

Beim "C-Test für erwachsene Zweitsprachler" wurde der Gesamtrohwert (maximal 120) sowie die Rohwerte für die einzelnen Texte ermittelt.

Die Auswertung erfolgte mit dem Statistik-Programmsystem SPSS/PC+ (V. 4.01).

2.2 Ergebnisse der Querschnittsuntersuchung

Die mittlere Deutsch-Endnote in der Gesamtstichprobe beträgt 1.95 (Standardabweichung: 1.18). Die entsprechenden Werte für die schriftliche und für die mündliche Deutschnote sind 2.56 (1.36) bzw. 1.69 (1.10). Die Häufigkeitsverteilungen der Deutschzensuren (insbesondere der mündlichen Note)

sind schief, sie haben ihren Gipfel – wie bereits die Mittelwerte ausweisen – im Bereich der guten Noten.

Die Gesamtstichprobe erreichte im C-Test einen Mittelwert von 72.6 (Standardabweichung: 22.7). Dies kann, wenn man es in das bisher vorliegende empirische Material zum "C-Test für erwachsene Zweitsprachler" einordnet, als befriedigendes Resultat gewertet werden.⁴ Die Häufigkeitsverteilung des C-Tests ähnelt sehr stark einer Normalverteilung; ein Modellanpassungstest wurde allerdings nicht berechnet.

Tabelle 1 zeigt die mittleren C-Test-Ergebnisse (Gesamtrohwerte) und Standardabweichungen für die einzelnen Leistungsgruppen:⁵

Tabelle 1
Mittlere C-Test-Ergebnisse der einzelnen Leistungsgruppen

Leistungsgruppe	Zahl der Schüler	mittleres C-Test-Ergebnis (Gesamtrohwert)	Standardabweichung
1	11	98.9	7.2
2	12	76.6	16.6
3	7	69.4	11.7
4	12	66.7	19.7
5	9	70.3	21.7
7	8	44.3	19.5

Unter der Voraussetzung, daß man das C-Test-Ergebnis als zuverlässige und gültige Schätzung des aktuellen Sprachstandes im Deutschen akzeptiert, läßt sich der Tabelle 1 entnehmen, daß die Schüler der Leistungsgruppe 1 erwartungsgemäß das beste Gesamtergebnis erzielten, die Schüler der Leistungsgruppe 7 hingegen das schlechteste; im mittleren Bereich sind die Verhältnisse jedoch weniger klar.

Während die Leistungsgruppe 2 sich noch etwas von Leistungsgruppe 3 abhebt, liegen die Leistungsgruppen 3, 4 und 5 beinahe auf einer Ebene.

⁴ Nach ersten Erhebungen an Volkshochschulen erreichen erfolgreiche Teilnehmer der Deutsch-Mittelstufenprüfung nach den Richtlinien des Goethe-Instituts im "C-Test für erwachsene Zweitsprachler" einen Mittelwert um 100.

⁵ Leistungsgruppe 1 steht für den besten Sprachstand im Deutschen, Leistungsgruppe 7 für den schlechtesten.

Die oben formulierte hypothetische Annahme, daß es einen engen Zusammenhang zwischen C-Test-Ergebnis und der Deutschzensur gibt, findet eine deutliche Bestätigung in den Korrelationskoeffizienten der Tabelle 2⁶.

Tabelle 2
Korrelationen zu den Deutschensuren (N = 55)

Koeffizient	Deutsch mündlich		Deutsch schriftlich		Endnote Deutsch		Halbjahresnote Deutsch	
	r	r _s	r	r _s	r	r _s	r	r _s
C-Test (Gesamtrohwert)	-.64**	-.52**	-.77**	-.73**	-.72**	-.65**	-.65**	-.59**
Leistungsgruppe	.37*	.27*	.60**	.54**	.50**	.44**	.45**	.42**

* $p < .01$ ** $p < .001$ (jeweils einseitige Tests)

Der C-Test korreliert mit den unterschiedlichen Deutschensuren durchgängig mittel bis hoch, während die Leistungsgruppe als Indikator für den Sprachstand nur mittlere Korrelationen aufweist.

Im Sinne der diskriminanten Validität ist es interessant zu sehen, ob Korrelationen zwischen dem C-Test und den anderen – weniger sprachbetonten – Unterrichtsfächern bestehen. Hierüber geben die Koeffizienten in Tabelle 3 Auskunft.

Erwartungsgemäß korrelieren die Variablen C-Test und Leistungsgruppe nicht signifikant mit den Zensuren in Mathematik und in Werken/Techn. Zeichnen. Offensichtlich spielt die Sprache in den Unterrichtsfächern Arbeitslehre und Erziehungskunde aber eine größere Rolle; diese Annahme ist plausibel und würde die signifikanten Zusammenhänge erklären, die hier gefunden werden konnten.

Es kann in Anbetracht der berichteten Korrelationskoeffizienten auch nicht weiter überraschen, daß die Verteilungen der mittleren Deutschensuren der einzelnen Leistungsgruppen weitgehend der oben dargestellten

⁶ Da man nicht unbedingt Intervallskalenniveau, auf jeden Fall aber ordinale Zusammenhänge unterstellen kann, wurde neben dem Produkt-Moment-Korrelationskoeffizienten r auch die Rangkorrelation r_s nach Spearman berechnet. Die negativen Korrelationen beim C-Test erklären sich dadurch, daß er anders "gepolt" ist als Schulnoten, d.h. hohe Werte kennzeichnen hohe Merkmalsausprägungen.

Tabelle 3
Korrelationen zu den anderen Unterrichtsfächern ($N = 55$)

Koeffizient	Endnote Mathematik		Endnote Arbeitslehre		Endnote Erziehungskunde		Endnote Werken/T.Z.	
	r	r_s	r	r_s	r	r_s	r	r_s
C-Test (Gesamtrohwert)	-.26	-.24	-.37*	-.33*	-.40*	-.24	-.06	-.03
Leistungsgruppe	.14	.12	.22	.25	.43**	.30*	.01	.00

* $p < .01$ ** $p < .001$ (jeweils einseitige Tests)

Verteilung der C-Test-Mittelwerte entsprechen (sie sind lediglich umgekehrt "gepolt").

Abbildung 1 zeigt die mittleren Deutschensuren (Endnote) für die einzelnen Leistungsgruppen. Die Abbildung 2 zeigt eine analoge Verteilung für die Deutsch-Halbjahresnoten. Aus Platzgründen wird auf eine grafische Darstellung der Verteilungen für die schriftlichen und die mündlichen Deutschnoten verzichtet; sie unterscheiden sich nur sehr gering von den Verteilungen der Deutsch-Endnoten und der Deutsch-Halbjahresnoten.

Abbildung 1
Mittlere Deutsch-Endnote
in den Leistungsgruppen ($N = 55$)

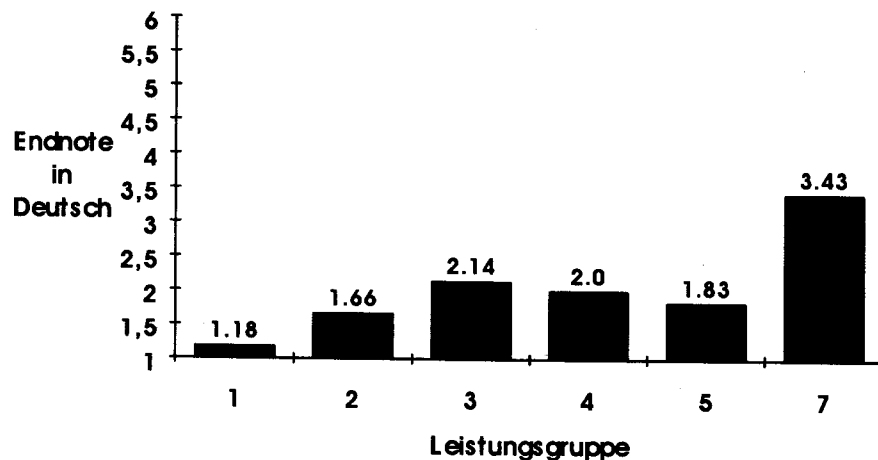
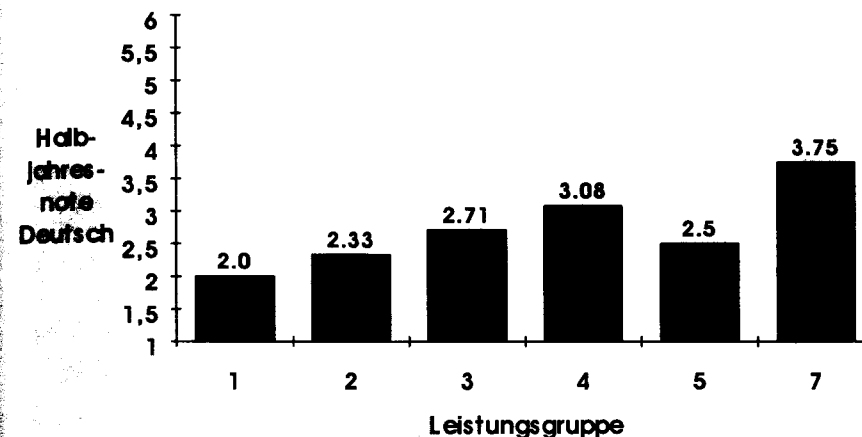


Abbildung 2
Mittlere Deutsch-Halbjahresnote
in den Leistungsgruppen ($N = 55$)



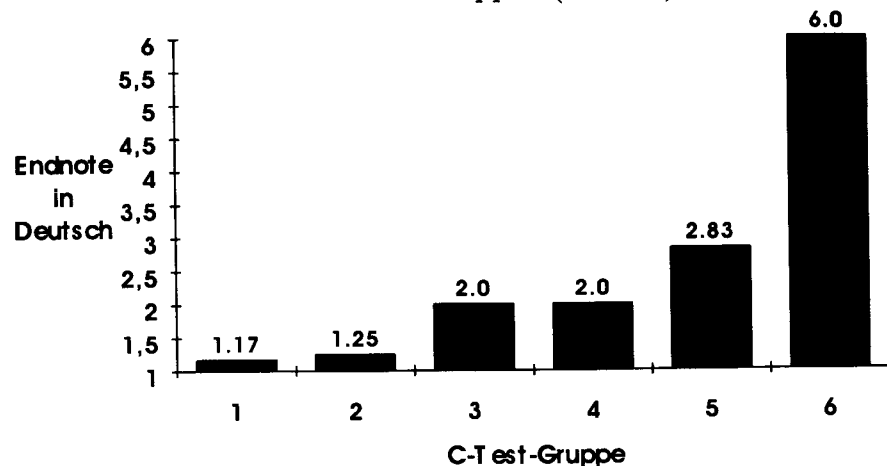
Es fällt – wie bei den C-Test-Mittelwerten – auf, daß die beste und die schlechteste Gruppe sich leicht identifizieren lassen; hingegen ist bei den Gruppen zwei bis fünf eine klare Differenzierung des sprachlichen Leistungsstandes kaum möglich.

Da erwogen wird, den "C-Test für erwachsene Zweitsprachler" zu einem späteren Zeitpunkt als Einstufungsverfahren bei der Bildung von Leistungsgruppen zu verwenden, wurde geprüft, wie sich die unterschiedlichen Deutschensuren verteilen, wenn man das C-Test-Resultat als Gruppierungskriterium verwendet. Es wurden die folgenden Gruppen gebildet:

- C-Test-Gruppe 1: Gesamtrohwert > 95
- C-Test-Gruppe 2: Gesamtrohwert > 80 und ≤ 95
- C-Test-Gruppe 3: Gesamtrohwert > 65 und ≤ 80
- C-Test-Gruppe 4: Gesamtrohwert > 50 und ≤ 65
- C-Test-Gruppe 5: Gesamtrohwert > 35 und ≤ 50
- C-Test-Gruppe 6: Gesamtrohwert ≤ 35

Die Abbildungen 3 und 4 zeigen die Verteilungen der Deutsch-Endnoten sowie der Deutsch-Halbjahresnoten auf der Grundlage der obigen Gruppeneinteilung nach dem C-Test-Gesamtrahwert. Auch hier gilt, daß die nicht dargestellten Verteilungen für die schriftlichen und die mündlichen Deutschnoten sich nur unwesentlich unterscheiden.

Abbildung 3
Mittlere Deutschendnote
in den C-Test-Gruppen (N = 55)



Die Abbildungen 3 und 4 veranschaulichen, daß zwischen der Gruppierungsvariablen "C-Test-Gruppe" und den Deutschzensuren eine monotone Beziehung besteht. Da die Gruppen Nr. 3 und 4 identische Mittelwerte bei drei von vier Zensuren in Deutsch aufweisen, kann allerdings keine streng monotone Beziehung konstatiert werden.

Tabelle 4 gibt die Interkorrelationen der sechs Texte, aus denen der "C-Test für erwachsene Zweitsprachler" besteht, wieder. Die fast durchweg hohen Korrelationskoeffizienten zeigen, daß der vorliegende C-Test sehr homogen ist.

Abschließend wurden die verschiedenen Zensuren, das C-Test-Ergebnis sowie die Variable Leistungsgruppe einer Faktorenanalyse unterzogen (Hauptachsenfaktorenanalyse bzw. Principal Axis Factoring (PAF) mit an-

Abbildung 4
Mittlere Deutsch-Halbjahresnote
in den C-Test-Gruppen (N = 55)

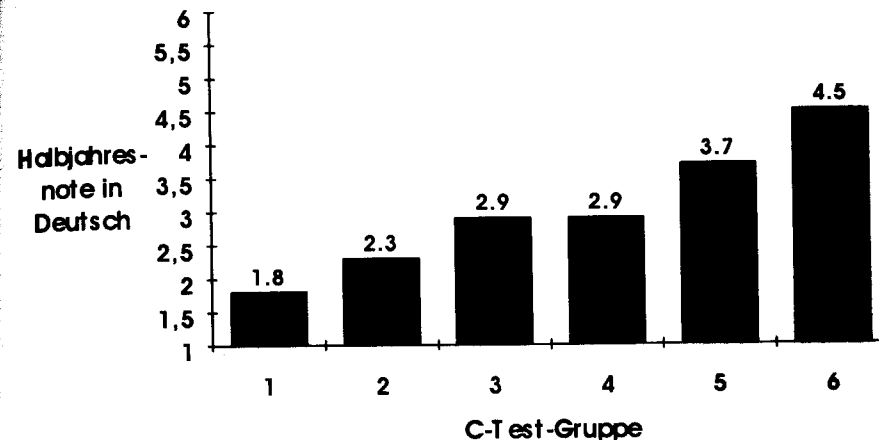


Tabelle 4
Interkorrelationen der C-Test-Texte
(Produkt-Moment-Korrelationen, N = 59)

	Text 1	Text 2	Text 3	Text 4	Text 5
Text 1					
Text 2	.76**				
Text 3	.68**	.69**			
Text 4	.66**	.72**	.76**		
Text 5	.59**	.64**	.83**	.74**	
Text 6	.70**	.77**	.80**	.78**	.71**

** $p < .001$ (einseitiger Test)

schließender Varimax-Rotation).⁷ Es wurden zwei Faktoren extrahiert, von denen der erste 53 Prozent der Varianz erklärt, der zweite 16 Prozent. Tabelle 5 gibt Aufschluß über die Faktorladungen der Variablen nach der Rotation.

⁷ Andere Extraktionsmethoden, die zu Vergleichszwecken angewandt wurden, erbrachten sehr ähnliche Resultate.

Tabelle 5
Faktorladungsmatrix

Variablen	Faktor 1	Faktor 2
C-Test-Gesamtergebnis	-.871	-.057
Schriftliche Deutschzensur	.860	.323
Mündliche Deutschzensur	.712	.438
Deutsch-Endnote	.838	.428
Deutsch-Halbjahresnote	.631	.464
Mathematikzensur	.216	.678
Arbeitslehrezensur	.383	.734
Erziehungskundezensur	.584	.281
Kunsterziehungzensur	.163	.636
Zensur in Werken/Techn. Zeichnen	-.005	.677
Leistungsgruppe	.699	-.055

Der erste Faktor läßt sich unschwer als allgemeiner Sprachfaktor interpretieren, für den der C-Test die höchste Ladung aufweist. Der zweite Faktor beinhaltet offensichtlich sowohl mathematisch-analytische als auch praktisch-anschauliche Anteile.

2.3 Diskussion der Ergebnisse der Querschnittsuntersuchung

Es konnte nachgewiesen werden, daß zwischen dem C-Test-Ergebnis und den verschiedenen Deutschnoten ein enger Zusammenhang besteht. Dies läßt den "C-Test für erwachsene Zweitsprachler" für den Einsatz als globales Einstufungsinstrument geeignet erscheinen – auch im Rahmen der Erfolgskontrolle. Signifikante Korrelationen bestehen auch zwischen der am Schuljahresanfang vorgenommenen Einteilung in Leistungsgruppen und den Deutschzensuren, jedoch in geringerer Höhe als beim C-Test.

Die mittleren Deutschnoten verteilen sich monoton, wenn man eine auf dem C-Test beruhende Gruppeneinteilung zugrundelegt. Bei der herkömmlichen Leistungsgruppeneinteilung zeigte sich hingegen, daß nur die beste und die schlechteste Gruppe klar identifiziert werden konnten; im mittleren Bereich war eine Differenzierung kaum möglich. Selbstverständlich kann mit den vorliegenden Daten nicht zweifelsfrei nachgewiesen werden, daß eine mit dem C-Test vorgenommene Leistungsgruppeneinteilung am Schuljahresanfang ebenfalls überlegen gewesen wäre: Es kann argumentiert

werden, daß die traditionelle Zuweisung zu Leistungsgruppen "eigentlich die richtige" war, die Schüler jedoch in unterschiedlichem Umfang vom Unterricht profitierten, so daß es im Verlauf des Schuljahres zu Verschiebungen der Rangreihe kam. Auf der anderen Seite: Ist es plausibel anzunehmen, daß die Gruppe 2 im Durchschnitt fast gar nicht, die Gruppe 5 jedoch sehr stark vom Unterricht profitierte (vgl. Abbildung 1) ?

Es gibt ein Indiz dafür, daß der C-Test tatsächlich als Einstufungsinstrument überlegen ist. Die Halbjahresnote in Deutsch weist nämlich etwa den gleichen zeitlichen Abstand zur herkömmlichen Leistungsgruppeneinteilung wie zur Untersuchung mit dem C-Test auf; trotzdem gelten die oben beschriebenen Zusammenhänge uneingeschränkt auch für diese Zensur (höhere Korrelation mit dem C-Test, monotone Verteilung der mittleren Noten auf der Grundlage von C-Test-Gruppen; vgl. Tabelle 2 sowie Abbildung 4).

Die nach der Querschnittserhebung noch bestehenden leichten Unsicherheiten hinsichtlich der Brauchbarkeit des C-Tests als Einstufungsinstrument sollten in der nächsten Phase der Studie ausgeräumt werden: Die anschließende Längsschnittuntersuchung sollte eine eindeutige Vergleichsrechnung hinsichtlich der unterschiedlichen Einstufungsmethoden erlauben.

3. Längsschnittuntersuchung

3.1 Beschreibung der Längsschnittuntersuchung

Die neuen Schüler des Schuljahres 1992/1993 wurden nach der herkömmlichen Methode in insgesamt fünf Leistungsgruppen eingeteilt. Ein bis zwei Wochen nach Unterrichtsbeginn, im September 1992, wurde den Schülern der "C-Test für erwachsene Zweitsprachler" vorgelegt. Das Testmaterial wurde vom Verfasser ausgewertet; die Ergebnisse waren den Lehrern nicht bekannt und wirkten sich nicht auf die Einteilung der Leistungsgruppen aus. Etwa vier Wochen nach Schuljahresbeginn wurden einige Umgruppierungen vorgenommen, was im wesentlichen darauf zurückzuführen war, daß sich die Gesamtzahl der Schüler durch einige Nachzügler erhöht hatte. Da die Umgruppierungen selbstverständlich nicht unabhängig waren von dem Eindruck, den die Lehrer im Unterricht von ihren Schülern gewannen, wurden sie in der vorliegenden Untersuchung nicht berücksichtigt; es wurden ausschließlich die ursprünglichen Gruppenzuweisungen berücksichtigt, um eine Konfundierung von unterschiedlichen Informationsquellen (herkömmliche Sprachstandsprüfung und Unterrichtserfahrung) zu vermeiden. Am

Schuljahresende, im Juli 1993, wurde der genannte C-Test erneut von den Schülern bearbeitet.

Insgesamt nahmen $N = 93$ Personen an der Untersuchung teil, darunter 38 Schülerinnen und 55 Schüler. Allerdings ist die "Kerngruppe" der Personen, bei denen der C-Test zweimal appliziert werden konnte, etwas kleiner, nämlich $N = 82$; bei den übrigen 11 Personen handelt es sich um Nachzügler, die z.T. erst Monate später den Unterricht beim DEB aufnahmen und bei denen deswegen auf die Ersttestung mit dem C-Test verzichtet wurde. Die wesentlichen Aussagen der Längsschnittuntersuchung beziehen sich naturgemäß auf die 82 Personen, bei denen eine Ersttestung mit dem C-Test möglich war. Die einzelnen Leistungsgruppen umfaßten die folgenden Schülerzahlen (einschließlich der Nachzügler): 1: 23, 2: 17, 3: 16, 4: 20, 5: 17. Das Durchschnittsalter betrug 16.7 Jahre. Zum Zeitpunkt der Ersttestung lebten die Schüler im Mittel seit einem Jahr (12.2 Monate) in Deutschland (Standardabweichung: 10.2 Monate). 42 Schüler (=45%) wurden bereits in ihrem Herkunftsland in Deutsch unterrichtet, durchschnittlich etwa vier Jahre lang (50 Monate; Standardabweichung: 20 Monate).

Es wurden dieselben Leistungsdaten wie bereits in der Querschnittsuntersuchung erhoben (siehe Abschnitt 2.1).

Von den 82 Schülern der "Kerngruppe" wurden 15 wegen schlechter Zensuren nicht zur Prüfung angemeldet; die Abschlußnoten dieser Personen wurden nicht bekanntgegeben. Da auf diese Weise gleichsam das untere Ende der Verteilung "abgeschnitten" und die Varianz erheblich eingeschränkt wurde, hätten ohne korrigierende Maßnahmen nur künstlich verminderte Zusammenhangsmaße berechnet werden können. Es erschien deswegen angemessen, bei den "Prüfungs-Aussteigern" die fehlende Deutsch-Endnote durch die Deutsch-Halbjahresnote zu ersetzen (die Subgruppe derjenigen, die nicht zur Prüfung angemeldet wurden, erzielte bei der Deutsch-Halbjahresnote einen mittleren Wert von 4.5, der Median liegt bei 5.0).

Die Auswertung erfolgte mit dem Statistik-Programmsystem SPSS für Windows (Version 5.02).

3.2 Ergebnisse der Längsschnittuntersuchung

Die mittlere Deutsch-Endnote in der Gesamtstichprobe beträgt 2.80 (Standardabweichung: 1.18)⁸. Die entsprechenden Werte für die schriftliche und

⁸ Bei dieser und bei allen folgenden Berechnungen, in denen die Deutsch-Endnote

für die mündliche Deutschnote sind 3.50 (1.35) bzw. 2.08 (0.86). Es fällt auf, daß die Schüler – zumindest im sprachlichen Bereich – schwächer einzuschätzen sind als die Vorjahresgruppe. Dies spiegelt sich auch in den mittleren C-Test-Ergebnissen wider: Die Gesamtstichprobe erzielte bei der Ersttestung einen Mittelwert von 42.3 (Standardabweichung: 25.1), bei der Zweittestung einen von 67.6 (Standardabweichung: 23.6); durchschnittlich waren die Schüler also im C-Test fünf Punkte schlechter als der vorausgegangene Jahrgang. Die beiden Häufigkeitsverteilungen für den C-Test ähneln, wie bei der Querschnitterhebung, sehr stark einer Normalverteilung; demgegenüber verteilen sich die Deutschzensuren tendentiell schief, allerdings weniger augenfällig als bei der Querschnittsuntersuchung.

Tabelle 6 zeigt die mittleren C-Test-Ergebnisse (Gesamtrohwerte) für die einzelnen Leistungsgruppen.

Tabelle 6
Mittlere C-Test-Ergebnisse der einzelnen Leistungsgruppen

Leistungsgruppe	Zahl der Schüler 1. Test/ 2. Test	mittleres C-Test-Ergebnis (Gesamtrohwert) Ersttestung	Standardabweichung	mittleres C-Test-Ergebnis (Gesamtrohwert) Zweittestung	Standardabweichung
1	19/21	71.5	13.7	90.5	13.7
2	16/14	51.4	12.7	69.1	17.1
3	14/13	39.6	9.7	64.5	17.6
4	16/18	20.1	15.1	51.8	21.3
5	17/15	24.0	24.2	55.7	23.2

Es fällt auf, daß zwar Unschärfen existieren (die Gruppen vier und fünf unterscheiden sich kaum im C-Test-Ergebnis), insgesamt scheint die Differenzierung aber etwas besser gelungen zu sein als bei der Querschnittsuntersuchung.

eine Rolle spielt, wurde die oben angegebene Korrektur durchgeführt. Bei der schriftlichen und der mündlichen Deutschnote wurden fehlende Werte nicht ersetzt.

Tabelle 7 gibt eine Übersicht über die Korrelationen der C-Test-Ergebnisse und der Leistungsgruppeneinteilung mit den verschiedenen Deutschnoten.⁹ Für den C-Test finden sich fast durchgängig die höheren Zusammenhänge. Lediglich die mündliche Deutschnote bildet eine Ausnahme; es kann gemutmaßt werden, daß sie hauptsächlich als "Noten-Korrektiv" eingesetzt wird (gute Zensuren überwiegen deutlich) und bei ihrer Findung weniger strenge Kriterien zum Tragen kommen.

Tabelle 7
Korrelationen zu den Deutschzensuren

Koeffizient	Deutsch mündlich		Deutsch schriftlich		Deutsch Endnote		Deutsch Halbjahresnote	
	r	r_s	r	r_s	r	r_s	r	r_s
C-Test (Gesamtrohwert)	-.31*	-.34**	-.58**	-.55**	-.63**	-.62**	-.65**	-.62**
Ersttestung	(N=63)	(N=63)	(N=64)	(N=64)	(N=80)	(N=80)	(N=80)	(N=80)
C-Test (Gesamtrohwert)	-.52**	-.52**	-.78**	-.74**	-.80**	-.79**	-.81**	-.80**
Zweittestung	(N=58)	(N=58)	(N=59)	(N=59)	(N=70)	(N=70)	(N=70)	(N=70)
Leistungsgruppe	.37**	.39**	.32**	.36**	.53**	.53**	.36**	.37**

* $p < .01$ ** $p < .001$ (jeweils einseitige Tests)

Auf eine Darstellung der Korrelationen zwischen dem C-Test und den Fächern Mathematik, Arbeitslehre, Erziehungskunde sowie Werken/Technisches Zeichnen wird hier verzichtet; sie unterscheiden sich in Höhe und Richtung nicht wesentlich von den in Tabelle 3 aufgelisteten Koeffizienten.

Auf Grundlage der Ersttestung mit dem "C-Test für erwachsene Zweitsprachler" wurden die in Tabelle 8 aufgeführten C-Test-Gruppen gebildet.

Die Verteilungen der mittleren Deutsch-Halbjahresnoten sowie der mittleren Deutsch-Endnoten für die Leistungsgruppen bzw. für die C-Test-Gruppen können den Abbildungen 5 bis 8 entnommen werden.

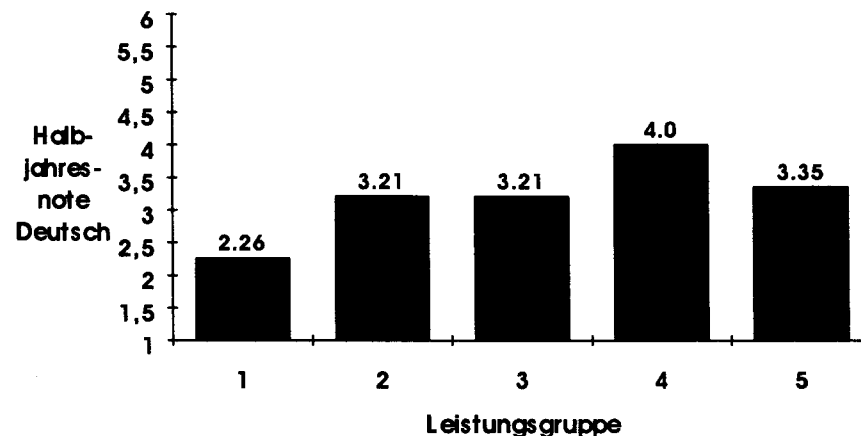
Der Abbildung 5 ist zu entnehmen, daß die Deutsch-Halbjahresnoten bei den Leistungsgruppen zwei, drei und fünf beinahe auf einer Ebene liegen. Im Grunde hebt sich nur die erste Gruppe deutlich von den anderen ab, die

⁹ Es wurde neben dem Produkt-Moment-Korrelationskoeffizienten r die Rangkorrelation r_s nach Spearman berechnet.

Tabelle 8
C-Test-Gruppen

C-Test-Gruppe	Rohwert-Intervall Gesamtrohwert C-Test Ersttestung	Prozentrang-Intervall
1	91-66	100 -81.7
2	65-47	81.6-61.0
3	46-37	60.9-42.7
4	36-18	42.6-22.0
5	17-0	21.9- 1.2

Abbildung 5
Mittlere Deutsch-Halbjahresnote
in den Leistungsgruppen (N = 80)



Gruppen zwei bis fünf bilden einen großen, kaum differenzierten "Rest". Etwas günstiger ist das Bild bei der Deutsch-Endnote insofern, als hier ein kontinuierlicher Anstieg der mittleren Noten bei den Leistungsgruppen eins bis vier erkennbar ist; allerdings sind die Unterschiede zwischen den Gruppen drei, vier und fünf nur gering (Abbildung 6).

Abbildung 6
Mittlere Deutsch-Endnote
in den Leistungsgruppen ($N = 80$)

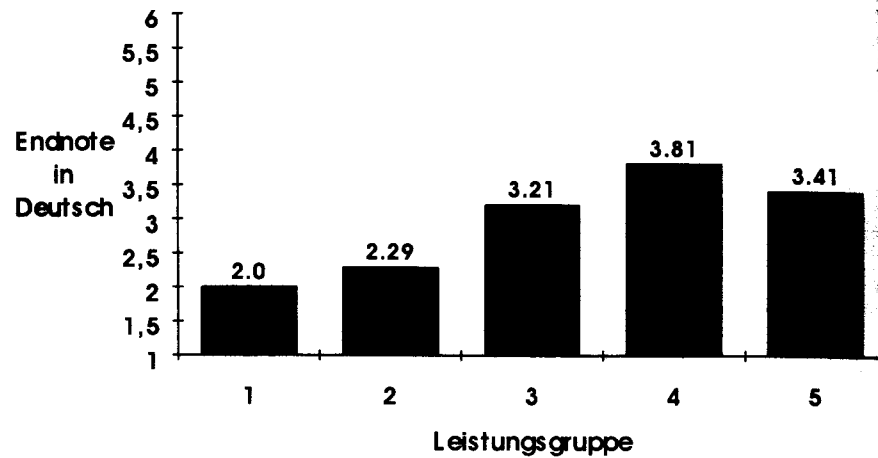


Abbildung 7
Mittlere Deutsch-Halbjahresnote
in den C-Test-Gruppen ($N = 80$)

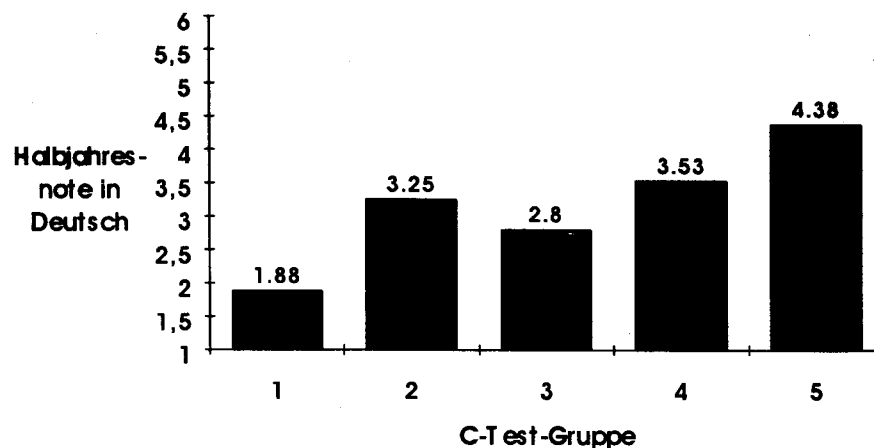
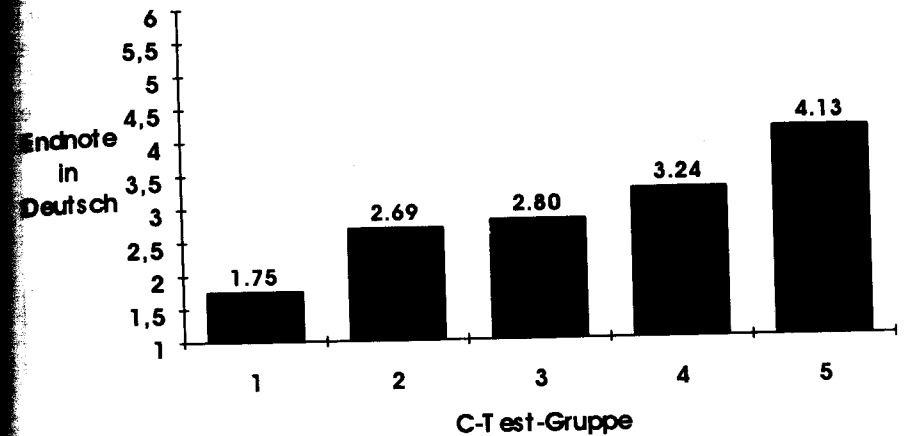


Abbildung 8
Mittlere Deutsch-Endnote
in den C-Test-Gruppen ($N = 80$)



Die Verteilung der mittleren Deutsch-Halbjahresnoten in den C-Test-Gruppen imponiert durch klar abgesetzte Extremgruppen; kritisch ist hier anzumerken, daß eine Differenzierung im mittleren Leistungsbereich nicht gelingt (Abbildung 7). Am überzeugendsten ist die Verteilung der mittleren Deutsch-Endnoten in den C-Test-Gruppen (Abbildung 8): Diese Häufigkeitsverteilung steigt streng monoton an, es wird eine in den Extrembereichen gute, ansonsten noch akzeptable Differenzierung der Gruppen erreicht.

Da C-Tests für ihre hohe Reliabilität bekannt sind, wurden Maße zur inneren Konsistenz mehr aus Gründen der Vollständigkeit berechnet. Der Alpha-Koeffizient nach Cronbach beträgt .92 für die Ersttestung, .94 für die Zweittestung. Die Test-Retest-Korrelation r_{tt} beträgt .82 ($N = 70$, $p < .001$, einseitiger Test), für den entsprechenden Rangkorrelationskoeffizienten r_s wurde ein Wert von .80 ermittelt ($N = 70$, $p < .001$, einseitiger Test). Diese Korrelationskoeffizienten sind verblüffend hoch, wenn man bedenkt, daß zwischen Erst- und Zweitmessung annähernd ein Jahr mit intensiver sprachlicher Förderung lag; sie deuten darauf hin, daß mit dem C-Test gebildete Rangreihen eine beträchtliche Stabilität aufweisen.

Der mittlere sprachliche Lernzuwachs in den C-Test-Gruppen, operationalisiert über die Differenz der beiden C-Test-Messungen (Gesamtroh-

wert Ersttestung minus Gesamtrahwert Zweittestung), kann der Tabelle 9 entnommen werden. Es zeigt sich, daß der Zuwachs in den anfänglich schwachen Gruppen am größten, bei den fortgeschrittenen Lernern am geringsten ist – zumindest rein rechnerisch betrachtet. Eine andere Frage ist, ob den Mittelwertsdifferenzen je nach Gruppenzugehörigkeit eine andere Qualität zukommt. Daß es sich übrigens um mehr als nur einen "Deckeneffekt" handelt, zeigt die Tatsache, daß die Gruppe 1 bei der ersten Vorgabe des C-Tests einen Mittelwert von 79.6 erreichte; theoretisch wäre also eine mittlere Steigerung von 40 Punkten möglich gewesen.

Tabelle 9
Mittlerer sprachlicher Lernzuwachs in den C-Test-Gruppen

C-Test-Gruppe	Zuwachs im C-Test	Standardabweichung
1	16.6	8.7
2	15.4	11.8
3	23.3	15.9
4	28.1	19.9
5	33.4	10.6

Um die faktorielle Stabilität zu prüfen, wurden abschließend alle Leistungsdaten sowie die Variable Leistungsgruppe einer Faktorenanalyse unterzogen (es wurde wiederum eine Hauptachsenfaktorenanalyse mit anschließender Varimax-Rotation durchgeführt). Es wurden zwei Faktoren extrahiert, von denen der erste 50 Prozent der Varianz erklärt, der zweite 17 Prozent. In Tabelle 10 werden die Faktorladungen der Variablen nach der Rotation aufgelistet. Der zweite Faktor läßt sich wegen der vielen mittleren bis hohen Ladungen in recht unterschiedlichen Variablen nicht zwanglos etikettieren. Der erste Faktor ist jedoch, wie in der Querschnittsuntersuchung, klar als sprachlicher Faktor zu identifizieren; wiederum weisen die C-Test-Gesamtrahwerte die höchsten Ladungen im ersten Faktor auf. Insgesamt kann festgestellt werden, daß die hier gefundene faktorielle Struktur weitgehend der in der Querschnittsuntersuchung erhobenen entspricht.

3.3 Abschließende Diskussion

Auch in der Folgeuntersuchung erwies sich der "C-Test für erwachsene Zweitsprachler" als überlegenes Instrument bei der Bildung von sprachlichen Leistungsgruppen. Zwar ist die Überlegenheit nicht so ausgeprägt wie

Tabelle 10
Faktorladungsmatrix

Variablen	Faktor 1	Faktor 2
C-Test-Gesamtergebnis – Ersttestung	-.831	-.028
C-Test-Gesamtergebnis – Zweittestung	-.803	-.414
Schriftliche Deutschzensur	.732	.535
Mündliche Deutschzensur	.580	.279
Deutsch-Endnote	.747	.451
Deutsch-Halbjahresnote	.628	.568
Mathematikzensur	.429	.585
Arbeitslehrezensur	.096	.826
Erziehungskundenzensur	.319	.670
Kunsterziehungszensur	.077	.611
Zensur in Werken/Techn. Zeichnen	.042	.643
Leistungsgruppe	.714	-.292

in der Querschnittsuntersuchung, wo die Bildung von C-Test-Gruppen ja auch auf Daten zum aktuellen sprachlichen Leistungsstand der Schüler beruhte, dennoch ist sie deutlich und zeigt sich übereinstimmend sowohl in Zusammenhangsmaßen zwischen C-Test und Deutschnoten als auch in der Verteilung der mittleren Deutsch-Endnoten, die nur in den C-Test-Gruppen einen monotonen Verlauf aufweisen. Der (zusätzliche) Einsatz des "C-Tests für erwachsene Zweitsprachler" bei der Bildung von sprachlichen Leistungsgruppen kann daher uneingeschränkt empfohlen werden. In beiden Untersuchungen, die beim Deutschen Erwachsenen-Bildungswerk durchgeführt wurden, konnte belegt werden, daß der C-Test – bei bekannt hoher Objektivität und Reliabilität – auch valide ist im Blick auf das Ziel, homogene Gruppen hinsichtlich des Merkmals "Sprachstand im Deutschen" zu bilden. Die gemessen am langen Zeitraum und an der dazwischenliegenden sprachlichen Förderung erstaunlich hohe Test-Retest-Korrelation zeigt darüber hinaus, daß mit dem C-Test vorgenommene Differenzierungen des sprachlichen Leistungsstandes durch eine hohe zeitliche Stabilität gekennzeichnet sind. Schließlich weisen die faktorenanalytischen Berechnungen den "C-Test für erwachsene Zweitsprachler" als ein Verfahren aus, das besonders geeignet erscheint, allgemeine Aspekte der Sprachbeherrschung zu erfassen.

Literaturverzeichnis

- Grotjahn, Rüdiger. (1992a). Der C-Test. Einleitende Bemerkungen. In Grotjahn (1992b), 1-18.
- Grotjahn, Rüdiger. (1992b). *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Bd. 1). Bochum: Brockmeyer.
- Jakschik, Gerhard. (1992). Zum Einsatz des C-Tests in den Psychologischen Diensten der Arbeitsämter. Ein C-Test für Deutsch als Zweitsprache. In Grotjahn (1992b), 297-311.
- Raatz, Ulrich & Klein-Braley, Christine. (1983). Ein neuer Ansatz zur Messung der Sprachleistung. Der C-Test: Theorie und Praxis. In Ralf Horn, Karlheinz Ingenkamp & Reinhold Jäger (Hrsg.), *Tests und Trends 1983, Jahrbuch der Pädagogischen Diagnostik* (S. 107-138). Weinheim: Beltz.

Grotjahn, Rüdiger. (Hrsg.). (1994). *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Bd. 2, S. 279-304). Bochum: Brockmeyer

Ulrich Germann und Rüdiger Grotjahn

Das Lösen von C-Tests auf dem Computer: Eine Pilotuntersuchung zu den Bearbeitungsprozessen

In der vorliegenden Arbeit wird anhand einer Pilotuntersuchung mit acht fortgeschrittenen Lernern des Englischen gezeigt, wie über die Protokollierung der Tastatureingaben mit Hilfe des Software-Packets CALIS auf ökonomische Weise Einblick in das Verhalten der Lerner bei der Bearbeitung von C-Tests auf dem Computer gewonnen werden kann. Es werden drei typische Bearbeitungsstrategien unterschieden und zu den C-Test-Leistungen der Probanden in Beziehung gesetzt. Es wird ferner gezeigt, daß die Computerprotokolle gewisse Einblicke in die dem beobachtbaren Lösungsverhalten zugrundeliegenden mentalen Prozesse und Wissensbestände ermöglichen.

1. Einleitung

Bei Untersuchungen zu psycholinguistischen Aspekten der Bearbeitung von C-Tests sind bisher vor allem introspektive Verfahren wie Lautes Denken und retrospektives Interview verwendet worden (vgl. z.B. Feldmann, Grotjahn & Stemmer, 1986; Feldmann & Stemmer, 1987; Grotjahn, 1986, 1987b; Stemmer, 1991, 1992). Bei diesen Untersuchungen wurden die Versuchspersonen (Vpn) u.a. aufgefordert, bei der Bearbeitung des Tests "laut zu denken". Ihre Äußerungen wurden auf Tonband aufgenommen.

Bei der Erhebung introspektiver Daten ergeben sich allerdings z.B. folgende Probleme:

1. Daten lassen sich nur gewinnen, wenn die mentalen Prozesse verbalisiert werden. Dazu müssen sie bis zu einem gewissen Grad bewußt ablaufen. Feldmann & Stemmer (1987) berichten in diesem Zusammenhang von langen Verbalisierungspausen, die weder vom Versuchsleiter noch von der Vp erklärt werden konnten. Die Autoren kommen zu dem Schluß: "In short, some problem-solving behaviour took place but we could not get at it" (Feldmann & Stemmer, 1987, S. 263).
2. Es kann aufgrund der Verbalisierung zu einer Veränderung des Bearbeitungsverhaltens kommen.