Anhang

Sie haben genau 30 Minuten Zeit. Füllen Sie bitte die Lücken orthographisch richtig aus. (Ein Strich kann ein oder mehrere Buchstaben bedeuten. Schrägstrich bedeutet Wechsel der Dialogpartner.)

La jeune fille dans le train./ Tu as v____, la blo____, là, pr____ de l____ fenêtre? O____, elle e____ "chouette"! On l____ parle?/ Attends, qu'e____ qu'elle fa____? Elle e____ mannequin?/ N____, elle e____ actrice!/ El____ habite e____ banlieue?/ Bien oui, el____ travaille____ Paris. O____ va vo____!/ Mademoiselle, v____ avez d____ beaux yeux, vous savez.

Pardon Monsieur, je cherche la gare. C'est l____?/ Non, c____ n'est p____ loin. Pre____ cette r____-là, dev____ vous, c'____ tout dr____ et pu____ à gauche./ C'____ bien l____ gare po____ Marseille?/ Ah n____, pour Marseille, c'____ la g____ de Lyon. I____ faut pre____ le mé____. C'est as____ loin. Al____ à la station Saint-Lazare.

Deux touristes, un jeune homme et une jeune fille, entrent chez Frédéric. Ils vo____ au b____ et rega____ le tar____ des consomm____. Il y a d____ petits déje____ à 4 F 50./ T____ en ve____ un? Oui, j____ veux bi____. J'ai fa____./ Ils vo____ à u____ table. L____ garçon arr____. Qu'est-ce q____ vous pre____, messieurs da____?/ Deux pet____ déjeuners, s'il vous plait.

J'ai enfin une correspondante allemande. Aujour____, je su____ sort____ avec el____. D'abord, no____ som____ allées s____ les gra____ boulevards. No____ avons reg____ les vitri____ des gra____ magasins, pu____ nous som____ entrées a____ Printemps e____ nous av____ acheté d____ souvenirs. A mi____, nous av____ mangé un sandwich à la terrasse d'un café.

Grotjahn, Rüdiger. (Hrsg.). (1992). Der C-Test. Theoretische Grundlagen und praktische Anwendungen (Bd. 1, S. 263–296). Bochum: Brockmeyer

Rüdiger Grotjahn, Christine Klein-Braley und Ulrich Raatz

C-Tests in der praktischen Anwendung. Erfahrungen beim Bundeswettbewerb Fremdsprachen

Dieser Beitrag beschreibt den Einsatz des C-Tests im Rahmen des Einzelwettbewerbs der Sekundarstufe I des Bundeswettbewerbs Fremdsprachen – einem nationalen Schülerwettbewerb im Bereich Fremdsprachen. Der C-Test wird dort zusammen mit anderen Tests (semi-kreative Aufgabe, Hörverständnistest und Mündliche Produktion) eingesetzt, wobei er zusätzlich die Aufgabe eines Screening-Tests erfüllt.

Im vorliegenden Beitrag werden u.a. folgende Problembereiche diskutiert: das Prinzip der reduzierten Redundanz als theoretische Basis von C-Tests, Authentizität von C-Test-Texten, Textinhalt und Textschwierigkeit, Textauswahl und C-Test-Format speziell im Rahmen des Bundeswettbewerbs Fremdsprachen, C-Tests und Multiple-Choice-Tests im Vergleich, Auswertungsobjektivität von C-Tests (insbesondere im Hinblick auf Orthographie und akzeptable Varianten), Akzeptanz von C-Tests bei Lehrern, Schülern und Gutachtern. Es wird gezeigt, daß der C-Test eine Reihe von Vorzügen im Vergleich zu den übrigen im Bundeswettbewerb Fremdsprachen eingesetzten Tests aufweist.

1. Einführung

Der Bundeswettbewerb Fremdsprachen[1] wird seit 1985 in der BRD in einer zunehmenden Zahl von Bundesländern in den Jahrgangsstufen 7 bis 13 in allen in den Schulen unterrichteten Fremdsprachen sowie im Rahmen des Mehrsprachenwettbewerbs der Sekundarstufe II auch in nicht an der Schule gelehrten Fremdsprachen durchgeführt. Seit kurzem richtet sich der Bundeswettbewerb Fremdsprachen auch an Auszubildende. Insgesamt gesehen soll durch den Bundeswettbewerb Fremdsprachen das Erlernen von Fremdsprachen gefördert werden, wobei sowohl eine Breiten- als auch eine Spitzenförderung beabsichtigt ist.

[1] Die folgenden Hinweise zum Bundeswettbewerb Fremdsprachen sind notwendigerweise selektiv. Detailliertere Informationen finden sich z.B. in Brandt, Hertel, Meiser & Schröder (1989), Finkenstaedt & Schröder (1989), Finkenstaedt & Weller (1988), Schröder (1988) und Schröder & Stütz (1988).

Der Bundeswettbewerb Fremdsprachen ist aus dem vom Stifterverband für die Deutsche Wissenschaft initiierten Schülerwettbewerb Fremdsprachen hervorgegangen und gehört – ebenso wie die Bundeswettbewerbe 'Mathematik' und 'Jugend musiziert' zu den großen nationalen Wettbewerben unter der Schirmherrschaft des Bundespräsidenten. Im Jahr 1987 betrug die Zahl der Teilnehmer bereits mehr als 4000, darunter mehr als 1000 Teilnehmer am Einzelwettbewerb. Im Jahr 1991 waren es schon mehr als 7000 Teilnehmer, darunter 2400 Teilnehmer am Einzelwettbewerb. Nach Finkenstaedt (1989, S. 178) ist der Bundeswettbewerb Fremdsprachen "das größte Experiment, das je im Bereich des Fremdsprachenunterrichts gemacht wurde".

Der Wettbewerb findet in der Sekundarstufe I als Gruppenwettbewerb (7. bis 10. Klasse) und als Einzelwettbewerb (8. bis 10. Klasse) in einer Fremdsprache statt. Er ist ganz bewußt schulformunabhängig konzipiert. In der Sekundarstufe II (11. bis 13. Klasse) wird der Bundeswettbewerb als Mehrsprachenwettbewerb fortgeführt.

Im Gruppenwettbewerb sollen die Schüler im Rahmen eines frei gewählten Projektes eher spielerisch und kreativ mit einer in der Schule erlernten Fremdsprache umgehen. Jede Gruppe oder Klasse kann sich dabei ihr Thema und die Art der Darstellung, z.B. als Hörspiel oder Videofilm, völlig frei wählen. Bedingung ist nur, daß der Beitrag aus einem schriftlichen und einem mündlichen Teil besteht.

Am Einzelwettbewerb können alle Schüler teilnehmen, die die gewählte Wettbewerbssprache über einen Zeitraum von vier bis sechs Jahren als erste Fremdsprache oder drei bis vier Jahren als zweite Fremdsprache gelernt haben. Die Prüfung besteht aus drei Tests und einem mündlichen Teil, der auf eine Kassette aufgenommen wird. Sie wird an zentral gelegenen Schulen durchgeführt und dauert ungefähr drei Stunden.

Im Mehrsprachenwettbewerb für die Jahrgangsstufen elf bis dreizehn sollen die Schüler ihre schulischen und außerschulischen Kenntnisse in mindestens zwei Fremdsprachen beweisen. Ziel ist die Förderung der Mehrsprachigkeit auf hohem Niveau. Die Prüfung besteht aus einem schriftlichen und einem mündlichen Teil und wird zentral durchgeführt. Die ersten Preisträger des Mehrsprachenwettbewerbs werden – sofern sie ein Hochschulstudium beginnen – bei freier Wahl der Studienfächer in die Förderung der Studienstiftung aufgenommen (vgl. auch Schröder, 1988 und Stütz, 1988).

Dem erst vor kurzem eingerichteten Fremdsprachenwettbewerb für Auszubildende liegt die Erkenntnis zugrunde, daß Fremdsprachenkenntnisse in zunehmenden Maße z.B. auch für den Facharbeiter oder für die Büroberufe wichtig werden. Da für diesen Wettbewerbsteil erst wenige Daten vorliegen, werden wir im weiteren nicht auf ihn eingehen.

2. Tests im Einzelwettbewerb der Sekundarstufe I

Der Einzelwettbewerb besteht aus folgenden Teilen:

semi-kreative Aufgabe	30 Punkte
C-Test	20 Punkte
Hörverständnistest	20 Punkte
mündliche Produktion	30 Punkte

Ein Teilnehmer kann also maximal 100 Punkte erhalten, und zwar 50 im schriftlichen und 50 im auditiv-mündlichen Teil, wobei die produktiv-kreativen Verfahren höher gewichtet werden. Die Aufgaben schließen an schulische Praxis an, kopieren sie jedoch nicht und sind deshalb auch nicht mit einer Klassenarbeit vergleichbar. Kommentierte Beispiele für Aufgaben und Lösungen finden sich in Schröder & Stütz (1988).

Bei der semi-kreativen Aufgabe müssen die Wettbewerbsteilnehmer nach bestimmten Vorgaben Texte produzieren. So sollen sie z.B. zu Bildern eine zusammenhängende Geschichte schreiben oder eine vorgegebene Geschichte weiterführen. Bewertet werden Inhalt und sprachliche Leistung, insbesondere die Ausdrucksfähigkeit (vgl. die detaillierten Hinweise in Kielhöfer, 1989 und Mundzeck, 1989).

Der C-Test besteht aus vier bis fünf authentischen, möglichst in sich abgeschlossenen Texten mit jeweils 25 bzw. 20 ausgelassenen Worthälften, die ergänzt werden müssen. Dafür stehen 25 Minuten zur Verfügung.

Beim Hörverständnistest wird den Schülern ein zusammenhängender Text, der von einem "native speaker" auf Band gesprochen ist, zweimal vorgespielt. Anschließend sind Fragen zum Inhalt zu beantworten (vgl. die detaillierten Hinweise in Herbst, 1989).

Bei der mündlichen Produktionsaufgabe wird den Schülern eine fiktive Gesprächssituation vom Tonband vorgespielt, wobei sie die Rolle des Kommunikationspartners übernehmen müssen. Sie sprechen die Antworten auf

10 bis 12 Fragen jeweils auf Band, wozu jedesmal ca. 45 Sekunden zur Verfügung stehen. Auch hierbei werden die inhaltliche und die sprachliche Leistung bewertet.

Die Wettbewerbsteilnehmer, die insgesamt mehr als 80 Punkte erhalten, sind preiswürdig.

In allen drei Wettbewerben (Gruppenwettbewerb, Einzelwettbewerb, Mehrsprachenwettbewerb) wurden am häufigsten die Sprachen Englisch, Französisch und Spanisch gewählt. In der Tabelle 1 sind für den Einzelwettbewerb der Sekundarstufe I die Häufigkeiten für die Jahre 1986 – 1989 dargestellt (vgl. Finkenstaedt, 1989).

Tabelle 1
Sprachen im Einzelwettbewerb

Englisch	1489
Französisch	464
Spanisch	48
Russisch	33
Italienisch	12
Türkisch	9
Niederländisch	5

Der Hörverstehenstest und die mündliche Produktionsaufgabe folgen in allen Sprachen demselben Schema. Die semi-kreative Aufgabe wird in den selteneren Fremdsprachen konkreter, stärker gelenkt und u.U. mit mehr Hilfestellung dargeboten. Größere Abweichungen ergeben sich nur beim C-Test. Im Spanischen und Italienischen, insbesondere aber im Russischen müssen leichtere Texte verwendet werden, oder es muß vom Strukturschema des C-Tests abgewichen werden (vgl. die Abschnitte 3 und 4). In der Wettbewerbssprache Türkisch ist das C-Prinzip wegen der besonderen Struktur dieser nicht-indogermanischen Sprache problematisch.[2] Der C-Test wurde deshalb hier durch andere Aufgaben ersetzt.

Seit 1989 wird der C-Test als Screening-Test eingesetzt, um in Anbetracht der ständig wachsenden Teilnehmerzahlen die Auswertung zu vereinfachen und die Ermittlung der Wettbewerbssieger zu beschleunigen. Diese Auswertungsstrategie wird in Abschnitt 7 genauer dargestellt und begründet.

[2] Vgl. jedoch Baur & Meder (1993), die das C-Test-Prinzip auch auf das Türkische angewendet haben.

3. Zur Theorie von C-Tests

3.1 Der Lerner

C-Tests werden in der Annahme eingesetzt, daß das Erlernen einer Sprache darin besteht, kontinuierlich ein Regelsystem aufzubauen und zu internalisieren. Chomsky hat dieses Regelwissen als Kompetenz bezeichnet. Sie bildet die Grundlage für jede sprachliche Aktivität – nach Chomsky die Performanz. Die Kompetenz entwickelt sich sowohl in der Muttersprache als auch in der Fremdsprache relativ kontinuierlich und, wie es scheint, auch sehr regelhaft.

In einem Übersichtsartikel kommt Carroll (1971) zu der Ansicht, daß der Sprachlernprozeß in der Muttersprache erst in der Spätpubertät abgeschlossen wird. Dies deckt sich mit den Ergebnissen von Raatz und Klein-Braley, die einen "ceiling effect"[3] für muttersprachliche C-Tests in Englisch und Deutsch bei ungefähr 15 Jahren entdeckten (vgl. Klein-Braley, 1985a, Raatz & Klein-Braley, 1983).

Da man die Kompetenz nicht direkt erfassen kann, müssen Verfahren entwickelt werden, die unverzerrte Stichproben der Performanz liefern, die dann Verallgemeinerungen in bezug auf die Kompetenz ermöglichen. Der C-Test ist ein solches Verfahren. Wir gehen dabei davon aus, daß die vollständige Rekonstruktion eines beschädigten Textes nur dann möglich ist, wenn die Kompetenz relativ weit fortgeschritten ist.

3.2 Der Text

3.2.1 Das Prinzip der reduzierten Redundanz

C-Tests bestehen, wie bereits erwähnt, aus mehreren kurzen Texten. Der erste Satz bleibt unbeschädigt. Danach wird die zweite Hälfte von jedem zweiten Wort getilgt, bis die erwünschte Anzahl von Streichungen erreicht

[3] Ein "ceiling effect" (Testobergrenzeneffekt) liegt vor, wenn die Schwierigkeit eines Tests so gering ist, daß er nicht mehr zwischen den Probanden zu differenzieren vermag.

worden ist (vgl. zum Konstruktionsprinzip die detaillierten Ausführungen in Grotjahn, 1987 und Raatz & Klein-Braley, 1985).

Das Vorlegen mehrerer Texte und die systematische und sehr häufige Löschung von Worthälften zielten darauf ab, über ein zweistufiges Stichprobenverfahren zu einer gültigen Aussage über die Sprachbeherrschung zu gelangen. C-Tests gehören zu den Tests, die auf dem Prinzip der Reduzierung sprachlicher Redundanz beruhen (vgl. Klein-Braley, 1985b). Die Redundanz entsteht dadurch, daß Einheiten des Textes mehrfachen Regeln unterliegen: Regeln der Morphologie, des Syntax, der Semantik, der Pragmatik, der Textkohäsion, der Textkohärenz usw. Man geht von der Annahme aus, daß die natürliche Redundanz einer jeden Sprache es den muttersprachlichen Sprechern dieser Sprache ermöglicht, auch beschädigte Sprachstücke zu rekonstruieren und damit zu verstehen.

Die Notwendigkeit der Rekonstruktion gestörter Sprache kommt in der realen Welt ziemlich oft vor, z.B. am Telefon, wenn die Leitung schlecht ist, oder im Bahnhof bei Lautsprecherdurchsagen. Auch bei geschriebenen Texten findet man das Phänomen des unvollständigen bzw. beschädigten Textes, z.B. durch schlechtes Fotokopieren.

Anhand der Zahl der von den Lernern korrekt gefüllten Lücken in den C-Test-Texten kann man Schlüsse über den relativen Stand der Sprachbeherrschung der Lerner ableiten. Wer eine Sprache besser beherrscht, wird eine bessere Leistung bei der Textrekonstruktion erbringen. Damit man nicht aus einer irgendwie verzerrten Stichprobe falsche Schlüsse ableitet, werden mehrere Texte benutzt.

3.2.2 Authentizität von C-Test-Texten

(a) *Muttersprachliche Lerner*

Das Kind extrahiert aus der Masse an Sprache, von der es umgeben ist, Regeln. Trotz aller Bemühungen ist man jedoch immer noch nicht in der Lage, umfassend zu sagen, welche Regeln wann und in welcher Reihenfolge gebildet werden. Wir können daher auch keinen Test erstellen, der die Erfassung der wichtigsten Regeln zum Ziel hat. Ebensowenig sind wir in der Lage, Aussagen darüber zu machen, mit welchen sprachlichen Elementen das Kind konfrontiert worden ist, d.h. welche Wörter, Konstruktionen, Redewendungen usw. das Kind vermutlich kennt. Eigentlich können wir nur von einem stochastischen Modell ausgehen: Gewisse Elemente werden in der Sprache häufiger benutzt als andere; die häufigeren Elemente wird das Kind eher

kennen. Der C-Test ist ein Ausweg aus diesem Dilemma. Auch bei einem C-Test wissen wir lediglich ansatzweise, welche Regeln zur Wiederherstellung des Textes angewendet werden müssen. Wir sehen jedoch, *wieviel* von dem Text erfolgreich rekonstruiert worden ist und nehmen die Anzahl richtiger Lösungen als eine grobe Maßzahl für das augenblickliche Ausmaß an sprachlicher Kompetenz. Diese Annahme ist vor allem dann berechtigt, wenn die Texte, die als Grundlage für die C-Tests dienen, authentisch sind, denn nur dann können sie ihre Funktion als Stichprobenverfahren optimal erfüllen.

Authentische Texte werden in der Regel der üblichen statistischen Zusammensetzung der Sprache entsprechen. Dies bedeutet, daß Wörter, syntaktische Fügungen oder auch Redewendungen, die in der Sprache häufig benutzt werden, auch häufig in den ausgewählten Texten vorkommen und daß in der Sprache seltene Einheiten auch wenig frequent im Text sind. Je 'normaler', oder linguistisch ausgedrückt, je unmarkierter die Texte sind, desto eher spiegeln sie auch die Sprache in ihrer Normalität wider. In einer Untersuchung zu Cloze-Tests hat Finn (1977-78) festgestellt, daß die Häufigkeitsverteilung der Wörter in Texten sehr stark der Verteilung des Wortschatzes in Corpora ähnelt. Allerdings gab es eine Ausnahme: Inhaltswörter, die die Textbedeutung tragen, sind häufiger in einem Text vertreten als in den Ranglisten der Corpora. Eigentlich wäre es von der Theorie her sogar wünschenswert, die C-Test-Texte anhand von einem Zufallsverfahren quasi automatisch auszusuchen. Das geht aus einer Reihe von Gründen nicht, so z.B. weil ein C-Test-Text trotz seiner Kürze eine in sich abgeschlossene Einheit bilden soll (vgl. auch Kirkwood, Wolfe, Maynes, Millar, Sword & Sword, 1980).

(b) *L2-Lerner*

Die Forderung, authentische Texte zu benutzen, gilt nicht nur für Kinder, die ihre eigene Sprache lernen, sondern auch für andere Lernergruppen, z.B. für Sprachlerner, die in der Familie ihre Muttersprache sprechen, ansonsten aber eine andere Sprache lernen und benutzen, wie z.B. die Kinder von ausländischen Arbeitnehmern in der Bundesrepublik Deutschland. Hier handelt es sich wie beim Erlernen der Muttersprache um einen natürlichen Sprachlernprozeß: Auch diese Lerner müssen aus der Sprache, mit der sie konfrontiert werden, ein Regelsystem aufbauen, und deshalb bilden auch für diese Gruppe authentische Texte das geeignete Testmaterial.

(c) *Fremdsprachenlerner*

Auch fortgeschrittene Fremdsprachlerner sollten mit authentischen C-Test-Texten konfrontiert werden, da sie in der Regel bereits über eine relativ breite Erfahrung mit authentischem fremdsprachlichen Material verfügen.

Für eine Gruppe sind u.E. jedoch authentische Texte nicht angemessen, nämlich für homogene schulische Fremdsprachenlerner auf unterem und mittlerem Niveau in der jeweiligen Sprache. Diese Gruppe bekommt die Sprache 'häppchenweise' und gestuft in einer als sinnvoll angesehenen didaktischen Progression präsentiert. Diese Sprachstichprobe ist in jeder Hinsicht eingeschränkt und weicht möglicherweise von der Alltagssprache recht deutlich ab. Es wäre nicht fair, wenn man diese Lerner in einem C-Test mit einer Vielzahl von sprachlichen Elementen und Strukturen konfrontieren würde, die ihnen noch nicht begegnet sind.

Bei der Entwicklung von C-Tests für Fremdsprachenlerner in den ersten Unterrichtsjahren sollten deshalb zunächst Listen der gelehrten Sprachelemente aufgestellt werden, die dann die Grundlage für neu zu schreibende Texte bilden. Dabei sollten vorzugsweise Muttersprachler zum Einsatz kommen, die keine Kenntnisse der eigentlichen Lehrtexte haben. Rump (1985) hat gezeigt, daß dieses Verfahren durchaus realisierbar ist.

3.2.3 Textinhalt

Die Texte in C-Tests sind für den Testhersteller nicht als Inhaltsträger, sondern in erster Linie in ihrer Eigenschaft als Elizitationsinstrument von Interesse. Wichtig ist, daß sie den Probanden die Gelegenheit geben, ihre Sprachbeherrschung zu zeigen. Der Sprachtestkonstrukteur sollte deshalb sorgfältig darauf achten, daß die Texte im linguistischen Sinne "unmarkiert" sind. Sie dürfen in keiner Weise auffallen: weder in ihrem Inhalt noch in ihrem pragmatischen Bezug und schon gar nicht in ihrer sprachlichen Gestaltung. Ferner sollten die Texte, die in einem C-Test verwendet werden, kurze, inhaltlich abgeschlossene Einheiten bilden.

C-Test-Texte sollten auch deshalb keinen zu speziellen oder gar fachlichen Inhalt aufweisen, weil sonst einzelne Probanden benachteiligt oder bevorzugt werden könnten. Es soll nicht Spezialwissen getestet werden, sondern Sprachbeherrschung. Dadurch, daß im C-Test im Gegensatz zum Cloze Test mehrere Texte verwendet werden, wird im übrigen die Gefahr, daß ein C-Test inhaltlich nicht fair ist, minimiert.

Weiterhin sollten die Texte für den Durchschnittsprobanden nicht intellektuell zu anspruchsvoll oder auch zu anspruchslos sein. Es soll ja keine Intelligenz getestet werden.

Es ist natürlich wünschenswert, daß die Texte vom Inhalt her für die Probandengruppe nicht nur angemessen, sondern auch interessant sind. Allerdings zählt für den Testkonstrukteur vor allem der Stichprobencharakter des Textes. Muß er zwischen einem interessanten, aber sprachlich etwas eigenwilligen Text und einem langweiligen, aber sprachlich 'normalen' Text wählen, würde der langweiligere Text den Zuschlag erhalten.

Eine Methode, die Normalität eines C-Test-Textes zu überprüfen, besteht darin, daß man ihn Muttersprachlern vorlegt. Wenn diese keine Schwierigkeiten beim Einsetzen der fehlenden Teile empfinden und ihnen auch sonst der Text nicht "eigenartig" vorkommt, dann kann der Text als prinzipiell geeignet angesehen werden.

Es ist allerdings für die Muttersprachler einer Sprache nicht immer sofort einsichtig, daß viele Texte einen starken kulturellen Bezug haben und daher Wissen voraussetzen, über das besonders ausländische Probanden normalerweise nicht verfügen. Bei C-Test-Texten ist es aber auch wichtig sicherzustellen, daß von den Probanden kein *implizites* Welt- oder Kulturwissen gefordert wird, über das sie nicht verfügen können. Beispielweise wäre es nicht angebracht, bei Englischtests für Kinder asiatischer Immigranten englisches oder europäisches Märchengut als Testgrundlage einzubringen.

Wenn landeskundliches Wissen im Fremdsprachenunterricht ein Nebenlehrziel hohen Ranges darstellt, dann kann es durchaus sinnvoll sein, gleichzeitig Sprache und landeskundliches Wissen zu überprüfen. Bei einem reinen Sprachtest sind jedoch spezifische landeskundliche und kulturelle Inhalte ein Störfaktor, der entsprechend der Testtheorie auszuschalten ist. Zudem ist es gerade in bezug auf die großen europäischen Sprachen nicht möglich, bei allen Lernern ein bestimmtes kulturelles Wissen vorauszusetzen. Dies gilt im besonderen Maße für Englisch in seiner Funktion als Lingua Franca. Aus all diesen Gründen sollten die Texte inhaltlich möglichst unmarkiert sein.

3.2.4 Textschwierigkeit

C-Test-Texte sollten weder zu leicht noch zu schwierig sein. Hier ergibt sich das Problem, daß C-Tests regelmäßig auch von erfahrenen Lehrern oder auch von mit dem C-Test wenig vertrauten Testkonstrukteuren als

zu schwierig eingeschätzt werden, obwohl die empirischen Ergebnisse anschließend zeigen, daß die Texte durchaus geeignet waren. Die Lehrer möchten, daß ihre Schüler/Studenten optimale Ergebnisse bei den Tests erzielen (ca. 100%!) und bewerten die Texte schon deshalb als zu schwierig. Für Testzwecke ist es aber notwendig, daß die Testwerte ausreichend streuen. Daher dürfen die Texte nicht vollständig lösbar sein.

Es ist allerdings möglich, die relative Schwierigkeit von englischen C-Test-Texten zueinander durch Regressionsgleichungen im voraus zu berechnen. Dabei werden je eine Maßzahl für die Komplexität der Grammatik (die durchschnittliche Satzlänge) und für die Breite des Wortschatzes (die *Type-Token Ratio*) als Prädiktoren eingesetzt. Der Grad der Übereinstimmung der auf diese Weise geschätzten Schwierigkeiten mit den empirisch ermittelten Schwierigkeiten ist hoch (Klein-Braley, 1984, 1985a). Zur Zeit werden weitere Untersuchungen in dieser Richtung angestellt (Klein-Braley, in Vorbereitung).

3.2.5 Veränderung von authentischen Texten

Wie bereits erläutert, sollten C-Test-Texte kurz sein, jedoch in sich abgeschlossen. Sie sollten weder sprachlich noch inhaltlich noch sonst irgendwie aus dem Rahmen fallen. Sie sollten nicht aus literarischen Quellen stammen – Literatur ist immer "geformte Sprache". Heikle Themen wie Religion, Drogen usw. sind zu vermeiden, ebenso wie manierierte oder humorvolle Texte. Der Wortschatz sollte nicht gewollt ausgefallen sein, und die Art der Textgestaltung sollte keine ungewöhnlichen Formulierungen oder Konstruktionen bedingen.

Texte, die diesen Ansprüchen genügen, sind nicht leicht zu finden. Manchmal findet man allerdings Texte, die bei minimalem Eingreifen durchaus geeignet wären, z.B. indem man ein ausgefallenes Wort durch einen Alltagsbegriff ersetzt oder einen Nebensatz, der nichts zum Text beiträgt, wegläßt, um auf die richtige Länge zu kommen. Solche Eingriffe scheinen uns legitim, wenn sie nur einzelne kleine Teile des Textes betreffen und wenn sie durch einen Muttersprachler der betreffenden Sprache vorgenommen werden.

4. Textauswahl und C-Test-Format im Bundeswettbewerb Fremdsprachen

Es wird wohl nie möglich sein, ein absolut objektives Verfahren der Textauswahl zu entwickeln. Das zeigt auch der gescheiterte Versuch einer quasimechanischen Auswahl von Cloze-Test-Texten durch Kirkwood et al. (1980). Das entscheidende Kriterium hinsichtlich der Adäquatheit der auszuwählenden Texte bleibt deshalb letztendlich die subjektive Einschätzung der Auswählenden. Damit grobe Fehleinschätzungen vermieden werden, sollte jedoch der Testersteller oder zumindest eine zur Textauswahl hinzugezogene Person die anvisierte Zielgruppe gut kennen.

Vermutlich können Entscheidungen in bezug auf die Gesamteignung von Texten unter Berücksichtigung der diskutierten Kriterien zur Zeit – und vermutlich bis auf weiteres – am ehesten von einer Expertengruppe getroffen werden. Beim Bundeswettbewerb Fremdsprachen wird das Problem der Textauswahl dadurch gelöst, daß die für die verschiedenen Sprachen zuständigen Testkonstrukteure zunächst bei einer Auswahlkommission eine Reihe von als C-Test eingerichteten Texten als Vorschlag einreichen. Die Kommission trifft dann die endgültige Auswahl. Allerdings neigt die Kommission immer wieder dazu, zu leichte Texte auszuwählen. Dies hängt auch mit der Akzeptanzproblematik von C-Tests zusammen (vgl. Abschnitt 9).

Im Bundeswettbewerb Fremdsprachen haben wir es mit einer Testgruppe zu tun, deren sprachliche Lerngeschichte nur sehr bedingt bekannt ist. Es handelt sich zudem nicht um eine einheitliche Gruppe von Fremdsprachenlernern, die ein vorbereitetes Programm absolviert haben, sondern um Schüler, die durchaus unterschiedlichen Lernverfahren ausgesetzt gewesen sind. Außerdem sind es häufig Schüler, die sich aufgrund ihres Interesses für Fremdsprachen zusätzlich zum schulischen Fremdsprachenunterricht um weitere Begegnungsmöglichkeiten mit der Fremdsprache bemüht haben, so z.B. um fremdsprachliche Bücher, Filme, Schallplatten oder Radiosendungen, um Auslandsaufenthalte, um Kontakte mit Ausländern usw. Aufgrund dieser Merkmale muß diese Gruppe ähnlich behandelt werden wie L1- und L2-Lerner, d.h. es kommt wiederum in erster Linie der authentische Text als Testgrundlage in Frage. Ein solcher Text zielt nicht auf ein bestimmtes Curriculum, er ist zudem anspruchsvoller als der übliche Schulbuchtext und insgesamt gesehen fairer für die vom Bundeswettbewerb Fremdsprachen erfaßte Probandengruppe.

Für die recht weit fortgeschrittenen Teilnehmer am Einzelwettbewerb Englisch konnten diese Forderungen weitgehend eingehalten werden. In den übrigen Sprachen des Einzelwettbewerbs mußten hingegen z.T. auch nichtauthentische Texte eingesetzt werden oder authentische Texte relativ stark bearbeitet werden, da sonst die schwächeren Teilnehmer zu weit überfordert gewesen wären.

Bei der Auswahl wurde weiterhin darauf geachtet, daß die Texte keine sprachlichen Strukturen enthielten, die Lernern (z.B. in der neunten Klasse) aufgrund des regulären schulischen Fremdsprachenunterrichts vermutlich unbekannt waren. So finden sich z.B. in den ausgewählten französischen C-Test-Texten keine "subjonctif"-Formen.

Betrachtet man die bisher erstellten C-Tests in den verschiedenen Wettbewerbssprachen, dann fällt auf, daß einzelne Testkonstrukteure besonders in den kleinen Sprachen die klassischen Testerzeugungsregeln nicht eingehalten haben (vgl. die entsprechende Kritik in Grotjahn, 1989 und Bauer, 1989). Explizite und zumindest ansatzweise wissenschaftlich abgesicherte Begründungen für die Unterschiede im Vorgehen liegen jedoch nicht vor. So gibt es z.B. Diskrepanzen in der Behandlung von einsilbigen Wörtern und von Eigennamen bei der Tilgung. Außerdem gibt es C-Test-Versionen, in denen im Widerspruch zum klassischen Konstruktionsprinzip nicht nur Teile von Wörtern, sondern zusätzlich auch ganze Wörter getilgt worden sind. Es ist nicht bekannt, inwieweit diese Veränderungen die Reliabilität und Validität der C-Tests beeinflussen.

Ferner wird in bestimmten Wettbewerbssprachen jeder getilgte Buchstabe durch einen einzelnen Punkt symbolisiert. Dadurch wird der entsprechende C-Test zwar leichter und damit gegebenenfalls addressatengerechter, gleichzeitig kommt es jedoch – zumindest in einem gewissen Umfang – zu einer im Rahmen einer fremdsprachlichen Aufgabenstellung unerwünschten Buchstabenzählerei auf seiten der Schüler. Dieser Nebeneffekt könnte vermieden werden, wenn man zur Senkung der Schwierigkeit z.B. *weniger* als die Hälfte jedes zweiten Wortes oder die Hälfte jedes *dritten* Wortes tilgen würde (vgl. Süßmilch, 1985).

Weiterhin gibt es Inkonsistenzen in bezug auf die Testinstruktion (vgl. Grotjahn, 1989). In kognitionspsychologischen Arbeiten ist vielfach nachgewiesen worden, daß das Verhalten der Probanden bei Konstanz der eigentlichen Problemlösungsaufgabe in entscheidender Weise von der jeweiligen Instruktion abhängen kann. In Anbetracht dieser Tatsache bedürfen u.E.

die vorhandenen Unterschiede in den C-Test-Instruktionen in den verschiedenen Wettbewerbssprachen einer expliziten Begründung.

So ist z.B. der Hinweis, daß die *Hälfte* eines jeden zweiten Wortes fehlt, nicht unproblematisch, da dies zu der bereits erwähnten Buchstabenzählerei führen kann. Belege, daß dies tatsächlich der Fall sein kann, finden sich in einigen an der Ruhr-Universität Bochum aufgenommenen Protokollen des Lauten Denkens beim C-Test-Lösen (vgl. hierzu Feldmann, Grotjahn & Stemmer, 1986).

Noch problematischer ist es, wenn die schriftliche Instruktion ganz weggelassen wird, wie dies im Wettbewerb 1988 bei den französischen, englischen, russischen und italienischen C-Tests der Fall war. In einem solchen Fall ist nicht sicher gestellt, daß alle Schüler die Texte unter den gleichen Voraussetzungen bearbeiten.

5. C-Test und Multiple-Choice-Test: Voruntersuchungen

Bis 1986 wurden im Einzelwettbewerb in der Sekundarstufe I anstelle des jetzt verwendeten C-Tests klassische Multiple-Choice-Tests (MC-Tests) in den jeweiligen Sprachen zur Messung von grammatikalischen Kenntnissen und der kommunikativen Kompetenz eingesetzt. Diese Tests wurden nur in Englisch vorerprobt.

Beim Einsatz und bei der Entwicklung dieser MC-Tests zeigten sich jedoch bald verschiedene Probleme, die einen allgemeinen Einsatz in Frage stellten, so z.B.:

- Der Inhalt von MC-Aufgaben ist nicht von vornherein definiert. Er muß zunächst auf der Grundlage einer Sprachtheorie, von Lehrplänen oder Lehrbüchern festgelegt werden.
- Solche MC-Aufgaben sind jedoch artifiziell, nicht authentisch und zudem unpädagogisch.
- Ein MC-Test, der gleichzeitig Grammatik, kommunikative Kompetenz und damit auch den Wortschatz messen soll, ist mehrdimensional.
- Ein MC-Test ist stark lehrplan- und schulbuchabhängig, was bei der Heterogenität der vorliegenden Zielpopulation sehr problematisch ist.
- Wegen eben dieser Heterogenität muß jeder Test sehr leichte und sehr schwere Aufgaben enthalten.

– Eine Vorerprobung eines jeden MC-Tests an repräsentativen Stichproben mit nachfolgender Itemanalyse ist unabdingbar.

C-Tests haben gegenüber MC-Tests eine Reihe von Vorteilen. Der Testinhalt ist vorgegeben, nämlich "Sprache", die anhand authentischen Materials theoriegeleitet und ganzheitlich getestet wird. C-Tests sind relativ unterrichtsunabhängig und lassen sich in den meisten Sprachen in jeweils beliebiger Zahl auf relativ einfache Weise entwickeln (vgl. die Abschnitte 3 und 4).

Wir haben deshalb zunächst in einer empirischen Untersuchung überprüft, ob C-Tests prinzipiell im Einzelwettbewerb der Sekundarstufe I eingesetzt werden könnten und ob sie evtl. den MC-Test ersetzen könnten (vgl. Raatz & Klein-Braley, 1986).

Für die empirische Untersuchung dieser sehr allgemeinen Fragestellung mußten einschränkende Randbedingungen vorgegeben werden, da der Aufwand sonst zu groß geworden wäre. Wir haben deshalb C-Tests und MC-Tests nur in der wichtigsten Wettbewerbssprache Englisch in 8. und 10. Klassen an Hauptschulen, Realschulen und Gymnasien verglichen.

Für die Entwicklung des MC-Tests lagen uns 50 Items mit jeweils vier Antwortmöglichkeiten vor. Diese Items stammten aus der "Kölner Itembank für Anglistikstudenten" von H. Bonheim und umfaßten meist grammatikalische Probleme, Wortschatz, Textverständnis, Landeskunde und den angemessenen Sprachgebrauch in sozialen Situationen (Kommunikative Kompetenz).

Für den C-Test haben wir 13 authentische englische Texte mit jeweils 25 Lücken zusammengestellt. Diese Texte waren bei englischen Schülern und teilweise auch bei Anglistikstudenten erprobt worden, in deutschen Schulen allerdings noch nie.

Aus dem Aufgabenmaterial wurden vier parallele Testformen mit jeweils 20 MC-Items und 4 C-Texten erstellt, die – aus Gründen der Vergleichbarkeit – bestimmte Items gemeinsam hatten.

Diese Tests wurden im Sommer 1985 bei insgesamt 1303 Schülern durchgeführt. Tabelle 2 zeigt die Zusammensetzung der Stichprobe. Tabelle 3 enthält die durchschnittlichen prozentualen Schwierigkeiten des MC-Tests und des C-Tests für die verschiedenen Teilstichproben.

Tabelle 2
Zusammensetzung der Stichprobe

	Klasse 8	Klasse 10
Hauptschule	128	182
Realschule	199	122
Gymnasium	303	269

Tabelle 3
Schwierigkeiten

	MC-Test		C-Test	
	Klasse 8	Klasse 10	Klasse 8	Klasse 10
Hauptschule	33	40	28	38
Realschule	37	44	37	57
Gymnasium	50	57	48	59

Die Ergebnisse dieser Untersuchung lassen sich wie folgt zusammenfassen:

(a) *Schwierigkeiten*

Erwartungsgemäß sind beide Tests in 10. Klassen leichter als in 8. Klassen, und ihre Schwierigkeit steigt vom Gymnasium über die Realschule zur Hauptschule an. Zwischen beiden Testformen ergeben sich prinzipiell keine Unterschiede.

(b) *Reliabilitäten*

Die Reliabilitäten wurden für den MC-Test nach der α-Formel von Cronbach abgeschätzt. Dabei wurden einmal alle Teilstichproben, einmal die Gesamtstichprobe zugrundegelegt.

Für die Teilstichproben ergaben sich im MC-Test Reliabilitäten zwischen -.22 und .57, im C-Test zwischen .19 und .84. In der Gesamtstichprobe betrug die Reliabilität des MC-Tests .57, die des C-Tests .83.

Die Koeffizienten streuten also stark, lagen aber beim C-Test im Einzelnen und insgesamt deutlich höher als beim MC-Test. Diese Ergebnisse waren in Anbetracht der Tatsache, daß beide Tests nicht vorerprobt waren, zu erwarten. Sie machen deutlich, daß auf eine Vorerprobung mit

anschließender Aufgabenanalyse nicht verzichtet werden kann. Allerdings scheint dieses Problem bei C-Tests nicht so gravierend zu sein.

(c) *Validität*

Die Korrelationen zwischen MC-Test und C-Test liegen für die einzelnen Teilgruppen und Testformen zwischen $-.18$ und .52, wobei die niedrigen Werte durch die geringen Reliabilitäten des MC-Tests bedingt sein dürften. Betrachtet man die Gesamtgruppe und wendet man eine doppelte Minderungskorrektur an, so liegt die geschätzte Interkorrelation bei .57. Das ist ein Hinweis darauf, daß beide Verfahren etwas ähnliches messen, nur unterschiedlich genau.

(d) *Akzeptanz*

Der MC-Test besaß bei den getesteten Schülern und auch bei den betroffenen Lehrern eine hohe Augenscheingültigkeit und Akzeptanz, der C-Test nur eine niedrige. Das mag daran liegen, daß MC-Aufgaben bekannt sind und immer wieder verwendet werden, C-Tests und das Konzept einer globalen Sprachstandsmessung aber weitgehend unbekannt sind (vgl. die detaillierteren Ausführungen zur Akeptanz in Abschnitt 9).

Aus den angeführten Ergebnissen haben wir damals folgende Schlußfolgerungen gezogen, die für die Ausgestaltung des Wettbewerbs von Bedeutung waren:

1. Prinzipiell können C-Tests im Einzelwettbewerb der Sekundarstufe I eingesetzt werden.
2. C-Tests messen wahrscheinlich ähnliche Aspekte der Fremdsprache wie heterogene MC-Tests, aber theoriegeleitet und reliabler.
3. C-Tests sind einfacher als MC-Tests zu entwickeln, und zwar u.a. deshalb, weil im Gegensatz zu MC-Tests eine Vorerprobung von C-Tests zwar empfehlenswert, aber nicht unbedingt notwendig zu sein scheint.
4. Die gegen C-Tests sprechende geringere Akzeptanz und die unökonomischere Auswertung wird durch diese Argumente mehr als aufgehoben.

Wir haben damals empfohlen, in die Testbatterie des Einzelwettbewerbs zusätzlich einen C-Test in der jeweiligen Fremdsprache aufzunehmen und später nach einer Erprobungsphase zu diskutieren, ob auf den MC-Test verzichtet werden kann.

6. C-Test und Multiple-Choice-Test im Einzelwettbewerb

Vom Jahr 1986 an wurden im Einzelwettbewerb in der Sekundarstufe I in den Wettbewerbssprachen Englisch und Französisch MC-Tests und C-Tests nebeneinander eingesetzt, ab 1987 auch in den Sprachen Italienisch, Spanisch und Russisch. Die Tabellen 4 und 5 enthalten einige Ergebnisse der von uns durchgeführten Analysen der Ergebnisse für die Sprachen Englisch und Französisch.[4] Die drei anderen Sprachen wurden so selten gewählt, daß eine statistische Analyse nicht sinnvoll war. Leider sind die Angaben in den Tabellen unvollständig, da zur Auswertung und Analyse jeweils nicht alle Daten zur Verfügung standen.

Tabelle 4
Ergebnisse der Analysen in Englisch

Jahr	N	Reliabilität		Interkorrelation
		MC-Test	C-Test	MC-Test/C-Test
1985	96	.80	.82	.72
1986	199	–	.89	–
1987	419	–	.84	.71

Tabelle 5
Ergebnisse der Analysen in Französisch

Jahr	N	Reliabilität		Interkorrelation
		MC-Test	C-Test	MC-Test/C-Test
1985	39	.86	.95	.86
1987	41	–	.89	.52

Die Testergebnisse des Wettbewerbs 1988 für Englisch wurden von Bauer (1989) analysiert. 1988 wurden MC-Tests und C-Tests letztmalig zusammen eingesetzt.

Bauer wertete die Daten von 420 Wettbewerbsteilnehmern aus und erhielt für den MC-Test eine Reliabilität von .87. Für den C-Test gab er keinen Koeffizienten an, wahrscheinlich deshalb, weil es in dem entsprechenden

[4] Zum C-Test-Französisch vgl. auch die detaillierteren Analysen in Grotjahn (1992a).

Jahrgang große Unsicherheiten bei der Testauswertung gab (Bauer konnte dies an einer Reihe von Beispielen zeigen). Bauer ermittelte außerdem die Korrelation beider Verfahren zum Gesamtpunktwert (inklusive dieser Verfahren) an. Diese betrug für den MC-Test .75, für den C-Test .70. Der letztere Wert ist in Anbetracht der problematischen Auswertungsobjektivität überraschend hoch.

Die Konsequenz, die Bauer seinerzeit aus diesen Ergebnissen zog, nämlich den MC-Test im Wettbewerb zu "stärken", wäre durchaus überlegenswert, wenn nicht der C-Test so hoch mit dem MC-Test korrelieren würde (siehe Tabelle 4) und wenn die Entwicklung von guten MC-Tests nicht vergleichsweise zu aufwendig und teuer wäre. Diese Tatsachen, nicht zuletzt aber auch die bessere theoretischen Fundierung der C-Tests, führten schließlich zu der Entscheidung, daß vom Jahr 1989 an auf den MC-Test zugunsten eines C-Tests verzichtet wurde.

Im Jahre 1991 wurde eine weitere Analyse der Ergebnisse des Einzelwettbewerbs Sekundarstufe I für mehrere Sprachen durchgeführt. Die entsprechenden Untersuchungen sind in Raatz & Klein-Braley (1993) dargestellt.

7. C-Tests als Screening-Tests

Die Auswertung der vier Wettbewerbstests ist insgesamt sehr zeitaufwendig, was besonders für die semi-kreative Aufgabe und die mündliche Produktion gilt. Es liegt deshalb nahe, zur Vereinfachung der Auswertung ein Screening-Verfahren anzuwenden. Dazu wählt man aus den Tests einen Test X aus, der möglichst hoch mit dem Gesamtergebnis korreliert und selbst möglichst ökonomisch und objektiv ausgewertet werden kann. Dann geht man folgendermaßen vor:

1. Man wertet zuerst den Test X aus.
2. Die P% schlechtesten Schüler scheiden aus dem Verfahren aus, ihre anderen Tests werden nicht mehr ausgewertet.
3. Bei den $(100 - P)$% besseren Schülern werden auch die drei anderen Tests ausgewertet und der Gesamtpunktwert ermittelt, der dann für die Verleihung eines Preises entscheidend ist.

Bei diesem Vorgehen sind zwei Arten von Fehlentscheidungen möglich:

- Ein potentieller Preisträger wird wegen eines schlechten C-Test-Ergebnisses ausgeschlossen.

- Ein Schüler bleibt wegen eines guten C-Test-Ergebnisses im Rennen, erhält aber wegen seiner geringen Gesamtleistung dann doch keinen Preis.

Der Anteil der erstgenannten Fehlentscheidungen sollte bei einem Screening-Verfahren möglichst bei Null liegen. Die andere Fehlentscheidung belastet zwar die Auswerter und senkt die Ökonomie dieser Strategie, nimmt den Schülern aber keine Chancen.

Beim Wettbewerb 1987 wurde diese Strategie erstmalig untersucht (Raatz & Klein-Braley, 1989). Als Screening-Test wurde der C-Test gewählt, da er von allen Tests damals am höchsten mit dem Gesamtpunktwert korrelierte (Englisch: $N = 419$, $r = .69$, Französisch: $N = 100$, $r = .74$). Als Auswertungsrate wurde $P = 50\%$ angesetzt. Der Anteil der Preisträger mit mehr als 80 Punkten war in Englisch 5.7% in Französisch 13.3%, in allen Sprachen zusammen 10%.

Der Anteil der Fehlentscheidungen der ersten Art wurde nun nach zwei Methoden bestimmt:

Einmal haben wir auf der Grundlage der folgenden Vorgaben (Selektionsrate 50%, Anteil der Geeigneten 10%, Validität .70) mit Hilfe der Taylor-Russell-Tabellen (vgl. Diehl & Kohr, 1991, S. 250ff.; Taylor & Russell, 1939) diesen Anteil abgeschätzt. Es ergab sich ein Fehler von 0.5%. Das würde bedeuten, daß man bei 1000 Wettbewerbsteilnehmern im Mittel 5 potentielle Preisträger ungerechtfertigt zurückweist.

Die zweite Methode besteht darin, daß man bei einer vorhandenen Datenmenge von vollständig ausgewerteten Testsätzen empirisch durch Auszählen die Anzahl der Fehlentscheidungen ermittelt. Nach dieser Methode hat sich bei dem damals vorliegenden Datensatz ($N = 519$ in fünf Sprachen) nur eine einzige Fehlentscheidung, und zwar in Englisch, ergeben.

Bauer (1989) diskutiert diese Strategie ebenfalls. Da in seiner Untersuchung der MC-Test eine höhere Korrelation zum Gesamtpunktwert als der C-Test besaß, schlug er natürlich seine Verwendung als Screening-Test vor. Die Kritik am Einsatz des C- Tests für diesen Zweck erläuterte er durch eine Abbildung (S. 15). Dabei geht er allerdings von falschen Selektionsraten aus und kommt deshalb zu für den C-Test negativen Schlüssen. Bei einem Mindestpunktwert für den Wettbewerbserfolg von 80 hätte sich auf der Grundlage seines Diagramms keine einzige Fehlentscheidung der ersten Art ergeben.

Bauer schlug vor, daß eine Kombination aus C-Test und MC-Test als Screening-Test eingesetzt werden sollte. In Anbetracht der empirischen Daten scheint das jedoch nicht notwendig zu sein.

Zur Zeit werden C-Tests in allen Wettbewerbssprachen als Screening-Tests eingesetzt. Dabei muß man sich jedoch darüber im Klaren sein, daß man bei diesem Vorgehen niemals mit festen Cut-off-Punkten arbeiten kann, sondern in jeder Wettbewerbssprache zunächst den Median der Verteilungen der C-Test-Ergebnisse ermitteln muß. Schüler mit einem schlechteren C-Test-Ergebnis fallen aus der Konkurrenz in dieser Sprache heraus.

Wenn man das so beschriebene Screening-Verfahren anwendet, muß man immer damit rechnen, daß man einige wenige Schüler ungerecht beurteilt, ihnen also einen möglichen Preis vorenthält. Wenn man das nicht will, muß man alle Tests auswerten und die sehr viel höheren Kosten in Kauf nehmen.

8. Zur Auswertungsobjektivität von C-Tests

Betrachtet man die von den Auswertern des Bundeswettbewerbs korrigierten C-Tests, so fällt auf, daß es sowohl innerhalb einzelner Sprachen als auch zwischen den Sprachen sowie zudem auch zwischen einzelnen Bundesländern nicht unerhebliche Inkonsistenzen bei der Bewertung der Lösungen gibt (vgl. zum Folgenden Grotjahn, 1989; zu Inkonsistenzen bei der Korrektur englischer C-Tests vgl. auch Bauer, 1989). So werden beispielsweise Rechtschreibefehler von manchen Auswertern als ganzer Fehler, von anderen Auswertern lediglich als halber Fehler gewertet, ohne zu prüfen, ob diese unterschiedliche Gewichtung z.B. negative Auswirkungen auf die Güte der jeweiligen C-Test-Version hat. Weiterhin scheint vielfach auch keine Einigkeit darüber zu bestehen, was überhaupt als Rechtschreibefehler zu werten ist. Ähnliches gilt für die Wertung von Varianten als korrekte Lösung.

Es muß betont werden, daß die festgestellten Inkonsistenzen kein Beleg dafür sind, daß C-Tests *prinzipielle* Mängel in ihrer Auswertungsobjektivität aufweisen. Die Inkonsistenzen sind vielmehr in erster Linie ein Beleg dafür, daß für den Einsatz von C-Tests exakte Auswertungsvorschriften zu formulieren sind und daß deren Einhaltung streng zu überwachen ist.

Das Problem der nicht ausreichenden Auswertungsobjektivität stellt sich übrigens – z.T. in weit stärkerem Maße – auch in anderen Testteilen des Wettbewerbs und ist, wie die Arbeiten von Bonheim (1983), Legenhausen (1988) und Mundzeck (1983) zeigen, ein Erbe aus dem Schülerwettbewerb

Fremdsprachen. Im Gegensatz zum C-Test ist die mangelnde Auswertungsobjektivität jedoch bei Testteilen wie der semi-kreativen Aufgabe oder der mündlichen Produktion ein prinzipielles Problem, das mit Hilfe von Auswertungsvorschriften nur sehr bedingt zu lösen ist.

Angesichts steigender Teilnehmerzahlen im Bundeswettbewerb Fremdsprachen und damit auch einer steigenden Zahl von zu korrigierenden C-Tests stellt sich die Frage, wie auf möglichst ökonomische Weise eine möglichst hohe Beurteilungsobjektivität bei der Auswertung von C-Tests sichergestellt werden kann. Zur Beantwortung dieser Frage hat Grotjahn (1989) insgesamt 80 französische C-Tests aus den Einzelwettbewerben 1987 und 1988 erneut ausgewertet. Dabei wurden folgende sechs Lösungskategorien unterschieden:

1. unausgefüllt
2. orthographisch richtiges Original
3. grammatisch und/oder inhaltlich nicht akzeptabel
4. orthographisch richtige Variante
5. orthographisch falsches Original
6. orthographisch falsche Variante

Es zeigte sich, daß es im Einzelfall weitreichender und häufig problematischer Interpretationen sowie vielfach auch einer detaillierten Kenntnis des Leistungsstandes der Testteilnehmer bedarf, um zu einer validen Trennung zwischen Orthographiefehlern und grammatisch und/oder inhaltlich nicht akzeptablen Lösungen zu kommen. Die Problematik entsprechender Lösungskategorien insbesondere im Fall wechselnder Korrektoren und bei Korrekturen einer großen Zahl von C-Tests durch ein und denselben Korrektor unter Zeitdruck (dies ist die für den Bundeswettbewerb Fremdsprachen typische Situation) ist offensichtlich.

Insgesamt gesehen wurde deutlich, daß die Berücksichtigung von Varianten und Orthographiefehlern – zumindest im Französischen – notwendigerweise zu Reliabilitätsproblemen bei der Auswertung führt.

Tabelle 6 zeigt die prozentualen Häufigkeiten der sechs Auswertungskategorien für die Jahrgänge 1987 und 1988 sowie für die zusammengefaßten Jahrgänge aufgeteilt am Leistungsmedian. Bei der Berechnung des Medians wurden lediglich orthographisch richtige Originale als korrekt gewertet (vgl. auch die entsprechenden Untersuchungen in Grotjahn, 1992a, Abschnitt 2.4.1).

Wie Tabelle 6 zeigt, spielen Varianten – zumindest bei den beiden untersuchten Testversionen – glücklicherweise praktisch keine Rolle. Bedeutsam ist, daß auch von Schülern der höheren Leistungsgruppe kaum Varianten gefunden werden. Dies steht im teilweisen Widerspruch zu den Befunden in Grotjahn (1987, 1992a), wo sich zumindest bei einigen Texten ein deutlich höherer Anteil von zulässigen Varianten ergeben hat. Auffallend ist, daß mehr als 6% grammatisch korrekter Lösungen Orthographiefehler enthalten, wobei erwartungsgemäß in der schwächeren Leistungsgruppe etwas mehr Orthographiefehler zu finden sind.

Tabelle 6
Häufigkeiten (%) von sechs Auswertungskategorien

Kategorie	1987 N = 41	1988 N = 39	87/88 N = 80	>Md 87/88 N = 39	<Md 87/88 N = 41
unausgefüllt	12.9	16.0	14.4	4.0	24.3
orthographisch richtiges Original	60.2	60.4	60.3	72.9	48.3
grammatisch/inhaltlich nicht akzeptabel	18.9	16.0	17.5	16.0	19.0
orthographisch richtige Variante	1.3	1.7	1.5	1.5	1.5
orthographisch falsches Original	6.5	5.6	6.1	5.4	6.7
orthographisch falsche Variante	0.1	0.3	0.2	0.2	0.2

Md: Median

In Tabelle 7 werden fünf Auswertungsmethoden, d.h. Methoden zur Berechnung des Gesamtpunktwertes unterschieden und Mittelwerte (M), Standardabweichungen (SD), Schwierigkeiten (p) und Reliabilitäten (Cronbachs α) berechnet. Als 'korrekt' gewertet wurden:

Methode A: Originale ohne Orthographiefehler
Methode B: Originale ohne Orthographiefehler;
Varianten ohne Orthographiefehler
Methode C: Originale mit oder ohne Orthographiefehler

Methode D: Originale mit oder ohne Orthographiefehler;
Varianten mit oder ohne Orthographiefehler
Methode E: von den Auswertern des Bundeswettbewerb Fremdsprachen vergebene Punktzahl

Tabelle 7 zeigt, daß bei Wertung von Orthographiefehlern als 'korrekt' der Test erwartungsgemäß leichter wird und es außerdem zu einem geringfügigen Absinken der Streuungen und – möglicherweise hierdurch bedingt – der Reliabilitäten kommt. Wie in Grotjahn (1989, S. 52) und in Grotjahn (1992a, Abschnitt 2.4.1) gezeigt wird, handelt es sich bei der Varianzreduzierung aufgrund der Wertung von Orthographiefehlern als 'korrekt' um keinen Zufallseffekt. Die Wertung von Orthographiefehlern als 'korrekt' hat somit eine potentiell negative Auswirkung auf die Meßgüte des Tests insbesondere im oberen Leistungsbereich. Dies sollte bei der Entscheidung über die künftige Korrekturpraxis im Bundeswettbewerb Fremdsprachen mitberücksichtigt werden.

Tabelle 7
Mittelwerte (M), Standardabweichungen (SD), Schwierigkeiten (p) und Reliabilitäten (α) für fünf Auswertungsmethoden

	1987 N = 41				1988 N = 39			
	M	SD	p	α	M	SD	p	α
Methode A	48.2	11.6	.60	.89	48.3	12.7	.60	.92
Methode B	49.2	11.5	.62	.89	49.7	12.8	.62	.92
Methode C	53.4	10.7	.67	.86	52.8	12.1	.66	.91
Methode D	54.5	10.6	.68	.85	54.4	12.3	.68	.91
Methode E	49.8	11.5	.62	.89	49.7	12.9	.62	.92

In Tabelle 8 sind zur Charakterisierung des Ausmaßes der Übereinstimmung zwischen den fünf Auswertungsmethoden in der oberen Dreiecksmatrix die Spearmanschen Rangkorrelationskoeffizienten r_s und in der unteren Dreiecksmatrix die entsprechenden Werte für den Kendallschen Rangkorrelationskoeffizienten τ_b aufgeführt. Zusätzlich wurden für Tabelle 8 und alle weiteren Tabellen auch noch Pearsonsche Produkt-Moment-Korrelationskoeffizienten berechnet. Da die erhaltenen Werte weitgehend mit den Werten für r_s übereinstimmten, wurde auf eine Wiedergabe verzichtet.

Tabelle 8
Spearmansche Rangkorrelationen (obere Dreiecksmatrix) und Kendallsche Rangkorrelationen (untere Dreiecksmatrix) zwischen fünf Auswertungsmethoden (1987; N = 41)

Methode	A	B	C	D	E
Methode A	–	.998	.981	.980	.986
Methode B	.983	–	.979	.980	.985
Methode C	.918	.909	–	.996	.976
Methode D	.919	.915	.974	–	.974
Methode E	.931	.921	.893	.892	–

Alle Spearmanschen Korrelationen in Tabelle 8 sind extrem hoch und unterscheiden sich numerisch nur minimal. Die Kendallschen Korrelationen zwischen Methode A und B sowie zwischen Methode C und D stimmen weitgehend mit den entsprechenden Spearmanschen Korrelationen überein. Hierin spiegelt sich die Tatsache wider, daß die Zahl der von den Wettbewerbsteilnehmern gefundenen akzeptablen Varianten sehr gering ist und daß als Folge Unterschiede in der Bewertung von akzeptablen Varianten keinen Einfluß auf die Rangfolge haben. Die übrigen Kendallschen Korrelationen sind jedoch geringfügig niedriger als die jeweiligen Spearmanschen Korrelationen. Insgesamt weisen die Korrelationen darauf hin, daß die Bewertung von Orthographiefehlern als 'korrekt' zu geringen Änderungen in der Rangfolge der Schüler zu führen scheint und daß damit zumindest im Einzelfall die Auswertungsmethode über die Preisvergabe mit entscheiden kann.

Auffallend sind auch die vergleichsweise etwas geringeren Korrelationen der Methoden A bis D mit Methode E, d.h. mit den Punktzahlen, die von den Korrektoren des Bundeswettbewerbs vergeben worden sind. In diesem Sachverhalt dürfte sich die bereits diskutierte inter- und intrapersonale Varianz der Bewertungsmaßstäbe bei der Korrektur durch den Bundeswettbewerb Fremdsprachen widerspiegeln.

Berechnet man die Korrelationen zwischen den fünf Auswertungsmethoden für das Jahr 1988, ergeben sich keine auffallenden Unterschiede zu 1987 (vgl. Tab. 4 in Grotjahn, 1989). Dies kann als ein Hinweis auf die Stichprobenunabhängigkeit der erhaltenen Koeffizienten gewertet werden.

In Tabelle 9 sind wiederum die Schüler aus den Jahren 1987 und 1988 zusammengefaßt und anhand des Medians der Methode A in zwei Leistungsgruppen aufgeteilt.

Tabelle 9
Spearmansche Rangkorrelationen (obere Dreiecksmatrix) und Kendallsche Rangkorrelationen (untere Dreiecksmatrix) aufgeteilt am Median der Auswertungsmethode A (1987/88; N = 80)

(a) obere Leistungsgruppe (N = 39)

Methode	A	B	C	D	E
Methode A	–	.982	.923	.919	.974
Methode B	.936	–	.906	.920	.964
Methode C	.797	.776	–	.989	.904
Methode D	.802	.800	.947	–	.904
Methode E	.896	.878	.766	.766	–

(b) untere Leistungsgruppe (N = 41)

Methode	A	B	C	D	E
Methode A	–	.991	.959	.959	.972
Methode B	.955	–	.945	.959	.969
Methode C	.866	.831	–	.989	.948
Methode D	.863	.859	.948	–	.951
Methode E	.895	.879	.829	.843	–

Es fällt auf, daß Unterschiede in der Wertung von Orthographiefehlern anscheinend eher in der oberen Leistungsgruppe und damit in dem für den Bundeswettbewerb entscheidenden Bereich zu Abweichungen in der Rangfolge der Schüler führen. Dieser Befund stimmt im übrigen in der Tendenz mit den Ergebnissen einer Untersuchung bei Studienanfängern des Französischen (Lehramt und Magister) an der Ruhr-Universität Bochum überein (vgl. Grotjahn, 1987, S. 241 und Grotjahn, 1992a, Abschnitt 2.4.1). Da die erhaltenen Unterschiede zwischen den Leistungsgruppen jedoch sehr gering sind, möchten wir, solange das Ergebnis nicht durch weitere Studien im Rahmen des Bundeswettbewerb Fremdsprachen bestätigt worden ist, auf eine weitergehende Interpretation verzichten.

Es ist nun zu fragen, welche Konsequenzen für die zukünftige Korrekturpraxis zu ziehen sind, falls sich die erhaltenen Resultate anhand größerer Stichproben replizieren lassen sollten.

Sollte nachgewiesen werden können, daß die aus den unterschiedlichen Korrekturmethoden resultierenden Abweichungen in der Rangfolge der Wettbewerbsteilnehmer nicht vor allem auf Reliabilitätsprobleme bei der Korrektur zurückzuführen sind, ist zu klären, ob und in welchem Maße der Bundeswettbewerb Entscheidungen über Preisträger mit abhängig machen will von der Beherrschung der Orthographie der jeweiligen Wettbewerbssprache.

Sollten Reliabilitätsprobleme jedoch die Hauptursache sein, dann böte es sich an, Orthographiefehler in Zukunft grundsätzlich als 'inkorrekt' zu werten. Eine Ausnahme könnte man eventuell im Fall von eindeutigen Akzentfehlern machen. Zusätzlich könnte man auch noch die ebenfalls mit Auswertungsproblemen behafteten akzeptablen Varianten als 'inkorrekt' werten oder zumindest nur solche Varianten als korrekt werten, die nach Überprüfung durch Muttersprachler oder vergleichsweise kompetente Beurteiler als akzeptable Varianten auf den den Korrektoren auszuhändigenden Lösungsbögen aufgelistet sind. Die Vorteile dieses Verfahrens, das die Auswertungsökonomie deutlich erhöhen würde, dürften angesichts steigender Teilnehmerzahlen unmittelbar einsichtig sein. Zudem hätte dieser Auswertungsmodus auch noch den wünschenswerten Effekt, daß die Tests schwerer würden und dadurch besser im oberen Leistungsbereich differenzieren würden.

9. Akzeptanz von C-Tests bei Lehrern, Schülern und Gutachtern

Beim Einsatz von C-Tests taucht ein Problem immer wieder auf: die mangelnde Akzeptanz dieser Testform (vgl. insbesondere die Untersuchung von Legenhausen, 1989 und die Bemerkung in Hood, 1990, S. 182). Zum einen ist die Ablehnung, die die Tests vor allen Dingen bei Lehrern erfahren, sicherlich auf den mangelnden Bekanntheitsgrad von modernen Sprachtestformaten schlechthin und insbesondere in der Bundesrepublik Deutschland zurückzuführen. In seiner Kritik zu C-Tests als Verfahren im Bundeswettbewerb Fremdsprachen weist Legenhausen (1989) ausdrücklich auf dieses Problem hin und stellt hierzu fest:

> "In der übergroßem Mehrheit der Fälle bleibt die Kritik an der überhöhten Schwierigkeit des C-Tests allerdings unspezifiziert, und man kann nur vermuten, daß sie sich dann auch auf das Aufgabenformat als solches bezieht. (S. 73) ... [Negative] Einstufungen ... scheinen nun tatsächlich zu belegen, daß es sich hier um eine reine Gültigkeitseinschätzung durch Augenschein handelt. Es wird nicht einmal mit der Möglichkeit gerechnet, daß das Verfahren testtheoretisch fundiert und abgesichert sein könnte." (S. 75)

Auch Bauer (1989) kritisiert die mangelnde Augenscheingültigkeit des C-Tests und will das ganze Wettbewerbsverfahren für den Außenstehenden durch Einbeziehung von Multiple-Choice-Grammatiktests glaubwürdiger machen.

Man erinnere sich in diesem Zusammenhang nur daran, wieviel Überzeugungsarbeit in den 60er Jahren durch Sprachtester wie Lado (1961) und Harris (1969) geleistet werden mußte, um die Multiple-Choice-Aufgabe überhaupt schulfähig zu machen. Dies zeigt, wie sehr die Augenscheingültigkeit eines Testverfahrens mit seinem Bekanntheitsgrad zusammenhängt.

Daß Multiple-Choice-Aufgaben tatsächlich in deutschen Schulen eine höhere Augenscheingültigkeit als der C-Test aufzuweisen scheinen, zeigt folgende Feststellung von Legenhausen (1989, S. 75):

> "Die in den Kommentaren zum Ausdruck kommende ziemlich einhellige Ablehnung des C-Tests muß vor allem deswegen überraschen, weil die Validität der Multiple-Choice-Aufgaben von den gleichen Kommentatoren vielfach nicht in Frage gestellt wird bzw. in vier Fällen sogar ausdrücklich positiv herausgestellt wird."

Weit wichtiger als das Problem der Augenscheingültigkeit ist allerdings ein Phänomen, das man als C-Test-Paradoxon bezeichnen könnte und das mit der Tatsache zusammenhängt, daß die mit dem Einsatz von C-Tests verknüpfte Zielsetzung sowohl bei Lehrern als auch bei Probanden falsch eingeschätzt wird. Der C-Test stellt nämlich einen normorientierten Test dar, d.h. der durchschnittliche Proband soll ungefähr 50% der möglichen Punkte erzielen. Sowohl Lehrer als auch Probanden halten den C-Test jedoch offensichtlich für einen lehrzielorientierten Test und erwarten einen 90 bis 100%-igen Erfolg. Daher meinen Probanden oft genug, schlecht abgeschnitten zu haben, obwohl ihr Erfolg im Test durchaus beachtlich war. Ferner wird ein Versagen im C-Test den schwächeren Testteilnehmern bereits während der Testdurchführung schmerzhaft deutlich. Die Probanden

merken nämlich sofort, wenn sie die Lücken nicht füllen können. Bei einem Multiple-Choice-Test können sie hingegen auch dann, wenn sie die richtige Lösung nicht kennen, durch Raten zur richtigen Lösung gelangen.

Das Problem wird auch von Legenhausen (1989) gesehen, der in bezug auf den Bundeswettbewerb Fremdsprachen feststellt:

> "Der C-Test erfüllt seine Funktionen als normorientierter Test besonders dann gut, wenn – bezogen auf den Einzeltext – etwa 50% der Lücken aufgefüllt werden, wobei die Textschwierigkeit bis etwa $P = 30$ gehen kann. Das hieße in diesem Fall, daß nur knapp ein Drittel der Lücken rekonstruiert würden. Damit läßt sich für allzuviele Schüler der Inhalt der jeweiligen Geschichte kaum mehr erahnen.
>
> Das dem Schüler unbekannte Aufgabenformat führt aber zu der Erwartungshaltung, daß dieser Text zumindest im Prinzip rekonstruierbar sein müßte, da man ihn – nach seiner subjektiven Einschätzung – sonst wohl nicht mit dieser Aufgabe konfrontiert hätte. Damit kommt es zu einer hohen Diskrepanz zwischen persönlichem Anspruchsniveau und testimmanenten Erfordernissen." (S. 77).

Von den Lehrern, die einen Kommentar zum C-Test abgegeben haben, meinte der überwiegende Teil, der Test sei zu schwer gewesen, obwohl die empirischen Ergebnisse deutlich zeigten, daß die untersuchten C-Tests eher zu leicht waren (vgl. auch Grotjahn, 1989 und Raatz & Klein-Braley, 1989). Dieses Problem kann nur durch ausreichende Vorinformation über C-Tests, deren Ziele und deren Konstruktionsprinzipien gelöst werden. In der Tat vermerkt Münzel (1989, S. 105) in bezug auf die Reaktionen der Testteilnehmer:

> "Der C-Test, der in den Vorjahren immer von Kopfschütteln, Stöhnen und Fluchen begleitet war, wurde diesmal fast routinemäßig erledigt. Hier machte sich offenbar die Vorbereitung auf diesen Testteil bemerkbar."

Weiterhin kritisiert nach Legenhausen (1989) eine Reihe von Lehrern, daß in den C-Tests einzelne Wörter oder Strukturen vorkommen, die im Unterricht noch nicht behandelt worden sind. Dieses Problem ist allerdings aufgrund der von uns aufgeführten Notwendigkeit, möglichst authentische Texte zu verwenden, nur sehr bedingt zu vermeiden. Auf keinem Fall können wir Legenhausen zustimmen, wenn er schreibt:

> "Sollte sich gar herausstellen, daß die Lehrwerke/Lehrpläne der einzelnen Schultypen die Strukturen und das Vokabular der C-Texte in ganz unterschiedlichem Maß abdecken, dann muß der häufiger geäußerte Vorwurf der Unfairness Schülern bestimmter Schultypen gegenüber als gerechtfertigt erscheinen. Hier kann sich nämlich ein dem C-Test immanentes Prinzip ganz nachteilig auswirken. Bereits ein einziges nicht-rekonstruierbares, weil unbekanntes Schlüsselwort kann die Rekonstruktionsleistung des ganzen Textes nachhaltig beeinflussen ... Ich halte es für unumgänglich, daß man bei einem schultypenunabhängigen Wettbewerb und einem Festhalten am C-Test die Vereinbarkeit der Textschwierigkeit mit *allen* Lehrplänen überprüft (1989, S. 73).

Ein weiterer Einwand gegen den Einsatz des C-Tests im Bundeswettbewerb Fremdsprachen findet sich in dem Gutachten von Doyé (1989). Doyé würde den C-Test lieber durch einen Leseverständnistest ersetzen, da der C-Test als Test der "general language proficiency" nicht in das kommunikative Konzept des Wettbewerbs passe. Drei der vier Fertigkeiten – Hören, Schreiben, Sprechen – seien durch spezielle Verfahren repräsentiert, aber ein eigentlicher Lesetest fehle. Der C-Test überprüfe zwar auch Lesen, tue dies jedoch nur "zusammen mit anderen Fähigkeiten" (1989, S. 37).

Doyé übersieht hier allerdings u.E. die Tatsache, daß die subjektiven Testteile nur wenig Varianz produzieren und daß deshalb eine genaue Differenzierung zwischen den Probanden anhand dieser Tests nicht möglich ist. Zudem gibt es zwar, wie Doyé richtig bemerkt, viele geeignete Aufgabenformen zur Erstellung von Lesetests. Ohne eine explizite Theorie des Konstrukts 'Leseverständnis' und vor allen Dingen ohne eine systematische Vorerprobung ist ein entsprechender Leseverständnistest jedoch äußerst problematisch. Dies zeigt u.a. der im Jahre 1986 im Bundeswettbewerb Fremdsprachen eingesetzte Hörverständnistest, der die Leistungsfähigkeit der Wettbewerbskandidaten gefährlich unterschätzt hat (vgl. Herbst, 1989).

Schließlich wird in dem Gutachten von Bauer (1989) die mangelnde Auswertungsobjektivität von C-Tests kritisiert. Wie jedoch in Abschnitt 8 gezeigt worden ist, ist dies kein prinzipieller Mangel des C-Tests.

10. Schlußbemerkung

Insgesamt gesehen hat sich der C-Test im Rahmen des Bundeswettbewerbs Fremdsprachen sehr bewährt. Für den Einsatz des C-Tests spricht insbesondere:

- Er ist das einzige Verfahren, das sich auf eine ausreichende theoretische Grundlage stützen kann und ausreichend empirisch untersucht worden ist.
- Beim C-Test ist im Gegensatz zu der semi-kreativen Aufgabe, dem Hörverständnistest und der mündlichen Produktion eine objektive und ökonomische Auswertung möglich.
- C-Tests sind leicht zu erstellen und beruhen in allen Wettbewerbssprachen auf dem gleichen Konstruktionsprinzip.
- Auch wenn die Tests zu leicht oder zu schwer sind, sind C-Tests im allgemeinen hinreichend reliabel und valide.
- C-Tests benötigen im Vergleich zu Multiple-Choice-Tests nur eine relativ geringe Vorerprobung, die auch an nicht vergleichbaren Gruppen vorgenommen werden kann.
- Es ist möglich, für jeden Wettbewerbsdurchlauf einen hinreichend gleichwertigen C-Test zu produzieren.
- C-Tests diskriminieren gut zwischen den Teilnehmern und besitzen eine höhere Reliabilität und Validität als die übrigen im Wettbewerb eingesetzten Verfahren.
- C-Tests eignen sich gut als Screening-Tests, da sie ökonomisch auszuwerten sind und hoch mit dem Gesamtpunktwert der Wettbewerbsteilnehmer korrelieren.
- Gerade die Andersartigkeit und Nicht-Schulkonformität des C-Tests unterstreichen die Zielsetzung des Wettbewerbs.

Literaturverzeichnis

Bauer, Hannspeter. (1989). MC-Test und C-Test: Die Philosophie und die Korrelation. In Finkenstaedt & Schröder (1989), 5-16.

Baur, Rupprecht S. & Meder, Gregor. (1993). C-Tests zur Ermittlung der globalen Sprachfähigkeit im Deutschen und in der Muttersprache bei ausländischen Schülern in der Bundesrepublik Deutschland. In Grotjahn (1993).

Bonheim, Helmut. (1983). Schrittweise zur Validität. In Thomas Finkenstaedt & Franz-Rudolf Weller (Hrsg.), *Der Schülerwettbewerb Fremdsprachen im Stifterverband für die Deutsche Wissenschaft* (S. 123-159). Augsburg: Universität (Augsburger I & I – Schriften Bd. 28).

Brandt, Horst, Hertel, Elke, Meiser, Siegfried & Schröder, Konrad. (1989). *Der Bundeswettbewerb Fremdsprachen. Gruppenwettbewerb und Einzelwettbewerb.* Bonn: Bildung und Begabung e.V.

Carroll, John B. (1971). Development of native language skills beyond the early years. In Carroll E. Reed (Hrsg.), *The learning of language* (S. 97-156). New York: Appleton Century Crofts.

Diehl, Joerg M. & Kohr, Heinz U. (1991), *Deskriptive Statistik* (9. Aufl.). Eschborn: Klotz.

Doyé, Peter. (1989). Die Aufgabenstellungen des Einzelwettbewerbs. Allgemeine Problematik. In Finkenstaedt & Schröder (1989), 30-40.

Feldmann, Ute, Grotjahn, Rüdiger & Stemmer, Brigitte. (1986). Was messen Sprachtests eigentlich? Überlegungen zur introspektiven Validierung von C-Tests. In Seminar für Sprachlehrforschung der Ruhr-Universität Bochum (Hrsg.), *Probleme und Perspektiven der Sprachlehrforschung. Bochumer Beiträge zum Fremdsprachenunterricht in Forschung und Lehre* (S. 325-338). Frankfurt/M.: Scriptor.

Finkenstaedt, Thomas. (1989). Bemerkungen zur Statistik des Bundeswettbewerbs Fremdsprachen. In Finkenstaedt & Schröder (1989), 168-177.

Finkenstaedt, Thomas & Schröder, Konrad. (Hrsg.). (1989). *Zwischen Empirie und Machbarkeit. Erstes Symposium zum Bundeswettbewerb Fremsprachen.* Augsburg: Universität (Augsburger I & I – Schriften Bd. 50).

Finkenstaedt, Thomas & Weller, Franz-Rudolf. (Hrsg.). (1988). *Schrittweise zur Validität: Der Schülerwettbewerb im Stifterverband für die*

Deutsche Wissenschaft, 1979-1984. Augsburg: Universität (Augsburger I & I – Schriften Bd. 41).

Finn, Patrick J. (1977-78). Word frequency, information theory, and cloze performance: a transfer feature theory of processing in reading. *Reading Research Quarterly, 13*, 508-537.

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 (Hrsg.), *Taking their measure: The validity and validation of language tests* (S. 219-253). Bochum: Brockmeyer.

Grotjahn, Rüdiger. (1989). Der C-Test im Bundeswettbewerb Fremdsprachen: Eignung und Probleme (Französisch). In Finkenstaedt & Schröder (1989), 41-56.

Grotjahn, Rüdiger. (1992a). Der C-Test im Französischen. Quantitative Analysen. In Grotjahn (1992b), 205-255.

Grotjahn, Rüdiger. (Hrsg.). (1992b). *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Bd. 1). Bochum: Brockmeyer.

Grotjahn, Rüdiger. (Hrsg.). (1993). *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Bd. 2). Bochum: Brockmeyer.

Harris, David P. (1969). *Testing English as a second language*. New York: McGraw-Hill.

Herbst, Thomas. (1989). Bemerkungen zum Hörverstehenstest Englisch im Bundeswettbewerb Fremdsprachen Sekundarstufe I in den Jahren 1986 – 1988. In Finkenstaedt & Schröder (1989), 149-167.

Hood, Mary Ann G. (1990). The C-Test: A viable alternative to the use of the cloze procedure in testing? In Louis A. Arena (Hrsg.), *Language proficiency: defining, teaching, testing* (S. 173-189). New York: Plenum.

Kielhöfer, Bernd. (1989). Über die Schwierigkeiten, eine Geschichte zu schreiben. Die semi-kreative Aufgabenstellung Französisch. In Finkenstaedt & Schröder (1989), 57-69.

Kirkwood, Kristian J., Wolfe, Richard G., Maynes, Florence, Millar, John, Sword, Karen & Sword, Marjory. (1980). *Matching students and reading materials. A cloze-procedure method for assessing the reading ability of students and the readability of textual materials*. Toronto: OISE.

Klein-Braley, Christine. (1984). Advance prediction of difficulty with C-Tests. 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 1983* (S. 97-112). Colchester: University of Essex, Dept. of Language and Linguistics.

Klein-Braley, Christine. (1985a). Advance prediction of test difficulty. In Klein-Braley & Raatz (1985), 23-41.

Klein-Braley, Christine. (1985b). Reduced redundancy as an approach to language testing. In Klein-Braley & Raatz (1985), 1-13.

Klein-Braley, Christine. (in Vorbereitung). *Readability and the C-Test* [Arbeitstitel]. Habilitationsschrift Universität Duisburg.

Klein-Braley, Christine & Raatz, Ulrich. (Hrsg.). (1985). *Fremdsprachen und Hochschule 13/14: Thematischer Teil: C-Tests in der Praxis*. Bochum: AKS.

Lado, Robert. (1961). *Language testing*. London: Longman.

Legenhausen, Lienhard. (1988). Fehler-Fuzziness und Bewertungsvarianz. In Finkenstaedt & Weller (1988), 211-233.

Legenhausen, Lienhard. (1989). Zur face validity der C-Tests: Lehrer- und Schülerurteile. In Finkenstaedt & Schröder (1989), 70-81.

Münzel, Helmut. (1989). Wie erleben Schüler den Klausurtag im Einzelwettbewerb? In Finkenstaedt & Schröder (1989), 103-111.

Mundzeck, Fritz. (1983). Das Korrekturverfahren. Theorie und Praxis im Kontext heutiger Diskussion um Prüfungen und Korrigieren. In Thomas Finkenstaedt & Franz-Rudolf Weller (Hrsg.), *Der Schülerwettbewerb Fremdsprachen im Stifterverband für die Deutsche Wissenschaft* (S. 161-212). Augsburg: Universität.

Mundzeck, Fritz. (1989). Gutachtermaßstäbe und Begutachtungskriterien des semi-kreativen Teils des Einzelwettbewerbs. In Finkenstaedt & Schröder (1989), 112-126.

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.

Raatz, Ulrich & Klein-Braley, Christine. (1985). How to develop a C-Test. In Klein-Braley & Raatz (1985), 20-22.

Raatz, Ulrich & Klein-Braley, Christine. (1986). Nachanalyse der Ergebnisse im Einzelwettbewerb Sek. I vom Herbst 1985 [Gutachten im Auftrag des *Bundeswettbewerb Fremdsprachen*] (Ms.).

Raatz, Ulrich & Klein-Braley, Christine. (1989). Der C-Test im Bundeswettbewerb Fremdsprachen: Eignung und Probleme (Englisch). In Finkenstaedt & Schröder (1989), 127-134.
Raatz, Ulrich & Klein-Braley, Christine. (1993). Analyse der Ergebnisse im Einzelwettbewerb des Bundeswettbewerb Fremdsprachen: Wettbewerb Sekundarstufe I Frühjahr 1991. In Grotjahn (1993).
Rump, Marina. (1985). C-Tests für Anfänger im Englischunterricht. In Klein-Braley & Raatz (1985), 128-129.
Schröder, Konrad. (Hrsg.) (1988). Fremdsprachenwettbewerbe [Themenheft]. *Die Neueren Sprachen, 87*(3).
Schröder, Konrad & Stütz, Wolfgang. (Hrsg.) (1988). *Der Bundeswettbewerb Fremdsprachen. Aufgaben, Lösungen, Kommentare aus den Jahren 1985 bis 1987.* Berlin: Cornelsen.
Süßmilch, Edgar. (1985). C-Tests für ausländische Schüler: Sprachdiagnose im Unterricht Deutsch als Zweitsprache. In Klein-Braley & Raatz (1985), 72-82.
Stütz, Wolfgang. (1988). Der Mehrsprachenwettbewerb der Sekundarstufe II im Bundeswettbewerb Fremdsprachen. *Die Neueren Sprachen, 87*, 289-295.
Taylor, H.C. & Russell, J.T. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection: discussion and tables. *Journal of Applied Psychology, 23*, 565-578.

Grotjahn, Rüdiger. (Hrsg.). (1992). Der C-Test. Theoretische Grundlagen und praktische Anwendungen (Bd. 1, S. 297–311). Bochum: Brockmeyer

Gerhard Jakschik

Zum Einsatz des C-Tests in den Psychologischen Diensten der Arbeitsämter. Ein C-Test für Deutsch als Zweitsprache

Die Psychologischen Dienste der Arbeitsämter betreiben Messungen im Bereich Deutsch als Zweitsprache, um die Arbeitsberater/Berufsberater in ihrer Arbeit zu unterstützen. Die eingesetzten Verfahren sind für diesen Zweck nur bedingt tauglich: Sie erfassen nur sprachliche Teilkompetenzen, haben nur wenig gemein mit alltagssprachlicher Kommunikation und sind vor allem zu schwierig. Im vorliegenden Aufsatz wird als notwendige Ergänzung ein C-Test für erwachsene Zweitsprachler diskutiert und dessen Entwicklung und Erprobung beschrieben. Die Ergebnisse sind vielversprechend: Der vorgelegte C-Test erweist sich als ökonomisch, objektiv und reliabel und liefert einen Globalindex für die Beherrschung der deutschen Sprache. Weitere empirische Untersuchungen sind notwendig, damit der C-Test als Standardverfahren im Arbeitsamts-Alltag eingesetzt werden kann.

1. Einführung

Die Frage, inwieweit die Deutschkenntnisse einer/eines Ratsuchenden für berufliche Bildungsmaßnahmen ausreichen – und falls ja: für welchen Berufsbereich und auf welchem Niveau? – ist nicht neu. Neu ist aber die Bedeutung, die diese Frage durch die gestiegene Zahl der anerkannten Asylbewerber und insbesondere der Aussiedler in den letzten Jahren gewonnen hat.

Die Kolleginnen und Kollegen der Berufsberatung und der Arbeitsvermittlung/Arbeitsberatung sind verständlicherweise häufig recht hilflos, wenn es darum geht, den Grad der Beherrschung der deutschen Sprache bei Ratsuchenden einzuschätzen. Und selbst wenn sie sich eine ungefähre Meinung zum Sprachstand der Ratsuchenden gebildet haben, stehen sie vor dem Problem, beurteilen zu müssen, ob die mutmaßlich vorhandenen Deutschkenntnisse für eine bestimmte anvisierte Bildungsmaßnahme hinreichend sind.

In dieser schwierigen Situation wenden sich viele Beraterinnen und Berater an den Psychologischen Dienst, im Vertrauen (bzw. in der Hoffnung) darauf, daß die Fachleute dort über ein geeignetes Instrumentarium verfügen,