

Ulrich Raatz und Christine Klein-Braley

Analyse der Ergebnisse im Einzelwettbewerb des Bundeswettbewerbs Fremdsprachen: Wettbewerb Sekundarstufe I Frühjahr 1991*

Der C-Test wird im Rahmen des Bundeswettbewerbs Fremdsprachen im Einzelwettbewerb Sekundarstufe I zusammen mit anderen Testverfahren eingesetzt. Außerdem dient er als Screening-Test und ist deshalb das einzige Verfahren, welches für alle Teilnehmer ausgewertet wird. Beim Wettbewerb 1991 wurden die Ergebnisse von 1283 Schülern in verschiedenen Wettbewerbssprachen genauer analysiert.

Folgende Ergebnisse waren dabei von besonderer Bedeutung: (1) Am Wettbewerb nahmen 70% Mädchen und 30% Jungen teil; (2) 77% der Wettbewerbsteilnehmer stammen aus Gymnasien; (3) 11 Wettbewerbssprachen wurden gewählt, und zwar an erster Stelle Englisch mit 63% und an zweiter Stelle Französisch mit 26%; (4) obwohl die C-Tests für die Gesamtgruppe in allen untersuchten Sprachen etwas zu leicht ausfielen, waren sie hoch reliabel und repräsentierten das Gesamtkonstrukt "Sprachbeherrschung" im Vergleich zu den anderen verwendeten Tests am besten; (5) Geschlechtsunterschiede in der Sprachbeherrschung waren nicht nachzuweisen; (6) in den Sprachen Englisch und Französisch ergaben sich die erwarteten Unterschiede zwischen den Schulformen in der Reihenfolge Gymnasium, Gesamtschule, Realschule, Hauptschule; (7) Schüler, bei denen ein Elternteil oder beide Elternteile die Wettbewerbssprache als Muttersprache sprechen, schnitten erwartungsgemäß besser ab.

1. Zum Auftrag

Im Sommer 1991 erhielt die Arbeitsgruppe C-Test an der Universität Duisburg von der Geschäftsstelle des Bundeswettbewerbs Fremdsprachen den Auftrag, die Ergebnisse des Einzelwettbewerbs Sekundarstufe I vom Frühjahr 1991 zu analysieren. Dabei sollten neben den Testgütekriterien und der Struktur der verwendeten Tests insbesondere mögliche Geschlechtsunterschiede untersucht und auf sonstige Auffälligkeiten bei den Ergebnissen geachtet werden.

* Dieser Text ist die veränderte Fassung eines Gutachtens für den Prüfungsausschuß Sekundarstufe I, Bundeswettbewerb Fremdsprachen. Der Abdruck erfolgt mit Genehmigung der Geschäftsstelle. Genauere Hinweise und Literaturangaben zum Bundeswettbewerb Fremdsprachen finden sich in Grotjahn, Klein-Braley & Raatz (1992).

2. Datensammlung

Nach der Auswertung der Tests in den einzelnen Bundesländern wurde das gesamte Material von den Länderbeauftragten an die Geschäftsstelle in Bonn geschickt, die es ohne weitere Bearbeitung nach Duisburg weiterleitete. Einzelne verspätete Sendungen gingen direkt nach Duisburg.

Die vollständigen Unterlagen eines jeden Wettbewerbsteilnehmers bestanden aus:

- einem Blatt mit persönlichen Daten
- einem Ergebnisblatt
- einem Auswertungsblatt der semi-kreativen Aufgabe
- dem Text der semi-kreativen Aufgabe
- dem C-Test
- dem Ergebnisblatt der mündlichen Produktion
- dem Antwortbogen des Hörverständnistests.

Leider fehlten in vielen Fällen Teile dieses Materials, z.B. einzelne Tests oder das Blatt mit den persönlichen Daten. Außerdem waren die komplexeren Tests zumeist nicht ausgewertet worden, wenn das Ergebnis im C-Test sehr niedrig war. Wegen dieser Unvollständigkeit der Daten sind die Stichprobenumfänge bei den Einzelanalysen unterschiedlich groß.

Die Datensammlung wurde von uns Mitte August beendet. Später eingehende Sendungen konnten nicht mehr berücksichtigt werden.

3. Beschreibung der Stichprobe

Wir konnten die Testergebnisse von insgesamt $N = 1283$ Wettbewerbsteilnehmern auswerten. Tabelle 1 zeigt die Verteilung auf die einzelnen Bundesländer.

Tabelle 2 enthält die Anzahl der Jungen und Mädchen. Auffällig ist, daß dreimal so viel Mädchen wie Jungen an dem Einzelwettbewerb teilgenommen haben.

Tabelle 3 zeigt das Geburtsjahr der Wettbewerbsteilnehmer. Die meisten Teilnehmer sind zwischen 15 und 17 Jahre alt. Falls diese Untersuchung wiederholt wird, sollte nicht nach dem Geburtsjahr, sondern nach dem Alter der Teilnehmer gefragt werden, da diese Information aussagekräftiger ist.

Tabelle 1

Anzahl der Teilnehmer in den Bundesländern

| | |
|---------------------|-----|
| Bremen | 22 |
| Nordrhein-Westfalen | 140 |
| Niedersachsen | 206 |
| Hamburg | 165 |
| Schleswig-Holstein | 165 |
| Rheinland-Pfalz | 363 |
| Hessen | 220 |
| keine Angabe | 2 |

Tabelle 2

Anzahl der Jungen und Mädchen

| | |
|--------------|-----|
| männlich | 373 |
| weiblich | 903 |
| keine Angabe | 7 |

Tabelle 3

Geburtsjahr der Teilnehmer

| | |
|-----------------|-----|
| 1972 und früher | 6 |
| 1973 | 53 |
| 1974 | 363 |
| 1975 | 587 |
| 1976 | 203 |
| 1977 und später | 12 |
| keine Angabe | 59 |

Tabelle 4

Schulformen

| | |
|--------------|-----|
| Gymnasium | 982 |
| Realschule | 151 |
| Gesamtschule | 55 |
| Hauptschule | 51 |
| Berufsschule | 4 |
| keine Angabe | 40 |

Tabelle 5
Klassen

| | |
|--------------|-----|
| 6 | 3 |
| 7 | 1 |
| 8 | 24 |
| 9 | 388 |
| 10 | 827 |
| 11 | 1 |
| keine Angabe | 39 |

Tabelle 4 zeigt, daß die meisten Wettbewerbsteilnehmer aus Gymnasien stammten. Daß die Gesamtschule so schwach vertreten ist, mag u.a. daran liegen, daß ihre Anzahl im Vergleich zu Realschulen und Gymnasien relativ gering ist. Hauptschüler fühlen sich vermutlich durch den Wettbewerb überfordert.

Tabelle 5 enthält die Angaben zu den Schulklassen der Wettbewerbsteilnehmer. Fast alle stammen aus den Klassen 9 und 10, also aus den letzten beiden Klassen der Sekundarstufe I.

Tabelle 6
Wettbewerbssprachen

| | |
|----------------|-----|
| Englisch | 809 |
| Französisch | 338 |
| Russisch | 46 |
| Dänisch | 41 |
| Spanisch | 17 |
| Polnisch | 12 |
| Italienisch | 7 |
| Portugiesisch | 3 |
| Türkisch | 3 |
| Schwedisch | 2 |
| Serbokroatisch | 2 |

Im Wettbewerb 1991 wurden insgesamt 11 Wettbewerbssprachen gewählt, in der Hauptsache Englisch und Französisch (vgl. Tabelle 6). Für Dänisch interessierten sich ausschließlich Schüler aus Schleswig-Holstein, was verständlich ist, da dort Dänisch Schulfremdsprache ist.

Die gewählte Wettbewerbssprache ist in 820 Fällen erste, in 350 Fällen zweite und 53 mal dritte Schulfremdsprache. In 20 Fällen war die gewählte Sprache keine in der Schule unterrichtete Fremdsprache.

Es fällt auf, daß auch Wettbewerbssprachen gewählt wurden, die offensichtlich keine Schulfremdsprachen sind. In nächster Zeit ist sicherlich verstärkt mit solchen Fällen zu rechnen. Es wäre zu überlegen, ob das dem Sinn des Wettbewerbs entspricht, da durch ihn der Fremdsprachenunterricht in der Schule gefördert werden soll.

In dem Personalfragebogen wurden die Wettbewerbsteilnehmer nach der Sprache des Vaters und der Mutter gefragt. Bei der Sprache des Vaters wurden außer deutsch insgesamt 25, bei der Sprache der Mutter 29 verschiedene Sprachen genannt. Tabelle 7 enthält die entsprechenden Zahlen in zusammengefaßter Form.

In Anbetracht der beschriebenen Situation wäre zu diskutieren, ob nicht ein eigener Wettbewerb für bilinguale Schüler und Schülerinnen eingerichtet werden sollte.

Tabelle 7
Sprache des Vaters und der Mutter

| | deutsch | nicht deutsch | keine Angabe |
|--------------------|---------|---------------|--------------|
| Sprache des Vaters | 1022 | 107 | 154 |
| Sprache der Mutter | 987 | 142 | 154 |

4. Kennwerte der verwendeten Testverfahren

Es konnten nur die Tests der vier am häufigsten gewählten Wettbewerbssprachen Englisch, Französisch, Dänisch und Russisch analysiert werden, da bei den anderen Sprachen die Stichprobenumfänge zu klein waren. Der Gesamtwert (Ges) ist die gewichtete Summe der vier verwendeten Tests. Er kann maximal 100 Punkte betragen. Der Anteil der vier Einzeltests am Gesamtwert beträgt: semi-kreative Aufgabe (SM) und mündliche Produktion (MP) jeweils maximal 30 Punkte, Hörverstehenstest (HV) und C-Test

(CT) jeweils maximal 20 Punkte. In den folgenden Tabellen wurden außer beim C-Test direkt die gewichteten Werte verwendet.

4.1 Kennwerte der einzelnen Tests und des Gesamtergebnisses

In den Tabellen 8 bis 11 sind für die vier am häufigsten gewählten Wettbewerbsprachen Englisch, Französisch, Russisch und Dänisch für die einzelnen Tests und das Gesamtergebnis Mittelwert (M), Standardabweichung (s) und der Stichprobenumfang (N) sowie zum Vergleich der maximal mögliche Punktwert (Max) angegeben.

Tabelle 8
Kennwerte der einzelnen Tests in Englisch

| Test | Max | M | s | N |
|------|-----|------|------|-----|
| SM | 30 | 18.4 | 5.2 | 560 |
| HV | 20 | 11.2 | 2.8 | 671 |
| MP | 30 | 18.7 | 6.3 | 430 |
| CT | 100 | 70.9 | 14.3 | 809 |
| Ges | 100 | 66.8 | 11.8 | 372 |

Tabelle 9
Kennwerte der einzelnen Tests in Französisch

| Test | Max | M | s | N |
|------|-----|------|------|-----|
| SM | 30 | 19.2 | 4.8 | 230 |
| HV | 20 | 13.2 | 3.2 | 321 |
| MP | 30 | 17.0 | 7.1 | 201 |
| CT | 100 | 74.1 | 12.9 | 338 |
| Ges | 100 | 67.2 | 12.4 | 175 |

Bei Durchsicht der Tabellen 8 bis 11 zeigt sich folgendes:

- a) In allen vier analysierten Sprachen ist der C-Test, der ja bei allen Schülern verwendet wurde, zu leicht. Das ist in Anbetracht der Entscheidungsstrategie des Wettbewerbsverfahrens ungünstig. Ein Test, der als Screening-Test zur Ausscheidung der schlechteren Wettbewerbsteilnehmer dienen soll, müßte eine mittlere Schwierigkeit haben, da er dann

Tabelle 10
Kennwerte der einzelnen Tests in Dänisch

| Test | Max | M | s | N |
|------|-----|------|------|----|
| SM | 30 | 15.2 | 4.8 | 32 |
| HV | 20 | 13.5 | 1.6 | 41 |
| MP | 30 | 12.7 | 5.6 | 41 |
| CT | 100 | 79.5 | 8.4 | 41 |
| Ges | 100 | 59.7 | 10.2 | 32 |

Tabelle 11
Kennwerte der einzelnen Tests in Russisch

| Test | Max | M | s | N |
|------|-----|------|------|----|
| SM | 30 | 22.8 | 5.2 | 44 |
| HV | 20 | 14.1 | 3.7 | 45 |
| MP | 30 | 19.6 | 7.3 | 41 |
| CT | 100 | 75.0 | 22.6 | 45 |
| Ges | 100 | 70.2 | 18.0 | 41 |

sowohl bei den guten als auch bei den schlechten Schülern gleich gut differenziert und außerdem eine höhere Meßgenauigkeit hat.

- b) Auch die anderen drei Testverfahren sind in allen vier Wettbewerbsprachen leicht. Dieses Ergebnis ist jedoch darauf zurückzuführen, daß die Tests der schlechteren Wettbewerbsteilnehmer nicht ausgewertet wurden. Es sollte deshalb nicht weiter interpretiert werden.
- c) In allen vier Sprachen hat der Hörverstehenstest eine zu geringe Standardabweichung. Dadurch werden die Korrelationen dieses Verfahrens mit den anderen Tests und mit dem Gesamtpunktwert herabgesetzt, wie aus den Tabellen 12 bis 15 ersichtlich wird. Das läßt vermuten, daß auch die Reliabilität dieses Tests vergleichsweise niedrig ist.
- d) Auch der dänische C-Test hat eine vergleichsweise zu niedrige Streuung. Hier ist ebenfalls ein Einfluß auf die Reliabilität des Tests zu erwarten.

4.2 Interkorrelationen

In den Tabellen 12 bis 15 werden die Interkorrelationen der vier verwendeten Tests und des Gesamtwertes angegeben. Da auch hier nur vollständige Datensätze berücksichtigt werden konnten, ist der Stichprobenumfang in allen Fällen niedriger als die Gesamtzahl der Testteilnehmer in der jeweiligen Sprache. Deshalb müssen diese Korrelationen mit großer Vorsicht interpretiert werden.

Tabelle 12
Interkorrelationen der Tests in Englisch (N = 304)

| | HV | MP | CT | Ges |
|----|------|------|------|------|
| SM | 0.29 | 0.46 | 0.51 | 0.79 |
| HV | | 0.34 | 0.42 | 0.59 |
| MP | | | 0.44 | 0.64 |
| CT | | | | 0.71 |

Tabelle 13
Interkorrelationen der Tests in Französisch (N = 159)

| | HV | MP | CT | Ges |
|----|------|------|------|------|
| SM | 0.07 | 0.42 | 0.37 | 0.69 |
| HV | | 0.42 | 0.47 | 0.55 |
| MP | | | 0.63 | 0.90 |
| CT | | | | 0.75 |

Tabelle 14
Interkorrelationen der Tests in Dänisch (N = 29)

| | HV | MP | CT | Ges |
|----|------|------|------|------|
| SM | 0.64 | 0.24 | 0.63 | 0.80 |
| HV | | 0.41 | 0.45 | 0.75 |
| MP | | | 0.14 | 0.76 |
| CT | | | | 0.59 |

Tabelle 15
Interkorrelationen der Tests in Russisch (N = 41)

| | HV | MP | CT | Ges |
|----|------|------|------|------|
| SM | 0.54 | 0.78 | 0.75 | 0.89 |
| HV | | 0.58 | 0.68 | 0.75 |
| MP | | | 0.79 | 0.93 |
| CT | | | | 0.91 |

Insgesamt gesehen sind diese Korrelationen niedriger als man sie bei Sprachtests allgemein vorfindet. Eine Erklärung dafür ist die Tatsache, daß die Stichproben durch die Selektion der besseren Wettbewerbsteilnehmer für die Auswertung homogenisiert worden sind. Eine Ausnahme bildet die Sprache Russisch mit höheren Korrelationen. Hier sind allerdings alle Tests vollständig ausgewertet worden; eine Selektion hat also nicht stattgefunden. Die Korrelationen des Hörverstehenstests mit allen anderen Verfahren sind wegen der geringeren Streuung besonders in Englisch und Französisch deutlich erniedrigt.

4.3 Reliabilitäten

Die Reliabilitäten konnten nur für den semi-kreativen Test und den C-Test berechnet werden, da hier Teilpunktwerte gegeben waren. Verwendet wurde die Alpha-Formel von Cronbach. Es ergaben sich für die vier genannten Wettbewerbssprachen die in Tabelle 16 aufgeführten Werte.

Tabelle 16
Reliabilitäten

| | SM | CT |
|-------------|------|------|
| Englisch | 0.91 | 0.91 |
| Französisch | 0.88 | 0.89 |
| Dänisch | 0.85 | 0.81 |
| Russisch | 0.91 | 0.95 |

Die Reliabilitäten sind sowohl für den semi-kreativen Test als auch für den C-Test in allen vier untersuchten Sprachen sehr hoch. Allerdings müssen diese Ergebnisse für die beiden Tests unterschiedlich interpretiert werden.

Beim C-Test sollen alle Teile dasselbe messen. Wenn diese Forderung erfüllt ist, müssen die Interkorrelationen und damit auch die Reliabilität hoch sein. Die in Tabelle 16 dargestellten Ergebnisse für den C-Test entsprechen also den Erwartungen. Beim semi-kreativen Test sollten verschiedene Aspekte der Sprachleistung, nämlich Originalität, Folgerichtigkeit, Themenbezug, Wortschatz, Idiomatik und Sprachrichtigkeit, beurteilt werden. Wenn es sich dabei wirklich um unterscheidbare Dimensionen der Sprache handelt, müßten die Interkorrelationen und damit auch die auf diese Weise berechnete Reliabilität deutlich niedriger sein. Das ist jedoch nicht der Fall. Deshalb lassen diese Ergebnisse nur die folgenden Schlüsse zu: Entweder kann eine sprachliche Leistung nicht in die genannten Dimensionen aufgeteilt werden, d.h. sie ist prinzipiell eindimensional. Oder die Beurteiler unterliegen einem starken Halo-Effekt, d.h. sie sind nicht in der Lage, die einzelnen Aspekte zu differenzieren. In beiden Fällen sollte man die Konsequenz ziehen, die Leistung im semi-kreativen Test in Zukunft nur noch global auf einer einzigen Skala beurteilen zu lassen. Dieser Vorschlag würde auch den Auswertungsaufwand stark reduzieren und damit zu einer größeren Ökonomie und Praktikabilität führen.

4.4 Faktorenanalysen

Die Faktorenanalyse ist eine Methode, die zur Konstruktvalidierung von Tests eingesetzt werden kann. Mit diesem Verfahren kann man überprüfen, ob das Konstrukt, das durch die verschiedenen Testverfahren operationalisiert wird, ein- oder mehrdimensional ist und wie hoch die Beiträge der einzelnen Tests (Ladungen) zu den verschiedenen Dimensionen (Faktoren) sind, wie gut also die verwendeten Tests das Gesamtmerkmal – im Falle nur eines Faktors – oder die verschiedenen Teilaspekte des Merkmals – im Falle von mehreren Faktoren – repräsentieren.

Bei der Analyse von Sprachtests ergibt sich im allgemeinen nur ein einziger allgemeiner Sprachfaktor ("General Language Proficiency"). Allerdings ergibt sich aus statistischen Gründen auch dann nur ein Faktor, wenn, wie in unserem Falle, nur vier Tests analysiert werden.

Die Tabelle 17 enthält für die vier analysierten Sprachen die Ladungen aller Tests in diesem allgemeinen Sprachfaktor. In Englisch, Französisch und Russisch zeigt sich, daß der C-Test mit seiner jeweils höchsten Ladung das Gesamtkonstrukt am besten repräsentiert. Eine Ausnahme bildet die Sprache Dänisch. Hier war der C-Test allerdings zu leicht, hatte eine zu niedrige Streuung und damit auch eine im Vergleich niedrigere Reliabilität.

Tabelle 17
Ladungen im ersten Faktor

| Sprache | SM | HV | MP | CT |
|-------------|------|------|------|------|
| Englisch | 0.71 | 0.67 | 0.74 | 0.81 |
| Französisch | 0.59 | 0.61 | 0.86 | 0.86 |
| Russisch | 0.88 | 0.79 | 0.90 | 0.92 |
| Dänisch | 0.88 | 0.84 | 0.57 | 0.80 |

4.5 Folgerungen aus den bisherigen Ergebnissen

Zusammenfassend können aus den bisherigen Ergebnissen folgende Schlüsse gezogen werden:

- a) Gegen die verwendeten Testverfahren ergeben sich nur geringfügige test-theoretische Einwände. So sollte in Zukunft bei der Entwicklung von C-Tests auf einen größeren Schwierigkeitsgrad geachtet werden. Diese Empfehlung bezieht sich insbesondere auf das Vorgehen der Textauswahlkommission, die in der Vergangenheit immer wieder die Neigung gezeigt hat, zu leichte Texte auszuwählen.
- b) Auch in dieser Untersuchung konnte wie schon früher gezeigt werden, daß die C-Tests hohe Reliabilitäten aufweisen und das Gesamtkonstrukt am besten erfassen. Dieses Ergebnis rechtfertigt noch einmal die eingeschlagene Strategie, C-Tests zur Vorselektion bei der Testauswertung einzusetzen. Es wäre zu überlegen, ob nicht die Gewichtung der einzelnen Tests zugunsten der C-Tests verändert werden sollte.
- c) Bei den dänischen C-Tests, insbesondere aber bei den Hörverstehentests in allen Sprachen müßte die Rohpunktstreuung größer sein. Das könnte man dadurch erreichen, daß man bei diesen Tests den Schwierigkeitsgrad der Items stärker abstuft. Die Ursache für dieses Problem und mögliche Auswege sollten in einer weiteren empirischen Untersuchung

analysiert bzw. diskutiert werden. Eine solche Untersuchung könnte nützliche Hinweise für die künftige Testentwicklung geben.

5. Geschlechtsunterschiede

Die Hauptfragestellung dieser Untersuchung war die Überprüfung von Geschlechtsunterschieden in den einzelnen Sprachen und in den verschiedenen Tests. Die Tabellen 18 bis 21 enthalten die Mittelwerte der Jungen und Mädchen sowie die jeweiligen Stichprobenumfänge.

Tabelle 18
Geschlechtsunterschiede in Englisch

| Test | Jungen | | Mädchen | |
|--------|--------|-----|---------|-----|
| | M | N | M | N |
| SM | 18.3 | 176 | 18.3 | 380 |
| HV | 11.2 | 217 | 11.2 | 450 |
| MP | 17.7 | 139 | 19.2 | 287 |
| CT | 71.1 | 262 | 70.8 | 543 |
| Gesamt | 65.8 | 117 | 67.1 | 251 |

Tabelle 19
Geschlechtsunterschiede in Französisch

| Test | Jungen | | Mädchen | |
|--------|--------|----|---------|-----|
| | M | N | M | N |
| SM* | 17.6 | 51 | 19.6 | 179 |
| HV | 13.6 | 71 | 13.0 | 250 |
| MP | 16.6 | 46 | 17.1 | 155 |
| CT | 75.2 | 75 | 70.8 | 263 |
| Gesamt | 66.4 | 37 | 67.4 | 138 |

* Unterschied statistisch signifikant: $p < .05$ (t -Test)

Faßt man die Ergebnisse in allen vier Sprachen zusammen, so zeigt sich, daß bei den einzelnen Tests siebenmal die Jungen und sechsmal die Mädchen ein höheres Ergebnis aufweisen. Dreimal sind die Leistungen gleich. Beim

Tabelle 20
Geschlechtsunterschiede in Dänisch

| Test | Jungen | | Mädchen | |
|--------|--------|---|---------|----|
| | M | N | M | N |
| SM | 15.2 | 5 | 15.0 | 26 |
| HV | 13.5 | 6 | 13.6 | 32 |
| MP | 12.7 | 6 | 12.7 | 32 |
| CT | 80.3 | 6 | 80.1 | 32 |
| Gesamt | 58.5 | 5 | 59.5 | 26 |

Tabelle 21
Geschlechtsunterschiede in Russisch

| Test | Jungen | | Mädchen | |
|--------|--------|----|---------|----|
| | M | N | M | N |
| SM | 22.0 | 15 | 23.3 | 29 |
| HV | 14.2 | 16 | 14.1 | 29 |
| MP | 21.0 | 13 | 18.9 | 28 |
| CT | 76.2 | 16 | 74.3 | 29 |
| Gesamt | 69.6 | 13 | 70.5 | 28 |

Gesamtpunktwert schneiden die Mädchen in allen vier Sprachen etwas besser ab.

Eine Signifikanzüberprüfung dieser Unterschiede ergab allerdings, daß mit Ausnahme des semi-kreativen Tests in Französisch kein einziger Mittelwertsunterschied signifikant und damit interpretierbar ist.

Beim ersten Blick widerspricht dieses Ergebnis den vorliegenden Ergebnissen in der Forschungsliteratur, nach denen gerade in dieser Altersstufe Mädchen bei Sprachtests deutlich besser abschneiden als Jungen. Allerdings gelten diese Aussagen nur für Zufallsstichproben aus der Gesamtpopulation. Die Stichproben von Jungen und Mädchen, die am Bundeswettbewerb teilgenommen haben, sind jedoch selbstselektiert und somit allein für den Bundeswettbewerb repräsentativ. Der geringe Anteil der Jungen (ca. 25%) läßt vermuten, daß sich nur die besonders sprachbegabten Schüler zum Wettbewerb melden, so daß die erwarteten Unterschiede zwischen Jungen und Mädchen verwischt werden.

6. Unterschiede zwischen den Schulformen

Bei der Entwicklung und Evaluation des Gesamtkonzepts des Wettbewerbs wurde immer wieder die Frage gestellt, ob Schüler aus verschiedenen Schulformen dieselbe Chance hätten, einen Preis zu erhalten, oder ob nicht z.B. Schüler aus Hauptschulen überfordert wären. Es wurde sogar diskutiert, ob man nicht verschiedene Wettbewerbe für die verschiedenen Schulformen veranstalten sollte. Aus diesem Grunde haben wir die Ergebnisse der unterschiedlichen Schulformen miteinander verglichen. Dabei konnten wir nur die Wettbewerbssprachen Englisch und Französisch berücksichtigen, da nur für diese beiden Sprachen Teilnehmer aus allen bzw. fast allen Schulformen gemeldet wurden.

Wir haben nicht nur das Gesamtergebnis analysiert, sondern auch die Ergebnisse der Einzeltests, da wir vermuteten, daß es Wechselwirkungen zwischen Schulform und spezieller Sprachleistung geben könnte.

Die Tabellen 22 und 23 enthalten Mittelwerte und Stichprobenumfänge. Die meisten aufgefundenen Unterschiede sind hoch signifikant ($p < .001$).

Tabelle 22
Unterschiede zwischen den Schulformen in Englisch

| Test | Gymnasium | | Realschule | | Hauptschule | | Gesamtschule | |
|--------|-----------|-----|------------|-----|-------------|----|--------------|----|
| | M | N | M | N | M | N | M | N |
| SM*** | 19.0 | 431 | 15.1 | 69 | 15.9 | 17 | 17.0 | 30 |
| HV*** | 11.5 | 527 | 10.4 | 68 | 9.0 | 33 | 11.6 | 30 |
| MP*** | 19.6 | 337 | 14.6 | 48 | 12.7 | 22 | 21.0 | 18 |
| CT*** | 74.0 | 614 | 60.3 | 102 | 50.2 | 48 | 76.5 | 30 |
| Ges*** | 68.0 | 306 | 58.7 | 32 | 54.4 | 11 | 66.4 | 18 |

*** Unterschied statistisch signifikant: $p < .001$ (Varianzanalyse)

Bei einer Analyse der Tabellen 22 und 23 ergibt sich:

- a) Geht man vom C-Test aus, der ja für alle Wettbewerbsteilnehmer ausgewertet worden ist, so zeigt sich in Englisch eine deutliche Reihenfolge von der Hauptschule über die Realschule und das Gymnasium bis zur Gesamtschule mit einer Gesamtdifferenz von 26.3 Punkten. Allerdings liegen Gymnasium und Gesamtschule in ihrer Leistung sehr eng zusammen. In Französisch ist die Reihenfolge Realschule, Gesamtschule, Gymnasium mit einer Differenz von 18.5 Punkten.

Tabelle 23

Unterschiede zwischen den Schulformen in Französisch

| Test | Gymnasium | | Realschule | | Gesamtschule | |
|-------|-----------|-----|------------|----|--------------|----|
| | M | N | M | N | M | N |
| SM | 19.3 | 214 | 16.6 | 5 | 18.4 | 7 |
| HV** | 13.4 | 280 | 12.8 | 13 | 11.1 | 20 |
| MP | 17.1 | 189 | 13.7 | 4 | 15.3 | 6 |
| CT*** | 75.5 | 296 | 57.0 | 14 | 69.0 | 20 |
| Ges | 67.7 | 164 | 56.8 | 4 | 60.8 | 5 |

** Unterschied statistisch signifikant: $p < .01$

*** Unterschied statistisch signifikant: $p < .001$ (Varianzanalyse)

- b) Die Ergebnisse der drei anderen Tests, sowie das Gesamtergebnis, wurden anhand einer reduzierten Stichprobe ermittelt, die nur noch die Ergebnisse der besseren Schüler enthält. Das ging in Englisch insbesondere zu Lasten der schlechten Hauptschüler. Erwartungsgemäß nähern sich jetzt in Englisch bei diesen Tests und dem Gesamtwert die Leistungen der Hauptschüler denen der Realschüler.
- c) In Englisch ergeben sich in den drei Tests semi-kreative Aufgabe, Hörverstehenstest und mündliche Produktion sowie im Gesamtergebnis zwei Gruppen: Gymnasium und Gesamtschule einerseits und Hauptschule und Realschule andererseits.
- d) In Französisch gilt die Reihenfolge Gymnasium, Gesamtschule, Realschule außer in der mündlichen Produktion, wo die Realschüler tendenziell besser als die Gesamtschüler abschneiden.

Die großen Leistungsunterschiede zwischen den Schulformen konnten also in dieser Untersuchung für alle verwendeten Testverfahren bestätigt werden. Vor dem Hintergrund von häufig geäußerten Vorurteilen gegen die Gesamtschule ist allerdings die hohe Leistung der Wettbewerbsteilnehmer aus diesen Schulen bemerkenswert.

7. Unterschiede zwischen den Bundesländern

Abschließend sollen noch Unterschiede zwischen den einzelnen Bundesländern angegeben werden. Dabei konnten nur die Sprachen Englisch und Französisch berücksichtigt werden. Tabelle 24 enthält die durchschnittlichen Gesamtpunktwerte sowie die jeweiligen Stichprobenumfänge.

Mittelwertsunterschiede zwischen den einzelnen Bundesländern können sicher nicht auf unterschiedliche Schulsysteme und Curricula oder die unterschiedliche Qualität des Fremdsprachenunterrichts zurückgeführt werden. Hiergegen spricht der teilweise zu geringe Stichprobenumfang sowie die Selbstselektion der Wettbewerbsteilnehmer.

Tabelle 24
Unterschiede zwischen den Bundesländern im Gesamtergebnis in Englisch und Französisch

| Land | Englisch | | Französisch | |
|---------------------|----------|-----|-------------|----|
| | M | N | M | N |
| Bremen | 70.9 | 5 | 73.0 | 1 |
| Nordrhein-Westfalen | 66.1 | 97 | 62.7 | 7 |
| Niedersachsen | 65.5 | 115 | 67.6 | 45 |
| Hamburg | 63.9 | 91 | 67.8 | 28 |
| Schleswig-Holstein | 65.7 | 50 | 61.7 | 32 |
| Rheinland-Pfalz | 70.0 | 96 | 71.9 | 48 |
| Hessen | 70.3 | 105 | 64.7 | 54 |

Außerdem besteht Grund zu der Annahme, daß die Durchschnittsleistung in einem Bundesland dann niedriger wird, wenn mehr Schüler – also nicht nur die allerbesten – am Wettbewerb teilnehmen und wenn z.B. der Anteil an Hauptschülern oder Schülern aus niedrigeren Klassenstufen höher ist.

Wir berichten diese Daten nur der Vollständigkeit halber, warnen aber vor einer allzu extensiven Interpretation.

Zum Einfluß der Mutter-/Vatersprache

In den Ausschußdiskussionen des Bundeswettbewerbs Fremdsprachen in den letzten Jahren wurde ab und zu die Vermutung geäußert, daß diejenigen Wettbewerbsteilnehmer bevorzugt wären, bei denen die Wettbewerbssprache die Mutter- oder Vatersprache wäre. Es dürfte wahrscheinlich relativ selten vorkommen, daß die Wettbewerbssprache der Mutter- oder Vatersprache entspricht. Der Effekt auf die Wettbewerbsleistung sollte aber trotzdem überprüft werden.

Wir haben für einige Sprachen diejenigen Wettbewerbsteilnehmer herausgesucht, bei denen die Wettbewerbssprache Mutter- oder Vatersprache war, und die Leistung im C-Test sowie die durchschnittliche Gesamtleistung mit der der anderen Wettbewerbsteilnehmer verglichen. Die Ergebnisse stehen in den Tabellen 25 bis 28.

Tabelle 25
Einfluß der Muttersprache auf das C-Test-Ergebnis

| | Muttersprache gleich | | Muttersprache ungleich | |
|---------------|----------------------|----|------------------------|-----|
| | Wettbewerbssprache | | Wettbewerbssprache | |
| | M | N | M | N |
| Englisch*** | 84.6 | 20 | 70.2 | 705 |
| Französisch** | 81.4 | 25 | 73.2 | 279 |
| Spanisch* | 89.4 | 5 | 72.0 | 10 |
| Dänisch* | 88.2 | 5 | 78.9 | 33 |

* Unterschied statistisch signifikant: $p < .05$

** Unterschied statistisch signifikant: $p < .01$

*** Unterschied statistisch signifikant: $p < .001$ (t -Test)

Es ergibt sich in allen Fällen ein teilweise sehr großer Vorteil dieser Schülerinnen und Schüler sowohl im Gesamtergebnis als auch speziell im C-Test. Der Unterschied ist trotz der kleinen Stichprobenumfänge in der Hälfte der Fälle statistisch signifikant. Leider konnten wir nicht überprüfen, ob diese Teilnehmer überproportional unter den Preisträgern zu finden sind, da uns die entsprechenden Informationen fehlten.

Prinzipiell wäre zu überlegen, ob Konsequenzen aus diesem Ergebnis gezogen werden sollten. Eventuell müßten solche Wettbewerbsteilnehmer beispielsweise bei der Preisvergabe als Sondergruppe behandelt werden.

Tabelle 26
Einfluß der Muttersprache auf das Gesamtergebnis

| | Muttersprache gleich | | Muttersprache ungleich | |
|--------------|-------------------------|----|---------------------------|-----|
| | Wettbewerbssprache | | Wettbewerbssprache | |
| | M | N | M | N |
| Englisch* | 81.7 | 15 | 66.3 | 305 |
| Französisch* | 78.3 | 16 | 65.6 | 132 |
| Spanisch | 79.4 | 4 | 67.2 | 5 |
| Dänisch* | 72.0 | 5 | 56.9 | 26 |

* Unterschied statistisch signifikant: $p < .05$ (t -Test)

Tabelle 27
Einfluß der Vatersprache auf das C-Test-Ergebnis

| | Vatersprache gleich | | Vatersprache ungleich | |
|-------------|------------------------|----|--------------------------|-----|
| | Wettbewerbssprache | | Wettbewerbssprache | |
| | M | N | M | N |
| Englisch | 77.5 | 15 | 70.5 | 710 |
| Französisch | 79.8 | 8 | 73.7 | 295 |
| Spanisch | 89.7 | 3 | 74.8 | 12 |
| Russisch | 96.0 | 1 | 76.0 | 25 |

Tabelle 28
Einfluß der Vatersprache auf das Gesamtergebnis

| | Vatersprache gleich | | Vatersprache ungleich | |
|-------------|------------------------|---|--------------------------|-----|
| | Wettbewerbssprache | | Wettbewerbssprache | |
| | M | N | M | N |
| Englisch* | 77.1 | 7 | 66.8 | 314 |
| Französisch | 75.4 | 5 | 66.6 | 142 |
| Spanisch | 79.4 | 2 | 70.7 | 7 |
| Russisch | 93.2 | 1 | 74.1 | 24 |

* Unterschied statistisch signifikant: $p < .05$ (t -Test)

9. Einige Schlußfolgerungen

Diese Untersuchung ist keine Untersuchung im streng wissenschaftlichen Sinne, sondern die Beschreibung der Daten, die sich bei einer selbst selektierten Stichprobe ergeben haben. Außerdem sind die Varianzen und Korrelationen in dieser Untersuchung systematisch erniedrigt, da die Tests der schlechteren Teilnehmer – mit Ausnahme des C-Tests, der als Screening-Test diente – nicht ausgewertet wurden. Aus diesem Grunde sollen nur einige vorsichtige Schlußfolgerungen gezogen werden, die den C-Test, aber auch den Wettbewerb insgesamt betreffen.

1. Der C-Test hat sich in fast allen untersuchten Sprachen als hochreliales und konstruktvalides Verfahren bewährt, das zu recht als Screening-Test eingesetzt werden kann. Seine Leistungsfähigkeit könnte noch gesteigert werden, wenn es gelingen würde, ihn in allen Sprachen etwas schwieriger zu machen. In der Praxis der Testkonstruktion ist es allerdings sehr schwer, diese Forderung zu realisieren, da erfahrungsgemäß sowohl die Testkonstruktoren als auch diejenigen, die die Tests für den Wettbewerb auswählen, C-Tests für schwerer halten, als sie in Wirklichkeit sind (vgl. auch Grotjahn, Klein-Braley & Raatz, 1992).
2. Der Hörverstehenstest ist zu leicht und hat wahrscheinlich deshalb von allen Verfahren die schlechtesten Kennwerte. Außerdem ist dieser Test technisch relativ aufwendig. Deshalb wäre es eigentlich zu empfehlen, auf dieses Verfahren im Rahmen des Wettbewerbs in Zukunft zu verzichten. Das wird allerdings aus Gründen der *face validity* nicht möglich sein.
3. Die differenzierte Auswertung des semi-kreativen Tests ist wegen der hohen Interkorrelationen der einzelnen Auswertungsdimensionen höchstwahrscheinlich unnötig. Ob der Grund dafür der Halo-Effekt bei den Beurteilern oder die prinzipielle Nichtaufteilbarkeit von Sprache ist, kann hier nicht entschieden werden. Die große Mühe, die in die praktische Auswertung dieser Testteile investiert wird, erhöht offensichtlich auch nur wieder die *face validity* dieses Tests. Es wäre zu überlegen, ob das Auswertungsverfahren nicht vereinfacht werden kann.
4. Der hohe Anteil der Mädchen am Wettbewerb ist bemerkenswert. In anderen eher technischen Wettbewerben überwiegen in der Regel die männlichen Teilnehmer. Das kann natürlich damit zusammenhängen, daß Mädchen sprachbegabter und motivierter sind und deshalb eher an

Wettbewerben dieser Art teilnehmen. Daß sie im Durchschnitt dann doch nicht besser sind als die Jungen, kann damit zusammenhängen, daß sich von den Jungen nur die besonders sprachbegabten beteiligt haben.

5. Wir halten es für bedenklich, daß die jeweilige Elternsprache einen so starken Einfluß auf die Ergebnisse im Wettbewerb hat, daß bestimmte Teilnehmer dadurch bevorzugt sind. Es ist keine Kunst, in Spanisch zu den besten zu gehören und eventuell einen Preis zu gewinnen, wenn zu Hause mit der Mutter Spanisch gesprochen wird. Es sollte überlegt werden, ob diese Schüler und Schülerinnen nicht aus dem Sekundarstufen I-Wettbewerb ausgeschlossen werden sollten. Entweder könnte für sie ein eigener Wettbewerb eingerichtet werden, oder sie sollten am Zwei-Sprachen-Wettbewerb der Sekundarstufe II teilnehmen.
6. Es sollte berücksichtigt werden, daß es sich bei dieser Untersuchung um eine typische Praxisstudie handelt, die auf einer selbstselektierten Stichprobe mit unvollständigen Daten basiert und bei der weder die Bedingungen noch die Testverfahren im Sinne einer Versuchsplanung manipuliert werden konnten. Diese Beschränkungen kommen in der Realität immer wieder vor, und man muß damit leben.

Literaturverzeichnis

- Grotjahn, Rüdiger, Klein-Braley, Christine, & Raatz, Ulrich. (1992). C-Tests in der praktischen Anwendung. Erfahrungen beim Bundeswettbewerb Fremdsprachen. In Rüdiger Grotjahn (Hrsg.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Bd. 1, S. 263-296). Bochum: Brockmeyer.