Language Testing and Evaluation 34

Rüdiger Grotjahn (Hrsg./ed.)

Der C-Test: Aktuelle Tendenzen

The C-Test: Current Trends

PETER LANG EDITION

Grotjahn, Rüdiger. (Ed.). (2014). Der C-Test: Aktuelle Tendenzen/ The C-Test: Current Trends. Frankfurt/M.: Lang.

Der C-Test: Aktuelle Tendenzen. Einleitung und Übersicht über den Band

Rüdiger Grotjahn*

This introductory chapter describes the contents and the structure of present anthology. To this end, it first outlines some current trends in C-Test research. Subsequently, it gives a general overview of the issues dealt with and the methods used. It then describes the contents of each chapter at some length, making an attempt to also establish relationships among the chapters and to provide the reader with some additional references to the pertinent literature.

1. Einleitung

In den letzten Jahren hat sich der C-Test weiter in der Testlandschaft etabliert (vgl. zum Folgenden auch die Hinweise in Grotjahn, 2010a). So hat es u.a. erfolgreiche Anwendungen des C-Test-Prinzips auf Sprachen wie Koreanisch, Polnisch, Griechisch, Niederländisch und Persisch gegeben (vgl. zum Persischen Baghaei, im vorliegenden Band). Weiterhin sind C-Tests in einer Reihe von Studien zum Bildungsmonitoring verwendet worden – so z.B. im Forschungsprojekt „Deutsch Englisch Schülerleistungen International – DESI", in der Hamburger Längsschnittstudie „Kompetenzen und Einstellungen von Schülerinnen und Schülern – KESS" oder im Rahmen einer großangelegten Bildungsstudie zu den fremdsprachlichen Kompetenzen von Schülern der Klasse 7 luxemburgischer Schulen (vgl. Reichert, Brunner & Martin, im vorliegenden Band). Auch in einer Reihe von Projekten zur Förderung von Deutsch als Zweitsprache bei Kindern, Jugendlichen und Erwachsenen sind C-Tests eingesetzt worden, darunter auch lernbehinderte Jugendliche (vgl. zur letztgenannten Gruppe Linnemann & Wilbert, im vorliegenden Band). Auch bei deutschen Muttersprachlern sind C-Tests verwendet worden, so z.B. von Wockenfuß (2009) zur „Diagnostik von Sprache und Intelligenz bei Jugendlichen und jungen Erwachsenen" (vgl. auch Wockenfuß & Raatz, im vorliegenden Band). Zudem ist eine Tendenz zu beobachten, verstärkt C-Tests für spezifische Diskursbereiche zu entwickeln. Schließlich werden in jüngerer Zeit zunehmend auch deutsche C-Tests mit Stammtilgung der linken Worthälfte zur ökonomischen Messung der Lesekompetenz im Deutschen eingesetzt (vgl. z.B. die Dissertation von

* **Korrespondenzadresse:** Prof. Dr. Rüdiger Grotjahn, Ruhr-Universität Bochum, Seminar für Sprachlehrforschung, D-44780 Bochum. E-mail: ruediger.grotjahn@rub.de.

Mashkovskaya, 2014 im Rahmen des Sprachkompetenzprojekts Diagnose & Förderung an der Universität Duisburg-Essen).

Auch die kommerzielle Vermarktung von C-Tests schreitet voran. So bietet der Cornelsen Verlag mittlerweile C-Tests als Diagnoseinstrument im Rahmen von schulischen Fördermaterialien für Deutsch als Zweitsprache an (vgl. Baur, Chlosta & Goggin, 2011, 2012). In der Schweiz werden vom Schulverlag plus C-Tests zur Evaluation von Fremdsprachenkompetenzen in Französisch und Englisch im 5. bis 9. Schuljahr auf der Plattform Lingualevel kostenpflichtig zur Verfügung gestellt (siehe http://www.lingualevel.ch/; Lenz & Studer, 2008). Weitere Beispiele für einen kommerziellen Einsatz sind folgende vom TestDaf-Institut bzw. der Gesellschaft für Akademische Studienvorbereitung und Testentwicklung (g.a.s.t) angebotenen online C-Tests: onDaF (Online-Einstufungstest Deutsch als Fremdsprache; siehe http://www.ondaf.de sowie Eckes, im vorliegenden Band), onSET (Online-Spracheinstufungstest für Englisch und weitere Sprachen; siehe http://www.onset.de) und onScreen (Teil des Tests für Ausländische Studierende – TestAS; http://www.testas.de/de/screening.htm).

Im Folgenden werde ich zunächst einen kurzen allgemeinen Überblick über die im vorliegenden Band behandelten Themen und die verwendeten Forschungsmethoden geben. Daran anschließend werde ich die einzelnen Beiträge relativ ausführlich beschreiben, Querverbindungen herstellen sowie einige ergänzende Literaturhinweise geben.

2. Der vorliegende Band

2.1. Allgemeiner Überblick

Im vorliegenden Band geht es in einem ersten Schwerpunkt wiederum um die zentrale Frage der Konstruktvalidität von C-Tests als Maß allgemeiner Sprachkompetenz – und zwar u.a. um folgende Aspekte: Eindimensionalität des C-Test-Konstrukts; Dimensionalität von C-Tests, die sowohl aus bildungs- als auch aus alltagssprachlichen Texten bestehen; Messäquivalenz von fremdsprachlichen C-Tests bei Probanden mit unterschiedlichen Muttersprachen; Gemeinsamkeiten sprachlicher Produktion und Rezeption; (reduzierte) Redundanz bei der Verarbeitung natürlicher Sprachen; inneres und privates Sprechen bei der Lösung von C-Tests; mentale Prozesse beim Lösen von C-Tests; konvergente und divergente Validität; Augenscheingültigkeit und Washback; Sensitivität von C-Tests bei der Messung von Lernzuwachs; Messäquivalenz von klassischem C-Test und computergestütztem Multiple-Choice C-Test; Beziehung von C-Test-Leistung, (verbaler) Intelligenz und Geschwindigkeit der Informationsverarbeitung; Beziehung zwischen C-Test-Leistung und Lebensalter bei Mutter-

sprachlern; Rolle des Mikro- und Makrokontextes bei der Rekonstruktion der C-Test-Lücken; Bestimmung optimaler Trennwerte für Kompetenzniveaus; psychometrische Merkmale von C-Tests zur Messung von Verstehenskompetenzen lernbehinderter deutscher Jugendlicher mit Deutsch als Mutter- oder Zweitsprache.

Darüber hinaus werden folgende Aspekte, die keinen unmittelbaren Bezug zur Frage der Konstruktvalidität aufweisen, thematisiert: Effektivität des onDaF als Screening-Verfahren für den Test Deutsch als Fremdsprache – TestDaF; C-Test-Ergebnisse als Eingangswerte für computer-adaptives Testen (CAT); Vorhersage der Schwierigkeit von C-Test-Texten; Anwendbarkeit des C-Test-Prinzips auf das Persische; Konstruktions- und Auswertungsspezifika bei spanischen C-Tests; spezifischer Wert schulischer C-Tests.

Zur Untersuchung der genannten Aspekte werden eine Vielzahl von Methoden und Designs verwendet: psychometrische Analysen auf der Basis der klassischen und probabilistischen Testtheorie unter Einschluss multidimensionaler Modelle; konfirmatorische Mehrgruppen-Faktorenanalysen; Optimierung von Trennwerten mit Hilfe von ROC-Analysen; Lautdenkprotokolle und retrospektive Interviews zu den (mentalen) Lösungsprozessen; Korrelationsanalysen mit diversen Außenkriterien; Erhebung der Einstellungen zu C-Tests bei Studierenden und Lehrenden mit Hilfe von Fragebögen und qualitativen Interviews; Mittelwertvergleiche zwischen den Ergebnissen im Einstufungstest und im Kursabschlusstest (ANOVA); Fehleranalysen der C-Test-Lösungen. Auffallend ist die Breite der methodologischen Zugänge sowie der zunehmende Einsatz auch komplexerer mathematischer Analyseverfahren.

2.2. Die einzelnen Beiträge

Der Beitrag „C-Tests und ‚allgemeine Sprachkompetenz': Theoretische Überlegungen und empirische Analysen" von **Yuki Asano** beschäftigt sich mit der grundlegenden Frage, inwieweit (deutsche) C-Tests als Maß allgemeiner Sprachkompetenz anzusehen sind (vgl. hierzu auch Eckes, im vorliegenden Band; Linnemann & Wilbert, im vorliegenden Band; Sumbling, Viladrich, Doval & Riera, im vorliegenden Band). Die Autorin definiert allgemeine Sprachkompetenz als eindimensionales Konstrukt, das im Sinne einer zentralen integrativen Kompetenz allen Sprachleistungen (unter Einschluss der klassischen vier Fertigkeiten) zugrunde liegt und das von kommunikativer Sprachkompetenz im Sinne eines komplexen mehrdimensionalen Konstrukts zu unterscheiden ist.

Für eine erste theoretische Klärung, warum ein schriftlicher Lückentest allgemeine Sprachkompetenz zu erfassen erlaubt, diskutiert die Autorin zunächst

Ollers Konzept einer internalisierten Erwartungsgrammatik und die Rolle sprachlicher Redundanz bei der Verarbeitung natürlicher Sprachen. Je höher die Sprachkompetenz einer Person ist, desto eher ist diese in der Lage, mit Hilfe ihrer Erwartungsgrammatik die sprachlichen Redundanzen zu nutzen und die Lücken eines C-Tests erfolgreich zu rekonstruieren.

Zur weiteren Klärung betrachtet die Autorin allgemeine Sprachkompetenz aus der Sicht der modernen Psycholinguistik. Dazu beschreibt sie zunächst psycholinguistische Modelle, die „von einem gemeinsamen kognitiven System für Sprachproduktion und -rezeption ausgehen" – mit einem Schwerpunkt auf dem bekannten Modell von Levelt. Weiterhin verweist Asano u.a. auf Libermans „Motor Theory of Speech Perception". Daran anschließend werden eine Reihe von Argumenten genannt, die eher für eine kognitive und biologische Trennung von Sprachproduktion und -rezeption sprechen.

Die theoretischen Überlegungen sind der Ausgangspunkt für die empirische Studie der Autorin. Diese besteht aus drei Teilstudien und zeichnet sich sowohl durch eine Daten- als auch eine Methodentriangulation aus. Anhand korrelationsstatistischer Analysen wird der Zusammenhang zwischen allgemeiner Sprachkompetenz (gemessen anhand von C-Tests) und den vier Fertigkeitsbereichen (gemessen anhand des TestDaF) untersucht. Mit Hilfe von Beobachtung und introspektiven Daten (Lautes Denken; retrospektives Interview) wird der Frage nachgegangen, ob beim schriftlichen Lösen von C-Tests ähnliche Prozesse wie beim Sprechen und Hören ablaufen. Untersucht wurden 10 bzw. 18 japanische Deutschlernende (Datenerhebungen 1 bzw. 2) sowie zwei japanische Deutschlernende und drei deutsche Muttersprachler (Datenerhebung 3) mit Hilfe eines selbst entwickelten C-Tests und des im TestDaF als Anker verwendeten C-Tests. Auf potentielle Einschränkungen u.a. aufgrund der kleinen Stichproben weist die Autorin selbst explizit hin.

In der korrelationsstatistischen Untersuchung ergaben sich signifikante, substantielle Pearson-Korrelationen zwischen den C-Tests und den vier Teiltests des Test-DaF. Die höchste Korrelation (0.81; $N = 18$) ergab sich zwischen dem TestDaF-Ankertest und dem Gesamtergebnis im TestDaF. Die Ergebnisse sprechen nach Auffassung der Autorin insgesamt dafür, „dass C-Tests nicht nur die Lesekompetenz oder die Schreibkompetenz messen, sondern vornehmlich integrative Fähigkeiten, die alle vier Fertigkeitsbereiche mehr oder weniger abdecken".

Die Analyse der verbalen Daten zur Lösung des selbst entwickelten C-Tests ergab ebenfalls eine Reihe von interessanten Befunden: So kamen z.B. Muttersprachler und Nichtmuttersprachler z.T. anhand unterschiedlicher Prozesse zu den gleichen Lösungen. Weiterhin scheint sich das Verstehen bei Nichtmutter-

sprachlern vorwiegend auf die Satzebene zu beziehen (vgl. für partiell unterschiedliche Befunde Babaii & Fatahi-Majd sowie Grotjahn & Schiller, im vorliegenden Band). Schließlich war die Herangehensweise der Testteilnehmer in Abhängigkeit von Faktoren wie Motivation und Testschwierigkeit unterschiedlich. Dies belegt nach Ansicht der Autorin Sigotts (2004; 2006) Konzept eines fluiden C-Test-Konstrukts. Die Fluidität des Testkonstrukts ist m.E. allerdings kein spezifisches Merkmal von C-Tests, sondern gilt auch in Bezug auf andere (standardisierte) Sprachtests.

Im folgenden 56-seitigen Beitrag „C-test as an indicator of general language proficiency in the context of a CBT (SIMTEST)“ beschreiben **Mick Sumbling**, **Carme Viladrich**, **Eduardo Doval** und **Laura Riera** im Detail die Funktion und psychometrischen Qualitäten eines an der Autonomen Universität von Barcelona (UAB) als Teil des Testsystems SIMTEST eingesetzten C-Tests. SIMTEST wurde an der Einheit für Evaluation und Zertifizierung des Sprachenzentrums der UAB entwickelt und existiert für Englisch und weitere Sprachen. Es handelt sich um ein computer-basiertes Testsystem (CBT) zur Einstufung in Sprachkurse und zur Zertifizierung am Kursende auf der Basis des Gemeinsamen europäischen Referenzrahmens für Sprachen (GER). Da von den Testergebnissen der Verlauf und Erfolg des Studiums abhängen kann, ist SIMTEST von seiner Funktion her ein high-stakes-Test. Der vorliegende Beitrag beschränkt sich auf die englische Version des Testsystems.

Neben einem nicht-adaptiven C-Test (20 Min.) besteht SIMTEST aus adaptiven Multiple-Choice-Tests zu Lexik und Grammatik (VGF-CAT; 10–20 Min.) sowie Hörverstehen (L-CAT; 10–20 Min.). Das Hörverstehen wird nur bei der Zertifizierung überprüft. SIMTEST wird ergänzt durch eine Schreibaufgabe im Papier-Bleistift-Format (Einstufung: 20 Min.; Zertifizierung: 90 Min.) und einer mündlichen Prüfung (Einstufung: Einzelprüfung, 10 Min.; Zertifizierung: Paarprüfung, 20 Min.). Die Testteilnehmer haben in der Regel noch keinen universitären Abschluss und sind zu 65% weiblich. Die Englischkurse sind eher allgemeinsprachlich ausgerichtet und den sechs GER-Niveaus A1, A2, B1, B2.1, B2.2 und C1.1 zugeordnet.

Der C-Test entspricht dem klassischen Konstruktionsprinzip und besteht aus jeweils vier Texten mit 25 Lücken pro Text. Die vorerprobten Texte werden per Zufall einer Datenbank entnommen und gelten als gegenseitig austauschbar. Das C-Test-Ergebnis dient als ein erster, vorläufiger Schätzwert für die englische Sprachkompetenz und als Startwert für die adaptiven Teile des SIMTEST. Der Aufbau einer Datenbank angemessen schwieriger, äquivalenter C-Test-Texte ist damit ein zentraler Aspekt des Testprojekts.

Die Autoren gehen zunächst kurz auf die Testspezifikationen und den institutionellen Hintergrund für die Testentwicklung ein. Anschließend beschreiben sie im Detail die durchgeführten psychometrischen Analysen. Es wurden fünf Datensätze analysiert: ein Papier-Bleistift-Prätest von 20 C-Test-Texten mit insgesamt 225 Studierenden (TE01); ein Einstufungstest mit 500 Studierenden (PL01); ein computerbasierter Prätest von 26 neuen C-Test-Texten mit insgesamt 396 Studierenden (TE05); ein Einstufungstest mit 809 Studierenden (PL05); eine Zertifikatsprüfung mit 816 Studierenden – darunter 441 aus der vorangehenden Einstufungsprüfung (CE06). Zur Analyse der internen Konsistenz wurde Cronbachs Alpha und der Intraklassen-Korrelationskoeffizient (ICC) berechnet, der nicht nur die Rangordnung, sondern auch die Schwierigkeit der Items (C-Test-Texte) berücksichtigt. Außerdem wurde der Standardmessfehler ermittelt. Die Dimensionalität wurde anhand der Itemtrennschärfen und der Übereinstimmung der Daten mit Andrichs Rating-Skalen-Modell beurteilt. Für das Rating-Skalen-Modell wurden zudem Reliabilitäts- und Separationsindizes berechnet.

Weiterhin wurden folgende Validitätsaspekte untersucht: (a) Korrelation zwischen C-Test-Ergebnissen und aktueller Kursstufe der Testteilnehmer; (b) Korrelation zwischen den Ergebnissen im C-Test und im VGF-CAT, Hörverstehenstest, Schreibaufgabe und mündlicher Prüfung; (c) Überprüfung der faktoriellen Struktur der eingesetzten Tests mit Hilfe konfirmatorischer Faktorenanalysen; (d) Korrelation zwischen C-Test-Ergebnissen und Selbsteinschätzungen der Testteilnehmer; (e) Zuwachs der C-Test-Ergebnisse in Abhängigkeit von der Kursstufe (ANOVA); (f) Güte der anhand der C-Tests auf der Basis von ROC-Analysen (*Receiver Operating Characteristic(s)*) ermittelten Trennwerte.

Bei den ROC-Analysen wurde davon ausgegangen, dass die mit einer falschen positiven und einer falschen negativen Entscheidung (Überschätzung vs. Unterschätzung des wahren Sprachniveaus) verbundenen Kosten gleich zu gewichten sind. Entsprechend wurden die Trennwerte so gewählt, dass die Differenz zwischen der sog. Sensitivität (Wahrscheinlichkeit einer richtigen positiven Entscheidung) und der sog. Spezifität (Wahrscheinlichkeit einer richtigen negativen Entscheidung) minimal war. Als Maß wurde die Fläche unter der ROC-Kurve zu Grunde gelegt (zur Verwendung von ROC-Kurven siehe auch Kniffka & Linnemann, im vorliegenden Band sowie Goldhammer & Hartig, 2007, S. 175–181).

Es ergaben sich u.a. folgende Ergebnisse: Die auf der Basis der klassischen Testtheorie und des Rating-Skalen-Modells der probabilistischen Testtheorie durchgeführten Analysen zeigten für die verwendeten C-Tests sehr zufrieden stellende psychometrische Eigenschaften. Nur wenige Texte erwiesen sich als

nicht modellkonform. Außerdem waren die Texte in ihrer Schwierigkeit weitgehend vergleichbar. Damit war eine wichtige Bedingung für die Austauschbarkeit der Texte im Hinblick auf ein Item-Banking erfüllt. Zudem konnten auf verlässliche Weise fünf verschiedene Niveaustufen anhand der C-Test-Ergebnisse unterschieden werden.

Weiterhin ergaben sich mittlere bis hohe Korrelationen zwischen C-Test-Ergebnissen und aktueller Kursstufe und statistisch signifikante, substanzielle Zuwächse im C-Test über die Kursstufen. Die mit Hilfe der ROC-Analysen ermittelten Trennwerte zeigten fast durchgehend hohe Werte für Sensitivität, Spezifität und AUC (*area under the curve*). Die Korrelationen zwischen den Testteilen waren in allen Stichproben stets größer als 0.5. Speziell für den C-Test ergaben sich folgende Korrelationen mit den übrigen Testteilen: $0.60 \leq r \leq 0.65$ (VGF-CAT); $0.56 \leq r \leq 0.69$ (Schreibaufgabe); $0.63 \leq r \leq 0.70$ (mündliche Prüfung); $r = 0.51$ (Hörverstehen). Die restlichen Korrelationen lagen im Bereich von 0.64 (Hörverstehen und VGF-CAT) bis 0.91 (Schreibaufgabe und mündliche Prüfung). Bemerkenswert sind hier u.a. die relativ hohe Korrelationen der C-Tests mit den mündlichen Prüfungen und die zugleich eher geringe Korrelation zwischen C-Test und Hörverstehen – ein Befund, der eher im Widerspruch zur bisherigen Forschung zum C-Test steht (siehe die Belege in Eckes & Grotjahn, 2006; Baghaei & Grotjahn, 2014; sowie auch Asano, im vorliegenden Band). Auch die Korrelationen der C-Test-Komponente mit den Selbsteinschätzungsskalen Hörverstehen, Produktion mündlich, Interaktion mündlich, Lesen und Schreiben sind mit ca. 0.50 relativ hoch. Zwischen C-Test und Gesamtwert der Selbsteinschätzungen beträgt die Korrelation sogar 0.57. Hier ist allerdings zu fragen, inwieweit diese Werte nicht auch die jeweils erreichte Kursstufe widerspiegeln.

Bemerkenswert sind auch die Ergebnisse der konfirmatorischen Faktorenanalysen (CFA). Ziel der CFA war, die von Eckes & Grotjahn (2006) im Hinblick auf die Komponentialität des Konstrukts „general language proficiency" formulierten fünf Modelle für die Beziehung zwischen C-Test und den vier Komponenten des TestDaF (Test Deutsch als Fremdsprache) in Bezug auf den SIMTEST zu überprüfen. Es ergaben sich für die drei untersuchten Stichproben trotz unterschiedlicher Anzahl von Variablen und Unterschiede in der Verteilung der Fähigkeiten weitgehend übereinstimmende Ergebnisse. In den beiden Einstufungstest-Stichproben mit den Komponenten C-Test, VGF-CAT, Schreiben und Sprechen ergab sich jeweils u.a. eine Einfaktorlösung. Besser passte allerdings ein Modell mit den hochkorrelierten Faktoren Produktion (operationalisiert durch mündliche Prüfung und Schreibaufgabe) und Rezeption (operationalisiert durch VGF-CAT und C-Test). Für die Zertifikatsstichprobe ergab eben-

falls das Modell mit der zusätzlichen Variablen Hörverstehen als Teil des Faktors Rezeption die beste Passung. Allerdings war die Anpassung hier deutlich schlechter als in den beiden anderen Stichproben. Insgesamt sprechen die Analysen nach Ansicht der Autoren dafür, dass die untersuchten C-Tests allgemeine Sprachkompetenz im Sinne eines weiter unterteilbaren Faktors „general language proficiency" messen. Dies stimmt weitgehend mit den Folgerungen in Eckes & Grotjahn (2006) überein. Dort wurde die beste Modellanpassung allerdings dann erreicht, wenn der C-Test nicht dem Faktor Rezeption, sondern dem Faktor Produktion (TestDaF-Subtests Schriftlicher Ausdruck und Mündlicher Ausdruck) zugeordnet wurde (vgl. auch die theoretischen Modelle von Sprachkompetenz für die englischen DESI-Daten in Jude, 2008, Kap. 6.2.2 und 6.2.3). Sumbling, Viladrich, Doval und Riera weisen auf die Möglichkeit hin, dass die Unterschiede in den Ergebnissen der beiden Studien möglicherweise auf Unterschiede in den Testteilnehmern (im Fall des TestDaF eine Vielzahl von Nationalitäten und Muttersprachen) sowie auch auf Unterschiede in den eingesetzten Itemformaten zurückzuführen sind.

Die Autoren schließen ihre psychometrischen Analysen mit der Betrachtung der Sensitivität des C-Test im Hinblick auf die Messung von Lernzuwachs ab. Auf den untersuchten Kursstufen B1, B2.1, B2.2 und C1.1 war der Kompetenzzuwachs nach 9-monatigem Englischunterricht jeweils signifikant und substanziell (mittlere bis hohe Effektstärken).

Im letzten Teil ihres Beitrags beschäftigen sich die Autoren ausführlich mit den Reaktionen der Testteilnehmer und der Lehrenden auf den C-Test sowie mit dessen Augenscheingültigkeit (*face validity*) – ein Aspekt, der auch im Hinblick auf den Washback-Effekt und die konsequenzielle Validität von C-Tests eine wichtige Rolle spielt (vgl. hierzu auch Mashkovskaya, 2014, Kap. 4.7; Mainzer, 2014; sowie Harsch & Schröder, im vorliegenden Band). Zur Untersuchung dieses empirisch bisher nur unzureichend erforschten Aspekts wurden anhand von Fragebögen mit geschlossenen und offenen Formaten Daten erhoben u.a. zur Schwierigkeit der C-Tests, zu Klarheit der Instruktionen, zu den Zeitvorgaben und zum Testkonstrukt. Schwierigkeit und Zeitvorgaben wurden tendenziell als weitgehend angemessen angesehen. Die offenen schriftlichen Kommentare der Studierenden zum C-Test waren in der Tendenz eher positiv, wobei negative Kommentare allerdings zuweilen extreme Ablehnung des C-Tests ausdrückten. Die Autoren belegen dies anhand einer Vielzahl aus dem Katalanischen ins Englische übersetzten Aussagen. Außerdem wurden mündliche Interviews mit sechs Studierenden zu deren Wahrnehmung des C-Tests und zu den eingesetzten Bearbeitungsstrategien durchgeführt, die ebenfalls im Detail von den Autoren kommentiert werden. Ein wichtiger Befund aus den Untersuchungen zu den Re-

aktionen der Lehrenden ist u.a., dass diese im Laufe der Zeit und mit zunehmender Vertrautheit mit C-Tests das C-Test-Verfahren zunehmend positiv beurteilten. Die Autoren stellen abschließend fest, dass trotz der nachgewiesenen sehr positiven psychometrischen Qualitäten von C-Tests das Problem der tendenziell nicht sehr hohen Augenscheingültigkeit weiterhin bestehe und dass Studierende, Lehrende und Sprachtester besser über die Vorzüge von C-Tests informiert werden sollten. Wie allerdings die nunmehr 33-jährige Geschichte der C-Test-Forschung belegt, sind Vorurteile gegenüber dem C-Test nicht leicht auszuräumen.

Monique Reichert, **Martin Brunner** und **Romain Martin** gehen in „Do test takers with different language backgrounds take the same C-test? The effect of native language on the validity of C-tests" der wichtigen Frage nach, inwieweit ein deutscher und ein französischer C-Test im Rahmen einer großangelegten Bildungsstudie bei Schülern der Klasse 7 in Luxemburg das (eindimensionale) Konstrukt „allgemeine Sprachkompetenz" (*general language proficiency*) in äquivalenter Weise in zwei verschiedenen Sprachgruppen zu messen erlaubt. Sollte keine ausreichende Messäquivalenz vorliegen, dann ist ein Vergleich zwischen Testteilnehmern aus unterschiedlichen Sprachgruppen oder auch zwischen den Sprachgruppen insgesamt problematisch.

Im ersten Teil ihres Beitrags diskutieren die Autoren zunächst den Einfluss von Eigenschaften der Testteilnehmer und insb. der Muttersprache auf die faktorielle Validität von Sprachtests. Die diskutierten Studien konnten u.a. zeigen, dass der Einfluss der Muttersprache tendenziell am stärksten am Anfang des Fremdsprachenlernprozesses war (vgl. hierzu auch Jude, 2008).

In Bezug auf den C-Test ist bisher der Einfluss insb. folgender Variablen auf die C-Test-Leistung und das C-Test-Konstrukt untersucht worden: kognitiver Stil, Intelligenz, Testerfahrung, Angst bei der Verwendung des Computers sowie Sprachstand. So konnte z.B. Sigott (2004, 2006) zeigen, dass sich weiter fortgeschrittene und weniger fortgeschrittene Englischlerner in Bezug auf die bei der Rekonstruktion der Lückenwörter verwendeten kognitiven Prozesse unterscheiden (vgl. hierzu Asano, im vorliegenden Band; Babaii & Fatahi-Majd, im vorliegenden Band; Grotjahn & Schiller, im vorliegenden Band). Welchen Einfluss die Muttersprache der Kandidaten auf die Lösung von C-Tests hat, ist jedoch bisher weitgehend ungeklärt. Eine Klärung ist im Hinblick auf den beschriebenen Einsatz von C-Tests in Luxemburg von großer Bedeutung: Luxemburg weist den größten Anteil an Migranten innerhalb der europäischen Union auf und hat zudem drei Amtssprachen, nämlich Luxemburgisch, Deutsch und Französisch.