

Zeitschrift für Interkulturellen Fremdsprachenunterricht

Didaktik und Methodik im Bereich Deutsch als Fremdsprache

ISSN 1205-6545 Jahrgang 18, Nummer 1 (April 2013)

Zertifikat B1 Deutschprüfung für Jugendliche und Erwachsene: Standard Setting. Ein Arbeitsbericht

Dr. Manuela Glaboniat

Universität Klagenfurt
Universitätsstr. 65-67
A-9020 Klagenfurt,
Telefon: +43 (0) 463 2700 2722
Fax: +43 (0) 463 2700 2799
E-Mail: manuela.glaboniat@aau.at

Dr. Michaela Perlmann-Balme

Goethe-Institut e.V.
Bereich 41 Sprachkurse und Prüfungen
Dachauer Straße 122
80637 München
Telefon: +49 89 15921-382
E-Mail: Perlmann-balme@goethe.de

Prof. Dr. Thomas Studer

Universität Freiburg Schweiz / Université de Fribourg Suisse
Mehrsprachigkeitsforschung und Fremdsprachendidaktik, Deutsch als Fremdsprache
Rue de Rome 1
CH-1700 Fribourg
Telefon: +41 26 300 7999
E-Mail: thomas.studer@unifr.ch

Abstract: Neue Sprachprüfungen haben in der Regel den Anspruch, sich auf ein externes Referenzsystem zu beziehen. In Europa ist dieses Referenzsystem seit mehr als einer Dekade der *Gemeinsame europäische Referenzrahmen für Sprachen* (GER). Um diesem Anspruch gerecht zu werden, sind Testanbieter bemüht, ihre Bezüge auf den GER mithilfe von Standards zum Sprachen-testen generell und Standards der *Association of Language Testers in Europe* (ALTE) im Besonderen nachzuweisen. Der Europarat empfiehlt Sprachtestanbietern einen methodologischen Leitfaden, das sogenannte *Manual for Relating Language Examinations to the Common Framework of Reference for Languages* (CEFR), der sie bei der „Eichung“ ihrer Sprachprüfung unterstützt. Kapitel 5 und 6 des Leitfadens beschreiben Methoden zum *Benchmarking* und *Standard Setting*. Das Ziel dieses Berichts ist es zu zeigen, mit welchen Methoden die neu entwickelte Prüfung für Deutsch als Fremdsprache *Zertifikat B1* auf die Stufe B1 des Referenzrahmens bezogen wurde und welche Ergebnisse dabei erreicht wurden.

As a rule, new language examinations, claim to be linked to an external reference system. In Europe this reference system has been the *Common European Framework of Languages* (CEFR) for more than a decade. In order to substantiate the validity of this claim, examination providers take pains to provide evidence of alignment to this system, guided by standards in language testing and, in particular, the standards set out by the *Association of Language Testers in Europe* (ALTE). The Council of Europe recommends a methodological guide for examination providers, the so-called *Manual for Relating Language Examinations to the Common Framework of Reference for Languages* (CEFR), which assists in the calibration of language tests. Chapters 5 and 6 of the Manual describe possible methods of benchmarking and standard setting. The aim of this paper is to report how the newly developed examination in German as a foreign language, *Zertifikat B1*, was certified to be at the level B1 of the CEFR; it explains the methodological steps as well as the results.

Schlagwörter: Deutschprüfung, Prüfung, Sprachprüfung, Standard Setting, Benchmarking, Leistungsbeispiele, Niveau B1, Test, Europäischer Referenzrahmen, Sprachkompetenzen, Deutsch als Fremdsprache

Kontext

Welches Ergebnis muss ein Lernender erreichen, um einen Test zu bestehen? Welche Leistungen sind ungenügend? Die Frage nach der Bestehensgrenze ist im Bereich des Prüfens und Testens zentral, weil darauf abgestützte Entscheidungen folgenreich sein können, für Teststeilnehmende ebenso wie für Entscheidungsträger, beispielsweise eine Schule, Kursanbieter oder eine Behörde, und diese Frage betrifft grundsätzlich jede Art von Test, nicht nur ‚große Prüfungen‘: Lernfortschrittskontrollen in der Schule, Einstufungstests als Grundlage für die Kurszuteilung, Qualifikationstests zur Feststellung von Sprachkompetenzen und auch diagnostische Tests, mit denen Stärken und Schwächen von Lernenden festgestellt und förderorientiert rückgemeldet werden – immer müssen Grenzwerte definiert werden, die darüber informieren, ob die gezeigten Leistungen den Erwartungen entsprechen oder nicht. Dazu wird das Leistungsspektrum unterteilt, sei es in zwei Hälften (bestanden oder nicht bestanden) oder mehrfach (erreichte bzw. nicht erreichte Stufe).

Im Rahmen der Entwicklung des *Zertifikats B1*, einer neuen, internationalen Qualifikationsprüfung (*proficiency test*) für Deutsch als Fremdsprache, hat die Frage nach der Bestehensgrenze aus zwei Gründen eine besondere Tragweite: Zum einen ist beim *Zertifikat B1* wie schon bei dessen Vorgänger, dem *Zertifikat Deutsch*, mit einer weltweit hohen Teilnehmerzahl zu rechnen und damit, dass dieser Prüfung in Ausbildungs- und Berufszusammenhängen eine entscheidende Bedeutung beigemessen wird (*high-stakes test*). Zum anderen wurde bei der Entwicklung das Ziel verfolgt, die Prüfung möglichst genau und transparent auf das Beschreibungssystem des *Gemeinsamen europäischen Referenzrahmens für Sprachen* (GER) und dessen Niveaubeschreibungen für die Stufe B1 zu beziehen. Aus Ersterem ergibt sich ein hohes Maß an Verantwortung für die Testentwickler, was die Einhaltung prüfungsrelevanter Qualitätsstandards angeht (vgl. die Richtlinien der EALTA von 2006 und die Minimalstandards der ALTE von 2007). Letzteres stellt eine fachwissenschaftliche Herausforderung dar, die u.a. eine Auseinandersetzung mit dem Begriff der Sprachkompetenz, dem Testkonstrukt und Operationalisierungen des Konstrukts in Aufgaben nach sich zieht. Welche Überlegungen dabei maßgeblich waren, wird im Handbuch zum *Zertifikat B1* (Glaboniat, Perlmann-Balme & Studer i.Dr.) genauer beschrieben.

Der Terminus *Standard Setting* meint in unserem Kontext eine Gruppe von strukturierten Verfahren, deren Ziel darin besteht, Lernerleistungen auf verbal definierte Niveaustufen wie diejenigen des GER zu beziehen (Kenyon 2013: 1). Die Auseinandersetzung mit diesen Praktiken hat eine lange Tradition in der Psychometrie, jüngeren Datums ist dagegen die Einsicht, dass für das Ermitteln von Bestehensgrenzen statistische Werkzeuge zwar gebraucht werden, der Prozess des *Standard Setting* selbst aber nicht in einem sozialen Vakuum stattfindet, sondern (mit-) bestimmt wird durch menschliche Urteile in Form von individuellen und in der Gruppe ausgehandelten Entscheidungen (Kantarcioglu & Papageorgiou 2011: 99f). Gegenstand von *Standard Settings* sind besonders rezeptive Leistungen, in der Regel Testresultate zum Hör- und Leseverstehen in Form von Punkten, die von einem aus Fachleuten zusammengesetzten Panel danach beurteilt werden, ob sie für das Erreichen einer Niveaustufe genügend und/oder typisch sind. Zwar können sich *Standard Settings* auch auf die Beurteilung schriftlicher und mündlicher Leistungen beziehen, doch ist hierfür der Begriff *Benchmarking* üblicher, wobei unter *Benchmarks* auch Beispiele von produktiven Lernerleistungen verstanden werden, die für das Training von Ratern eingesetzt und dazu gebraucht werden, Lernende mit dem Testinhalt vertraut zu machen (ebd.: 94).

Die bei der Entwicklung des *Zertifikats B1* eingesetzten Verfahren des *Standard Setting* und *Benchmarking* stützen sich auf die Fachliteratur, darunter neben der bereits erwähnten v.a. das *Manual for relating language examinations to the Common European Framework of Reference for Languages* (Council of Europe 2009), Béréšová et al. 2011, Kaftandjieva 2010, Figueras & Noijons 2009, Cizek & Bunch 2007 und Karantonis & Sireci 2006.

Die Erfahrungen, die wir mit dem *Standard Setting* und *Benchmarking* gemacht haben, veröffentlichen wir in Form dieses Arbeitsberichts, um das Vorgehen und die Entscheidungsfindungen offen zu legen und nachvollziehbar zu

machen. Das scheint uns angezeigt, obwohl der Bericht zweifellos auch diskutabile Passagen enthält und gerade *weil* die erzielten Resultate urteilsbasierter Natur sind (Kantarcioğlu & Papageorgiou 2011). Kenyon (2013: 1) empfiehlt in diesem Zusammenhang:

Careful documentation of each study includes the background and qualifications of panelists, detailed description of the procedures and materials used, the analysis of the data and results, and an evaluation by participants of their experience. Such documentation allows others to judge the quality of the study ...

Schließlich verstehen wir diesen Bericht auch als Beitrag zu einem *assessment use argument* im Sinne von Bachman & Palmer 2010, genauer, als empirisch abgestütztes Validitätsargument für die in der neuen Prüfung angesetzten und auf den GER bezogenen Bestehensgrenzen. Dabei ist klar, dass das Beibringen von Evidenz für die Validität eines Tests immer nur ein Schritt in einem prinzipiell nie abschließbaren Prozess ist (vgl. Weir 2005).

Anhang. [Zertifikat B1 Deutschprüfung für Jugendliche und Erwachsene Standard Setting. Ein Arbeitsbericht.](#)

Literatur

- Association of Language Testers in Europe (ALTE) (Hrsg.) (2007), *Minimum standards for establishing quality profiles in ALTE examinations*. [Online: http://www.alte.org/attachments/files/minimum_standards.pdf. 1.3.13].
- Bachman, Lyle & Palmer, Diana (2010), *Language Assessment in Practice*. Oxford: Oxford University Press (= Applied Linguistics).
- Bérešová, Jana; Breton, Gilles; Noijons, José & Szabó, Gábor (2011), *Relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR). Highlights from the Manual*. Graz: ECML. [Online: <http://www.ecml.at/tabid/277/PublicationID/67/Default.aspx>. 01.03.2013].
- Cizek, Gregory J. & Bunch, Michael B. (2007), *Standard Setting. A Guide to Establishing and Evaluating Performance Standards on Tests*. Thousand Oaks, CA: Sage.
- Council of Europe (Hrsg.) (2009), *Relating Language Examinations to the Common European Framework of References for Languages: Learning, Teaching, Assessment. A Manual*. Strasbourg, Language Policy Division: Council of Europe.
- EALTA (2006), *Guidelines for good practice in language testing and assessment* (Adopted 20th May 2006). [Online: <http://www.ealta.eu.org/guidelines.htm>. 01.03.2013].
- Europarat (Hrsg.) (2001), *Gemeinsamer europäischer Referenzrahmen für Sprachen: lernen, lehren, beurteilen*. Berlin: Langenscheidt.
- Figueras, Neus & Noijons, José (Hrsg.) (2009), *Linking to the CEFR levels: Research perspectives*. Arnheim: Cito, EALTA.
- Glaboniat, Manuela; Perlmann-Balme, Michaela & Studer, Thomas (2013, in Druck), *Zertifikat Deutsch. Deutschprüfung für Jugendliche und Erwachsene. Prüfungsziele, Testbeschreibung*. Ismaning: Hueber.
- Kaftandjieva, Felianka (2010), *Methods for Setting Cut Scores in Criterion-Referenced Achievement Tests. A Comparative Analysis of Six Recent Methods with an Application to Tests of Reading in EFL*. Arnheim: EALTA Cito.
- Kantarcioğlu, Elif & Papageorgiou, Spiros (2011), Benchmarking and standards in language tests. In: O'Sullivan, Barry (Hrsg.), *Language Testing. Theories and Practices*. New York: Palgrave, 94-110.
- Karantonis, Ana & Sireci, Stephen G. (2006), The Bookmark Standard Setting Method: A literature review. *Educational Measurement: Issues and Practice* 25, 4–12.

Kenyon, Dorry (2013), Standard setting on language tests. In: Chapelle, Carol A. (Hrsg.), *The Encyclopedia of Applied Linguistics*. Blackwell, 1-5. [Online: <http://onlinelibrary.wiley.com/doi/10.1002/9781405198431.wbeal1113/pdf>. 23.03.2013].

Weir, Cyril J. (2005), *Language Testing and Validation: An Evidence-Based Approach*. Basingstoke: Palgrave/Macmillan.