

INQUIRIES IN LANGUAGE LEARNING

Forschungen zu Psycholinguistik und Fremdsprachendidaktik

Edited by / Herausgegeben von Christiane Bongartz / Jutta Rymarczyk

5

Henning Rossa

Mentale Prozesse beim Hörverstehen in der Fremdsprache

Eine Studie zur Validität der Messung
sprachlicher Kompetenzen



PETER LANG

Internationaler Verlag der Wissenschaften

INQUIRIES IN LANGUAGE LEARNING

Forschungen zu Psycholinguistik und Fremdsprachendidaktik

Edited by / Herausgegeben von Christiane Bongartz / Jutta Rymarczyk

5

Henning Rossa

Mentale Prozesse beim Hörverstehen in der Fremdsprache

Eine Studie zur Validität der Messung
sprachlicher Kompetenzen

LESEPROBE



PETER LANG

Internationaler Verlag der Wissenschaften

1 Das Hörverstehen und die Validität von Testaufgaben: Desiderata der Fremdsprachenforschung im Kontext der Kompetenzorientierung

Seit ungefähr zehn Jahren werden bildungspolitische und wissenschaftliche Diskurse zunehmend von der Forderung dominiert, das Lehren und Lernen in der Schule und in anderen Bildungskontexten solle durch eine Neuorientierung an Kompetenzen, Standards und Lernergebnissen gefördert und verbessert werden. Dieser Reformprozess, der häufig mit dem Begriff ‚Kompetenzorientierung‘ umschrieben wird, geht einher mit dem verstärkten Bemühen, die fachspezifischen und übergreifenden Kompetenzen der Lernenden zu messen und die Bedingungen zu erforschen, die den Kompetenzerwerb unterstützen.

Ausgelöst durch die zum Teil beunruhigenden Ergebnisse internationaler Vergleichsstudien wie TIMSS (Beaton et al. 1996), PISA (Deutsches PISA-Konsortium 2001) und IGLU (Bos et al. 2003) werden in Deutschland mittlerweile auf unterschiedlichen Ebenen des Bildungssystems regelmäßig Kompetenztests, Lernstandserhebungen, Vergleichsarbeiten und zentrale Prüfungen eingesetzt, um zu messen, inwiefern die angestrebten Ziele erreicht werden (vgl. Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland 2006).

Als Folge dieser Entwicklung lässt sich auch im Kontext der Fremdsprachenforschung in Deutschland eine intensivierte Ausrichtung des Forschungsinteresses auf die Messung und Beschreibung fremdsprachlicher Kompetenzen beobachten. Bislang sind drei Bezugspunkte von besonderer Bedeutung für die Entwicklung der Diskussion um die Kompetenzorientierung des Fremdsprachenunterrichts gewesen: erstens die Veröffentlichung des Gemeinsamen europäischen Referenzrahmens für Sprachen (vgl. Europarat 2001), zweitens die Implementierung und Überprüfung nationaler Bildungsstandards für die erste Fremdsprache (vgl. Harsch, Pant & Köller 2010; Porsch, Tesch & Köller 2010; Rupp, Vock, Harsch & Köller 2008; Zydatiß 2005) und drittens die Rezeption der Schulleistungsstudie DESI und ihrer Forschungsergebnisse zu den Kompetenzen deutscher Schülerinnen und Schüler in der Fremdsprache Englisch und zur Wirksamkeit des Englischunterrichts (vgl. DESI-Konsortium 2008¹).

1 Das DESI-Konsortium besteht aus Eckhard Klieme (DIPF, Sprecher), Wolfgang Eichler (Oldenburg), Andreas Helmke (Landau), Rainer H. Lehmann (Berlin), Günter Nold (Dortmund), Hans-Günter Rolff (Dortmund), Günther Thomé (Osnabrück) und Heiner Willenberg (Hamburg).

Mit Blick auf die vorangestellte These Ciceros muss jedoch konstatiert werden, dass Erkenntnisse über die Gültigkeit dieser Kompetenzmessungen – etwa durch die Erforschung der Frage, wie die Messergebnisse zustande kommen – bislang nur sehr eingeschränkt vorliegen. Wegen der bedeutsamen Auswirkungen auf das Fremdsprachenlernen in der Schule, die aus der Interpretation der Testergebnisse nationaler und internationaler Schulleistungsstudien erwachsen (‘washback effect’, vgl. Bailey 1999; Cheng 2005; Spratt 2005), muss es ein vornehmliches Ziel der Fremdsprachenforschung werden, die Erfassung fremdsprachlicher Kompetenzen durch Testaufgaben kritisch zu begleiten und die Angemessenheit der Messinstrumente und der Interpretation der Messergebnisse zu überprüfen.

Die hier vorgelegte Forschungsarbeit bezieht sich auf diese Desiderata und untersucht die Validität der Testaufgaben zum Hörverstehen in der Fremdsprache Englisch, die in der DESI-Studie eingesetzt wurden (Nold & Rossa 2007; Nold, Rossa & Hartig 2008b). Die zentrale Forschungsfrage der Untersuchung lautet: Inwiefern gelingt es, mit den Testaufgaben die Facetten des theoretisch spezifizierten Testkonstrukts ‚Hörverstehen in der Fremdsprache Englisch‘ zu erfassen?

Dieses Erkenntnisinteresse berührt zwei bedeutsame Problemfelder, die von Brindley (1998: 127 ff.) als Schnittstelle zwischen den Gegenständen der Zweit-sprachenerwerbsforschung und der Sprachtestforschung beschrieben werden: die theoretische Modellierung und Beschreibung der Sprachkompetenz und die Spezifizierung von Messinstrumenten, die valide Messergebnisse ermöglichen sollen. Pienemann und Keßler (2007: 257) argumentieren, dass sich Studien zur Sprachkompetenz bzw. zum Spracherwerb an dieser Schnittstelle positionieren müssen. Der Versuch einer Positionierung gleicht jedoch einem Dilemma (‘scope-precision dilemma’), da die Erfassung eines komplexen Konstrukt, wie etwa der kommunikativen Sprachkompetenz, zwangsläufig zu Einbußen in der Präzision der Messung führt (‘precision’), während eine präzise Messung nur die Untersuchung eines sehr beschränkten Ausschnitts (‘scope’) der Sprachverwendung erlaubt.

In der Sprachtestforschung existiert bereits seit einiger Zeit ein Bewusstsein für die Einsicht, dass sich die Entwicklung von Tests auf ein möglichst genaues Verständnis der gemessenen Attribute stützen muss, wenn die Beengung des zuvor beschriebenen Dilemmas zukünftig aufgelöst werden soll.

Die valide Operationalisierung dieses theoretischen Verständnisses hängt wiederum von der Frage ab, inwiefern die Qualität der Interaktion zwischen den

Testteilnehmern² und den Anforderungen des Tests dem für die Testaufgaben spezifizierten Fähigkeitskonstrukt entspricht (vgl. Alderson & Banerjee 2002; Bachman 2000a; Brindley 1998b; Freedle & Kostin 1999; Green 1998; Grotjahn & Kleppin 2001; Lazaraton 2002; Mislevy 2008; Weir & Shaw 2008).

Die Forschungsgegenstände der hier dargelegten Studie – das Hörverstehen (Kapitel 2) und die Validität von Sprachtestaufgaben (Kapitel 3) – nehmen im wissenschaftlichen Diskurs paradoxe Positionen ein. So klagt Messick (1992: 89), auf dessen grundlegende Arbeiten sich gegenwärtige Validitätskonzeptionen in der Sprachtestforschung beziehen, “[m]any test makers acknowledge a responsibility for providing general validity evidence of the instrumental value of a test but very few actually do it”. In einer späteren Publikation bestätigt Weir (2005: 11), dass dieses Missverhältnis weiterhin besteht.

Bachman und Alderson beschreiben im Vorwort zu Bucks Handbuch zum Testen des Hörverständens ein ähnliches Paradox: “The assessment of listening abilities is one of the least understood, least developed and yet one of the most important areas of language testing and assessment” (Buck 2001: preface X).

Die ungenügenden Forschungsanstrengungen zur Validität und zum Hörverständen sind hauptsächlich darauf zurückzuführen, dass es sehr anspruchsvoll ist, diese Konstrukte in einem Forschungsdesign zu definieren und methodisch zu operationalisieren. Das Hörverständen ist ein komplexer, interaktiver Konstruktionsprozess, der zu einem großen Teil auf automatisierten mentalen Operationen beruht, die nicht direkt beobachtet werden können. Das Konzept der Validität wurde indessen von den Vertretern verschiedener mit der Messung von Konstrukten befasster Disziplinen in einer Reihe von theoretisierenden Veröffentlichungen zu einem umfassenden und gleichsam in seiner zunehmenden Komplexität kaum erfassbaren Gütekriterium entwickelt (vgl. Bachman 2004, 2005; Chapelle 1998; Kane 2001; Kunan 2000; Messick 1989, 1996; Mislevy 1996).

Die hier vorliegende Studie stützt sich auf eine Validitätskonzeption, die erstens die Bedeutung des theoretisch fundierten Fähigkeitskonstruktur für eine adäquat begründete Kompetenzmessung betont (vgl. Weir 2005) und zweitens annimmt, dass die Ausprägungen der zu messenden Fähigkeit auf Seiten der Testteilnehmer die Ausprägungen ihrer Testergebnisse bestimmen (vgl. Borsboom, van Heerden & Mellenbergh 2004).

2 Aus Gründen der besseren Lesbarkeit werden drei besonders häufig verwendete Begriffe in dieser Arbeit nur in der männlichen Form gebraucht: Die ‚Testteilnehmer‘ werden im Rahmen der Datenerhebung und -analyse als ‚Informanten‘ und als ‚Hörer‘ bzw. ‚Leser‘ bezeichnet. In allen drei Fällen sind jeweils auch ‚Testteilnehmerinnen‘, ‚Informantinnen‘ und ‚Hörerinnen‘ bzw. ‚Leserinnen‘ gemeint.

Das Forschungsdesign der Studie (Kapitel 4) orientiert sich einerseits an den zentralen Annahmen und forschungsmethodischen Implikationen dieser Validitätskonzeption und berücksichtigt andererseits den erweiterten Forschungskontext der Studie. Aus diesen Überlegungen ergibt sich die Notwendigkeit eines mehrperspektivischen Forschungsansatzes, der qualitative und quantitative Methoden der Datenerhebung und -analyse miteinander verbindet.

Das forschungsmethodische Vorgehen (Kapitel 5) besteht aus der Erhebung introspektiver Daten zu den mentalen Prozessen, die mit der Bearbeitung der Hörverstehensaufgaben in der Testsituation einher gehen. Diese Daten wurden vor dem Hintergrund des theoretisch spezifizierten Testkonstrukts und in Verbindung mit den quantitativ ermittelten psychometrischen Parametern der Testaufgaben und den Testleistungen der Testteilnehmer analysiert.

Die Ergebnisse der Studie werden im Hinblick auf die in Kapitel 3 postulierte Validitätskonzeption dargestellt und diskutiert. Sie bestehen erstens aus einer umfassenden Beschreibung der mentalen Prozesse, die von Testteilnehmern während der Bearbeitung der Testaufgaben aktiviert werden (Kapitel 6). Zweitens werden die identifizierten Prozesse hinsichtlich ihrer Relevanz für das spezifizierte Testkonstrukt untersucht. Die Ausprägungen der konstruktrelevanten bzw. -irrelevanten Prozesse werden anschließend in Bezug zur allgemeinen fremdsprachlichen Kompetenz der Informanten gesetzt und mit Blick auf die Bearbeitung unterschiedlicher Hörtextsorten verglichen (Kapitel 7). Drittens werden Zusammenhänge zwischen der Aktivierung von konstruktrelevanten bzw. -irrelevanten Prozessen und der Testleistung der Informanten überprüft (Kapitel 8).