

# Bewertung der Multikompetenz – die Herausforderung eines Paradigmenwechsels

*Philip Herdina*

## 1 Einleitung

Obwohl es Mode geworden ist, von einer „linguistischen Wende“ in der Philosophie und dementsprechend von einem „multilingual turn“ und einem „neuen Paradigma“ in der angewandten Linguistik zu sprechen, scheint es an Klarheit darüber zu mangeln, was dies sowohl in der Sprachtheorie als auch in der Unterrichtspraxis tatsächlich bedeuten soll. Nach Kuhn (1962) besteht ein Paradigma oder eine disziplinäre Matrix aus einem Komplex etablierter theoretischer Begriffe und entsprechender wissenschaftlicher Praktiken, die am ehesten dem entsprechen, was Bourdieu (1992) als „Habitus“ bezeichnet. Darüber hinaus neigen, wiederum nach Kuhn, die eingefahrenen Konzepte und Praktiken der paradigmatischen Normalwissenschaft eher dazu auszusterben, als einen abrupten Wandel zu durchlaufen. Es ist also wahrscheinlicher, dass wir in jedem Forschungsbereich, einschließlich der angewandten Linguistik, eher Paradigmenverschiebung (*paradigm shift*) als wissenschaftliche Revolutionen beobachten können. Lakatos (1970) entwickelte ein Konzept der Normalwissenschaft, das zwischen progressiven und degenerativen Forschungsprogrammen unterscheidet. Progressive Forschungsprogramme können als Programme charakterisiert werden, bei denen die vom Forschungsprogramm angebotenen Lösungen die Anzahl der Probleme übertreffen, während degenerative Programme dadurch gekennzeichnet sind, dass die Anzahl der Forschungsprobleme die vom jeweiligen

Herdina, Philip (2024): Bewertung der Multikompetenz – die Herausforderung eines Paradigmenwechsels. In: Allgäuer-Hackl, Elisabeth; Geiger, Daniel; Hufeisen, Britta; Meirer, Eva & Schlabach, Joachim (Hrsg.) (2024): *Using alli mini Sprocha – bien sûr ! Beiträge zum Schulentwicklungsprojekt „formatio-plurilingual-digital“*.

Darmstadt: Technische Universität sowie Triesen: formatio Privatschule.

<https://www.daf.tu-darmstadt.de/media/daf/dateien/fpd/bewertung.pdf>

Forschungsprogramm angebotenen Lösungen zunehmend überwiegen, was auf die Notwendigkeit eines Paradigmenwechsels in der Forschung hinweist. Sowohl Lakatos als auch Kuhn sind sich darin einig, dass jeder Paradigmenwechsel – auch aufgrund eines Phänomens, das als *confirmation bias* (Bestätigungsfehler) bezeichnet wird – den Widerstand der normalen Wissenschaftler und der etablierten Wissenschaft überwinden muss. Damit es zu einem solchen Wechsel kommt, müssen sich die Forscher der Anhäufung ungelöster Probleme im Forschungsbereich, die mit den bestehenden Instrumenten und Konzepten, die das vorherrschende Paradigma bereitstellt, nicht zu überwinden sind, bewusst werden. Dieser Beitrag wird sich daher auf diejenigen Probleme konzentrieren, die durch einsprachige Tests entstanden sind, bevor er die Kriterien, die an das multilinguale Testen und Bewerten gestellt werden und die Bewertungskriterien spezifizieren, die ein neues Forschungsparadigma erfüllen müsste, determiniert. Es muss auch anerkannt werden, dass der vielzitierte Paradigmenwechsel im Sprachenlehren und Sprachentesten noch nicht stattgefunden hat, da dies nicht nur eine Revolution in den Konzepten, die von der normalen angewandten Sprachwissenschaft verwendet werden, sondern auch eine Veränderung der entsprechenden Praktiken beinhalten müsste, die nachweislich noch nicht stattgefunden hat<sup>1</sup>. Daher kann der gegenwärtige Stand des Wandels, wie festgestellt wurde, bestenfalls als präparadigmatisch bezeichnet werden<sup>2</sup>.

## 2 Die Herausforderung der mehrsprachigen Bewertung

Die Tatsache, dass es bis dato nicht gelungen ist, eine standardisierte Bewertung der Mehrsprachigkeit im Hinblick auf die etablierten Testkriterien der Objektivität, Zuverlässigkeit, Authentizität und vor allem Validität (einschließlich prädiktiver Validität) zu definieren, ist auf zwei Tatsachen zurückzuführen: Zum einen gibt es nach wie vor keinen Konsens darüber, **WAS** und **WER** bewertet werden soll, was sich in der fehlenden präzisen Unterscheidung zwischen „Zweisprachigen“, „Mehrsprachigen“, „Polyglotten“ und „Plurilingualen“ widerspiegelt, und zum

- 
- 1 Obwohl dynamische, vielfältige und konstruktive Diskussionen über mehrsprachiges Lehren und Lernen derzeit im Bereich der Sprachausbildung stattfinden, wird das Phänomen im Bereich der Bewertung völlig übersehen, der Sprache weiterhin als monolinguales, homogenes und oft noch muttersprachliches Konstrukt betrachtet (Shohamy 2011: 419).
  - 2 In diesem Beitrag wird argumentiert, dass trotz der mehrsprachigen Wende im Sprachunterricht Tests und Beurteilungen im Großen und Ganzen weiterhin einsprachige Unterfangen sind, und wir versuchen, auch die Ursachen dafür aufzuzeigen. Die wenigen Forschungsprojekte, die in letzter Zeit in verschiedenen Teilen der Welt entstanden sind, haben keinen gemeinsamen Diskurs entwickelt, und ihre Merkmale sind oft unterschiedlich, auch wenn sie ein ähnliches Ziel anstreben. Der substanzielle Dialog zwischen den Forschern ist unzureichend, was dazu führt, dass das mehrsprachige Testen und Bewerten ein fragmentierter und etwas unzusammenhängender disziplinärer Bereich ist (Dendrinis 2019: 8).

anderen z.B. in der fehlenden Berücksichtigung des Unterschieds zwischen *age of acquisition* (Spracherwerbssalter) und *duration of exposure* (Dauer des Sprachkontakts) sowie des Unterschieds zwischen Erst- und Zweitspracherwerb und Primär- und Sekundärsprache – um nur einige Punkte zu nennen –, wobei diese Begriffe vielfach und fälschlicherweise als synonym verwendet werden<sup>3</sup>. In Beantwortung der Frage, wer getestet wird, muss auch darauf hingewiesen werden, dass Forschungen und Tests, die anhand von Zweisprachigen und für Zweisprachige entwickelt wurden (die zugegebenermaßen zur Verfügung stehen), nicht unmittelbar auf Kandidaten übertragbar sind, die als mehrsprachig oder plurilingual einzustufen sind. Es scheint auch wenig Konsens darüber zu bestehen, **WIE** solche Tests durchgeführt werden sollen, wobei zu bedenken ist, dass die Lösung nicht in der Anwendung mehrerer einsprachiger Tests liegen kann, da Mehrsprachige im Vergleich zu Einsprachigen nach den etablierten einsprachigen Testkriterien notgedrungen benachteiligt sein werden<sup>4</sup>. Dieses Problem wird noch dadurch verschärft, dass die herkömmlichen Sprachdeskriptoren, die in Beurteilungsverfahren verwendet werden, immer noch weitgehend von einer im Wesentlichen strukturalistischen Interpretation von Sprache bestimmt werden, die den Sprachunterricht durchdrungen hat und sich in erster Linie auf Lexik, Morphologie und Syntax konzentriert, anstatt auf die kommunikative Funktion der Sprache und auf Kriterien, die den kommunikativen Erfolg messen sollen<sup>5</sup>. Die Verwendung analytischer Kriterien zur Bestimmung der sprachlichen Leistung kann zum Teil durch die Notwendigkeit motiviert sein, präzisen Teststandards zu entsprechen und die einzelnen Ergebnisse der Bewertungen zu quantifizieren, um das Kriterium der Vergleichbarkeit der Ergebnisse zu erfüllen – was erforderlich zu

- 
- 3 Es gibt zwar gemeinsame Fäden, die sich durch das farbenfrohe Bild ziehen, aber es gibt auch erhebliche Diskrepanzen, z.B.: (a) in der Art und Weise, wie Begriffe wie Mehrsprachigkeit, Plurilingualismus und Heteroglossie verstanden und verwendet werden, (b) in der Art und Weise, wie Konzepte wie Translanguaging, Code-Switching und Mediation verwendet oder vermieden werden, (c) in der Art und Weise, wie die Forschungsteilnehmer betrachtet werden, welche Subjektpositionen sie einnehmen und ob sie als sinnstiftende Akteure wahrgenommen werden, und (d) in den ideologischen und politischen Grundlagen ihrer Ausgangspunkte und ihrer letztendlichen Ziele (Dendrinos 2019: 8). Zur Frage der geeigneten Terminologie siehe: VanPatten/Benati (2010) und Requal (2021).
- 4 Der wichtigste Punkt, der hier gemacht wird, ist, dass, solange solche nationalen Tests von Nicht-Muttersprachlern verlangen, dass sie wie Einheimische sein müssen und in einsprachigen Tests in der dominanten Sprache gemessen werden, diese Schüler nicht in der Lage sein werden, angemessene Ergebnisse zu erzielen (Shohamy 2011: 422).
- 5 Es bestand die Tendenz, das zu prüfende Wissen zu atomisieren und zu dekontextualisieren und Wissensaspekte isoliert zu testen... diese Praxis des Testens separater, einzelner Wissenspunkte, bekannt als diskrete punktuelle Tests (*discrete point testing*), wurde in Theorie und Praxis innerhalb der Psychometrie verstärkt (McNamara 2000: 14). Die Mehrsprachigkeit hätte den Prüfern neue und umfassendere Definitionen der Sprache liefern sollen, die von den Prüfern bei der Erstellung von Sprachtests berücksichtigt werden müssen. Solche Tests werden angesichts der mehrsprachigen Praktiken benötigt, insbesondere im Kontext von Migration und Globalisierung (Shohamy 2011: 420).

sein scheint<sup>6</sup>. Dies ist ein entscheidender Punkt bei der Überprüfung von Mehrsprachigkeit, ebenso wie die Anforderung, dass die Verfahren zur Sprachbewertung vergleichbare Ergebnisse für Einsprachige und Mehrsprachige liefern müssen. Dies stellt uns vor ein „Problem der Erhebung von Mehrsprachigkeit“, da einsprachige Tests nicht gleichermaßen auf ein- und mehrsprachige Personen angewendet werden können und gleichzeitig die Kriterien der Testvalidität usw. erfüllen können. Einsprachige Tests, die auf Mehrsprachige angewandt werden, werden zwangsläufig keine gültigen Ergebnisse für Lernende mit einem Hintergrund in anderen Sprachen liefern, ganz zu schweigen von der heiklen Frage der prädiktiven Validität, die offensichtlich einen wichtigen Aspekt der Sprachbewertung darstellt. Wir müssen nicht nur in der Lage sein festzustellen, wie Personen zu einem bestimmten Zeitpunkt sprachlich abschneiden, sondern auch, was diese Ergebnisse über die Fähigkeit der Prüflinge aussagen, in Zukunft bei Sprachstandserhebungen entsprechend abzuschneiden, d.h. so etwas wie die *zone of proximal development* (ZPD) (Zone der proximalen Entwicklung) zu bestimmen, um ein Konzept von Vygotsky zu verwenden (Lantolf 2000)<sup>7</sup>. Die Notwendigkeit Beurteilungskriterien eines Sprachtests festzulegen, die gültige Prädiktoren für die Sprachentwicklung liefern, legt nahe, dass Sprachtests, die als statische/punktuelle Tests verstanden werden, und nur die Leistung zu einem bestimmten Zeitpunkt messen (d.h. Sprachstandserhebungen), nicht ausreichen, um das Lernpotenzial der Lernenden zu ermitteln. Wir sind daher eher geneigt, den Einsatz dynamischer Bewertungsverfahren zur Ermittlung der Sprachentwicklung zu erwägen, da sich die Kenntnisse in einer bestimmten Sprache (gemessen an der Leistung) zwischen Muttersprachlern und Fremdsprachlern unterscheiden, da diese mit der Sprache, in der sie getestet werden, natürlich viel weniger Kontakt hatten. Wir können also nicht einen ‚Einheitsansatz‘ verfolgen und gleichzeitig einsprachige Beurteilungsverfahren anwenden, in der Erwartung, dass wir für alle Lernenden gültige Ergebnisse erhalten – eine Anforderung, die die traditionelle Sprachbeurteilung offensichtlich erfüllen muss<sup>8</sup>.

- 
- 6 Diese Hauptüberlegungen lassen sich unter dem Begriff der Quantifizierung zusammenfassen und werden in den Sozialwissenschaften häufig als „quantitativer Imperativ“ bezeichnet (Michell 1997, 1999; Niaz 2005). Eine Neubewertung wird als notwendig erachtet, obwohl die Bewertungsdiskussion seit ihren Anfängen über das zwanzigste Jahrhundert bis ins einundzwanzigste Jahrhundert fortgeführt wurde. Von Anfang an wird ausdrücklich darauf hingewiesen, dass die mathematischen Formulierungen, logischen Ableitungen, statistischen Konzeptionen und die Theoreme, Beweise, Axiome und Nebenschlüsse, die sich aus den grundlegenden und bewährten Lehren ergeben, nicht in Frage gestellt werden (Murphy 2011: 77). Siehe auch: Paran & Sercu (2010).
- 7 Theorie der Zone der Proximalen Entwicklung (*zone of proximal development* oder ZPD). Diese Metapher wird von Vygotsky definiert als „der Unterschied zwischen dem, was eine Person erreichen kann, wenn sie allein handelt, und dem, was dieselbe Person erreichen kann, wenn sie mit der Unterstützung einer anderen Person und/oder kultureller Artefakte handelt“ (Lantolf 2000: 17).
- 8 Bezieht sich auf den Begriff der Konstruktvalidität: Damit Tests konstrukt-valide sind, müssen sie auf einem Konstrukt basieren, das dem aktuellen Sprachverständnis und den

Dieses Problem kann auch nicht dadurch gelöst werden, dass man eine Mehrsprachigkeitsprüfung für mehrsprachige Sprecher einführt, während man einsprachige Bewertungsverfahren für Kandidaten beibehält, von denen man annimmt, dass sie einsprachig sind. Diese Herausforderung wird durch die Tatsache verschärft, dass die jeweiligen Populationen von Einsprachigen und/oder Mehrsprachigen keine homogene Gruppe bilden. Bei näherer Betrachtung kann sich herausstellen, dass Probanden, die als einsprachig eingestuft werden, häufig mehr als eine Sprache beherrschen, was sich in der Tatsache widerspiegelt, dass die Mehrheit der Weltbevölkerung als mehrsprachig und nicht als einsprachig angesehen wird, und dass die als mehrsprachig eingestuften Lernenden alle Arten von Sprachenkonstellationen oder Multikompetenzen aufweisen können. Dieses Problem lässt sich an der Notwendigkeit verdeutlichen, zwischen natürlicher Mehrsprachigkeit (die den gleichzeitigen Erwerb von mehr als einer Sprache im Allgemeinen in der Familie oder im lokalen Umfeld als Teil der primären Sozialisation widerspiegelt) und institutionell induzierter oder geforderter bzw. aufgezwungener Mehrsprachigkeit (bei der der Erwerb von mehr als einer Sprache durch lokale oder nationale Bildungsanforderungen bestimmt wird, wie z.B. die Notwendigkeit, eine (curricular vorgeschriebene) zweite Sprache bereits im Vorschulalter zu erwerben) zu unterscheiden, um zu versuchen, die verschiedenen Arten von Mehrsprachigkeit und ihre entsprechenden Unterschiede zu kategorisieren<sup>9</sup>. Dieses Problem kann nur durch die Einführung einer Sprachbewertung auf mehrsprachiger Basis für alle, die gewährleistet, dass die erzielten Ergebnisse tatsächlich vergleichbar sind, gelöst werden. Natürlich gibt diese Forderung nur den Rahmen für die Sprachbewertung vor und beantwortet nicht die Frage, **WIE** solche Bewertungsverfahren durchgeführt werden sollen und noch weniger, wie diese Ergebnisse zu bewerten und zu vergleichen sind. Ohne die Fähigkeit, die beiden letztgenannten Kriterien zu erfüllen, wird die mehrsprachige Bewertung jedoch nicht den Bedürfnissen der jeweiligen Bildungseinrichtungen entsprechen, die die Ergebnisse der Sprachlehrpläne bestimmen und standardisierte Kriterien für die Sprachbewertung bereitstellen müssen<sup>10</sup>.

---

Sprachtheorien (Grundlagen) entspricht. So die hier aufgestellte Behauptung (Shohamy 2011: 420).

- 9 Simultan-Zweisprachige lernen beide Sprachen vor dem Alter von 3 Jahren (Paradis 2010). Im Gegensatz dazu ist bei sequentiellen Bilingualen die erste Sprache (L1) im Allgemeinen gut entwickelt, bevor sie eine zweite Sprache (L2) lernen (Paradis 2010). Die dominante Sprache ist die Sprache, mit der sie am meisten in Berührung gekommen sind. Die Entwicklung in der zweiten Sprache ist nicht vergleichbar mit der Sprachentwicklung gleichaltriger monolingualer Gleichaltriger (Bedore/Pena 2008), und die Entwicklung in der L1 kann ins Stocken geraten, wenn die L2 komplexer wird. Der Sprachkontakt mit der L2 und die soziokulturellen Erfahrungen der Kinder beeinflussen die Sprachentwicklung (Nelson 1990) (Glazzard 2017: 2).
- 10 Mehrsprachige Beurteilungen werden im Unterricht häufig in Form von formativen und klassenbasierten Beurteilungen eingesetzt. Allerdings sind Fälle, in denen mehrsprachige Bewertungen in standardisierten und groß angelegten Testsituationen durchgeführt werden, sehr selten (Garcia/Li 2014). Da mehrsprachige Dimensionen komplexe Merkmale umfassen und die Sprache als ein dynamisches System betrachtet werden muss (Canagarajah 2011;

Im Gegensatz zur ein- oder mehrsprachigen Kompetenz ist die plurilinguale Kompetenz wie folgt definiert:

Mehrsprachige Kompetenz, wie sie im GERS definiert ist, ... bezieht sich auf die Fähigkeit der Lernenden, effektiv zu kommunizieren, indem sie ihr gesamtes sprachliches Repertoire nutzen, um Bedeutungen auf eine Art und Weise zu schaffen, bei der die Gesamtheit der semiotischen Ressourcen der Lernenden (die Sprachen, Diskurse, Dialekte und Modalitäten), über die sie verfügen, für eine effektive interkulturelle Kommunikation ins Spiel gebracht werden und somit nicht als separate Einheiten, sondern als miteinander verbunden und interaktiv behandelt werden (Dendrinos 2019: 2).

Wenn wir die Tatsache akzeptieren, dass einsprachige Tests nicht auf mehrsprachige Individuen angewandt werden können, da die erzielten Ergebnisse weder valide Messwerte für die Sprachleistung noch valide Prädiktoren für die Sprachentwicklung bei Sprechern von mehr als einer Sprache darstellen, und wenn wir zudem akzeptieren, dass wir in der Lage sein müssen, die Sprachbewertung so anzuwenden, dass die Ergebnisse sowohl den Testanforderungen als auch den definierten Sprachstandards entsprechen, die objektiv spezifizierbar sind, müssen wir zu dem Schluss kommen, dass die standardisierte Sprachbewertung nicht einfach durch mehrsprachige Bewertungsverfahren ergänzt werden kann<sup>11</sup>. D.h. die jeweiligen Tests müssen uns insgesamt vergleichbare Messungen liefern, die auf vergleichbaren Maßstäben beruhen. Dies ist offensichtlich das Ziel und der Grund für den Erfolg des Gemeinsamen europäischen Referenzrahmens (GERS) bei der Festlegung von Normen sowohl für Sprachlehrer als auch für Sprachprüfer<sup>12</sup>. Alles andere würde nämlich darauf hinauslaufen eine Methode anzuwenden, die Anthropologen und Soziologen als *Bricolage* bezeichnen würden, d.h. einem mehrsprachigkeitsorientierten Flickens am bestehenden System<sup>13</sup>. Es ist offensichtlich, dass eine

---

Garcia/Li 2014), ist die Entwicklung von Tests und Rubriken, die auf den mehrsprachigen Konstrukten von Sprache basieren, eine Herausforderung (Shohamy 2011). McNamara (2015) weist darauf hin, dass zu den Herausforderungen bei der Durchführung mehrsprachiger Bewertungen der Mangel an mehrsprachiger Kompetenz der Lehrkräfte, an Ressourcen für die Entwicklung, Verwaltung und Auswertung dynamischer Beurteilungen sowie die Überwindung des Widerstands der vorherrschenden monolingualen Ideologie hinzukommt (Wang/Li 2020).

- 11 Darüber hinaus gibt es nur sehr wenige bildungspolitische Maßnahmen, die mehrsprachige Beurteilungspraktiken begünstigen, keine curricularen Anforderungen für zwei- oder mehrsprachige Tests und einen gravierenden Mangel an evidenzbasierter Forschung, die für den positiven Rückkopplungseffekt mehrsprachiger Tests auf die mehrsprachige Bildung spricht.
- 12 Sogar das University of Cambridge Language Examination Syndicate (UCLES) musste seine Sprachprüfungskriterien anpassen, um den vom CEFR festgelegten Normen und Standards zu entsprechen.
- 13 Mehrsprachige Kinder können beim Erreichen typischer Entwicklungsmeilensteine ein breites Leistungsspektrum aufweisen (Goral/Conner 2013), und die Leistung kann durch Variablen wie Alter, Sprachstatus, Sprachinput, Sprachkontaktmuster (sequentiell oder simultan) und

vollständige Überarbeitung der etablierten Normen für die Sprachbewertung erforderlich ist, um die folgenden Punkte zu berücksichtigen:

Erstens müssen wir in der Lage sein zu spezifizieren, WAS bewertet werden soll. Wie Shohamy (2012)<sup>14</sup> darlegt, fehlt der Bewertung von Multikompetenz noch immer eine angemessene theoretische Grundlage. Solange wir nicht spezifizieren können, was wir sowohl unter Sprache als solcher als auch unter Sprachkompetenz im Kontext von Mehrsprachigkeit verstehen, können wir natürlich auch keine angemessenen Maßstäbe für Multikompetenz entwickeln. Ein Überblick über die im deutschsprachigen Raum existierenden Modelle zur Beurteilung von Mehrsprachigkeit hinterlässt einen nachhaltigen Eindruck von der Divergenz dessen, was tatsächlich als adäquates Kriterium für die Beurteilung von Sprachkompetenz angesehen wird (vgl. Schneiders/Komar 2007). Es ist uns demnach noch nicht gelungen zu bestimmen, was wir von mehrsprachigen Individuen auf den verschiedenen Stufen der Entwicklung von Mehrsprachigkeit erwarten. Offensichtlich geben der GERS und die PISA-Richtlinien keine befriedigenden Antworten auf diese Frage, da sie überwiegend auf monolingualen Auffassungen von Sprachkompetenz beruhen<sup>15</sup>. Wir stellen zudem fest, dass Programme wie *No Child Left Behind* (NCLB 2001) noch immer so sehr auf einsprachige Annahmen ausgerichtet sind, dass das Programm entgegen der öffentlichen Meinung dafür sorgt, dass mehrsprachige Schüler tatsächlich zurückbleiben. Der GER-Ergänzungsband versucht zwar, das Problem der Mehrsprachigkeit zu thematisieren, bietet aber keine angemessenen Bewertungskriterien<sup>16</sup>. Wie zu beobachten ist, unterliegen zudem die meisten Bewertungsverfahren dem „Linguismus“, d.h. der ausschließlichen Konzentration auf sprachliche Merkmale bei der Bewertung zum Nachteil anderer Faktoren, die

---

Sprachkontaktfrequenz beeinflusst werden (Goral/Conner 2013). Daher ist die Unterscheidung zwischen mehrsprachigen Kindern mit typischer und atypischer Sprachentwicklung inhärent komplex (Goral/Conner 2013) und unsicher (Anderson/Marquez 2009), da es Überschneidungen bei Fehlern zwischen den beiden Gruppen geben kann, zum Beispiel bei der Verwendung von Artikeln (Glazzard 2017: 4).

- 14 Die Sprachprüfung als Disziplin ist an einem Punkt angelangt, an dem sie ihre zugrundeliegenden Konstrukte im Einklang mit den Entwicklungen im Bereich des Sprachenlernens, -lehrens und SE überprüfen muss. Es ist an der Zeit, dass die Sprachprofis über ein Konstrukt nachdenken, das dem aktuellen Stand der Forschung und des Wissens besser entspricht und nicht den Ideologien der Nationalstaaten dient, die vor allem an homogenen und monolingualen Bürgern interessiert sind (Shohamy 2011: 428).
- 15 Das Programm zur internationalen Schülerbeurteilung (PISA) ... verlangt von allen Schülern, dass sie ihre Leistungen ausschließlich in den Landessprachen der Länder erbringen. Schließlich ist es wichtig, die weit verbreiteten Instrumente wie den Gemeinsamen Europäischen Referenzrahmen (GER) und die Richtlinien des American Council on the Teaching of Foreign Languages (ACTFL) zu untersuchen ... diese basieren ausschließlich auf einsprachigen, homogenen Konstrukten, die es anderen Sprachen verbieten, sich „einzuschmuggeln“ oder „aufzublühen“ (Shohamy 2011: 421).
- 16 Siehe den kürzlich vom Europarat veröffentlichten Begleitband, Gemeinsamer europäischer Referenzrahmen für Sprachen: Lernen, Lehren, Bewerten - Begleitband, Europarat Verlag, Straßburg (2020).

die gemessenen Ergebnisse beeinflussen könnten, was wiederum zu Fragen hinsichtlich der Gültigkeit der angewandten Bewertungsverfahren führt. Wie von Vygotsky behauptet ist das, was wir im Laufe der Sprachentwicklung beobachten, ein Hinweis auf eine Koevolution von Sprache und Kognition, bei der die Sprache zunehmend eine kognitive Funktion erfüllt und umgekehrt kognitive Prozesse den Prozess des Spracherwerbs zunehmend beeinflussen, was sich in der Entwicklung eines größeren metalinguistischen Bewusstseins bei Lernenden mehrerer Sprachen zeigt (vgl. Jessner 2008). Die kognitive Funktion der Sprache in primär nicht-sprachlichen Fächern wie Mathematik zeigt sich auch in den von Wright und Xiaoshi (2008) zitierten Befunden, wonach Sprachdefizite nicht-sprachliche kognitiven Leistungen ebenfalls beeinträchtigen<sup>17</sup>. Ein befriedigendes Beurteilungsverfahren muss daher sowohl die kognitive als auch die sprachliche Leistung berücksichtigen und ebenso unterscheiden können, um dem Anspruch der Testvalidität gerecht zu werden. Es gibt zahlreiche Belege dafür, dass bei einsprachigen Tests, die auf Mehrsprachige angewandt werden, häufig kognitive und sprachliche Aspekte miteinander vermengt werden, was dazu führt, dass intellektuell Hochbegabte als kognitiv beeinträchtigt eingestuft werden und Prüflinge aufgrund mangelnden Verständnisses der jeweiligen Aufgabenstellung in Sprachtests versagen<sup>18</sup>. Diese Problematik macht deutlich, dass es notwendig ist, zwischen *delayed* (sprachentwicklungsverzögerten) und *impaired* (beeinträchtigten) Spracherwerbsprozessen zu unterscheiden (vgl. SLI oder *specific language impairment*).

Zweitens müssen wir wissen, **WER** beurteilt wird. Wie von Paikeday (1985) und Kramsch (1998)<sup>19</sup> vorgeschlagen, müssen wir die Vorstellung des idealen Muttersprachlers als Norm, aber auch die des idealen Sprachenlerner, wie sie im Konzept des Erst- und Zweitspracherwerbs und in der SLA-Forschung impliziert ist, zunehmend ablehnen, da diese davon ausgeht, dass es einen standardisierten

---

17 Wright und Li (2008) wiesen beispielsweise nach, dass Schüler, die in ihrem Heimatland gute Leistungen in Mathematik erbrachten, im neuen Land aufgrund der neuen Sprache schlechter abschnitten (Shohamy 2011: 425).

18 Am häufigsten werden standardisierte, normbezogene Tests verwendet, die jedoch in der Regel nicht für zweisprachige Probandengruppen standardisiert sind und von SLTs mit dem Hinweis verwendet werden, dass die Normen möglicherweise nicht auf eine bestimmte Person anwendbar sind, für die Englisch eine zweite oder zusätzliche Sprache ist. Die Ergebnisse werden qualitativ oder deskriptiv verwendet, ohne jedoch zu berücksichtigen, dass die Erhebung der Daten von der kulturellen Vertrautheit mit den Materialien abhängt und dass die Anweisungen möglicherweise nicht perfekt verstanden werden oder das Kind eine ungewohnte Aufgabe ausführen muss (De Lamo White/Jin 2011). Insbesondere der Wortschatz und die erzählerischen Fähigkeiten sind mit den früheren Erfahrungen eines Kindes verknüpft, was dazu führt, dass die Testergebnisse eher die Lebenserfahrungen und den sozioökonomischen Status widerspiegeln als die Sprachkenntnisse (Lidz/Peña 2009) (Hasson et al. 2012: 58).

19 Der Muttersprachensprecher der Linguisten und Sprachlehrer ist in der Tat eine Abstraktion, die auf willkürlich gewählten Merkmalen des Aussehens und des Auftretens beruht ... der Muttersprachler ist außerdem eine monolinguale, monokulturelle Abstraktion: Er/sie ist jemand, der nur seine/ihre standardisierte Muttersprache spricht und in einer standardisierten nationalen Kultur lebt (Kramsch 1998: 79-78). Vgl. auch Paikeday (1985).

(und häufig linearen) Prozess des Erst- und Zweitspracherwerbs und somit vorhersehbare und unveränderliche Ergebnisse beim Erst- und Zweitspracherwerb gibt, wie von der sogenannten kartesischen Linguistik<sup>20</sup> (vgl. Chomsky 2009) angenommen. Der ideale Muttersprachler kann nicht länger als Maßstab für die erwartete Sprachleistung dienen, und auch der Sprachlerner kann nicht nur nach altersgemäßen Spracherwerbsstufen gemessen werden (vgl. die Hypothese der natürlichen Ordnung des Spracherwerbsprozesses in Krashen 1987)<sup>21</sup>. Forscher, die Sprache als ein komplexes dynamisches System<sup>22</sup> betrachten, müssen berücksichtigen, dass die jeweilige Sprachleistung vom Anfangszustand abhängt und dass die jeweils erworbenen Sprachen miteinander interagieren, was zu emergenten Eigenschaften wie metalinguistischem Bewusstsein führt, die ihrerseits die Dynamik des komplexen Sprachsystems beeinflussen (vgl. Herdina/Jessner 2002). Daraus ergibt sich die Einschätzung der möglichen *zone of proximal development* (ZPD) für jeden Sprachlerner und damit das Ausmaß des *Scaffolding*, das erforderlich ist, um das sprachliche Potenzial des jeweiligen Sprachlerner zu realisieren. An dieser Stelle könnte man erwarten, dass *dynamic assessment* (DA) (dynamische Bewertung) eine fertige Antwort auf das genannte Problem liefert<sup>23</sup>. Bei näherer Betrachtung müssen wir jedoch feststellen, dass die von DA etablierten dynamischen Beurteilungsmethoden in mehr als einer Hinsicht hinter unseren Erwartungen zurückbleiben.

Schließlich müssen wir die Frage beantworten, **WIE** die Bewertung durchgeführt werden soll. Es scheint allgemeines Einvernehmen darüber zu herrschen, dass eine bloß statische Beurteilung (d.h. Sprachstandserhebung) unzureichend ist. Vor allem die dynamische Beurteilung (DA) behebt einige der Unzulänglichkeiten der

---

20 Beurteilungen der mehrsprachigen Kompetenz, die eine monoglossische Perspektive widerspiegeln, „versuchen, die ultimative muttersprachliche Kompetenz in allen Sprachen zu berücksichtigen“ und „gehen davon aus, dass die Mehrsprachigkeit die Summe der muttersprachlichen monolingualen Kompetenz in jeder Sprache ist“ (Stavans/Hoffmann 2015: 157) Lopez/Turkan/Guzman-Orth (2016).

21 Der monoglossischen Perspektive liegt die Annahme zugrunde, dass es keinen Unterschied zwischen der Sprachentwicklung einsprachiger und mehrsprachiger Menschen gibt. So haben mehrsprachige Menschen Zugang zu mehreren voneinander getrennten Sprachsystemen, die sich linear entwickeln (Grosjean 1989). Daher werden Sprachen, so wie sie in den Köpfen zweier oder mehrsprachiger Individuen existieren, als getrennte Einheiten und nicht als ein einheitliches System behandelt (Lopez/Turkan/Guzman-Orth 2016: 3) Vgl. auch: Krashen (1987).

22 Sprache ist mehr als der scheinbar einfache Klangstrom, der aus der Zunge des Muttersprachlers fließt; sie ist mehr, als der Muttersprachler denkt, dass sie ist. Sie ist ein komplexes Kommunikationssystem mit verschiedenen Komplexitätsstufen, das eine komplizierte Auswahl und Anordnung von Bedeutungen, Lauten und größeren Einheiten und Anordnungen beinhaltet (Lado 1961: 2).

23 Laut Goral und Conner (2013: 132) "ist die dynamische Beurteilung ein vielversprechendes Instrument zur Unterscheidung zwischen mehrsprachigen Kindern mit PLI (*Primary Language Impairment*) und [denen mit] TLD (*Typical Language Development*). Eine statische Beurteilung ist möglicherweise nicht genau, da mehrsprachige Kinder eine große Bandbreite an Leistungen in ihren aktuellen sprachlichen Fähigkeiten aufweisen können (Goral/Conner 2013) Glazzard (2017: 4).

statischen Beurteilung bei der Vorhersage der Sprachentwicklung. Die dynamische Beurteilung behebt diesen Mangel durch die Einführung eines so genannten „Sandwich“-Verfahrens, das aus einer Beurteilungssequenz bestehend aus Vortest, Intervention, Training und Posttest besteht und sich nicht so sehr auf den Sprachstand eines Lernenden zu einem bestimmten Zeitpunkt konzentriert, sondern vielmehr auf die Fähigkeit des Lernenden, seinen Sprachstand zu verbessern. Dies geschieht nach einer Intervention durch den *more knowledgeable other* MKO (wissenden Anderen, vgl. Vygotsky), in der Regel dem Sprachlehrer in einem Klassenzimmer. Hier liegt der Schwerpunkt auf dem Grad der Veränderung in der Sprachkompetenz innerhalb eines bestimmten Zeitintervalls und nicht auf der Kenntnis der Sprache oder dem, was zu einem bestimmten Zeitpunkt als Sprachkompetenz anzusehen ist, da dies nicht als besonders aussagekräftiges Testkonstrukt angesehen wird. Leider gibt die DA nur wenige Antworten auf das Problem der Beurteilung der Multikompetenz und des Sprachpotenzials von Sprechern anderer Sprachen. Sieht man einmal von den interventionistischen Aspekten der Beurteilung ab, die durch das *Scaffolding* vorgeschlagen werden, so bietet das Dynamic Assessment (DA) kein ausreichendes Problemlösungspotenzial. Andererseits liegt ein positiver Aspekt in der stärkeren Fokussierung auf die kommunikative Kompetenz und dem größeren Potenzial für den Einsatz authentischer Beurteilungsverfahren, die von DA vorgeschlagen werden.

### 3 Paradigmenwechsel

Im Gegensatz zur impliziten Annahme vieler Forschungsansätze zur Mehrsprachigkeit und Multikompetenz müssen wir bedenken, dass der vorausgesetzte multikompetente Sprecher nicht einer homogenen Population von Mehrsprachigen entspricht. Es gibt so viele unterschiedliche Wege und Mittel, um mehrsprachig zu werden, dass wir nicht von einem gemeinsamen Typus oder Lehrplan für (multikompetente) Sprachenlerner ausgehen können und daher auch kein gültiges Konstrukt wie den des idealen Mehrsprachensprechers entwickeln können. So bequem es für die Theorie der Sprachbewertung auch erscheinen mag, wir müssen die Vorstellung vom Sprachensprecher als einer Abstraktion, einem kartesischen *homo loquens* in Analogie zur kartesischen *res cogitans*, die von der *res extensa* – der äußeren Welt der Objekte - getrennt zu halten ist, über Bord werfen, d.h. die Vorstellung, dass Sprachwissen gleichbedeutend ist mit *Words in the Mind* (Aitchison 2002), und statt dessen den Begriff der situierten Sprache in Analogie zum Konzept der *situated cognition* (situierten Kognition) anzunehmen. Um die jeweilige Sprachkompetenz angemessen beurteilen zu können, müssen wir auch intervenierende Variablen berücksichtigen, die zur Komplexität des Sprachsystems beitragen, wie z.B. das Lebensalter zum Zeitpunkt des Spracherwerbs, die Dauer des Sprachkontakts, die Intensität des

Sprachkontakts, die relative Sprachdominanz, die Sprachdrift, die kognitive Reifung im Vergleich zur sprachlichen Reifung, die sprachübergreifende Interaktion und die beteiligten Sprachen, um die relative Komplexität der bewerteten Sprache(n) für den Lernenden in Abhängigkeit von den bereits erworbenen Sprachen zu ermitteln usw. Der unterschiedliche Komplexitätsgrad der Test- oder Lernaufgaben hängt von den (sprachlichen) Vorkenntnissen des Lernenden ab, wie die sprachspezifischen Anforderungen an die Lese- und Schreibfähigkeit zeigen. Der Unterschied lässt sich anhand der Herausforderung verdeutlichen, die der Erwerb der Lese- und Schreibfähigkeit im Chinesischen mit seinen maximal ca. 106.000 Zeichen darstellt, im Vergleich zu der leichteren Aufgabe, die Lese- und Schreibfähigkeit in den meisten europäischen Sprachen mit einem Alphabet aus 20 bis 26 Buchstaben zu erlangen. Auch die Herausforderungen der Phonemik hängen von den Rechtschreibkonventionen ab, wobei es erhebliche Unterschiede zwischen dem Englischen und anderen Sprachen wie Koreanisch oder Arabisch gibt, die im Gegensatz zum Englischen eher nach dem Prinzip der phonemischen Schreibweise vorgehen. Auch der leichter zu quantifizierende Erwerb des Wortschatzes kann zwischen den einzelnen Sprachen erheblich variieren, vom klassischen Latein mit 40–50.000 Wörtern bis zum Koreanischen mit einem Wortschatz von über einer Million, auch wenn die angegebenen Zahlen je nach Quelle variieren können. Es liegt daher auf der Hand, dass die Deskriptoren, die zur Messung des Fortschritts beim Spracherwerb verwendet werden, sprachenspezifisch sein müssen und daher nicht universell über die ‚Grenzen‘ von Sprachen hinweg anwendbar sind. Die relative Komplexität der Sprachenlernaufgaben hängt von den bereits erworbenen Sprachen ab, wie sich leicht am Unterschied zwischen Sprachen, die dieselbe Schrift teilen, und Sprachen, deren Schriftsysteme und Schreibkonventionen sich erheblich unterscheiden, erkennen lässt.

Es wird vorgeschlagen, anstelle einer monolithischen Sichtweise von Sprachkompetenz, die entsprechend bewertet werden kann, eine gebündelte Sichtweise von Kompetenz anzunehmen, um das Konzept der Multikompetenz angemessen zu reflektieren. Eine Bündeltheorie der Multikompetenz würde es uns ermöglichen, einen analytischen Ansatz zu wählen, der für die Testtheorie erforderlich ist, und gleichzeitig eine ganzheitliche Sicht auf die Sprachenkompetenz oder -fertigkeit, die bewertet wird, beizubehalten. Eine Beurteilung von Multikompetenz muss die Anforderungen der Konstruktvalidität und der prädiktiven Validität, der Reliabilität, der Authentizität, der Objektivität und im Wesentlichen der Quantifizierbarkeit erfüllen, um die Vergleichbarkeit der

Ergebnisse und damit die faire und gerechte Zuordnung von Sprachqualifikationen zu gewährleisten<sup>24</sup>. Dies stellt natürlich eine erhebliche Herausforderung dar, wie Aquiles Iglesias betont:

Obwohl wir in verschiedenen Bereichen einige Fortschritte gemacht haben, sind unsere theoretischen Arbeitsmodelle nach wie vor zu einfach und erfassen nicht das komplexe dynamische System, in dem das zweisprachige Kind Sprache lernt. Diese komplexeren, datengestützten Modelle würden uns ein besseres Verständnis dafür vermitteln, wie die große Variabilität in der zweisprachigen Bevölkerung zu verstehen ist, und würden uns den besten Weg weisen, auf die Bedürfnisse sprachbehinderter zweisprachiger Lerner einzugehen.<sup>25</sup>

## 4 Schlussfolgerung

Das Hauptaugenmerk der Schlussfolgerungen muss auf den Mängeln der bestehenden Praktiken (vgl. *best practice*) bei der Sprachprüfung und -bewertung und der daraus resultierenden dringenden Notwendigkeit eines Paradigmenwechsels liegen.

Zur Frage, **WER** beurteilt wird: Solange einsprachige Tests auf Lernende angewandt werden, die wir sowohl als ein- wie mehrsprachig einstufen, werden diese Tests kein zufriedenstellendes Maß an Validität erreichen (d.h. Testvalidität, Konstruktvalidität, prädiktive Validität). Wenn man Ein- und Mehrsprachen-sprecher nach denselben (einsprachigen) Standards testet, entsteht der falsche Eindruck, dass einsprachige Prüflinge bessere Leistungen erbringen, als dies tatsächlich der Fall ist, und dass mehrsprachige Personen die erwarteten sprachlichen Ergebnisse nicht erzielen und daher zu den Leistungsschwachen gehören, die irgendeine Art von Abhilfemaßnahme, wie Einbindung in einen Förderunterricht, benötigen<sup>26</sup>. Durch den Vergleich erhalten wir also ein verzerrtes Bild von den sprachlichen Ressourcen, die einsprachigen und mehrsprachigen Lernern zur Verfügung stehen.

Da einsprachige Tests per definitionem nicht mehrsprachige Ressourcen testen, und zwangsläufig andere (Fremd-)Sprachen ignorieren, können sie außerdem

---

24 Die oben genannten Daten und Argumente zeigen, dass die weit verbreiteten Tests, die auf einem monolingualen Konstrukt beruhen, zu verzerrten und oft diskriminierenden Ergebnissen führen, da viele Faktoren nicht berücksichtigt werden (Shohamy 2011: 425).

25 Aquiles Iglesias, Language Impairment in Bilingual Children: From Theory to Practice, In: Seminars in Speech and Language, 36 (02), 87, Thieme: New York 2015.

26 Die mehrsprachige Bevölkerung ist insofern heterogen, als die individuellen Umstände in Bezug auf das Alter des Zweitspracherwerbs und das Ausmaß des Kontakts mit der Sprache variieren (Paradis et al. 2011). Es gibt einige Unterschiede in der Art und Weise, wie einsprachige und mehrsprachige Kinder Sprache und Sprechfähigkeit erwerben (Grech/McLeod 2012), und diese Unterschiede können zu falschen Annahmen führen, dass mehrsprachige Kinder eine Sprach- oder Sprechstörung aufweisen (Glazzard 2017: 2).

keine wesentlichen Informationen über die getesteten Lernenden liefern. Auf der Grundlage der Ergebnisse können wir also nicht einmal ausschließen, dass unsere so genannte einsprachige Gruppe nicht Lernende enthält, die als mehrsprachig eingestuft werden sollten, und dass die mehrsprachige Gruppe Lernende enthält, die tatsächlich weniger mehrsprachig sind als einige der Einsprachigen. Aus dieser Perspektive sagen uns die konventionellen und etablierten Sprachtestverfahren nicht genug über die getesteten Lernenden, um die erzielten Ergebnisse angemessen zu interpretieren und entsprechende sinnvolle Maßnahmen zu ergreifen.

Wir müssen auch anerkennen, dass wir bei unserem Versuch, zwischen Einsprachensprechern und Mehrsprachensprechern zu unterscheiden, einen Kategorienfehler begehen, wie ihn Gilbert Ryle (1973) beschrieben hat, indem wir davon ausgehen, dass man entweder einsprachig sein muss, damit der Test anwendbar ist, oder mehrsprachig, damit der Test ungültig ist. Das Konzept der Mehrsprachigkeit legt nahe, dass Einsprachige und Mehrsprachige nicht so sehr getrennte Kategorien bilden, sondern vielmehr Teil eines Kontinuums bilden, das Grade der Mehrsprachigkeit ausdrückt<sup>27</sup>. Dies würde uns zu dem Schluss führen, dass einsprachige Tests in der Tat auf keine der beiden Gruppen anwendbar sind. Dieses Argument ist stärker als das in der Literatur zu findende Argument, dass einsprachige Tests nicht auf Mehrsprachige anwendbar sind, da sie die anderen Sprachressourcen der Lernenden nicht berücksichtigen. Umgekehrt gilt auch: Einsprachige Tests, die keine anderen Sprachen berücksichtigen, sind nicht auf Lerner anwendbar, die als einsprachig eingestuft werden, da diese Tests nicht feststellen können, ob diese Sprecher nicht tatsächlich mehrsprachig sind. Dieser Sachverhalt spiegelt die Tatsache, dass nicht Sprecher, sondern primär Sprachen getestet werden.

In dem Bemühen, quantifizierbare und vergleichbare Ergebnisse zu erzielen, stützen sich die meisten Tests auf Deskriptoren und Konzepte, die aus der curricular verankerten Sprachtheorie abgeleitet sind (z.B. präskriptive Grammatik, strukturalistische Linguistik oder gelegentlich Universalgrammatik), die uns ebenso wenig über die Fähigkeit des Einzelnen, in Alltagssituationen effektiv zu kommunizieren, aussagen, wie eine Antwort auf den Sinn des Lebens liefert (für diejenigen unter uns, die auf Douglas Adams stehen)<sup>28</sup>. Man beachte die behauptete Divergenz zwischen den von Cummins identifizierten grundlegenden interpersonellen kommunikativen Fähigkeiten (BICS), die den Sprachbedürfnissen

---

27 Shohamy (2011) vertritt eine heteroglossische Sichtweise, die den Standpunkt vertritt, dass das Sprachrepertoire einer mehrsprachigen Person als ein einheitliches System funktioniert, und stellt die Bewertung von Mehrsprachigkeit auf ein Kontinuum (Lopez/Turkan/Guzman-Orth 2016: 5).

28 Ein Sprachtest ist nur so gut wie die Sprachtheorie, auf der er basiert, und in diesem Bereich der theoretischen Untersuchung des Wesens von Sprache und Kommunikation müssen wir unsere Fähigkeit entwickeln, die nächste Frage zu stellen (McNamara 2000: 86).

von Nicht-Muttersprachlern entsprechen, und den kognitiven akademischen Sprachkenntnissen (CALP), die den Schwerpunkt des standardsprachlichen Lehrplans stellen (z.B. im Deutschen als Bildungssprache bezeichnet) (Cummins 2007: 487-499). Eine wirksame mehrsprachige Wende muss daher mit einer entsprechenden pragmatischen Wende in den Methoden der Sprachbewertung einhergehen, da die Standardtestaufgaben weit von der alltäglichen Spracherfahrung der meisten mehrsprachigen Sprecher entfernt sind und daher das Kriterium der Authentizität nicht erfüllen<sup>29</sup>. Vergleiche die Probleme die als *teaching to the test* und als *washback* Effekt identifiziert werden (vgl. Tsangari 2011; Cheng 2005).

Als nächstes stellt sich das Problem **WIE** die Tests durchgeführt werden sollen. Statische/punktuelle Tests oder so genannte „Sprachstandserhebungen“, die von Bildungseinrichtungen in Deutschland oder Österreich eingesetzt werden, lassen den sprachlichen Hintergrund und die Sozialisation des Lernenden außer Acht (was sowohl auf mangelnde Kenntnisse über die vom Sprecher gesprochenen Sprachen als auch auf mangelnde spezifische Kenntnisse über die unterschiedliche Beschaffenheit dieser Sprachen zurückzuführen ist) und geben daher nur wenig oder gar keinen Aufschluss über die sprachlichen Ressourcen oder das Sprachenpotenzial des jeweiligen Lernenden, ebenso wenig wie eine Momentaufnahme etwas über die Flugbahn eines sich bewegenden Objekts aussagen kann<sup>30</sup>. Eine effektive Bewertung muss folglich dynamisch sein, wobei zu berücksichtigen ist, dass bei jedem Testverfahren der Grad der Veränderung im Laufe der Zeit und nicht das Leistungsniveau als gültiger Prädiktor für die Sprachentwicklung untersucht wird.

In Analogie zu der Argumentation David Humes, der behauptete, dass jedes Buch, das weder mathematische Beweise noch empirische Erkenntnisse enthält, in die Flammen gehört, können wir behaupten, dass jedes Testverfahren, das nur monolinguale Standards verwendet, traditionelle Deskriptoren einsetzt, anstatt sich auf die kommunikative Multikompetenz zu konzentrieren, und nur die Sprachleistung und nicht die Sprachentwicklung testet, auf den Müllhaufen der sprachdidaktischen Forschungsgeschichte gehört. In Anbetracht des so

---

29 Mehrsprachige Beurteilungen sollten Sprachpraktiken widerspiegeln, die dynamisch und fließend sind, und somit den Testteilnehmern die Möglichkeit geben, Sprachmerkmale aus ihrem Sprachenrepertoire so auszuwählen, dass sie ihren kommunikativen Bedürfnissen entsprechen (Garcia und Wei 2014, Lopez et al. 2014). Dazu ist ein Paradigmenwechsel – von einer monolingualen/monoglossischen/fraktionellen Sichtweise zu einer mehrsprachigen/mehrsprachigen/holistischen Sichtweise – erforderlich (Shohamy 2013) Lopez/Turkan/Guzman-Orth (2016: 8).

30 In gewisser Weise ist ein guter Test wie ein Experiment, in dem Sinne, dass er alle fremden Variationsquellen eliminieren oder zumindest konstant halten muss. Wir wollen, dass unsere Tests nur die spezielle Art von Variation im Wissen oder in den Fähigkeiten widerspiegeln, an der wir gerade interessiert sind (Carroll 1961: 319).

genannten Rückkopplungseffekts haben unangemessene Bewertungsmaßnahmen auch einen negativen Einfluss auf die Lehrplangestaltung und die Unterrichtspraxis, weshalb die Frage der Bewertung der Mehrsprachigkeit nicht von der Frage des Sprachunterrichts getrennt werden kann<sup>31</sup>.

Wenn sie auf Mehrsprachensprecher angewendet werden, erfüllen einsprachige Tests nicht die Kriterien, die erforderlich sind, damit Tests aussagekräftige Informationen über die sprachlichen Leistungen oder das Potenzial der Lernenden liefern. Die Tests werden nicht valide sein, da sie nicht das testen, was sie testen sollen. Sie werden auch nicht zuverlässig sein, da sich die Ergebnisse der Lernenden kurzfristig ändern können. Schließlich sind die Tests nicht authentisch, da sie zu weit von der kommunikativen Erfahrung der Sprachbenutzer mit gemischtem Sprachhintergrund entfernt sind. Sie zudem nicht objektiv, da die erzielten Ergebnisse in erheblichem Maße von den Sprachkenntnissen und kulturellen Dispositionen derjenigen abhängen, die die jeweiligen Prüfungen festlegen oder durchführen<sup>32</sup>.

Auf der Grundlage dieser Überlegungen und der Erkenntnis dieser Unzulänglichkeiten hat sich die Dyme-Forschungsgruppe daran gemacht, ein dynamisches Modell zur Beurteilung von Mehrsprachigkeit (*Dynamic Model of Multilingual Assessment*, kurz DMMA) zu entwickeln, das sich auf die Erkenntnis stützt, dass die Sprache mehrsprachiger Sprecher ein komplexes dynamisches System bildet, sowie auf die Erkenntnisse über die Natur des Mehrspracherwerbs. Es liegt auf der Hand, dass dieses Modell zu Beurteilungsverfahren führen muss, die sowohl den Kriterien als auch den Erwartungen entsprechen, die in diesem Artikel dargelegt werden<sup>33</sup>.

Hier können wir aus der Geschichte lernen: So wie die großen Erfolge von Napoleons *Grande Armée* auf ihrer Fähigkeit beruhten, sich mit überraschender Geschwindigkeit zu bewegen und große Gebiete zu erobern – was auf der

---

31 Angesichts des Rückkopplungseffekts (*washback*) von Tests und Bewertungen auf das Lehren und Lernen (Cheng 2005; Tsangari 2011) und der Kontrolle, die Tests auf die Neudefinition von Wissen ausüben (Shohamy 2001), erfordern Veränderungen im Sprachunterricht auch Veränderungen in der Sprachprüfungs- und Bewertungspraxis (Dendrinou 2019: 4).

32 Studien haben Hinweise auf kulturelle Verzerrungen in Tests gefunden (Sattler 2001). So werden kulturelle Inhalte und kulturspezifisches Wissen häufig in Testaufgaben eingebettet (Warren 2006), was sich nachteilig auf die Leistungen von Kindern mit mehrsprachigem Hintergrund auswirken kann (Schon et al. 2008). Dies kann dazu führen, dass Schüler mit ethnischen Minderheitenhintergrund in der Sonderpädagogik überproportional vertreten sind, was seit über 30 Jahren ein Problem darstellt (Strand/Lindsay 2009) Glazzard (2017: 3).

33 Obwohl ein eindeutiger Bedarf besteht, Beurteilungen von mehrsprachiger Kompetenzen zu entwickeln, die eine ganzheitliche oder heteroglossische Perspektive widerspiegeln, ist weitere Forschung erforderlich, um valide Beurteilungen mehrsprachiger Kompetenzen zu entwickeln. So müssen beispielsweise die Konstrukte, die in mehrsprachigen Tests gemessen werden sollen, klar definiert werden (Lopez/Turkan/Guzman-Orth 2016: 10).

Bereitstellung eines guten Paares Stiefel für jeden Soldaten beruhte, die sowohl einen linken als auch einen rechten Stiefel aufwies, im Gegensatz zu der Gewohnheit der alliierten Truppen, einen einzigen Stiefel bereitzustellen, der unterschiedslos am linken oder rechten Fuß getragen werden konnte oder musste –, so wird die Bereitstellung eines geeigneten Beurteilungsinstruments für mehrsprachige Lernende, das sowohl für einsprachige als auch für mehrsprachige Sprecher geeignet ist, sowohl die Geschwindigkeit des Fortschritts als auch die Effektivität bei der Erreichung spezifischer Sprachziele für mehrsprachige Lernende erhöhen und somit die Unterschiede zwischen einsprachigen und mehrsprachigen Sprechern ausgleichen und die Diskriminierung von Lernenden mit einer abweichenden Sprachsozialisation unterbinden.

## Literatur

Aitchison, Jean (2002): *Words in the mind. An introduction to the Mental Lexicon*. Hoboken: John Wiley & Sons.

Anderson, Raquel T. & Márquez, Alejandra (2009): The article paradigm in Spanish-speaking children with SLI in language contact situations. In: Grinstead, John (Ed.): *Hispanic child languages: Typical and impaired developments*. Amsterdam & Philadelphia: John Benjamins, 29-55.

Bourdieu, Pierre (1992): *The logic of practice*. Stanford: Stanford University Press.

Bourdieu, Pierre (1991): *Language and symbolic power*. Cambridge Mass.: Harvard University Press.

Canagaraiah, Suresh (2006): Changing communicative needs. Revised assessment objectives: Testing English as an international language. *Language Assessment Quarterly* 3: 3, 229-242.

Carroll, John B. (1961): Fundamental considerations in testing for English language proficiency of foreign students. In: Allen, Harold B. & Campbell, Russell N. (Eds.): *Teaching English as a Second Language: A Book of Readings*. New York: McGraw Hill, 313-320.

Cheng, Liying (2005): *Changing language teaching through testing.: A washback study*. Cambridge: Cambridge University Press.

Chomsky, Noam (2009): *Cartesian linguistics: A chapter in the history of rationalist thought*. Cambridge: Cambridge University Press.

Council of Europe (2001): *Common European Framework of reference for languages: learning, teaching, assessment*. Cambridge et al.: Cambridge University Press.

Council of Europe (2012): *FREPA – A framework of reference for pluralistic approaches to languages and cultures: Competences and resources*. [Prepared by Candelier, Michel et al.]. Strasbourg: Council of Europe.

- Council of Europe (2007): *From linguistic diversity to plurilingual education: Guide for the development of language education policies in Europe*. Strasbourg: Council of Europe.
- Council of Europe (2018): *CEFR – Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. [Companion Volume]. Strasbourg: Council of Europe.
- Cummins, Jim (2007): BICS and CALP: Empirical and theoretical status of the distinction. In: Hornberger, Nancy H. (Ed.): *Encyclopedia of Language and Education*. Boston: Springer, 487-499. [https://doi.org/10.1007/978-0-387-30424-3\\_36](https://doi.org/10.1007/978-0-387-30424-3_36)
- De Lamo White, Caroline & Jin, Lixian (2011): Evaluation of speech and language assessment approaches with bilingual children. *International journal of language and communication disorders* 46: 6, 613-627.
- Dendrinos, Bessie (2019): Multilingual Testing and Assessment for Plurilingual Education. [ECSPM position paper].
- Dendrinos, Bessie & Gotsoulia, Voula (2015): Setting standards for multilingual frameworks in foreign language education. In: Spolsky, Bernard; Inbar-Lourie, Ofra & Tannenbaum, Michal (Eds.): *Recognizing the Individual in Language Studies*. London: Routledge.
- Garcia, Ofelia & Wei, Li (2014): *Translanguaging: Language, Bilingualism and Education*. London: Palgrave Macmillan. <https://doi.org/10.1057/9781137385765>
- Glazzard, Jonathan (2017): The necessity for assessment and management of speech, language and communication needs to take account of cultural and multilingual diversity. *International Journal of Learning, Teaching and Educational Research* 16: 2, 11-21.
- Goral, Mira & Conner, Peggy S. (2013): Language Disorders in Multilingual and Multicultural Populations. *Annual Review of Applied Linguistics* 33, 128-161.
- Grech, Helen & McLeod, Sharynne (2012): Multilingual speech and language development and disorders. In: Battle, Dolores E. (Ed.): *Communication Disorders in Multicultural and International Populations*. St Louis, MO: Elsevier, 119-147.
- Hasson, Natalie; Camilleri, Bernard; Jones, Caroline; Smith, Jodie & Dodd, Barbara (2012): Discriminating disorder from difference using dynamic assessment with bilingual children. *Language Teaching and Therapy* 29:1, 57–75.
- Herdina, Philip & Jessner, Ulrike (2002): *A dynamic model of multilingualism. Perspectives of Change in Psycholinguistics*. Clevedon: Multilingual Matters.
- Iglesias, Aquiles (2015) Language Impairment in Bilingual Children: From Theory to Practice, in: *Seminars in Speech and Language*, 36 (02), 87, Thieme: New York 2015

- Jessner, Ulrike (2008): A DST model of multilingualism and the role of metalinguistic awareness. *Modern Language Journal* 92: 2, 270-283.
- Kramsch, Claire (1998): *Language and culture*. Oxford: Oxford University Press.
- Krashen, Stephen D. (1987): *Principles and Practice in Second Language Acquisition*. Hoboken: Prentice-Hall International.
- Kuhn, Thomas (1962): *The structure of scientific revolutions*. Chicago & London: University of Chicago Press.
- Lado, Robert (1961): *Language testing: The construction and use of foreign language tests: A teacher's book*. Bristol: Longmans, Green and Company.
- Lakatos, Imre (1970): The theory of scientific research programmes. In: Lakatos, Imre & Musgrave, Alan (Eds.): *Criticism and the growth of knowledge*. Cambridge: Cambridge University Press, 191-197.
- Lantolf, James P. (2000): *Sociocultural Theory and Second Language Learning*. Oxford: Oxford University Press.
- Lantolf, James P. & Poehner, Matthew E. (2007): Dynamic assessment. In: Hornberger, Nancy H. (Ed.): *The Encyclopedia of Language and Education*. [vol. 7]. Cambridge: Cambridge University Press, 273-285.
- Lopez, Alexis A; Turkan, Sultan & Guzman-Orth, Danielle (2016): Assessing multilingual competence, In: Shohamy, Elena et.al. (Eds.): *Language Testing and Assessment, Encyclopedia of Language and Education*. New York: Springer, 1-12.
- Lidz, Carol S. & Peña, Elizabeth D. (2009): Response to intervention and dynamic assessment: Do we just appear to be speaking the same language? *Seminars in Speech and Language* 30: 2, 121–133.
- McNamara, Tim (2000): *Language testing*. Oxford: Oxford University Press.
- Murphy, Raegan (2011): *Dynamic assessment, intelligence and measurement*. London: Wiley-Blackwell.
- Paikeday, Thomas (1985): *The Native Speaker Is Dead: An Informal Discussion of a Linguistic Myth with Noam Chomsky and Other Linguists, Philosophers, Psychologists, and Lexicographers*. [2<sup>nd</sup> Edition]. Toronto & New York: Paikeday Press.
- Paradis, Johanne; Genesee, Fred & Crago, Martha B. (2011): *Dual Language Development and Disorders: A Handbook on Bilingualism and Second Language Learning*. Baltimore, MD: Paul H. Brookes.
- Paran, Amos & Sercu, Lies (Eds.) (2010): *Testing the untestable in language education. New perspectives in language and education*. Bristol: Multilingual Matters.
- Requal, (2021): *Glossary of basic theoretical concepts for teaching and learning in multilingual contexts, Requalification of recently immigrated refugee teachers in Europe*, Cologne.

- Ryle, Gilbert (1973): *The Concept of Mind*. Harmondsworth: Penguin University Press.
- Sattler, Jerome M. (2001): *Assessment of children: Cognitive Applications*. [4th edition]. San Diego, CA: Jerome M. Sattler Publishing.
- Schneiders, Guido & Komor, Anna (2007): Eine Synopse aktueller Verfahren der Sprachstandserhebung. In: Bundesministerium für Bildung und Forschung (Hrsg.): *Anforderungen an Verfahren der regelmässigen Sprachstandsfeststellung als Grundlage für die frühe und individuelle Förderung von Kindern mit und ohne Migrationshintergrund*. Bonn & Berlin: BMBF, 261-342.
- Schon, Jacqueline; Shaftel, Julia & Markham, Paul (2008): Contemporary Issues in the Assessment of Culturally and Linguistically Diverse Learners. *Journal of Applied School Psychology* 24: 2, 163-189. <https://doi.org/10.1080/15377900802089395>
- Shohamy, Elana; Or, Iair G. & May, Stephen (Eds.) (2016): *Language testing and assessment*. New York: Springer.
- Shohamy, Elana (2011): Assessing multilingual competencies: Adapting construct valid assessment policies. *The Modern Language Journal* 95: 3, 418-429. <https://doi.org/10.1111/j.1540-4781.2011.01210.x>
- Strand, Steve & Lindsay, Geoff (2009): Evidence of Ethnic Disproportionality in Special Education in an English Population. *The Journal of Special Education* 43: 3, 174-190. <https://doi.org/10.1177/0022466908320461>
- Tsangari, Dena (2011): The Washback Effect of Language Proficiency Testing on Education and Learning Processes. RCEl eperiodica. ([http://rcel.enl.uoa.gr/periodical/prev\\_articles\\_3\\_en.htm](http://rcel.enl.uoa.gr/periodical/prev_articles_3_en.htm))
- Wang, Yangting & Li, Shikun (2020): Issues, challenges, and future directions for multilingual assessment. *Journal of language teaching and research* 11: 6, 914-919. <http://dx.doi.org/10.17507/jltr.1106.06> © 2020 ACADEMY PUBLICATION
- VanPatten, Bill & Benati, Alessandro G. (2010): *Key terms in second language acquisition*. London & New York: Continuum.
- Warren, Cortney S. (2006): Incorporating multiculturalism into undergraduate psychology courses: Three simple active learning activities. *Teaching of Psychology* 33: 2, 105-109. [https://doi.org/10.1207/s15328023top3302\\_5](https://doi.org/10.1207/s15328023top3302_5)
- Wright, Wayne E. & Li, Xiaoshi (2008): High-stakes math tests: How No Child Left Behind leaves newcomer English language learners behind. *Language Policy* 7, 237-266.
- Yuzar, Ella (2020): Incorporating communicative competence in assessment and English language teaching in multilingual settings, *Journal of Research and Innovation in Language* 2: 1, 8-13. <https://doi.org/10.31849/reila.v2i1.3864>