

## 26 Sprachliche Teilbereiche als Indikatoren der Sprachkompetenz

### 26.1 Faktorenanalyse

Eine Faktorenanalyse bietet Informationen sowohl über ein Persönlichkeitsmerkmal wie über das zu analysierende Verfahren. Zunächst wurden mit Hilfe der Faktorenanalyse Merkmale auf ihre Ein- oder Mehrdimensionalität hin untersucht. Die Theorie von Spearman (1904) über einen vermuteten Generalfaktor bei der Intelligenz neben spezifischen Elementarfähigkeiten bildete den Ausgangspunkt für eine Reihe von Ansätzen der Faktorenanalyse, durch die Spearmans Hypothese dann widerlegt bzw. relativiert wurde. Die Faktorenanalyse ist eine multivariate Methode zur Klassifizierung mehrerer Variablen in einer geringeren Anzahl von Kategorien. Da die korrelativen Beziehungen dabei berücksichtigt werden, findet bei dieser datenreduzierenden Methode kein großer Informationsverlust statt. Da inzwischen jedoch mehrere Verfahren der Faktorenanalyse entwickelt wurden, die als faktoranalytische Modelle, Verfahren der Faktorenextraktion und der Faktorrotation zu unterscheiden sind, hängt die Anzahl der zu ermittelnden Faktoren nicht nur von dem untersuchten Merkmal, sondern auch von der Wahl der dabei benutzten Methode ab.

Trotz gewisser Unterschiede bestimmen alle Methoden die faktorielle Struktur des Merkmals bzw. des Verfahrens. Dadurch kann aufgrund der Analysedaten je nach dem Zweck des Verfahrens ein faktoriell homogenes oder heterogenes Verfahren zusammengestellt oder zur Validitätserhöhung absichtlich eine Suppressor-Variable aufgenommen werden. Außerdem bietet die Analyse für eine eventuelle Revision Entscheidungshilfen bei der Frage, welche der untersuchten Variablen als fundamental anzusehen und daher zwecks ausreichender Genauigkeit beizubehalten sind und welche bei einer Revision vernachlässigt werden können.

Ist die Dimensionalität eines Merkmals einigermaßen bekannt, ermöglicht die Faktorenanalyse auch die Überprüfung eines Diagnoseverfahrens hinsichtlich der hinreichenden Berücksichtigung der faktoriellen Struktur des Merkmals. In Bezug auf die Sprachkompetenz ist ein Verfahren meines Erachtens als geeignet zu betrachten, wenn es das komplexe Sprachverhalten auch entsprechend mehrdimensional erfasst. Mitunter wird zwar eher ein allgemeiner Sprachfähigkeitsfaktor (*General Language Proficiency Factor*) angenommen, doch diese in den 70er Jahren vor allem von Spolsky und Oller vehement vertretene Position ist sehr umstritten. Ob Untersuchungen zur Struktur von Fremdsprachenkompetenz eher die Hypothese von der Eindimensionalität oder die Hypothese der Mehrdimensionalität stützen, hängt anscheinend sehr stark von den bei der Datenerhebung als auch bei der Auswertung eingesetzten

Verfahren ab. Bei einem Vergleich mehrerer Studien zur Dimensionalität der Fremdsprachkompetenz ist festzustellen, dass eine Tendenz zum Generalfaktor nahezu ausschließlich besteht, wenn nur wenige Variablen – und vor allem keine Daten zur mündlichen Produktion – einbezogen und die Faktoren außerdem nach der Hauptkomponentenmethode bestimmt werden. So erweist sich bei genauerer Analyse ein angeblicher Generalfaktor oft nur als Pseudogeneralfaktor, zumal wenn der gefundene Faktor nur knapp 50% der Varianz aufklären kann.<sup>260</sup> Ein solches Ergebnis muss als Indiz dafür angesehen werden, dass wichtige sprachliche Dimensionen von dem Verfahren nicht erfasst wurden. Wenn auch die Debatte über die Anzahl und Art der Faktoren einer Sprachkompetenz bzw. der Disposition zum Lernen einer Fremdsprache noch andauert, hinsichtlich der Sprachperformanz bestehen allgemein keine Zweifel, dass daran mehrere Fähigkeiten (*skills*) beteiligt sind.

Grundlage jeder Faktorenanalyse ist eine Matrix von Daten, die bei Probanden für eine Anzahl von verschiedenen Variablen gemessen oder beobachtet worden sind. Bei der nacheinander erfolgenden Extraktion von Faktoren wird die sogenannte Ladung der Variablen bei diesem Faktor ermittelt. Die Summe der quadrierten Ladungen ergibt dann dessen Eigenladung  $\lambda$ , der dem Anteil der Gesamtvarianz entspricht, die durch diesen Faktor aufgeklärt wird. Anhand der Eigenwerte ist festzustellen, wie viele Faktoren notwendig sind, um das beobachtete Verhalten hinreichend genau zu beschreiben. Bei einem Diagnoseverfahren sollten in der Regel so viele Faktoren zu identifizieren sein, deren Summe 80% der Gesamtvarianz aufklärt. Grundsätzlich werden zur Aufklärung der Gesamtvarianz weniger Faktoren benötigt, wenn die Variablen wie bei einem homogenen Verfahren alle hoch miteinander korrelieren. Bei einer Matrix, deren Interkorrelationen den Wert 1 aufweisen, ist auch die Dimensionalität der Variablen gleich 1 und ein Faktor würde ausreichen, alle Interkorrelationen und beobachteten Messwerte hinreichend zu erklären.

Die Summe der quadrierten Ladungen einer Variablen über die aufgefundenen Faktoren ergibt hingegen deren Kommunalitätsindex  $h^2$  und sagt aus, in welchem Maße diese Variable durch die Faktoren bereits erfasst bzw. aufgeklärt wird. Theoretisch kann die Kommunalität maximal den Wert 1 annehmen, doch wird die Extraktion der Faktoren wegen ihrer zunehmenden Bedeutungslosigkeit meistens vorher abgebrochen. Das ist bei der Hauptkomponentenmethode der Fall, wenn der Eigenwert eines Faktors unter dem Wert 1 liegt, weil der betreffende Faktor dann weniger Varianz aufweist als eine einzelne Variable. Der verbleibende Rest bei einer Variablen ist dann durch die Fehlervarianz oder eine Spezifität zu erklären, die nicht zu

---

<sup>260</sup> Helmut J. Vollmer, *Spracherwerb und Sprachbeherrschung*, Untersuchungen zur Struktur

den wesentlichen Kriterien des untersuchten Merkmals gehört. Ansonsten ist eine nur bei einem Faktor auftretende hohe Ladung bei einer Variablen als Hinweis auf deren Spezifität in Bezug auf eine Komponente des Merkmals zu interpretieren.

Die in dieser Untersuchung nach Lienert<sup>261</sup> durchgeführte Faktorenanalyse beginnt nach den Regeln der Zentroidmethode, basiert im Wesentlichen aber auf der von Hotelling entwickelten Hauptkomponentenmethode, die im Gegensatz zur Zentroidmethode keine Schätzwerte einsetzt und daher als rechnerisch objektiv gilt. Der Vorzug dieser Methode besteht darin, eine Korrelationsmatrix auf die geringstmögliche Anzahl von orthogonalen Faktoren zu reduzieren, hat jedoch zur Folge, dass die Gemeinsamkeit der resultierenden Faktoren oft kaum erkennbar und ihre inhaltliche Interpretation daher schwierig ist. Bei einer grundsätzlichen Studie zur Dimensionalitätsstruktur der Sprachkompetenz, die hier aber nicht anliegt, wäre diese Rechenmethode weniger geeignet.

Wie schon bei den anderen auf das Verfahren angewendeten Analysen wird das Verfahren wiederum nach Sprache und Sprachstatus getrennt auf seine faktorielle Struktur untersucht. Die Faktorenanalyse wird beispielhaft bei der Partnersprache Italienisch (Berechnung siehe Tabellenblatt 16), bei Deutsch als starker Sprache unter Einbezug der monolingualen Kontrollgruppe (Berechnung siehe Tabellenblätter 17, 18) durchgeführt, desgleichen bei beiden Sprachen Deutsch und Italienisch als starker Sprache einschließlich der Kontrollgruppen gemeinsam (Berechnung siehe Tabellenblätter 19, 20). Allerdings werden zur Reduzierung der erforderlichen Rechenvorgänge nicht alle ausgewerteten Variablen einbezogen. Aus der 21 Variablen umfassenden Korrelationsmatrix bei Deutsch bzw. Italienisch als starker Sprache wurden 12 Variablen unter Berücksichtigung folgender Aspekte ausgewählt:

alle Bereiche, die nur einmal ausgewertet wurden:

Lesefertigkeit, kommunikatives Sprachverhalten, Begriffsbildung (Konzepte)

bei Bereichen, die bei zwei Untersuchungsanordnungen ausgewertet wurden, das aus dem Mittelwert bestehende Gesamtergebnis für diesen Bereich:

Phonematik/Prosodie, Hörverständnis, Wortschatz, Morphosyntax, mündlicher Sprachgebrauch

bei starken Abweichungen der Korrelationen zum Gesamtergebnis bei einem Bereich jedoch die einzelnen Komponenten oder zusätzlich eine Komponente

Diktat und motorische Schreibfertigkeit (nicht Schreibfertigkeit Gesamt)  
Textverständnis (Komponente von Hörverständnis)  
Phonematik/Prosodie Bild

Entsprechend wurden bei der Partnersprache von 13 Variablen nur 5 Variablen der Faktorenanalyse unterworfen:

---

von Fremdsprachenfähigkeit, Tübingen 1982

<sup>261</sup> Lienert 1969, a.a.O., S.546-551, S.559

Hörverständnis - gesamt  
 Phonetik/ Prosodie - gesamt  
 Mündlicher Sprachgebrauch - gesamt  
 Lesefertigkeit  
 Kommunikatives Sprachverhalten

Aus der Faktorenanalyse ergeben sich folgende faktorielle Strukturen des Verfahrens:

Partnersprache Italiano	Ladung A <sub>0i</sub>	Ladung B <sub>0i</sub>	h <sub>i</sub> <sup>2</sup> = Kommunalität
Hörverständnis Gesamt	0,87	-0,17	0,79
Lesefertigkeit	0,58	0,73	0,87
Phon./ Pros. Gesamt	0,90	0,26	0,88
Mündl. Sprachgebrauch G	0,91	-0,14	0,85
Kom. Sprachverhalten	0,52	-0,73	0,80
Quadratsumme = $\lambda$ (5 Variablen)	3,00	1,18	Summe h <sub>i</sub> <sup>2</sup> = 4,18
% der Gesamt-Varianz	60%	24%	84%

starke Sprache D	Ladung A <sub>0i</sub>	Ladung B <sub>0i</sub>	Ladung C <sub>0i</sub>	h <sub>i</sub> <sup>2</sup> = Kommunalität
Phon./ Pros. Bild B	0,79	-0,38	0,31	0,86
Textverständnis	0,74	0,36	-0,25	0,74
mot. Schreibfertigkeit	0,54	0,47	0,02	0,51
Diktat	0,72	0,46	-0,16	0,76
Konzepte	0,34	-0,42	-0,72	0,81
Hörverständnis Gesamt	0,83	0,36	0,02	0,82
Lesefertigkeit	0,82	0,50	0,04	0,92
Phon./ Pros. Gesamt	0,85	0,32	0,09	0,83
Wortschatz Gesamt	0,78	-0,44	0,09	0,81
Morphosyntax Gesamt	0,72	-0,58	0,27	0,93
Mündl. Sprachgebrauch G	0,83	-0,55	0,14	1,01
Kom. Sprachverhalten	0,45	-0,38	-0,60	0,71
Quadratsumme = $\lambda$ (12 Variablen)	6,1993	2,342	1,174	Summe h <sub>i</sub> <sup>2</sup> = 9,715
% der Gesamt-Varianz	52%	20%	10%	81%

starke Sprache D/ I	Ladung A <sub>0i</sub>	Ladung B <sub>0i</sub>	Ladung C <sub>0i</sub>	h <sub>i</sub> <sup>2</sup> = Kommunalität
Phon./ Pros. Bild B/ A	0,74	-0,22	0,51	0,86
Textverständnis	0,73	0,16	-0,54	0,85
mot. Schreibfertigkeit	0,28	-0,70	-0,23	0,62
Diktat	0,51	-0,55	-0,32	0,67
Konzepte	0,57	0,35	-0,09	0,46
Hörverständnis Gesamt	0,82	0,03	-0,24	0,73
Lesefertigkeit	0,53	-0,69	0,06	0,76
Phon./ Pros. Gesamt	0,74	-0,47	0,32	0,87
Wortschatz Gesamt	0,82	0,33	0,03	0,78
Morphosyntax Gesamt	0,83	0,30	0,20	0,82
Mündl. Sprachgebrauch G	0,89	0,35	0,10	0,92
Kom. Sprachverhalten	0,56	0,20	-0,09	0,36
Quadratsumme = $\lambda$ (12 Variablen)	5,712	2,048	0,938	Summe h <sub>i</sub> <sup>2</sup> = 8,698
% der Gesamt-Varianz	48%	17%	8%	72%

Je nachdem auf welche Zielgruppe sich das Verfahren bezieht, zeigt die faktorielle Struktur des Verfahrens durchaus Abweichungen. Im gemeinsamen Faktor A weisen aber die Bereiche Phonematik/Prosodie, Hörverständnis/Textverständnis und mündlicher Sprachgebrauch sowie dessen Komponenten Wortschatz und Morphosyntax durchwegs hohe Ladungen auf, weshalb dieser Faktor eindeutig der Erfassung der rezeptiven und produktiven mündlichen Sprachkompetenz zuzuordnen ist. Außerdem ist das kommunikative Sprachverhalten am Faktor A bei allen Zielgruppen mäßig positiv vertreten, während es bei den anderen Faktoren eine negative oder nur sehr geringe Ladung aufweist, ohne annähernd aufgeklärt zu werden, besonders wenn Deutsch und Italienisch als starke Sprache gemeinsam der Faktorenanalyse unterworfen werden. Das mit dem Verfahren erfasste kommunikative Sprachverhalten repräsentiert daher keinen eigenständigen Aspekt der Sprachkompetenz, sondern ist an den anderen Aspekten des mündlichen Sprachgebrauchs beteiligt, bezieht sich aber darüber hinaus auch noch spezifisch auf nonverbale Kommunikationsfähigkeiten, wie die niedrigen Interkorrelationen mit den anderen Bereichen anzeigen. Das ist nun keineswegs verwunderlich, sondern bestätigt die Annahmen über die integrative Struktur des kommunikativen Sprachverhaltens sowie die Beteiligung von weiteren Persönlichkeitsmerkmalen an der Kommunikationsfähigkeit.

Weniger eindeutig sind jedoch die anderen Faktoren zu interpretieren, da dort die Ladungen der Variablen je nach Zielgruppe bei verschiedenen Faktoren variieren. Bedingt durch das Verfahren der Hauptkomponentenmethode und sprachstrukturelle Unterschiede werden bei den anderen Faktoren weitere Sprachaspekte nicht einheitlich erfasst, zumal das Verfahren bei der Partnersprache bereits mit der Extraktion eines zweiten Faktors zur Erklärung der Varianz hinreichend auskommt, da ein Diagnoseverfahren ab 80% Varianzaufklärung für das untersuchte Verhalten als geeignet gilt. Gegenüber dem Faktor A ist bei den Faktoren B oder C die stärkere Berücksichtigung einer Diskriminationsfähigkeit von Lauten und Buchstaben erkennbar, da die Bereiche Phonematik/ Prosodie, Schreiben und Lesefertigkeit dort höhere Ladungen aufweisen als andere Bereiche. Jedenfalls ist festzustellen, dass das Verfahren die Sprachkompetenz in allen Leistungsbereichen mehrdimensional erfasst und damit die Anforderungen an ein Sprachdiagnoseverfahren grundsätzlich erfüllt. Da aber die bei der Untersuchung an der SESB benutzte Pilotfassung bei ihrer Vielzahl von Untersuchungsanordnungen und wegen der mehrfachen Auswertung bei einigen Bereichen äußerst zeitaufwendig ist, sollen nun durch die Selektion von unverzichtbaren Untersuchungsbereichen Ansätze für die Zusammenstellung eines im Umfang reduzierten Beobachtungsverfahrens gewonnen werden.

## 26.2 Selektion von Untersuchungsbereichen

Bei der Auswahl geeigneter Untersuchungsbereiche für eine überarbeitete Fassung des Beobilingua-2dit-Verfahrens stehen vielfältige Analysedaten zur Verfügung, da zu allen Bereichen Informationen über die Trennschärfe und Schwierigkeit, die sich daraus ergebende Reliabilität sowie über die Validität, Dimensionalität und Kovarianz mit anderen Bereichen vorliegen. Die Selektion der Untersuchungsanordnungen darf aber keineswegs schematisch anhand dieser Daten erfolgen, sondern hat neben dem Zeitaufwand auch das Alter der Zielgruppe zu berücksichtigen. Damit die etwa 8-jährigen Probanden nicht allzu sehr empfinden, einer Überprüfungssituation ausgesetzt zu sein, ist eine Reduzierung auf Einzelaufgaben zu vermeiden, obwohl sie sich zur Beobachtung der Sprachfähigkeiten mehr bewährt haben als die Gruppen- oder Partneraufgaben. Außerdem ist zu bedenken, dass jeder Verzicht auf Beobachtungen zu einem – wenn auch weniger fundamentalen – sprachlichen Aspekt und vor allem die nur einmalige Beobachtung zu einem Verlust an der Zuverlässigkeit und Validität der Daten führen, da Sprachverhalten stark situativ bedingten Leistungsschwankungen unterliegt. Deshalb sollte die Anwendung eines reduzierten Verfahrens wenigstens von mehrfachen Hospitationen begleitet werden, damit die Ergebnisse spezieller Aufgaben anhand von Eindrücken in natürlicheren Situationen überprüft werden können.

Zeitökonomisch wäre es zu begrüßen, wenn sich ein Untersuchungsbereich allein schon hinreichend als allgemeiner Indikator der Sprachkompetenz erweisen würde. Aber kein Bereich erfüllt die Bedingung für eine solche Basisvariable, bei allen Faktoren eine einigermaßen hohe Ladung aufzuweisen, auch nicht das Hörverständnis, das gern als einziger Bereich zur Überprüfung der Sprachkompetenz herausgegriffen wird. Bei Anwendung der Kriterien für das Hörverständnis bei Deutsch als starker Sprache ergibt sich zwar bei allen Faktoren eine positive Ladung, doch fällt diese im Faktor C mit 0,02 dermaßen gering aus, dass sie nicht in Betracht gezogen werden kann. Außerdem spricht die Anwendung derselben Kriterien auf Lerngruppen, deren Sprachkompetenz eine größere Varianz aufweist, dafür, dass das Hörverständnis eher sehr spezifisch nur die rezeptiven Fähigkeiten im Mündlichen erfasst, denn bei Italienisch als Partnersprache und als sogenannter starker Sprache liegen allein bei Faktor A sehr hohe positive Ladungen vor. Bei allen Lerngruppen fallen auch die Interkorrelationen zu anderen Sprachbereichen recht mäßig aus (unter 0.70), weshalb Beobachtungen allein zum Hörverständnis kaum Rückschlüsse auf den Grad der Sprachkompetenz rechtfertigen. So führt die Beurteilung der Sprachfähigkeiten in einer erst schwach beherrschten Sprache aufgrund einer auf das Hörverständnis beschränkten Untersuchung in der Regel zu einer überhöhten Einschätzung, wie

beispielsweise anhand der Beobachtungen an der SESB zu den Sprachfähigkeiten in der Partnersprache Italienisch (Kapitel 16, S.248) nachvollzogen werden kann.

Als allgemeine Indikatoren erscheinen eher die Bereiche Phonematik/Prosodie oder mündlicher Sprachgebrauch, vor allem die Komponente Morphosyntax geeignet. Diese Bereiche korrelieren auch bei allen Lerngruppen relativ hoch mit dem externen Validitätskriterium und dem Gesamtergebnis und weisen in der Partnersprache auch zu anderen Sprachbereichen mit Ausnahme des kommunikativen Sprachverhaltens und der Lesefertigkeit durchwegs hohe Interkorrelationen auf. Außerdem bestätigt die Faktorenanalyse die Mehrdimensionalität dieser beiden Bereiche, da immerhin neben hohen Ladungen im Faktor A auch mäßige Ladungen (bis zu 0,51 bei Phon./Pro.) in einem oder gar beiden anderen Faktoren auftreten. Entsprechend hoch fallen denn auch die Kommunalitätsindizes aus (zwischen 0,82 und 0,93). Zur halbwegs gültigen Einschätzung der Sprachkompetenz kann daher auf Beobachtungen zum mündlichen Sprachgebrauch nicht verzichtet werden.

Ähnlich mehrdimensional erweist sich auch die Lesefertigkeit, die aber wiederum aufgrund ihrer Spezifität als alleinige Variable absolut auszuschließen ist. Abgesehen von Deutsch als starker Sprache, wo auch zum Hör- und Textverständnis sowie dem Diktat respektable Interkorrelationen auftreten, besteht bei den übrigen Gruppen lediglich zum Bereich Phonematik/Prosodie eine aussagekräftige Beziehung, wobei sich diese Kovarianz vermutlich verfahrensintern durch die auch bei der Leseübung vorgenommene Auswertung in diesem Bereich erklärt. Da eine Leseprobe Fähigkeiten der graphemischen Diskrimination und den Transfer in die phonologische Performanz integriert, stellt sie eine geeignete Ergänzung zu Beobachtungen des mündlichen Sprachgebrauchs dar. Allerdings ist der Bereich der Lesefertigkeit angesichts der zu geringen Schwierigkeit (starke Sprachen It: 0.80, D: 0.70 - Ps It + D: 0.59) und der daraus resultierenden mangelhaften Trennschärfe in Bezug auf die Textvorlage und die Bewertungskriterien noch zu überarbeiten.

Auch das Textverständnis weist positive Ladungen in mehr als einem Faktor auf, wobei der Schwerpunkt eindeutig beim Faktor A liegt, also schon von anderen Variablen des mündlichen Sprachgebrauchs, vor allem vom Hörverständnis erfasst wird (Interkorrelation 0,87 bei Deutsch und Italiano als starker Sprache). Auf mündliche Beobachtungen zum Textverständnis kann daher ohne großen Informationsverlust verzichtet werden.

Der Bereich Begriffsbildung (Konzepte-Fragenkatalog) erweist sich ähnlich wie das kommunikative Sprachverhalten mäßig an mehreren Faktoren beteiligt, wobei die höchsten Ladungen im Faktor A aufzufinden sind, was sich durch die Verwandtschaft

zum Bereich Wortschatz (besonders Begriffsbildung) und Morphosyntax (stärker beim kom. Sprachv.) erklärt. Da die Ladungen aber sowohl bei der Begriffsbildung als auch beim kommunikativen Sprachverhalten ziemlich niedrig oder sogar bei den Faktoren B und/oder C negativ ausfallen, werden die Erwartungen bestätigt, dass beide Bereiche jeweils noch spezifische, weniger sprachabhängige Persönlichkeitsmerkmale erfassen wie einerseits ein die Kommunikation förderndes Interaktionsverhalten oder im Fall der Begriffsbildung Aspekte der kognitiven Entwicklung. Unter Abwägung zeitökonomischer und sprachlicher Gesichtspunkte kann bei einer Straffung des Verfahrens auf den Bereich Begriffsbildung verzichtet werden, der in der Untersuchung Beobachtungen zur Auswirkung des Sachunterrichts in der nichtdeutschen Sprache diene.

Der Konzepte-Fragenkatalog zeichnet sich zwar durch die bei den Schülern der SESB vorliegende mittlere Schwierigkeit von 0,47 (Idealwert 0,50) als sehr gut konstruiert aus, unterscheidet aber bei einer Trennschärfe von 0,52 nicht eindeutig zwischen guten und schwachen Sprachleistungen. Die mit der Begriffsbildung verknüpfte kognitive Disposition, die das Lernen einer Zweitsprache begünstigt, interessiert bei der Beobachtung der Sprachperformanz weniger als der Bereich des kommunikativen Sprachverhaltens, bei dem ebenfalls ein guter Schwierigkeitsindex von 0,48 und immerhin eine Trennschärfe von 0,64 (D + It., st. Sprache + Ps) vorliegen. Auf Beobachtungen zum kommunikativen Sprachverhalten in beiden Sprachen sollte nicht verzichtet werden, weil sie auch zur Einschätzung des bilingualen Sprachverhaltens wichtig sind. Sie verlängern auch keineswegs die Phase der Datenerhebung, da sich besondere Untersuchungsanordnungen erübrigen, wenn bei anderen Bereichen und den Hospitationen nebenher aufmerksam protokolliert wird.

Hinsichtlich der Schreibfertigkeit ist zwischen dem Bereich Rechtschreibung (Diktat) und der motorischen Schreibfertigkeit zu differenzieren. Wie sich bei der Untersuchung ergeben hat, scheint zwischen der motorischen Schreibfertigkeit und der allgemeinen Sprachkompetenz kaum ein Zusammenhang zu bestehen, so dass dieser Bereich, bei dem wohl eher die Entwicklung der Feinmotorik und der optischen Diskrimination von Symbolen eine ausschlaggebende Rolle spielt, für eine Sprachbeobachtung ohne Einbußen an Validität entfallen kann. Die statistischen Analysen ergeben für diesen Bereich eine mangelnde Trennschärfe (D: 0,32 - It.: - 0,12) und bei Betrachtung beider Sprachen eine starke nichtsprachliche Spezifität, da im Faktor A nur eine geringfügige Ladung von 0,28 und bei den anderen Faktoren sogar negative Ladungen (B: -0,70 - C: -0,23) auftreten.

Der Bereich Rechtschreibung hingegen erfasst auch einen spezifischen Aspekt der Sprachfähigkeit, da er ähnlich wie die Lesefertigkeit und der Bereich Phonematik/



Prosodie außer im Faktor A nicht unerhebliche Ladungen in einem weiteren Faktor aufweist. Er erreicht in Bezug auf die deutsche Sprache eine optimale Trennschärfe von 0.91, trägt aber auch im Deutschen nur mäßig zur Erklärung der Varianz bei, jedenfalls solange nur die Ergebnisse der SESB betrachtet werden. Bei der italienischen Gruppe ergibt sich beim Diktat eine noch geringere Korrelation von 0.44 zum Gesamtergebnis, so dass dieser Bereich ebenfalls ohne große Abstriche an der Validität des Verfahrens im 2. Schuljahr entfallen kann, wenn die Mehrdimensionalität des Sprachverhaltens durch Beobachtungen in dafür in Frage kommenden Bereichen weiterhin gewährleistet bleibt. Ob sich das Diktat zur Überprüfung von Kenntnissen in einer Fremdsprache als Prüfungsform eignet, ist ohnehin umstritten.

Während Oller ein Diktat – jedenfalls in der von ihm und Cohen vorgeschlagenen Form, die sich durch die Angleichung an eine normale Sprechgeschwindigkeit, längere Texteinheiten und die Bewertung nur sinnentstellender Rechtschreibfehler von herkömmlichen Darbietungen unterscheidet, – für einen allgemeinen Indikator der Sprachkompetenz hält, weil es die internalisierte Erwartungsgrammatik eines Lerners aktiviert und substantielle Informationen über die Genese von Fehlern und damit über psycholinguistisch komplexe Prozesse auf verschiedenen sprachlichen Ebenen bietet, erhebt Valette meines Erachtens berechtigte Einwände.<sup>262</sup> Hiernach erscheint ein Diktat allenfalls für fortgeschrittene Schüler und auch nur in Sprachen, bei denen keine starke phonemisch-graphemische Übereinstimmung auftritt, als Prüfungsinstrumentarium geeignet. Auch bei der Untersuchung an der SESB wurden die sprachstrukturellen Unterschiede evident; das Diktat erscheint im Deutschen nur leichter, weil die deutschen Schüler zum Ausgleich der stärkeren Abweichungen zwischen Graphemen und Lauten gegenüber der italienischen Sprache den Text nach deutschen Schulgepflogenheiten zuvor geübt haben. Außerdem zeigte die Untersuchung auch ziemlich deutlich, dass in Hinblick auf den im 2. Schuljahr noch nicht gesicherten Erwerb der motorischen Schreibfertigkeit (Interkorrelationen von 0.50 [it. Sprachgr.] bis 0.66 [dt. Sprachgr.]) ein Diktat für die Diagnose der Sprachkompetenz in der Primarstufe noch wenig angebracht ist.

Für den Verzicht auf ein Diktat bei der Revision des Verfahrens spricht auch, dass durch die Eliminierung dieser Aufgabe zweifellos die Durchführungsobjektivität erhöht wird, denn die abweichende Vorbereitung hat die italienischen Schüler, denen der Text vorher nicht präsentiert wurde, sicherlich benachteiligt. Da es sich beim Diktat um eine

---

<sup>262</sup> Zu Oller siehe: Sang / Vollmer, *Allgemeine Sprachfähigkeit und Fremdsprachenerwerb. Zur Struktur von Leistungsdimensionen und linguistischer Kompetenz des Fremdsprachenlerner*, Berlin 1978, S.23f. – Siehe auch: Bolton, *Die Gütebestimmung kommunikativer Tests*, Tübingen 1985, S.46 - Valette, *Tests im Fremdsprachenunterricht*, Berlin 1971, S.163ff.

Gruppenaufgabe handelte und die Auswertung durch Auszählen der Wörter relativ schnell erfolgen konnte, wird der Zeitaufwand für die Durchführung und Auswertung des Verfahrens ohne das Diktat nur unwesentlich verringert, aber da nach Ausschluss beider Untersuchungsanordnungen im Bereich Schreiben Beobachtungen zu diesem nur in der starken Sprache durchgeführten Bereich entfallen, bleiben unterschiedliche Bezugssysteme und somit die Berücksichtigung der auf die jeweiligen Bezüge (25 oder 22) transformierten Werte erspart, wodurch Vergleiche der Sprachleistungen in der starken Sprache und in der Partnersprache enorm vereinfacht werden.

Bei Entfallen des Konzepte-Fragenkatalogs, der Abschreibaufgabe und des Diktats erhöht sich die Varianzaufklärung durch das Verfahren auch im Bereich der starken Sprachen, und zwar vor allem im Faktor A, der den mündlichen Sprachgebrauch erfasst, auf 56% und damit insgesamt auf 77%. Nach einer Überarbeitung der Bewertungskriterien in den zu erhaltenden, aber wenig trennscharfen Bereichen, dürfte das neu zusammengestellte Verfahren in der Lage sein, auch in der starken Sprache über 80% der in einem 2. Schuljahr zu beobachtenden Varianz der Sprachfähigkeiten mit weniger Zeitaufwand zu erfassen.

Wenn außerdem das Textverständnis nicht mehr mündlich überprüft wird, entfällt zusätzlich die bei dieser Untersuchungsanordnung vorgenommene Auswertung im Bereich Wortschatz und Morphosyntax. Da aber in diesen Bereichen keine hohen Interkorrelationen zwischen den Ergebnissen im Textverständnis und beim Interview zu beobachten waren, trug die mehrfache Datenerhebung ohnehin nicht wie erwartet zu einer zuverlässigeren Beobachtung bei. Um die intraindividuelle Variabilität beim Sprechen reliabler zu erfassen, müsste der mündliche Sprachgebrauch wohl doch bei mehreren freien Gesprächen beobachtet werden, was in der Praxis jedoch höchstens bei Untersuchungen möglich sein dürfte, die auf wenige Individuen begrenzt sind. Schließlich wären dann bei der Erfassung der Sprachfähigkeiten in 2 Sprachen 4 Interviews durchzuführen und auszuwerten. Da nun das Hör- bzw. Textverständnis bei beiden Leistungsbereichen ähnlich hohe Ladungen im Faktor A aufweisen wie die produktiven Ebenen des mündlichen Sprachverhaltens, wird statt einer zweifachen Beobachtung des mündlichen Sprachgebrauchs zur Erhöhung der Reliabilität vorgeschlagen, in jeder Sprache die produktiven Aspekte nur einmalig, das Hörverständnis aber weiterhin bei zwei verschiedenen Untersuchungsanordnungen zu beobachten.

Beim Zusammenstellen eines heterogenen Verfahrens sind allgemein solche Subtests bzw. Bereiche auszuwählen, die untereinander schwache Interkorrelationen aufweisen, sich aber durch hohe Korrelationen mit dem Außenkriterium als valide erwiesen haben. Nun hat sich das Lehrerurteil in einigen Fällen jedoch als ziemlich fragwürdig erwiesen

(siehe Abschnitt 25.3), so dass eine Orientierung höchstens an den zuverlässigeren Beurteilungen der deutschen Kontrollgruppe und in Bezug auf Italienisch nur bei der Lehrerin y gerechtfertigt erscheint. Aus diesen Gründen sind zur Beobachtung einer bilingualen Sprachkompetenz von 8-jährigen Kindern unter Berücksichtigung aller Analysedaten Erhebungen zu den Bereichen Phonematik/Prosodie (sowohl beim Lesen wie beim freien Sprechen), Textverständnis (starke Sprache – aber nicht unbedingt mündlich), Hörverständnis, Wortschatz und Morphosyntax beim Bildimpuls-gesteuerten Interview sowie zur Erfassung spezifischer Aspekte zu den Bereichen Lesefertigkeit und (bilinguales) kommunikatives Sprachverhalten unbedingt beizubehalten.

## **27 Ansätze zur Überarbeitung des Beobachtungsverfahrens**

Die hier vorgelegte Pilotfassung des Beobachtungsverfahrens hat sich bei der Untersuchung am deutsch / italienischen Standort der SESB durchaus als geeignetes Instrumentarium zur Erfassung bilingualer Sprachfähigkeiten von 8 bis 9-jährigen Kindern bewährt. Trotz seiner schon beachtlichen Qualität in Bezug auf die Auswertungsobjektivität, Validität und Reliabilität kann die Wirksamkeit des Verfahrens aber noch aufgrund der dokumentierten Erfahrungen und Analysen gesteigert werden. Die im letzten Abschnitt vorgeschlagene Beschränkung auf besonders sprachimmanente Untersuchungsbereiche leistet dabei vor allem einen Beitrag zur Minderung des Zeitaufwands. Im Folgenden sollen noch einige weitere, aus der Erprobung abgeleitete Änderungen umrissen werden, die zur Qualitätsverbesserung des Verfahrens beitragen können. Eine detaillierte Ausarbeitung würde den Rahmen der vorliegenden Arbeit übersteigen und muss daher auf einen späteren Zeitpunkt verschoben werden.

Zur Beibehaltung ist zu empfehlen:

Der Einsatz von Beobachtungs-Teams à zwei Personen, die sich hinsichtlich ihrer muttersprachlichen Kompetenz in den relevanten Sprachen ergänzen;  
die Protokollführung durch den jeweils nicht mit der Durchführung der Untersuchungsanordnung betrauten Beobachter;  
die Tonbandmitschnitte bei den Interviews, den Leseproben und beim Spiel 1;  
die konsequente Trennung von Beobachtungs- und Auswertungsphase;  
eine quantifizierende Punktezuordnung erst nach einer qualitativen Analyse anhand operationalisierter Auswertungsbögen.

Zur Änderung wird erwogen:

## **Der Zeitpunkt der Durchführung**

Wegen des Einbezugs der spezifischen Sprachleistung der Lesefertigkeit, die mehr Schülern, als vor der Untersuchung vermutet wurde, im 2. Schuljahr noch erhebliche Schwierigkeiten bereitet, sollte das Verfahren statt gegen Ende des zweiten erst gegen Mitte des 3. Schuljahres durchgeführt werden. Wenn sich die Lesefertigkeit stärker gefestigt hat, kann auch die Sinnentnahme eher schriftlich im Klassenverband überprüft werden, was wiederum den Zeitaufwand der Datenerhebung reduziert. Eine solche zeitliche Verschiebung ist durch eine Änderung der Bewertungskriterien in den übrigen Bereichen – vor allem bei den morphosyntaktischen Sprachleistungen – zu berücksichtigen.

Aus der im Abschnitt 26.2 begründeten Minimierung der Untersuchungsbereiche, die mit hinreichender Genauigkeit bilinguale Sprachfähigkeiten bei 8 bis 9-jährigen Kindern erfassen, folgen:

- **Eine Reduzierung der Untersuchungsbereiche**

Die Bereiche Begriffsbildung und Schreibfertigkeit (starke Sprache) sowie Mündlicher Sprachgebrauch bei der Aufgabe zum Textverständnis können entfallen.

Folgende Bereiche sollten bei beiden Sprachen erhalten bleiben:

- Phonematik/Prosodie bei Spontansprache
- Phonematik/Prosodie beim Lesen
- Hörverständnis bei Spontansprache
- Lesefertigkeit
- Kommunikatives Sprachverhalten
- Mündlicher Sprachgebrauch bei Spontansprache (Wortschatz u. Morphosyntax)

Diese Beobachtungen sind zu ergänzen durch:

- Textverständnis (nur starke Sprache)
- Hörverständnis bei Spiel 1 (nur Partnersprache)
- Bilinguales kommunikatives Verhalten

- **Eine Reduzierung der Untersuchungsanordnungen** von 10 auf 7 pro Kind

Dabei sollen die Bildimpuls-gesteuerten Interviews und die Leseübung individuell, das Spiel 1 als Partneraufgabe, die schriftlich zu beantwortenden Fragen zum Textverständnis (starke Sprache) und das Spiel 2 in Gruppen durchgeführt werden.

- **Eine Reduzierung des Zeitaufwands**

Bei einer Reduzierung der Untersuchungsanordnungen von 10 auf 7 verkürzt sich die notwendige Zeit für die Durchführung pro Schüler von 148 Minuten auf 80

Minuten (auf mehrere Tage zu verteilen), wovon aber 50 Minuten bei jeder untersuchten Lerngruppe nur zweimal anfallen (je 20 Minuten in Deutsch und der nichtdeutschen Sprache für das Textverständnis in der starken Sprache, je 30 Minuten für das Spiel 2 in der Teilungsgruppe). Einzeln würden die Sprachfähigkeiten bei jedem Schüler 20 Minuten lang überprüft werden, und zwar je 10 Minuten in jeder Sprache an verschiedenen Tagen, damit kein Sprachwechsel nach dem Interview stattfinden muss, der das Lesen eventuell ungünstig beeinflussen würde.

Beispiel: Ausgehend von der für die Pilotfassung veranschlagten Beobachtungsdauer (S.154) würde sich bei ebenfalls 36 Probanden die Datenerhebung um ca. 14 Zeitstunden bzw. 17 Schulstunden verringern. Mit organisatorischen Pausen würden für diese Schülerzahl noch etwa 30 Schulstunden benötigt, womit die Beobachtungen bei nur zwei zur Verfügung stehenden Teams innerhalb einer Woche durchführbar sein müssten.

Weitere Reduzierungen der Untersuchungsanordnungen, insbesondere der Verzicht auf eine mündliche Sprachprobe in der schwachen und starken Sprache sind bei dem Anspruch einer validen Erfassung bilingualer Sprachfähigkeiten in diesem Alter nicht vertretbar.

Stark vermindern würde sich die notwendige Auswertungszeit, da pro Kind nur noch 2 Transkriptionen (Interviews) anzufertigen und 30 Minuten Tonaufzeichnung abzuhören wären, bevor die Auswertung auf den vorstrukturierten Bögen erfolgen kann. Je nach Übung würden dazu nur noch etwa 4 bis 5 Zeitstunden (gegenüber ca. 10 Stunden bei der Pilotfassung) benötigt.

### **Eine Überprüfung der Gewichtung und Bewertung**

Sowohl in der Untersuchung zum Schulanfang als auch in der Untersuchung gegen Ende des 2. Schuljahres wurde bei der Bewertung eine implizite Gewichtung nach der vermuteten Relevanz der Untersuchungsbereiche für die allgemeine Beurteilung der Sprachkompetenz vorgenommen. Da der Beitrag eines Bereichs zum Gesamtergebnis aber ausschließlich von seiner Variabilität und nicht etwa von der Anzahl der auf ihn maximal entfallenden Punktwerte abhängt, ist eine solche implizite Wägung bei Diagnoseverfahren jedoch nur dann zweckmäßig, wenn die auftretende Streuung eines Bereichs der zuerkannten Bedeutung proportional entspricht.

Betrachtet man nun beim Beobilingua-2dit-Verfahren die Standardabweichungen der einzelnen Untersuchungsbereiche nach Sprachstatus getrennt (Tabellenblätter 13 + 14), fallen, abgesehen von einer allgemeinen Untergewichtung des kommunikativen Sprachverhaltens in der starken Sprache und der Partnersprache, in der starken Sprache leichte Unterbewertungen nur in den Bereichen Wortschatz/ Bild ( $s = 0,62$ ; 4 Punkte bei Bezug 25 entsprechen stattdessen nur 0,47) und Lesefertigkeit (0,58 statt

0,35 bei 3 Punkten) auf. Ansonsten ergeben sich bei der starken Sprache hinsichtlich der vorgenommenen und unter Berücksichtigung der tatsächlich aufgetretenen Streuung angemessenen Gewichtung kaum Diskrepanzen. Die Bewertung in den Bereichen Phonematik/Prosodie und Hörverständnis entspricht auch bei der Partnersprache der nach Überprüfung der Streuung angebrachten Gewichtung. Die Bewertung der Lesefertigkeit mit maximal 3 Punkten erscheint dort ebenfalls angebracht. Beim mündlichen Sprachgebrauch tritt aber bei der Partnersprache eine viel größere Varianz auf als bei der starken Sprache, so dass die Gewichtung der Pilotfassung im schwachen Leistungsbereich relativ stark von einer angemessenen Gewichtung abweicht (Ps Wortschatz:  $s = 1,49$  gegenüber  $0,82$  bei 4 Punkten/ Bezug 22; Ps Morphosyntax:  $s = 1,75$  statt  $1$  bei 5 Punkten).

Nun sollte ein Verfahren zur Beobachtung einer bilingualen Sprachkompetenz zwecks Vergleichbarkeit der Ergebnisse aber im schwachen wie im starken Leistungsbereich denselben Maßstab anlegen, so dass bei einer Überarbeitung der Gewichtung die Standardabweichungen des Verfahrens bei allen Schülern insgesamt herangezogen werden sollten (siehe Tabellenblatt 24). Entsprechend erscheint dann eine leicht geringere Gewichtung bei den Bereichen Lesefertigkeit und Hörverständnis sowie eine stärkere Betonung beim kommunikativen Sprachverhalten, dem Bereich Phonematik/Prosodie und vor allem bei den Bereichen Wortschatz und Morphosyntax (mündlicher Sprachgebrauch) angebracht. Die zu niedrige Gewichtung der Pilotfassung im Bereich des mündlichen Sprachgebrauchs bei der Partnersprache wird durch diese Änderung korrigiert.

Von einer expliziten Wägung anhand der Validitäts- oder Reliabilitätswerte der Untersuchungsbereiche wird abgesehen, denn die Reliabilitätskoeffizienten der Untersuchungsbereiche scheiden als Gewichtungskriterium aus, weil sich die Werte des Verfahrens gegenüber der Pilotuntersuchung durch verschiedene Änderungen bei der Bewertung unterscheiden dürften und die von der auftretenden Streuung abhängende Reliabilität der Bereiche erst bei einer Erprobung der revidierten Fassung ermittelt werden kann. Auch eine multiple Korrelation mit dem Validitätskriterium wird nicht erwogen, da sich die implizite Wägung der Pilotfassung als ziemlich zutreffend erwiesen hat und daher eher als Anhaltspunkt zur Bestimmung der Gewichte dienen kann als das teilweise ziemlich unzuverlässige Außenkriterium Lehrerurteil.

Obwohl auch die Faktorenladungen bei einer erneuten Durchführung des Verfahrens leicht von denen der Pilotuntersuchung abweichen werden, wurde zur Überprüfung der Gewichtung anhand der Standardabweichungen berechnet, welche Gewichte sich bei einer expliziten Wägung unter Anwendung einer von Spearman entwickelten Formel

anhand der Ladungen der Untersuchungsbereiche im gemeinsamen Faktor A ergeben würden (Tabellenblatt 24). Danach stimmt die Gewichtung bis auf den Bereich des kommunikativen Sprachverhaltens tendenziell überein. Auch entsprechend der Ladung ist den Komponenten des mündlichen Sprachgebrauchs ein weitaus größeres Gewicht zu geben als den übrigen Bereichen, denn auf den lexikogrammatischen Bereich entfällt die höchste Ladung. Dieses Ergebnis entspricht übrigens auch den Ladungsanteilen bei kanadischen Untersuchungen zur bilingualen Sprachkompetenz der Immersionsschüler (6. Schuljahr). Dem Hörverständnis gebührt nach der Ladung wiederum ein ähnlicher Anteil wie dem Bereich Phonematik/Prosodie. Diese Anteile wirken sich nach der Berechnung anhand der Ladung gegenüber dem mündlichen Sprachgebrauch anscheinend stärker aus, was aber daran liegt, dass weder der Wortschatz noch die Morphosyntax beim Bildimpuls-gesteuerten Interview einzeln in die Faktorenanalyse einbezogen wurden und in Ermangelung der Daten stattdessen die Ladung des aus diesen Komponenten zusammengesetzten Bereichs zur Berechnung herangezogen wurde. Der Lesefertigkeit wäre nach der Ladung ebenfalls ein geringeres Gewicht einzuräumen als den anderen bereits erwähnten Bereichen. Eine starke Diskrepanz tritt allerdings hinsichtlich des kommunikativen Sprachverhaltens auf, das nach der Ladung ähnlich niedrig zu gewichten wäre wie die Lesefertigkeit – statt etwas höher als das Hörverständnis und der Bereich Phonematik/Prosodie nach s. Insofern erscheint eine kleine Abminderung des sich nach der Standardabweichung ergebenden Gewichts angebracht (4,6 abgerundet zu 4 statt aufgerundet zu 5). Ansonsten ist meines Erachtens den Gewichtsanteilen, die empirisch anhand der Standardabweichungen ermittelt wurden, der Vorzug zu geben. Gleichzeitig empfiehlt sich eine Erhöhung der maximalen Gesamtpunktzahl auf 30, da die Abstufungen dann leichter ganzen Zahlen zuzuordnen sind.<sup>263</sup>

Eigentlich wäre angesichts der Mehrdimensionalität des komplexen Merkmals der Sprachkompetenz, die mit dem vorgestellten strukturierten Beobachtungsverfahren entsprechend heterogen erfasst wird, auch die Erstellung eines Leistungsprofils angemessener als die Addition der bei Teilbereichen erreichten Punkte zu einem Gesamtergebnis.<sup>264</sup> Dementsprechend wäre nach der getrennten Auswertung bei den einzelnen Sprachbereichen ein Sprachleistungsprofil zu erstellen.

---

<sup>263</sup> Die bei der Rundung entstehenden Ungenauigkeiten sind wohl zu verschmerzen, da die revidierte Gewichtung von den Standardabweichungen der Untersuchung am deutsch/italienischen Standort der SESB abgeleitet wurde und bei einer anderen Zusammensetzung der Probandengruppe mit dem Auftreten anderer Varianzen zu rechnen ist. Angaben zur Gewichtung mit mehreren Kommastellen würden nur eine Scheingenauigkeit vortäuschen. Aber die Orientierung an den Daten der Pilotstudie dürfte die Gewichtung und somit auch die Reliabilität des Verfahrens noch erhöhen.

<sup>264</sup> Siehe Jürgen Bortz, *Statistik für Sozialwissenschaftler*, Berlin u.a. 1989<sup>3</sup>, S.618.

Unabhängig davon, ob nun am Ende der Untersuchung die Beobachtungen in einem Leistungsprofil oder einem Gesamtergebnis zusammengefasst werden, sollten zur Steigerung der Reliabilität bei einigen Untersuchungsbereichen die Trennschärfe und die Schwierigkeit erhöht werden. Hier dürfte es bei den meisten Bereichen ausreichen, unter Beibehaltung der Auswertungskriterien der Pilotfassung strengere Maßstäbe bei der Bewertung anzulegen. Lediglich beim Bereich Hörverständnis (Bildimpuls-gesteuertes Interview) ist außerdem die Ergänzung durch weitere Kriterien zu empfehlen.

Da das Verfahren jedoch ein breites Spektrum von Spracherwerbsstadien erfassen soll, werden nicht für alle Zielgruppen gleichzeitig mittlere Schwierigkeiten zu erreichen sein. Grundsätzlich werden die Kriterien für die deutsche Sprache wegen des Umgebungsfaktors eher zu erfüllen sein als bei der nichtdeutschen Sprache als Zweitsprache. Aber auch bei den Untersuchungsbereichen, die sich vor allem in der deutschen Sprache als wenig trennscharf erwiesen, lässt sich die Trennschärfe durch eine leichte Veränderung der Auswertungsschlüssel erhöhen. Beim Hörverständnis bei Bild B (Trennschärfe 0) und Bild A (0,30) hat sich eindeutig ein Zusammenhang zwischen geringer Trennschärfe und geringer Schwierigkeit (Ps I: 0,50 [bei Bild A], 0,89 [Ps D u. st. D bei Bild B] bis 0,92 [st. I bei Bild A] ergeben. Hier wird das Heraufsetzen der Schwierigkeit auch die Trennschärfe steigern.

Bei der Lesefertigkeit indessen lässt sich aufgrund der anlässlich der vorliegenden Untersuchung erneut bestätigten Spezifität dieser Sprachfähigkeit die Trennschärfe nicht wesentlich verbessern, da eine gute Lesefertigkeit – zumindest bei der für das Verfahren maßgeblichen Altersstufe von 8 bis 9-jährigen Kindern – nicht zwangsläufig mit einem weit entwickelten mündlichen Sprachgebrauch einhergeht. Hier lässt sich die Trennschärfe allenfalls durch eine Erhöhung der Schwierigkeit etwas anheben, was sich dann jedoch auf Beurteilungen von Leistungen vor allem in der Partnersprache auswirken wird. Durch strengere Bewertungsmaßstäbe wird sich die Schwierigkeit für die Zielgruppe Partnersprache insgesamt erhöhen, was aber durchaus mit der Intention des Verfahrens in Einklang steht, besonders im schwächeren Bereich zu differenzieren. Solange insgesamt eine mittlere Schwierigkeit des Diagnose-instrumentariums erreicht wird, ist es durchaus legitim, extrem schwierig oder leicht zu bewältigende Kriterien zu integrieren. Die Anforderungen dürfen nur nicht so extrem erhöht werden, dass die Kinder, die sich noch in schwächeren Stadien des Spracherwerbs befinden, keine Chance haben, wenigstens die eine oder andere Aufgabe ansatzweise zu bewältigen, denn sonst wäre das Gegenteil einer optimalen Trennschärfe erreicht: eine Null-Differenzierung.



## Die Veränderung einiger Untersuchungsanordnungen

Die Bildimpuls-gesteuerten Interviews in beiden Sprachen sowie das Spiel 1 zur Überprüfung des Hörverständnisses in der Partnersprache haben sich hinsichtlich des Bildmaterials wie auch des Ablaufs während der Untersuchung bereits gut bewährt, so dass, abgesehen von der Eliminierung der gleichzeitigen Erhebungen zum mündlichen Sprachgebrauch beim Spiel 1 über die Realisierung einiger elementarer Sprechakte, auf die in diesem Bericht bereits bei der Darstellung zum Ablauf der Untersuchungsanordnung hingewiesen wurde, keine Änderungen vorgeschlagen werden. Doch bei folgenden Untersuchungsanordnungen wäre vermutlich die Beobachtung der dabei relevanten Sprachleistungen durch einen veränderten Ablauf zu optimieren:

### Textverständnis (nur starke Sprache)

Statt der ehemaligen Partneraufgabe, die mitunter, wenn die Reihenfolge der Antworten im Protokoll nicht genau vermerkt worden war, auch viel Mühe bei der Auswertung bereitete, wird nunmehr – unter der Voraussetzung einer Durchführung gegen Mitte des 3. Schuljahres – für die Überprüfung des Textverständnisses in der starken Sprache eine im Klassenverband durchführbare Aufgabe vorgeschlagen, die das sinnentnehmende Lesen und schriftliche Kurzantworten integriert. Zur Vermeidung von Zufallseinflüssen wird das freie Ergänzen von Schlüsselwörtern einer gebundenen Fragebeantwortung vorgezogen. Um wiederum einen hemmenden Einfluss auf die Antworten wegen einer eventuellen Rechtschreibunsicherheit zu verhindern, werden Rechtschreibfehler nicht gewertet, worauf die Schüler hinzuweisen sind. Da sich der in der deutschen Sprachgruppe dargebotene pointierte, in sich geschlossene Text *„Fremdsprachen sollte man können...“* nach Erfahrung einiger Beobachter besser bewährt hat als das verkürzte Märchen *Prezzemolina*, das den italienischen Schülern zum Teil wenig interessant erschien, sollte der deutsche Text in entsprechender Übersetzung auch bei der anderen Sprachgruppe eingesetzt werden, wodurch gleichzeitig auch die Vergleichbarkeit optimiert wird. Statt der mündlichen Befragung wird nun folgender Ablauf vorgeschlagen: Wie zuvor wird die Geschichte mündlich, aber diesmal vom vertrauten Lehrer und vor der ganzen Lerngruppe vorgetragen. Nachdem an einem Beispiel die Form der geforderten Kurzantwort geklärt wurde, erhält jeder Schüler einen Bogen mit 7 Fragen zum Text, die er still erlesen und anschließend schriftlich beantworten soll. Da zum Textverständnis als Komponente im Bereich Hörverständnis maximal 2 Punkte vorgesehen sind, wird als Bewertung festgelegt: 2 Punkte bei 7 - 1,5 Punkte bei 6 - 1 Punkt bei 5 - 0,5 bei 4 - 0 Punkte bei weniger als 4 kohärenten Antworten.

Beispiel: **Welche** Maus hat Angst vor der Katze? – erwartete Antwort: *die große Maus.*

### Lesefertigkeit und Phonematik/Prosodie beim Lesen

Die in der Pilotfassung vorgenommene Leseübung in der starken Sprache anhand eines literarischen Textes in der Großgruppe entfällt. Stattdessen sollten die Schüler beider Sprachgruppen (starke Sprache und Partnersprache) zur besseren Vergleichbarkeit jeweils einzeln dieselbe Textvorlage, die sprachspezifisch schwierige Phoneme enthält, laut vorlesen. Die Beispielsätze der Pilotfassung Ps-D bzw. Ps-I sollten aber im Umfang reduziert werden. Bei der deutschen Vorlage sind zur besseren Beobachtung prosodischer Fähigkeiten die Aussagesätze teilweise in Fragen bzw. Aufforderungen umzuformen.

Alternative, die den Kindern gewiss mehr Spaß bereiten würde, aber zusätzliches Bildmaterial erfordert, keine Beobachtungen zur Satzintonation zulässt und außerdem eine Beschränkung auf leichter zu visualisierende Nomen auferlegt: Ein Stapel mit Bildkarten, auf denen auf der Rückseite jeweils das passende Wort geschrieben steht, wird vor den Augen des Kindes so gemischt, dass mal das Bild, mal das Schriftbild zu sehen ist. Das Kind soll dann die Karten der Reihe nach abheben und entweder zum Bild das entsprechende Wort nennen oder erst das Schriftbild erlesen. Danach oder bei Leseschwierigkeiten darf das Kind dann die Karte auf die Bildseite umdrehen.

## Spiel 2

Bei diesem Spiel sollen gezielt Beobachtungen zum bilingualen Sprachverhalten ermöglicht werden. Es wird in einer gemischtsprachigen Lerngruppe als Wettkampf zwischen zwei Schülergruppen, die sich in gleicher Anzahl aus Kindern beider Sprachgruppen (etwa je 6 Kinder) zusammensetzen, in der Art einer Staffel durchgeführt. Vor Spielbeginn werden (nicht im Beisein der Kinder) verschiedene Gegenstände im Klassenraum versteckt. Dann nehmen die Schüler einer Gruppe hintereinander auf Stühlen Platz. Beim stark veränderten Spielablauf erhalten die Plüschtiere „Känguru“ und „Storch“, die sich bei der ursprünglichen Version zur Anregung verbaler Kommunikation bewährt haben, eine tragende Funktion. Die vertrauten Lehrkräfte übernehmen jeweils eine der Handpuppen. Nach der Vereinbarung, dass die Tiere jeweils nur die Sprache der für diesen Unterricht zuständigen Lehrkraft verstehen, kann das Spiel beginnen. Damit die Sprachäußerungen von den Beobachtern verstanden werden können, führen die Gruppen die Aufgabe aber nicht gleichzeitig, sondern jeweils im Wechsel aus. Der Reihe nach dürfen die Schüler einer Gruppe bei einem Tier – natürlich in dessen Sprache – eine Information (Merkmale wie Größe, Farbe; Funktion; Lage des Verstecks) zum gesuchten Gegenstand erfragen. Diese Information geben sie an das nächste Kind weiter, das dann beim anderen Tier weitere Informationen einholt. Das letzte Kind der Gruppe soll dann nach den gesammelten Angaben den Gegenstand holen. Danach werden die Plätze getauscht, so dass jedes Kind im Verlauf der verschiedenen Suchaktionen mal die eine, mal die andere Sprache

390

benutzen muss. Die Schüler einer Gruppe dürfen sich untereinander helfen. Bei dieser veränderten Aufgabenstellung, die einen mehrfachen Sprachwechsel impliziert, können sich die Schüler der sprachlichen Kommunikation nicht mehr entziehen.

Der Einsatz dieses strukturierten Beobachtungsverfahrens ist nicht auf eine deutsch/italienische Sprachkombination limitiert. Während die Bildvorlage B wegen der dabei visualisierten Aktivitäten, die eine Benutzung von im Deutschen trennbaren Verben nahe legt, speziell für Erhebungen zum deutschen Sprachgebrauch geeignet ist, lässt sich das übrige Material bei Adaptation der Textvorlagen (Sätze zum Lesen, erzählter Text für die Aufgabe zum Textverständnis) und der Auswertungskriterien bei den Bereichen Phonematik/Prosodie und Morphosyntax auch bei anderen Sprachkombinationen anwenden. Somit ist die überarbeitete Fassung des bei der Untersuchung am deutsch/italienischen Standort der SESB eingesetzten Pilotverfahrens als Entwurf für ein allgemeines Verfahren zur Evaluierung einer bilingualen Kompetenz anzusehen. Unter Berücksichtigung der Überarbeitungsvorschläge würde sich das Verfahren zur Beobachtung bilingualer Sprachfähigkeiten bei 8 bis 9-jährigen Kindern wie folgt zusammensetzen:

Untersuchungsanordnungen und -bereiche – Übersicht						
Untersuchungsanordnung		Organisation	Zeit	Untersuchungsbereich	Bewertung	Sprache
Bild A (I) Bild B (D)	Bildimpuls- gesteuertes Interview	einzel	je 7 Min.	Phonem./Prosodie Hörverständnis mündl. Sprachgebrauch: Wortschatz Morphosyntax	max. 2 max. 2 max. 7 max. 8	D + It
Sätze	Leseübung	einzel	je 3 Min.	Phonem./Prosodie Lesefertigkeit	max. 2 max. 3	D + It
erzählter Text (+ schriftl. Fragen + Kurzantworten)		in gesamter Sprachgruppe	20 Min.	Textverständnis	max. 2	nur starke Sprache
Spiel 1	Spielplan	Partner derselben Sprachgruppe	10 Min.	Hörverständnis		nur Ps
Spiel 2	Spielsituation	in gemischtspra. Teilungsgruppe	30 Min.	bilinguales kommunikat. Sprachverhalten	nur mit Symbolen	D + It
Protokoll bei allen Hospitationen und Untersuchungsanordnungen			-	kom. Sprachverhalten	max.4	D + It
					max. 30 Punkte	je Sprache