

Teil V

Ergebnisse der Untersuchung (Ende 2. Schuljahr) am deutsch/italienischen Standort der SESB

15 Vorbemerkungen zur Interpretation der Untersuchungsergebnisse

Vor der Darstellung und Interpretation der an der SESB beobachteten Entwicklung der Sprachfähigkeiten ist zunächst die Aussagekraft des für die Untersuchung entwickelten Instrumentariums anhand der wichtigsten Gütekriterien zu überprüfen. Da die Analyse der Pilotfassung des Beobilingua-2dit-Verfahrens mittels geeigneter statistischer Verfahren detailliert im Teil VI beschrieben wird, sollen zur Vermeidung von Wiederholungen an dieser Stelle jedoch allgemeine Angaben als Voraussetzung für ziemlich gesicherte Aussagen über die gegen Ende des 2. Schuljahres bei den Schülern beobachteten deutschen und italienischen Sprachfähigkeiten sowie die herausgefundenen Zusammenhänge bei der sprachlichen Entwicklung genügen. Wie die Analyse ergeben hat, ist die Qualität der an der SESB durchgeführten Untersuchung durchaus zufriedenstellend, wenngleich auch einige bei einer späteren Überarbeitung vorgesehene Veränderungen die Reliabilität noch steigern können. Schon die Pilotfassung des Beobachtungsverfahrens weist eine sehr gute Auswertungsobjektivität mit Korrelationen von über 0.95, Kriteriumsvaliditäten von 0.65 (starke Sprache) und 0.84 (Partnersprache) sowie Reliabilitätskoeffizienten zwischen 0.83 (Partnersprache Deutsch) und 0.94 („starke“ Sprache Italienisch) auf.

Obwohl sowohl bei der Durchführung als auch bei der Auswertung ein Höchstmaß an Objektivität, Reliabilität und Validität angestrebt wurde und entsprechend den Analysedaten auch weitgehend eingehalten werden konnte, ist dennoch ein gewisser Vorbehalt bei der Interpretation der Daten angebracht. Im Vergleich zu anderen longitudinalen Untersuchungen im Bereich der Entwicklung von bilingualer Sprachkompetenz, die sich meistens auf nur wenige Fallstudien beschränken, haben an der vorliegenden Untersuchung zwar ziemlich viele Kinder teilgenommen, statistisch gesehen handelt es sich aber immer noch um eine relativ geringfügige Anzahl von Probanden (zum Schulanfang 51 Schüler der SESB, zum Ende des 2. Schuljahres noch 36 von diesen 51 Schülern der SESB und 19 Schüler aus monolingualen deutschen bzw. italienischen Kontrollgruppen). Außerdem befindet sich die Entwicklung der Sprachkompetenz besonders in der Partnersprache, wenn sie die Zweitsprache darstellt, im 2. Schuljahr erst noch am Anfang. Entsprechend sind alle in Rohwerte transponierten Beobachtungen, Korrelationen sowie statistischen Kennwerte nicht als endgültige, absolut geltende Ergebnisse aufzufassen, sondern nur als Zwischenbericht zu verstehen, der nachweisbare Entwicklungstendenzen lediglich bei einem Jahrgang in einer Sprachkombination der SESB aufzeigt. Allgemeine Schlüsse in Bezug auf das Konzept des Schulversuchs der Staatlichen Europa-Schule Berlin aufgrund dieser empirischen Erkenntnisse wären übereilt.

Die Untersuchungsergebnisse können zwar ohne Bedenken zu gruppeninternen Vergleichen an der SESB und zum Vergleich mit monolingualen Lerngruppen oder anderen bilingualen Schulmodellen herangezogen werden, sie sollten aber keinesfalls zur Bewertung einzelner Schüler benutzt werden. Auch die sich aus der Untersuchung ergebenden Interkorrelationen der Sprachfähigkeiten sind absolut nicht als eindeutig bewiesen zu betrachten, denn Interkorrelationsberechnungen reagieren mitunter sehr empfindlich auf Abweichungen bei den Verhältnissen zwischen zwei Variablen. Ab einer bestimmten Größenordnung spielen vereinzelt Abweichungen zwar kaum noch eine Rolle, aber bei nur wenigen Probanden (die Partnersprachgruppen umfassen jeweils nur 18 Schüler) wirken sich selbst nur geringfügige Inkonsistenzen sehr stark aus.²⁰² Daher wird in statistischen Lehrbüchern zur Vermeidung von Verzerrungen häufig empfohlen, sogenannte Ausreißer nicht zu berücksichtigen²⁰³. Aber bei einer solchen Vorgehensweise sehe ich die Gefahr einer willkürlichen Manipulation, denn zur Herstellung erwarteter Zusammenhänge könnten doch beliebig viele Werte als „Ausreißer“ betrachtet werden. Daher habe ich bis auf wenige Ausnahmen, die jeweils als solche gekennzeichnet wurden (z. B. bei der Ermittlung der Vergleichsnormen für eine altersgemäße Sprachkompetenz ohne Berücksichtigung der beiden Schüler, die das 2. Schuljahr wiederholen müssen), von der Eliminierung solcher Ausreißerwerte abgesehen, zumal bei der an der SESB durchgeführten Untersuchung Erkenntnisse über allgemeingültige Zusammenhänge bei der Entwicklung von Bilingualität nicht im Vordergrund stehen. Hauptsächlichste Intention der Studie ist die Beobachtung der unter den organisatorischen und konzeptionellen Bedingungen der SESB bei einer zufälligen Schülerzusammensetzung tatsächlich feststellbare Sprachentwicklung.

Hinsichtlich der Punktwert-Resultate muss noch auf Folgendes hingewiesen werden: Vor Beginn der statistischen Analysen erwies es sich als notwendig, die bei der Auswertung ermittelten Rohwerte auf eine einheitliche Skala zu transformieren. Dabei wurde angesichts des Erprobungscharakters des an der SESB benutzten Beobachtungsverfahrens auf eine Normierung unter Verwendung von Standardwerten oder Prozenträngen verzichtet. Zum Vergleich zwischen den an der Untersuchung beteiligten Lerngruppen reichen statistische Berechnungen anhand der Rohwerte auch aus, sofern ihnen dasselbe Bezugssystem zugrunde liegt. Doch bei der Konstruktion der Pilotfassung wurde die Einhaltung eines direkten Vergleichsmaßstabs nicht hinreichend bedacht, weshalb vor der Anwendung von Rechenverfahren eine

²⁰² In der Tat bewirkte eine probeweise durchgeführte Berechnung der Interkorrelation zwischen Textverständnis und Hörverständnis bei Bild B eine Erhöhung von $r = 0.10$ auf $r = 0.75$, als die Werte von dem durch seine Inkonsistenz mehrfach auffälligen Kind C1 dem Sprachverhalten der Mitschüler im Verhältnis leicht angeglichen wurden.

²⁰³ Z.B. bei Lienert 1969, a.a.O., S.280

nachträgliche Transformation der bei der starken Sprache bzw. der Partnersprache erzielten Punktwerte auf ein einheitliches Bezugssystem angebracht erschien.²⁰⁴

Bei der Pilotfassung des Verfahrens war bei der starken Sprache wie bei der Partnersprache eine Bewertung mit einem festgelegten Maximum von 25 Punkten vorgesehen. Doch wurden in der Partnersprache noch keine Beobachtungen zur Schreibfertigkeit erhoben, weil das im 2. Schuljahr der SESB wegen der konzeptionell noch nicht vorgesehenen Übung verfrüht gewesen wäre. Bei Vergleichen zwischen den Leistungen in der starken Sprache und der Partnersprache darf der auf die Schreibfertigkeit entfallende Anteil also nicht berücksichtigt werden. Leider genügt es aber nicht, den dafür vorgesehenen Anteil von maximal 3 Punkten bei Vergleichen zwischen den Lerngruppen bei den Ergebnissen der starken Sprache abzuziehen, denn die in der Pilotfassung festgelegte Bewertung sah trotz teilweise identischer Untersuchungsanordnungen bei mehreren Sprachleistungen in der Partnersprache eine von der starken Sprache abweichende Bewertung vor. Statt bei den Lerngruppen unterschiedliche, um die Punkte für die Schreibfertigkeit variierende Maximalwerte festzulegen, wurde für die maximal 3 auf die Schreibfertigkeit entfallenden Punkte bei der starken Sprache ein rechnerischer Ausgleich in der Partnersprache durch Erhöhung der maximalen Punktwerte (um je 1) in drei Bereichen (Hörverständnis, Morphosyntax und Lesen) vorgenommen, womit sich sowohl bei der starken Sprache wie bei der Partnersprache zwar der gleiche maximale Punktwert von 25 ergab, in dieses Ergebnis aber unterschiedlich bewertete Leistungen einfließen.

Diese nunmehr als problematisch erkannte unterschiedliche Bewertung erschien mir bei der Entwicklung des Verfahrens sachlich hinreichend begründet. Einerseits sollten die zum Schulanfang annähernd bilingualen Schüler bei der Untersuchung – zum Zwecke des Entwicklungsvergleichs – auch in der Partnersprache den Maximalwert von 25 erreichen können, ohne dass bei ihnen die Schreibfertigkeit in der schriftlich noch nicht geübten Sprache beobachtet würde. Außerdem wollte ich gerade bei guten

²⁰⁴ Die Berechnung von Interkorrelationen würde zwar von unterschiedlichen Bezugssystemen nicht beeinflusst, da es hierbei allein auf das Verhältnis zwischen den Datenreihen ankommt. Doch schon die Beziehung zum Gesamtergebnis ändert sich je nachdem, ob ein Maximum von 25 Punkten (unter Einbezug der Schreibfertigkeit bei der starken Sprache) oder von 22 Punkten den Bezugswert bildet. Bei allen statistischen Rechenverfahren, die den Mittelwert und die Standardabweichung der Rohwert-Verteilungen verwenden, ist ein einheitliches Bezugssystem für den Vergleich unabdingbar.

Leistungen den Lernerfolg unter weniger geförderten Bedingungen anerkennen. Als Pädagogin ließ ich mich dazu wohl durch ein diffuses „Gerechtigkeitskonzept“ verleiten, das bei der Bewertung der Leistungen von Kindern, die unter ungünstigen Bedingungen aufwachsen, deren Lernbedingungen einbezieht und oftmals die tatsächliche Leistung verhältnismäßig überbetont. Derlei pädagogisch vertretbare Ansichten sind bei einer wissenschaftlichen Untersuchung, deren Ergebnisse statistisch überprüfbar sein sollen, aber nicht angebracht. Obendrein erübrigt sich bei dieser Untersuchung zur Entwicklung der Sprachkompetenz, jedenfalls bei der Auswertung, die hervorhebende Anerkennung der Leistung in der Partnersprache/ Zweitsprache, da die beobachteten Schüler an dieser Phase nicht beteiligt sind. Während der Beobachtungsphase sollten die Schüler zwecks Ermunterung zu Sprachäußerungen jedoch unbedingt unabhängig von der tatsächlichen Leistung immer wieder gelobt werden, wie das die studentischen Mitarbeiter entsprechend meiner Anleitung auch während der Untersuchung an der SESB befolgt haben.

Bei der Überarbeitung des Verfahrens ist die Vergleichbarkeit der Ergebnisse leicht zu optimieren, indem von vornherein konsequent derselbe Maßstab zur Bewertung der Leistungen festgelegt wird. Hinsichtlich der Pilotfassung sind vergleichbare Bezüge nachträglich herzustellen, wobei die von den studentischen Mitarbeitern entsprechend der ursprünglichen Anleitung vorgenommene Bewertung (Tabellenblatt 3, Spalten 13 bis 15) als Grundlage dient. Bei der starken Sprache ist das Gesamtergebnis durch ein Ergebnis unter Ausklammerung der Schreibfertigkeit zu ergänzen. Bei der Partnersprache sind die ermittelten Punktwerte beim mündlichen Sprachgebrauch (nur im Bereich Morphosyntax, $6 \rightarrow 5$), dem Hörverständnis ($5 \rightarrow 4$) und der Lesefertigkeit ($4 \rightarrow 3$) der vorgesehenen Bewertung in der starken Sprache anzugleichen. Nach dieser Umrechnung liegen dann bei den Lerngruppen der SESB neben den Angaben mit dem Bezug 25 zur Berücksichtigung der Schreibfertigkeit bei der starken Sprache auch vergleichbare Resultate mit einem maximalen Gesamtwert von 22 Punkten vor (Tabellenblatt 3, Spalten 16 bis 18).

In diesem Zusammenhang der Vergleichbarkeit der sprachlichen Leistungen an der SESB erscheint noch ein weiterer Hinweis auf eine gewisse Konstruktionsschwäche des Verfahrens angebracht, ohne damit jedoch unterschiedliche Aufgabenstellungen angesichts abweichender Lernvoraussetzungen an dieser Stelle grundsätzlich in Frage zu stellen. Bei der Interpretation der Ergebnisse ist jedoch zu beachten, dass die Leistungen in der Partnersprache und der starken Sprache trotz der Transformation auf eine gleiche Bewertungsskala dennoch nur eingeschränkt vergleichbar sind, da die Untersuchungsanordnungen zum Textverständnis (nur starke Sprache) bzw. zum Hörverständnis beim Spiel 1 (nur Partnersprache) sowie zur Lesefertigkeit unter-

schiedliche Anforderungen enthalten. Zwar ergibt die Aufgabenanalyse bei einigen Aufgaben ähnliche Schwierigkeitsgrade, etwa beim Textverständnis und dem Hörverständnis beim Spiel 1 einen viel zu leichten Schwierigkeitsgrad um 0,9 in der deutschen Sprache, aber bei der italienischen Sprache wirken sich die verschiedenen Lernvoraussetzungen der Lerngruppen weitaus stärker aus. Streng betrachtet sind daher lediglich die Leistungen bei den Bildimpuls gesteuerten Interviews (Darbietung derselben Bildvorlagen und Anwendung derselben Auswertungskategorien bei den Bereichen Hörverständnis, Phonematik/ Prosodie, Wortschatz und Morphosyntax) miteinander vergleichbar.

Die Transformation auf den Bezugswert 22 wirkt sich auf die beiden Lerngruppen verschieden aus. Während sich bei der starken Sprache einige Rangplätze gemäß dem Mittelwert verschieben, da nunmehr die Berücksichtigung der Schreibfertigkeit entfällt, bleiben die Verhältnisse bei der Partnersprache durch die Umwandlung mittels Multiplikation mit Konstanten erhalten. Hier ergeben sich jedoch vereinzelt wegen der an der altersgemäßen Sprachkompetenz orientierten Grenzwerte andere Zuordnungen zu den Lerngruppen. Drei deutsche Schüler, die beim ursprünglich vorgesehenen Bezugssystem als annähernd bilingual einzustufen gewesen wären, gehören nunmehr noch der Lerngruppe 2 sPi an. Der nachstehende Auszug aus dem Tabellenblatt 3 soll die Auswirkungen der Transformationen veranschaulichen.

Partnersprache Italiano				starke Sprache Italiano				
C1	C12	C8	D5	C6	C16	C22	C23	Code
16,5	4,19	9,64	19,57	10,88	16,25	20,31	23,88	Bezug: max. Gesamtpunktwert 25
14,04	3,43	8,34	17,25	8,38	14,25	18,44	21,75	Bezug: max. Gesamtpunktwert 22

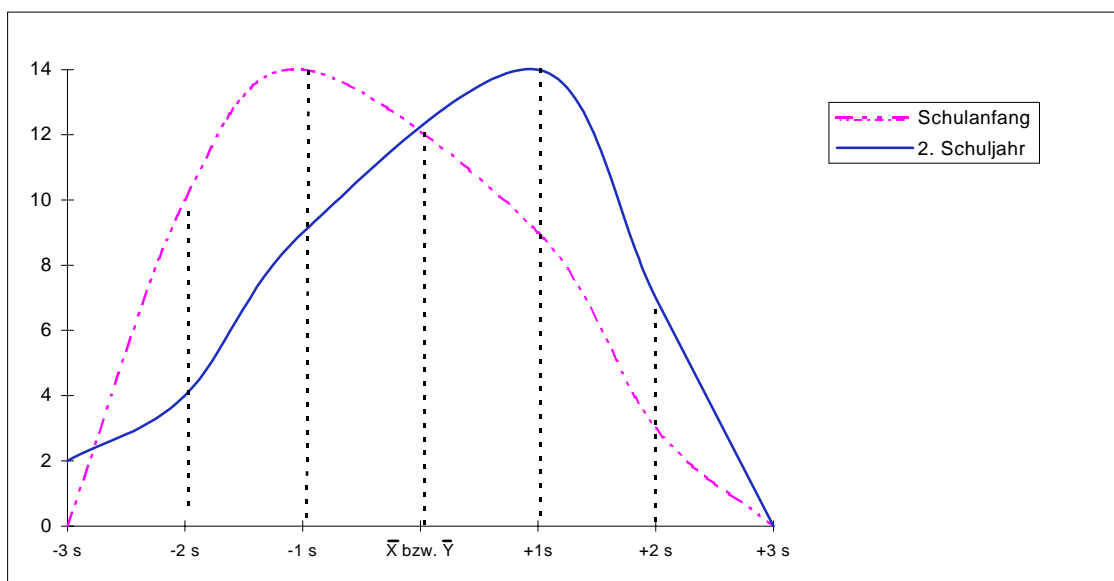
Bei Vergleichen der SESB-Lerngruppen in der starken Sprache mit monolingualen Kontrollgruppen wird unter Einbezug der Schreibfertigkeit selbstverständlich das Maximum von 25 als Bezugswert beibehalten, ebenfalls bei Betrachtungen und Berechnungen, die sich nur auf die Sprachkompetenz in der starken Sprache beziehen. Wegen der dadurch verbleibenden unterschiedlichen Bezugssysteme wird bei den folgenden Ausführungen zur Vermeidung von Missverständnissen der jeweilige Bezug soweit erforderlich angegeben.

Übrigens werden auf dem Tabellenblatt 3 in den Spalten 10 und 11 die sich nach dem Mittelwert bei Schulanfang und bei Ende des 2. Schuljahres ergebenden Rangplätze über die sprachliche Entwicklung zwar angeführt, Berechnungen anhand von Rangplätzen aber vermieden. Da bei Rangplätzen keine gleichen Abstände vorliegen,

dürfen sie mathematisch auch nicht wie Intervallwerte behandelt werden, wodurch viele statistische Methoden nicht anhand von Rängen angewendet werden können. Daher werden zur statistischen Analyse der Ergebnisse – soweit zulässig – Produktmomentkorrelationen gegenüber Rangkorrelationen bevorzugt.

Schließlich ist noch darauf hinzuweisen, dass statistische Rechenoperationen bei dem Beobilingua-2dit-Verfahren überhaupt nur sehr begrenzt eingesetzt werden können, weil die gebräuchlichsten Methoden häufig eine Normalverteilung voraussetzen. Aber zumindest in Deutschland ist Zweisprachigkeit im Primarbereich noch nicht so allgemein verbreitet, als dass hinsichtlich der Bilingualität eine normalverteilte Population vorliegen könnte. Erwartungsgemäß zeigt denn auch das nachstehende Häufigkeitspolygon, bei dem die zum Schulanfang bzw. im 2. Schuljahr an der SESB beobachteten Mittelwerte als Standardwerte eingetragen wurden, im Vergleich mit einer Normalverteilung, deren glockenförmiger Verlauf als bekannt vorausgesetzt und daher nicht abgebildet wird, außer Asymmetrien positive bzw. negative Schiefen.

Mittelwerte der SESB - Klassen c + d im Vergleich zur Normalverteilung



In Anbetracht des nicht normalverteilten Untersuchungsgegenstandes kommen hauptsächlich nur verteilungsfreie statistische Verfahren zur Anwendung, was die statistische Analyse insofern erschwert, als sie bis auf den Chi-Quadrat-Test weniger gebräuchlich sind und daher kaum Tabellen vorliegen, die zur Vereinfachung der Rechenoperationen beitragen würden.