

## „Vom Wiegen wird das Schwein nicht fett!" – Was müssen Sprachtests heute leisten?

In Kitas und Schulen wird seit Jahren beklagt, dass viele Kinder sprachlich nicht das Niveau erreichen, das sie für den Schulalltag bräuchten. Immer wenn ein Bericht – wie jetzt die neueste PISA-Studie – veröffentlicht wird, herrscht große Aufregung, aber auch Ratlosigkeit: Wir wissen jetzt, **dass** es schlecht um die Bildung unserer Kinder und Jugendlichen bestellt ist, aber nicht genau, **was** zu tun ist.

5 bis 8 Prozent der Kinder eines Jahrgangs sind von einer umschriebenen Sprachentwicklungsstörung (SES) betroffen und haben aufgrund ihrer neurobiologischen Veranlagung Probleme bei der Laut- und Satzbildung, ein begrenztes Vokabular oder Grammatikschwächen, während sie in anderen Entwicklungsbereichen altersgemäß entwickelt sind. Doch viel mehr Kinder – nämlich rund 50 Prozent in Berlin-Mitte – haben bereits bei der Einschulung sprachliche Probleme. Da Sprache in der Bildung eine Schlüsselposition hat, ist die Folge sprachlicher Probleme auch schlechte Leistungen in anderen Bildungsbereichen. Schnell wird dann gefordert: „Mehr Sprachtests! Einheitliche Sprachtests! Gute Sprachtests!“ Doch was ist ein guter Sprachtest? Wir wissen doch schon lange „Vom Wiegen wird das Schwein nicht fett.“ „Vom Testen wird das Kind nicht klug!“ Liefert der Sprachtest Antworten auf die Frage, ob es sich um eine umschriebene SES oder eine umweltbedingte Sprachstörung handelt? Gibt der Sprachtest Anhaltspunkte, wo Förderung ansetzen kann? Passt der Sprachtest zu dem veränderten Sprachverhalten in unserer Gesellschaft und dem damit verbundenen niedrigeren Sprachniveau?

### Zwischen Test und Wirklichkeit

In der quantitativen wissenschaftlichen Forschung gelten für Tests strenge Gütekriterien: **Objektivität**, **Reliabilität** und **Validität**. Sie sollen sicherstellen, dass Testergebnisse exakt, verlässlich und vergleichbar sind. Doch Sprache ist kein Blutdruckmesswert – sie ist lebendig, situativ und zutiefst menschlich.

- **Objektivität?** Das Testergebnis soll unabhängig von der Person sein, die den Test durchführt.
- **Reliabilität?** Der Test ist zuverlässig und exakt, wenn er bei Wiederholung unter gleichen Bedingungen die gleichen Ergebnisse bringt.
- **Validität?** Der Test misst tatsächlich das Merkmal, das er laut Zielsetzung messen soll.

Sprache ist aber immer subjektiv, richtet sich immer von einem Subjekt an ein anderes. Eine Äußerung bekommt eine andere Bedeutung, je nachdem, wer sie wann, wo und an wen richtet. Sie kann sich nicht identisch wiederholen. Selbst wenn wir zitieren oder die Rede anderer wiederholen, bekommt sie in der Wiederholung eine andere Bedeutung, die sich auf die aktuelle Äußerungssituation bezieht. In der einen Situation kann eine Äußerung noch angebracht sein, in der anderen schon nicht mehr. Denken wir beispielsweise an eine Vorstellungsrunde in der Klasse, wo jedes Kind seinen Vornamen und eine Lieblingsbeschäftigung mit dem Anfangsbuchstaben seines Vornamens nennen soll. Die Lehrerin reagiert erfreut, als sie von Ronny hört: „Ich bin Ronny, ich mag Rechnen.“ Als Luca,

der als Nächstes dran ist, sagt: „Ich bin Luca, ich mag Rechnen auch“, reagiert sie schon nicht mehr ganz so erfreut. Was bei R wie Ronny als sprechendes Subjekt noch gut war, passt bei L wie Luca schon nicht mehr.

Man muss sich bei einer Sprachstandserhebung also überlegen, welche sprachlichen Fertigkeiten man testen will, ob und wie diese operationalisiert werden können und wie die somit erhaltenen Ergebnisse interpretiert werden. Sprachliche Fertigkeiten liegen auf mehreren Ebenen: Ein Kind erwirbt Regeln der Lautbildung und Lautunterscheidung (Phonetik/Phonologie), der Wortformen (Morphologie) und des Satzbaus (Syntax), baut sein mentales Lexikon auf (Wortschatz), erkennt Bedeutungen auf Wort-, Satz- und Textebene (Semantik) und lernt die Regeln des Sprachhandelns (Pragmatik). Bei jeder sprachlichen Äußerung braucht es Fertigkeiten auf allen diesen Ebenen. Eine Operationalisierung einzelner herausgegriffener Fertigkeiten, beispielsweise der Aussprache, ist eventuell möglich, sagt aber kaum etwas über die gesamte Sprachfähigkeit aus

Andere Sprachfähigkeiten lassen sich nicht operationalisieren oder den Gütekriterien unterwerfen, beispielsweise das Sprachhandeln (Pragmatik). Pragmatische Fähigkeiten umfassen das situations- und kontextangemessene Sprachhandeln, also die Fähigkeit, auf die Gesprächsperson einzugehen, das eigene sprachliche Handeln an das sprachliche Gegenüber anzupassen. In standardisierten Sprachtests darf die Testleitung nur in engen Grenzen sprachlich handeln, also nicht helfen, korrigieren oder kooperieren, wie in einem natürlichen Gespräch. Von dem getesteten Kind wird nun verlangt, dass es in den Bereichen Wortschatz und Grammatik möglichst genau auf die Fragen der Testleitung eingeht, aber in der Pragmatik nicht. Ein Kind, das sich an das stereotype Verhalten der Testleitung anpasst, und sich genauso einsilbig zeigt wie die Testleitung, liefert schlechte Ergebnisse.

Ein standardisierter Sprachtest ist also ein Widerspruch in sich. Sprache und Gespräche sind gemeinschaftliche Ko-Konstruktionen von zwei (oder mehr) Personen. Man kann die Kompetenzen einer Person nicht testen, wenn die andere nicht kooperiert. Das wäre so, als ob man versucht, die Kompetenzen im Paartanz zu testen, während einer der Partner durch einen Besenstiel ersetzt wird.

Zusammengefasst: Die Gütekriterien für standardisierte Tests sind auf (Kinder-)Sprache kaum anwendbar, weil sprachliche Fähigkeiten sich nicht sinnvoll quantifizieren lassen – Sprachtests können erst dann valide sein, wenn die getestete Person über so viel bildungssprachliche und metalinguistische Kompetenz verfügt, dass sie die Pragmatik der Testsituation erfasst, was bei Kindern im Kita- und Grundschulalter nicht der Fall ist. Deswegen sind standardisierte Sprachtest zur Ermittlung der umweltbedingten Sprachverzögerung reine Zeitverschwendung. Die Testleitung müsste erst ein vom sprachförderlichen Sprechen abweichendes Sprachverhalten erlernen und dann in kostbaren 1:1-Situationen mit dem Kind als Test anwenden. Davon würde niemand profitieren.

Es wäre also zu fragen, welche Kriterien zugrunde gelegt werden sollen, wenn es um einen Untersuchungsgegenstand wie die Sprache geht, bei dem **qualitative** (und keine quantitativen) Erkenntnisse gewonnen werden sollen, man also wissen will, über welche sprachlichen Fähigkeiten ein Kind bereits verfügt und wo genau die Förderung ansetzen kann. Hier greifen die sogenannten Nebengütekriterien **Ökonomie**, **Akzeptanz**, **Nützlichkeit** und **Ethik**, die bei jedem Erkenntnisgewinn Anwendung finden sollten.

**Ökonomie:** Die Messinstrumente sollten so sein, dass Daten einfach, schnell und ressourcensparend erhoben werden können. Eine Sprachstandserhebung sollte sich ohne viel zeitlichen und personellen Aufwand in den Kita- oder Schulalltag integrieren lassen.

**Akzeptanz** beschreibt die positive Einstellung der Personen zur Untersuchungsmethode. Die Sprachstandserhebung sollte beiden Seiten, der Testleitung und vor allem dem Kind, Spaß machen und guttun.

**Nützlichkeit:** Eine Datenerhebung ist nur gut, wenn die Daten zeigen, was im Anschluss zu tun ist, wenn man praktisches Handeln ableiten kann. Eine Sprachstandserhebung ist nur sinnvoll und nützlich, wenn sie zeigt, wie das Kind im Anschluss gefördert werden kann.

**Ethik:** Ein Test sollte zumutbar, transparent und fair sein. Im Bereich der Sprachstandserhebung bedeutet dies, dass mit positivem Blick auf das geschaut wird, was das Kind kann und wo Förderung ansetzen kann. Standardisierte Tests fragen aber nach den Abweichungen von der Musterlösung. Dabei kommt es vor, dass individuelle Abweichungen als falsch gewertet werden, obwohl sie hohe sprachliche Fähigkeiten zeigen, die aber in diesem Setting nicht erfragt wurden. Auch die Fähigkeiten von Kindern, die in mehrsprachigem Kontext aufwachsen, lassen sich schwer quantifizieren. Diese zu negieren widerspricht dem ethischen Gütekriterium.

### **Aufbauen statt abfragen: das Prinzip der „Aufbauenden Sprache“**

Wie ein Vorgehen aussehen kann, das diese vier Kriterien einlöst, zeigt das Konzept der Aufbauenden Sprache, das am Sprachförderzentrum Berlin-Mitte entwickelt wurde. Es teilt einige Eigenschaften mit dem Konzept der Leichten Sprache. Doch während Leichte Sprache statisch ist, weil sie davon ausgeht, dass manche Menschen – etwa aufgrund einer geistigen Einschränkung – dauerhaft nur bis zu einem gewissen Sprachniveau kommen, setzt die dynamische Aufbauende Sprache auf Entwicklung. Sie gestaltet Sprache so, dass Kinder sie in jedem Entwicklungsstadium verstehen, anwenden – und gleichzeitig erweitern können. Das Sprachniveau ist der Schlüssel zum Kind.

Die pädagogische Fachkraft beginnt jede Lernsituation sprachlich „bei null“. Alle Kinder können mitmachen, und Schritt für Schritt wird sprachlich aufgebaut – durch Fragen, Wiederholungen, gemeinsame Aktionen. Dabei beobachtet und notiert die Fachkraft, wie weit das einzelne Kind mitgeht, welche Fortschritte sichtbar werden. Beobachtung und Förderung gehen also Hand in Hand. Manche Kinder werden dem Sprachangebot bzw. der Förderung nur bis zu einem gewissen Grad folgen können, andere bleiben bis zum Schluss dabei. Wichtig ist, dass Übungen, Spiele, Erzählungen etc. angeboten werden, bei denen zunächst alle mitmachen können, dann auf ihrem persönlichen Höchstniveau verweilen und dieses festigen können. Es geht darum, die Wohlfühlstufe jedes Kindes zu ermitteln – die Stufe, auf der es sich mit Leichtigkeit bewegen kann und Ressourcen frei hat, weitere Stufen zu erklimmen. So nutzt man die Zone der nächsten Entwicklung. Genau dieses Prinzip liegt dem Dynamic Assessment zugrunde, Testverfahren, bei denen nicht die isolierte Leistung, sondern die Lernfähigkeit unter Anleitung erfasst wird. Die Lehrkraft gibt während des Tests gezielte individuelle Hilfestellungen und korrekatives Feedback und notiert, was das Kind von alleine kann und was es mit wie viel Hilfe schafft und passt den Test dem Niveau des Kindes an. Test und Förderung sind keine Gegensätze mehr – sie folgen demselben pädagogischen Prinzip.

Setzen Test, Förderung oder das sprachliche Angebot dagegen auf einem zu hohen Niveau an, dann profitieren nur die Kinder davon, die dieses Niveau bereits beherrschen; bei anderen Kindern kommt nichts an, oder – was noch schlimmer ist – die Förderung verwirrt sie (Warum findet die Lehrerin bei Ronny das Rechnen gut, bei Luca nicht – soll Luca nicht rechnen?).

Viele diagnostische Verfahren gehen von einem hohen sprachlichen Ausgangsniveau aus. Doch die Realität vieler Vorschul- und Grundschulkinder – vor allem in sozial benachteiligten Stadtteilen – sieht anders aus. Tests, die zu komplexe Anforderungen stellen, zeigen dann nur, was Kinder nicht können, aber nicht, wo ihre Entwicklung ansetzen könnte. Das demotiviert sowohl Kinder als auch Fachkräfte.

Beispielsweise ist Laube, die Lernausgangslage Berlin für Kinder der ersten Klasse, ein quantifizierbarer Test und bietet inzwischen auch eine bequeme Eingabe in eine Computermaske, so dass man die Werte für das einzelne Kind im Vergleich zur Klasse oder dem landesweiten Durchschnitt erhält. Die Erhebung der Daten zur Validierung von Laube fand 2014 statt. Das ist 12 Jahre her. Seitdem ist das sprachliche Niveau der Schulanfänger laut Einschätzung vieler Fachkräfte deutlich gesunken. Laube geht davon aus, dass der Durchschnitt der Kinder 70 Prozent des erfragten Vokabulars beherrscht und – unabhängig von deren Bedeutung – bereits über die lautliche Gestalt von Wörtern nachdenken kann, also Wörter in Silben gliedern, Anlaute heraushören oder Reime finden kann. Kinder, die nicht die nötigen Voraussetzungen dafür haben, werden durch den Test nur verwirrt und wertvolle Zeit in einer 1:1 Situation bleibt für die Förderung ungenutzt.

Das Sprachförderzentrum zeigt mit Laube-Light eine Version, die wesentliche Elemente der Lernausgangslage in einem Dynamic-Assessment-Setting erfasst. Die Reihenfolge der Testteile sind so angeordnet, dass die Testleitung das Wohlfühlniveau des Kindes ermittelt und entscheiden kann, ob die Voraussetzungen für eine standardisierte Durchführung gegeben sind. Sind sie nicht gegeben, nutzt die Testleitung die Zeit für den Aufbau einer guten Lernbeziehung zum Kind. Wenn die Lehrkraft weiß, bis zu welchem Niveau die Kinder ihrer Klasse Äußerungen verstehen oder selbst bilden können (rezeptiv oder produktiv beherrschen), kann sie alle Angebote ihres Unterrichts daran ausrichten.

### **Fazit: Sprachdiagnostik sollte kein Selbstzweck sein.**

Die Durchführung standardisierter Tests, die den klassischen Gütekriterien entsprechen, hilft in der momentanen Situation der deutschen Bildungslandschaft wenig. Letztlich will man gar nicht ganz genau wissen, wie gut oder schlecht der Sprachstand eines Kindes ist, sondern man will, dass das Kind am Ende tatsächlich besser spricht – das Testergebnis ist nur ein Indikator dafür. Ein standardisierter Test macht kein Kind klüger – im Gegenteil, ein Test kann verwirren und demotivieren. Klug und sprachlich gewandt wird ein Kind, wenn es Gelegenheit bekommt, Sprache in sinnvollen, emotional bedeutsamen Situationen zu erleben, anzuwenden und zu erweitern – das ist das eigentliche Kriterium. Hat also ein Kind bereits während des Tests die Möglichkeit, seine Sprachkenntnisse zu erweitern, dann erfüllt der Test dieses Kriterium.

Wenn also Sprachbeobachtung und Sprachförderung eins werden, dann gilt plötzlich das Gegenteil des alten Sprichworts: Vom Testen wird das Kind klug – wenn der Test zugleich Förderung und Förderung zugleich ein Test ist.