

Diagnostik der Sprachkompetenz

Gl/sh Schülern Sprache beizubringen ist eine komplexe Angelegenheit, und ganz besonders für einen Klassenlehrer (im Gegensatz zu Experten im Bereich der Sprachtherapie). Wie wir bereits gesehen haben, sind die individuellen Unterschiede zwischen den Kindern sehr groß, und die Entwicklung der Sprachkompetenz ist viel schwieriger vorherzusagen als bei hörenden Kindern. Um den Unterricht auf die Stärken und Schwächen der gl/sh Schüler zuzuschneiden, müssen wir diagnostischen Sprachunterricht betreiben. Diagnostischer Unterricht ist eine Form des Unterrichts, in der es sowohl für den individuellen Schüler als auch für die gesamte Klasse ein ständiges Zusammenspiel zwischen der Beurteilung von Fähigkeiten und Fortschritten und der weiteren Unterrichtsplanung gibt. Aus dieser Sicht ist das Ziel der Diagnostik, Orientierungen und Richtlinien für die Unterrichtung gl/sh Schüler zu bekommen. Die Diagnostik kann auch anderen Zielen dienen, zum Beispiel kann sie zur Entscheidung für ein Bildungssetting oder zur Beurteilung eines Bildungsprogramms beitragen, aber unserer Ansicht nach ist die Priorität klar: Die Verwendung von diagnostischen Informationen für eine Beurteilung des Erfolgs von Bildungsmaßnahmen kann nur ein Nebenprodukt sein. Priorität hat die Diagnose der Schüler für praktische Entscheidungen in Bezug auf deren Förderung und Platzierung in für sie adäquate Bildungsmaßnahmen. Mitunter können dieselben diagnostischen Instrumentarien Informationen für alle drei Ziele liefern, aber angesichts der großen Unterschiede zwischen den gl/sh Schülern wird das nicht immer und nicht einmal oft der Fall sein. Diese Heterogenität stellt ganz im Gegenteil eine erhebliche Herausforderung für Lehrer, Psychologen, Sprachtherapeuten und Audiologen dar. Es müssen für jeden einzelnen Schüler Entscheidungen darüber getroffen werden, welches diagnostische Vorgehen den Zielen der Diagnostik am besten dient und welche Art von diagnostischen Instrumenten (Tests, Beobachtung etc.) eingesetzt werden sollten (*Jamieson & Simmons 2011*).

Für die Auswahl eines geeigneten Instruments zur Sprachüberprüfung stehen verschiedenen Arten von Verfahren zur Verfügung. Eine erste nützliche Unterscheidung ist die zwischen formeller und informeller Beurteilung. Bei formellen Verfahren sind die einzelnen Aufgaben und die Vorgehensweise bei der Durchführung ganz genau vorgeschrieben. Jede Abweichung von den standardisierten Vorgaben für die Testdurchführung kann die Interpretation der Daten ernsthaft beeinflussen und die gesamte Validität und Reliabilität der Testung in Frage stellen. Im Falle standardisierter Tests gibt es üblicherweise eine reichhaltige Forschungsbasis, durch die belegt ist, dass das Testkonstrukt das misst, was es zu messen verspricht (Validität), und dass diese Messungen verlässlich sind (Reliabilität) – dass also unabhängig davon, wer die Daten erhebt, ähnliche Ergebnisse erzielt werden, so lange die standardisierten Vorgaben für die Testdurchführung genau befolgt werden (vgl. Kap. 6 für ähnliche Probleme in Bezug auf Intelligenztests). Damit die Ergebnisse verglichen werden können, müssen solche Tests Normen haben, die auf einer großen Referenzgruppe basieren.

Eine informelle Einschätzung der Sprachkompetenz ist sehr viel stärker prozessorientiert als formelle Diagnostik. Auch wenn es hier standardisierte Vorgehenswei-

sen geben kann, umfasst informelle Diagnostik keine Tests im eigentlichen Sinne. Der Sprachgebrauch einer bestimmten Person wird mit informellen Verfahren im Kontext beurteilt (Jamieson & Simmons 2011). Typische Beispiele für informelle Einschätzungen sind die Verwendung von Beobachtungen, Checklisten, Elizitierungsstrategien, die Kinder zu bestimmten sprachlichen Äußerungen anregen sollen, und die »dynamische Beurteilung«, eine Kombination aus diagnostischer Beurteilung und individueller Förderung.

Eine weitere nützliche Unterscheidung bei der Diagnostik der Sprachkompetenz ist die zwischen lehrplanunabhängigen und lehrplangebundenen Beurteilungsverfahren. Mit letzteren möchte man herausfinden, wie viel ein Schüler mit einem bestimmten Lehrplan oder einer bestimmten Methode gelernt hat. Die relevanten Beurteilungen werden üblicherweise zu zwei oder mehreren Zeitpunkten vorgenommen, meist eine zu Beginn der Lehrplaneinheit oder Fördereinheit und eine weitere irgendwann zu einem späteren Zeitpunkt, vielleicht auch mehrere in regelmäßigen Abständen (zum Beispiel einmal monatlich). Mit diesem diagnostischen Vorgehen lässt sich der Fortschritt eines Schülers im Vergleich zu den Fortschritten messen, die andere Schüler (aus derselben Lerngruppe oder einer anderen Lerngruppe) im selben Unterricht gemacht haben. Lehrplanunabhängige Diagnoseinstrumente ermöglichen es dagegen, allgemeine Fähigkeiten oder Kompetenzen von Schülern zu beurteilen. Anders als bei lehrplangebundenen Diagnosemethoden lässt sich die Interpretation des Fortschritts allerdings nicht direkt mit der jeweiligen Unterrichtseinheit in Verbindung bringen.

Wenn eine Lehrkraft einen vollständigen Überblick über die sprachlichen Kompetenzen der Schüler in ihrer Klasse gewinnen möchte, müssen alle sprachlich relevanten Bereiche beurteilt werden, nicht nur der Wortschatz (der am häufigsten überprüft wird), sondern auch Phonologie, Morphologie, Syntax und Pragmatik, sowie Sprachverständnis und Sprachproduktion. Idealerweise umfasst eine solche Beurteilung sowohl formelle Tests als auch die Auswertung von sogenannten Spontansprachproben. Dafür wird die Kommunikation zwischen Diagnostiker und Kind aufgezeichnet, transkribiert und nach bestimmten Merkmalen analysiert, um ein Bild von der tatsächlich in Unterhaltungen verwendeten Sprache und ihrer Qualität zu erhalten. Bei formellen Tests erfolgt die Beurteilung eher unter laborähnlichen Bedingungen, was vielleicht einen breiter gefächerten Eindruck davon gibt, was und wie viel an Sprache der Schüler verstehen oder produzieren könnte, aber nur wenig Informationen über die Wahrscheinlichkeit und Fähigkeit des Schülers dazu liefert, dass sie im Alltag tatsächlich verwendet wird. Spontansprachproben können einen umfassenden Überblick über die Sprachproduktion liefern, bieten aber weit weniger Informationen zum Sprachverständnis.

Welche Art von diagnostischer Erhebung (formell/informell) und welche diagnostischen Instrumente bei einer Überprüfung der sprachlichen Entwicklung zum Einsatz kommen, sollte hauptsächlich vom Ziel der Beurteilung abhängig gemacht werden. Geht es darum, wie viel Sprache der einzelne Schüler versteht? Oder geht es darum, wie viel er oder sie an Sprache produzieren kann? Liegt der Fokus nur auf dem Wortschatz oder auch auf der Syntax? Soll der Fortschritt innerhalb einer Lern- oder Fördereinheit ermittelt werden, oder geht es um einen Überblick über die gene-

rellen sprachlichen Kompetenzen? Will man die Leistung eines Schülers mit den Leistungen einer Referenzgruppe vergleichen oder mit einem vorher festgelegten Kriterium? Und soll die Evaluation die Fähigkeiten unter »realen Alltagsbedingungen« beurteilen, oder eher allgemein, d. h. unabhängig von spezifischen Kommunikationssituationen? Wichtig ist auch eine Reihe praktischer Erwägungen. Beispielsweise kann der Zeitraum, der für eine diagnostische Beurteilung zur Verfügung steht, die Wahl des bevorzugten Untersuchungsinstruments bestimmen, einfach weil das Erheben von Spontansprachproben im Vergleich zu Sprachtests sehr viel mehr Zeit braucht.

Bis zu diesem Punkt haben wir die Frage der Diagnostik unabhängig vom Hörstatus erörtert. Sowie es um gl/sh Schüler geht, wird die Sache noch weit komplizierter. Die erste Frage ist, welche Sprache denn beurteilt werden soll. Für Lautsprachen wie Englisch, Niederländisch oder Deutsch sind etliche Tests verfügbar, die häufig auch gute psychometrische Eigenschaften aufweisen (vgl. zum Überblick für deutsche Sprachtests zum Beispiel *Kany & Schöler* 2007). Die meisten dieser Messinstrumente wurden allerdings speziell für gut hörende Menschen entwickelt. Eine Beurteilung der Gebärdensprachkompetenz ist schwieriger, weil wir relativ wenig Informationen zu verschiedenen Aspekten der Gebärdensprachentwicklung haben (jedenfalls im Vergleich zu Forschungsergebnissen zum Lautspracherwerb), weil es in Bezug auf die Gebärdensprachkompetenz von muttersprachlichen und nicht-muttersprachlichen Gebärdennutzern eine extrem große Bandbreite gibt, weil nur relativ wenige standardisierte Messinstrumente existieren und weil es nur wenig Messinstrumente mit guten psychometrischen Eigenschaften gibt, die sich als valide und reliabel erwiesen haben (vgl. zum Beispiel *Haug & Mann* 2008; *Singleton & Supalla* 2011).

In letzter Zeit wurden einige Tests für die Beurteilung der Sprachkompetenz bei Kindern entwickelt, die diverse Gebärdensprachen verwenden, zum Beispiel American Sign Language (ASL; *Maller, Singleton, Supalla & Wix* 1999; *Strong & Prinz* 1997), British Sign Language (BSL; *Herman, Holmes & Woll* 1999), Australian Sign Language (Auslan; *Schembri et al.* 2002; aktuell: <http://www.deafeducation.vic.edu.au/Research/Pages/Auslan-Assessment-Tool-now-published.aspx>) und Sign Language of the Netherlands (SLN; *Hermans, Knoors & Verhoeven* 2009). Einen Überblick über die meisten verfügbaren Messinstrumente findet man in *Singleton* und *Supalla* (2011) und auf der Website <http://www.signlang-assessment.info>. Im Vergleich zu den Tests, die zur Beurteilung von Lautsprachkompetenz entwickelt wurden, sind die Normierungen für diese Tests aber immer noch recht schwach, wenn denn überhaupt welche verfügbar sind. So sind es zum Beispiel 138 Gebärdensprachnutzer in den Normen für einen BSL-Test (*Herman et al.* 1999) und 300 muttersprachliche und nicht-muttersprachliche gehörlose Kinder in den Normen für einen SLN-Test (*Hermans et al.* 2009). Für den deutschen Sprachraum stehen derzeit zwei Verfahren zur Verfügung, um Aspekte der Gebärdensprachentwicklung von gl/sh Kindern zu überprüfen. Der Perlesko (Prüfverfahren zur Erfassung der lexikalisch-semantischen Kompetenz, *Bizer & Karl* 2002) ermöglicht objektive und vergleichbare Aussagen über die lexikalisch-semantische Kompetenz gehörloser Schüler im Grundschulalter (3.-5. Klasse). Es ist ein standardisiertes Instrument, das die passive Wortschatzkenntnis anhand einer repräsentativen Auswahl von Nomen, Verben und Adjektiven zu gleichen Anteilen und in Form von drei Sprachmodalitäten (Ge-

bärdensprache, Schriftsprache, Lautsprache) getestet. Es existieren Normen in Form von Prozentrangwerten und linearen Standardwerten. Als Normierungsgruppe wurden 112 gl/sh Kinder aus sieben deutschen Schulen für Gehörlose und Schwerhörige und einer Schule in der deutschsprachigen Schweiz herangezogen. *Haug* (2011) hat eine deutsche computerbasierte Adaption des englischen BSL RST (British Sign Language Receptive Skills Test) entwickelt und den daraus entstandenen DGS-Verständnistest mit 54 gehörlosen Kindern im Alter zwischen 3;9 und 10;10 Jahren auf inhaltliche Tauglichkeit, Reliabilität und Validität überprüft. Eine Normierung des Verfahrens, damit es in den Schulen Verwendung finden kann, ist derzeit in Kooperation mit zahlreichen Schulen und der Humboldt-Universität zu Berlin sowie der Ludwig-Maximilians-Universität München in Planung (es ist auch eine webbasierte Version geplant, vgl. www.signlanguagetest.com). *Haug* und *Hintermair* (2003) konnten in einer Umfrage an deutschen Bildungseinrichtungen für gl/sh Kinder, an der sich 203 (vorwiegend) Lehrkräfte aus fast der Hälfte aller Einrichtungen beteiligten, den Bedarf an Testverfahren zur Überprüfung der gebärdensprachlichen Fähigkeiten der Kinder bestätigen. 87 Prozent der Befragten gaben an, dass trotz des Einsatzes von Gebärden in der Unterrichtung von Gebärden bzw. Gebärdensprache keine regelmäßige Überprüfung der Entwicklung in diesem Bereich erfolgt so wie dies in anderen Bereichen (kognitive Entwicklung, visuelle Wahrnehmung, Teilleistungsstörungen) üblich ist.

Für die diagnostische Überprüfung der beruflichen Eignung mit jungen Erwachsenen hat sich in den letzten Jahren das Aachener Testverfahren zur Berufseignung von Gehörlosen (ATBG, *Grote, Louis-Nouvelné & Sieprath* 1997) etabliert und in der Praxis bewährt (<http://www.gebaerdensprache.de/1712.html>). Der Vorteil dieses computerbasierten Verfahrens besteht darin, dass es den gl/sh Probanden die Möglichkeit eröffnet, die Testanweisungen in der von ihnen präferierten Sprachmodalität (Lautsprache, LBG, DGS) abzurufen und so einen Beitrag zu einer erhöhten Reliabilität und Validität der Überprüfung leistet. Zudem steht eine mittlerweile große Vergleichsstichprobe mit gl/sh Testpersonen zur Verfügung.

Die Beurteilung der Lautsprachkompetenz bei gl/sh Kindern ist ein weiteres komplexes Thema. Viele Tests sind für die Beurteilung der Lautsprache entwickelt worden, die hörende Schüler verwenden. Da sich aber die Lautsprachkompetenz gl/sh Kinder erheblich von der ihrer hörenden Peers unterscheidet, stellt sich die Frage, wie angemessen es ist, diese Instrumente zur Beurteilung der Lautsprachkompetenz gl/sh Schüler einzusetzen. Messen wir hier dieselben Konstrukte? Welchen Wert hat es, Normen für Hörende für die Auswertung der sprachlichen Leistungen gl/sh Kinder heranzuziehen? Die Antworten auf diese Fragen hängen davon ab, welche spezifische Absicht die jeweilige diagnostische Beurteilung verfolgt. Soll beispielsweise die Lautsprachkompetenz gl/sh Kinder in Regelschulklassen beurteilt werden, könnte es angemessen sein, Normen für Hörende heranzuziehen, selbst wenn bei der Auswertung und weiteren Verwendung des Testergebnisses einige zusätzliche Aspekte zu berücksichtigen sind. Ein paar Tests zur Lautsprachkompetenz wurden speziell für die Beurteilung gl/sh Schüler entwickelt, wie zum Beispiel der Rhode Island Test of Language Structure (*Engen & Engen* 1984). Solche Tests könnten hilfreich sein, um gl/sh Schüler mit gl/sh Peers zu vergleichen, aber die geringe Anzahl gl/sh Kinder in der Bevölkerung hat zur Folge, dass die Normie-

rungsgruppen meist relativ klein sind (siehe oben). Solange eine Lehrkraft sich mehr für die Fähigkeiten eines individuellen Schülers interessiert als für eine Beurteilung im Vergleich zu einer Gruppe, ist das kein Problem. Wenn das Ziel der Beurteilung aber den Bezug zu einer Normierungsgruppe erforderlich macht, ist Vorsicht angezeigt.

Ein weiteres Problem ist, wie gut die Items eines Tests für gl/sh Schüler zugänglich sind. Wenn man einen gl/sh Schüler in Lautsprache testet, steht unweigerlich nicht nur die Lautsprachkompetenz sondern gleichzeitig auch die auditive und visuelle Sprachwahrnehmung auf dem Prüfstand. Ein bestimmtes Testergebnis spiegelt deshalb nicht unbedingt die Sprachkompetenz wider, sondern sollte als Kombination aus Sprachkompetenz und Sprachwahrnehmungsfähigkeiten verstanden werden. Leider gibt es keinen wirklichen Weg, um dieses Problem zu umgehen. Manche Lehrer oder andere diagnostisch tätige Fachleute passen die Testverfahren an, indem sie die Tests zum Beispiel in Schriftform, gesprochener und geschriebener Sprache oder in simultaner Kommunikation durchführen. Aber solche Anpassungen gefährden die Validität und Reliabilität eines Tests (Cawthon 2011; Qi & Mitchell 2012). Die Referenznormen sind dann wertlos, weil sie unter anderen Durchführungsbedingungen gewonnen wurden. Und zudem können solche Anpassungen das Problem womöglich gar nicht aus der Welt schaffen. Vermeidet man bei der Beurteilung der Lautsprachkompetenz die Anwendung gesprochener Sprache, weil damit auch Sprachwahrnehmung und Absehfähigkeit getestet würden, und setzt stattdessen die Schriftform ein, dann wird der Test zu einem Lesetest. Angesichts der Herausforderung, die das Lesen für viele gl/sh Schüler bedeutet (vgl. Kap. 8), schafft die Umgehung des einen Problems schlicht ein neues. Auf ähnliche Weise wird die Validität eines Lautsprachtests auch durch den Einsatz lautsprachunterstützender Gebärden und simultaner Kommunikation in mehrfacher Hinsicht gefährdet. Zum einen misst der Test dann nicht nur das Lautsprachverständnis sondern auch das Gebärdenverständnis. Zum anderen kann der Sprachtest aufgrund der Ikonizität mancher Gebärden (d. h. Gebärden, die so aussehen, wie das, was sie bezeichnen, wie zum Beispiel BASEBALL oder KAMERA in ASL, vgl. White & Tischler 1999) zu einer recht einfach zu lösenden Aufgabe werden, vor allem dann, wenn die Identifizierung von Bildern gefragt ist (wie zum Beispiel beim Peabody Picture Vocabulary Test). Es wird somit vielleicht gar nicht die Lautsprachkompetenz gemessen.

Manchmal werden Tests von Lehrern im Hinblick auf die Schwierigkeit der Items angepasst, weil sie davon ausgehen, dass verschiedene Items für gl/sh Schüler zu schwer seien. Entweder werden dann bestimmte Items verändert, oder Lehrer lassen gl/sh Schüler Tests absolvieren, die für jüngere Schüler bestimmt sind, was zum sogenannten »out-of-level testing« führt. »Out-of-level testing« ist in den USA weit verbreitet, gl/sh Kinder absolvieren dort standardisierte Leistungstests wie den Stanford Achievement Test und diverse schulische Tests der einzelnen Bundesstaaten. Wie gut gemeint dieser Ansatz auch immer scheinen mag, er bietet keine Garantien für eine bessere Genauigkeit oder Validität bei der Beurteilung der Fähigkeiten gl/sh Schüler. Oft fehlen empirische Belege für die Effektivität von Anpassungen, und man hat die Wirksamkeit verschiedener Tests und Testanpassungen in Frage gestellt (Cawthon 2011; Qi & Mitchell 2012).

Ein weiteres Problem bei Beurteilungen der Lautsprachkompetenz ist, dass gl/sh Schüler auch aus Familien stammen können, in denen eine andere Lautsprache gesprochen wird als die in der Schule verwendete Landessprache (*Jamieson & Simmons* 2011). Das kann in den USA Spanisch sein, Türkisch in Deutschland oder eine Berbersprache in den marokkanischen Familien in den Niederlanden. Um ein vollständiges Bild von der Lautsprachkompetenz eines individuellen Schülers zu erhalten, sollte auch die Kompetenz in der zu Hause gesprochenen Sprache beurteilt werden. Für manche dieser Sprachen gibt es vielleicht keine Kompetenztests oder nur Tests, die anhand anderer Gruppen normiert wurden, und es kann schwierig sein, eine Person zu finden, die die Sprache beherrscht und den Test durchführen kann. Oft spielen alle drei Faktoren eine Rolle. In manchen Ländern verbessert sich diese Situation leicht, weil spezifische Messinstrumente entwickelt oder bestimmte Verfahren und Best-Practice-Richtlinien geschaffen wurden, oder auch Dolmetscher bei der Testdurchführung eingesetzt werden (*American Speech and Hearing Association* 2003; *Blumenthal* 2009).

Aus all diesen bislang berichteten Problemen sollte man aber nicht schließen, dass die Diagnostik der Sprachkompetenz gl/sh Schüler so schwierig sei, dass sie den Aufwand nicht lohne oder Fachleute sich dadurch entmutigen lassen sollten. Die Sprachdiagnostik ist im Gegenteil ein außerordentlich wichtiges Element im Sprachunterricht für gl/sh Schüler. Man muss aber bei der Durchführung von Sprachdiagnostik die Ziele der Beurteilung sorgfältig definieren und die besten verfügbaren Messinstrumente auswählen und sollte Verfahren und Items nicht leichtfertig verändern. Wenn Testitems oder -verfahren angepasst werden, müssen die Veränderungen so genau wie möglich dokumentiert werden, damit eine Wiederholung unter vergleichbaren Bedingungen möglich ist. Angesichts der komplexen Sprachsituation, in der sich viele gl/sh Schüler befinden, wäre es vielleicht besonders geschickt, ihre Sprachentwicklung auf Grundlage mehrerer unterschiedlicher Beurteilungsverfahren darzustellen und dabei ihr unterschiedliches (multimodales und multilinguales) Repertoire zu dokumentieren und ihre unterschiedlichen Ressourcen zu berücksichtigen. Und schließlich muss man bei der Interpretation der Testergebnisse bei gl/sh Schülern noch vorsichtiger sein als bei hörenden Schülern. Wenn man all die benannten potenziellen Herausforderungen der Testdurchführung und -interpretation mit gl/sh Kindern zur Kenntnis nimmt und angemessen berücksichtigt, kann die Diagnostik sprachlicher Kompetenzen so vorgenommen werden, dass sie größtmögliche Validität und Reliabilität gewährleistet.