

Kapitel 2: Werkzeuge zur Umsetzung von Konzepten: Woran erkennt man gute Tools?

Sabrina Geyer, Cora Titz, Susanne Weber, Anna Ropeter & Marcus Hasselhorn

In Konzepten zur sprachlichen Bildung und Förderung können Fördermaßnahmen und die für ihre Optimierung erforderliche Sprachdiagnostik mit Hilfe verschiedener Werkzeuge (Tools) umgesetzt werden. Im vorliegenden Kapitel werden Qualitätskriterien guter Diagnostik- und Förder-Tools definiert und erläutert. Zunächst wird mit der *Passung* von Tool, Ziel und Zielgruppe ein wesentliches Prinzip bei der Auswahl von Diagnostik- und Förder-Tools beschrieben. Getrennt für die beiden Bereiche Diagnostik und Förderung werden anschließend zentrale Qualitätskriterien beschrieben, die Fachkräfte bei der Bewertung und Auswahl von Tools unterstützen können. Diagnostik-Tools dienen im Rahmen sprachlicher Bildung der Feststellung und Förderung des aktuellen Entwicklungsstandes von Kindern und Jugendlichen sowie ihres sprachlichen Entwicklungsverlaufs. Die Qualität dieser Tools ist umso höher, je geringer die Risiken für eine Fehleinschätzung des Entwicklungsstandes von Kindern und Jugendlichen sind. Es sollten daher Diagnostik-Tools verwendet werden, die objektiv, reliabel (zuverlässig) und valide (inhaltlich gültig) sind und eine Normierung aufweisen. Darüber hinaus sind Kriterien aus Praxis-Perspektive zu berücksichtigen. Bei der Auswahl von Tools zur Sprach- und Leseförderung ist neben Kriterien aus Perspektive der Praxis auch zu prüfen, inwiefern ein Tool über eine theoretische Fundierung verfügt sowie ob und welche Nachweise zu seiner Wirksamkeit vorliegen.

Einleitung

Seit dem »Pisa-Schock« zu Beginn unseres Jahrhunderts, ausgelöst durch die Offenlegung ungenügender Lesekompetenzen bei rund einem Viertel aller 15-jährigen Schülerinnen und Schüler in Deutschland, wurde eine Vielzahl von Maßnahmen zur Sprach- und Leseförderung initiiert. Die Erkenntnis reifte, dass sprachliche und schriftsprachrelevante Kompetenzen von Kindern bereits vor der Einschulung gefördert werden sollten, um ihre Chance auf eine erfolgreiche Teilhabe am Bildungssystem zu erhöhen. Entsprechend ist mit der Ausweitung der Bemühungen um eine effektive Sprach- und Leseförderung in den letzten Jahren auch die Anzahl

der zur Verfügung stehenden Werkzeuge (im Folgenden Tools) angestiegen, die zur Diagnose und Förderung von Sprach- und Lesekompetenzen genutzt werden können. Auch in den Verbänden¹ der Initiative »Bildung durch Sprache und Schrift« (BiSS) wurde eine Vielzahl solcher Tools eingesetzt, um die jeweiligen Ziele der sprachlichen Bildung und Förderung zu erreichen. Dazu gehören Materialien zur Diagnostik sprachlicher Kompetenzen und zu ihrer Förderung, die die pädagogischen Fachkräfte in ihren Handlungskompetenzen unterstützen sollen. Zu den Diagnostik-Tools zählen Beobachtungs-, Befragungs- oder Testverfahren, die zur Erfassung des Entwicklungsstandes bzw. zur Evaluation von zuvor durchgeführten Fördermaßnahmen eingesetzt werden. Förder-Tools sind u. a. spezifisch gestaltete Förderprogramme zu bestimmten Zielbereichen (z. B. phonologische Bewusstheit, Leseflüssigkeit) sowie Methoden und Techniken, die als Bausteine eines Konzepts genutzt werden können (z. B. dialogisches Lesen, Scaffolding, Sprachlehrstrategien wie Modellieren).

Mittlerweile befindet sich eine kaum mehr überschaubare Vielzahl an Tools auf dem Markt. In verschiedenen Expertisen wurde daher insbesondere für den Elementarbereich versucht, Tools zur Diagnose und Förderung zu sichten und die Qualität von Werkzeugen zur Sprachdiagnostik und Sprachförderung einzuschätzen (zur Diagnostik z. B. Dietz & Lisker, 2010; Neugebauer & Becker-Mrotzek, 2013; zur Förderung z. B. Lisker, 2011; Jampert, Best, Guadatiello, Holler & Zehnauer, 2007). Dabei zeigte sich u. a., dass für etliche Verfahren gar nicht bekannt ist, inwiefern sie die für professionelle Verfahren festgelegten Qualitätskriterien erfüllen. Das hat zur Folge, dass es für pädagogische Fachkräfte in der Praxis oftmals kaum möglich ist, gute von mangelhaften Tools zu unterscheiden. Um eine Orientierungshilfe zu geben, werden in diesem Kapitel sowohl für Diagnostik- als auch für Förder-Tools Qualitätskriterien formuliert, deren Berücksichtigung aus wissenschaftlicher Sicht bei der Bewertung und Auswahl geeigneter Tools empfohlen wird.

1 Wann ist ein Tool ein passendes Tool?

Die Frage, ob ein Tool geeignet ist, lässt sich nicht pauschal beantworten. Zentral für die Auswahl eines Tools zur Diagnostik bzw. Förderung ist zunächst die Frage danach, wofür es genau eingesetzt werden soll. Kurzum: Das Tool muss zu den Zielen und der Zielgruppe, für die es eingesetzt werden soll, passen.

¹ In der Initiative BiSS haben sich verschiedene Bildungseinrichtungen zu sogenannten Verbänden zusammengeschlossen, die gemeinsam Konzepte zur Sprach- und Lesediagnostik und -förderung entwickeln und umsetzen. Ein Verbund besteht i. d. R. aus drei bis zehn Kindertageseinrichtungen oder Schulen sowie weiteren Partnern wie beispielsweise Universitäten oder Bibliotheken.

In Bezug auf das Ziel ist daher relevant, welche sprachlichen Kompetenzen zu welchem Zweck erfasst bzw. gefördert werden sollen. Diagnostik hat keinen Selbstzweck in sich, sondern dient immer einem bestimmten Ziel. Dieses Ziel kann darin liegen, Veränderungen im Entwicklungsverlauf durch eine durchgängige Dokumentation zu erfassen. Es kann aber auch in der Feststellung und Identifikation individueller Förderbedarfe in bestimmten sprachlichen Bereichen bestehen, die noch nicht ausreichend entwickelt sind – etwa im Vergleich zur Gruppe der Gleichaltrigen. In diesem Fall sollten aus dem Ergebnis einer Sprachdiagnostik auch konkrete Förderentscheidungen abgeleitet werden können, die sich auf die spezifischen Zielbereiche beziehen. Die Komplexität des Erwerbs sprachlicher sowie schriftsprachlicher Kompetenzen setzt eine differenzierte Beschäftigung mit dem Erwerbsgegenstand voraus. So umfasst die menschliche Sprache verschiedene sprachliche Ebenen, die in der Kommunikation zusammenwirken, jedoch jeweils eigenen Regeln folgen. In der Sprachwissenschaft werden mindestens die fünf Ebenen der Phonologie, Morphologie, Syntax, Semantik sowie Pragmatik unterschieden (für eine ausführliche Beschreibung siehe Müller, Schulz & Tracy, 2018). Die Zielbereiche für den Einsatz eines Tools zur Diagnose bzw. Förderung mündlicher Sprachkompetenzen können sich auf all diese Ebenen beziehen. Dabei sollten jeweils sowohl rezeptive als auch produktive Kompetenzen berücksichtigt werden, da diese im Spracherwerb nicht synchron erworben werden. Zur Diagnose und Förderung schriftsprachlicher Kompetenzen sind typische Zielbereiche zum Beispiel die (frühe) Literalität, die Leseflüssigkeit oder das Leseverständnis. Darüber hinaus sind weitere Zielbereiche denkbar, wie z. B. die Diagnose und Förderung im Bereich der Orthographie oder Textproduktion. Innerhalb der Zielbereiche lassen sich weitere Spezifizierungen vornehmen, z. B. im Bereich der Grammatik auf die Verbstellung in Haupt- und Nebensätzen oder im Bereich des Leseverstehens auf die Nutzung ganz bestimmter Lesestrategien.

Auch zur Überprüfung der Passung eines Tools zur Zielgruppe sind verschiedene Aspekte zu beachten. So muss bei der Auswahl von Diagnostik-Tools zunächst geklärt werden, ob ein vom Zielbereich her passendes Tool für den Alters- und Entwicklungsbereich der konkreten Zielgruppe geeignet ist. Innerhalb des gewählten Altersbereichs ist ferner relevant, ob das Diagnostik-Tool bei allen Kindern der Altersgruppe zur Bestimmung ihrer sprachlichen Ausgangslagen herangezogen werden soll, oder ob es bei einer vorausgewählten Subgruppe (z. B. Kinder mit Deutsch als Zweitsprache) oder zur Identifikation einer Subgruppe (z. B. lesechwache Schülerinnen und Schüler) zur näheren Bestimmung des Förderbedarfs dient. Von dieser Entscheidung wird auch die Auswahl einer Methode zur Sprachdiagnostik geprägt. Im ersten Fall wäre etwa ein diagnostisches Instrument geeignet, das Fachkräfte bei der Beobachtung und Dokumentation der Sprachentwicklung aller Kinder unterstützt, wie beispielsweise ein standardisierter Beobachtungsbogen. Im zweiten Fall dagegen sollte ein Testverfahren gewählt werden, mit dessen Hilfe genau ermittelt werden kann, bei wem und in welchem Zielbereich welcher Förderbedarf gegeben ist. Um Fehleinschätzungen zu vermeiden, ist beispielsweise bei Kindern oder Jugendlichen mit Deutsch als Zweitsprache bei der Diagnostik auch deren spezifische Spracherwerbssituation, d. h. das Alter bei Erwerbsbeginn der deutschen Sprache sowie die Kontaktzeit zum Deutschen,

zu berücksichtigen (vgl. Grimm & Schulz, 2016; Müller et al., 2018). Bei der Auswahl eines Förder-Tools muss die Passung zum zugrunde liegenden Diagnostik-Tool geprüft werden. Nur wenn im Förder-Tool genau die Zielbereiche berücksichtigt werden, die im Diagnostik-Tool erfasst wurden, können identifizierte Förderbedarfe in der Förderung auch aufgegriffen werden. Daneben sollte die Passung zur Zielgruppe bei der Förderung berücksichtigt werden. Auch bei den Förder-Tools ist ergänzend zum Alter bzw. dem Sprachstand der Zielgruppe relevant, ob sich das Tool an alle Kinder und Jugendlichen einer Altersgruppe richtet oder ob es für eine bestimmte Subgruppe entwickelt wurde.

Neben der Passung des Tools zu Zielen und Zielgruppen der Diagnostik und Förderung sind aus wissenschaftlicher Sicht jeweils für Diagnostik- und Förder-Tools spezifische Qualitätskriterien heranzuziehen, die im Folgenden erläutert werden. Deren Kenntnis kann Fachkräfte bei der evidenzbasierten Auswahl eines geeigneten Tools unterstützen.

2 Welche Kriterien kennzeichnen gute Diagnostik-Tools?

Detaillierte Kenntnisse über den Sprachentwicklungsstand eines Kindes bzw. Jugendlichen gelten als Voraussetzung für die individuelle Förderung sprachlicher und schriftsprachlicher Kompetenzen und somit als wichtiger Ausgangspunkt für das pädagogische Handeln von Fachkräften (vgl. Titz, Ropeter & Hasselhorn, 2018). Der Erfolg von Fördermaßnahmen hängt also maßgeblich von einer gelingenden Diagnostik ab, für die je nach Ziel und Zielgruppe ein passendes sowie qualitativ hochwertiges Instrument gewählt werden sollte.

Bei der Auswahl eines Tools zur Sprachdiagnostik ist es für Anwenderinnen und Anwender daher zentral zu wissen, wie sehr sie sich auf die mit dem Verfahren erhaltenen Ergebnisse verlassen können. Bei jedem Versuch, die individuelle Ausprägung von Kompetenzen zu erfassen, gibt es Fehlerquellen. Eine zentrale Forderung zur Minimierung der damit verbundenen Gefahr von Fehleinschätzungen ist deshalb, dass bei der Auswahl von Diagnostik-Tools auf die Einhaltung wissenschaftlich definierter Qualitätskriterien geachtet wird. Im Folgenden werden daher zunächst diese Qualitätskriterien (auch Gütekriterien genannt) vorgestellt, bevor auf (zusätzliche) Bedarfe aus Praxis-Perspektive eingegangen wird.

2.1 Wissenschaftliche Gütekriterien

Die Einhaltung wissenschaftlicher Gütekriterien soll gewährleisten, dass ein Diagnoseverfahren individuelle Kompetenzen, Fertigkeiten oder Merkmale möglichst eindeutig und genau erfasst. Handbücher bzw. Manuale von Diagnostik-Tools sollten deshalb Informationen zur Überprüfung verschiedener Qualitätsbereiche

beinhalten. Die Qualitätsbereiche oder Gütekriterien, die zwingend erläutert sein sollten, sind die *Objektivität*, die *Reliabilität* (Genauigkeit) und die *Validität* (inhaltliche Gültigkeit). Daneben sollten die im Individualfall erhaltenen Ergebnisse eindeutig interpretierbar sein, was über eine *Normierung* erreicht werden kann.

Objektivität

Ein Testverfahren muss objektiv sein. Das bedeutet, dass die Durchführung und Auswertung des Verfahrens sowie die Interpretation der Ergebnisse eindeutig sind. Man unterscheidet verschiedene Facetten der Objektivität, insbesondere die *Durchführungsobjektivität*, *Auswertungsobjektivität* und *Interpretationsobjektivität*.

Eine hohe *Durchführungsobjektivität* bedeutet, dass eine Testleiterin bzw. ein Testleiter das Testverfahren immer wieder auf die gleiche Weise durchführt bzw. dass unterschiedliche Testleiterinnen bzw. Testleiter sich in ihrer Durchführung nicht voneinander unterscheiden. Grundlage hierfür ist ein Manual, das möglichst keinen Interpretationsspielraum bei der Durchführung erlaubt. Maximal standardisiert sind Verfahren, bei denen die Durchführung von einem Computer-Algorithmus gesteuert wird.

Die *Auswertungsobjektivität* ist gewährleistet, wenn das Testergebnis unabhängig von der Auswerterin bzw. dem Auswerter ist – jede Person also bei einem individuellen Ergebnisprotokoll zum gleichen Ergebnis kommt. Bei Testverfahren mit vorgegebenen Antwortmöglichkeiten ist es einfach, eine hohe Auswertungsobjektivität zu erreichen. Schwieriger stellt es sich bei Verfahren dar, bei denen die getesteten Kinder bzw. Jugendlichen freie Antworten geben dürfen. Hier muss exakt definiert werden, welche Antwort als richtig bzw. falsch zu bewerten ist (Bühner, 2011).

Bei der *Interpretationsobjektivität* geht es darum, inwiefern ein Testergebnis eindeutig zu interpretieren ist – ebenfalls unabhängig von der Person, die es interpretiert. Am besten zu erreichen ist Interpretationsobjektivität, wenn es sich bei den Testverfahren um normierte Leistungstests oder Fragebögen handelt. In diesen wird nach der Auswertung ein numerischer Wert ermittelt, der in einer Tabelle bzw. Skala eine eindeutige Position der Probandin oder des Probanden darstellt, die sie oder ihn bezüglich ihrer oder seiner Leistung und damit der zu vermutenden zugrunde liegenden Kompetenz eindeutig einordnet (Lienert & Raatz, 1998).

Reliabilität

Unter Reliabilität ist die Zuverlässigkeit eines Testverfahrens zu verstehen. Je reliabler ein Test ist, desto unabhängiger ist das erzielte Testergebnis von situativen Bedingungen, die nicht unmittelbar mit der gemessenen Kompetenz des Kindes bzw. des Jugendlichen zu tun haben, wie zum Beispiel der Tageszeit oder der Tagesform. Unterschiedliche Ergebnisse von Testpersonen lassen bei einem hoch-

reliablen Test also auf tatsächliche Unterschiede des zu erfassenden Merkmals schließen. Die Reliabilität lässt sich mit verschiedenen Methoden empirisch ermitteln.

Die *Retest-Reliabilität* wird bestimmt, indem die getesteten Personen einer Stichprobe zu zwei unterschiedlichen Zeitpunkten mit demselben Instrument getestet werden. Geprüft wird dann die Stärke des statistischen Zusammenhangs in Form einer sogenannten Korrelation zwischen den Ergebnissen der beiden Zeitpunkte. Die Werte, die bei einer Korrelation errechnet werden, können zwischen -1 und $+1$ liegen. Ein Wert um 0 herum zeigt an, dass kein statistischer Zusammenhang zwischen den Verfahren vorliegt. Ein Wert, der nahe an $+1$ oder -1 heranreicht, deutet auf einen starken linearen positiven oder negativen Zusammenhang hin.

Ebenfalls durch eine Korrelation bestimmt wird die Reliabilität, wenn man sie mit Hilfe der *Paralleltest-Methode* überprüft. Hier werden die Ergebnisse zweier ähnlich konstruierter Tests, die mit unterschiedlichen Items (Fragen oder Testaufgaben) dieselbe Eigenschaft oder Fähigkeit messen sollen, miteinander korreliert. Eine hohe Korrelation bedeutet eine hohe Übereinstimmung der Ergebnisse und somit eine hohe Paralleltest-Reliabilität.

Eine dritte Methode zur Bestimmung der Reliabilität stellt die Ermittlung der *internen Konsistenz* dar. Bei dieser Methode der Reliabilitätsbestimmung wird geprüft, ob der Test in sich konsistent ist – ob also alle Teilaufgaben geeignet sind, das gleiche Merkmal zu erfassen. Um die interne Konsistenz zu berechnen, kann die Konsistenzanalyse angewendet werden (Lienert & Raatz, 1998). Es wird hierbei jede einzelne Aufgabe des Tests mit dem Gesamtergebnis in Beziehung gesetzt. Wenn die Ergebnisse bei einzelnen Aufgaben mit dem Gesamtergebnis übereinstimmen, ist die Homogenität der Aufgaben und somit die interne Konsistenz hoch. Das Ergebnis dieser Berechnung ist ein sogenannter Konsistenzkoeffizient. Der bekannteste Konsistenzkoeffizient ist Cronbachs-Alpha (Cronbachs- α). Lienert und Raatz (1998) ordnen der internen Konsistenz auch die sogenannte *Testhalbierungs-* oder *Split-Half-Methode* zu. Bei ausreichender Testlänge (d. h. bei einer ausreichenden Anzahl von Aufgaben) kann sie berechnet werden, indem die Aufgaben des Diagnostik-Tools zunächst vollständig bearbeitet werden. Sie werden dann in zwei Hälften geteilt und das Testergebnis für beide Testhälften einzeln ermittelt. Die Ergebnisse der beiden Testhälften werden dann miteinander korreliert. Der ermittelte Wert lässt sich anschließend mit Hilfe einer einfachen Formel (der Spearman-Brown-Formel) korrigieren, so dass die Reliabilität für einen Test der ursprünglichen Länge berechnet werden kann. Ausführliche Informationen zur Berechnung finden sich bei Fisseni (1997). Reliabilitätswerte unter $.70$ gelten weithin als nicht ausreichend für Zwecke der Individualdiagnose.

Validität

Unter Validität versteht man die inhaltliche Gültigkeit eines Diagnostik-Tools. Validitätsmaße geben an, in welchem Ausmaß der jeweilige Test tatsächlich die Kompetenz bzw. das Merkmal misst, die bzw. das er zu messen vorgibt. Ein Test zur

Erfassung der Leseflüssigkeit z. B. ist dann valide, wenn die Testergebnisse ausschließlich durch die Leseflüssigkeit eines Kindes und nicht zusätzlich in nennenswertem Ausmaß durch seine visuellen Fähigkeiten bestimmt werden. Es wäre sonst schwierig zu bestimmen, auf welchen der beiden Bereiche (Leseflüssigkeit oder visuelle Fähigkeiten) die im Test gezeigte Leistung zurückzuführen ist. Bei einem validen Test lässt das Testergebnis also genaue Rückschlüsse auf das zu erfassende Merkmal zu (Lienert & Raatz, 1998). Auch bei der Validität lassen sich unterschiedliche Aspekte voneinander abgrenzen.

Die *theoretische Validität* ist gegeben, wenn jede einzelne Aufgabe eines Diagnostik-Verfahrens bzw. die Aufgaben im Gesamten aus einer zugrunde gelegten Theorie abgeleitet werden können. Im thematischen Fokus dieses Bandes könnten dies etwa Theorien zum (Zweit-)Spracherwerb oder zum Lesererwerb sein. Eine erwerbstheoretische Fundierung ist für eine zuverlässige Diagnostik deshalb von so großer Bedeutung, weil jeder erfasste Entwicklungsstand bei einem Kind nur vor dem Hintergrund einer entsprechenden Entwicklungstheorie interpretiert werden kann. Einzelne Aufgaben im Test sollten daher das zugrunde liegende Erwerbsmodell nachvollziehbar abbilden (Kany & Schöler, 2007). Auch die Berücksichtigung aktueller Forschungsergebnisse ist für die theoretische Validität zentral, wenn zum Beispiel empirisch belegt werden kann, dass das erfasste Merkmal für die zu erfassende Kompetenz relevant ist. So gilt z. B. die phonologische Bewusstheit, also die Einsicht in die Lautstruktur einer Sprache, in der Forschung als zentraler Prädiktor für die spätere angemessene Aneignung des Lesens und Rechtschreibens. Ihre diagnostische Erfassung wurde daher als wichtiger Schritt für die Prävention von Lese-Rechtschreibschwierigkeiten beschrieben (Schneider & Marx, 2008). Von professionellen Tools darf man daher erwarten, dass ihre Manuale nicht nur die Ziele, Zielbereiche und Zielgruppe beschreiben, für die das Tool geeignet ist, sondern dass sie auch die theoretischen Grundlagen des Tools explizit erläutern sowie die Relevanz einzelner Testaufgaben für die zugrunde gelegte Theorie darlegen.

Fehlt eine zugrunde liegende Theorie, tritt bisweilen die *Augenscheinvalidität* an die Stelle der theoretischen Validität: Hier vertraut man darauf, dass es offenkundig ist, dass die gewählten Testaufgaben das erfassen, was sie erfassen sollen. Die theoretische Validität wird nicht durch einen numerischen Wert gekennzeichnet und wird deshalb in der Regel von unabhängigen Expertinnen und Experten eingeschätzt.

Es gibt auch die Möglichkeit, die *empirische Validität* zu ermitteln, d. h. auf empirischem Wege zu prüfen, ob ein Diagnose-Verfahren tatsächlich den Inhaltsbereich abdeckt, den es abdecken soll. In den Bereich der empirischen Validität fällt die *Konstruktvalidität*, die wiederum in die *konvergente* und *diskriminante Validität* unterteilt werden kann. Die *konvergente Validität* wird ermittelt, indem die Testergebnisse des betreffenden Tools mit denen anderer Tests verglichen werden, die den Anspruch haben, den gleichen sprachlichen Kompetenzbereich zu erfassen. Bei der *diskriminanten Validität* wird der statistische Zusammenhang zu den Ergebnissen eines Testverfahrens ermittelt, das ausdrücklich *nicht* das gleiche Merkmal erfasst, sondern einen inhaltlich anderen Bereich (z. B. Korrelation zwischen einem Leseflüchtigkeits- und einem Aufmerksamkeitstest). Erwartet wird in

diesem Fall ein möglichst niedriger statistischer Zusammenhang, da die miteinander verglichenen Verfahren Unterschiedliches messen sollten.

Die *Kriteriumsvalidität*, die ebenfalls der empirischen Validität zugeordnet wird, gibt die Beziehung zwischen dem mit einem Diagnostik-Tool ermittelten Ergebnis und einem Außenkriterium an, von dem man annimmt, dass es mit dem zu messenden Merkmal in Zusammenhang steht. So wird bei der Bestimmung der Kriteriumsvalidität von Sprachtests als Außenkriterium beispielsweise oftmals die Einschätzung der sprachlichen Kompetenz der getesteten Kinder durch Sprachförderexpertinnen und -experten herangezogen. Je höher der statistische Zusammenhang zwischen Testergebnis und einem Außenkriterium ist, desto valider ist das Instrument für dieses Kriterium.

Normierung

Um das Testergebnis eines Kindes bzw. Jugendlichen adäquat interpretieren zu können, ist es entscheidend, seine Leistung mit der einer geeigneten Vergleichsgruppe in Beziehung zu setzen. Aus diesem Grund werden bei Diagnostik-Tools in der Regel Normwerte angegeben, die an repräsentativen Stichproben der Zielgruppe(n) des Diagnostik-Tools ermittelt wurden. Erst ein Vergleich mit diesen ermöglicht es, Aussagen über die relative Position eines Kindes bzw. Jugendlichen innerhalb der interessierenden Zielgruppe zu treffen, z. B. wie angemessen der aktive Wortschatz eines (monolingual deutschsprachigen) Kindes im Alter von drei Jahren im Vergleich zu anderen Kindern seiner (monolingual deutschsprachigen) Altersgruppe ausfällt. Das absolute Ergebnis eines solchen Tests in Form eines Rohwertes lässt eine Einschätzung über die Angemessenheit einer Leistung nicht zu, sondern beschreibt nur, über wie viele der vom Test geprüften Wörter der Wortschatz des Kindes umfasst.

Für die Normierung eines Diagnostik-Tools muss eine Stichprobe ausgewählt werden, die hinsichtlich verschiedener Merkmale repräsentativ für die Zielgruppe(n) des Testverfahrens ist. Diese Merkmale können beispielsweise das Alter, das Geschlecht, die Schultypen oder den soziokulturellen Hintergrund umfassen. Bei der Diagnostik sprachlicher Kompetenzen sollten zudem Merkmale zum Spracherwerbshintergrund des Kindes bzw. Jugendlichen, das Alter bei Erwerbsbeginn mit dem Deutschen sowie die jeweilige Kontaktdauer zum Deutschen berücksichtigt werden (vgl. Grimm & Schulz, 2016). Die sprachlichen Leistungen eines Kindes bzw. Jugendlichen in einem Test sollten daher immer mit einer Normierungsgruppe verglichen werden, die dem getesteten Kind bzw. Jugendlichen hinsichtlich dieser Merkmale ähnelt.

Amelang und Zielinski (2002) weisen darauf hin, dass Normdaten bisweilen schnell veralten und daher möglichst aktuell sein sollten. Es gibt genügend Beispiele dafür, dass sich die Mittelwerte und Verteilungen von Testleistungen bei der gleichen Zielgruppe über die Zeit systematisch verändern können. So hat man in den letzten Jahrzehnten etwa im Bereich der Rechtschreibleistungen systematische Leistungsverlechterungen im Grundschul- und Sekundarstufenalter festgestellt. Veränderungen dieser Art unterstreichen die Notwendigkeit, möglichst aktuelle Normdaten heranzuziehen und darauf zu achten, immer die aktuellste Version

eines Instruments zu verwenden. Die Herausgeberinnen und Herausgeber von Schultests in Deutschland plädieren also aus gutem Grunde dafür, die Normen von schulrelevanten Leistungstests alle zehn Jahre neu zu überprüfen.

2.2 Praxis-Perspektive

Erfüllt ein Diagnostik-Tool die beschriebenen wissenschaftlichen Gütekriterien, so ist damit zwar eine notwendige Bedingung dafür gegeben, dass es ein gutes Tool ist. Ob es dann jedoch im Alltag auch tatsächlich eingesetzt wird, hängt von weiteren Kriterien ab, insbesondere von seiner Durchführbarkeit und seiner Akzeptanz in der Praxis.

Durchführbarkeit

Die Entscheidung für oder gegen ein diagnostisches Verfahren hängt auch davon ab, wie gut sich das Tool in die Abläufe in der pädagogischen Praxis in Kita und Schule integrieren lässt. Da diagnostische Tätigkeiten in der pädagogischen Praxis oftmals als nicht zentral eingestuft werden, spielt der Zeitfaktor für die Akzeptanz eines Diagnostik-Tools eine entscheidende Rolle, d. h. Tools, für deren Durchführung und Auswertung weniger Zeit benötigt wird, werden in der Regel im Praxisfeld eher akzeptiert. Gleichzeitig ist wichtig, ob ein Tool nur als Einzelverfahren oder als Gruppenverfahren einsetzbar ist, weil Gruppenverfahren in ihrer Umsetzung weniger zeitliche Ressourcen beanspruchen. Ein Test, der zu viele zeitliche und personelle Ressourcen bindet, wird es schwer haben, in der pädagogischen Praxis dauerhafte Anwendung zu finden.

Solche Ökonomie-Anforderungen aus der Praxis stellen für Testentwicklerinnen und -entwickler eine große Herausforderung dar, da Testverkürzungen mit Einbußen in den Qualitätsmerkmalen einhergehen, die wir mit den wissenschaftlichen Gütekriterien (► Abschnitt 2.1) beschrieben haben. So führt die Kürzung eines Tests beispielsweise fast immer zu Einbußen in der Messgenauigkeit (Reliabilität).

Ein weiterer wichtiger Punkt ist, ob die Durchführung des Verfahrens für die Anwenderin oder den Anwender einfach nachzuvollziehen ist. Dazu gehören klar und verständlich formulierte Anweisungen in Handbüchern bzw. Manualen, die für jeden anspruchsvollen Test vorliegen. Je konkreter ausformuliert ist, wie die Durchführung zu erfolgen hat, desto eher wird sich eine Anwenderin oder ein Anwender zutrauen, einen Test selber durchzuführen. Hierfür sind detaillierte Beschreibungen der Aufgabenabläufe hilfreich (Wie soll die Aufgabe dem Kind erklärt werden? In welcher Reihenfolge werden die Aufgaben vorgegeben? Wie werden die Antworten kodiert? etc.). Zudem sollte genau angegeben sein, welche Materialien für die Durchführung der Diagnostik benötigt werden. Ein weiterer positiver Effekt einer konkreten Durchführungsbeschreibung ist, dass Fehler in der Durchführung reduziert werden, die zu nicht validen Ergebnissen führen können (► Abschnitt 2.1). Auch eine einfach und nachvollziehbar beschriebene Auswertung und Interpretation der Ergebnisse vereinfacht die Durchführbarkeit und erhöht deutlich die Wahrscheinlichkeit, dass ein Test im Alltag angewendet wird.

Akzeptanz

Für die Akzeptanz eines Tests in der Praxis ist es wichtig, dass den Anwenderinnen und Anwendern Ziel und Nutzen des Tests verständlich und überzeugend vermittelt werden. Das bedeutet, dass eindeutig formuliert werden muss, für wen der Test gedacht ist (Zielgruppe) und was genau mit dem Test erfasst werden soll (Zielbereich). Zudem sollte deutlich werden, was den Einsatz des Tests in der Praxis rechtfertigt, d. h. welcher zusätzliche Nutzen mit dem Test einhergeht. Pädagogische Fachkräfte in Kitas und Schulen erhalten durch ihre Beobachtungen und Interaktionen mit den Kindern bzw. Jugendlichen letztlich laufend diagnostische Informationen über diese, die sie bewerten und einordnen. Die Durchführung von Diagnostik-Tools bindet selbst bei weniger aufwändigen Tests zeitliche und personelle Ressourcen. Der Einsatz muss sich aus der Perspektive der Erzieherinnen und Erzieher bzw. Lehrerinnen und Lehrer daher lohnen. Diagnostik-Tools, in denen klar beschrieben wird, welche Informationen mit der Durchführung gewonnen werden können (vgl. Titz, Ropeter & Hasselhorn, 2018), haben den Vorteil, dass Fachkräfte besser abschätzen können, ob sich für sie der Einsatz lohnt oder nicht. Eine hohe Augenscheinvalidität (► Abschnitt 2.1 *Validität*) eines Tests ist dabei sicherlich hilfreich, kann aber durch eine nachvollziehbare und überzeugende Beschreibung von Zweck und Nutzen eines Tests ausgeglichen werden. Ob die Bemühungen von Testanbieterinnen und -anbietern in dieser Hinsicht dann aber auch tatsächlich auf Akzeptanz bei den Anwenderinnen und Anwendern stoßen, hängt letztlich von den Werten und Einstellungen potenzieller Anwenderinnen und Anwender in der pädagogischen Praxis ab.

3 Welche Kriterien kennzeichnen gute Förder-Tools?

Die Auswahl eines geeigneten Werkzeugs zur Sprachförderung stellt für Fachkräfte nicht nur aufgrund der nahezu unüberschaubaren Fülle an Programmen, Konzepten und Materialien eine anspruchsvolle Aufgabe dar. Vorliegende Tools zur Umsetzung von Sprach- und Lesefördermaßnahmen unterscheiden sich zudem teilweise erheblich hinsichtlich der definierten Zielgruppen (etwa in Bezug auf Alter, Spracherwerbstyp oder Sprachstand), ihrer Zielsetzung und den damit verbundenen vorgeschlagenen Zielbereichen, der inhaltlichen und konzeptionellen Ausrichtung und des Grades ihrer Strukturierung (Paetsch, Wolf, Stanat & Darso, 2014; Redder et al., 2011).

Im Gegensatz zu den in Abschnitt 2 genannten Kriterien für Diagnostik-Tools liegen für die wissenschaftliche Beurteilung von Förder-Tools bislang weniger fest etablierte Standards vor. Dennoch finden sich dazu durchaus Anhaltspunkte in der einschlägigen Literatur. Zur Identifizierung qualitativ hochwertiger Tools werden daher im Folgenden unstrittige Qualitätskriterien aus wissenschaftlicher Sicht

erläutert, bevor Anforderungen an gute Förder-Tools aus Praxis-Perspektive beschrieben werden.

3.1 Wissenschaftliche Gütekriterien

Bei Förder-Tools sind aus wissenschaftlicher Perspektive vor allem zwei Qualitätskriterien relevant: Zum einen sollte ein qualitativ hochwertiges Tool über eine hinreichende theoretische Fundierung verfügen, zum anderen sollten empirische Nachweise zu seiner Wirksamkeit vorliegen.

Theoretische Fundierung

Das Qualitätsmerkmal der *theoretischen Fundierung* erinnert zu Recht an die Überlegungen zur Validität von Diagnostik-Tools, die wir oben dargelegt haben. Von einer wissenschaftlich fundierten Förderung erwartet man eine Erklärung dafür, wie der eingeschlagene Förderweg zu den gewünschten Effekten führt. Daher sollte sich ein Förder-Tool auf wissenschaftlich geprüfte und dabei bewährte Theorien sowie aktuelle Forschungsergebnisse (z. B. zum Spracherwerb oder zum Leseerwerb) beziehen. Ist eine in sich plausible theoretische Fundierung gegeben, stellt sich die Frage, ob die im Tool genannten Fördermaßnahmen auch erkennbar aus dieser Theorie abgeleitet wurden. Die Form der Förderung und die Auswahl einzelner Fördermaßnahmen sind stets von der zugrunde liegenden Theorie abhängig und werden maßgeblich von theoretischen Vorannahmen bestimmt (Kany & Schöler, 2007). Gute Förder-Tools sollten daher auch eine genaue Beschreibung über die Ableitung einzelner Förderelemente aus der zugrunde liegenden Theorie sowie eine Begründung der Relevanz dieser Elemente für die ausgewählte Zielgruppe enthalten. Von großer Bedeutung ist es, Sprachförder-Tools linguistisch zu fundieren, d. h. bei der Beschreibung einzelner Förderbausteine sollten differenzierte Angaben und Begründungen dazu vorliegen, welche sprachlichen Phänomene innerhalb welcher Zielbereiche gefördert werden sollen. Bei der Auswahl eines Tools ist daher immer zu prüfen, ob sich die Konzeption, die Inhalte sowie die Auswahl der Fördermaterialien eines in Betracht gezogenen Tools auch aus Sicht der beschriebenen theoretischen Grundlagen für die angestrebten Förderziele eignen und an relevanten Erwerbsprozessen der Zielgruppe orientieren. Dies erhöht die Wahrscheinlichkeit, dass eine individuelle und adaptive Förderung gelingt, die an bereits erworbene Kompetenzen anknüpft. In einem Handbuch zum Tool sollte der theoretische Zugang zum Förderansatz daher stets expliziert und darauf aufbauend Ableitungen aus der Theorie auf die Ziele und die Zielgruppe der Förderung reflektiert und erläutert werden. Hinweise auf positive Erfahrungen aus der Praxis ersetzen keine theoretische Fundierung.

Wirksamkeitsnachweise

Neben der theoretischen Fundierung stellt auch die *empirische Fundierung*, d. h. der Nachweis der Wirksamkeit, ein zentrales Kriterium zur Beurteilung der Qualität

eines Förder-Tools dar. Ein Förder-Tool ist dann empirisch fundiert, wenn sich das Tool in Evaluationsstudien als prinzipiell wirksam hinsichtlich der im Tool angegebenen Zielbereiche erwiesen hat. Es muss also ein gültiger Nachweis darüber vorliegen, dass die sprachlichen Kompetenzen der mit dem Tool geförderten Kinder und Jugendlichen auch tatsächlich ansteigen. Positive Eindrücke in Form informeller Erfahrungsberichte über die Wirksamkeit eines Tools, oder Urteile zur Wirksamkeit aus Anwenderinnen- und Anwenderbefragungen reichen als Beleg für die Wirksamkeit eines Tools nicht aus. Die empirische Fundierung eines Tools gilt nur dann als bestätigt, wenn wenigstens ein erfolgreicher empirischer Nachweis der Wirksamkeit für eine vergleichbare Zielgruppe unter geeigneten kontrollierten Bedingungen vorliegt.

Zur Prüfung der Wirksamkeit sind unterschiedliche Designs verbreitet. Eine kontrollierte Anwendung des zu prüfenden Förder-Tools im Rahmen einer Intervention ist allerdings bei allen Design-Varianten unabdingbar. In Abhängigkeit von den zu prüfenden Hypothesen können neben einer Experimentalgruppe, in der die Intervention durchgeführt wird, Kontrollgruppen und weitere Experimentalgruppen mit Alternativinterventionen eingeführt werden. Letztere eignen sich in erster Linie dazu, die Stärke von Auswirkungen einer neuen Intervention gegenüber bereits etablierten Interventionen abzuschätzen. Kontrollgruppen hingegen werden häufig dazu verwendet, die Veränderung im relevanten Bildungsmerkmal im Verlaufe der Interventionsdurchführung in einen interventionsverursachten Anteil und einen auf natürliche Entwicklung zurückführbaren Anteil aufzuteilen. Im Rahmen der Datenerhebung muss mindestens das relevante Merkmal – also die sprachliche Kompetenz, die gefördert werden soll – vor (Prätest) und nach (Posttest) der Durchführung der Fördermaßnahme gemessen werden, um die angestrebte Veränderung quantifizierbar machen zu können. Häufig sind allerdings nicht nur kurzfristige, sondern vor allem mittel- oder langfristige Auswirkungen einer Intervention von Interesse. Daher sind in vielen Designs sogenannte Follow-up-Messungen des relevanten Merkmals im Abstand von mehreren Wochen oder Monaten zum Ende der Intervention Bestandteil des Erhebungsplans (Hager, 2000; Hasselhorn & Hasselhorn, 2017; Rost, 2007).

Bis dato gibt es für viele Förder-Tools kaum belastbare Studien mit überzeugenden Wirksamkeitsnachweisen. Eine der wenigen Ausnahmen sind Wirksamkeitsnachweise für Trainings zur phonologischen Bewusstheit im Elementarbereich (Jäger et al., 2012; Pröscholdt et al., 2013). Im Hinblick auf die Verbesserung allgemeiner Sprachkompetenzen sind bisher dokumentierte Ergebnisse jedoch eher ernüchternd, da die Sprachentwicklung von Kindern in vielen Studien nicht auf die Förderung mit spezifischen Tools zurückgeführt werden konnte (z. B. Gasteiger-Klicpera, Knapp & Kucharz, 2010; Hofmann, Polotzek, Roos & Schöler, 2008). Auch im Schulbereich liegen bis dato kaum eindeutige Befunde zur Wirksamkeit von Maßnahmen zur Sprachförderung vor (Limbird & Stanat, 2006; Paetsch et al., 2014).

Woran liegt es, dass so wenige bisherige Prüfversuche zur Wirksamkeit von Tools zur Sprachförderung erfolgreich waren? Hierfür wurden zum einen die oftmals mangelnde theoretische Fundierung, zum anderen methodische Unzulänglichkeiten von Evaluationsstudien verantwortlich gemacht (Kaltenbacher, 2011;

Limbird & Stanat, 2006; Redder et al., 2011; Tietze et al., 2014). So blieb in vielen Untersuchungen offen, ob und wie die zu evaluierenden Tools im Rahmen der Intervention z. B. hinsichtlich Dauer, Regelmäßigkeit und Gruppengröße sachgemäß umgesetzt wurden. Darüber hinaus ist ein Wirksamkeitsnachweis für ein Förder-Tool immer nur ein Hinweis darauf, dass das Tool bei sachgerechter Nutzung wirksam sein *kann*. Ob dies im Einzelfall auch funktioniert, hängt auch von der Kompetenz der pädagogischen Fachkräfte zur erfolgreichen Umsetzung der Sprachförderung ab (Baumann & Becker-Mrotzek, 2014; Müller, 2014). Inwiefern die Durchführenden in den bislang untersuchten Studien über hinreichende Sprachförderkompetenzen verfügten, um die Tools sachgerecht nutzen zu können, wurde in vielen der derzeit vorliegenden Studien nicht thematisiert.

3.2 Praxis-Perspektive

Aus Perspektive der Praxis werden im Folgenden analog zur Darstellung in Abschnitt 2 wesentliche Aspekte zur Durchführbarkeit und Akzeptanz von Werkzeugen zur Sprach- und Leseförderung erläutert.

Durchführbarkeit

Um Bedarfen aus der Praxis gerecht zu werden, sollten Förder-Tools klare und eindeutige Angaben zu ihrer Durchführung beinhalten. Eine sachgemäße Verwendung eines Tools setzt voraus, dass in einem Manual oder Handbuch beschrieben ist, welche Zielgruppe wie in der Umsetzung zu adressieren ist, wie sich die oder der Durchführende vorzubereiten hat, welche Settings herzustellen sind und welche Methoden und/oder Materialien wie einzusetzen sind.

In Bezug auf die Zielgruppe eines Förder-Tools werden konkrete Angaben darüber benötigt, für welche Adressatengruppe sich das Förder-Tool eignet. Falls die Zielgruppe spezifisch eingeschränkt ist (z. B. leseschwache Kinder und Jugendliche), sollte im Handbuch zum Tool angegeben werden, auf Grundlage welcher diagnostischen Information die Zugehörigkeit eines Kindes oder Jugendlichen zu dieser Subgruppe festgestellt werden kann (► Abschnitt 2). Neben der zu fördernden Zielgruppe sollten auch konkrete Informationen darüber bereitgestellt werden, ob der Einsatz des Tools besondere Qualifikationen der durchführenden Person voraussetzt und deshalb nur von einem eingeschränkten Personenkreis verwendet werden sollte (z. B. pädagogische Fachkräfte, Deutsch- oder Fachlehrkräfte). Damit eine fachgerechte Umsetzung eines Förder-Tools gewährleistet werden kann, sollte in einem Manual oder Handbuch auch angegeben werden, ob und welche Fortbildungen oder Schulungen zum adäquaten Einsatz des Tools erforderlich sind. Aus den Durchführungshinweisen muss darüber hinaus möglichst klar hervorgehen, unter welchen kontextuellen Bedingungen bzw. in welchem Setting das Tool umzusetzen ist. Hinweise zum Grad der Strukturierung sowie zur Häufigkeit, Dauer und Intensität der Förderung erleichtern Fachkräften die sachgemäße Umsetzung.

Während Sprachförderprogramme meist aus vorgefertigten und curricular aufgebauten Fördereinheiten bestehen, die in einer festgelegten Reihenfolge zeitlich begrenzt umgesetzt werden, basieren Förderkonzepte häufig eher auf allgemeinen Förderprinzipien. Entsprechend können bei Programmen und Konzepten keine identischen Anforderungen an den Grad der Strukturierung angelegt werden: Konzepte sind in der Regel allgemeiner und stellen anstatt eines festen Curriculums eher das Handeln der Fachkräfte in den Mittelpunkt, die für die Durchführung der Fördermaßnahme verantwortlich sind. Dies wiederum stellt hohe Anforderungen an die Professionalität der Fachkräfte. Hinweise zur Häufigkeit, Dauer und Intensität der Förderung sind in Konzepten daher häufig weniger spezifisch als bei Programmen. Präzise Hinweise zum Schwierigkeitsgrad einzelner Förderelemente sowie zu einer sinnvollen Reihenfolge einzelner Bausteine in konkreten Fördersituationen helfen den Durchführenden bei der sachgemäßen Umsetzung des Tools. Hinsichtlich ihrer Organisationsform können sowohl Förderkonzepte als auch Förderprogramme entweder alltags- bzw. unterrichtsintegriert und/oder additiv in Einzel- oder Kleingruppenförderung organisiert sein. In einigen vielversprechenden Ansätzen werden alltags- bzw. unterrichtsintegrierte Fördermaßnahmen und additive Fördermaßnahmen miteinander kombiniert. Dabei entstehen Mischformen, indem beispielsweise sogenannte Fokuskinder mit ausgewiesenem Förderbedarf identifiziert werden, die innerhalb der Gesamtgruppe gezielt gefördert werden können (vgl. Geyer, Schwarze & Müller, 2018).

Ein weiteres Merkmal für die praktische Durchführbarkeit eines Förder-Tools stellen eindeutige Angaben zu Methoden und Materialien dar, mit denen das Tool umzusetzen ist. Während Methoden in Sprachförderprogrammen häufig als fester curricularer Baustein im Tool integriert und benötigte Materialien im Tool enthalten sind, erfolgen Hinweise zu geeigneten Methoden und Materialien in Sprachförderkonzepten häufig auf einer allgemeineren Ebene. Dabei ist der adaptive Einsatz, d. h. die Passung der angewendeten Materialien und Methoden zum Ziel sowie zur zu fördernden Zielgruppe, zentral (Hardy et al., 2011; Klieme & Warwas, 2011; Müller, 2015), da je nach verfolgtem Ziel sowie je nach Ausgangslage der zu fördernden Individuen unterschiedliche Methoden geeignet sein können. Bedeutsam ist darüber hinaus, dass ausgewählte Materialien thematisch und inhaltlich zur Lebenswelt der Kinder und Jugendlichen passen und altersgerecht aufbereitet sind.

Akzeptanz

Ein wichtiger Erfolgsgarant für ein Förder-Tool ist dessen Akzeptanz in der Praxis. Bei der Beurteilung der Qualität eines Förder-Tools ist daher auch die Frage relevant, ob ein bestimmtes Tool aus Sicht der Anwenderinnen und Anwender sinnvoll ist, da es nur bei einer entsprechenden positiven Beurteilung eine gute Chance hat, von der Praxis aufgegriffen und nachhaltig genutzt zu werden. Wichtig ist daher eine nachvollziehbare und überzeugende Beschreibung des Nutzens eines Förder-Tools.

Bislang liegen keine unabhängigen und repräsentativen Befragungen zur Beurteilung bestimmter Sprachförderinstrumente und -konzepte durch Anwende-

rinnen und Anwender in der Praxis vor. Die Aussagemöglichkeiten zur Akzeptanz von Tools in der Praxis sind daher derzeit sehr begrenzt.

4 Fazit und Ausblick

Viele Diagnostik-Tools und Förder-Tools sind derzeit auf dem Markt verfügbar. Für pädagogische Fachkräfte in der Praxis stellt sich daher die Frage, ob ein konkretes Tool gut bzw. für das eigene Förderanliegen passend ist oder nicht. Dieser Beitrag gibt Hilfestellung und Orientierung für die Klärung dieser Frage, indem er Qualitätskriterien guter Tools zur Diagnose und Förderung von Sprach- und Lesekompetenzen erläutert, die pädagogischen Fachkräften bei der Bewertung und Auswahl geeigneter Tools helfen. Neben der Passung von Tool, Zielbereich und Zielgruppe der Förderung, die als grundlegendes Prinzip bei der Auswahl von Diagnostik- und Förder-Tools gilt, wurden im Beitrag getrennt für Diagnostik- und Förder-Tools sowohl wissenschaftliche Qualitätskriterien als auch Kriterien erläutert, die die Perspektive der Praxis berücksichtigen.

Folgende Leitfragen dienen als Hilfestellung bei der Auswahl qualitativ hochwertiger Diagnostik-Tools:

- Sind Durchführung, Auswertung sowie Interpretation des Tools eindeutig beschrieben bzw. standardisiert? (Objektivität)
- Gibt es nachvollziehbare Angaben dazu, ob das Tool zuverlässig ist? (Reliabilität)
- Verfügt das Tool über eine theoretische und empirische Fundierung? (Validität)
- Verfügt das Tool über eine geeignete Normierung?
- Ist die Umsetzung klar und nachvollziehbar beschrieben? (Durchführbarkeit)
- Gibt es überzeugende Argumente dafür, dass das Tool gut in die Praxis integrierbar ist? (Akzeptanz)

Für die Bewertung von Förder-Tools schlagen wir folgende Leitfragen vor:

- Basiert das Tool auf einer geeigneten Theorie (etwa zum (Zweit-)Spracherwerb oder zum Lesererwerb) und werden die einzelnen Förderschritte aus der Theorie abgeleitet? (theoretische Fundierung)
- Liegen empirische Nachweise zur prinzipiellen Wirksamkeit des Tools vor? (Wirksamkeitsnachweise)
- Ist die Durchführung eindeutig und nachvollziehbar beschrieben? (Durchführbarkeit)
- Gibt es überzeugende Argumente dafür, dass das Vorgehen sinnvoll und gut in die Praxis integrierbar ist? (Akzeptanz)

Wir verbinden mit diesem Beitrag die Hoffnung einer zukünftig besser möglichen Auswahl von Werkzeugen zur Diagnostik und Förderung von Kindern und Jugendlichen in der pädagogischen Praxis. Dabei sind wir uns darüber im Klaren, dass die Auswahl guter Tools bei aller Bedeutsamkeit nur die halbe Miete auf dem Weg zu einer erfolgreichen Förderung in der Praxis darstellt. Erst die professionellen Kompetenzen der durchführenden Fachkräfte machen aus guten Tools auch tatsächlich wirksame.

Literatur

- Amelang, M. & Zielinski, W. (2002). *Psychologische Diagnostik und Intervention*. Heidelberg: Springer.
- Baumann, B. & Becker-Mrotzek, M. (2014). *Sprachförderung und Deutsch als Zweitsprache an deutschen Schulen: Was leistet die Lehrerbildung? Überblick, Analysen und Handlungsempfehlungen*. Köln: Mercator-Institut für Sprachförderung und Deutsch als Zweitsprache.
- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion* (3. aktual. u. erw. Aufl.). München: Pearson Studium.
- Dietz, S. & Lisker, A. (2010). *Sprachstandsfeststellung und Sprachförderung im Kindergarten. Expertise im Auftrag des Deutschen Jugendinstituts*. Verfügbar unter: http://www.dji.de/fileadmin/user_upload/bibs/Sprachstandsfeststellung_Dietz_Lisker.pdf [06.03.2017].
- Fisseni, H.-J. (1997). *Lehrbuch der psychologischen Diagnostik* (2. Aufl.). Göttingen: Hogrefe.
- Gasteiger-Klicpera, B., Knapp, W. & Kucharz, D. (2010). *Abschlussbericht der Wissenschaftlichen Begleitung des Programms »Sag´ mal was – Sprachförderung für Vorschulkinder«*. Pädagogische Hochschule Weingarten.
- Geyer, S., Schwarze, R. & Müller, A. (2018). Sprachförderung im Elementarbereich. In C. Titz, S. Geyer, A. Ropeter, H. Wagner, S. Weber & M. Hasselhorn (Hrsg.), *Konzepte zur Sprach- und Schriftsprachförderung entwickeln (Bildung durch Sprache und Schrift, Band 1, S. 161–178)*. Stuttgart: Kohlhammer.
- Grimm, A. & Schulz, P. (2016). Warum man bei mehrsprachigen Kindern dreimal nach dem Alter fragen sollte: Sprachfähigkeiten simultan-bilingualer Lerner im Vergleich mit monolingualen und frühen Zweitsprachlernern. *Diskurs Kindheits- und Jugendforschung, 11*(1), 27–42.
- Hager, W. (2000). Zur Wirksamkeit von Interventionsprogrammen: Allgemeine Kriterien der Wirksamkeit in einzelnen Untersuchungen. In W. Hager, J. L. Patry & H. Brezing (Hrsg.), *Evaluation psychologischer Interventionsmaßnahmen* (S. 153–168). Bern: Huber.
- Hardy, I., Hertel, S., Kunter, M., Klieme, E., Warwas, J., Büttner, G. & Lühken, A. (2011). Adaptive Lerngelegenheiten in der Grundschule. Merkmale, methodisch-didaktische Schwerpunktsetzungen und erforderliche Lehrerkompetenzen. *Zeitschrift für Pädagogik, 57*(6), 819–833.
- Hasselhorn, M. & Hasselhorn, J. (2017). Quantitative methodische Zugänge der empirischen Bildungsforschung: Funktionen für den Erkenntnisgewinn und Relevanz für die Bildungspraxis. In M. Heinrich, C. Kölzer & L. Streblov (Hrsg.), *Forschungspraxen der Bildungsforschung* (S. 39–53). Münster: Waxmann.
- Hofmann, N., Polotzek, S., Roos, J. & Schöler, H. (2008). Sprachförderung im Vorschulalter - Evaluation dreier Sprachförderkonzepte. *Diskurs Kindheits- und Jugendforschung, 3*, 291–300.

- Jäger, D., Faust, V., Blatter, K., Schöppe, D., Artelt, C., Schneider, W. & Stanat, P. (2012). Kompensatorische Förderung am Beispiel eines vorschulischen Trainings der phonologischen Bewusstheit. *Frühe Bildung*, 1 (4), 202–209.
- Jampert, K., Best, P., Guadatiello, A., Holler, D. & Zehnauer, A. (2007). *Schlüsselkompetenz Sprache. Sprachliche Bildung und Förderung im Kindergarten. Konzepte – Projekte – Maßnahmen*. Weimar/Berlin: Verlag das netz.
- Kaltenbacher, E. (2011). Zur Problematik der Evaluation von Sprachfördermaßnahmen. In N. Hahn & T. Roelcke (Hrsg.), *Grenzen überwinden mit Deutsch* (S. 163–178). Göttingen: Universitätsverlag.
- Kany, W. & Schöler, H. (2007). *Fokus: Sprachdiagnostik. Leitfaden zur Sprachstandsbestimmung im Kindergarten*. Berlin: Cornelsen Scriptor.
- Klieme, E. & Warwas, J. (2011). Konzepte der Individuellen Förderung. *Zeitschrift für Pädagogik*, 57 (6), 805–818.
- Lienert, G. A. & Raatz, U. (1998). *Testaufbau und Testanalyse* (6. Aufl.). Weinheim: Beltz.
- Limbird, C. K. & Stanat, P. (2006). Sprachförderung bei Schülerinnen und Schülern mit Migrationshintergrund: Ansätze und ihre Wirksamkeit. In J. Baumert, P. Stanat & R. Watermann (Hrsg.), *Herkunftsbedingte Disparitäten im Bildungswesen: Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit. Vertiefende Analysen im Rahmen von PISA 2000* (S. 257–307). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Lisker, A. (2011). *Additive Maßnahmen zur vorschulischen Sprachförderung in den Bundesländern. Expertise im Auftrag des Deutschen Jugendinstituts*. Verfügbar unter: <https://www.deutsche-digitale-bibliothek.de/binary/Z6VFVEX35GTE32TQHLLVISY-JYCOF2C6K/full/1.pdf> [17.07.2017].
- Müller, A. (2014). Profession und Sprache: Die Sicht der (Zweit-)Spracherwerbsforschung. In T. Betz & P. Cloos (Hrsg.), *Kindheitspädagogische Beiträge. Kindheit und Profession – Konturen und Befunde eines Forschungsfeldes* (S. 66–83). Weinheim: Beltz Juventa.
- Müller, A. (2015). Spracherwerbstheoretische Aspekte der Zweitsprachdidaktik. In H. Klages & G. Pagonis (Hrsg.), *Linguistisch fundierte Sprachförderung und Sprachdidaktik: Grundlagen, Konzepte, Desiderate* (S. 123–139). Berlin: Mouton de Gruyter.
- Müller, A., Schulz, P. & Tracy, R. (2018). Spracherwerb. In C. Titz, S. Geyer, A. Ropeter, H. Wagner, S. Weber & M. Hasselhorn (Hrsg.), *Konzepte zur Sprach- und Schriftsprachförderung entwickeln (Bildung durch Sprache und Schrift, Band 1, S. 53–68)*. Stuttgart: Kohlhammer.
- Neugebauer, U. & Becker-Mrotzek, M. (2013). *Die Qualität von Sprachstandsverfahren im Elementarbereich. Eine Analyse und Bewertung*. Herausgegeben vom Mercator-Institut für Sprachförderung und Deutsch als Zweitsprache. Verfügbar unter http://www.mercator-institut-sprachfoerderung.de/fileadmin/user_upload/Mercator-Institut_Qualitaet_Sprachstandsverfahren_Web_03.pdf [06.03.2017].
- Paetsch, J., Wolf, K. M., Stanat, P. & Darsow, A. (2014). Sprachförderung von Kindern und Jugendlichen aus Zuwandererfamilien. *Zeitschrift für Erziehungswissenschaft*, 17, 315–347.
- Pröscholdt, M. V., Michalik, A., Schneider, W., Duzy, D., Glück, D., Souvignier, E. & Penner, Z. (2013). Effekte kombinierter Förderprogramme zur phonologischen Bewusstheit und zum Sprachverstehen auf die Entwicklung der phonologischen Bewusstheit von Kindergartenkindern mit und ohne Migrationshintergrund. *Frühe Bildung*, 2 (3), 122–132.
- Redder, A., Schwippert, K., Hasselhorn, M., Forschner, S., Fickermann, D. & Ehlich, K. (2011). *Bilanz und Konzeptualisierung von strukturierter Forschung zu »Sprachdiagnostik und Sprachförderung«*. ZUSE-Berichte Band 2. Hamburg: Hamburger Zentrum für Unterstützung der wissenschaftlichen Begleitung und Erforschung schulischer Entwicklungsprozesse (ZUSE).
- Rost, D. H. (2007). *Interpretation und Bewertung pädagogisch-psychologischer Studien*. Weinheim: Beltz.
- Schneider, W. & Marx, P. (2008). Früherkennung und Prävention von Lese-Rechtschreibschwierigkeiten. In F. Petermann & W. Schneider (Hrsg.), *Enzyklopädie der Psychologie, Band 7: Angewandte Entwicklungspsychologie* (S. 237–273). Göttingen: Hogrefe.

- Tietze, W., Kammermeyer, G., Laier, M., Roux, S., Tracy, R., Weber, A., Jahn, M., King, S., Lee, H.-J., Metz, A., Nunnenmacher, S., Raible, G., Roux, S., Stolberg, D. & Vogelmann, U. (2014). Erfahrungen und Reflexionen: Diskussion des Programms »Sprachliche Bildung für Kleinkinder«. In Baden-Württemberg Stiftung (Hrsg.), *Sag' mal was – Sprachliche Bildung für Kleinkinder. Sprachförderansätze: Erfahrungen und Reflexionen* (S. 125–137). Tübingen: Francke.
- Titz, C., Ropeter, A. & Hasselhorn, M. (2018). Zum Mehrwert von Diagnostik. In C. Titz, S. Geyer, A. Ropeter, H. Wagner, S. Weber & M. Hasselhorn (Hrsg.), *Konzepte zur Sprach- und Schriftsprachförderung entwickeln (Bildung durch Sprache und Schrift, Band 1, S. 87–100)*. Stuttgart: Kohlhammer.