Buchwald, K., Ebenbeck, N. & Gebhardt, M. (2022). Screenings, Status- und adaptive Tests in der schulischen Diagnostik. In M. Gebhardt, D. Scheer & M. Schurig (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik. Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung* (S. 305-312). Regensburg: Universitätsbibliothek. https://doi.org/10.5283/epub.53149

Screenings, Status- und adaptive Tests in der schulischen Diagnostik

Katharina Buchwald, Nikola Ebenbeck & Markus Gebhardt

Pädagogische Diagnostik ermöglicht im Unterricht die Lernausgangslage oder die Lernentwicklung durch standardisierte Tests objektiv, valide und reliabel festzustellen und aus den Ergebnissen Informationen für die Praxis abzuleiten (Hartung et al., 2021). Tests haben in der sonderpädagogischen Praxis eine lange Tradition und wurden schon zu Zeiten der Hilfsschule zur Feststellung der Hilfsschulbedürftigkeit und nun in Zeiten der inklusiven Schule zur Feststellung des sonderpädagogischen Unterstützungsbedarfs verwendet (Gebhardt et al., 2021). Die Einsatzmöglichkeiten von Tests sind jedoch vielfältig und gehen über Feststellungsdiagnostik weit hinaus. In diesem Beitrag sollen daher zunächst verschiedene Kategorien von Test dargestellt werden, um dann insbesondere den Nutzen und die Anwendungen von adaptiven Tests für die sonderpädagogische Praxis zu vertiefen.

1 Vom Screening zur LVD: Vier Kategorien von Schulleistungstests

Je nach Einsatz und Häufigkeit der Messung gibt es verschiedene Formen von Tests, welche in der Praxis häufig eingesetzt werden. Nach Hasbrouck und Tindal (2006) kann man vier verschiedene Arten an diagnostischen Instrumenten je nach Einsatzgebiet unterscheiden:

Screenings sind kurze ökonomische Tests, welche reliabel Risikokinder identifizieren möchten und einen schnellen Überblick über einen spezifischen Leistungsstand für eine Person oder eine Klasse geben. Screenings messen möglichst zu einem Messzeitpunkt einen Teil eines Lernbereichs, welcher für das Erlernen einer Kompetenz ausschlaggebend ist. Aus den Ergebnissen kann der Ist-Stand eines Kindes im jeweiligen Lernbereich abgeleitet werden, weshalb der Einsatz zu Beginn des Schuljahres empfohlen wird (Hasbrouck & Tindal, 2006). Lehrkräfte verwenden Screenings, um einen schnellen und einfachen Überblick über den Leistungsstand einer Gruppe oder Person zu bekommen und didaktische Entscheidungen treffen zu können. Die ökonomische Handhabung steht im Vordergrund, damit die Tests z. B. einfach und schnell in inklusiven Klassen durchgeführt werden. So können Kinder mit Schwierigkeiten frühzeitig erkannt und alternative Formen der Förderung abgeleitet werden (Gebhardt, 2021; Hartung et al., 2021).

Ein kostenfreies Screening für den Bereich Mathematik ist beispielsweise die Reihe Mathes, welche auf dem Portal lernlinien.de (Blumenthal, 2021), erschienen ist. Es wurde für die Einschulung bis zur 4. Klasse entworfen und geprüft. Das Verfahren dauert je nach Klassenstufe zwischen 30 und 45 Minuten (ebd.). Da das Screening mehrere Aufgabengruppen und somit Kompetenzbereiche umfasst, dauert es sehr lang. Screenings können auch mit einer kürzeren Durchführungszeit konstruiert sein, dann messen sie jedoch ein weniger breites Testprofil. Screenings selbst wollen nur einen Einblick geben und sind daher kürzer sowie schneller durchführbar als umfangreiche Schulleistungstest. Für eine umfassende Einzelfalldiagnostik und die Feststellung eines sonderpädagogischen Gutachtens sind Screenings daher nicht geeignet. Screenings dienen in erster Linie der didaktischen Entscheidungsfindung. Ebenso können Risikokinder frühzeitig identifiziert werden, um diesen Kindern präventiv passendere und bessere Förderung zukommen zu lassen.

Schulleistungstests (Statustest; *diagnostic measures*) sind normierte Tests, welche ebenfalls zu einem Messzeitpunkt einen Kompetenzbereich oder die Kompetenzen eines ganzen Schulfaches messen (Hasbrouck & Tindal, 2006). Sie werden durchgeführt, wenn eine umfassende Analyse der Fähigkeiten des Kindes notwendig ist, um ein sonderpädagogisches Gutachten zu erstellen (Gebhardt, 2021). So kann es sein, dass ein Kind schon mittels Screenings identifiziert worden ist und trotz intensiver pädagogischer Bemühungen keine Fortschritte zu erkennen sind oder mehr diagnostische Informationen erfasst werden müssen. Dann wäre der nächste Schritt ein umfassendes Profil des Kindes mit einem oder mehreren normierten Testverfahren zur Testung des aktuellen Status zu erstellen. Im Unterschied zu Screenings soll bei Schulleistungstests eine Kompetenz und der Ist-Zustand möglichst umfassend geprüft werden. Der zeitliche Faktor und die Ökonomie sind dementsprechend weniger wichtige Kriterien. Deshalb bestehen diese Tests aus mehreren Dimensionen (Subtests), d.h. sie sind meist multidimensional, beinhalten verschiedene Aufgabenbereiche mit mehreren Aufgaben (Items) und prüfen verschiedene Bereiche der Kompetenz. Damit kann ein genaueres Profil des Kindes im Vergleich zu einer Norm skizziert und angegeben werden, ob ein Kind im durchschnittlichen oder unter- bzw. überdurchschnittlichen Bereich einer Kompetenz im Altersbereich liegt. Schulleistungstests dauern aufgrund ihres Aufbaus meist ein bis zwei Schulstunden und benötigen bei der Interpretation eine fachliche Expertise, um ein vollumfängliches Profil über die Stärken und Bedürfnisse von Schüler:innen zu erstellen. Häufig werden diese Tests im Einzelsetting durchgeführt, um das Antwortverhalten des Kindes beobachten und sicherzustellen zu können, dass die Instruktionen korrekt verstanden werden. Da der Testprozess dementsprechend aufwendig ist, werden Schulleistungstests meist für Gutachten und zur Klärung individuell wichtiger pädagogischer Fragestellungen durchgeführt (Gebhardt, 2021; Hartung et al., 2021).

Im Rahmen solcher Gutachten werden oft Schulleistungstests wie beispielsweise der MBK1+ (Ennemoser et al., 2017) und auch Intelligenztests wie der WISC-V (Wechsler, 2017) verwendet.

Kriteriumsbezogene Tests oder Leistungstests (Outcome Measures) sind umfassende Schulleistungstests und Ergebnismessungen, um zu bestimmen, ob die Schüler:innen das Jahrgangsstufenziel erreicht haben. Diese Tests werden entweder am Ende einer Intervention oder am Ende des Schuljahres durchgeführt (Hasbrouck & Tindal, 2006). In Deutschland legen beispielsweise im Primarbereich die Bildungsstandards der KMK (2005) fest, welche Kompetenzen bis zu einer bestimmten Jahrgangsstufe erreicht werden sollen. Mittels kriteriumsorientierte Tests kann man das Erreichen des Kriteriums prüfen, wie es z. B. die Materialien und Tests der Individuelle Lernstandsanalysen (ILeA) ermöglichen (Liebers et al., 2019). Auf der anderen Seite gibt es Leistungstests in internationalen Studien wie IGLU (Hussmann et al., 2017) oder in den Län-

dervergleichsstudien des Instituts für Qualitätsentwicklung im Bildungswesen in Deutschland (Stanat et al., 2019), welche für das Bildungsmonitoring Kompetenzen auf Populationsebene prüfen, aber für die Einzelfalldiagnostik nicht geeignet sind.

Lernverlaufsdiagnostik besteht aus mindestens drei, meist kürzeren Tests, welche in einem kontinuierlichen (wöchentlichen oder monatlichen) Abstand erhoben werden, um den Lernverlauf eines Kindes über die Zeit darzustellen. Hierbei gibt es die Ziele: a) die Lernentwicklung des Kindes einzuschätzen, b) Kinder zu identifizieren, welche nicht adäquate Fortschritte machen, c) verschiedene Instruktionen für Kinder mit Risiko zu evaluieren (Hasbrouck & Tindal, 2006). Im Gegensatz zu den ersten drei diagnostischen Verfahren hat Lernverlaufsdiagnostik nicht das Ziel den Ist-Zustand zu messen, sondern durch mehrmalige Testungen den Lernverlauf über die Zeit hinweg möglichst fair und reliabel zu dokumentieren (Gebhardt et al., 2021; Klauer, 2014).

2 Herausforderungen an eine inklusive Diagnostik

Feststellungsdiagnostik, also oft (Schul-) Leistungstests, benötigt viel Zeit für die Testung und die häufig folgende Erstellung von FörderGutachten. Es bleibt daher weniger Zeit für die Förderung selbst sowie deren Evaluation. Aus diesem Grund fokussieren sich moderne inklusive Ansätze, wie der Response-to-Intervention-Ansatz (Fuchs & Fuchs, 2006), das Rügener Inklusionsmodell (Hartke, 2017) oder das Throughput Modell (Preuss-Lausitz, 2016) auf förderdiagnostische Instrumente, wie Screenings oder Lernverlaufsdiagnostik, die auf eine kurze und ökonomische Durchführung setzen. Dadurch kann die eigentliche Zeit mit dem Kind besser genutzt und die Testungen gut in den Unterrichtsalltag integriert werden. Da SchülerInnen mit sonderpädagogischem Unterstützungsbedarf und/oder Intelligenzminderung häufig auch Konzentrationsschwierigkeiten haben, ist ökonomisches Testen für sie besonders relevant, um eine möglichst faire Testung zu gewährleisten (Schurig et al., 2021). Betrachtet man diese Schüler:innen als Zielgruppe inklusiver Diagnostik, die präzise erkannt werden muss, dann steigen die Anforderungen an inklusive Diagnostik. Die Tests müssen fair und ökonomisch testen, um so effizient wie möglich zu sein. Außerdem müssen sie einfach in den Unterricht integrierbar und sowohl für SchülerInnen mit und ohne Förderbedarf geeignet sein.

Adaptives Testen und besonders computerisiertes adaptives Testen (CAT) erfüllt diese Herausforderungen an eine inklusive Diagnostik, wie erste Ergebnisse zeigen (Stone & Davey, 2011; Smith, 2015). Als Nebeneffekt steigert CAT die Motivation zur Testbearbeitung von Schüler:innen mit sonderpädagogischem Unterstützungsbedarf (Betz & Weiss, 1976).

3 Adaptives Testen

Adaptive Tests passen während der Testung ihre Schwierigkeit an die Fähigkeit der getesteten Person an. Hierfür wird häufig mit einem Item mit mittlerer Schwierigkeit gestartet. Kann dieses Item richtig beantwortet werden, wird der Testperson als nächstes ein schwierigeres Item vorgelegt. War das Item falsch, wird dementsprechend ein einfacheres Item zur Bearbeitung vorgegeben. Auf diese Art und Weise pendelt sich die Schwierigkeit des Tests schnell auf der höchst möglichen Schwierigkeit ein, die die Testperson mit ihren Fähigkeiten noch bearbeiten kann. Durch dieses Pendeln zwischen Aufgaben, die gerade nicht mehr gelöst werden können und solchen, die gerade noch gelöst werden können – sowie aufgrund der Basis der

Item-Response-Theory – liegt die Lösungswahrscheinlichkeit eines vorgelegten Items bei 50 %. Die Auswahl zur Bearbeitung vorgelegter Items am vorherigen Antwortverhalten der Testperson gibt keine starre Bearbeitungsreihenfolge vor, sondern verfolgt das Ziel, möglichst viel diagnostische Information über eine Testperson zu erhalten (Frey, 2020).

Aus pädagogischer Sicht muss vor Beginn der Testung thematisiert werden, dass der Test sowohl sehr leichte und sehr schwere Aufgaben enthält und auch gute SchülerInnen nicht alle Items lösen können. Insbesondere für Kinder mit guten Leistungen kann es daher ungewohnt sein, dass sie Aufgaben bekommen, welche sie zu 50% nicht lösen können. Für Kinder mit sonderpädagogischem Förderbedarf hingegen kann die Testsituation eher eine positive Erfahrung darstellen, da in den meist linear konstruierten Tests diese Kinder nur einen geringen Anteil an Aufgaben lösen würden. Adaptive Tests sind daher besonders für Kinder geeignet, welche nicht nahe dem Mittelwert der Norm liegen. Ebenso ermöglichen sie eine kürzere Testdurchführung bei vergleichbar hoher Reliabilität.

Die analoge Durchführung von adaptiven Tests ist zeitlich aufwendig und kann die Reliabilität z. B. aufgrund von Testleitereffekten kaum gewährleisten (Kubinger, 2021). Daher hat sich schon früh eine digitale Variante, das computerisierte adaptive Testen (CAT), durchgesetzt. CAT ist eine Weiterentwicklung des adaptiven Testens, das sich auf die Testkonstruktion und Testdurchführung am Computer bezieht. Durch einen vorher definierten bzw. programmierten Algorithmus werden die jeweils nächsten Aufgaben für die Testperson immer so ausgewählt, dass sie die größte Aussagekraft über die Fähigkeiten der Person treffen können und dementsprechend am besten zu ihren Kompetenzen passen. Neben der Ziehung spezifiziert man anhand des Algorithmus unter anderem auch, wann die Testung beendet ist oder abgebrochen wird (Magis et al., 2017; Frey, 2020).

Für die Konstruktion eines CATs wird eine speziell kalibrierte Sammlung von Items (Itempool) benötigt. Jedes Item im Itempool muss einen Schwierigkeitswert zugeordnet haben, um später passend durch den Algorithmus gezogen werden zu können. Um die Items im Itempool kalibrieren zu können, werden die Items zunächst mit mehreren Testpersonen, z. B. mit Schulklassen, bearbeitet, um Ergebnisse für jedes Item zu erhalten. Darauf folgend wird der Schwierigkeitswert durch die Passung zu einem Modell der Item-Response-Theory, z. B. dem Rasch-Modell, berechnet (Magis et al., 2017).

Im Vergleich zu regulären, nicht adaptiven Tests ist die Testkonstruktion eines CAT aufgrund der notwendigen Definierung und Anpassung des Algorithmus aufwendiger. Da jedoch Testleitereffekte entfallen und die Testung selbst aufgrund des Ziehalgorithmus kürzer ist, testet CAT im Vergleich zu analogem Testen objektiver, ökonomischer, reliabler und im inklusiven Bereich potenziell fairer.

4 Fazit

Für die pädagogische Diagnostik gibt es eine Reihe an unterschiedlichen Instrumenten, welche für unterschiedliche Zwecke konstruiert wurden. Insbesondere durch die Digitalisierung der Schulen ist der Einsatz von Tests durch die automatische Auswertung und Darstellung der Ergebnisse leichter geworden. Komplexe Auswertungen und umfangreiche Testdurchführungen sind so einfacher möglich. Dies gilt für alle Kategorien von pädagogischen Tests. Insbesondere adaptive Tests benötigen digitale Unterstützung, da eine analoge Durchführung sehr aufwendig wäre. Mit fortschreitender Digitalisierung werden standardisierte Auswertungen in pädagogi-

schen Materialien leichter für Testkonstrukteure und Verlage umzusetzen. Deshalb wird die schulische Praxis in Zukunft eher mit adaptiven Verfahren konfrontiert, da diese insbesondere bei einer heterogenen Schülerschaft genauer und sensitiver Messen können, wenn sie nach theoretischen Vorgaben konstruiert und anhand der Zielgruppe auch geprüft wurden. Eine Einschätzung der wissenschaftlichen Güte der eingesetzten Verfahren ist daher Pflicht für alle Lehrkräfte. Dies ist notwendig, da neben geprüften Verfahren auch Instrumente am Markt sind, welche für Kinder mit sonderpädagogischen Unterstützungsbedarf wenig bis gar nicht geeignet sind. Generell ist es aber zu begrüßen, dass mittelweile ein breites Angebot an verschiedenen Tests und informellen Instrumenten besteht, welche entweder aus der Wissenschaft, aus den Schulen selbst oder von kommerziellen Verlagen entwickelt wurden. Der adaptive Ansatz findet sich nicht nur in der Diagnostik, sondern auch in der Förderung. Jungjohann, Anderson und Gebhardt (2020) haben beispielsweise Materialien zur adaptiven Leseförderung umgesetzt, bei der individuell für jeden Schüler und jede Schülerin passendes Material vorgeschlagen wird. Dadurch kann auch bei der Förderung individueller gearbeitet werden (Jungjohann et al., 2020).

Literatur

Betz, N. E. & Weiss, D. J. (1976). *Psychological Effects of Immediate Knowledge of Results and Adaptive Ability Testing.: Research Report 76-4*.

Blumenthal (2021). *Lernlinie*. Universität Rostock. https://www.lernfortschrittsdokumentation-mv.de/_lernlinie/index.htm

Ennemoser, M., Krajewski, K. & Sinner, D. (2017). *MBK1+: Test mathematischer Basiskompetenzen ab Schuleintritt*. Hogrefe.

Frey, A. (2020). Computerisiertes adaptives Testen. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion*. Springer, 3. Aufl., 501-525. https://doi.org/10.1007/978-3-662-61532-4_20

Fuchs, D. & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it?. *Reading Research Quarterly, 41*(1), 93–99.

Gebhardt, M. (2021). *Inklusiv- und sonderpädagogische Pädagogik im Schwerpunkt Lernen. Eine Einführung (Version 0.2)*. Universität Regensburg. https://doi.org/10.5283/EPUB.45609

Gebhardt, M., Jungjohann, J. & Schurig, M. (2021). *Lernverlaufsdiagnostik im förderorientierten Unterricht: Testkonstruktionen, Instrumente, Praxis*. Ernst Reinhardt Verlag. http://www.reinhardt-verlag.de/de/titel/54852/

Hartke, B. (2017). Gelingende Inklusion – das Rügener Inklusionsmodell (RIM). In B. Hartke (Hrsg.), *Handlungsmöglichkeiten Schulische Inklusion: Das Rügener Modell kompakt* (S. 11–19). Kohlhammer Verlag.

Hartung, N., Schurig, M., Vossen, A. & Gebhardt, M. (2021). Pädagogische Diagnostik im Rahmen des RTI-Modells. In J. Kuhl, A. Vossen, N. Hartung & C. Wittich (Hrsg.), *Evidenzbasierte Förderung bei Lernschwierigkeiten in der Grundschule*. Ernst Reinhardt Verlag, 28-39.

Hasbrouck, J. & Tindal, G. A. (2006). Oral Reading Fluency Norms: A Valuable Assessment Tool for Reading Teachers. *The Reading Teacher, 59*(7), 636–644. https://doi.org/10.1598/RT.59.7.3

Hussmann, A., Wendt, H., Bos, W., Bremerich-Vos, A., Kasper, D., Lankes, E.-M., McElvany, N., Stubbe, T. C. & Valtin, R. (Hrsg.). (2017). *Waxmann-E-Books Empirische Erziehungswissenschaft. IGLU 2016: Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Waxmann Verlag. https://elibrary.utb.de/doi/book/10.31244/9783830987000

Jungjohann, J., Anderson, S. & Gebhardt, M. (2020). *Adaptive Leseförderung zur Steigerung der Leseflüssigkeit und des basalen Leseverständnisses »Levumis Leseabenteuer«*. Technische Universität Dortmund. https://doi.org/10.17877/DE290R-20992

Kubinger, K. (2021, 17. Februar). *Adaptives Testen*. URL: https://dorsch.hogrefe.com/stichwort/adaptives-testen – letzter Aufruf: 11.11.2021.

Klauer, K. J. (2014). Formative Leistungsdiagnostik: Historischer Hintergrund und Weiterentwicklung zur Lernverlaufsdiagnostik. In M. Hasselhorn, W. Schneider & U. Trautwein (Hrsg.), *Tests und Trends: Neue Folge Band 12. Lernverlaufsdiagnostik* (Bd. 12, S. 1–18). Hogrefe.

KMK. (2005). *Beschlüsse der Kultusministerkonferenz. Bildungsstandards im Fach Mathematik für den Primarbereich*. Beschluss vom 15.10.2004. München, Neuwied. https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2004/2004_10_15-Bildungsstandards-Mathe-Haupt.pdf

Liebers, K., Kanold, E. & Junger, R. (2019). Digitale Lernstandsanalysen in der inklusiven Grundschule? In S. Bartusch et al. (Hrsg.), *Lernprozesse begleiten*. (S.209-221). VS Verlag.

Magis, D., Duanli, Y., Davier, A. A. von & Yan, D. (2017). *Computerized Adaptive and Multistage Test-ing with R: Using Packages catR and mstR. Use R!*. Springer.

Preuss-Lausitz, U. (2016). Throughput instead of Input. Herausforderungen beim Wegfall der Fest-stellungsdiagnostik in den Förderbereichen Lernen, emotionale und soziale Entwicklung und Sprache. *Zeitschrift für Heilpädagogik*(5), 204–214.

Schurig, M., Jungjohann, J. & Gebhardt, M. (2021). Minimization of a Short Computer-Based Test in Reading.* Frontiers in Education, 6*, Artikel 684595. https://doi.org/10.3389/feduc.2021.684595

Smith, M. (2015). *The Usefulness of Alternative Testing Environments with Students with a Specific Learning Disability in mathematics*. Northwest Missouri State University Missouri.

Stanat, P., Schipolowski, S., Mahler, N., Weirich, S., Henschel, S. & Lorz, R. A. (Hrsg.). (2019). *IQB-Bildungstrend 2018: Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I im zweiten Ländervergleich*. Waxmann.

Stone, E. & Davey, T. (2011). Computer-Adaptive Testing for Students with Disabilities: A Review of the Literature. *ETS Research Report Series, 2011*(2), i-24.

Wechsler, D. (2017). *WISC-V: Wechsler Intelligence Scale for Children – Fifth Edition*. Deutsche Bearbeitung hrsg. Franz Petermann. Pearson.

Katharina Buchwald ist studierte Sonderpädagogin mit dem Schwerpunkt Sprache und der sonderpädagogischen Qualifikation Lernen. https://orcid.org/ 0000-0001-7570-7068

Nikola Ebenbeck ist studierte Sonderpädagogin mit dem Schwerpunkt Geistige Entwicklung. Sie ist Mitarbeiterin am Lehrstuhl für Lernbehindertenpädagogik einschließlich inklusiver Pädagogik an der Universität Regensburg. In Ihrem Promotionsprojekt entwickelt und evaluiert sie ein digitales und adaptives Lesescreening für die Grundschule, welches mit automatisierten Förderempfehlungen verbunden wird. https://orcid.org/ 0000-0002-4167-981X

Prof. Dr. Markus Gebhardt ist Sonderpädagoge und Lehrstuhlinhaber für Lernbehindertenpädagogik einschließlich inklusiver Pädagogik an der Universität Regensburg. https://orcid.org/ 0000-0002-9122-0556