

Für jedes Heft wird ein Repräsentant oder eine Repräsentantin eines heilpädagogischen Fachgebiets dazu eingeladen, einen persönlich gefärbten Kurzbeitrag zu voraussichtlichen Trends der kommenden Jahre in Forschung, Theorie und Praxis zu verfassen.

Tests werden besser

Reimer Kornmann

Pädagogische Hochschule Heidelberg

Psychometrische Tests erlauben meistens nur recht grobe Diagnosen und Klassifikationen und tragen wenig zur Klärung der praktisch wichtigen Frage bei, wie konkrete pädagogische oder therapeutische Situationen möglichst erfolgreich zu gestalten sind. Genau an dieser Fragestellung orientieren sich Vertreter des förderungsdiagnostischen Ansatzes. Sie gehen davon aus, dass Fragen nach Diagnosen und Klassifikationen umso seltener gestellt werden, je besser die pädagogischen Prozesse ablaufen: Unterricht, der allen Lernenden einer Klasse weitgehend gerecht wird, gibt seltener Anlass für diagnostische Untersuchungen, um Schwierigkeiten einzelner Schülerinnen oder Schüler diagnostisch abzuklären. Gleichwohl gründen sich erfolgreiche Unterrichtsprozesse auf Erkenntnisse, die als Ergebnisse förderungsdiagnostischer Arbeit gelten können. Daher ist es verständlich, wenn Vertreter der Förderdiagnostik den Entwicklungen der Psychometrie wenig Beachtung schenken mögen. Hinzu kommt, dass die Fülle wichtiger förderungsdiagnostischer Aufgaben in Praxis und Theorie die verfügbare Arbeitskraft vollauf binden kann. Auch war die lange Jahre vorherrschende Dominanz von methodisch und inhaltlich recht anspruchslos konzipierten psychometrischen Tests wenig dazu angetan, diesen diagnostischen Ansatz allzu ernst zu nehmen. Gerade in diesem Punkt zeichnen sich aber einige Entwicklungen ab, die auch für die Sonderpädagogik interessant und wichtig sein dürften.

Durch die internationalen Vergleichsstudien wird ein leistungsfähiges diagnostisches Messkonzept über die engeren Fachgrenzen hinaus bekannt

Die Konstruktionsprinzipien nahezu aller psychometrischen Tests, die für den Einsatz in Kindergärten und Schulen entwickelt worden sind, orientieren sich am Messkonzept der „Klassischen Testtheorie“. Dabei sind mehrere Aufgaben mittlerer Schwierigkeit zu bearbeiten, deren Lösung entweder als „richtig“ oder „falsch“ bewertet wird. Der Schwierigkeitsgrad einer jeden Testaufgabe wird bestimmt durch den prozentualen Anteil der Personen, die sie gelöst haben, an der Gesamtzahl derer, die sie zu bearbeiten hatten. Die so ermittelten Schwierigkeits-Indices hängen also direkt von der Leistungsstärke der dafür untersuchten Stichprobe ab. Die Summe der richtigen Lösungen der zu bearbeitenden Aufgaben ergibt den individuellen Rohwert für jede untersuchte Person. Anhand von Normtabellen, die sich auf bestimmte Verteilungsparameter einer als repräsentativ geltenden Eichstichprobe beziehen, wird die relative Position des individuellen Testwerts auf einer Standard-Skala metrisch bestimmt. Sowohl die Feststellung der Aufgabenschwierigkeiten als auch die Messung selbst beziehen sich auf eine bestimmte Population, welche die Beurteilungsmaßstäbe für die Einschätzung der gemessenen Testleistung liefert. Solche popu-

lationsabhängigen Ergebnisse begrenzen den Interpretationsrahmen und werfen weitere messtechnische Probleme auf, auf die hier nicht eingegangen werden kann. Unbefriedigend ist weiterhin der Tatbestand, dass lediglich die Anzahl der gelösten Aufgaben entscheidend für das Ergebnis ist, und zwar unabhängig davon, welche Items der Testskala jeweils gelöst wurden. Wegen der fehlenden Berücksichtigung möglicher „Muster“ der gelösten Testaufgaben wird also die Menge der erhaltenen Information nicht ausgeschöpft. Diese Mängel haben den dänischen Mathematiker Rasch (1960) veranlasst, ein alternatives Messmodell zu entwerfen. Es wurde trotz seiner wiederholt herausgestellten Vorteile (siehe schon Fischer 1974) nur selten für die Konstruktion psychometrischer Tests aufgegriffen. Der bekannteste dieser Tests, der vor dem Hintergrund sonderpädagogischer Fragen entstanden ist, ist sicher das „Adaptive Intelligenz Diagnostikum 2 (AID 2)“ von Kubitinger und Wurst (2001).

Bei den verschiedenen internationalen Vergleichsstudien zur schulischen Leistungsfähigkeit wurde nun aber auf das Messmodell von Rasch (1960) zurückgegriffen. Dadurch bestehen gute Aussichten, dass seine Vorteile einen größeren Bekanntheitsgrad erhalten und bei der Konstruktion neuer Tests berücksichtigt werden. Es werden dann Leistungsvergleiche zwischen verschiedenen Personen möglich, ohne dass diese die gleichen Aufgaben bearbeiten müssen. Zur Abschätzung der individuellen Leistungsstärke genügt es nämlich, die jeweils schwierigste Aufgabe, die die untersuchte Person noch lösen konnte, zu ermitteln. Neben den messtechnischen und testökonomischen Vorteilen ist der Ansatz auch aus pädagogischen Gründen zu begrüßen: Er lenkt den Blick sehr konsequent auf die Klasse solcher konkreter Aufgaben, die die Untersuchten bereits bewältigen und die somit auf deren verfügbare Kompetenzen schließen lassen. Zugleich richtet er die Perspektive auf die Gruppen von Aufgaben, welche die nächst höhere Kompetenzstufe repräsentieren.

Einige neuere Tests gründen sich auf Theorien zu gegenstandsspezifischem Kompetenzerwerb und ermöglichen förderungsorientierte Auswertungen

Die meisten der zahlreichen Tests für den vorschulischen und schulischen Bereich, die seit den 70er Jahren veröffentlicht wurden, zielten lediglich auf die Erfassung fächer- oder bereichsspezifischer Performanzkriterien. Ihre Aufgabenformate waren oft sehr einförmig und phantasielos gestaltet, und manche Aufgabenstellungen und Vorschläge zur Auswertung und Interpretation ließen Zweifel an der fachdidaktischen Kompetenz der Testautoren aufkommen. Hiervon heben sich einige der neueren Tests deutlich ab. Bei diesen fließt in die Testentwicklung neben psychometrischer Fachkompetenz auch fachdidaktisch und entwicklungstheoretisch begründetes Wissen über allgemeine und spezifische Erwerbsstrukturen ein, wie die nachfolgenden Beispiele zeigen:

- Die „Schultestbatterie zur Erfassung des Lernstandes in Mathematik, Lesen und Schreiben“ (SBL 1 und SBL 2) und die „Hamburger Schreibprobe“ (HSP) von May (1998) ermöglichen eine kompetenzorientierte Analyse der Fehlermöglichkeiten bei jedem einzelnen Testitem und lenken damit den Blick auf schon verfügbare Denk- und Lösungsstrategien bei der Auseinandersetzung mit der jeweiligen Aufgabe.
- Für jeweils gleiche Anforderungsbereiche finden sich unterschiedliche Aufgabenformate, die von verschiedenen Kindern unterschiedlich gut bearbeitet werden können. Dies ermöglicht intraindividuelle Vergleiche unter der Fragestellung, welche Aufgabenformate den Fähigkeiten der untersuchten Kinder entgegenkommen oder ihnen schwer fallen – so wiederum bei der SBL 1 und SBL 2.
- Bei dem „Diagnostikum: Basisfähigkeiten im Zahlenraum 0 bis 20 (DBZ 1)“ von Wagner und Born (1994) werden inhaltlich gleiche Aufgaben auf unterschiedlichen Ebe-

nen der Repräsentation vorgegeben (mündlich/schriftlich/bildlich), und die Lösungen können auf verschiedenen Ebenen der Repräsentation dargestellt werden (schriftlich-symbolisch/zeichnerisch-ikonisch/handelnd-konkret). So lässt sich für einzelne Gruppen von strukturell gleichen Aufgabebereichen ermitteln, auf welcher Niveaustufe der Repräsentation sie gegeben und mit Aussicht auf Erfolg bearbeitet werden können. Eine solche differenzierte Testung ermöglicht es, die „Zone der aktuellen Entwicklung“ einzugrenzen und Hypothesen über die „Zone der nächsten Entwicklung“ (Wygotski 1977) als Ansatzpunkt für eine individualisierende Förderung zu formulieren.

- In „Knuspels Leseaufgaben“ von Marx (2000) sind die Untertests so zusammengestellt, dass sie theoretisch begründete Erwerbssequenzen des zu erfassenden Gegenstandsbereichs abbilden und auf die Ansatzpunkte individueller Fördermaßnahmen verweisen.

Es werden pädagogisch sinnvolle Klassifikationen im Hinblick auf Indikationen für erfolgversprechende Fördermaßnahmen angestrebt

Als ein Meilenstein in der Entwicklung pädagogischer Tests, der in verschiedener Hinsicht Maßstäbe setzt, ist das „Bielefelder Screening zur Früherkennung von Lese-Rechtschreibschwierigkeiten (BISC)“ von Jansen, Mannhaupt, Marx und Skowronek (1999) anzusehen. Inzwischen dürfte der Test so bekannt sein, dass es genügt, seine besonderen Qualitätsmerkmale stichwortartig zu skizzieren:

- Seine Entwicklung beruht auf den Ergebnissen sehr gründlicher Literaturstudien zu der Frage, mittels welcher Merkmale sich bereits im Vorschulalter spätere Lese- und Rechtschreibschwierigkeiten prognostizieren lassen.

- Im Rahmen einer sorgfältig geplanten Längsschnittstudie zwischen dem zehnten Monat vor der Einschulung bis zum Ende des zweiten Grundschuljahres wurden die besten Prädiktoren zur Vorhersage vorab definierter Kriterien ermittelt. Dabei haben die Autoren zwischen spezifischen Kriteriumsleistungen (Lesen, Rechtschreiben) und unspezifischen Kriteriumsleistungen (Testintelligenz, Mathematik) ebenso wie zwischen spezifischen, unspezifischen und konfundierten Prädiktoren unterschieden. Identifiziert wurden also genau die Prädiktoren, die höher mit den spezifischen Kriteriumsleistungen korrelierten als mit den unspezifischen und die solche Merkmalsbereiche umfassen, die erstens noch nicht durch Lese- und Schreibfertigkeiten beeinflusst sind und zweitens Voraussetzungen prüfen, die überwiegend für den Erwerb der Schriftsprache bedeutsam sind und weniger den Erwerb anderer Kompetenzen begünstigen.
- Als Validitätskriterium wurde die Quote der richtigen Vorhersagen, gemessen am Kriterium des erfolgreichen Schriftspracherwerbs, angegeben. Die ermittelte Quote ist sehr günstig. Solche harten Bewährungsproben von Tests sind äußerst selten, sollten aber zum Standard der Testgütekriterien zählen.
- Die zehn Monate vor der Einschulung als auffällig diagnostizierten Kinder haben gute Chancen, durch gezielte Förderung mit dem Programm von Küspert und Schneider (2000) so weit gebracht zu werden, dass sie vier Monate vor der Einschulung nicht mehr auffällig sind und eine günstige Prognose für den Lernerfolg im schriftsprachlichen Bereich haben.

Besonders erfreulich ist, dass dieses gute Beispiel Schule macht. So informiert nun Krajewski (2003) über wichtige Ergebnisse eines Forschungsprogramms, das methodisch von den gleichen Prinzipien ausgeht wie die Arbeiten zur Entwicklung des BISC, inhaltlich aber den mathematischen Lernbereich betrifft.

Literatur

- Fischer, G. (1974): Einführung in die Theorie psychologischer Tests. Bern: Huber
- Jansen, H.; Mannhaupt, G.; Marx, H.; Skowronek, H. (1999): Bielefelder Screening zur Früherkennung von Lese-Rechtschreibschwierigkeiten (BISC). Göttingen: Hogrefe
- Kautter, H.; Storz, L.; Munz, W. (2000): Schultestbatterie zur Erfassung des Lernstandes in Mathematik, Lesen und Schreiben (SBL 1). Weinheim: Beltz
- Kautter, H.; Storz, L.; Munz, W. (2002): Schultestbatterie zur Erfassung des Lernstandes in Mathematik, Lesen und Schreiben (SBL 2). Weinheim: Beltz
- Krajewski, K. (2003): Vorhersage von Rechenschwäche in der Grundschule. Hamburg: Kovac
- Kubinger, K. D.; Wurst, E. (2001): Adaptives Intelligenz Diagnostikum 2 (AID 2). Weinheim: Beltz
- Marx, H. (2000): Knuspels Leseaufgaben. Göttingen: Hogrefe
- May, P. (1998): Hamburger Schreib-Probe (HSP). Hamburg: Verlag für pädagogische Medien
- Rasch, G. (1960): Probabilistic models for some intelligence and attainment tests. Copenhagen: Pædagogiske Institut
- Wagner, H.-J.; Born, C. (1994): Diagnostikum: Basisfähigkeiten im Zahlenraum 0 bis 20 (DBZ 1). Weinheim: Beltz
- Wygotski, L. S. (1977): Denken und Sprechen. Frankfurt/M.: Fischer

Prof. Dr. Reimer Kornmann
Pädagogische Hochschule Heidelberg
Institut für Sonderpädagogik
Postfach 104 240
D-69032 Heidelberg
E-Mail: kornmann@ph-heidelberg.de