

Trends und Innovationen pädagogisch-psychologischer Diagnostik

**Ulrich Trautwein
Marcus Hasselhorn
Claudia Mähler
Tobias Richter**
(Hrsg.)

Tests und Trends
der pädagogisch-
psychologischen
Diagnostik

 **hogrefe**

Trends und Innovationen pädagogisch-psychologischer Diagnostik

**Tests und Trends
der pädagogisch-psychologischen Diagnostik
Band 21**

Trends und Innovationen pädagogisch-psychologischer Diagnostik

Prof. Dr. Ulrich Trautwein, Prof. Dr. Marcus Hasselhorn, Prof. Dr. Claudia Mähler,
Prof. Dr. Tobias Richter

Die Reihe wird herausgegeben von:

Prof. Dr. Marcus Hasselhorn, Prof. Dr. Claudia Mähler,
Prof. Dr. Tobias Richter, Prof. Dr. Ulrich Trautwein

Ulrich Trautwein
Marcus Hasselhorn
Claudia Mähler
Tobias Richter
(Hrsg.)

Trends und Innovationen pädagogisch-psychologischer Diagnostik

Prof. Dr. Ulrich Trautwein. 1992–1999 Studium der Psychologie. 1999 Diplom in Psychologie. 2002 Promotion. 2005 Habilitation. Seit 2008 Universitätsprofessor für Empirische Bildungsforschung an der Universität Tübingen, seit 2015 geschäftsführender Direktor des Hector-Instituts für Empirische Bildungsforschung.

Prof. Dr. Marcus Hasselhorn. 1977–1983 Studium der Psychologie in Göttingen und Heidelberg. 1986 Promotion. Seit 2007 Professor für Psychologie an der Goethe-Universität Frankfurt und Direktor der Abteilung Bildung und Entwicklung am DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation in Frankfurt am Main.

Prof. Dr. Claudia Mähler. 1981–1987 Studium der Psychologie in Göttingen. 1994 Promotion. 2006 Habilitation. Seit 2008 Professorin für Pädagogische Psychologie und Diagnostik an der Universität Hildesheim. Leitung der Hochschulambulanz für Kinder und Jugendliche KiM – Kind im Mittelpunkt.

Prof. Dr. Tobias Richter. 1992–1998 Studium der Psychologie und Philosophie in Frankfurt am Main. 2003 Promotion, 2006 Habilitation. Seit 2016 Professor für Pädagogische Psychologie und Leiter des Lehrstuhls für Psychologie IV, seit 2019 auch Leiter der Begabungspsychologischen Beratungsstelle an der Julius-Maximilians-Universität Würzburg.

Wichtiger Hinweis: Der Verlag hat gemeinsam mit den Autor:innen bzw. den Herausgeber:innen große Mühe darauf verwandt, dass alle in diesem Buch enthaltenen Informationen (Programme, Verfahren, Mengen, Dosierungen, Applikationen, Internetlinks etc.) entsprechend dem Wissensstand bei Fertigstellung des Werkes abgedruckt oder in digitaler Form wiedergegeben wurden. Trotz sorgfältiger Manuskripterstellung und Korrektur des Satzes und der digitalen Produkte können Fehler nicht ganz ausgeschlossen werden. Autor:innen bzw. Herausgeber:innen und Verlag übernehmen infolgedessen keine Verantwortung und keine daraus folgende oder sonstige Haftung, die auf irgendeine Art aus der Benutzung der in dem Werk enthaltenen Informationen oder Teilen davon entsteht. Geschützte Warennamen (Warenzeichen) werden nicht besonders kenntlich gemacht. Aus dem Fehlen eines solchen Hinweises kann also nicht geschlossen werden, dass es sich um einen freien Warennamen handelt.

Alle Rechte, auch für Text- und Data-Mining (TDM), Training für künstliche Intelligenz (KI) und ähnliche Technologien, sind vorbehalten. All rights, including for text and data mining (TDM), Artificial Intelligence (AI) training, and similar technologies, are reserved.

Copyright-Hinweis:

Das E-Book einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlags unzulässig und strafbar.

Der Nutzer verpflichtet sich, die Urheberrechte anzuerkennen und einzuhalten.

Hogrefe Verlag GmbH & Co. KG
Merkelstraße 3
37085 Göttingen
Deutschland
Tel. +49 551 999 50 0
info@hogrefe.de
www.hogrefe.de

Satz: Beate Hautsch, Göttingen
Format: PDF

1. Auflage 2026

© 2026 Hogrefe Verlag GmbH & Co. KG, Göttingen

(E-Book-ISBN [PDF] 978-3-8409-3371-4; E-Book-ISBN [EPUB] 978-3-8444-3371-5)

ISBN 978-3-8017-3371-1

<https://doi.org/10.1026/03371-000>

Nutzungsbedingungen:

Durch den Erwerb erhalten Sie ein einfaches und nicht übertragbares Nutzungsrecht, das Sie zum privaten Gebrauch des E-Books und all der dazugehörigen Dateien berechtigt.

Der Inhalt dieses E-Books darf vorbehaltlich abweichender zwingender gesetzlicher Regeln weder inhaltlich noch redaktionell verändert werden. Insbesondere dürfen Urheberrechtsvermerke, Markenzeichen, digitale Wasserzeichen und andere Rechtsvorbehalte im abgerufenen Inhalt nicht entfernt werden.

Das E-Book darf anderen Personen nicht – auch nicht auszugsweise – zugänglich gemacht werden, insbesondere sind Weiterleitung, Verleih und Vermietung nicht gestattet.

Das entgeltliche oder unentgeltliche Einstellen des E-Books ins Internet oder in andere Netzwerke, der Weiterverkauf und/oder jede Art der Nutzung zu kommerziellen Zwecken sind nicht zulässig.

Das Anfertigen von Vervielfältigungen, das Ausdrucken oder Speichern auf anderen Wiedergabegeräten ist nur für den persönlichen Gebrauch gestattet. Dritten darf dadurch kein Zugang ermöglicht werden. Davon ausgenommen sind Materialien, die eindeutig als Vervielfältigungsvorlage vorgesehen sind (z.B. Fragebögen, Arbeitsmaterialien).

Die Übernahme des gesamten E-Books in eine eigene Print- und/oder Online-Publikation ist nicht gestattet. Die Inhalte des E-Books dürfen nur zu privaten Zwecken und nur auszugsweise kopiert werden.

Diese Bestimmungen gelten gegebenenfalls auch für zum E-Book gehörende Download-Materialien.

Die Inhalte dürfen nicht zur Entwicklung, zum Training und/oder zur Anreicherung von KI-Systemen, insbesondere von generativen KI-Systemen, verwendet werden. Das Verbot gilt nicht, soweit eine gesetzliche Ausnahme vorliegt.

Inhaltsverzeichnis

Kapitel 1

Pädagogisch-psychologische Diagnostik zwischen Tradition und Innovation

Ulrich Trautwein, Marcus Hasselhorn, Claudia Mähler & Tobias Richter 1

Kapitel 2

Feststellung von Lernfortschritten

Elmar Souvignier 17

Kapitel 3

Testfairness und Bias: Implikationen für die Migrationsgesellschaft

Ronja Runge 33

Kapitel 4

Validität und psychometrische Eigenschaften im Dynamischen Testen:

Analyse der Trainingsdaten des Graduated Prompt Ansatzes

Moritz Börnert-Ringleb, Jochanan Veerbeek & Jürgen Wilbert 49

Kapitel 5

Adaptives Testen

Andreas Frey, Christoph König & Aron Fink 65

Kapitel 6

Wie kontinuierliche Normierung die Datenerhebung erleichtert und gleichzeitig die Testgüte verbessert

Alexandra Lenhard & Wolfgang Lenhard 81

Kapitel 7

Digitale Verhaltensspurdaten zur Erfassung von Lernprozessen:

Bisherige Forschung und aktuelle Herausforderungen

Luise von Keyserlingk, Hannah Deininger & Florian Berens 101

Kapitel 8

Intelligente tutorielle Systeme

Cora Parrisius, Heiko Holz, Sarah Bez, Samuel Merk & Marc Philipp Janson 123

Kapitel 9

Ambulatory Assessment

Florian Schmiedek, Theresa Serratore, Lena M. Wieland & Friederike Blume 141

Kapitel 10
Automatisierte Beurteilung schriftlicher Argumentationen mithilfe künstlicher Intelligenz
Jennifer Meyer & Thorben Jansen 163

Kapitel 11
Gestaltung Computergestützter Diagnostik zur Theoriebasierten Nutzung von Interaktionsdaten: Das VIAE-Modell
Ulf Kroehne 183

Kapitel 12
Diagnostik auf Klassen- und Schulebene
Ulrich Trautwein & Benjamin Nagengast 217

Anhang (CC BY-ND 4.0)
Das Kompetenzstufenmodell des Schriftsprach- und Rechenerwerbs im Hilfssystem der Online-Plattform LONDI: Grundlage individualisierter Diagnostik- und Förderempfehlungen für Kinder mit Lernschwierigkeiten
Jelena Marković, Anne Fischbach, Kathleen Thomas, Katharina Galuschka, Kristina Moll, Gerd Schulte-Körne, Marcus Hasselhorn, Stefan Haberstroh & Susanne Volkmer 235

Die Autorinnen und Autoren dieses Bandes 255

Kapitel 1

Pädagogisch-psychologische Diagnostik zwischen Tradition und Innovation

Ulrich Trautwein, Marcus Hasselhorn, Claudia Mähler & Tobias Richter

1.1 Die Ausgangslage

Zwei Stühle, ein Tisch in einem ruhigen Raum, dazu zwei Personen, ein Kind und eine erwachsene Person, genauer gesagt ein:e Psycholog:in, auf dem Tisch ein Testheft und ein oder mehrere Papierbögen mit dem eingedruckten Logo eines Verlags sowie ein Stift und eine Uhr: Ein solches Setting wird in unzähligen pädagogisch-psychologischen Tests in den Anweisungen zur Durchführung beschrieben. Diese Szene erscheint schnell vor dem inneren Auge, wenn man an pädagogisch-psychologische Diagnostik denkt, und hat Eingang in die allgemeinen Vorstellungen von Diagnostik gefunden hat: Ähnliche Settings und Szenen werden auch in Spielfilmen und Romanen verwendet.

Denkbar wäre für die Zukunft jedoch auch folgendes Szenario: Lernende sitzen vor ihrem eigenen Computer und bearbeiten online Aufgaben, die ihnen zur Verfügung gestellt wurden. Wenn es sich um Leistungstests handelt, wird aus den Antworten, die sie geben, automatisch ein Fähigkeitsprofil erstellt, das dazu genutzt wird, weitere Aufgaben zur Verfügung zu stellen, die zum individuellen Fähigkeitsstand am besten passen. Zudem werden weitere Informationen, beispielsweise zum Zeitpunkt der Bearbeitung der Aufgaben sowie die Art der Bearbeitung, dazu verwendet, um ein motivationales Profil zu erstellen und automatische Angebote zur Lernunterstützung (z.B. Erinnerungen zur Bearbeitung von Hausaufgaben) zu generieren; wenn Lernende dies wollen, können sie sich während der Aufgabenbearbeitung auch von der im Rechner eingebauten Kamera aufzeichnen lassen und automatisch Hinweise darauf bekommen, bei welchen Lerninhalten ihre Aufmerksamkeit nachgelassen hat.

Unterschiedlicher, so scheint es, können zwei Szenarien kaum sein. Aber gemeinsam ist beiden, dass sie Formen der pädagogisch-psychologischen Diagnostik beschreiben, die letztlich die gleiche allgemeine Zielsetzung verfolgen: pädagogisch-psychologisch relevante Zustände, Prozesse und individuelle Personenmerkmale möglichst reliabel, valide und fair zu erfassen als Voraussetzung dafür, Lern- und Entwicklungsprozesse zu unterstützen bzw. zu verbessern. Was hat sich bewährt in der pädagogisch-psychologischen Diagnostik? Wo stößt sie an ihre Grenzen?

Welche Entwicklungen und Chancen sind zu erkennen? Wie ändern sich die Kontexte von pädagogisch-psychologischer Diagnostik? Welche Anforderungen wird die Zukunft mit sich bringen? Und was bedeutet dies ggf. auch für die Ausbildung von Diagnostiker:innen und den Kontexten, in denen pädagogisch-psychologische Diagnostik stattfindet? In diesem Beitrag soll es darum gehen, wichtige Trends und Innovationen herauszuarbeiten, einzuordnen und kritisch zu reflektieren.

1.2 Status Quo, Trends und Innovationen

Die pädagogisch-psychologische Diagnostik hat im Laufe der vergangenen Jahrzehnte beeindruckende Fortschritte gemacht (vgl. Spinath & Brünken, 2016). Inzwischen liegt eine Vielzahl von Tests für eine große Bandbreite von diagnostisch relevanten Phänomenen vor – allein bei der Hogrefe Testzentrale sind mehrere hundert Verfahren für Kinder und Jugendliche verzeichnet. Eine steigende Anzahl dieser Tests wurde auf der Basis hoher Qualitätsstandards entwickelt: Sie basieren auf empirisch bewährten Lern- und Entwicklungstheorien, wurden in mehrschrittigen Prozessen entwickelt, anhand von strengen psychometrischen Kriterien validiert und in vielen Fällen auch aufwändig normiert. Sie erfassen mit teils erstaunlicher Präzision und hoher prognostischer Validität zentrale Aspekte des Denkens, Fühlens und Handelns von Kindern und Jugendlichen und erlauben in vielen Fällen eine punktgenaue Intervention durch Lehrkräfte bzw. weiteres entsprechend geschultes Personal. Für Wissenschaft und Praxis steht somit heute eine breite Palette von Verfahren zur Verfügung, mit deren Hilfe ein tieferes Verständnis für den Lern- und Entwicklungsstand von Schülerinnen und Schülern gewonnen werden kann.

Ungeachtet dieser positiven Entwicklungen lassen sich auch eine Reihe von Herausforderungen, Defiziten und bisher nicht ausreichend genutzten Chancen festhalten. Das wahrscheinlich augenfälligste Problem: Trotz all der Fortschritte in der Qualität pädagogisch-psychologischer Verfahren kommen diese weit seltener zur Anwendung als es sinnvoll wäre und man vielleicht vermuten würde. Dazu trägt in Deutschland ganz wesentlich ein Mangel an Schulpsycholog:innen sowie an Kinder- und Jugendlichen-Psychotherapeut:innen bei. So sind in allen Bundesländern die Schulpsycholog:innen jeweils für viele tausend Schüler:innen zuständig, so dass Schüler:innen oft monatelang auf einen Termin warten müssen (vgl. https://www.bdp-schulpsychologie.de/aktuell/2024/2024_versorgungszahlen.pdf). Gleichzeitig lässt sich beobachten, dass auch an den Universitäten die pädagogisch-psychologische Diagnostik und die Ausbildung von Schulpsycholog:innen eine eher untergeordnete Rolle spielt; nur an wenigen Standorten gibt es entsprechende Spezialisierungen. Zudem fehlt es an einer überzeugenden Strategie zur besseren Ausbildung und Einbindung von Lehrkräften bei pädagogisch-psychologischer Diagnostik.

Neben personellen Engpässen tragen jedoch auch weitere Faktoren dazu bei, dass man bei der pädagogisch-psychologischen Diagnostik eine gewisse Stagnation beobachten kann. So sind standardisierte Verfahren der pädagogisch-psychologischen Diagnostik oftmals sehr zeitaufwändig sowie in der Durchführung und im Erwerb teuer. Inwieweit die in vielen Tests berichteten Normen aktuell sind, lässt sich teilweise schwer abschätzen – insbesondere bei schulnahen Leistungsbereichen wurden über die vergangenen Jahre in Monitoring-Studien wie dem IQB-Bildungstrend (vgl. Stanat et al., 2023) substanzielle Leistungsunterschiede zwischen einzelnen Kohorten beobachtet, die sich mit einer gewissen Wahrscheinlichkeit auch in veränderte Normen bei pädagogisch-psychologischen Tests niederschlagen würden. Auch lässt sich u. a. beobachten, dass der Großteil der heutzutage verwendeten Verfahren auf der klassischen Testtheorie beruht, die im Vergleich beispielsweise mit der Item-Response-Theorie eine Reihe von Nachteilen bei der Testökonomie sowie der Erfassung von Lernverläufen hat (siehe bereits Rost, 1996; vgl. Frey, König & Fink, 2026/in diesem Band). Zudem gibt es zahlreiche „blinde Flecken“ im vorhandenen Testangebot – so erhalten beispielsweise Schüler:innen, deren Schulleistung unauffällig ist, in der Regel keine psychometrisch abgesicherte Analyse zu den Stärken, Schwächen und Verbesserungspotenzialen bei ihrem Lernverhalten, und auch die Diagnostik im Hochbegabungsbereich fristet, vielleicht mit Ausnahme der Intelligenzdiagnostik, eher ein Schattendasein.

Allerdings würde ein falsches Bild entstehen, wenn nicht neben den Stagnationsphänomenen auch die Innovationsdynamiken beschrieben würden, die zu beobachten sind, auch wenn sie in der diagnostischen Praxis bzw. in populären Lehrbüchern (z. B. Ingenkamp & Lissmann, 2008; Schmidt-Atzert et al., 2021; Spinath & Brünken, 2016) bislang noch keine Hauptrolle spielen. Wir nennen im Folgenden sieben Bereiche, die ein hohes Potenzial besitzen, die pädagogisch-psychologische Diagnostik zukünftig zu prägen. Einige dieser Entwicklungen sind nicht per se „neu“, aber teilweise haben Rahmenbedingungen wie eine nur langsam vorankommende Digitalisierung ihre Implementation in die Praxis behindert. Viele der in den Kapiteln des vorliegenden Bands vertieft beschriebenen Innovationen beruhen zumindest teilweise darauf, dass digitale Technologien zunehmend breiter zur Verfügung stehen. Es dürfte nur eine Frage der Zeit sein, bis sich das entsprechende Innovationspotenzial in einer beschleunigten Weiterentwicklung des Feldes niederschlägt.

1.2.1 Analyseebenen der pädagogisch-psychologischen Diagnostik

Der traditionelle Schwerpunkt pädagogisch-psychologischer Diagnostik liegt auf der Individualdiagnostik, wobei auf der Basis eines bestimmten Anlasses bzw. Auftrags von bzw. über eine Person zu einem bestimmten Zeitpunkt mehr oder weni-

ger viele Daten gesammelt werden und diese von einer psychologisch geschulten Person in einer systematischen Auswertung (ggf. in einem Gutachten) verdichtet werden. Ohne Frage wird diese Art pädagogisch-psychologischer Diagnostik auch in Zukunft ihre Berechtigung haben und eine wichtige Rolle spielen.

Daneben wurden in den vergangenen Jahren jedoch zunehmend weitere Analyseebenen für die pädagogisch-psychologische Diagnostik erschlossen. Hierbei ist zum einen die Erfassung von „Zuständen“ (anstatt von Personenmerkmalen) zu nennen, die oftmals als wiederholte Messung von Denken, Fühlen und Verhalten einzelner Personen zur Anwendung kommt und die in der klinischen Psychologie sowie in der Forschung gut etabliert ist (vgl. Ebner-Priemer & Trull, 2009), aber zunehmend auch im pädagogisch-psychologischen Bereich Fuß fasst (z. B. Bugl et al., 2015) und in Schmiedek, Serratore, Wieland & Blume (2026/in diesem Band) beschrieben wird. Der Grundgedanke des Ambulatory Assessments besteht darin, psychologische und physiologische Daten von Personen im natürlichen Lebensumfeld und in Echtzeit (und damit ohne retrospektive Verzerrungen) zu erfassen, wobei die Erfassung häufig in Form von wiederholten Momentaufnahmen (Experience Sampling) bzw. in Form von kontinuierlichen Messungen (z. B. unter Nutzung von Sensoren) erfolgt. Neben der Möglichkeit, die Datenpunkte über die Zeit zu aggregieren, erlaubt es das Ambulatory Assessment auch, bestimmte zeitliche Verläufe bzw. Auftretenshäufigkeiten zu identifizieren.

Zum anderen widmet sich die pädagogisch-psychologische Diagnostik aber auch vermehrt der Erfassung der Qualität von Lernumgebungen, wie sie beispielsweise Klassen oder Schulen darstellen (vgl. Rauer & Schuck, 2021, für ein bereits bewährtes Verfahren). Solche Einschätzungen können Ausgangspunkt wirksamer Interventionen zur Verbesserung des Lernens vieler Schüler:innen sein und haben entsprechend ein hohes Potenzial. Wie von Trautwein und Nagengast (2026/in diesem Band) ausgeführt wird, stellen sich bei dieser Form pädagogisch-psychologischer Diagnostik jedoch auch eine Reihe von methodischen Herausforderungen, deren Überwindung die Entwicklung entsprechender Fachkompetenz voraussetzt.

1.2.2 Interaktion und Adaptivität im diagnostischen Prozess

Eine weitere innovative Entwicklung bezieht sich darauf, dass in einer Klasse von Verfahren der diagnostische Prozess interaktiv bzw. adaptiv angelegt ist. In Abgrenzung zu klassischen Verfahren, bei denen bereits vor dem diagnostischen Akt entschieden ist, genau welche Aufgaben zur Diagnostik eingesetzt werden, „reagieren“ adaptive Instrumente in der Aufgabenauswahl auf die Antworten der Kinder bzw. Jugendlichen.

Bereits seit einigen Jahrzehnten ist die Idee des Adaptiven Testens (siehe Frey et al., 2026/in diesem Band) zumindest theoretisch gut etabliert. Die Idee hinter

dem Ansatz ist plausibel und bestechend einfach: Items, die für eine Testteilnehmerin bzw. einen Testteilnehmer viel zu schwierig oder viel zu leicht sind, werden erst gar nicht vorgegeben; vielmehr nutzt man die Antworten auf die bereits bearbeiteten Aufgaben, um eine zunehmend genauere Einschätzung der Leistungsstärke vorzunehmen und nur noch Aufgaben aus diesem Schwierigkeitsbereich vorzulegen. Adaptives Testen verspricht eine hohe Testökonomie bei gleichzeitiger Abdeckung eines großen Itemuniversums.

Ein weitere Innovation, bei der im diagnostischen Prozess auf die Antworten der Testteilnehmer:innen eingegangen wird, ist das dynamische Testen (vgl. Börnert-Ringleb et al., 2026/in diesem Band). Das dynamische Testen ist ein diagnostisches Verfahren, das letztlich nicht primär den aktuellen Leistungsstand einer Person erfasst, sondern insbesondere an ihrem Potenzial und ihrer Fähigkeit, durch gezielte Anleitung und Unterstützung zu lernen und sich zu verbessern, interessiert ist. Trotz des einleuchtenden Grundgedankens des dynamischen Testens gibt es bislang wenige Verfahren, die es in hoher Qualität umsetzen.

1.2.3 Der Fokus auf Förderung von Lernprozessen

Während beim dynamischen Testen die Diagnostik im Vordergrund steht, stellen eine Reihe weiterer Instrumente das Lernen in den Mittelpunkt, wobei aber aktiv diagnostische Daten der Lernenden für die Gestaltung bzw. Verbesserung von Lernprozessen genutzt werden.

In diesem Zusammenhang sind insbesondere Instrumente zu nennen, die wiederholt den Lernfortschritt messen und den Lernenden adaptiv Materialien zur Verfügung stellen, mit denen sie festgestellte Lerndefizite überwinden können. Im Beitrag von Souvignier (2026/in diesem Band) wird das entsprechende Vorgehen genauer beschrieben, und es werden Möglichkeiten aufgezeigt, wie Diagnostik mit Intervention verknüpft werden kann.

Intelligente tutorielle Systeme (vgl. Parrisius et al., 2026/in diesem Band) verfolgen im Grunde eine ähnliche Strategie, aber bei ihnen ist der Gedanke von Lernfortschrittsdiagnose und Intervention noch enger verwoben: Während Lernende mit einer digitalen Lernumgebung lernen und Lern- und Übungsaufgaben bearbeiten, geben die Systeme punktgenaues Feedback, beurteilen die Leistungsfähigkeit der Lernenden und wählen automatisch die Art und den Schwierigkeitsgrad der nachfolgenden Aufgaben aus.

Aus pädagogisch-psychologischer Sicht sind auch Learning-Management-Systeme von hohem Interesse, da sie eine große Menge von diagnostisch relevanten Informationen in authentischen Lernsituationen sammeln (vgl. von Keyserlingk et al., 2026/in diesem Band). Das entsprechende diagnostische Potenzial ist bislang weitgehend ungenutzt, was auch daran liegt, dass eine valide Interpretation

von Interaktionsdaten mit digitalen Endgeräten teilweise eine enorme Herausforderung darstellt (vgl. Kroehne, 2026/in diesem Band).

1.2.4 Nutzung von Daten aus authentischen Lernkontexten für die pädagogisch-psychologische Diagnostik

Auffällig – und teilweise eng verbunden mit anderen, bereits beschriebenen Innovationssträngen – ist der Trend, für diagnostische Prozesse zunehmend Daten zu verwenden, die bereits vorhanden sind bzw. ohnehin anfallen. Daten aus Intelligenten tutoriellen Systemen (Parrisius et al., 2026/in in diesem Band) oder Learning-Management-Systemen (von Keyserlingk et al., 2026/in diesem Band) sind Paradebeispiele für diesen Trend. Aber auch Überlegungen, Daten, die im Rahmen von Monitoring-Studien im Bildungssystem anfallen, als Screening-Daten für individuelle Entwicklungen zu verwenden, gehören in diesen Bereich, genauso wie die Verwendung von Daten aus Schulleistungsstudien für Statusgespräche mit Schulleitungen bzw. für die Identifikation von Schulen für besondere Fördermaßnahmen (Trautwein & Nagengast, 2026/in diesem Band). Die „Zweitnutzung“ von Daten reduziert den Aufwand für alle Beteiligten und erhöht in diesem Sinne auch die Akzeptanz diagnostischer Prozesse, birgt jedoch, wie unten noch beschrieben werden wird, auch hohe Anforderungen an die Transparenz des Vorgehens und kann datenschutzrechtliche Fragen aufwerfen (siehe auch Börnert-Ringleb et al., 2026/in diesem Band).

1.2.5 Implementation geeigneter statistischer Verfahren

Prägend für die pädagogisch-psychologische Diagnostik war und ist die Implementation geeigneter statistischer Verfahren bei der Konzeption von Tests und der Feststellung der Testleistung der Testteilnehmer:innen. Seit den 1980er Jahren erwarteten viele Experten die Entwicklung und den zunehmenden Einsatz von diagnostischen Verfahren, die auf der probabilistischen Testtheorie aufbauen. Obwohl dieser Ansatz im Rahmen des Adaptiven Testens (vgl. Frey et al., 2026/in diesem Band) der Klassischen Testtheorie deutlich überlegen ist, ist er bis heute bei den vorliegenden Instrumenten der Diagnostik nur vereinzelt auszumachen. Allerdings basieren viele Lernfortschrittsanalysen (vgl. Souvignier, 2026/in diesem Band) und Intelligente tutorielle Systeme (vgl. Parrisius et al., 2026/in diesem Band) auf Ansätzen der probabilistischen Testtheorie.

Die sorgfältige Normierung charakterisiert einen guten Test; gleichzeitig ist es immer schwieriger geworden, die für die Normierung notwendigen großen Stichproben zu rekrutieren. Hier setzen Verfahren (vgl. Lenhard & Lenhard, 2026/in diesem Band) an, die die Normierung klassischer pädagogisch-psychologischer

Tests effizienter machen. Aber auch große Schulleistungsstudien (vgl. Trautwein & Nagengast, 2026/in diesem Band) sowie Intelligente tutorielle Systeme (vgl. Parrisius et al., 2026/in diesem Band) können Vergleichsdaten liefern, die im Rahmen diagnostischer Prozesse von Relevanz sind.

1.2.6 Künstliche Intelligenz

Der vielleicht bedeutsamste Trend, dessen Folgen noch gar nicht abschätzbar sind, dürfte der absehbare Siegeszug der Künstlichen Intelligenz sein. Eine Reihe von Entwicklungen dürfte direkt an die Verwendbarkeit Künstlicher Intelligenz im diagnostischen Prozess gekoppelt werden. So darf man beispielsweise davon ausgehen, dass in den kommenden Jahren eine Vielzahl von verbesserten Intelligenz tutoriellen Systemen auf den Markt kommen werden. Die Verwendung künstlicher Intelligenz scheint darüber hinaus enorme Möglichkeiten bei der Analyse eher komplexer Stimuli (z. B. längere Texte oder Verhalten in komplexen Situationen) aufzuweisen, die u. a. aufgrund des hohen Aufwands bei Auswertungen in diagnostischen Prozessen oftmals nicht verwendet werden. In Meyer und Jansen (2026/in diesem Band) sowie in einem Abschnitt in Trautwein und Nagengast (2026/in diesem Band) finden sich Beispiele für die Verwendung Künstlicher Intelligenz in diagnostischen Prozessen.

1.2.7 Fairness

Die Bedeutung von Bildung für den gesamten Lebensweg (vgl. Autorengruppe Bildungsberichterstattung, 2018) ist gut dokumentiert, und so kann es nicht verwundern, dass es auch im Bildungsbereich „Verteilungskämpfe“ gibt, bei denen diagnostische Entscheidungen eine wichtige Rolle spielen. Beispielsweise möchten Eltern eine bestimmte Übertrittsempfehlung für ihre Kinder erhalten, die Anwendung eines Tests hat einen Effekt darauf, ob ein Kind einen Nachteilsausgleich bekommt, oder die Begabung eines Kindes mit Flucht- und Zuwanderungshintergrund wird mithilfe eines Potenzialtests erfasst. Es ist in solchen Situationen von großer Bedeutung, dass diagnostische Verfahren eine hohe Fairness aufweisen, und der Aspekt der Testfairness hat zunehmend an Bedeutung gewonnen (siehe American Educational Research Association et al., 2014). Hierbei ist es wichtig zu verstehen (vgl. Runge, 2026/in diesem Band), dass die Testfairness ein breites Konstrukt ist und u. a. die Auswahl von Tests, die Entwicklung von Tests sowie die Identifikation von geeigneten Items umfasst. Inzwischen ist es im Übrigen bekannt, dass Überlegungen zur Testfairness auch beim Einsatz von Künstlicher Intelligenz im diagnostischen Prozess (vgl. Meyer & Jansen, 2026/in diesem Band) eine bedeutende Rolle spielen müssen (siehe Kotek et al., 2023).

1.3 Die Rolle der Testgütekriterien

Die Entwicklung und Berücksichtigung der Hauptgütekriterien Objektivität, Reliabilität und Validität sowie einer Reihe von Nebengütekriterien psychologischer Tests war und ist ganz wesentlich für die Akzeptanz und den Erfolg pädagogisch-psychologischer Diagnostik verantwortlich. Es hat sich bewährt, die Testgütekriterien in den Mittelpunkt der Bewertung von psychologischen Testverfahren zu stellen, wie dies auch vom Diagnostik- und Testkuratorium der Deutschen Gesellschaft für Psychologie sowie der Föderation Deutscher Psychologinnenvereinigungen (vgl. Testkuratorium, 2024) gemacht wird. Nicht zuletzt wird damit auch die federführende Rolle der Psychologie bei der Erstellung und Durchführung von Tests betont.

Gleichzeitig besteht immer auch die Gefahr, dass ein Festhalten an bislang bewährten Praktiken zu einem Innovationshemmnis wird, weil mögliche Optimierungen sowie neue Anwendungsfelder übersehen werden. Betrachtet man die oben beschriebenen Innovationen, so wird offensichtlich, dass auch eine kritische Reflexion des Umgangs mit Testgütekriterien notwendig wird, wenn alle neuen Möglichkeiten genutzt werden sollen. Dies soll an wenigen Beispielen illustriert werden.

So wirft beispielsweise die Sammlung von diagnostisch relevanten Daten über das Lernverhalten von Schüler:innen bzw. Studierenden mithilfe von Learning-Management-Systemen (siehe von Keyserlingk et al., 2026/in diesem Band) im Sinne der Objektivität Fragen auf, da die Lernenden typischerweise in sehr unterschiedlichen Settings ihre Aufgaben bearbeiten. Eine Durchführungsobjektivität im klassischen Sinne ist hier nicht gewährleistet, da die „Testsituation“ nicht für alle Lernenden identisch („standardisiert“) ist bzw. sie ihre Lernumgebung sogar selbst gestalten. Aber gerade diese Offenheit der Situation erlaubt es, bestimmte Konstrukte wie „Selbstregulationsfähigkeiten“ und „Prokrastination“ ökologisch valide zu erfassen.

Ein weiteres interessantes Beispiel sind Auswertungen von Daten mithilfe von Large Language Modellen (LLM; vgl. Meyer & Jansen, 2026/in diesem Band) wie beispielsweise ChatGPT. Diese werden u. a. dafür eingesetzt, Aufsätze von Schüler:innen zu bewerten bzw. Unterrichtsqualität zu analysieren. Wie inzwischen verschiedentlich gezeigt werden konnte (vgl. Meyer & Jansen, 2026/in diesem Band; Trautwein & Nagengast, 2026/in diesem Band), erzielen LLMs bei der Auswertung von komplexen Datensätzen teilweise psychometrische Kennwerte, die denen von menschlichen Auswerter:innen ebenbürtig sind. Allerdings besteht hier bei den derzeit gängigen Large Language Modellen wie ChatGPT die Herausforderung, dass völlig unklar ist, wie es mit der Auswertungsobjektivität bzw. der Reproduzierbarkeit von Testergebnissen über längere Zeit bestellt ist, denn diese LLMs „lernen“ über die Zeit selbst dazu, so dass sich die resultierenden Werte ändern und auch die ursprünglichen Werte nicht mehr reproduziert werden kön-

nen; einen möglichen Ausweg bieten hier nicht-lernende Modelle, die auf einem Standardkorpus beruhen.

Auch in Hinblick auf die Validität zeigen sich teilweise neue Herausforderungen. Man denke nur beispielsweise an Intelligente tutorielle Systeme (vgl. Parrius et al., 2026/in diesem Band) bzw. Instrumente zur Verlaufsdiagnostik (vgl. Souvignier, 2026/in diesem Band), bei denen lernbegleitend eine Diagnostik der Leistungen der Lernenden vorgenommen wird, um darauf aufbauend die nächsten Aufgaben zur Verfügung zu stellen mit dem Ziel, den Lernerfolg zu optimieren. Die Validität der Ermittlung des Leistungsstands sollte in solchen Systemen sinnvollerweise anhand der bei den Lernenden erzielten Leistungssteigerung erfasst werden – und damit würde der Validierungsprozess letztlich eine Überprüfung der Effektivität des Instruments beinhalten, beispielsweise im Rahmen eines randomisierten Feldexperiments (vgl. Trautwein et al., 2022). Eine Validitätsprüfung für die lernbegleitende Diagnostik im Rahmen eines Intelligenzen tutoriellen Systems könnte beispielsweise erfolgen, indem man die Lernwirksamkeit und -effizienz mit der eines Lernprogramms vergleicht, das dieselben Inhalte in nicht-adaptiver Form (und ohne Diagnostik) enthält.

Auch bei den Nebengütekriterien ergeben sich eine Reihe von Fragen, die einer neuen Aufmerksamkeit bedürfen und hier nur stichpunktartig genannt werden sollen. So stellt sich beispielsweise die Frage, inwieweit die kontinuierlich im Rahmen von Intelligenzen tutoriellen Systemen gesammelten Daten zur Normierung von Testinstrumenten bzw. den erfassten diagnostischen Größen genutzt werden können. Bei der Verwendung von Künstlicher Intelligenz im Rahmen von Diagnostik ergeben sich unmittelbar Fragen nach der Testfairness sowie nach der Zumutbarkeit bzw. Akzeptanz. In Hinblick auf die Ökonomie/Sparsamkeit eines Verfahrens werden neue Gesichtspunkte relevant, sobald ein Verfahren sowohl diagnostische Anteile enthält, als auch selbst bereits eine Intervention darstellt.

Mit den Testgütekriterien besitzt die Pädagogische Psychologie ein zentrales Werkzeug, um diagnostische Instrumente zu verbessern und zu beurteilen. Auch in Zukunft wird es unverzichtbar sein, die psychometrische Qualität der Verfahren systematisch und kritisch zu prüfen. Für eine breite Nutzung pädagogisch-psychologischen Wissens dürfte es – wie nachfolgend argumentiert wird – wesentlich darauf ankommen, dieses Wissen in Zukunft noch systematischer direkt ins Schulsystem einfließen zu lassen.

1.4 Institutionelle Rahmenbedingungen: Die Schule der Zukunft

Das Schulsystem in Deutschland gehört zu den großen Anwendungsfeldern der pädagogisch-psychologischen Diagnostik. Dieses ist derzeit großen Veränderun-

gen und Herausforderungen ausgesetzt. Um nur einige wenige zu nennen: In internationalen Schulleistungsstudien findet sich seit rund einem Jahrzehnt eine deutlich negative Leistungsentwicklung (vgl. OECD, 2023). Die Zusammensetzung der Schülerschaft hat sich innerhalb von zwei Jahrzehnten stark geändert, wobei u. a. der starke Anteil von Schüler:innen, die einen Flucht- bzw. Zuwanderungshintergrund aufweisen (Stanat et al., 2023), herauszustellen ist. Ganztagsschulen haben in ihrer Bedeutung zugenommen, und ab dem Schuljahr 2026/27 besteht bundesweit ein Recht auf Ganztagsförderung in der Grundschule (vgl. <https://www.recht-auf-ganztag.de/gb/politik>); Schulen kommt damit auch eine größere Verantwortung für die breitere Entwicklung der Kinder zu. Und im Rahmen der Digitalpakete von Bund und Ländern (vgl. <https://www.digitalpaktschule.de/index.html>) verfügen immer mehr Schulen über eine digitale Ausstattung sowie über schnellen Internetzugang, weshalb technologie-inspirierte Innovationen in Zukunft eine deutlich größere Rolle spielen könnten.

Eine Konsequenz der diskutierten Veränderungen und Herausforderungen ist ein Bedarf nach qualitativ hochwertiger pädagogisch-psychologischer Diagnostik. Zugleich bieten sich Chancen für eine breitere Nutzung pädagogisch-psychologischer Expertise. Ein wichtiger Trend ist hierbei die vermehrte Nutzung von standardisierter Diagnostik in der Schule. Zu nennen sind hierbei zunächst die Instrumente zur Erfassung von Schulleistung im Rahmen von Monitoringstudien, Lernstandserhebungen sowie Potenzialanalysen. In Deutschland erfolgt die Entwicklung dieser Instrumente dabei primär am Institut zur Qualitätsentwicklung im Bildungssystem (IQB) in Berlin sowie an den entsprechenden Landesinstituten der Bundesländer. Es ist zu erwarten, dass die Bedeutung dieser Verfahren in der Zukunft eher noch weiter zunehmen wird. Aus pädagogisch-psychologischer Sicht ist es vordringlich, dass bei der Entwicklung und beim Einsatz standardisierter Verfahren die etabliert hohen Ansprüche an die psychometrische Qualität gestellt werden. Diese Forderung scheint selbstverständlich zu sein, ist aber insbesondere bei vielen der landesspezifischen Verfahren bislang nur teilweise gegeben.

Zu erwarten ist als möglicher zweiter Trend zudem, dass im Bildungssystem bestimmte Herausforderungen deutlicher benannt werden und eine Nachfrage nach „geprüften“ bzw. „empfohlenen“ Verfahren entsteht, die bei der Bewältigung der entsprechenden Herausforderungen helfen können. Dies betrifft beispielsweise den Sprachstand der Kinder vor Einschulung bzw. nach dem Eintritt von Kindern ins deutsche Schulsystem, der in vielen Bundesländern zunehmend systematischer erfasst wird (Autor:innengruppe Bildungsberichterstattung, 2024), obschon es noch an einer Einheitlichkeit mangelt und kritische Rückfragen zur Qualität der Verfahren angebracht sind (vgl. Neugebauer & Becker-Mroczek, 2013). Auch im Bereich von Lernstörungen finden sich Entwicklungen, die zu einer stärkeren Systematisierung von Verfahren, die teilweise auch von Lehrkräften an Schulen administriert werden können, sowie zu einer stärkeren Nutzung entsprechender Instrumente führen könnten, wie sich anhand der Online-Plattform londi