



POSTGRADUALE FERNSTUDIENGÄNGE  
SCIENCE & ENGINEERING



STUDIENBRIEF AM01\_01 ➤ PSYCHOLOGIE KINDLICHER LERN- UND ENTWICKLUNGSAUFFÄLLIGKEITEN

# PÄDAGOGISCH-PSYCHOLOGISCHE DIAGNOSTIK

AUTOR ➤

PROF. EM. DR. KARL JOSEF KLAUER

ÜBERARBEITUNG UND ERGÄNZUNG

PROF. DR. MARIA KLATTE



TECHNISCHE UNIVERSITÄT  
KAISERSLAUTERN

DISTANCE AND INDEPENDENT  
STUDIES CENTER



Leseprobe

Das Werk ist urheberrechtlich geschützt. Die dadurch begründeten Rechte, insbesondere das Recht der Vervielfältigung und Verbreitung sowie der Übersetzung und des Nachdrucks, bleiben, auch bei nur auszugsweiser Verwertung, vorbehalten. Kein Teil des Werkes darf außerhalb der im Urheberrecht geregelten Erlaubnisse in irgendeiner Form (Druck, Fotokopie, Mikrofilm oder ein anderes Verfahren) ohne schriftliche Genehmigung der Technischen Universität Kaiserslautern reproduziert oder unter Verwendung elektronischer Systeme verarbeitet, vervielfältigt oder verbreitet werden.

Titelbild: Julia A. Heinrich

2. Auflage 2016

# Inhaltsverzeichnis

<b>Glossar</b>	<b>iii</b>
<b>Über die Autoren</b>	<b>v</b>
<b>Literaturverzeichnis</b>	<b>vii</b>
<b>Lernziele des Lehrbriefes</b>	<b>xv</b>
<b>1 Einführung in die pädagogisch–psychologische Diagnostik</b>	<b>1</b>
1.1 Lernziele des Kapitels	1
1.2 Gegenstand und Aufgaben der pädagogisch–psychologischen Diagnostik	1
1.3 Methoden der pädagogisch–psychologischen Diagnostik	4
1.3.1 Verhaltensbeobachtung	5
1.3.2 Gespräch und Interview	6
1.3.3 Testverfahren	6
1.4 Bezugsnormorientierung bei Lehrkräften	11
1.5 Diagnostische Kompetenz und Urteilsfehler	13
1.6 Zur Geschichte der pädagogisch–psychologischen Diagnostik	16
1.7 Übungsaufgaben	18
<b>2 Testtheorie</b>	<b>19</b>
2.1 Lernziele des Kapitels	19
2.2 Klassische und probabilistische Testtheorie	19
2.2.1 Die zentralen Annahmen der klassischen Testtheorie	20
2.2.2 Folgerungen zur Reliabilität von Tests	22
2.3 Probabilistische Testtheorie	25
2.4 Übungsaufgaben	29
<b>3 Aufgabenformen</b>	<b>31</b>
3.1 Lernziele des Kapitels	31
3.2 Einführung	31
3.3 Freie Aufgabenformen	33
3.4 Gebundene Aufgabenformen	35
3.5 Itemformen in Einstellungs– und Persönlichkeitstests	41
3.6 Bewertung von Testaufgaben	44
3.7 Übungsaufgaben	44
<b>4 Aufgabenanalyse</b>	<b>47</b>
4.1 Lernziele des Kapitels	47
4.2 Allgemeines Vorgehen bei der Aufgabenanalyse	47
4.3 Schwierigkeitsindex	48
4.4 Trennschärfe	50
4.5 Selektionsindex	53
4.6 Änderungssensitivität	54
4.7 Distraktorenanalyse	55

4.8	Interkorrelation der Items	57
4.9	Aufgabenselektion	58
4.10	Übungsaufgaben	59
<b>5</b>	<b>Testanalyse</b>	<b>61</b>
5.1	Lernziele des Kapitels	61
5.2	Verteilung der Testscores und Testnormierung	61
5.3	Gütekriterien von Tests: Ein Überblick	66
5.3.1	Objektivität	67
5.3.2	Reliabilität	69
5.3.2.1	Verfahren zur Schätzung der Reliabilität	70
5.3.2.2	Retest–Reliabilität	70
5.3.2.3	Paralleltest–Reliabilität	72
5.3.2.4	Testhalbierungs–Reliabilität	73
5.3.2.5	Konsistenzanalyse	75
5.3.3	Validität	78
5.3.3.1	Inhaltliche Validität	79
5.3.3.2	Kriterienbezogene Validität	81
5.3.3.3	Konstruktvalidität	88
5.3.3.4	Validität und Reliabilität	91
5.3.4	Nebengütekriterien	92
5.4	Übungsaufgaben	94
	<b>Lösungen zu den Übungsaufgaben</b>	<b>97</b>
	<b>Index</b>	<b>107</b>

## Glossar

Bitte nutzen Sie, falls erforderlich, das Glossar aus dem Buch „Testtheorie und Testkonstruktion“ von Moosbrugger und Kelava (2012), welches als E-Book über die Universitätsbibliothek verfügbar ist.

Leseprobe

Leseprobe

## Über die Autoren

### **Klauer, Karl Josef (Erstautor)**

Jahrgang 1929. Ausbildung als Volksschullehrer von 1948 – 1950. Ausbildung als Sonderschullehrer an der Universität Mainz 1952 – 1953. Studium der Psychologie an der Universität Mainz 1954 – 1958. Promotion 1958 (Hauptfach Psychologie, Nebenfächer Philosophie und Psychopathologie). Habilitation für Pädagogische Psychologie 1967 an der Mathematisch–naturwissenschaftlichen Fakultät der Universität Düsseldorf.

Berufliche Tätigkeiten: 1950 – 1960 Praxis als Volks– und Sonderschullehrer zu Bernkastel/Mosel, Jünkerath/Eifel und Bad Kreuznach/Nahe. 1956 – 1957 Beurlaubung vom Schuldienst zur Fortsetzung des Psychologiestudiums. 1960 Dozent. 1963 Professor für Lernbehindertenpädagogik am Heilpädagogischen Institut Köln. 1965– 1968 Dekan der inzwischen umgewandelten Heilpädagogischen Fakultät. 1968 Annahme eines Rufs als ordentlicher Professor und Direktor des Seminars für Pädagogik der Technischen Universität Braunschweig. 1976 Annahme eines Rufs als ordentlicher Professor und Direktor des Instituts für Erziehungswissenschaft der RWTH Aachen. Abgelehnte Rufe an die Universität Marburg, an die Universität zu Köln (zwei Mal), an die Universität Freiburg/Schweiz und an die Ludwig–Maximilians–Universität München. 1994 Emeritierung.

### **Klatte, Maria (Überarbeitung und Ergänzung)**

Jahrgang 1962. Studium der Psychologie in Bielefeld und Oldenburg 1985 – 1990. Promotion Universität Oldenburg 1996. Habilitation Universität Oldenburg 2007. 2010 Umhabilitation und Tätigkeit als Privatdozentin TU Kaiserslautern. 2011 Ernennung zur außerplanmäßigen Professorin TU Kaiserslautern.

Berufliche Tätigkeiten: 1992 – 1996 Doktorandin im Graduiertenkolleg „Psychoakustik“ der Universität Oldenburg. 1997 – 2006 wissenschaftliche Mitarbeiterin am Institut für Psychologie, Universität Oldenburg. 2006 – 2007 Lehrbeauftragte im Fach Entwicklungspsychologie der LMU München und Projektleiterin im Projekt „Learning to Think“ zur kognitiven Förderung von Kindern in Kindertagesstätten der Parmenides–Stiftung, München. 2008 – 2009 wissenschaftliche Mitarbeiterin TU Kaiserslautern. 2010 Privatdozentin TU Kaiserslautern. Seit 2011 außerplanmäßige Professorin TU Kaiserslautern, Fachgebiet Kognitive und Entwicklungspsychologie. Lehrgebiete: Entwicklungs–, Kognitive und Pädagogische Psychologie; Forschungsgebiete: Kognitive Entwicklung, Schriftspracherwerb, Arbeitsgedächtnis, Lärmwirkungsforschung.

Leseprobe



# Literaturverzeichnis

## Einführende Literatur

- **Amelang, M. & Schmidt–Atzert, L. (2012). Psychologische Diagnostik (5. Auflage). Heidelberg: Springer.\***
- Ingenkamp, K. & Lissmann, U. (2008). Lehrbuch der Pädagogischen Diagnostik. Weinheim: Beltz.
- Kubinger, K. D. (2009). Psychologische Diagnostik. Theorie und Praxis psychologischen Diagnostizierens (2. Auflage). Göttingen: Hogrefe.
- Leutner, D. (2006). Pädagogisch–psychologische Diagnostik. In: Rost, D. (Hrsg.). Handwörterbuch Pädagogische Psychologie (3. Auflage, S. 559–568). Weinheim: BeltzPVU.
- Lienert, G. A. & Raatz, U. (1998). Testaufbau und Testanalyse (6. Auflage). Weinheim: BeltzPVU.
- **Moosbrugger, H. & Kelava, A. (2012). Testtheorie und Fragebogenkonstruktion (2. Auflage). Heidelberg: Springer Medizin Verlag.\***
- Pospeschill, M. & Spinath, F. M. (2009). Psychologische Diagnostik. München: Reinhardt.
- Rost, J. (2004). Lehrbuch Testtheorie – Testkonstruktion (2. Auflage). Bern: Huber.

## Weiterführende Literatur

- Jäger, R. S. & Petermann, F. (Hrsg.) (1999). Psychologische Diagnostik (3. Auflage). Weinheim: BeltzPVU.
- Klauer, K. J. (Hrsg.) (1978). Handbuch der Pädagogischen Diagnostik (4 Bände). Düsseldorf: Schwann.
- Krauth, J. (1995). Testkonstruktion und Testtheorie. Weinheim: BeltzPVU.
- Langfeldt, H.–P. & Tent, L. (1999). Pädagogisch–psychologische Diagnostik. Band 2: Anwendungsbereiche und Praxisfelder. Göttingen: Hogrefe.
- Lord, F. M. & Novick, M. R. (1968). Statistical theories of mental test scores. Reading, Mass.: Addison–Wesley.
- Tent, L. & Stelzl, I. (1993). Pädagogisch–psychologische Diagnostik. Band 1: Theoretische und methodische Grundlagen. Göttingen: Hogrefe.

- Weinert, F. E. (Hrsg.) (2001). Leistungsmessungen in Schulen. Weinheim: Beltz.
  - Westhoff, K. & Hagemeister, C. (2005). Konzentrationsdiagnostik. Lengerich: Pabst Science Publishers.
- \* *Online-Zugriff für Studierende über die Bibliothek der TU Kaiserslautern*

### Zitierte Literatur

- Baumert, J., Lehmann, R., Lehrke, M., Schmitz, B., Clausen, M., Hosenfeld, I., Köller, O. & Seipp, B. (1997). TIMSS – Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich. Opladen: Leske + Budrich.
- Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U. & Schneider, W. (2001). PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. Opladen: Leske & Budrich.
- Borkenau, P. & Ostendorf, F. (2008). NEO – FFI. NEO – Fünf-Faktoren-Inventar. Göttingen: Hogrefe.
- Brand, S., Reimer, T. & Opwis, K. (2007). How do we learn in a negative mood? Effects of a negative mood on transfer and learning. *Learning and Instruction*, 17, 1–16.
- Breuer, H. & Weuffen, M. (2006). Lernschwierigkeiten am Schulanfang (7. Auflage). Weinheim: Beltz.
- Bröder, M. (2004). Gesprächsführung in Kita und Kindergarten. Freiburg: Herder.
- Brügelmann, H.: PISA 2000: Befunde, Deutungen, Folgerungen. In: *Zeitschrift für Pädagogik*, 54(2) (2002), S. 40–43.
- Bullinger, M., von Mackensen, S. & Kirchberger, I.: KINDL – ein Fragebogen zur Erfassung der gesundheitsbezogenen Lebensqualität von Kindern. In: *Zeitschrift für Gesundheitspsychologie*, 2(1) (1994), S. 64–77.
- Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81–105.
- Cattell, R. B. (1971). *Abilities: Their structure, growth and action*. Boston: Houghton Mifflin.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37–46.

- Cohen, J. (1977). *Statistical Power Analysis for the Behavioral Sciences*. New York: Academic Press.
- Dahmer, H. & Dahmer, F. (2003). *Gesprächsführung. Eine praktische Anleitung* (5. Auflage). Stuttgart: Thieme.
- Deter, B. (1982). Zum Einfluss von Übung und Training auf die Leistung im „Test für medizinische Studiengänge“. Braunschweig: Agentur Pedersen.
- Ebbinghaus, H.: Über eine neue Methode zur Prüfung geistiger Fähigkeiten und ihre Anwendung bei Schulkindern. In: *Zeitschrift für Psychologie und Physiologie der Sinnesorgane*, 13 (1897), S. 401–457.
- Fink, E. (2002). *Souveräne Gesprächsführung und Moderation*. Frankfurt/M.: Campus.
- Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin*, 101, 171–191.
- Fricke, R. (1974). *Kriteriumsorientierte Leistungsmessung*. Stuttgart: Kohlhammer.
- Friedrich, H. F. (1987). Multiple–Choice–Prüfungen im Lehr–Lernsystem „Funkkolleg“. In: Horn, R., Ingenkamp, K. & Jäger, R. S. (Hrsg.). *Tests und Trends* 6. S. 128–175. Weinheim: Beltz.
- Friedrich, H. F., Klemt, W. & Schubring, G.: Ein Rate–Modell für Mehrfachauswahl – Aufgaben. In: *Zeitschrift für erziehungswissenschaftliche Forschung*, 14 (1980), S. 189–208.
- Gerrig, R. (2015). *Psychologie – das Übungsbuch: Das Prüfungstraining zum Zimbardo*. Pearson Studium.
- Gittler, G. (1990). *Dreidimensionaler Würfeltest*. Göttingen: Hogrefe.
- Göllitz, D., Roick, T. & Hasselhorn, M. (2005). Deutsche Mathematiktests für dritte und vierte Klassen (DEMAT 3+ und DEMAT 4). In: Hasselhorn, M., Marx, H. & Schneider, W. (Hrsg.). *Diagnostik von Mathematikleistungen*. S. 167–186. Göttingen: Hogrefe.
- Grewe, W. & Wentura, D. (1997). *Wissenschaftliche Beobachtung*. Weinheim: BeltzPVU.
- Gustafsson, J.–E. (1988). Hierarchical models of individual differences in cognitive abilities. In R. J. Sternberg (Ed.). *Advances in the Psychology of Human Intelligence* (Vol. 4, pp. 35–71). Hillsdale, NJ: Erlbaum.
- Haffner, J., Baro, K., Parzer, P. & Resch, F. (2005). *Heidelberger Rechentest (HRT 1–4)*. Göttingen: Hogrefe.

- Hasselhorn, M., Roick, T. & Gölitz, D. (2005). Stabilitäten und prognostische Validitäten der Mathematikleistungen: Eine Längsschnittstudie mit der DEMAT – Reihe in der Grundschule. In: Hasselhorn, M., Marx, H. & Schneider, W. (Hrsg.). Diagnostik von Mathematikleistungen. S. 187–198. Göttingen: Hogrefe.
- Heller, K. A., Kratzmeier, H., Lengfelder, A. (1998). Matrizen-Test-Manual, Band 1. Ein Handbuch mit deutschen Normen zu den Standard Progressive Matrices von J. C. Raven. Frankfurt/Main: Pearson Assessment.
- Helmke, A. & Renkl, A.: Das Münchener Aufmerksamkeitsinventar: Ein Instrument zur systematischen Verhaltensbeobachtung der Schülersaufmerksamkeit im Unterricht. In: Diagnostica, 38 (1992), S. 130–141.
- Helmke, A., Hosenfeld, I. & Schrader, F. (2004). Vergleichsarbeiten als Instrument zur Verbesserung der Diagnosekompetenz von Lehrkräften. In: Arnold, R. & Griese, C. (Hrsg.). Schulmanagement und Schulentwicklung. Hohengehren: Schneider-Verlag.
- Horn, W. (1972). Begabungstestsystem B-T-S. Göttingen: Hogrefe.
- Hornke, L. F., Etzel, S. & Rettig, K. (1997). Adaptiver Matrizentest (AMD). Mödling: Schuhfried.
- Institut für Test- und Begabungsforschung (Hrsg.) (1990). Der neue TMS. Göttingen: Hogrefe.
- Jansen, H., Mannhaupt, G., Marx, H. & Skowronek, H. (2002). BISC. Bielefelder Screening zur Früherkennung von Lese-Rechtschreibschwierigkeiten. Göttingen: Hogrefe.
- Klatte, M., Steinbrink, C., Prölß, A., Estner, B., Christmann, C. & Lachmann, T. (2014). Effekte des computerbasierten Trainingsprogramms "Lautarium" auf die phonologische Verarbeitung und die Lese-Rechtschreibleistungen bei Grundschulkindern. In: Schulte-Körne, G. (Hrsg.). Legasthenie und Dyskalkulie – Neue Methoden zur Diagnostik und Förderung (S. 127–144). Bochum: Winkler.
- KMK (2002). PISA 2000: Zentrale Handlungsfelder. Beschluss der 299. Kultusministerkonferenz vom 17./18.10.2002.  
[https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen\\_beschluesse/2002/2002\\_10\\_07-Pisa-2000-Zentrale-Handlungsfelder.pdf](https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2002/2002_10_07-Pisa-2000-Zentrale-Handlungsfelder.pdf) (letzter Download 21.3.2016).
- Klauer, K. J. (Hrsg.) (1978). Handbuch der Pädagogischen Diagnostik (4 Bände). Düsseldorf: Schwann.
- Klauer, K. J. (1987). Kriteriumsorientierte Tests. Lehrbuch der Theorie und Praxis lehrzielorientierten Messens. Göttingen: Hogrefe.

- Klauer, K. J. (1993). Learning potential testing. The effect of retesting. In J. H. M. Hamers, K. Sijtsma & A. J. J. M. Ruissenaars (Eds.). *Learning potential assessment* (pp. 135–153). Amsterdam: Swets & Zeitlinger.
- Klauer, K. J. (1994). *Diagnose- und Förderblätter. Rechenfertigkeiten 3. Schuljahr*. Berlin: Cornelsen.
- Klauer, K. J. (2001). Wie misst man Schulleistungen? In: Weinert, F. E. (Hrsg.). *Leistungsmessungen in Schulen*. S. 103–115. Weinheim: Beltz.
- Klauer, K. J. (2006). Intelligenz und Begabung. In: Rost, D. H. (Hrsg.). *Handwörterbuch Pädagogische Psychologie*. S. 275–280. Weinheim: Beltz-PVU.
- Klauer, K. J. (2008). *Denksport für Ältere. Geistig fit bleiben*. Bern: Huber.
- Klauer, K. J. (2010). *Transfer des Lernens*. Stuttgart: Kohlhammer.
- Klauer, K. J., Fricke, R., Herbig, M., Rupprecht, H. & Schott, F. (1972). *Lehrzielorientierte Tests*. Düsseldorf: Schwann.
- Klauer, K. J. & Leutner, D. (2012). *Lehren und Lernen: Einführung in die Instruktionspsychologie*. Weinheim: Beltz.
- Krampen, G. (2007). *KKA. Kaseler – Konzentrations–Aufgabe für 3– bis 8–Jährige*. Göttingen: Hogrefe.
- Krauth, J. (2005). *Testkonstruktion und Testtheorie*. Weinheim: BeltzPVU.
- Kubinger, K. D. (2009). *Psychologische Diagnostik. Theorie und Praxis psychologischen Diagnostizierens (2. Auflage)*. Göttingen: Hogrefe.
- Kurth, E. & Büttner, G. (1999). *TPK. Testreihe zur Prüfung der Konzentrationsfähigkeit*. Göttingen: Hogrefe.
- LeGagnoux, G., Michael, W. B., Hocevar, D. & Maxwell, V. (1990). Retest effects on standardized structure – of – intellect measures for a sample of elementary school children. *Educational and Psychological Measurement*, 50, 475–492.
- Lienert, G. A. & Ratz, U. (1998). *Testaufbau und Testanalyse*. Weinheim: BeltzPVU.
- Lissmann, U. (2010). Schultests. In: Rost, D. H. (Hrsg.). *Handwörterbuch Pädagogische Psychologie*. S. 737–751. Weinheim: Beltz.
- Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, Mass.: Addison–Wesley.

- Marx, H.: Methodische und inhaltliche Argumente für und wider eine frühe Identifikation und Prädiktion von Lese–Rechtschreibschwierigkeiten. In: *Diagnostica*, 38 (1992), S. 249–268.
- Marx, P. & Lenhard, W. (2010). Diagnostische Merkmale von Screeningverfahren. In: Hasselhorn, M. & Schneider, W. (Hrsg.). *Frühprognose schulischer Kompetenzen*. Göttingen: Hogrefe.
- Marx, P. & Weber, J.: Vorschulische Vorhersage von Lese– und Rechtschreibschwierigkeiten. In: *Zeitschrift für Pädagogische Psychologie*, 20 (2006), S. 251–259.
- Mattejat, F. & Remschmidt, H. (2006). *Inventar zur Erfassung der Lebensqualität bei Kindern und Jugendlichen (ILK)*. Bern: Huber.
- McElvany, N., Kortenbruck, M. & Becker, M.: Lesekompetenz und Lesemotivation. In: *Zeitschrift für Pädagogische Psychologie*, 22 (2008), S. 207–219.
- Mund, H. A. (1993). *Naturwissenschaftlich denken. Aachener Beiträge zur Pädagogik*. Aachen: Augustinus Buchhandlung.
- Neumann, O.: Theorien der Aufmerksamkeit: Von Metaphern zu Mechanismen. In: *Psychologische Rundschau*, 43 (1992), S. 83–101.
- Neumann, O. (1996). *Theories of attention*. Amsterdam: Elsevier.
- Petermann, F. (2010). *Kaufman – Computerized Assessment Battery K–ABC*. Frankfurt: Pearson.
- Pospeschill, M. & Spinath, F. M. (2009). *Psychologische Diagnostik*. München: Reinhardt.
- Proyer, R. T., Wagner–Menghin, M. M. & Grafinger, G. (2006). *Leseverständnistest*. Mögling: Schuhfried.
- Putnam, S. P. & Rothbart, M. K. (2006). Development of Short and Very Short forms of the Children's Behavior Questionnaire. *Journal of Personality Assessment*, 87(1), 103–113.
- Raatz, U. & Klein–Braley, C. (1983). Ein neuer Ansatz zur Messung der Sprachleistung. In: Horn, R., Inenkamp, K. & Jäger, R. S. (Hrsg.). *Tests und Trends. 3. Jahrbuch der Pädagogischen Diagnostik*. S. 107–137. Weinheim: Beltz.
- Raatz, U., Voss, B. & Klein–Braley, C. (1991). Diagnose der Fremdsprachenleistung in der Schule. In: Inenkamp, K. & Jäger, R. S. (Hrsg.). *Tests und Trends. 9. Jahrbuch der Pädagogischen Diagnostik*. S. 43–79. Weinheim: Beltz.

- Raven, J. (1998). *Advanced Progressive Matrices (APM)*. Göttingen: Hogrefe.
- Rheinberg, F. (2001). Bezugsnormorientierung. In: Rost, D. H. (Hrsg.). *Handwörterbuch Pädagogische Psychologie*. S. 61–68. Weinheim: Beltz.
- Richter, T., Naumann, J. & Horz, H.: Eine revidierte Fassung des Inventars zur Computerbildung (INCOBI-R). In: *Zeitschrift für Pädagogische Psychologie*, 24 (2010), S. 23–37.
- Rost, D. & Schermer, F. (2007). *Differentielles Leistungsangst-Inventar (DAI)*. Göttingen: Hogrefe.
- Sacher, W. (2005). Deutsche Leistungsdefizite bei PISA: Bedingungsfaktoren in Unterricht, Schule und Gesellschaft. In: Frederking, V., Heller, H. & Scheunpflug, A. (Hrsg.). *Nach PISA. Konsequenzen für Schule und Lehrerbildung nach zwei Studien*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Schmidt-Atzert, L., Büttner, G. & Bühner, M. (2004). Theoretische Aspekte von Aufmerksamkeits-/Konzentrationsdiagnostik. In: Büttner, G. & Schmidt-Atzert, L. (Hrsg.). *Diagnostik von Konzentration und Aufmerksamkeit*. S. 3–22. Göttingen: Hogrefe.
- Schmitz, L. (2009). *Lösungsorientierte Gesprächsführung*. Dortmund: Verlag Modernes Lernen.
- Schöne, C., Dickhäuser, O., Spinath, B. & Stiensmeier-Pelster, J. (2002). *SESSKO. Skalen zur Erfassung des schulischen Selbstkonzepts*. Göttingen: Hogrefe.
- Schrader, F.-W. (2006). Diagnostische Kompetenz von Eltern und Lehrern. In: Rost D. (Hrsg.). *Handwörterbuch Pädagogische Psychologie (3. Auflage)*, S. 95–100. Weinheim: BeltzPVU.
- Schrader, F.-W.: Diagnostische Kompetenz von Lehrpersonen. In: *Beiträge zur Lehrerbildung*, 31(2) (2013), S. 154–165.
- Schrader, F.-W. & Helmke, A. (2001). Alltägliche Leistungsbeurteilung durch Lehrer. In: Weinert, F. E. (Hrsg.). *Leistungsmessungen in Schulen*. S. 45–58. Weinheim: Beltz.
- Spearman, C. (1904). „General intelligence“ objectively determined and measured. *American Journal of Psychology*, 15, 72–101.
- Spearman, C. (1923). *The Nature of Intelligence and the Principles of Cognition*. London: Macmillan.
- Spinath, B.: Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz. In: *Zeitschrift für Pädagogische Psychologie*, 19 (2005), S. 85–95.

- Spinath, B., Stiensmeier-Pelster, J., Schöne, C. & Dickhäuser, O. (2002). SELMO. Skalen zur Erfassung der Lern- und Leistungsmotivation. Göttingen: Hogrefe.
- Steinbrink, C., Schwanda, S., Klatter, M., & Lachmann, T.: Sagen Wahrnehmungsleistungen zu Beginn der Schulzeit den Lese-Rechtschreiberfolg in Klasse 1 und 2 voraus? – Zur prognostischen Validität der Differenzierungsproben 1 und 2 nach Breuer und Weuffen. In: Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 42 (2010), 188–200.
- Strathmann, A. M. & Klauer, K. J.: Diagnostik des Lernverlaufs. Eine Pilotstudie am Beispiel der Entwicklung der Rechtschreibkompetenz. In: Sonderpädagogik, 18 (2008), S. 5–24.
- Strathmann, A. M. & Klauer, K. J.: Lernverlaufsdagnostik: Ein Ansatz zur längerfristigen Lernfortschrittsmessung. In: Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 42 (2010), S. 111–122.
- Strathmann, A.M. & Klauer, K. (2012). Lernverlaufsdagnostik Mathematik für die Klassenstufen 2–4 (LD-M-2-4). Göttingen: Hogrefe.
- Strathmann, A. M., Klauer, K. J. & Greisbach, M.: Lernverlaufsdagnostik. Dargestellt am Beispiel der Entwicklung der Rechtschreibkompetenz in der Grundschule. In: Empirische Sonderpädagogik, 2 (2010), S. 64–77.
- Südkamp, A., Kaiser, J. & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. Journal of Educational Psychology, 104, 743–762.
- von Aster, M., Bzufka, M. & Horn, W. (2009). ZAREKI-K. Neuropsychologische Testbatterie für Zahlenverarbeitung und Rechnen bei Kindern – Kindergartenversion. Göttingen: Hogrefe.
- Walter, J. (2010). Lernfortschrittsdiagnostik Lesen. Göttingen: Hogrefe.
- Weiß, R. H. (2008). Grundintelligenztest Skala 2 – Revision (CFT 20-R) mit Wortschatztest und Zahlenfolgetest – Revision (WS/ZF-R). Göttingen: Hogrefe.
- Westhoff, K. & Kluck, M.-L. (1994). Psychologische Gutachten schreiben und beurteilen. Berlin: Springer.
- Wettstein, A. (2008). BASYS. Beobachtungssystem zur Analyse aggressiven Verhaltens in schulischen Settings. Göttingen: Hogrefe.
- Willson, V. L. & Putnam, R. R. (1982). A meta-analysis of pretest sensitization effects in experimental design. American Educational Research Journal, 19, 249–258.



## Lernziele des Lehrbriefes

Der Lehrbrief „Pädagogisch–psychologische Diagnostik“ bietet eine Einführung in die Grundlagen der psychologischen Diagnostik und ihrer Anwendung im pädagogisch–psychologischen Bereich. Im Zentrum stehen die Prinzipien der Konstruktion und Anwendung psychologischer Tests, wie sie für alle Anwendungsfelder erforderlich sind. Soweit es um konkrete Testverfahren und um spezielle Diagnosen geht, wird aber der Bereich der Anwendung im pädagogisch–psychologischen Bereich im Vordergrund stehen.

Eingangs werden einige Grundlagen und Begriffe zu klären sein, die für das weitere Verständnis erforderlich sind. Weiterhin werden Sie Aufgabenformen mit ihren besonderen Vor– und Nachteilen kennenlernen, die wichtig sind, wenn Sie selbst einen Test konstruieren oder einen neuen Test beurteilen wollen. Anschließend werden Sie die wesentlichen Schritte der Testentwicklung kennenlernen, die erforderlich sind, um ein solches Messinstrument nutzbringend einsetzen zu können. Sie werden feststellen, dass mathematisch–statistische Verfahren bei der Konstruktion, Prüfung und Anwendung von Tests eine wesentliche Rolle spielen. Die Inhalte der von Ihnen bereits absolvierten Methodenmodule sind also beim Studium dieses Lehrbriefes von großer Bedeutung. Bitte schlagen Sie bei Bedarf noch einmal in den entsprechenden Lehrmaterialien nach! Mit diesem Rüstzeug versehen, sind Sie dann in der Lage, die wesentlichen Analysen, die Testautoren im Manual (Testhandbuch) üblicherweise darlegen, nachzuvollziehen und die Güte eines Tests kritisch zu beurteilen.

Nach dem Studium dieses Lehrbriefes ...

- kennen Sie Gegenstand, Aufgaben und Ziele der pädagogisch–psychologischen Diagnostik,
- verstehen Sie die Grundbegriffe der Testdiagnostik und können diese anwenden,
- verstehen Sie die klassische Testtheorie in ihren Grundzügen,
- wissen Sie, wie man Mess– und Entscheidungsfehler minimieren kann,
- verstehen Sie, durch welche Techniken man gewährleisten kann, dass ein Test tatsächlich auch das Merkmal misst, das er messen soll,
- sind Sie in der Lage, Testaufgaben zu entwickeln und zu erproben.

Nach Durcharbeitung des Lehrbriefes sollten Sie in der Lage sein, selbstständig einen Ihnen noch unbekanntem Test daraufhin zu beurteilen, ob bzw. inwieweit er den Gütekriterien der klassischen Testtheorie entspricht und vertretbar eingesetzt werden kann, und Sie sollten in der Lage sein, den Test mithilfe des Manuals richtig durchzuführen, auszuwerten und die Ergebnisse zu interpretieren.

Leseprobe

# 1 Einführung in die pädagogisch–psychologische Diagnostik

## 1.1 Lernziele des Kapitels

Dieses Kapitel führt Sie in die Grundlagen und Grundbegriffe der pädagogisch–psychologischen Diagnostik ein. Es bildet die Basis für das Verständnis der spezielleren Themengebiete des Lehrbriefes. Neben der Erläuterung zentraler Grundbegriffe werden die wichtigsten Aufgabengebiete der pädagogisch–psychologischen Diagnostik umrissen, und es werden Befunde zur schulischen Leistungsbeurteilung und zur diagnostischen Kompetenz von Lehrkräften dargestellt, die in der Folge wesentlich zur Verbesserung der Lehramtsausbildung im diagnostischen Bereich beigetragen haben. Das Kapitel schließt mit einem Abriss der geschichtlichen Entwicklung des Gebiets.

**Lernziele des Kapitels**

Wenn Sie dieses Kapitel durchgearbeitet haben, wissen Sie,

- womit sich die pädagogisch–psychologische Diagnostik befasst,
- was latente und manifeste Variablen sind,
- was man unter einem theoretischen Konstrukt versteht und warum man überhaupt von solchen spricht,
- welche Methoden in der Diagnostik eingesetzt werden,
- was psychologische Tests sind,
- welche Arten von Tests in der pädagogisch–psychologischen Diagnostik häufig eingesetzt werden,
- welche Urteilsfehler besonders zu beachten sind,
- was man unter Bezugsnormen versteht,
- wie es um die diagnostische Kompetenz von Lehrkräften steht.

## 1.2 Gegenstand und Aufgaben der pädagogisch–psychologischen Diagnostik

Die **Psychologische Diagnostik** stellt eine zentrale Anwendungsdisziplin der Psychologie dar. Sie befasst sich mit dem systematischen Erfassen und Aufbereiten von Informationen, mit dem Ziel, Entscheidungen und darauf basierende Maßnahmen (z. B. eine Psychotherapie oder Fördermaßnahme) zu begründen und zu optimieren (Jäger & Petermann, 1999, S. 11). In aller Regel beginnt der **diagnostische Prozess** mit einer Fragestellung eines Klienten oder Auftraggebers, auf

**Psychologische Diagnostik**

**diagnostische Prozess**

die eine Lösung oder Beantwortung durch den Diagnostiker erwartet wird. Zur Beantwortung dieser Frage werden systematisch Daten erhoben, wobei verschiedene Methoden zum Einsatz kommen: Testverfahren (z. B. Entwicklungs-, Intelligenz- und Schulleistungstests, Persönlichkeitstests), Dokumentenanalyse (z. B. Einsichtnahme in Zeugnisse und bereits vorliegende Gutachten), Interviewmethoden (z. B. Gespräche mit Eltern, Lehrkräften, dem betroffenen Kind), Beobachtungsmethoden (z. B. Beobachtung eines Kindes im Unterricht und/oder bei der Erledigung der Hausaufgaben). Im nächsten Schritt werden die gesammelten Informationen integriert, und es wird eine Antwort auf die Frage formuliert. Je nach Fragestellung kann diese Antwort in einer Diagnose (z. B. „Leserechtschreibstörung liegt vor“), einem Entscheidungsvorschlag für eine Maßnahme oder einer Vorhersage bestehen (z. B. im Falle der Eignungsdiagnostik: „Von den Bewerberinnen und Bewerbern wird Kandidatin X die zu besetzende Position am besten ausfüllen“). Wenn es möglich ist, sollte im letzten Schritt eine Evaluation erfolgen, d. h. der Diagnostiker sollte im Nachhinein überprüfen, ob seine Diagnose bzw. Handlungsempfehlung korrekt war und zum Erfolg geführt hat.

**pädagogisch-  
psychologische  
Diagnostik**

Die **pädagogisch-psychologische Diagnostik** ist nichts anderes als Psychologische Diagnostik, die im Dienst pädagogischer Entscheidungen eingesetzt wird. Dabei kann es sich um Entscheidungen handeln, die sich auf einzelne Personen beziehen, etwa auf ein einzelnes Kind, oder auf ganze Gruppen von Individuen, wie etwa auf eine Schulklasse. Die Entscheidungen können den gegenwärtigen Zustand betreffen, vielfach geht es aber auch darum, längerfristige Entwicklungen zu dokumentieren oder den Erfolg einer Intervention, einer Fördermaßnahme nachzuweisen.

**Definition 1.1:**  
pädagogisch-  
psychologische  
Diagnostik

Pädagogisch-psychologische Diagnostik umfasst das Insgesamt von Erkenntnisbemühungen im Dienste aktueller pädagogischer Entscheidungen (Klauer, 1978, S. 5).

Bei jeder Diagnostik geht es zunächst um Erkenntnis, also um die Feststellung von Sachverhalten, von Eigenschaften, Bedingungen und dergleichen mehr. Allerdings unterscheidet sich die pädagogisch-psychologische Diagnostik von der wissenschaftlichen Forschung insofern, als es nicht um die Entdeckung allgemeiner Zusammenhänge, sondern um Erkenntnisbemühungen geht, die wohlbegründete, „gute“ pädagogische Entscheidungen ermöglichen sollen. Dabei kann es um Planungsentscheidungen gehen, aber auch um Handlungsentscheidungen oder um Feststellungs- und Bewertungsentscheidungen. Ferner handelt es sich meist um aktuell anstehende Entscheidungen, die sich auf bestimmte Fragestellungen und auf einzelne Lernende oder auf Gruppen von Lernenden beziehen. Wenn es aber darum geht, den Leistungsstand von Fünftklässlern international vergleichend zu erheben, neue Curricula zu erproben oder die Effekte bestimmter Lehrmethoden zu analysieren, so können solche Untersuchungen im Dienste bildungspolitischer Entscheidungen erfolgen, wobei es eben nicht darum geht, pädagogische Ent-

scheidungen zugunsten einzelner Individuen begründet zu fällen. Gleichwohl werden die Instrumente der pädagogisch–psychologischen Diagnostik auch in solchen Untersuchungen eingesetzt.

Die meisten pädagogischen Entscheidungen, die im Rahmen eines diagnostischen Prozesses zu treffen sind, erfordern die Erfassung des gegenwärtigen Ist–Standes hinsichtlich der Ausprägung bestimmter Merkmale eines Kindes (z. B. Intelligenz, Lesefähigkeit) oder einer Schulklasse (z. B. durchschnittliches Leistungsniveau in Mathematik). Wenn es beispielsweise um Schullaufbahnentscheidungen geht oder um die Entscheidung, ob ein Kind besonderer Hilfen bedarf, so wird man zunächst einmal zu klären suchen, wie sich der Status des Kindes zum gegenwärtigen Zeitpunkt darstellt. In solchen Fällen spricht man von **Statusdiagnostik**. Hier geht es darum, den aktuellen Stand zu dokumentieren, was in aller Regel eine einmalige Diagnose erfordert. Anders liegen die Dinge bei der Evaluation einer Interventionsmaßnahme, um festzustellen, ob sie bei einzelnen Probanden oder Gruppen von Probanden bereits erfolgreich war oder noch fortgesetzt werden sollte. Solche Untersuchungen erfordern mindestens eine zweimalige Datenerhebung (vor und nach der Intervention). Wollen aber Lehrpersonen oder Therapeuten entscheiden, ob ihre Lehrmethode bzw. Intervention so weitergeführt werden kann wie bisher oder einer Modifikation bedarf, so müssen Entwicklungen oder Lernverläufe erfasst werden. Hierzu sind längerfristig zu wiederholende Erhebungen über den Zeitraum der Intervention erforderlich. Dies bezeichnet man als **Verlaufsdagnostik**.

**Statusdiagnostik**

**Verlaufsdia-  
gnostik**

In der Diagnostik wird also zunächst der Ist–Stand von Probanden erhoben. Es ist natürlich nicht möglich, das „Wesen“ eines Menschen insgesamt diagnostisch zu erfassen. Stattdessen werden Probanden hinsichtlich der Ausprägung einzelner Merkmale diagnostiziert. Je nach diagnostischer Fragestellung kann es sich bei diesen Merkmalen um kognitive Fähigkeiten (z. B. Intelligenz, Aufmerksamkeit, Gedächtnis, Schulleistungen, Sprachentwicklungsstand), Einstellungen (z. B. Schuleinstellungen), Persönlichkeitsmerkmale (z. B. Leistungsmotiv; „Big Five“, Ängstlichkeit) oder auch um Merkmale der Lern– und Entwicklungsumwelt handeln (z. B. Bindung zwischen Kind und Bezugsperson, soziale Integration eines Kindes in die Lerngruppe). Diese Merkmale werden unter standardisierten Bedingungen erhoben, um die Vergleichbarkeit der Befunde zu gewährleisten. Die Bedingungen sind standardisiert, wenn sie in allen für wesentlich erachteten Gesichtspunkten gleichgehalten werden.

Die im Rahmen eines diagnostischen Prozesses zu erfassenden Merkmale sind Variablen, also Größen, die in zwei oder mehr Ausprägungen vorkommen können. Dass die Person XY ein Mensch ist, unterscheidet sie nicht von anderen Menschen. Insofern handelt es sich bei der Eigenschaft „Mensch“ in dem Zusammenhang nicht um eine **Variable**. Dass XY aber eine Frau ist und 27 Jahre alt, das unterscheidet sie sehr wohl von anderen Personen. Das Geschlecht und das Lebensalter stellen also Variablen dar. Hierbei handelt es sich um **manifeste**, also mehr oder minder klar nachweisbare Variablen. Für die Psychologie mindes-

**latente und  
manifeste  
Variablen**

tens genauso interessant sind aber Variablen, die nicht manifest sind, sondern latent, nicht direkt beobachtbar. **Latente Variablen** sind z. B. die Intelligenz, die Ängstlichkeit, die Lesefertigkeit oder die Leistungsmotivation.

**theoretische Konstrukte**

Latente Variablen sind also nicht direkt beobachtbar. Man bezeichnet sie auch als **theoretische Konstrukte**. Ihre Ausprägung lässt sich nur erschließen aufgrund bestimmter Beobachtungen. Beispielsweise schließt man auf eine hoch ausgeprägte Intelligenz, wenn jemand sehr gut in der Lage ist, neue Probleme zu lösen und neue Informationen zu erlernen. Aus diesen Beobachtungen – den manifesten Variablen – schließt man also auf zugrunde liegende Eigenschaften oder Merkmale, auf latente Größen. Und um der Vergleichbarkeit willen geschieht die Beobachtung unter standardisierten Bedingungen: Im Falle der Intelligenz würde man einen standardisierten Intelligenztest durchführen, um anhand der beobachtbaren Antworten der Person auf die Testaufgaben (manifeste Variablen) auf die Ausprägung des Konstrukts „Intelligenz“ (latente Variable) schließen zu können.

Eine andere Frage ist allerdings, ob es tatsächlich Eigenschaften wie Intelligenz, Ängstlichkeit oder Leistungsmotivation gibt. Im Grunde lassen sich solche Fragen nicht entscheiden. Stattdessen postuliert oder behauptet man ihre Existenz, wenn es dafür gute Argumente gibt. Um also keinen Missverständnissen ausgesetzt zu sein, wird den latenten Persönlichkeitsmerkmalen der Status eines theoretischen Konstrukts zuerkannt. Das bedeutet nichts anderes, als dass theoretisch die Existenz eines solchen Merkmals angenommen oder behauptet wird. Dies hat den Vorteil, bei neuen Erkenntnissen gegebenenfalls auch andere Persönlichkeitseigenschaften annehmen zu können (wie etwa die Informationsverarbeitungsgeschwindigkeit oder die Arbeitsgedächtniskapazität, an die früher kaum jemand dachte und die nach Meinung einiger Forscher anstelle der Intelligenz treten könnten). In der Praxis bedeutet dies, dass ein theoretisches Konstrukt erst dann von der Scientific Community, also von der Mehrheit der Wissenschaftler, akzeptiert wird, wenn es gute empirische Gründe gibt, weiterhin mit dem Konstrukt zu arbeiten.

### 1.3 Methoden der pädagogisch–psychologischen Diagnostik

Wie im vorigen Abschnitt bereits dargelegt, erfordert die Entscheidungsfindung im Rahmen eines diagnostischen Prozesses die Sammlung und Integration von Informationen aus unterschiedlichsten Quellen. Die wichtigsten Datenerhebungsmethoden in der psychologischen und pädagogisch–psychologischen Diagnostik umfassen

- die Verhaltensbeobachtung,
- Gesprächs– und Interviewverfahren sowie
- Testverfahren.

Diese Methoden sollen im Folgenden kurz dargestellt werden.

### 1.3.1 Verhaltensbeobachtung

**Verhaltensbeobachtungen** können zu wichtigen Erkenntnissen führen, auch wenn sie unsystematisch und außerhalb des regulären Rahmens erfolgen, etwa auf dem Schulhof oder im familiären Umfeld. Grundsätzlich kommen auch bei der Verhaltensbeobachtung **Urteilstendenzen**, d. h. Tendenzen zur Abgabe verzerrter Beurteilungen, ins Spiel. Diese können bei verschiedenen Beurteilern unterschiedlich stark ausgeprägt sein, aber auch bei einzelnen Beurteilern in verschiedenen Situationen variieren. Bekannt sind die **Milde-** und **Strengtendenz**, d. h. die Tendenz zu sehr positiven bzw. sehr negativen Beurteilungen, und der **Halo-Effekt**, bei dem ein positiver bzw. negativer Eindruck bezüglich einzelner Merkmale auf die Beurteilung anderer Merkmale „überstrahlt“ (so wird beispielsweise bei gepflegter Erscheinung und sicherem Auftreten auf Einsatzbereitschaft und Intelligenz geschlossen). Solche Urteilstendenzen wird der einzelne Beurteiler kaum bemerken, am ehesten noch im Vergleich mit den Ergebnissen anderer Urteile.

**Verhaltens-  
beobachtung**

Um Beobachtungsfehler zu minimieren und zu möglichst genauen Aussagen zu gelangen, werden im Rahmen eines diagnostischen Prozesses **systematische Verhaltensbeobachtungen** durchgeführt. Eine systematische Verhaltensbeobachtung ist gekennzeichnet durch einen klar definierten Beobachtungsgegenstand (z. B. Mitarbeit des Schülers X im Unterricht) und ein regelgeleitetes, standardisiertes Vorgehen. Dem Beobachter werden genaue Regeln vorgegeben, die beschreiben, welche Verhaltensweisen (z. B. X meldet sich, X schaut aus dem Fenster, X unterhält sich mit Tischnachbarn) wann (z. B. alle 5 Minuten für 10 Sekunden) zu beobachten sind und wie diese Beobachtungen zu protokollieren sind. Hierdurch soll erreicht werden, dass die Beobachtungen – anders als unsystematische Gelegenheitsbeobachtungen – objektiv erfolgen und eine möglichst repräsentative Stichprobe der für die Fragestellung relevanten Verhaltensweisen erhoben wird.

**systematische  
Verhaltens-  
beobachtung**

Derartige Beobachtungsinstrumente oder „**Beobachtungssysteme**“ können für eine spezifische Fragestellung vom Diagnostiker selbst konstruiert werden, etwa nach den Anleitungen von Grewe und Wentura (1997). Oftmals bietet es sich jedoch an, auf bereits vorhandene und geprüfte Beobachtungsinstrumente zurückzugreifen. Beispiele für solche standardisierten Beobachtungssysteme sind das „Münchener Aufmerksamkeitsinventar“ (Helmke & Renkl, 1992) zur Erfassung des Aufmerksamkeitsverhaltens von Schülerinnen/Schülern im Unterricht und das Beobachtungssystem zur Analyse aggressiven Verhaltens in schulischen Settings (BASYS, Wettstein, 2008), welches die systematische Beobachtung aggressiver und oppositioneller Verhaltensweisen von Schülerinnen/Schülern durch Lehrkräfte oder Fremdbeobachter sowie die Ableitung von Interventionsschritten ermöglicht. Die Qualität der Beobachterübereinstimmung lässt sich mittels des Koeffizienten **Kappa von Cohen** (1960), bestimmen. Hierbei wird der Prozentanteil der Übereinstimmungen zwischen mehreren Beurteilern in Relation zu dem Prozentanteil gesetzt, der sich bei rein zufälligen Beurteilungen ergeben würde. Der Parameter Kappa kann Werte zwischen 0 (Übereinstimmung zwischen den Beurtei-

**Beobachtungs-  
systeme**

**Kappa von Cohen**

lern geht nicht über den Zufall hinaus) und 1 (vollkommene Übereinstimmung) annehmen. Nach Greve und Wentura (1997) weisen Werte über 0.75 auf eine hohe Beurteilerübereinstimmung hin, während Werte unter 0.4 mit Skepsis zu betrachten sind. Die Übereinstimmung kann deutlich erhöht werden, indem die Beobachter im Umgang mit dem Beobachtungssystem geschult werden (z. B. mittels Beobachtung von Videoszenen) und auf Übersichtlichkeit und Handhabbarkeit der Protokollbögen geachtet wird (während der Beobachtung ist keine Zeit, lange herumzublättern!).

### 1.3.2 Gespräch und Interview

#### Gespräch und Interview

**Gespräche** und ausführlichere **Interviews** sind im pädagogischen Kontext unerlässlich. Techniken der Gesprächsführung sollten daher eigentlich zum Standardrepertoire der Ausbildung von Pädagogen gehören. Für Führungskräfte in Organisationen und Wirtschaft existiert eine Fülle von Seminarangeboten, in denen solche Techniken insbesondere auch zur Motivierung der Mitarbeiter oder zur Konfliktlösung eingeübt werden. Entsprechende Literatur steht ebenfalls zur Verfügung (vgl. etwa Fink, 2002). In der Medizin kommt dem Gespräch mit Patienten ebenfalls eine zentrale Rolle zu, und auch in diesem Bereich gibt es hilfreiche Literatur (Dahmer & Dahmer, 2003). Selbst für Erzieherinnen und Erzieher in Kindertagesstätten gibt es praxisnahe Hilfestellungen (etwa Bröder, 2004), so insbesondere zur „Lösungsorientierten Gesprächsführung“ (Schmitz, 2009). Für Psychologen existieren entsprechende Anleitungen zu diagnostischen Gesprächen und Interviews (Westhoff & Kluck, 1994). Dagegen wird die Gesprächskultur im Schulbereich keineswegs so gepflegt, wie dies wünschenswert wäre.

Gespräche dienen vielfach dazu, kleinere oder größere Konflikte zu bewältigen, Lernende zu stärkerem Einsatz zu motivieren oder folgenreiche Entscheidungen begründet zu fällen. Die Gütekriterien, die für Tests gelten und weiter unten behandelt werden, müssten für diagnostische Gespräche ebenfalls gelten. In der Praxis hat es sich bei Interviews bewährt, zuvor einen möglichst präzisen und differenzierten Leitfaden zu erstellen, an dem man sich dann im Gespräch auch orientieren sollte, um von jedem Probanden die Information zu erhalten, um die es gerade geht. Ein solcher Leitfaden kann den Inhalt und Verlauf des Gesprächs in gewisser Weise steuern, dem Gespräch also ein Mindestmaß an Objektivität sichern. Da auf diese Weise aber auch alle wichtigen Aspekte des Problems systematisch angesprochen werden, darf man damit rechnen, gültigere und verlässlichere Befunde zu erzielen als bei einem spontan geführten Gespräch, bei dem die Teilnehmer ihren eigenen Impulsen folgen.

### 1.3.3 Testverfahren

#### Testverfahren

Im Bereich der **Testverfahren** unterscheidet man standardisierte und nicht standardisierte Tests. Bei den Letzteren handelt es sich um Verfahren, wie sie manche



Lehrkräfte sozusagen für den Hausgebrauch einsetzen (z. B. schriftliche Kurzttests zur Hausaufgabenüberprüfung; mündliches „Abfragen“). Man bezeichnet sie auch als informelle Verfahren. Die in der psychologischen und pädagogisch–psychologischen Diagnostik eingesetzten Tests sind dagegen nach wissenschaftlichen Grundsätzen entwickelt und hinsichtlich ihrer Qualität erprobt.

Die standardisierten Tests der pädagogisch–psychologischen Diagnostik werden gerne in diese vier Gruppen eingeteilt:

- **Intelligenztests** lassen sich noch differenzieren in Tests der allgemeinen Intelligenz und in Tests, die spezielle Begabungen erfassen sollen.
- **Entwicklungstests** dienen der Erfassung des allgemeinen Entwicklungsniveaus und werden vor allem in der frühen Kindheit eingesetzt.
- **Leistungstests** erfassen spezielle Leistungen, wie die Konzentrationsfähigkeit, das Sprachverständnis oder die Motorik, aber auch Schulleistungen, wie die Lesefertigkeit oder mathematische Leistungen.
- **Persönlichkeitstests** werden herangezogen, um Persönlichkeitseigenschaften, wie etwa die Interessen, die Ängstlichkeit, die Leistungsmotivation oder Einstellungen, zu messen.

Die Anwendung solcher Tests kann die beste Begründung für eine diagnostische Vorgehensweise liefern, wie später noch deutlich werden wird. Allerdings sind auch hier Einschränkungen anzumerken: Erstens gibt es für viele pädagogisch wichtige Entscheidungen (noch) keine Testverfahren, und zweitens gibt es auch Tests, die nicht so gut sind, wie sie sein sollten oder könnten.

Unter einem Test versteht man in der Psychologie ein wissenschaftliches Prüfverfahren, bei dem unter standardisierten Bedingungen Verhaltensmerkmale von Personen erhoben werden, welche als Indikatoren (manifeste Variablen) für bestimmte Merkmale (latente Variablen; theoretische Konstrukte) betrachtet werden. Ziel ist es, aus dem Testergebnis (Rohwert) eine Aussage über den Grad der Ausprägung des latenten Merkmals bei dieser Person abzuleiten.

Dies wird verständlich, wenn man sich vergegenwärtigt, dass der Testrohwert selbst zunächst keine Beurteilung der Leistung ermöglicht. Angenommen, ein Kind hat im Rechtschreibtest genau acht Wörter falsch geschrieben. Die beobachtete Leistung ist damit recht präzise umschrieben – aber ob dies eine gute oder eine schlechte Leistung darstellt, bleibt völlig offen. Eine Leistungsbeurteilung wird erst dann möglich, wenn ein Wertmaßstab angelegt wird, eine **Bezugsnorm**, an der man die tatsächliche Leistung misst. Bei der Bewertung von Leistungsständen kommen in der pädagogisch–psychologischen Diagnostik drei Bezugsnormen zur Anwendung (vgl. Rheinberg, 2001):

- die **soziale Bezugsnorm**: die in einer Vergleichsgruppe erhobenen Testwerte,

**Bezugsnormen bei  
psychologischen  
Tests**

- die **sachliche** oder **kriteriumsbezogene Bezugsnorm**: Erreichen oder Nicht-Erreichen eines vorab definierten Kriteriums (z. B. Lernziel),
- die **individuelle Bezugsnorm**: die vorherigen Leistungen der untersuchten Person selbst.

**sozialnormorientierte/bezugsgruppenorientierte Tests**

Tests, bei denen eine Vergleichsgruppe als Bezugsmaßstab dient (soziale Bezugsnorm), wurden vorab mit einer großen Stichprobe von Personen („Normstichprobe“, auch „Eichstichprobe“ genannt) durchgeführt. Die Aussage über die Merkmalsausprägung einer mit dem Test untersuchten Person besteht dann in einem Vergleich ihrer Testwerte mit den in der Normstichprobe erhobenen Werten. Man bezeichnet solche Tests auch als **sozialnormorientierte** oder **bezugsgruppenorientierte Tests**.

Um normorientierte Aussagen zu ermöglichen, sind im Rahmen der Testkonstruktion umfangreiche Normdaten zu erheben. Im Prinzip sollten die Eichstichproben repräsentativ für das Bundesgebiet einerseits und für die jeweiligen Altersstufen andererseits erfasst werden; dies wird allerdings in der Praxis nicht immer so konsequent gehandhabt. Der Aufwand ist also beträchtlich. Hinzu kommt, dass Normdaten nicht auf Dauer gültig bleiben. Die meisten veralten nach etwa einem Jahrzehnt, sodass dann eigentlich neu geeicht werden müsste.

Die weitaus meisten Testverfahren der pädagogisch–psychologischen Diagnostik, wie etwa Intelligenztests und viele Schulleistungstests, sind sozialnorm- bzw. bezugsgruppenorientierte Tests. Im klassischen Lehrbuch von Lienert und Ratz (1998) wird der Begriff „Test“ in diesem Kontext wie folgt definiert (bitte beachten Sie: der Begriff „Persönlichkeitsmerkmal“ bezieht sich in dieser Definition nicht nur auf die bekannten Persönlichkeitseigenschaften wie „Big Five“ und Ängstlichkeit, sondern umfasst auch Einstellungs- und Leistungsmerkmale):

**Definition 1.2:**  
Test

„Ein **Test** ist ein wissenschaftliches Routineverfahren zur Untersuchung eines oder mehrerer empirisch abgrenzbarer Persönlichkeitsmerkmale mit dem Ziel einer möglichst genauen quantitativen Aussage über den relativen Grad der individuellen Merkmalsausprägung“ (Lienert & Ratz, 1998, S. 1).

Diese zunächst recht abstrakt anmutende Definition soll kurz erläutert werden. Mit dem Ausdruck „wissenschaftliches Routineverfahren“ ist gemeint, dass die Tests nach vorgegebenen Richtlinien und Regeln entwickelt und erprobt und nach festgelegten Anweisungen durchgeführt und ausgewertet werden. Diese Regeln und Richtlinien basieren auf der klassischen Testtheorie, der wir uns im folgenden Kapitel zuwenden werden. Mit dem relativen Ausprägungsgrad ist gemeint, dass der Test eine Aussage über die Position des Einzelnen innerhalb seiner Population zulässt, also etwa, ob der Betreffende bei diesem Merkmal durchschnittlich, über- oder unterdurchschnittlich einzuordnen ist. Quantitativ präzisiert ist die Aussage, wenn diese relative Position numerisch genauer fixiert ist, z. B. durch einen T-Wert oder einen Prozentrangwert.

**Kriteriumsbezogene Tests** erlauben Aussagen darüber, ob eine Person ein vorab festgelegtes Kriterium (ein inhaltlich definiertes Lern- oder Therapieziel oder eine bestimmte Mindest-Punktzahl) erreicht hat oder nicht. Als Bezugsnorm für die Leistungsbeurteilung wird also ein vorab gesetztes Kriterium herangezogen. Solche Verfahren werden nahezu ausschließlich im schulischen Bereich eingesetzt. Mit ihrer Hilfe soll ermittelt werden, ob bzw. inwieweit Lernende ein bestimmtes Lehrziel erreicht haben, weswegen man auch gerne von **lehrzielorientierten Tests** spricht. Sozialnormorientierte Schulleistungstests gestatten zwar eine Einschätzung darüber, wie die Einzelnen in der Gesamtheit der Normstichprobe einzuordnen sind, daraus geht aber nicht hervor, ob das Lehrziel erreicht worden ist oder nicht. Diese Fragestellung ist im Kontext des Lehr-Lern-Prozesses aber besonders interessant, denn sie ermöglicht beispielsweise die Entscheidung, ob der Unterricht mit der nächsten Lektion weitergeführt werden kann oder ob einzelne Schülerinnen und Schüler noch zusätzlicher Übung bedürfen. Lehrzielorientierte Tests sind daher im Rahmen des Unterrichts sinnvoll einzusetzen. Darüber hinaus sind solche Tests hilfreich, wenn es darum geht festzustellen, ob ein übergreifendes Lehrziel, wie es im Lehrplan festgelegt ist, von den Lernenden etwa am Ende des Schuljahres beherrscht wird. Bundesweit sind lehrzielorientierte Tests natürlich nur dann einsetzbar, wenn das jeweilige Lehrziel auch bundesweit vermittelt und angestrebt wird.

**kriteriums-  
bezogene Tests**

**lehrzielorientierte  
Tests**

Lehrzielorientierte Tests setzen in jedem Fall zweierlei voraus:

- die inhaltliche Präzisierung des Lehrziels sowie
- die quantitative Festlegung des Lehrziels, also die genaue Angabe des geforderten Kompetenzgrads.

Mit dem letzteren Aspekt beginnend, ist zunächst festzuhalten, dass erst durch die Festlegung eines geforderten Kompetenzgrads entschieden werden kann, ob jemand das Lehrziel erreicht hat oder nicht. Dabei geht es also um eine Klassifikation, genauer, um die am Lehrziel orientierte Klassifikation. Eine einfache Möglichkeit besteht darin, das Lehrziel durch einen Prozentsatz zu quantifizieren, durch den Prozentsatz von Aufgaben, der beherrscht werden muss, um das Lehrziel erreicht zu haben. Sicherlich ist es nicht sinnvoll, 100 % richtige Lösungen zu erwarten, weil Fehler auch bei perfekter Kompetenz nie ganz auszuschließen sind. Man wird also eine gewisse Fehlertoleranz einbauen und den geforderten Kompetenzgrad beispielsweise auf 90 % festlegen. Tatsächlich gibt es bei manchen Lehrzielen auch gute Gründe, ein noch niedrigeres Kriterium zu setzen (wie etwa die Lösung von mindestens 50 % der Aufgaben in der Klausur zu diesem Studienmodul). Inzwischen sind mehrere Verfahren entwickelt worden, um die Festsetzung des Kompetenzgrades zu objektivieren, worauf hier nur verwiesen werden soll (Klauer, 1987, S. 62 ff).

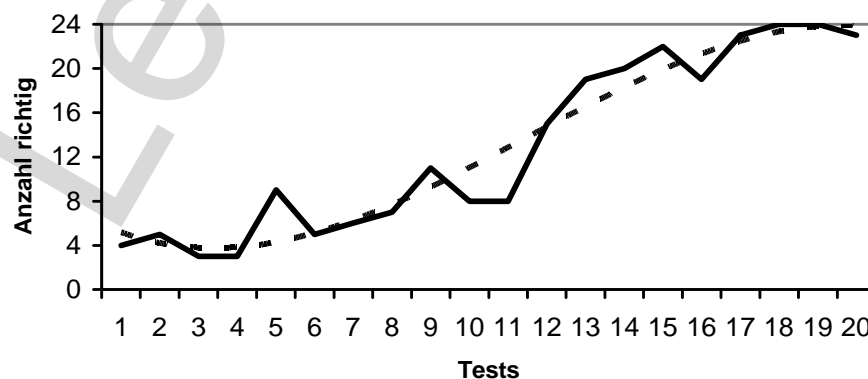
Die inhaltliche Präzisierung des Lehrziels kann Verfahren erfordern, die für Pädagogen ungewohnt sind. Hier geht es um den Aspekt der inhaltlichen Validität, der im späteren Kapitel über Testgütekriterien noch ausführlich behandelt wird. Vor-

wegenommen sei hier, dass ein Lehrziel wie beispielsweise die Beherrschung der Grundrechenarten als **Aufgabenmenge** definiert werden kann. Aus dieser Menge können dann Zufallsstichproben von Aufgaben erzeugt werden. Hierdurch entstehen äquivalente Stichproben von Aufgaben, die als Tests eingesetzt werden können und die gleiche Kompetenz messen. Auf diese Weise lässt sich sowohl der Kompetenzgrad der Einzelnen ermitteln als auch die lehrzielorientierte Klassifikation vornehmen, also entscheiden, wer das Ziel erreicht hat und wer nicht.

Theoretisch sind kriteriumsorientierte Verfahren gut fundiert, aber praktisch spielen sie keine allzu große Rolle. Das dürfte hauptsächlich darauf zurückzuführen sein, dass Hauptinteressenten eigentlich Lehrkräfte sein müssten, die aber faktisch nicht besonders interessiert sind an solchen Verfahren, zumal diese indirekt ja auch den Lehrerfolg abtesten.

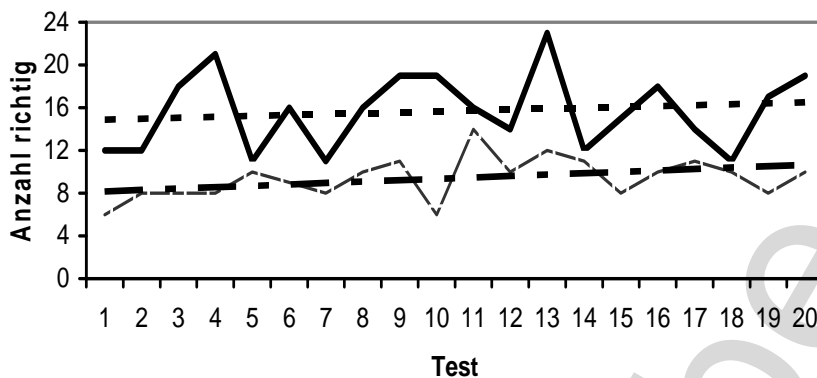
Geht es um Aussagen über den individuellen Entwicklungsverlauf der Testperson (z. B. im Rahmen einer Lerntherapie), so wird das aktuelle Testergebnis nur im Hinblick auf die vorherigen Leistungen der Testperson bewertet; es wird somit eine individuelle Bezugsnorm angelegt. Für eine solche **Lernverlaufsdagnostik** entwickelte Tests werden über längere Zeit regelmäßig durchgeführt und bilden so die Lernverläufe individueller Probanden oder ganzer Schulklassen ab. Dabei werden typische Verläufe unterscheidbar, etwa negativ beschleunigte Verläufe, wenn sich Lernende dem Lehrziel annähern. Es kann aber auch vorkommen, dass über längere Zeit hinweg keinerlei Fortschritt zu verzeichnen ist oder sogar Rückschritte auftreten. Die Abbildungen 1.1a und 1.1b zeigen Beispiele für solche Lernverlaufskurven aus einer Studie von Strathmann und Klauer (2010). Die Daten wurden mittels des inzwischen publizierten Verfahrens „Lernverlaufsdagnostik Mathematik für die Klassenstufen 2–4“ (LD-M-2-4, Strathmann & Klauer, 2012) erhoben.

**Abbildung 1.1a:**  
Lernverlaufskurve  
eines Mädchens  
der 2. Klasse (die  
ausgezogene Linie  
zeigt die Rohdaten,  
die strichlierte  
Linie die  
angepasste  
Regressionslinie)



In Abbildung 1.1a zeigt das Mädchen in den ersten Wochen praktisch keinen Lernzuwachs, später aber sehr deutliche Fortschritte, sodass sie am Ende des Schuljahres alle 24 Aufgaben des jeweiligen Tests löst. Es gibt aber auch andere Beispiele, wie aus Abbildung 1.1b hervorgeht. Die beiden Kinder aus Abbildung 1.1b zeigen nahezu keinen Lernzuwachs während des ganzen Schuljahres.

Möglicherweise ist dies den Lehrpersonen entgangen, denn sie hatten bisher keine Möglichkeit, den Lernverlauf so konsequent zu dokumentieren.



**Abbildung 1.1b:** Lernverlaufskurve eines Drittklässlers (obere Linie) und eines Viertklässlers (untere Linie) (beide Abbildungen nach Strathmann & Klauer, 2010).

Ein Instrument zur Lernverlaufsdagnostik im Lesen stellt die „Lernfortschrittsmessung Lesen (LDL)“ von Walter (2010) dar. Beim Lesen ist es nicht so leicht, eine ausreichende Anzahl gleich schwieriger Lesetexte zu erzeugen, aus der dann zufällig gezogen werden kann. Deshalb hat Walter 28 Lesetexte ausgewählt, die zwischen 188 und 259 Wörter umfassen und gleich schwer sein sollen. Die Aufgabe des Kindes besteht darin, für genau eine Minute laut vorzulesen, wobei die Anzahl der richtig gelesenen Wörter ermittelt wird. Walter kann eindrucksvolle Lernverlaufskurven vorlegen, die zeigen, welche Vorteile mit dieser Art von Diagnostik verbunden sind.

#### **Lernfortschrittsmessung Lesen – LDL**

Natürlich sind die drei in diesem Kapitel behandelten Bezugsmaßstäbe nicht völlig unabhängig voneinander. So kann die Leistung einer Person in einem normierten Schulleistungstest natürlich auch mit ihrer früheren Leistung in diesem Test verglichen werden, und bei der Lernverlaufsdagnostik kann die zu einem Zeitpunkt erreichte Leistung auch mit einem vorab gesetzten Kriterium verglichen werden.

Wichtig ist nun, dass für sozialnormorientierte, kriterienorientierte und verlaufsorientierte diagnostische Verfahren jeweils spezifische Konstruktionsbedingungen und eigene Gütekriterien gelten, die entsprechend zu beachten sind. In den nachfolgenden Kapiteln wird darauf differenziert eingegangen. Wer diesen Lehrbrief studiert, soll in die Lage versetzt werden zu beurteilen, ob bei einem Testverfahren die jeweilige Strategie angemessen umgesetzt und die erforderlichen Gütekriterien eingehalten worden sind.

## **1.4 Bezugsnormorientierung bei Lehrkräften**

Im vorigen Abschnitt wurde bereits auf die verschiedenen Bezugsnormen eingegangen, die bei der Beurteilung einer Leistung herangezogen werden können.