Enzyklopädie der Psychologie

Leistungs-, Intelligenz- und Verhaltensdiagnostik

Psychologische Diagnostik

3

Hogrefe · Verlag für Psychologie
Göttingen · Bern · Toronto · Seattle

Enzyklopädie der Psychologie

Dieses Dokument ist nur für den persönlichen Gebrauch bestimmt und darf in keiner Form vervielfältigt und an Dritte weitergegeben werden.
Aus L.F. Hornke, M. Amelang & M. Kersting (Hrsg.): Leistungs-, Intelligenz- und Verhaltensdiagnostik (ISBN 9783840915260)
© 2011 Hogrefe Verlag, Göttingen.

ENZYKLOPÄDIE DER PSYCHOLOGIE

In Verbindung mit der
Deutschen Gesellschaft für Psychologie

herausgegeben von

Prof. Dr. Niels Birbaumer, Tübingen
Prof. Dr. Dieter Frey, München
Prof. Dr. Julius Kuhl, Osnabrück
Prof. Dr. Wolfgang Schneider, Würzburg
Prof. Dr. Ralf Schwarzer, Berlin

Themenbereich B

Methodologie und Methoden

Serie II

Psychologische Diagnostik

Band 3

Leistungs-, Intelligenz- und Verhaltensdiagnostik

Hogrefe • Verlag für Psychologie
Göttingen • Bern • Toronto • Seattle

Dieses Dokument ist nur für den persönlichen Gebrauch bestimmt und darf in keiner Form vervielfältigt und an Dritte weitergegeben werden.
Aus L.F. Hornke, M. Amelang & M. Kersting (Hrsg.): Leistungs-, Intelligenz- und Verhaltensdiagnostik (ISBN 9783840915260)
© 2011 Hogrefe Verlag, Göttingen.

Leistungs-, Intelligenz- und Verhaltensdiagnostik

herausgegeben von

Prof. em. Dr. Lutz F. Hornke, Aachen
Prof. em. Dr. Manfred Amelang, Heidelberg
Prof. Dr. Martin Kersting, Münster

Hogrefe • Verlag für Psychologie
Göttingen • Bern • Toronto • Seattle

Dieses Dokument ist nur für den persönlichen Gebrauch bestimmt und darf in keiner Form vervielfältigt und an Dritte weitergegeben werden.
Aus L.F. Hornke, M. Amelang & M. Kersting (Hrsg.): Leistungs-, Intelligenz- und Verhaltensdiagnostik (ISBN 9783840915260)
© 2011 Hogrefe Verlag, Göttingen.

© 2011 Hogrefe Verlag GmbH & Co. KG
Göttingen • Bern • Wien • Paris • Oxford • Prag • Toronto
Cambridge, MA • Amsterdam • Kopenhagen • Stockholm
Rohnsweg 25, 37085 Göttingen

http://www.hogrefe.de
Aktuelle Informationen • Weitere Titel zum Thema • Ergänzende Materialien

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.

Satz: ARThür Grafik-Design & Kunst, Weimar
Format: PDF

ISBN 978-3-8409-1525-3

Dieses Dokument ist nur für den persönlichen Gebrauch bestimmt und darf in keiner Form vervielfältigt und an Dritte weitergegeben werden.
Aus L.F. Hornke, M. Amelang & M. Kersting (Hrsg.): Leistungs-, Intelligenz- und Verhaltensdiagnostik (ISBN 9783840915260)
© 2011 Hogrefe Verlag, Göttingen.

Nutzungsbedingungen:

Der Erwerber erhält ein einfaches und nicht übertragbares Nutzungsrecht, das ihn zum privaten Gebrauch des E-Books und all der dazugehörigen Dateien berechtigt.

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

Der Nutzer ist nicht berechtigt, das E-Book – auch nicht auszugsweise – anderen Personen zugänglich zu machen, insbesondere es weiterzuleiten, zu verleihen oder zu vermieten.

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.

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 Audiodateien.

Dieses Dokument ist nur für den persönlichen Gebrauch bestimmt und darf in keiner Form vervielfältigt und an Dritte weitergegeben werden.
Aus L.F. Hornke, M. Amelang & M. Kersting (Hrsg.): Leistungs-, Intelligenz- und Verhaltensdiagnostik (ISBN 9783840915260)
© 2011 Hogrefe Verlag, Göttingen.

Autorenverzeichnis

Prof. em. Dr. Manfred Amelang

Universität Heidelberg
Psychologisches Institut
Hauptstraße 47
69117 Heidelberg
E-Mail: manfred.amelang@
psychologie.uni-heidelberg.de

Prof. Dr. Bernad Batinic

Johannes Kepler Universität Linz
Altenbergerstraße 69
4040 Linz
Österreich
E-Mail: bernad.batinic@jku.at

Prof. Dr. André Beauducel

Universität Hamburg
Fachbereich Psychologie
Differentielle Psychologie und
Psychologische Diagnostik
Von-Melle-Park 5
20146 Hamburg
E-Mail:
andre.beauducel@uni-hamburg.de

Prof. Dr. Klaus Bös

Institut für Sport und
Sportwissenschaft
Karlsruher Institut für Technologie
(KIT)
Kaiserstraße 12
76131 Karlsruhe
E-Mail: boes@kit.edu

Ass.-Prof. Dr. Pia Deimann

Institut für Entwicklungspsychologie
und Psychologische Diagnostik
Liebiggasse 5
1010 Wien
Österreich
E-Mail: pia.deimann@univie.ac.at

Prof. Dr. Joachim Funke

Universität Heidelberg
Psychologisches Institut
Hauptstraße 47
69117 Heidelberg
E-Mail: joachim.funke@psychologie.
uni-heidelberg.de

Dr. Timo Gnambs

Johannes Kepler Universität Linz
Altenbergerstraße 69
4040 Linz
Österreich
E-Mail: timo.gnambs@jku.at

PD Dr. Carmen Hagemeister

TU Dresden
Fachrichtung Psychologie
Professur Diagnostik und
Intervention
01062 Dresden
E-Mail: Carmen.Hagemeister@
tu-dresden.de

Dieses Dokument ist nur für den persönlichen Gebrauch bestimmt und darf in keiner Form vervielfältigt und an Dritte weitergegeben werden.
Aus L.F. Hornke, M. Amelang & M. Kersting (Hrsg.): Leistungs-, Intelligenz- und Verhaltensdiagnostik (ISBN 9783840915260)
© 2011 Hogrefe Verlag, Göttingen.

Prof. Dr. Guido Hertel

Westfälische Wilhelms-Universität
Münster
Institut für Psychologie
Fliednerstraße 21
48149 Münster
E-Mail: ghertel@uni-muenster.de

Prof. em. Dr. Lutz F. Hornke

Institut für Psychologie
RWTH Aachen University
Jägerstraße 17/19
52056 Aachen
E-Mail: lutz.hornke@psych.
rwth-aachen.de

Ass.-Prof. Dr.
Ursula Kastner-Koller

Institut für Entwicklungspsychologie
und Psychologische Diagnostik
Liebiggasse 5
1010 Wien
Österreich
E-Mail:
ursula.kastner-koller@univie.ac.at

Prof. Dr. Martin Kersting

Fachhochschule des Bundes
für öffentliche Verwaltung
Fachbereich Finanzen
Gescher Weg 100
48161 Münster
E-Mail: Martin@Kersting-internet.de

Prof. Dr. Klaus D. Kubinger

Institut für Entwicklungspsychologie
und Psychologische Diagnostik
Liebiggasse 5
1010 Wien
Österreich
E-Mail: klaus.kubinger@univie.ac.at

Dr. Jonas W. B. Lang

Maastricht University
Department of Work & Social
Psychology
Postbus 616
6200 MD Maastricht
Niederlande
E-Mail:
jonas.lang@maastrichtuniversity.nl

Prof. Dr. Thomas Rammsayer

Universität Bern
Institut für Psychologie
Abteilung Persönlichkeits-,
Differentielle Psychologie und
Diagnostik
Muesmattstraße 45
3012 Bern
Schweiz
E-Mail:
thomas.rammsayer@psy.unibe.ch

Prof. Dr. Bernd Reuschenbach

Katholische Stiftungsfachschule
München
Preysingstraße 83
81667 München
E-Mail:
bernd.reuschenbach@ksfh.de

Prof. Dr. Birgit Spinath

Psychologisches Institut –
Pädagogische Psychologie
Universität Heidelberg
Hauptstraße 47–51
69117 Heidelberg
E-Mail: birgit.spinath@psychologie.
uni-heidelberg.de

Dieses Dokument ist nur für den persönlichen Gebrauch bestimmt und darf in keiner Form vervielfältigt und an Dritte weitergegeben werden.
Aus L.F. Hornke, M. Amelang & M. Kersting (Hrsg.): Leistungs-, Intelligenz- und Verhaltensdiagnostik (ISBN 9783840915260)
© 2011 Hogrefe Verlag, Göttingen.

Prof. Dr. Jutta Stahl

Universität zu Köln
Professur für Differentielle Psychologie und Psychologische Diagnostik
Höninger Weg 115
50969 Köln
E-Mail: jutta.stahl@uni-koeln.de

Prof. Dr. Heinz-Martin Süß

Otto-von-Guericke-Universität Magdeburg
Institut für Psychologie I
Abteilung für Psychologische Methodenlehre, Psychodiagnostik und Evaluationsforschung
Universitätsplatz 2, Gebäude 24
39106 Magdeburg
E-Mail: heinz-martin.suess@ovgu.de

PD Dr. Susanne Tittlbach

Universität Bayreuth
Institut für Sportwissenschaft
95440 Bayreuth
E-Mail:
Susanne.Tittlbach@uni-bayreuth.de

Dr. Otto B. Walter

Institut für Psychologie
RWTH Aachen University
Jägerstraße 17/19
52056 Aachen
E-Mail:
walter@psych.rwth-aachen.de

Prof. Dr. Karl Westhoff

TU Dresden
Fachrichtung Psychologie
Professur Diagnostik und Intervention
01062 Dresden
E-Mail:
Karl.Westhoff@tu-dresden.de

Dieses Dokument ist nur für den persönlichen Gebrauch bestimmt und darf in keiner Form vervielfältigt und an Dritte weitergegeben werden.
Aus L.F. Hornke, M. Amelang & M. Kersting (Hrsg.): Leistungs-, Intelligenz- und Verhaltensdiagnostik (ISBN 9783840915260)
© 2011 Hogrefe Verlag, Göttingen.

Dieses Dokument ist nur für den persönlichen Gebrauch bestimmt und darf in keiner Form vervielfältigt und an Dritte weitergegeben werden.
Aus L.F. Hornke, M. Amelang & M. Kersting (Hrsg.): Leistungs-, Intelligenz- und Verhaltensdiagnostik (ISBN 9783840915260)
© 2011 Hogrefe Verlag, Göttingen.

Vorwort

Eine Enzyklopädie ist zunächst sehr vereinfacht viel Arbeit, die auf viele Köpfe und fleißige Hände verteilt ist. Dabei handelt es sich um eine besondere Textsorte, die mit einer ganz besonderen fachlichen Absicht erstellt wurde.

Seit dem 5. Jh. v. Chr. versteht man unter „Enzyklopädie" die universelle Bildung oder doch zumindest die Zusammenstellung derselben (siehe dazu den Eintrag in der wohl größten Enzyklopädie der Gegenwart, der Wikipedia, die zudem noch internetbasiert und mit vielen Querverweisen organisiert und allseits zugänglich ist). Enzyklopädisten meinten damit vom Wortstamm her zunächst einen geschlossenen Kreis an Wissen, und später stand die Idee des kreisförmig immer wiederkehrenden und zu erweiternden bzw. erneuernden Wissens im Vordergrund.

Sicher wird niemand von den vorliegenden Enzyklopädiebänden eine vollständige Darstellung von Fragen, Methodiken und Antworten der Psychologischen Diagnostik erwarten, wohl aber Hinweise auf neue Trends und Ansätze. Die Herausgeber haben versucht, dies zu ermöglichen und mit weitgefassten Vorgaben an die Autoren sowie entsprechenden Begutachtungen und Hinweisen die Grundfragen und -antworten der Psychologischen Diagnostik zu Beginn des 21. Jahrhunderts vorzustellen. Wenngleich andere, historische Enzyklopädien viel umfangreicher waren, so kann sich diese hier doch sehen lassen: 47 Artikel von 74 Autorinnen und Autoren.

Dass eine solche Arbeit mehrere Jahre in Anspruch nahm, verwundert nicht. Autoren und Herausgeber haben mit großem Elan begonnen, aber manche mussten zwischenzeitlich ihre anfänglichen Zeitplanungen angesichts ihrer auch vielfältigen anderen Aufgaben revidieren. Doch auch die traurigen und erfreulichen privaten und beruflichen Lebensereignisse verzögerten die Arbeit: Wissenschaftler, auch und gerade Psychologen, sind Menschen; Herausgeber haben das zu akzeptieren und hoffen, all dem Beruflichen und Privaten durch Benevolenz begegnet zu sein. Die Leser werden das sicher respektieren.

Jeder Beitrag ist wie bei Zeitschriften auch entsprechend begutachtet worden. Dabei war stets eine Rückmeldung anonym, und die Herausgeber haben sich entschlossen, offen ihre Korrekturen, Kommentare und Diskussionen anzu-

Dieses Dokument ist nur für den persönlichen Gebrauch bestimmt und darf in keiner Form vervielfältigt und an Dritte weitergegeben werden.
Aus L.F. Hornke, M. Amelang & M. Kersting (Hrsg.): Leistungs-, Intelligenz- und Verhaltensdiagnostik (ISBN 9783840915260)
© 2011 Hogrefe Verlag, Göttingen.

bringen. Hieraus ergab sich mit einigen Autoren ein sehr fruchtbarer Dialog; die Beiträge haben dadurch fraglos gewonnen, wenngleich die Autorinnen und Autoren jeweils den eigenen Beitrag selber verantworten, denn bei ihnen liegt letztlich die Expertise, wegen derer sie zur Mitarbeit eingeladen wurden.

Im Ergebnis ist durch die Aufgeschlossenheit und die Bemühungen aller eine Enzyklopädie der Psychologischen Diagnostik entstanden, die wie ihre Vorgänger fachlich über Fragen, Methoden und Antworten informiert. Auch wenn nicht alle anfangs geplanten Beiträge geschrieben werden konnten, so liegt doch nun ein Überblick vor, der allgemein Interessierte, Studierende und Fachwissenschaftler gleichermaßen, wenngleich aus unterschiedlicher Perspektive, ansprechen kann und soll. Zumindest hat die Psychologische Diagnostik wieder eine Plattform gefunden, von der aus Gegenwarts- und Zukunftsfragen gestellt und beantwortet werden können. Die hier vorliegende Enzyklopädie zeigt die Nähe aller Autorinnen und Autoren zueinander und zum Fach. Sie ist eine Art Zugehörigkeitsausweis; denn alle, die sich hier zu Wort melden, sind die Mitgestalter und -träger zukünftiger Entwicklungen in dieser Teildisziplin der Psychologie. Dass vieles zwar auch international mitteilbar ist, aber manches doch nur im deutschsprachigen Raum rezipiert wird, das liegt an der Natur des Gebietes: Diagnostik menschlichen Erlebens und Verhaltens geht zwar in seinen allgemeinen Grundfragen über den eigenen Sprachraum hinaus, aber die gesellschaftliche Anwendung vollzieht sich nun einmal dominant innerhalb einer (Sprach-) Kultur. Beides aber bedingt sich und ist allein genommen eher problematisch.

Nach allem gebührt aus Sicht der Herausgeber allen Autorinnen und Autoren der Dank, an dieser Enzyklopädieserie „Psychologische Diagnostik" mitgewirkt zu haben. Die vielen Leser werden ihren Dank dadurch abstatten, dass sie die Artikel zitieren, die Ideen aufgreifen, diskutieren und weiterentwickeln. Dazu gehört auch die Frage, ob die Textsorte Enzyklopädie zukünftig nicht doch auch einer anderen Präsentationsform oder einer Ergänzung durch diese bedarf.

Die Gesellschaft aber ist eingeladen, das Nachstehende zur Kenntnis zu nehmen und dort, wo es nicht unmittelbar verständlich ist, die Autorinnen und Autoren zu fragen, anzusprechen und sich mit ihnen auseinanderzusetzen. Gerade die vielen jungen Autorinnen und Autoren garantieren, dass fachlich „nachgehalten" werden kann. Nur so wird die eigentliche Absicht einer *Enkyklios Paideia* eingelöst: Psychologie und Psychologische Diagnostik im Besonderen ist der gesellschaftlichen Wirklichkeit und Aufklärung verpflichtet; hier entfaltet sie ihre vornehmsten Wirkungen.

Lutz F. Hornke, Aachen,
Manfred Amelang, Heidelberg und
Martin Kersting, Münster

Dieses Dokument ist nur für den persönlichen Gebrauch bestimmt und darf in keiner Form vervielfältigt und an Dritte weitergegeben werden.
Aus L.F. Hornke, M. Amelang & M. Kersting (Hrsg.): Leistungs-, Intelligenz- und Verhaltensdiagnostik (ISBN 9783840915260)
© 2011 Hogrefe Verlag, Göttingen.

Inhaltsverzeichnis

1. Kapitel: Testorientierte Leistungsdiagnostik: Grundlagen und Probleme, Verfahrensinventar und besondere Einsatzgebiete

Von Klaus D. Kubinger

Dieses Dokument ist nur für den persönlichen Gebrauch bestimmt und darf in keiner Form vervielfältigt und an Dritte weitergegeben werden.

Aus L.F. Hornke, M. Amelang & M. Kersting (Hrsg.): Leistungs-, Intelligenz- und Verhaltensdiagnostik (ISBN 9783840915260)

© 2011 Hogrefe Verlag, Göttingen.

2. Kapitel: Konzentrationsdiagnostik
Von Carmen Hagemeister und Karl Westhoff

Dieses Dokument ist nur für den persönlichen Gebrauch bestimmt und darf in keiner Form vervielfältigt und an Dritte weitergegeben werden.
Aus L.F. Hornke, M. Amelang & M. Kersting (Hrsg.): Leistungs-, Intelligenz- und Verhaltensdiagnostik (ISBN 9783840915260)
© 2011 Hogrefe Verlag, Göttingen.

3. Kapitel: Intelligenztests und ihre Bezüge zu Intelligenztheorien

Von Heinz-Martin Süß und André Beauducel

Dieses Dokument ist nur für den persönlichen Gebrauch bestimmt und darf in keiner Form vervielfältigt und an Dritte weitergegeben werden.
Aus L.F. Hornke, M. Amelang & M. Kersting (Hrsg.): Leistungs-, Intelligenz- und Verhaltensdiagnostik (ISBN 9783840915260)
© 2011 Hogrefe Verlag, Göttingen.

Dieses Dokument ist nur für den persönlichen Gebrauch bestimmt und darf in keiner Form vervielfältigt und an Dritte weitergegeben werden.
Aus L.F. Hornke, M. Amelang & M. Kersting (Hrsg.): Leistungs-, Intelligenz- und Verhaltensdiagnostik (ISBN 9783840915260)
© 2011 Hogrefe Verlag, Göttingen.

4. Kapitel: Wissensdiagnostik: Allgemeine und spezielle Wissenstests
Von André Beauducel und Heinz-Martin Süß

5. Kapitel: Entwicklungstests
Von Ursula Kastner-Koller und Pia Deimann

Dieses Dokument ist nur für den persönlichen Gebrauch bestimmt und darf in keiner Form vervielfältigt und an Dritte weitergegeben werden.
Aus L.F. Hornke, M. Amelang & M. Kersting (Hrsg.): Leistungs-, Intelligenz- und Verhaltensdiagnostik (ISBN 9783840915260)
© 2011 Hogrefe Verlag, Göttingen.

6. Kapitel: Schultests
Von Birgit Spinath

7. Kapitel: Funktionendiagnostik
Von Susanne Tittlbach und Klaus Bös

Dieses Dokument ist nur für den persönlichen Gebrauch bestimmt und darf in keiner Form vervielfältigt und an Dritte weitergegeben werden.
Aus L.F. Hornke, M. Amelang & M. Kersting (Hrsg.): Leistungs-, Intelligenz- und Verhaltensdiagnostik (ISBN 9783840915260)
© 2011 Hogrefe Verlag, Göttingen.

8. Kapitel: Grundlagen des adaptiven Testens
Von Otto B. Walter

Dieses Dokument ist nur für den persönlichen Gebrauch bestimmt und darf in keiner Form vervielfältigt und an Dritte weitergegeben werden.
Aus L.F. Hornke, M. Amelang & M. Kersting (Hrsg.): Leistungs-, Intelligenz- und Verhaltensdiagnostik (ISBN 9783840915260)
© 2011 Hogrefe Verlag, Göttingen.

9. Kapitel: Computer-adaptives Testen
Von Jonas W. B. Lang

10. Kapitel: Internetbasierte psychologische Diagnostik
Von Timo Gnambs, Bernad Batinic und Guido Hertel

Dieses Dokument ist nur für den persönlichen Gebrauch bestimmt und darf in keiner Form vervielfältigt und an Dritte weitergegeben werden.
Aus L.F. Hornke, M. Amelang & M. Kersting (Hrsg.): Leistungs-, Intelligenz- und Verhaltensdiagnostik (ISBN 9783840915260)
© 2011 Hogrefe Verlag, Göttingen.

11. Kapitel: Apparative Diagnostik
Von Thomas Rammsayer und Jutta Stahl

Dieses Dokument ist nur für den persönlichen Gebrauch bestimmt und darf in keiner Form vervielfältigt und an Dritte weitergegeben werden.
Aus L.F. Hornke, M. Amelang & M. Kersting (Hrsg.): Leistungs-, Intelligenz- und Verhaltensdiagnostik (ISBN 9783840915260)
© 2011 Hogrefe Verlag, Göttingen.

12. Kapitel: Ambulantes Assessment
Von Bernd Reuschenbach und Joachim Funke

Dieses Dokument ist nur für den persönlichen Gebrauch bestimmt und darf in keiner Form vervielfältigt und an Dritte weitergegeben werden.
Aus L.F. Hornke, M. Amelang & M. Kersting (Hrsg.): Leistungs-, Intelligenz- und Verhaltensdiagnostik (ISBN 9783840915260)
© 2011 Hogrefe Verlag, Göttingen.

13. Kapitel: Einsatz technischer Mittel in der psychologischen Diagnostik
Von Joachim Funke und Bernd Reuschenbach

Dieses Dokument ist nur für den persönlichen Gebrauch bestimmt und darf in keiner Form vervielfältigt und an Dritte weitergegeben werden.
Aus L.F. Hornke, M. Amelang & M. Kersting (Hrsg.): Leistungs-, Intelligenz- und Verhaltensdiagnostik (ISBN 9783840915260)
© 2011 Hogrefe Verlag, Göttingen.

Dieses Dokument ist nur für den persönlichen Gebrauch bestimmt und darf in keiner Form vervielfältigt und an Dritte weitergegeben werden.
Aus L.F. Hornke, M. Amelang & M. Kersting (Hrsg.): Leistungs-, Intelligenz- und Verhaltensdiagnostik (ISBN 9783840915260)
© 2011 Hogrefe Verlag, Göttingen.

1. Kapitel

Testorientierte Leistungsdiagnostik: Grundlagen und Probleme, Verfahrensinventar und besondere Einsatzgebiete

Klaus D. Kubinger

1 Einleitung

Psychologisch-diagnostische Verfahren werden zumeist eingeteilt in solche zur Persönlichkeitsdiagnostik und solche zur „Leistungsdiagnostik"; demgemäß spricht man bei letzteren von „Leistungstests". Allerdings ist dies unscharf: Einerseits wird vieles in der modernen Persönlichkeitsdiagnostik primär leistungsbezogen erfasst (vgl. dazu die sog. „Objektiven Persönlichkeits*tests sensu R. B. Cattell*" bei Kubinger, 2003b, bzw. Ortner, Proyer & Kubinger, 2006); andererseits sind die Ergebnisse von Leistungstests durch die Persönlichkeit der Testperson stark (mit-)beeinflusst. Und schließlich beobachten wir zwar aktuelle Leistungen, interessieren uns aber für überdauernde Fähigkeiten – d. h., die beobachtbare Leistung wird als Ausdruck einer eigentlich zu messen beabsichtigten, jedoch nicht direkt beobachtbaren Fähigkeit aufgefasst. Unter diesen Einschränkungen soll hier die Bezeichnung „Leistungstest" beibehalten und verstanden werden.

Ausgehend von der allgemeinen Definition aller psychologisch-diagnostischer Verfahren (vgl. z. B. Kubinger, 2009a) ist ein Leistungstest wie folgt definiert: Ein psychologischer Leistungstest erhebt unter standardisierten Bedingungen eine Verhaltensstichprobe über die Testperson, indem mit systematisch erstellten Aufgaben die interessierenden Verhaltensweisen oder psychischen Vorgänge ausgelöst, beobachtet und bewertet werden.

Dieses Dokument ist nur für den persönlichen Gebrauch bestimmt und darf in keiner Form vervielfältigt und an Dritte weitergegeben werden.
Aus L.F. Hornke, M. Amelang & M. Kersting (Hrsg.): Leistungs-, Intelligenz- und Verhaltensdiagnostik (ISBN 9783840915260)
© 2011 Hogrefe Verlag, Göttingen.

2 Einteilung von Leistungstests

Psychologische Leistungstests sind vornehmlich durch die Klasse der „Intelligenztests" bekannt. Wobei dem einzelnen Test jeweils eine spezifizierte Intelligenztheorie zugrunde liegt (zu einer Übersicht gängiger Intelligenztheorien vgl. z. B. Amelang, Bartussek, Stemmler & Hagemann, 2006; zu einer beispielhaften Verankerung eines bestimmten Intelligenztests innerhalb der von Intelligenztests vgl. Kubinger, Litzenberger & Mrakotsky, 2006); und zwar auch dann, wenn dem Test bloß ein pragmatischer Ansatz zugrunde liegt und (nur) versucht wird, „die Gesamtheit aller kognitiven Voraussetzungen" zu erfassen, „die notwendig sind, um Wissen zu erwerben und Handlungskompetenzen zu entwickeln" (Kubinger, 2009b, S. 23). Mit „Kognition" wird jeder Prozess gemeint, durch den sich der Mensch seiner Umwelt bewusst wird bzw. sich über sie Kenntnisse verschafft (also vor allem Wahrnehmen und Erkennen, Merken, Lernen und Urteilen sowie Vorstellen und Denken).

Zumeist handelt es sich bei Intelligenztests eigentlich um eine Zusammenstellung von mehreren Untertests zu einer Testbatterie. Über diese Intelligenz-Testbatterien hinaus gibt es spezielle Leistungstests, zum Beispiel für die Messung bestimmter neuropsychologischer Funktionstüchtigkeiten. Manche solcher speziellen Leistungstests sind insofern ebenfalls den Intelligenztests zuzurechnen, als sie einen einzelnen etablierten Intelligenzfaktor erfassen. Umgekehrt existieren auch Testbatterien mit mehreren Untertests, die nur auf spezielle Leistungen bzw. Fähigkeiten abzielen und mit Intelligenz im engeren Sinn weniger zu tun haben, wie zum Beispiel sogenannte „Konzentrationstests". Eine mögliche Einteilung der speziellen Leistungstests könnte sein:

- Sprachverständnis,
- räumliches Vorstellungsvermögen,
- Gedächtnis,
- Verarbeitungsgeschwindigkeit,
- logisches Denken,
- intellektuelle Lernfähigkeit,
- Aufmerksamkeit und Konzentration,
- technisches Verständnis.

Intelligenz-Testbatterien wie spezielle Leistungstests sind also dadurch gekennzeichnet, dass sie die kognitiven Fähigkeiten eines Menschen erfassen wollen – insofern wäre auch der Oberbegriff „Kognitionstest" angemessen.

Dieses Dokument ist nur für den persönlichen Gebrauch bestimmt und darf in keiner Form vervielfältigt und an Dritte weitergegeben werden.
Aus L.F. Hornke, M. Amelang & M. Kersting (Hrsg.): Leistungs-, Intelligenz- und Verhaltensdiagnostik (ISBN 9783840915260)
© 2011 Hogrefe Verlag, Göttingen.

3 Voraussetzungen von Tests zur Leistungsdiagnostik

Mit der *Psychologischen Diagnostik* hat sich ausgehend von der Testkonstruktionslehre Gulliksens (1950) die sogenannte „Klassische Testtheorie" entwickelt. Sie formuliert maßgebliche Haupt- und Nebengütekriterien, die für in der Praxis einsetzbare Tests gegeben sein müssen. Darunter fallen vor allem die bekannten testtheoretischen Ansprüche wie Objektivität, Reliabilität und Validität. Nicht nur wegen der methodischen Weiterentwicklung wurden die ursprünglichen Gütekriterien um andere erweitert, sondern vor allem, weil heutzutage der Konsumentenschutz auch für psychologische Tests relevant ist. Das heißt, der Konsument (der Begutachtete, der Klient oder Patient, die Testperson) sollte die Garantie haben, fachgemäß psychologisch behandelt und begutachtet zu werden. Die gegenwärtig wohl bedeutendste rechtsnahe Vorgabe ist die DIN 33430 (DIN Deutsches Institut für Normung e. V., 2002). Sie regelt konkret die „Anforderungen an Verfahren und deren Einsatz bei der berufsbezogenen Eignungsbeurteilung", wobei sie sich in wesentlichen Teilen auf die bereits angesprochenen Gütekriterien bezieht (vgl. Westhoff et al., 2005).

Ein weiteres, relativ neues Gütekriterium ist die der „Modernen Testtheorie" (Kubinger, 1989b) entstammende Skalierung, die u. a. Tests zur Leistungsdiagnostik erfüllen sollten (vgl. Testkuratorium, 2006). Hierunter versteht man (vgl. z. B. Kubinger, 2009a, S. 82): „Ein Test erfüllt das Gütekriterium Skalierung, wenn die laut Verrechnungsvorschriften resultierenden Testwerte die empirischen Verhaltensrelationen adäquat abbilden." Gemeint ist damit, dass die vom Testautor festgelegte Regel, wie die einzelnen Testleistungen bzw. -reaktionen einer Testperson zu einem numerischen Testwert zu verrechnen sind, „angemessen" insofern ist, als sie im Sinn der Theorie des Messens faktisch gerechtfertigt und nicht nur „apodiktisch" bzw. willkürlich festgesetzt ist.

Die angesprochene Moderne Testtheorie oder „Item-Response-Theorie" (IRT) bietet nun Erkenntnisse und mathematisch-statistische Modelle, die im Wesentlichen auf eindimensionale Messungen abzielen (vgl. eine darüber hinaus gehende mehrdimensionale Messung, insbesondere das sog. *multidimensional Rasch model* bei Adams, Wilson & Wang, 1997). Freilich versuchen die meisten Leistungstests der *Psychologischen Diagnostik*, (je Untertest) eindimensional zu messen. Dies ist nicht nur daraus abzuleiten, dass die Messintentionen von Tests regelmäßig „eindimensional" begrifflich benannt/beschrieben werden (z. B. „schlussfolgernde Denkfähigkeit"), sondern auch daraus, dass bei der Interpretation der erhaltenen Testwerte es regelmäßig um das Abwägen als „mehr" oder „weniger" geht. Bei der Personalauswahl zum Beispiel geht es

Dieses Dokument ist nur für den persönlichen Gebrauch bestimmt und darf in keiner Form vervielfältigt und an Dritte weitergegeben werden.
Aus L.F. Hornke, M. Amelang & M. Kersting (Hrsg.): Leistungs-, Intelligenz- und Verhaltensdiagnostik (ISBN 9783840915260)
© 2011 Hogrefe Verlag, Göttingen.

darum, einen möglichst hohen Testwert für eine positive Entscheidung zugrunde zu legen und einen zu niedrigen Testwert als Indiz für zu geringe berufliche Eignung anzusehen. Umgekehrt ist es offensichtlich kritisch, wenn die bei einem Test zu erbringende Leistung nicht nur von einer einzigen, bestimmten Fähigkeit abhängt, sondern auch noch von anderen.

Dann ergibt sich das Problem, dass der berechnete Testwert für den Anwender oft wertlos ist, und somit die gegebene psychologisch-diagnostische Fragestellung nicht beantwortet werden kann. Soll ein Test „Lernfähigkeit" erfassen, aber misst im hohen Grad auch „Lesefähigkeit" (weil es sich um syntaktisch besonders anspruchsvolle Texte handelt), dann ist ein niedriger Testwert nicht eindeutig zu interpretieren: Entweder die Testperson verfügt über eine geringe Lernfähigkeit oder sie verfügt über eine geringe Lesefähigkeit; oder es mangelt ihr gar an beidem. Misst zum Beispiel ein Test mit der Messintention „Konzentrationsfähigkeit" im hohen Grad auch „Instruktionsverständnis/sprachliches Auffassungsvermögen" (weil es sich um semantisch besonders anspruchsvolle Texte handelt), dann ist ein niedriger Testwert ebenfalls nicht eindeutig zu interpretieren: Entweder die Testperson verfügt über eine geringe Konzentrationsfähigkeit oder sie verfügt über ein geringes Instruktionsverständnis; oder es mangelt ihr gar an beidem. Eindimensionalität der gemessenen Fähigkeit ist also fast immer eine wesentliche Voraussetzung von Tests zur Leistungsdiagnostik.

Aufbauend auf dem Konzept eindimensionalen Messens geht es nun innerhalb der IRT im Zusammenhang mit dem Gütekriterium Skalierung darum, je Verrechnungsvorschrift ein Modell anzubieten, welches einen zur Diskussion stehenden psychologischen Test in Bezug auf die oben angesprochene „Angemessenheit" zu prüfen erlaubt. Das bedeutet: Modelle, die erstens empirisch prüfbar sind und die zweitens mindestens eine hinreichende, besser die notwendige Bedingung darstellen, die erfüllt sein muss, damit die jeweilige Verrechnungsvorschrift fair ist – also die empirisch gegebenen Verhaltensrelationen bei den einzelnen Testaufgaben sowohl innerhalb einer Testperson als auch zwischen verschiedenen Testpersonen vermittels der resultierenden Testwerte adäquat abgebildet werden. Einen Katalog solcher Modelle bzw. Verrechnungsvorschriften gibt Kubinger (1989a). Im Folgenden soll es genügen, ein einziges solches Modell näher zu betrachten. Dieses ist das sogenannte *Rasch-Modell* und bezieht sich auf die Verrechnungsvorschrift: Testwert ist gleich Anzahl gelöster Aufgaben. Es handelt sich dabei sowohl um den einfachsten Fall einer Verrechnungsvorschrift als auch um den am weitesten verbreiteten.

Das Rasch-Modell beschreibt die Wahrscheinlichkeit, dass die Testperson v Item i löst („+"), in Abhängigkeit eines (eben eindimensionalen) Personenpara-

Dieses Dokument ist nur für den persönlichen Gebrauch bestimmt und darf in keiner Form vervielfältigt und an Dritte weitergegeben werden.
Aus L.F. Hornke, M. Amelang & M. Kersting (Hrsg.): Leistungs-, Intelligenz- und Verhaltensdiagnostik (ISBN 9783840915260)
© 2011 Hogrefe Verlag, Göttingen.

meters, ξ_v, als der (wahren) Fähigkeit von *v*, und eines (gleichfalls eindimensionalen) Aufgabenparameters, σ_i, als der (wahren) Schwierigkeit von *i*. Das heißt, eine bestimmte Fähigkeit ξ_v bedingt nicht deterministisch, ob es zu einer Lösung kommt oder nicht, sondern nur probabilistisch in der Hinsicht, dass die Lösungswahrscheinlichkeit für größere ξ, bei konstantem σ, ebenfalls größer wird. Konkret lautet das Modell:

$$P\left(+\left|\xi_v,\sigma_i\right.\right)=\frac{e^{\xi_v-\sigma_i}}{1+e^{\xi_v-\sigma_i}}. \qquad (1)$$

Fischer (1995) hat bewiesen, dass dieses Modell notwendigerweise gelten muss, damit die genannte Verrechnungsvorschrift fair ist. Die Tragweite des Beweises liegt im Attribut „notwendigerweise". Mit ihm ist nämlich der Umkehrschluss zwingend, dass Tests nicht verrechnungsfair sind, für die das Rasch-Modell nicht gilt. Möglich wäre es zwar, dass schlicht ein anderer Verrechnungsmodus verhaltensadäquat ist, zum Beispiel wenn die einzelnen Antworten aufgabenspezifisch gewichtet würden; bezüglich des festgelegten Verrechnungsmodus: Testwert ist gleich Anzahl gelöster Aufgaben, sind sie es jedenfalls dann nicht. Und hier ist es auch tatsächlich möglich, die Geltung des Modells empirisch zu prüfen – im Gegensatz zu vielen anderen (testtheoretischen) Modellen, für die sogenannte *Goodness-of-fit*-Tests lediglich feststellen, wie gut die beobachteten Daten durch das Modell erklärt werden können. Wie dabei im Sinne eines Standards Modelltests zum Rasch-Modell vorzugehen ist, zeigt Kubinger (2005).

4 Gestaltungsweisen von Tests zur Leistungsdiagnostik

Grundsätzlich können die Tests zur Leistungsdiagnostik nach vier Gesichtspunkten eingeteilt werden:

1. freies Antwortformat versus Multiple-Choice-Format,
2. Power- versus Speed-and-Power-Test,
3. Gruppen- versus Individualtest,
4. Papier-Bleistift-Test versus Computertest.

4.1 Antwortformat

Grundsätzlich ist das sogenannte „freie Antwortformat", manchmal auch als „offenes" Antwortformat bezeichnet, leicht zu beschreiben: Die Testperson reagiert auf die gestellte Aufgabe beim Lösungsversuch meist mit einer münd-

Dieses Dokument ist nur für den persönlichen Gebrauch bestimmt und darf in keiner Form vervielfältigt und an Dritte weitergegeben werden.
Aus L.F. Hornke, M. Amelang & M. Kersting (Hrsg.): Leistungs-, Intelligenz- und Verhaltensdiagnostik (ISBN 9783840915260)
© 2011 Hogrefe Verlag, Göttingen.

lichen, zumindest sprachlichen Aussage in mehr oder weniger detaillierter selbstgewählter Ausführlichkeit. Es sind aber auch nonverbale Tätigkeiten, zum Beispiel mathematische Ausarbeitungen oder gewisse (fein-)motorische Aktionen denkbar. Das gezeigte Verhalten muss der Testleiter dann nach vorgegebenen Bewertungsregeln verrechnen, und zwar wieder am einfachsten nach „richtig" oder „falsch". Zur Illustration gelte folgendes Item: *„Warum sollte man Obst vor dem Essen waschen?"* (Lösungen: „um Schmutz abzuwaschen" oder „um Pflanzenschutzmittel abzuwaschen").

Allerdings gibt es eine besondere Variante des freien Antwortformats, die sehr einfach zu verrechnen ist: Kubinger, Holocher-Ertl und Frebort (2006) bezeichnen es als „Kästchen"-Format, wenn bei einer Testaufgabe das numerische Ergebnis einer Berechnung in vorgegebene, dem Stellenwert entsprechende Kästchen eingetragen werden muss (prinzipiell möglich ist auch das Eintragen von kurzen Wörtern). Werden die Kästchen als Lesebereich definiert, ist eine Auswertung auch über Scanner bzw. Computer möglich. Als Beispiel sei folgendes Item entworfen: *„Wie lautet die um 2 größere Zahl von 29?"*

Als „gebundenes" Antwortformat wird das bekannte Multiple-Choice-Format in all seinen Variationen bezeichnet: Aus einer Liste von Antwortmöglichkeiten (Lösungsvorschlägen) muss die Testperson auswählen, welche sie davon für richtig erachtet. Herkömmlich beinhaltet eine Aufgabe im Multiple-Choice-Format eine einzige Lösung, die übrigen Antwortmöglichkeiten nennt man „Distraktoren": Letztere sollen der Lösung möglichst nahe kommen, und trotzdem in gewisser Hinsicht falsch sein. Außerhalb der *Psychologischen Diagnostik* wird das Multiple-Choice-Format zumeist in Form von vier Antwortmöglichkeiten benutzt, eine Lösung samt drei Distraktoren. Um erfolgreiches Raten seitens der Testperson wenig wahrscheinlich zu machen, verwendet man allerdings innerhalb der *Psychologischen Diagnostik* besser mehr Antwortmöglichkeiten.

Damit die Rateeffekte so klein wie möglich gehalten werden, gibt es die Variante, unter den Antwortmöglichkeiten mehr als nur eine einzige Lösung zu bieten. Kubinger, Holocher-Ertl und Frebort (2006) haben vorgeschlagen, im Gegensatz zum klassischen Multiple-Choice-Format „1 aus 5" oder „1 aus 6", „2 aus 5" oder „x aus 5" zu verwenden. Gemeint ist in den beiden ersten Fällen, dass eine Antwort der fünf bzw. sechs Antwortmöglichkeiten richtig ist, bzw. bei den letzten Fällen, das entweder genau zwei der fünf Antwortmöglichkeiten richtig sind (und die Testperson darüber explizit informiert ist) oder mehrere der fünf Antwortmöglichkeiten richtig sind: Keine Antwortmöglichkeit, eine, zwei, drei, vier oder sogar alle fünf Antwortmöglichkeiten (worüber die Testperson gleichfalls informiert sein muss). Zur Illustration sei das Item in Kasten 1 entworfen.

Dieses Dokument ist nur für den persönlichen Gebrauch bestimmt und darf in keiner Form vervielfältigt und an Dritte weitergegeben werden.
Aus L.F. Hornke, M. Amelang & M. Kersting (Hrsg.): Leistungs-, Intelligenz- und Verhaltensdiagnostik (ISBN 9783840915260)
© 2011 Hogrefe Verlag, Göttingen.

Kasten 1:
Beispielitem mit dem Multiple-Choice-Format „x aus 5“
(in Klammer sind die richtigen Antwortmöglichkeiten mit Häkchen gekennzeichnet)

An Deutschland grenzt/grenzen:

a) Slowenien ( )
b) Polen (✓)
c) Italien ( )
d) Österreich (✓)
e) Niederlande (✓)

4.2 Zeitbegrenzung

Sogenannte „Power-Tests“ messen eine bestimmte (kognitive) Fähigkeit ohne Zeitbegrenzung, die die Leistung beeinflussen könnte; „Speed-Tests“ dagegen messen als Fähigkeit die Verarbeitungsgeschwindigkeit, ohne anspruchsvolle (sonstige) kognitive Anforderung. „Speed-and-Power-Tests“ erfassen sowohl eine bestimmte (kognitive) Fähigkeit als auch (Verarbeitungs-)Geschwindigkeit, d. h. sie stellen anspruchsvolle Leistungsanforderungen unter Zeitdruck. Genau genommen ist dabei zu differenzieren zwischen solchen Speed-and-Power-Tests, bei denen die insgesamte Bearbeitungszeit beschränkt ist, so dass nicht alle Testpersonen tatsächlich alle Items bearbeiten, und solchen, bei denen jedes einzelne Item mit einer Zeitbegrenzung versehen ist, d. h., jede Testperson bearbeitet das Item zeitlich begrenzt, ohne es notwendigerweise auch (richtig oder falsch) zu beantworten. Als Beispiel sei das in Abbildung 1 dargestellte Item aus dem AID 2 (Kubinger, 2009b) angeführt: *„Wenn du (alle) diese Teile richtig aneinandersetzt, wird eine Figur daraus …“* (verfügbare Lösungszeit für eine Verrechnung von 1 Punkt: 1 Minute 30 Sekunden, für 2 Punkte: 20 Sekunden).

Abbildung 1:
Item „①-1“ des Untertests *Antizipieren und Kombinieren*-figural aus dem AID 2;
als Lösungsfigur ergibt sich „Teddybär“

Dieses Dokument ist nur für den persönlichen Gebrauch bestimmt und darf in keiner Form vervielfältigt und an Dritte weitergegeben werden.
Aus L.F. Hornke, M. Amelang & M. Kersting (Hrsg.): Leistungs-, Intelligenz- und Verhaltensdiagnostik (ISBN 9783840915260)
© 2011 Hogrefe Verlag, Göttingen.

4.3 Sozialform

Trivialerweise können Gruppentests gleichzeitig, auch ohne Nutzung eines Computers mehreren Testpersonen vorgegeben werden. Bei Individualtests gilt, dass die besondere Vorgabemodalität bzw. besondere Materialien eine Testvorgabe an mehr als eine einzige Testperson gleichzeitig praktisch nicht ermöglichen. Computertests sind in gewisser Weise als Individualtests aufzufassen, auch wenn mehrere Computer-Testplätze in ein und demselben Raum verfügbar sind, und mehr als nur eine einzige Testperson getestet wird. Ein wesentlicher Vorteil der Individualtestung ist nämlich hier erfüllt, nämlich die Möglichkeit, individuell verschiedene Tests vorzugeben und nicht zwingend mit jedem Test zum selben Zeitpunkt zu beginnen. Was Gruppentests betrifft, gilt erfahrungsgemäß, dass pro Testleiter nicht mehr als sechs bis höchstens acht Testpersonen sachgemäß getestet werden können; dies trifft wohl auch für Computer-Testplatzsysteme zu (vgl. zur Illustration die Raumanordnung eines in der Praxis bewährten Computer-Testplatzsystems in Abb. 2).

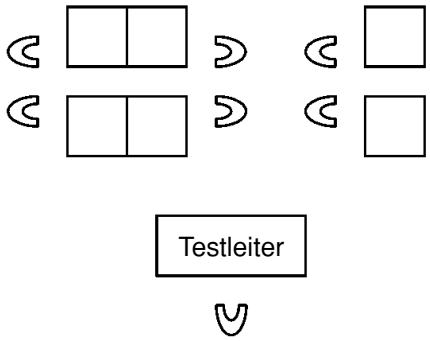

Abbildung 2:
Raumanordnung eines in der Praxis bewährten Computer-Testplatzsystems.

4.4 Technik

Papier-Bleistift-Tests und Computertests unterscheiden sich ganz offensichtlich und selbst bei solchen Tests, die ursprünglich zur Testung mit Papier und Bleistift vorgesehen waren, kann die Übertragung auf den Computer psychologisch relevant abweichende Ergebnisse bedingen; sogenannte Äquivalenzprüfungen (vgl. Wagner-Menghin, 2003a) sind bei solchen Übertragungen unabdingbar. Prinzipiell gibt es zwischen Papier-Bleistift-Tests und Computertests ähnliche Unterschiede wie zwischen Gruppentests und Einzeltests. Dies ist nämlich dadurch bedingt, dass nicht jede Vorgabemodalität mit beiden Techniken realisierbar ist, nicht jedes Testmaterial am Computer umsetzbar ist, und nicht jede Population mit beiden Techniken zumutbar getestet werden kann (vgl. das Gütekriterium Zumutbarkeit z. B. bei Kubinger, 2009a). Zur Illustration sei der

Dieses Dokument ist nur für den persönlichen Gebrauch bestimmt und darf in keiner Form vervielfältigt und an Dritte weitergegeben werden.
Aus L.F. Hornke, M. Amelang & M. Kersting (Hrsg.): Leistungs-, Intelligenz- und Verhaltensdiagnostik (ISBN 9783840915260)
© 2011 Hogrefe Verlag, Göttingen.

Test *Signal Detection* (Schuhfried, 1986) angeführt, bei dem jeweils dann eine Reaktionstaste gedrückt werden muss, wenn in der laufend wechselnden, komplexen Punktkonfiguration ein Quadrat bestimmter Größe auftaucht (vgl. Abb. 3).

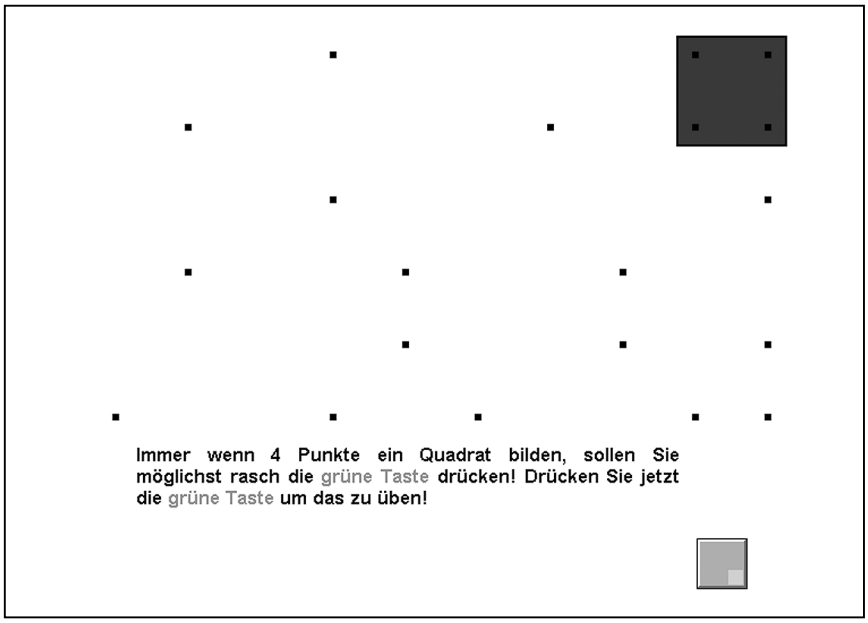

Abbildung 3:
Screenshot aus der Instruktion zum Test *Signal Detection*
(der Abdruck erfolgt mit freundlicher Genehmigung von Schuhfried GmbH)

Grundsätzlich sind die verschiedenen Gestaltungsweisen nicht beliebig kombinierbar. Ein freies Antwortformat (mit Ausnahme des Kästchen-Formats, vgl. Abschnitt 4.1) findet sich heutzutage noch (!) eher nur bei (Papier-Bleistift-)Individualtests, kaum bei Gruppen- und Computertests, bei welchen dieses Antwortformat schlicht unpraktisch weil unökonomisch ist. Umgekehrt verbietet sich das Multiple-Choice-Format eigentlich für (Papier-Bleistift-)Individualtests; wenn sich schon ein Testleiter ausschließlich einer einzigen Testperson widmen muss, dann sollte auch die regelmäßig beim freien Antwortformat zusätzlich anfallende Information für den diagnostischen Prozess genutzt werden. (Papier-Bleistift-)Gruppentests wiederum machen Speed-and-Power-Tests (bei denen die insgesamte Bearbeitungszeit beschränkt ist) deshalb praktisch notwendig, selbst wenn man ursprünglich an der Messung einer bestimmten (kognitiven) Fähigkeit ohne leistungsbeeinflussende Zeitbegrenzung interessiert ist, weil dabei, anders als bei Individual- und Computertests, ein individuelles Zeitbudget bei der Testbearbeitung schwerlich zugestanden werden kann. Allerdings erfordern wiederum solche Speed-and-Power-Tests, bei denen jedes einzelne Item mit einer Speed-Komponente versehen ist, im Fall eines Papier-Bleistift-Tests auch einen Individualtest, weil anders die Zeitkontrolle nicht möglich ist.

Dieses Dokument ist nur für den persönlichen Gebrauch bestimmt und darf in keiner Form vervielfältigt und an Dritte weitergegeben werden.
Aus L.F. Hornke, M. Amelang & M. Kersting (Hrsg.): Leistungs-, Intelligenz- und Verhaltensdiagnostik (ISBN 9783840915260)
© 2011 Hogrefe Verlag, Göttingen.

5 Technische Möglichkeiten und Grenzen der Leistungsdiagnostik

Die Praxis *psychologischen Diagnostizierens* läuft manchmal an den Konzepten und Erkenntnissen der *Psychologischen Diagnostik*, insbesondere ihrer Grundlagenforschung, vorbei. Was die eben angesprochenen unterschiedlichen Gestaltungsweisen betrifft, wird darauf weiter unten nochmals eingegangen. Davon abgesehen finden sich aber in der geübten Praxis etliche Unzulänglichkeiten, die zu irreführenden Interpretationen führen; das betrifft folgende Themenbereiche:

- Messgenauigkeit,
- Ökologische Validität,
- Stabilität des gemessenen Merkmals,
- Globalisierungsanspruch,
- Anforderungsmodell.

5.1 Messgenauigkeit

Zwar ist die Messgenauigkeit grundlegender Ausbildungsstoff wohl in jedem Studiengang Psychologie, trotzdem ist ihre kritische Reflexion/Bewertung kaum in psychologischen Gutachten der Praxis zu finden.

Innerhalb der Klassischen Testtheorie wird bekanntlich der „Standardmessfehler" über die geeignet ermittelte Reliabilität eines Tests bestimmt. Damit kann ein Konfidenzintervall für den wahren, gesuchten Messwert *(ω)* der Testperson errechnet werden. Nach üblicher Festlegung beinhaltet dann in 95 % der Fälle das solcherart berechnete Konfidenzintervall den gesuchten, wahren Messwert irgendwo in diesem Intervall; in 5 % der Fälle liegt der wahre Wert aber außerhalb. Mit ϱ als der Reliabilität eines Tests und damit 1-ϱ als der „Unreliabilität" sowie mit σ als der Standardabweichung der geeichten Testwerte ergibt sich der angesprochene Standardmessfehler *(δ)* als $\sigma \times \sqrt{1-\varrho}$. Das fragliche Konfidenzintervall für die angedachte Irrtumswahrscheinlichkeit von $\alpha = .05$ bestimmt sich dann für einen beobachteten Testwert *(X)* wie folgt:

$$\omega_{\text{min, max}} = X \pm 1.96 \times \delta \approx X \pm 2\delta. \qquad (2)$$

In der Praxis wird regelmäßig übersehen, dass dieses Konfidenzintervall für übliche Testreliabilitäten extrem groß und daher interpretationsbezogen ziemlich uninformativ ist (vgl. z. B. Kubinger, 2009a): Für einen Test mit einer Reliabilität $\varrho = .80$ folgt etwa in IQ-Werten (mit $\sigma = 15$) bei einem beobachteten Testwert $IQ = 100$, dass wegen $\omega_{\text{min, max}} \approx X \pm 2\delta = 100 \pm 2 \times 15 \times \sqrt{1-.80} = 100 \pm 13.4$ der wahre Messwert ω im Intervall von 86.6 bis 113.4 angenommen werden muss

Dieses Dokument ist nur für den persönlichen Gebrauch bestimmt und darf in keiner Form vervielfältigt und an Dritte weitergegeben werden.
Aus L.F. Hornke, M. Amelang & M. Kersting (Hrsg.): Leistungs-, Intelligenz- und Verhaltensdiagnostik (ISBN 9783840915260)
© 2011 Hogrefe Verlag, Göttingen.

– und das in 5 von 100 Fällen auch noch falsch ist. Da bekanntlich die IQ-Werte extra so geeicht sind, dass zwischen $IQ = 90$ und $IQ = 110$ die leistungsbezogen mittleren 50 % einer Population zu finden sind, ist die Aussage „der wahre Messwert liegt zwischen 86.6 bis 113.4" ziemlich „leer". Sie beinhaltet nämlich nicht nur die (konventionelle) Interpretation, dass der wahre Messwert als „durchschnittlich" bezeichnet werden kann, sondern auch die Möglichkeit, dass der zu $IQ = 100$ gehörige wahre Messwert als unterdurchschnittlich (unter dem Prozentrang 25) zu beurteilen ist, wie sogar die Möglichkeit, dass der wahre Messwert als überdurchschnittlich (über dem Prozentrang 75) zu beurteilen ist. Für den analogen Fall einer Testreliabilität von $\varrho = .95$ ergibt sich übrigens ein Konfidenzintervall von 93.3 bis 106.7. Wagner-Menghin (2003b) bietet genauer eine Grafik, ab welcher Reliabilität das Konfidenzintervall um einen exakt durchschnittlichen beobachteten Testwert nicht mehr den Prozentrang 25 bzw. 75 unter- bzw. überschreitet: .87. Aus der Grafik ist auch zu entnehmen, dass für einen Test mit Reliabilität $\varrho = .87$ der IQ-Wert $IQ \geq 122.5$ betragen müsste, um unter dem einkalkulierten Risiko von 5 % (nämlich dass diese Schlussfolgerung falsch ist) den wahren Testwert als überdurchschnittlich interpretieren zu können – oder $IQ \leq 77.5$, um den wahren Testwert als unterdurchschnittlich zu interpretieren.

Für die Praxis bedeutet das aber, dass Testwerte von wenig reliablen Tests fahrlässig unsachlich interpretiert werden. Die Reliabilität eines Tests ist also keine Kenngröße, deren Abweichung vom Ideal $\varrho = 1$ beliebig toleriert werden kann.

Ein methodischer Zugang, um bei weniger reliablen Tests eher trotzdem aussagekräftige Interpretationen zu erhalten, ist der hypothesengeleitete Ansatz, nur einseitige Konfidenzintervalle im Blick auf eine konkrete Entscheidungsfindung zu bestimmen. Bei gleichen Risiko (nämlich dass die letztlich getroffene Schlussfolgerung falsch ist), wird dabei die Intervallbreite im relevanten Testwertebereich kleiner. Für das angesprochene Risiko von wieder 5 % ergibt sich zum Beispiel $\omega_{\max} = X + 1.645 \times \delta \approx X + 1.5 \times \delta = 85 + 1.5 \times 15 \times \sqrt{1 - .87} = 93$. Damit sind etwa Fragestellungen beantwortbar, ob trotz unterdurchschnittlich beobachteten Testwerts (im Beispiel $IQ = 85$) der wahre Testwert als durchschnittlich (im Sinn eines Prozentrangs größer als 25) bezeichnet werden kann.

Wenig empfehlenswert ist dagegen die jüngst manchmal feststellbare Vorgehensweise, das eingegangene Risiko (nämlich für eine letztlich falsch getroffene Schlussfolgerung) zu erhöhen, zum Beispiel auf 10 % – statt $\omega_{\text{min, max}} \approx X \pm 2\delta = 100 \pm 2 \times 15 \times \sqrt{1 - .80} = 100 \pm 13.4$ zum Beispiel ergäbe sich dabei nämlich $\omega_{\text{min, max}} = X \pm 1.645 \times \delta \approx 100 \pm 1.5 \times 15 \times \sqrt{1 - .80} = 100 \pm 8.1$. Es scheint kaum verantwortlich, durchschnittlich jede zehnte Testperson bei einer solchen Vorgehensweise in Bezug auf ihren wahren Testwert fälschlich zu interpretieren.

Dieses Dokument ist nur für den persönlichen Gebrauch bestimmt und darf in keiner Form vervielfältigt und an Dritte weitergegeben werden.
Aus L.F. Hornke, M. Amelang & M. Kersting (Hrsg.): Leistungs-, Intelligenz- und Verhaltensdiagnostik (ISBN 9783840915260)
© 2011 Hogrefe Verlag, Göttingen.

Besonders kritisch wird die Missachtung der Messgenauigkeit bei einer sogenannte „Profilinterpretation". Dabei geht es darum, die einzelnen Untertestleistungen einer Testperson zueinander in Relation zu setzen, also „Höhen" und „Tiefen" zu bestimmen. Offensichtlich gibt es bei wenig reliablen Untertests kaum Testwerte, die sich, kalkuliert man den wahrscheinlichen Messfehler mit ein, tatsächlich unterscheiden. Es gibt andernorts (vgl. z. B. Huber, 1973) publizierte Formeln, um die „Signifikanz" eines Testprofils zu bestimmen, d. h. auszurechnen, ob sich unter Miteinbeziehung des Messfehlers tatsächlich mindestens zwei Testwerte unterscheiden. Diese Formeln beruhen allerdings auf sehr strikten Annahmen, die kaum je praktisch erfüllt sind, so dass der oben angesprochene hypothesengeleitete Ansatz unter Berechnung einseitiger Konfidenzintervalle in Bezug auf bestimmte vermutete Höhen und Tiefen pragmatisch vorzuziehen ist.

Hinsichtlich des möglichen Messfehlers eines Tests ist jedoch über alles bisher Gesagte hinaus überhaupt der Ansatz der Klassischen Testtheorie grundsätzlich kritisch. Er beruht, wie ausgeführt, auf dem Konzept der sogenannte „Reliabilität", also im Wesentlichen auf einer Korrelation, wie sie in einer (mehr oder weniger für die fragliche Population repräsentativen) Stichprobe bestimmt wurde. Abgesehen davon, dass die Korrelation als eine statistische Methode extrem stichprobenabhängig ist (vgl. Kubinger, 2009a), ist es anschaulich sehr unplausibel anzunehmen, dass die Messgenauigkeit eines Tests für alle Testwerte im gesamten Testwertbereich gleich ist, wie das eben die Klassische Testtheorie mit ihrer Bestimmung des Standardmessfehlers annimmt. Vielmehr ist es erfahrungsgeleitet wahrscheinlich, dass im oberen und unteren Testwertbereich wesentlich ungenauer gemessen wird, weil dort jeweils weit weniger Items zwischen ähnlichen Fähigkeitsausprägungen differenzieren können. Dem gegenüber ist der entsprechende Ansatz der IRT bestechend: Dieser ermöglicht einen sogenannten „(Standard-)Schätzfehler" spezifisch, pro konkret erbrachter Leistung einer Testperson zu berechnen: Geht es zum Beispiel beim Rasch-Modell darum, den (unbekannten) Fähigkeitsparameter ξ_v zu schätzen (vgl. Abschnitt 3), so ist über *Fisher*s Theorie der *„information in the sample"* für die diesbezügliche Maximum-Likelihood-Schätzung zunächst die person- und itemspezifische „Information" $I(i,v) = \frac{[P'("+"|i,v)]^2}{P("+"|i,v).\ P("-"|i,v)}$ für Person v und Item i zu bestimmen, und daraus der Standardschätzfehler für ξ_v:

$$\varepsilon(\xi_v) = \sqrt{\frac{1}{\sum_i^k I(i,v)}} = \left[\sum_i^k \frac{e^{\xi_v-\sigma_i}}{1+e^{\xi_v-\sigma_i}} \times \frac{1}{1+e^{\xi_v-\sigma_i}}\right]^{-\frac{1}{2}}, \qquad (3)$$

Dieses Dokument ist nur für den persönlichen Gebrauch bestimmt und darf in keiner Form vervielfältigt und an Dritte weitergegeben werden.
Aus L.F. Hornke, M. Amelang & M. Kersting (Hrsg.): Leistungs-, Intelligenz- und Verhaltensdiagnostik (ISBN 9783840915260)
© 2011 Hogrefe Verlag, Göttingen.