



Universität Zürich
Psychologisches Institut,

Zürich, 21.07.2006

Mag. René Proyer

Fachrichtung Persönlichkeitspsychologie und Diagnostik

Universität Zürich

Binzmühlestrasse 14/7

8050 Zürich

r.proyer@psychologie.unizh.ch

**Gutachten zum aktuellen Stand und Erfahrungen
bei der Umsetzung von Papier-Bleistift-Tests
auf Internet-basierte Tests**

Erstellt für

Dr. Daniel Jungo

Präsident der Diagnostikkommission des SVB

(Schweizerischer Verband für Berufsberatung)



Zieldefinition

Ziel des vorliegenden Gutachtens ist, eine Übersicht über den aktuellen Stand und Erfahrungen bei der Umsetzung von Papier-Bleistift-Tests auf Internet-basierte Tests zu geben. Dabei soll, ausgehend von einer umfangreichen Aufarbeitung der Literatur, der aktuelle Stand zum Thema reflektiert werden. Standards, die in der Literatur diskutiert werden, werden vorgestellt. Als Beispiel wird eine Umsetzung eines Papier-Bleistift-Tests in eine Online Form diskutiert. Dabei handelt es sich um den Interessentest von Egloff (2005). Am Ende soll eine Checkliste vorgestellt werden, die als praktisches Hilfsmittel bei der Umsetzung genutzt werden kann.

Einleitung

Mit zunehmender Verbreitung und leichterem allgemeiner Verfügbarkeit von Internetzugängen werden auch Anwendungen für die Psychologie, vor allem für die Umsetzung psychologischer Tests, interessant (vgl. Buchanan, 2002; Buchanan & Smith, 1999ab; Drasgow & Chuah, 2006; Garb, 2000; McKenna & Bargh, 2000). Darüber hinaus finden sich in der Literatur Hinweise darauf, dass computerisiert durchgeführte Tests seitens der Testpersonen „ausserordentlich gut akzeptiert werden“ (Kubinger, 2006, S. 120), was für die Anwendung solcher Verfahren spricht. In der Literatur wird aber auch deutlich gemacht, dass bei der Umsetzung psychologischer Tests im Internet die Einhaltung von Qualitätsstandards sichergestellt werden muss (Coyne & Bartram, 2006; Naglieri et al., 2004). Weiter ist anzuführen, dass Computer längst Einzug in den Alltag des psychologischen Diagnostizierens gefunden haben und in manchen Teilgebieten bereits obligat sind; z. B. in der neuropsychologischen Diagnostik, wie Willmes (2003) festhält. Wichtig ist allerdings, bereits an dieser Stelle darauf hinzuweisen, dass sich aus dem Einsatz des Computers in der Psychologischen Diagnostik nicht nur Vorteile, sondern auch eine Reihe von Problemen ergeben, die es zu bedenken gibt (vgl. Kubinger, 1993; Kubinger & Proyer, 2005). Dennoch scheint der Computer aus dem Alltag diagnostischen Arbeitens nicht mehr weg zu denken zu sein und der nächste Schritt, Testungen auch internetbasiert durchzuführen, ist dazu als logische Fortsetzung zu



sehen. In letzter Instanz werden auch Online-Beratungen (z. B. Mallen, Vogel & Rochlen, 2005), Online-Mitarbeiter-Rekrutierung (z. B. Wottawa & Woike, 2002) sowie (therapeutische) Online-Interventionen (z. B. Emmelkamp, 2005; Tate & Zabinski, 2004) diskutiert.

Literaturübersicht zu speziellen Fragen

Äquivalenz von Testformen (Papier-Bleistift und Online)

Lievans und Harris (2003) berichten in ihrer Literaturübersicht, dass Untersuchungen grossteils Hinweise auf die Äquivalenz von Papier-Bleistift und Internet-basierten Testformen liefern. Die Autoren weisen aber auch darauf hin, dass diese Ergebnisse mit Vorsicht zu interpretieren sind, da zum Zeitpunkt der Veröffentlichung der Arbeit, relativ wenig empirische Befunde vorgelegen sind. Bartram und Brown (2004) berichten von keinen Unterschieden in den Testergebnissen für den Vergleich eines ipsativen Interessenfragebogens (Occupational Personality Questionnaire, OPQ 32i, SHL, 1999) in klassischer Papier-Bleistift-Vorgabe mit einer web-basierten Form. Die Testergebnisse differieren bis zu maximal einem Viertel einer Standardabweichung. Die Autoren konnten zeigen, dass es sowohl bezogen auf die Reliabilität als auch auf die Skalen Interkorrelationen zu geringen Unterschieden kommt.

Buchanan (2002) weist darauf hin, dass die Äquivalenz (was die psychometrischen Eigenschaften sowie Gütekriterien betrifft) von Papier-Bleistift-Version und Internet-Version ein und desselben Fragebogens nicht ohne vorhergehende Prüfung angenommen werden kann. Zwar bezieht er sich vor allem auf den klinischen Bereich, dennoch kann diese Aussage auch für andere Anwendungsbereiche generalisiert werden, worauf eine Reihe von Literaturzitaten hinweisen (z. B. Buchanan & Smith, 1999b; Green, Bock, Humphreys, Linn & Reckase, 1984).

Im Bereich von Leistungstests gibt es Hinweise darauf, dass für computerisierte Versionen von Tests die Papier-Bleistift-Tests nicht automatisch übertragbar sind (z. B. Kubinger & Farkas, 1991; Kubinger, Formann & Farkas, 1991; Mead & Drasgow, 1993); wobei es hier in der Literatur auch Gegenbeispiele gibt (z. B. Klinck, 2002;



Preckel & Thiemann, 2003). Bei Mead und Drasgow fanden sich wesentliche Unterschiede vor allem für Tests mit einer stark ausgeprägten Speed-Komponente (tempobetonte Leistungstests). Zu diskutieren ist, ob sich diese Unterschiede durch technische Verbesserungen und anwenderfreundliche Umsetzungen aufgrund neuer technischer Möglichkeiten ausgleichen lassen. Ergebnisse dazu stehen allerdings noch aus. Für Fragebogen deuten die Hinweise in eine entgegengesetzte Richtung. Beispielsweise berichten Austin, Carlbring, Richards und Andersson (2006), Bader, Hofmann und Kubinger (1993), Beckmann, Brähler und Richter (1991), Finger und Ones (1999), Merten und Ruch (1996) sowie Merten und Siebert (1997) von weitgehender Übereinstimmung zwischen computerisierter Fassung und Papier-Bleistift-Fassung von Fragebogen. Ähnliches findet sich auch bei Verfahren, die in der Klinischen Psychologie zum Einsatz kommen (s. Schmitz, Hartkamp, Brinschwitz & Michalek, 1999; Schulenberg & Yutrzenka, 2001). Allerdings finden sich in der Literatur auch Befunde, die darauf hinweisen, dass es Unterschiede zwischen Papier-Bleistift-Tests unter Standardbedingungen den verschiedenen Umgebungen, in denen Personen Internettests bearbeiten, geben kann (z. B. Buchanan & Smith, 1999b). Zur Äquivalenzprüfung gibt Wagner-Menghin (2003) eine Übersicht. Aus der Literatur kann die Schlussfolgerung gezogen werden, dass eine Äquivalenzprüfung unbedingt durchgeführt werden sollte.

Vorteile bezogen auf die Äquivalenz ergeben sich dann, wenn Tests nicht auf Normvergleiche zurückgreifen, sondern kriteriumsorientiert arbeiten oder sich allein auf die Interpretation intraindividuelle Unterschiede beziehen.

Neben der Normierung sind auch andere Gütekriterien zu berücksichtigen. Bezogen auf die Objektivität bleibt festzuhalten, dass sich grundsätzlich keine Änderungen gegenüber der Papier-Bleistift-Form ergeben. Voraussetzung dafür ist allerdings, dass die Testung unter möglichst standardisierten Bedingungen abläuft. Rahmenbedingungen dafür sind in der Instruktion zu beschreiben. Besondere Vorteile ergeben sich hinsichtlich der Auswertungsobjektivität, also der Verrechnungssicherheit. Vorausgesetzt der Auswertalgorithmus wurde korrekt programmiert, passieren keine Fehler bei der Auswertung, wie das bei manuellen



Auswertungen durch verschiedene Faktoren (mangelnde Konzentration, mehrdeutige Markierungen durch die Testperson, abgenutzte Schablonen oder ähnliches) passieren kann (vgl. Allard, Butler, Faust & Shea, 1995; Hebenstreit, 2000). Bezogen auf die Reliabilität und Validität sollten begleitende Studien durchgeführt werden, die zeigen, dass auch bezogen auf diese Gütekriterien Äquivalenz besteht (Butcher, Perry & Hahn, 2004), wobei es in der Literatur Hinweise darauf gibt, dass die Konstruktvalidität mit jener bei herkömmlichen Papier-Bleistift-Testungen vergleichbar ist (Buchanan & Smith, 1999b).

Einstellung

Die Autoren mancher Computertests führen als Vorteile ihrer Tests an, dass diese die Motivation der Testpersonen durch den Einsatz eines modernen, interessanten Mediums steigern. Kubinger (2006) weist darauf hin, dass sich daraus Einflüsse auf das Gütekriterium Fairness ergeben können. Schliesslich ist es denkbar, dass Testpersonen in unterschiedlichem Ausmass auf den Computer reagieren (positive vs. negative Einstellung). Daraus ergibt sich allerdings die Frage, ob die Testpersonen mit der gleichen Einstellung bzw. Motivation an die Testungen bei Computer- und Papier-Bleistift-Versionen ein und desselben Tests herangehen. Wie Buchanan und Smith (1999a) festhalten, können unterschiedliche Umgebungen bei der Beantwortung von Fragebogen zu unterschiedlichen Testergebnissen führen (siehe auch Maurice & Day, 2004; Pasveer & Ellard, 1998; Richman, Kiesler, Weisband & Drasgow, 1999). Daraus ergibt sich die Forderung, dass bei Testungen im Internet Hinweise auf die Rahmenbedingungen gegeben werden müssen, die die Testperson einhalten soll. Das betrifft beispielsweise das ruhige, ungestörte Arbeiten bei der Testbearbeitung oder den Hinweis, dass keine Ablenkungen die Testdurchführung stören sollen (z. B. Natel, Fernseher und Radio abschalten usw.). Letztlich gibt es durch die computerisierte Anwendung auch die Möglichkeit, eine Reihe von Zusatzinformationen, wie etwa die Bearbeitungszeit für Items (entweder allgemein oder in Abhängigkeit, ob das Item positiv oder negativ beantwortet wurde), mit auszuwerten (s. z. B. Furnham, Forde & Cotter, 1998; Wagner-Menghin, 2002).

Antwortformat



Wie Butcher et al. (2004) anführen ist das Auftreten von Antwortverzerrungen im Sinne bestimmter Antwortstile bei Tests, die im Internet durchgeführt werden, noch nicht ausreichend untersucht. Daraus kann die Empfehlung abgeleitet werden, dass eine einfache Kontrolle der Antworten hinsichtlich des Antwortmusters eingesetzt werden sollte. Das mehr oder weniger ausschliessliche Verwenden extremer Antwortmöglichkeiten sowie das ausschliessliche Verwenden einer neutralen Mitte-Angabe kann dadurch leicht entdeckt werden. Eine analoge Vorgehensweise findet sich im Gießen-Test (GT; Beckmann, Brähler & Richter, 1991). Von Vorteil wäre es daher bei der Auswertung bestimmte Antwortmuster (z. B. nur Extremantworten in allen Skalen oder nur „Mitte“-Antworten in allen Skalen bzw. wenn ein bestimmter Cut-Off-Wert überschritten wird) zu berücksichtigen. Dies kann zum Beispiel in Form eines automatisch generierten Texts in der Rückmeldung geschehen, aus dem der Testperson ersichtlich wird, dass sie ein auffälliges Antwortmuster gezeigt hat. Weiter sollte an dieser Stelle dann auch der Hinweis gegeben werden, dass für eine genauere und weitere Ausdifferenzierung der eigenen Interessen ein Gespräch mit einem professionellen (Berufs-)Berater hilfreich ist.

Technische Umsetzung

In Studien konnte gezeigt werden, dass bei der Implementierung von web-basierten Tests verschiedene demographische Variablen relevant sind. Buchanan und Reips (2001) konnten beispielsweise zeigen, dass Testpersonen in Abhängigkeit vom Bildungsgrad Umsetzungen mit oder ohne JavaScripts bevorzugen bzw. PC und Mac-Nutzer sich in bestimmten Persönlichkeitseigenschaften (z. B. Offenheit) voneinander unterscheiden.

Testangst

In der Literatur wird der Einfluss von Ängstlichkeit/Angst bzw. Testängstlichkeit/Testangst auf Leistungen in Testsituationen kontrovers diskutiert (vgl. z. B. Bedell & Marlowe, 1995; Helmke, 1983; Hembree, 1988; Keogh & French, 2001; Pekrun, 1984; Seipp, 1991). Themen sind dabei Einflüsse des Geschlechts - in der Metaanalyse von Hembree (1988) zeigten Frauen höhere Testangst



(Test-Anxiety), die sich allerdings nicht in Leistungsunterschieden auswirkte - sowie von potentiellen Moderatorvariablen, wie zum Beispiel Art des Leistungsmasses, Messzeitpunkt, Art der Testsituation, Art der Testvorgabe oder Einstellungen gegenüber dem Test bzw. dem Diagnostiker (vgl. Bethge, Carlson & Wiedl, 1982; Jäger, 1986; Klinck, 2002). Bei Klinck (2002) zeigte sich, dass das Medium (Computer im Vergleich zu Papier-Bleistift-Tests) zu keinen unterschiedlichen Ausprägungen in der Testängstlichkeit geführt hatte. Hochängstliche erzielten bei Computertests gleich hohe Ängstlichkeitswerte wie bei Papier-Bleistift-Tests und niedrigängstliche in beiden Testformen entsprechend niedrige(re) Werte. Jedenfalls ist die Testangst stets als eine relevante Variable zu berücksichtigen und vor allem bei Leistungstests empfiehlt es sich, vor Beginn der eigentlichen Testung einige Aufwärmaufgaben bzw. so genannte „Eisbrecheritems“ vorzugeben. Als Problem für die Durchführung von Tests im Internet ergibt sich hier, dass es keinen Testleiter gibt, bei dem man Rückfragen stellen kann. Wenngleich dadurch auch sog. „Testleitereffekte“ reduziert werden können, so mag sich das Fehlen eines direkten Ansprechpartners bei manchen Testpersonen negativ auswirken. Demgemäss wäre das Anführen einer Kontaktadresse (z. B. eine Email-Adresse) für Rückfragen zum Test von Vorteil.

Gleichermassen können Vorerfahrungen oder angstinduzierende Vorstellungen über die Institution, an der die Tests durchgeführt werden und die zum Auftreten von Testangst führen können (Sieveking & Chappell, 1970), durch die Vorgabe von Tests im Internet, wo der Test in gewohnter Umgebung durchgeführt werden kann, unter Umständen reduziert werden - allerdings können die Umgebungsbedingungen auch als Störvariable wirksam werden, wenn die Standardbedingungen zur Testdurchführung verletzt werden!

Qualifikation und Aussagegrenzen

Viele psychologische Tests stellen hohe Anforderungen an den Psychologen, bezogen auf die Interpretation der Ergebnisse. Theoretisches Wissen, fundierte psychometrische Kenntnisse und Erfahrung bzw. hohe Vertrautheit im Umgang mit verschiedenen Verfahren sind dabei erforderlich. Bei Internettestungen ist sicher



zu stellen, dass die Testperson durch die Ergebnismrückmeldung nicht überfordert wird und alle Ergebnisse in einer für Laien verständlichen, altersgemässen Sprache transportiert werden. Darüber hinaus sollten bereits in der Ergebnismrückmeldung Informationen darüber gegeben werden, wo der Testnutzer weitere Hinweise zur Interpretation der Ergebnisse bzw. weitere Unterstützung bekommen kann (z. B. Berufsberatungszentren, Informationsstellen).

Die Internettestung kann eine ausführliche professionelle Beratung nicht ersetzen. Komplexe Themen, wie beispielsweise die Berufs-, Laufbahn- oder Karriereberatung erfordern neben hoher Expertise des Beraters auch einen begleitenden psychologisch-diagnostischen Prozess (vgl. Jäger, 2003), der letztlich zur Ableitung eines konkreten Massnahmenvorschlags bzw. einer konkreten Empfehlung führt. Dazu kann die Testung im Internet einen wichtigen Ausgangspunkt geben. Eine Beratung kann sie allerdings nicht ersetzen. Positive Auswirkungen sind dadurch zu erwarten, dass sich die zu beratenden Personen durch den Test im Internet bereits eingehender und strukturierter mit verschiedenen Themen und Fragestellungen beschäftigt haben und so besser vorbereitet in den Beratungsprozess eintreten können und bereits erste Informationen und Hilfestellungen in der Ergebnismrückmeldung erhalten haben.

Gleichzeitig soll an dieser Stelle auch darauf hingewiesen werden, dass darauf zu achten ist, Urheber- und Autorenhinweise klar zu platzieren. Daten, die im Internet frei zugänglich sind, können je nach Aufbau, durch mehr oder weniger geübte Anwender leicht kopiert werden, so dass dem Schutz des geistigen Eigentums der Urheber besondere Aufmerksamkeit geschenkt werden muss.

Weiter soll für den Testanwender auch klar ersichtlich sein, dass das Angebot durch einen seriösen Betreiber bereitgestellt wurde. Das bedeutet, dass an leicht erkennbarer Stelle Logos platziert werden sollten und Hinweise auf den Eigentümer der Seite (z. B. mit Hyperlink zur Homepage) eingefügt werden sollten. Dadurch soll einerseits eine Abgrenzung zu weniger seriösen „Konkurrenten“, die ebenfalls Tests im Internet anbieten, gelingen und andererseits die Schwelle, Kontakt mit der Institution aufzunehmen (z. B. zu



einem individuellen Beratungstermin oder Rückfragen zum Test bzw. den Testergebnissen) möglichst niedrig gehalten werden.

Soziale Erwünschtheit

Grundsätzlich ist festzuhalten, dass Fragebogen, unabhängig davon in welchem Format (Papier-Bleistift oder computergestützt) sie vorgegeben werden, in gleicher Art und Weise in Richtung sozialer Erwünschtheit verfälschbar sind (Fox & Schwartz, 2002). Das bedeutet, dass für Internet-basierte Tests in gleichem Ausmass Kontrollskalen zur Beschreibung von Antwortstilen oder sozial erwünschten Antworten einzusetzen, sind wie bei herkömmlichen Papier-Bleistift-Fragebogen. Allenfalls ist durch eine Anmerkung in der Instruktion im Sinne einer „Ehrlichkeitsinstruktion“ darauf hinzuweisen, dass die Ergebnisse nach Verfälschung, unabhängig davon in welche Richtung, weniger aussagekräftig sind.

Auch die Hoffnung, dass Personen einem „anonymen“ Computer gegenüber ehrlicher antworten würden als einem realen Testleiter (vgl. Hofer, 1985) konnte sich nicht bestätigen lassen (Menghin & Kubinger, 1996). Das bedeutet, dass bei der Auswahl der Items gleiche Massstäbe wie an Items für Papier-Bleistift-Tests gelegt werden müssen bzw., dass bei der Übertragung von Papier-Bleistift-Tests nicht davon ausgegangen werden kann, dass Testpersonen hier offener antworten oder eigene Schwächen verstärkt zugeben würden.

Ergebnisdarstellung und Interpretation

In der Ergebnisdarstellung ist darauf zu achten, dass die Testpersonen ausreichende Informationen zum Verständnis der Testergebnisse erhalten. Dadurch soll einerseits die korrekte Interpretation der Ergebnisse ermöglicht werden und andererseits eine allzu intuitive Interpretation verhindert werden. Bei der Darstellung ist auf Verständlichkeit in der entsprechenden Population (Schüler, Jugendliche, junge Erwachsene, Senioren, ...) zu achten. Werden (psychologische) Fachbegriffe verwendet, dann sind diese kurz und allgemein verständlich zu definieren.

Weiter ist darauf zu achten, ausführliche Erklärungen und gegebenenfalls auch ein Beispiel einzufügen, das die Verständlichkeit und Nachvollziehbarkeit der Ergebnisse erhöht. Zu



verhindern ist, dass sich die Testperson mit ihren Testergebnissen „alleine gelassen“ fühlt bzw. keine Möglichkeit für Rückfragen vorfindet. Daher ist zu empfehlen, auf der Ergebnismeldung Kontaktmöglichkeiten zu professionellen Beratungseinrichtungen (z. B. Berufsberatung) anzuführen. Zu achten ist weiter auf eine strikte Sprachregelung, um bei der Testperson keine falschen Erwartungen hervorzurufen. So sollte zum Beispiel darauf hingewiesen werden, wo die Möglichkeiten und Grenzen eines Verfahrens liegen, das zur selbständigen Bearbeitung im Internet zur Verfügung gestellt wird. Wichtig wäre beispielsweise bei der Bearbeitung eines Interessentests, dass man durch den Test Informationen über die eigene Interessenstruktur sammeln kann, dass man dadurch aber noch nichts über die Eignung hinsichtlich anderer Anforderungspunkte, die nicht mit Interessen in Verbindung stehen (z. B. Leistungsaspekte), sagen kann. Zur Abklärung dieser zusätzlich relevanten Punkte sollte dann wieder der Hinweis auf professionelle Beratungseinrichtungen gegeben werden.

Testarten

In der psychologisch-diagnostischen Praxis werden, je nach Anforderungen und Fragestellung, verschiedene Arten von Tests eingesetzt. Man kann gut zwischen Interessentests, Leistungstests, Persönlichkeitstests, Selbstexplorationsverfahren und Projektiven Verfahren unterscheiden. Dabei sind jeweils Besonderheiten der verschiedenen Testarten zu beachten. Allgemein ist festzuhalten, dass bei Leistungstests, und hier vor allem bei umfangreichen Intelligenztestbatterien, wie etwa dem Intelligenz-Struktur-Test (I-S-T 2000 R, Amthauer, Brocke, Liepman & Beauducel, 2001), die Auswertung und der Umgang mit den Normen komplexe Anforderungen an die Testperson stellen, die ohne professionelle Unterstützung kaum zu bewältigen scheinen. Bei komplexen Leistungsanforderungen wirken sich wohl auch nicht-standardisierte Umgebungsbedingungen gravierender aus, als das bei (Persönlichkeits-) Fragebogen der Fall ist. Abgesehen davon sind viele Aufgaben aus Individualverfahren nicht in gleicher Art und Weise („deckungsgleich“) von Papier-Bleistift- auf Computeranwendungen übertragbar. Zu denken ist hier beispielsweise an das Figurenlegen oder den Mosaik-Test aus dem



Hamburg-Wechsler Intelligenztest für Erwachsene (HAWIE-R, Tewes, 1991).

Interessen- und Persönlichkeitstests sowie Selbstexplorationsverfahren scheinen im Vergleich dazu wesentlich besser umsetzbar. Aufbau und Antwortformat können in der Regel mehr oder weniger deckungsgleich zur Papier-Bleistift-Version übernommen werden. Dennoch ist, wie zuvor dargestellt, auch hier eine Prüfung der Äquivalenz erforderlich. Für Explorix (Jörin, Stoll, Bergmann & Eder, 2003) liegen Erfahrungen mit der Umsetzung eines Selbstexplorationsverfahren im Internet vor. Über die Homepage (<http://www.explorix.ch/berufsberatung.html>) sind auch Beispielerückmeldungen, Hinweise zu theoretischen Grundlagen sowie Hinweise auf Beratungsstellen in Deutschland, Österreich und der Schweiz verfügbar. Für die Umsetzung eines Interessentests im Internet wird auf Seite 11 ein konkretes Beispiel vorgestellt.

Projektive Verfahren scheinen aufgrund ihrer besonderen Charakteristika am wenigsten geeignet, als internet-basierte Tests umgesetzt zu werden. Freie Antworteingabe und Vervollständigen oder Anfertigen von Zeichnungen stellen einerseits hohe technische Anforderungen und brauchen viel Erfahrung in der Interpretation, so dass die Testperson damit nicht alleine gelassen werden sollte. Darüber hinaus wird der Nutzen von Projektiven Verfahren in der Literatur durchwegs kontrovers diskutiert. Verschiedene Autoren, wie etwa Amelang und Schmidt-Atzert (2006), Kubinger (2006) oder Rollett (2003), kritisieren Projektive Verfahren vor allem aufgrund von Mängeln hinsichtlich der Testgütekriterien. Die Autoren sehen sie eher als hypothesengenerierende Verfahren oder als Explorationshilfe, deren Aussagen noch durch psychometrisch gut abgesicherte Verfahren abgestützt werden müssen (s. a. Kubinger, 1997; Punter & Kubinger, 2002).



Psychologisch-diagnostischer Prozess

Ein über das Internet administrierter Test kann keine professionelle Beratung ersetzen. Gemäss Jäger (2003) ist psychologisches Diagnostizieren als ein Prozess zu verstehen, der nach der Klärung der Fragestellung, der Auswahl der diagnostischen Verfahren, der Anwendung und Auswertung der diagnostischen Verfahren zur Interpretation und Gutachtenerstellung führt, die letztlich ein Festsetzen der Intervention mit einschliesst. Daraus wird ersichtlich, dass die Testdurchführung ein einzelner Baustein in einem grösseren Gesamtprozess ist. Dennoch kann die Testperson aus dieser Testung wichtige Informationen ziehen. Neben ausführlichen Interpretationshinweisen sollten sich daher auch Hinweise auf Beratungseinrichtungen mit professioneller Unterstützung in der Auswertung finden. Die Testdurchführung im Internet kann somit den Anstoss geben, eine Beratung aufzusuchen oder eine Entscheidung, die man bereits getroffen hat, durch ein wissenschaftlich fundiertes Verfahren absichern, das „Gesamtpaket“ des psychologisch-diagnostischen Prozesses kann aber letztlich nur durch eine geschulte Person gegeben werden.

Standards

Für den Einsatz von psychologischen Tests wurden zahlreiche Standards entwickelt (z. B. DIN 2002; Fernandez-Ballesteros & Steyer, 2001; Häcker, Leutner & Amelang, 1998; Kersting & Hornke, 2006; Westhoff et al. 2005), die für die Nutzung von Internet-basierten Tests angepasst werden müssen. Nach Fisher und Fried (2003) sind aus ethischen Gesichtspunkten folgende Aspekte zu beachten (übersetzt und geringfügig angepasst durch den Verfasser)

1. Wert ist auf verschlüsselte Datenübertragung, passwortgeschützte Datensicherung, sowie die Verwendung einer Firewall bzw. Sicherheitsvorkehrungen zum Schutz der Daten vor unbefugtem Zugriff zu legen.
2. Sicherheitsmassnahmen müssen mit geschultem Personal (Experten) diskutiert und entwickelt werden, wenn Daten auf einem gewöhnlichen Server gespeichert werden oder an einem Ort



gespeichert werden, der mehreren Personen zugänglich ist (z. B. an einem institutionellen System).

3. Wenn das Internet zur Supervision genutzt wird, dann müssen Mitarbeiter entsprechende Anweisungen gegeben werden, damit die vertraulichen Daten von Klienten/Patienten adäquat geschützt werden.
4. Disketten, CDs und andere (externe) Speichermedien müssen an sicheren Orten abgelegt werden bzw. mit einem Passwort geschützt werden, damit unbefugte Personen keinen Zugriff über den Computer haben.
5. Bei Darstellungen (der Klienten/Patienten) ist die Stimme zu verzerren bzw. das Gesicht unkenntlich zu machen, um die Vertraulichkeit der Daten zu gewährleisten.
6. Daten sollen vernichtet werden, wenn sie nicht länger gebraucht werden. Das gilt, wenn ihre Zerstörung nicht anderen ethischen oder gesetzlichen Vorschriften widerspricht, die sich auf die Aufbewahrung von wissenschaftlichen oder professionellen Daten beziehen.
7. Passwörter sollen nicht gemeinsam (von verschiedenen Mitarbeitern bzw. für verschiedene Datenquellen) genutzt werden und häufig geändert werden.
8. Auch bei der Nutzung von drahtlosen Verbindungen ist zu gewährleisten, dass die Sicherheitsstandards eingehalten werden.
9. Man soll vermeiden, in Emails und instant messages vertrauliche Informationen zu schicken, wenn man keine Verschlüsselungstechniken einsetzt.
10. Sicherheitsmassnahmen zum Schutz transportabler Geräte (wie Laptops, Notebooks, Palms/Palmtops oder Smartphones) sollten, unabhängig davon, ob sie sich im Büro oder in Privaträumen befinden, getroffen werden.
11. Es sollten Bildschirme verwendet werden, die verhindern, dass dritte Personen am Bildschirm mitlesen können bzw. sollten die Bildschirme sichtgeschützt aufgestellt werden.



12. Virusschutz-Software sowie andere Sicherheitsmassnahmen (z. B. Firewalls) für Computer an denen Mitarbeiter arbeiten oder Dateneingaben erfolgen (kurz: überall wo vertrauliche Daten gespeichert werden), sind regelmässig auf dem neuesten Stand zu halten (regelmässige Updates).
13. Werden Computer nicht mehr genutzt oder recycelt, ist darauf zu achten, dass alle Datensätze vorher sorgfältig entfernt werden. Hier ist in den meisten Fällen die Hilfe von Fachpersonen erforderlich, denn auch nach dem Löschen oder Formatieren eines Datenspeichers können (zumindest Spuren von) Daten darauf zurückbleiben und mit spezieller Software wieder lesbar gemacht werden.

Neben diesen technischen Hinweisen werden in Literaturübersichten (z. B. Dillman & Bowker, 2001) auch gestalterische Vorschläge gemacht, die sich auf das Design von Internet-tauglichen Tests beziehen. Die Darstellung der relevantesten Punkte für das vorliegende Gutachten stützt sich auf die Arbeit von Umbach (2004). Die Verwendung von zu vielen Farben soll vermieden werden, um die Lesbarkeit und die Navigierbarkeit nicht zu beeinträchtigen (1¹). Meist wird in der Literatur die Empfehlung gegeben, Befragungen möglichst simpel aufzubauen. Diese Vorgehensweise empfiehlt sich vor allem auch, wenn man Studien mit Populationen durchführen möchte, die möglicherweise über wenig Computererfahrung verfügen (etwa mit Senioren). Die erste Frage sollte ein Item sein, das einfach zu beantworten ist und auf der ersten Bildschirmseite zu sehen ist. Bei dieser Empfehlung lässt sich leicht ein Bezug zu den Ausführungen zu „Eisbrecheritems“ bei Leistungstests machen. Vor allem bei Personen mit hoher Testängstlichkeit bzw. einer reservierten oder distanzierten Einstellung dem Test gegenüber, kann sich eine einfache Einstiegsfrage positiv auswirken. Daraus kann man auch die Forderung ableiten, dass die Bildschirmseite, bei der die Testpersonen willkommen geheissen werden möglichst motivierend gestaltet werden sollte (2). Weiter wird empfohlen, ein konventionelles Format, ähnlich wie bei einem herkömmlichen

¹ Die Hinweise werden nummeriert, um dann später wieder Bezug darauf nehmen zu können.



Fragebogen, einzusetzen (3). Bezogen auf das Antwortformat, das eingesetzt wird, sollte darauf geachtet werden, keinen Antwortmodus zu verwenden, bei dem sich auch bei Papier-Bleistift-Fragebogen gezeigt hat, dass es dabei zu Schwierigkeiten kommen kann (z. B. der Einsatz des freien Antwortformats) (4). Weiter werden bei Umbach Studien zitiert, in denen dazu geraten wird, keine drop-down Antwortfelder (also eine Auswahlmöglichkeit aus einer Liste, die sich beim Anklicken graphisch „herausklappt“ bis eine Antwortmöglichkeit ausgewählt wurde oder der Anwender in eine andere Aktion ausführt) einzusetzen (5). Bezogen auf die Länge ist darauf zu achten, dass die Gesamtdauer/-länge begrenzt wird (6). Dadurch soll verhindert werden, dass Testpersonen Antworten überspringen. Zu Beginn sind klare Instruktionen zu geben, wie man innerhalb des Tests navigieren kann (z. B. „weiter“-Felder, Felder, mit denen man ein - oder mehrere Felder - zurückblättern kann, o. ä.) und wie die Antworten einzugeben sind (7). Darüber hinaus führt Umbach eine Studie an, in der empfohlen wird, auf den Einsatz eines Zählers, der die Zahl der Besucher auf der Seite registriert, zu verzichten (8). Weiter wird vorgeschlagen, die Studie auf verschiedenen Computern (PC, Mac sowie auf Festgeräten und Laptops) und mit verschiedenen Browsern zu testen, um zu verhindern, dass Testpersonen unterschiedliche Bilder auf dem Bildschirm sehen, in Abhängigkeit davon, welche Konfiguration ihr Computer aufweist oder welchen Browser sie zum Einstieg ins Internet nutzen (9). Auch der Einsatz einer Anzeige, über die die Testperson informiert ist, wie weit sie mit der Bearbeitung ist und wie lange es noch dauern wird, bis sie die Testbearbeitung abgeschlossen hat, hat sich in Studien als positiv erwiesen (10). Weiter wird der Hinweis gegeben, dass nicht für jede Frage eine Antwort eingefordert werden sollte, bevor die Testperson mit der Testbearbeitung fortfahren kann (11). Letztlich wird bei Umbach noch die Empfehlung gegeben, lange dauernde, umfangreiche Studien in einzelne Abschnitte zu unterteilen (12).

Für die Durchführung von Befragungen im Internet wurden durch den Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute Standards vorgestellt (2001). Für die Anforderungen der Psychologischen Diagnostik aber wesentlich relevanter, sind die International guidelines on computer-based and Internet delivered



testing der International Test Commission (ITC, 2005 - die Standards sind online verfügbar, die URL wird im Literaturverzeichnis wiedergegeben; s. a. Coyne & Bartram, 2006). Die Passagen, die sich auf die Übertragung von Papier-Bleistift-Tests auf Online-Tests beziehen (Abschnitt 2c in den Richtlinien), werden in der Folge zusammengefasst. In den ITC-Richtlinien heisst es, dass durch den/die Testentwickler klare Belege dafür zu erbringen sind, dass die Internetfassung und die nicht-computerisierte Fassung äquivalent sind (v. a. bezogen auf die Reliabilitäten, dass sie untereinander im - auf Basis der Reliabilitäten - erwarteten Bereich korrelieren, mit anderen Tests und externen Kriterien in vergleichbarem Ausmass korrelieren, und vergleichbare Mittelwerte bzw. Standardabweichungen liefern bzw. entsprechend kalibriert wurden, um vergleichbare Scores zu produzieren). Darüber hinaus ist darauf zu achten, dass die Testperson in der Internet-Version die gleiche „Kontrolle“ über den Test ausüben kann als in der Papier-Bleistift-Version (also z. B. Items überspringen oder zurück „blättern“). Ebenso sollen die Items auf gleiche Art und Weise präsentiert werden wie in der Ursprungsversion. Weiter gilt, dass das Antwortformat gleich sein sollte.

Für alle Studien, die durchgeführt werden (z. B. zur Bestimmung der Äquivalenz oder für Normierungszwecke) gilt, dass für die Internet-Version die selben Bedingungen simuliert werden sollten, wie sie dann beim tatsächlichen Einsatz für die Zielpopulation gelten (also zum Beispiel unbeaufsichtigtes Bearbeiten des Fragebogens, nicht standardisierte Testbedingungen). Auf Seiten der Durchführungsbedingungen ist darauf zu achten, dass Hard- und Software-Eigenschaften auf Teilnehmerinnen mit Behinderungen oder speziellen Bedürfnissen abgestimmt sind. Die Standards für andere Durchführungsbedingungen (s. die oben angeführten) sind zu beachten. Modifikationen an den Tests sollen sich speziell auf die Bedürfnisse der Testnutzer beziehen und sich innerhalb eines akzeptablen Rahmens bewegen, so dass die Validität der Ergebnisse nicht beeinflusst wird.

Darüber hinaus wird in den ITC-Standards darauf hingewiesen, dass für Testpersonen mit Behinderungen auch andere Erhebungsprozeduren in Betracht gezogen werden sollten. Das



bedeutet, dass Situationen und Fälle denkbar sind, wo für bestimmte Personengruppen nach wie vor eine Papier-Bleistift-Vorgabe des entsprechenden Tests empfohlen werden sollte.

Beispiel für die Umsetzung eines Papier-Bleistift-Tests in eine Internet-taugliche Fassung

Im Internet gibt es eine Vielzahl von unterschiedlich fundierten Tests und Fragebogen. Ein Beispiel eines wissenschaftlich gut abgesicherten Verfahrens soll den Abschluss dieser Stellungnahmen bilden, um zuvor gegebene Ausführungen zu veranschaulichen. Der Interessenfragebogen von Egloff (2005) ist über <http://www.feelok.ch> („Tests“ → „Zeigt die Tests“ → „Berufseignungstest“) im Internet verfügbar.

Einige Anmerkungen. Zunächst fällt auf, dass der Interessentest als Berufseignungstest angeboten wird. Das kann zu Missverständnissen und falschen Erwartungen führen, da unter der Eignung auch Leistungsaspekte zu verstehen sind. Hier wird aber die Interessenstruktur der Testperson erfasst. Aussagen über die Leistungsfähigkeit der Testperson sind allerdings nach Abschluss des Tests nicht möglich. Darüber hinaus wird empfohlen, eher von Interessenfragebogen als von Interessentest zu sprechen, da der Begriff „Test“ bei den Teilnehmern vielleicht die Vermutung wecken könnte, dass es sich tatsächlich um einen Test handelt, also etwas „geprüft“ wird und es richtige und falsche Antworten geben kann. Diese Argumentation folgt Kubinger (2006) und wird dort ausführlicher vorgestellt.

Die Beschreibung der Ziele (Möglichkeiten und Grenzen) des Fragebogens wird auch in der Zieldarstellung im Internet dargestellt: „Er erleichtert es dir, deine Interessen selber zu beurteilen und dein eigenes Interessenbild darzustellen. Du kannst die Stärke deiner Interessen mit andern Jugendlichen vergleichen. Du kannst aufgrund deiner Interessen Verbindungen mit Berufsfeldern herstellen.“ Dieser einfühlsam und gut verständlich geschriebene



Text ist als besonders positiv hervorzuheben. Die Teilnehmer erhalten so eine genaue Vorstellung von dem Fragebogen und werden gut in die Testung eingeführt.

Im Bereich der optischen Gestaltung finden sich gewiss Verbesserungsmöglichkeiten, die zum Beispiel die Präsentation der Antwortmöglichkeiten oder den optischen Aufbau der Seite betreffen.

Ein wesentlicher Unterschied zur Papier-Bleistift-Version scheint zu sein, dass es hier nicht möglich ist, keine Antwort einzugeben. Technisch wäre es kein Problem, auch fehlende Antworten zu verrechnen. Es bleibt die Frage, ob durch den Zwang, bei jeder Frage eine Antwort eingeben zu müssen, sog. „Reaktanzphänomene“ (vgl. Karner, 2002) auftreten, die sich letztlich auf die Angaben der Testperson im Sinne von Antwortverzerrungen auswirken können. Darüber hinaus ist anzumerken, dass kein „Zurückblättern“ gestattet ist, was wiederum einen Unterschied zur Papier-Bleistift-Form darstellt. Technisch wäre auch das Zurückblättern umsetzbar. Somit wurden zwei Unterschiede zur Papier-Bleistift-Version in der Umsetzung identifiziert. Welche und ob diese Änderungen Auswirkungen haben, muss in einer Äquivalenzstudie geprüft werden (vgl. Boejti & Sommer, in Vorb.). Argumente für und gegen diese Veränderungen in der Umsetzung, sollten in der angesprochenen Arbeit ebenfalls diskutiert werden.

Nicht umgesetzt wurde ein Algorithmus, der Antwortmuster prüft. Gibt man bei jeder Frage „nein“ ein (bei jeder Frage in jedem Interessengebiet), dann gibt es keine Warnung oder Meldung, dass man versuchen soll, differenzierter Auskunft zu geben, da die Ergebnisse so nur eingeschränkt interpretierbar sind. Technisch gesehen ist (zumindest) die Auswertung von Extremantworten (bzw. eines entsprechenden Cut-Off-Werts) wenig problematisch. In der Rückmeldung an die Testperson mit auffälligem Antwortmuster kann auch der Hinweis eingefügt werden, dass die Rücksprache mit einem professionellen Berufsberater hilfreich ist, um eigene Interessen sowie Stärken und Schwächen besser kennen zu lernen.

Die Ergebnisdarstellung ist sehr ausführlich gestaltet und bietet fundierte Informationen. Der Hinweis auf professionelle Beratungsstellen fehlt allerdings. Auch eine Checkliste für weitere



Schritte (z. B. wie gehe ich vor bei der Suche nach einer Lehrstelle in dem Gebiet, das mich interessiert) wäre für die Testpersonen gewiss hilfreich, wobei das auch durch den Hinweis auf entsprechende Beratungsstellen erfolgen kann.

Ein Hinweis betrifft noch die Beurteilung der einzelnen Skalenausprägungen. Hier bedarf es eigentlich der Berechnung sog. kritischer Differenzen. Natürlich ist dem Laien weder die Berechnung zumutbar noch davon auszugehen, dass er/sie versteht, warum das erforderlich ist. Technisch wäre es allerdings leicht möglich, eine Rangliste, mit den am höchsten ausgeprägten Skalen zu präsentieren, aus der für die Testperson ersichtlich ist, welche Skalen am höchsten, welche weniger hoch und welche gleich ausgeprägt sind.

Abschliessende Bemerkung. Wie gezeigt wurde, findet sich hier eine gute Umsetzung eines Papier-Bleistift-Tests in eine Internet-taugliche Fassung. Einige Punkte, die verbesserungswürdig erscheinen, wurden angeführt und einige Anregungen für Verbesserungen gegeben. Vorausgesetzt, die Studie von Boejti und Sommer (in Vorb.) zeigt, dass die beiden Testformen äquivalent sind, kann die Umsetzung hier als Beispiel für zukünftige Vorgehensweisen gesehen werden.



Checkliste

Wie zu Beginn erwähnt, sollte am Ende dieses Gutachtens eine Checkliste gegeben werden, anhand derer geprüft werden kann, welche Schritte bei der Umsetzung eines Papier-Bleistift-Tests in einen Internet-tauglichen Fragebogen hilfreich sind und was dabei beachtet werden sollte. Die Checkliste gliedert sich in die Teilbereiche Technik, Umsetzung und Ergebnisdarstellung.

I) Technik

- Aus technischer Sicht sind die bei Fisher und Fried (2003) genannten 13 Punkte von Bedeutung. Weiter ist der bei Umbach (2004) unter Punkt 9 genannte Aspekt zu beachten: Vor der Freigabe der Testung sollte ein umfangreiches Testprogramm durchgeführt werden, um sicher zu stellen, dass Anwender, die verschiedene Computer, Browser oder Konfigurationen verwenden, alle dieselben Testbedingungen vorfinden.
- Falls es Ausschlusskriterien für Nutzer aus technischer Sicht (wenn z. B. ein bestimmter Browser verwendet werden muss oder ähnliches) gibt, dann sind diese rechtzeitig bekannt zu geben.
- Weiter ist die Umsetzung so zu gestalten, dass keine Anforderungen gestellt werden, die handelsübliche Computer nicht erfüllen können; z. B. die Ladezeiten oder die graphische Aufbereitung betreffend. Von zentraler Bedeutung ist in diesem Zusammenhang auch der Hinweis darauf, dass bei der Testentwicklung bedacht werden muss, dass Nutzer der Web-Seite an Computern mit unterschiedlichen Verbindungsgeschwindigkeiten (z. B. Modem, Breitband, usw.) arbeiten. Deswegen ist zu fordern, dass die Web-Seite so zu gestalten ist, dass sie auch mit einer langsamen Internet-Verbindung noch gut zu bearbeiten ist (z. B. was die Ladezeiten für den Seitenaufbau betrifft).



II) Umsetzung

- In Übereinstimmung mit den ITC-Guidelines (2005; Coyne & Bartram, 2006) ist zunächst eine Studie (bzw. falls erforderlich: Studien) zur Äquivalenz der Testformen durchzuführen. Liegt in dieser Studie (den Studien) der Nachweis der Übereinstimmung der Testformen (gemäss den in den Richtlinien angeführten Punkten, s. o.), kann der Test zur standardmässigen Nutzung (unter denselben Bedingungen, wie die Papier-Bleistift-Form) empfohlen werden.
- Gemäss ITC-Richtlinien ist weiter darauf zu achten, dass die Testperson in der Internet-Version die gleiche Kontrolle über den Test ausüben kann wie in der Papier-Bleistift-Version („z. B. Zurückblättern“, Auslassen von Fragen) und, dass Übereinstimmung hinsichtlich der Vorgabebedingungen und des Antwortformats bestehen soll.
- Aus den bei Umbach (2004) angeführten Punkten sind 1-8 sowie 10-12 für die Umsetzung relevant.
- In der Instruktion sollten Hinweise auf die Umgebungsbedingungen gegeben werden, unter denen die Testung durchgeführt werden sollte (z. B. möglichst wenig Ablenkung durch andere Medien - TV, Radio o. ä. - oder Natel ausschalten). Weiter sollte in der Instruktion auch das Ziel des Tests genau beschrieben und ein Ausblick auf die Ergebnisse gegeben werden.
- Autoren- und Urheberrechte sowie Kontaktinformationen zur Organisation, die den Test ins Internet gestellt hat, sollten klar sichtbar angebracht werden.
- Der Einsatz von Verrechnungsvorschriften oder eigenen Skalen zur Kontrolle von bestimmten Antwortmustern, -stilen oder sozial erwünschten Antworten sollte bei der Erstellung des Tests diskutiert werden.

III) Ergebnisdarstellung



- Auf eine übersichtliche, klare und (für Laien und die entsprechende Zielpopulation) leicht verständliche Sprache ist zu achten. Informationen über den theoretischen Hintergrund des Verfahrens sollten gegeben werden. Werden Fachbegriffe verwendet, so sind diese zu erklären. Wird mit Normen gearbeitet, so ist die Bedeutung des Normvergleichs genau zu erklären.
- Eine graphische Ergebnisdarstellung erhöht die Verständlichkeit der Ergebnisse.
- Wenn es inhaltlich Sinn macht und wichtige Zusatzinformation für die Testperson bringt, dann ist die Berechnung von kritischen Differenzen für Skalenunterschiede eine hilfreiche Ergänzung bei der Gestaltung der Auswertung.
- Hinweise auf professionelle Beratungsstellen sollten leicht verfügbar und an gut erkennbarer Stelle platziert werden.



Literatur

- Allard, G., Butler, J., Faust, D. & Shea, M. T. (1995). Errors in hand scoring objective personality tests: The case of the Personality Diagnostic Questionnaire-Revised (PDQ-R). *Professional Psychology: Research and Practice*, 26, 304-308.
- Amelang, M. & Schmidt-Atzert, L. (2006). *Psychologische Diagnostik und Intervention* (4. Aufl.). Heidelberg: Springer.
- Anthauer, R., Brocke, B., Liepman, D. & Beauducel, A. (2001). *Intelligenz-Struktur-Test 2000 R (I-S-T 2000 R)*. Göttingen: Hogrefe.
- Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute (2001, Mai). *Standards zur Qualitätssicherung für Online-Befragungen*. Online verfügbar (07.07.2006): http://www.adm-ev.de/pdf/Onlinestandards_D.PDF.
- Austin, D. W., Carlbring, P., Richards, J. C. & Andersson, G. (2006). Internet administration of three commonly used questionnaires in panic research: Equivalence to paper administration in Australian and Swedish samples of people with panic disorder. *International Journal of Testing*, 6, 25-39.
- Bader, P., Hofmann, K. & Kubinger, K. D. (1993). Zur Brauchbarkeit der Normen von Papier-Bleistift-Tests für die Computer-Vorgabe: Ein Experiment am Beispiel des Gießen-Tests. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 14, 129-135.
- Bartram, D. & Brown, A. (2004). Online testing: Mode of administration and the stability of OPQ 32i scores. *International Journal of Selection and Assessment*, 12, 278-284.
- Beckmann, D., Brähler, E. & Richter, H. E. (1991). *Der Gießen-Test (GT)* (4. Aufl.). Göttingen: Hogrefe.
- Bedell, J. R. & Marlowe, H. A. (1995). An evaluation of test anxiety scales: Convergent, divergent, and predictive validity. In C. D. Spielberger & P. R. Vagg (Eds.), *Test anxiety - theory, assessment, and treatment* (pp. 35-45). Washington, DC: Taylor & Francis.



- Bethge, H.-J., Carlson, J. S. & Wiedl, K. H. (1982). The effects of dynamic assessment procedures on Raven Matrices performance, visual search behavior, test anxiety and test orientation. *Intelligence, 6*, 89-97.
- Boejti, B. & Sommer, D. (in Vorbereitung). *Äquivalenzstudie zum Interessenfragebogen von Egloff (Arbeitstitel)*. Masterarbeit NABB-8.
- Buchanan, T. (2002). Online assessment: Desirable or dangerous? *Professional Psychology-Research and Practice, 33*, 148-154.
- Buchanan, T. & Reips, U.-D. (2001). Platform-dependent biases in online research: Do Mac users really think different? In K. J. Jonas, P. Breuer, B. Schauenburg & M. Boos (Eds.), *Perspectives on Internet research: Concepts and methods*. Retrieved July 4, 2006, from <http://server3.uni-psych.gwdg.de/gor/contrib/buchanan-tom>.
- Buchanan, T. & Smith, J. L. (1999a). Research on the Internet: Validation of a World Wide Web mediated personality scale. *Behavior Research Methods, Instruments, & Computers, 31*, 565-571.
- Buchanan, T. & Smith, J. L. (1999b). Using the Internet for psychological research: Personality testing on the World Wide Web. *British Journal of Psychology, 90*, 125-144.
- Butcher, J. N., Perry, J. & Hahn, J. (2004). Computers in clinical assessment: Historical developments, present status, and future challenges. *Journal of Clinical Psychology, 60*, 331-345.
- Coyne, I. & Bartram, D. (2006). Introduction to the special issue on the ITC guidelines on computer-based and Internet-delivered testing [Special Issue]. *International Journal of Testing, 6*(2).
- Dillman, D. A. & Bowker, D. (2001). The web questionnaire challenge to survey methodologists. In U.-D. Reips & M. Bosnjak (Eds.), *Dimensions of Internet science*. Lengerich: Pabst.
- DIN (2002). DIN 33430: *Anforderungen an Verfahren und deren Einsatz bei berufsbezogenen Eignungsbeurteilungen*. Berlin: Beuth.



- Drasgow, F. & Chuah, S.C. (2006). Computer-based testing. In M. Eid & E. Diener (Eds.), *Handbook of multimethod measurement in psychology* (pp. 87-100). Washington, DC: American Psychological Association.
- Egloff, E. (2005). *Berufswahltagbuch*. Buchs: Lehrmittelverlag des Kantons Aargau.
- Emmelkamp, P. M. G. (2005). Technical innovations in clinical assessment and psychotherapy. *Psychotherapy and Psychosomatics*, 74, 336-343.
- Fernandez-Ballesteros, R. & Steyer, R. (2001). Special Section on psychological assessment standards and guidelines - Introduction [Special Section]. *European Journal of Psychological Assessment*, 17(3).
- Finger, M. S. & Ones, D. S. (1999). Psychometric equivalence of the computer and booklet forms of the MMPI: A meta-analysis. *Psychological Assessment*, 11, 58-66.
- Fisher, C. B. & Fried, A. L. (2003). Internet-mediated psychological services and the American Psychological Association Ethics Code. *Psychotherapy: Theory, Research, Practice, Training*, 40, 103-111.
- Fox, S. & Schwartz, D. (2002). Social desirability and controllability in computerized and paper-and-pencil personality questionnaires. *Computers in Human Behavior*, 18, 389-410.
- Furnham, A., Forde, L. & Cotter, T. (1998). Personality scores and test taking style. *Personality and Individual Differences*, 24, 19-23.
- Garb, H. N. (2000). Computers will become increasingly important for psychological assessment: Not that there's anything wrong with that! *Psychological Assessment*, 12, 31-39.
- Green, B. F., Bock, R. D., Humphreys, L. G., Linn, R. L. & Reckase, M. D. (1984). Technical guidelines for assessing computerized adaptive tests. *Journal of Educational Measurement*, 21, 347-360.
- Häcker, H., Leutner, D. & Amelang, M. (1998). Standards für pädagogisches und psychologisches Testen. *Supplementum 1/1998*



der „*Diagnostica*“ und der „*Zeitschrift für Differentielle und Diagnostische Psychologie*“.

- Hebenstreit, G. K. (2000). *Die Fehleranfälligkeit der Auswertung beim Adaptiven Intelligenz Diagnostikum (Kubinger & Wurst, 1991)*. Unveröffentlichte Diplomarbeit: Universität Wien.
- Helmke, A. (1983). Prüfungsangst. *Psychologische Rundschau*, 34, 7-47.
- Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. *Review of Educational Research*, 58, 47-77.
- Hofer, P. J. (1985). Developing standards for computerized psychological testing. *Computers in Human Behavior*, 1, 301-315.
- International Test Commission (2005). *International Guidelines on computer-based and Internet delivered testing*. ITC (available online: <http://www.intestcom.org/guidelines>).
- Jäger, R. S. (1986). Measuring examiner and examinee reactions to each other and to the psychodiagnostic situation. In B. Nevo & R. S. Jäger (Eds.), *Psychological testing: The examinee perspective* (pp. 129-149). Göttingen: Hogrefe.
- Jäger, R. S. (2003). Prozess, diagnostischer. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe Psychologischer Diagnostik* (S. 348-354). Weinheim: Beltz/PVU.
- Jörin, S., Stoll, F., Bergmann, C. & Eder, F. (2003). *Explorix - Das Werkzeug zur Berufswahl und Laufbahnplanung*. Bern: Hans Huber.
- Karner, T. (2002). The volunteer effect of answering personality questionnaires. *Psychologische Beiträge*, 44, 42-49.
- Keogh, E. & French, C. C. (2001). Test anxiety, evaluative stress, and susceptibility to distraction from threat. *European Journal of Personality*, 15, 123-141.
- Kersting, M. & Hornke, L. F. (2006). Improving the quality of proficiency assessment: The German standardization approach. *Psychology Science*, 48, 85-98.
- Klinck, D. (2002). *Computergestützte Diagnostik*. Göttingen: Hogrefe.



- Kubinger, K. D. (1993). Testtheoretische Probleme der Computerdiagnostik. *Zeitschrift für Arbeits- und Organisationspsychologie*, 37, 130-137.
- Kubinger, K. D. (1997). Testrezensionen: 25 einschlägige Verfahren [Sonderheft]. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 18(1-2).
- Kubinger, K. D. (2006). *Psychologische Diagnostik - Theorie und Praxis psychologischen Diagnostizierens*. Göttingen: Hogrefe.
- Kubinger, K. D. & Farkas, M. G. (1991). Die Brauchbarkeit der Normen von Papier-Bleistift-Tests für die Computer-Vorgabe: Ein Experiment am Beispiel der SPM von Raven als kritischer Beitrag. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 12, 257-266.
- Kubinger, K. D., Formann, A. K., & Farkas, M. G. (1991). Psychometric shortcomings of Raven's Standard Progressive Matrices (SPM) in particular for computerized testing. *European Review of Applied Psychology*, 41, 295-300.
- Kubinger, K. D. & Proyer, R. (2005). Basisinformationen zum Einsatz des Computers in der berufsbezogenen Eignungsbeurteilung. In K. Westhoff, L. J. Hellfritsch, L. F. Hornke, K. D. Kubinger, F. Lang, H. Moosbrugger, A. Püschel & G. Reimann (Hrsg.), *Grundwissen für die berufsbezogene Eignungsbeurteilung nach DIN 33430* (S. 166-172) (2. Aufl.). Lengerich: Pabst.
- Lievans, F. & Harris, M. M. (2003). Research on Internet recruitment and testing: Current status and future directions. In I. Robertson & C. Cooper (Eds.), *The international review of industrial and organizational psychology* (pp. 131-165). Chichester, England: Wiley.
- Mallen, M. J., Vogel, D. L. & Rochlen, A. B. (2005). The practical aspects of Online counseling: Ethics, training, technology, and competency. *Counseling Psychologist*, 33, 776-818.
- Maurice, S. A. & Day, R. L. (2004). Online testing technology: Important lessons learned. *International Journal of Engineering Education*, 20, 776-818.



- McKenna, K. Y. A. & Bargh, J. A. (2000). Plan 9 from cyberspace: The implications of the Internet for personality and social psychology. *Personality and Social Psychology Review*, 4, 57-75.
- Mead, A. & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin*, 114, 449-458.
- Menghin, S. & Kubinger, K. D. (1996). Zur Legende: "Testpersonen beantworten dem Computer persönliche und intime Fragen offener als einem Testleiter" - Ergebnisse eines Experiments. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 17, 163-169.
- Merten, T. & Ruch, W. (1996). A comparison of computerized and conventional administration of the German versions of the Eysenck Personality Questionnaire and the Carroll Rating Scale for depression. *Personality and Individual Differences*, 20, 281-291.
- Merten, T. & Siebert, K. (1997). A comparison of computerized and conventional administration of the EPQ-R and the CRS: further data on the Merten and Ruch (1996) study. *Personality and Individual Differences*, 22, 283-286.
- Naglieri, J. A., Drasgow, F., Schmitt, M., Handler, L., Prifitera, A., Margolis, A., et al. (2004). Psychological testing on the Internet: New problems, old issues. *American Psychologist*, 59, 150-162.
- Pasveer, K. A. & Ellard, J. H. (1998). The making of a personality inventory: Help from the WWW. *Behavior Research Methods, Instruments, & Computers*, 30, 309-313.
- Pekrun, R. (1984). An expectancy-model of anxiety. In H. M. van der Ploeg, R. Schwarzer & C. D. Spielberger (Eds.), *Advances in test anxiety research* (pp. 52-73). Hillsdale, NJ: Erlbaum.
- Preckel, F. & Thiemann, H. (2003). Online- versus paper-pencil-version of a high potential intelligence test. *Swiss Journal of Psychology*, 62, 131-138.



- Punter, J. F. & Kubinger, K. D. (2002). Was ist aus der Kritik der "Testrezensionen: 25 einschlägige Verfahren" (Zeitschrift für Differentielle und Diagnostische Psychologie, 18, Heft 1-2) geworden? *Psychologie in Österreich, 22*, 24-33.
- Richman, W. L., Kiesler, S., Weisband, S. & Drasgow, F. (1999). A meta-analytic study of social desirability distortion in computer administered questionnaires, traditional questionnaires, and interviews. *Journal of Applied Psychology, 84*, 754-775.
- Rollett, B. (2003). Projektive Verfahren. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 340-348). Weinheim: Beltz/PVU.
- Schmitz, N., Hartkamp, N., Brinschwitz, C. & Michalek, S. (1999). Computerized administration of the Symptom Checklist (SCL-90-R) and the Inventory of Interpersonal Problems (IIP-C) in psychosomatic outpatients. *Psychiatry Research, 87*, 217-221.
- Schulenberg, S. E. & Yutrzenka, B. A. (2001). Equivalence of computerized and conventional versions of the Beck Depression Inventory-II (BDI-II). *Current Psychology, 20*, 216-230.
- Seipp, B. (1991). Anxiety and academic performance: A meta-analysis of findings. *Anxiety Research, 4*, 27-41.
- Sieveking, N. A. & Chappell, J. E. (1970). Reactions to the names "counseling center" and "psychological center". *Journal of Consulting and Clinical Psychology, 34*, 124-127.
- SHL. (1999). *OPQ32 manual and user's guide*. Thames Ditton, UK: SHL group plc.
- Tate, D. F. & Zabinski, M. F. (2004). Computer and Internet applications for psychological treatment: Update for clinicians. *Journal of Clinical Psychology, 60*, 209-220.
- Tewes, U. (1991). *Hamburg-Wechsler Intelligenztest für Erwachsene-Revision (HAWIE-R)*. Bern: Hans Huber.
- Umbach, P. D. (2004). Web surveys: Best practices. *New Directions for Institutional Research, 121*, 23-38.



- Wagner-Menghin, M. M. (2002). Towards the identification of non-scalable personality questionnaire respondents: Taking response time into account. *Psychologische Beiträge, 44*, 62-77.
- Wagner-Menghin, M. M. (2003). Äquivalenzprüfung. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 32-34). Weinheim: Beltz/PVU.
- Westhoff, K., Hellfritsch, L. J., Hornke, L. F., Kubinger, K. D., Lang, F., Moosbrugger, H., Püschel, A., & Reimann, G. (Hrsg.) (2005). *Grundwissen für die berufsbezogene Eignungsbeurteilung nach DIN 33430* (2. Aufl.). Lengerich: Pabst.
- Willmes, K. (2003). Neuropsychologische Diagnostik. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 287-297). Weinheim: Beltz/PVU.
- Wottawa, H. & Woike, J. H. (2002). Internet-Recruiting und Assessment. Eine Chance, die die Wirtschaftspsychologen nutzen sollten! *Wirtschaftspsychologie, 9*, 33-38.