

Internetbasierte psychologische Diagnostik

Timo Gnambs*, Bernad Batinic* & Guido Hertel**

Johannes Kepler Universität Linz* & Westfälische Wilhelms-Universität Münster**

1 EINLEITUNG	2
2 INTERNETBASIERTE ERHEBUNGSTECHNIKEN	6
2.1 INTERNETBASIERTE TESTS UND FRAGEBÖGEN	7
2.2 INTERNETBASIERTE INTERVIEWS UND GRUPPENDISKUSSIONEN	8
2.3 INTERNETBASIERTE BEOBACHTUNGEN	10
3 UMSETZUNG DIAGNOSTISCHER AUFGABEN MITHILFE NEUER MEDIEN	13
3.1 EXPLORATION UND ANAMNESE	13
3.2 PERSÖNLICHKEITSDIAGNOSTIK	15
3.3 INTELLIGENZ- UND LEISTUNGSDIAGNOSTIK	16
3.4 KLINISCH-PSYCHOLOGISCHE DIAGNOSTIK	18
3.5 BERUFLICHE EIGNUNGSDIAGNOSTIK	20
4 GÜTEKRITERIEN INTERNETBASIERTER METHODEN	23
4.1 OBJEKTIVITÄT	23
4.2 RELIABILITÄT	24
4.3 VALIDITÄT	29
4.4 NORMIERUNG	32
4.5 FAIRNESS	34
4.6 VERFÄLSCHBARKEIT	35
4.7 AKZEPTANZ	41
5 RESÜMEE UND AUSBLICK	44
6 LITERATUR	47

1 Einleitung

Die rasche Entwicklung neuer Kommunikationstechnologien in den letzten beiden Jahrzehnten, insbesondere des Internets, hat die Möglichkeiten der psychologischen Diagnostik deutlich erweitert. Während die Einführung computergestützter Diagnoseverfahren anfangs vor allem davon geprägt war, Zeit und Geld zu sparen, entfalteten sie zunehmend ihr eigentliches Innovationspotenzial mit der Entwicklung eigenständiger Testverfahren. Im Gegensatz zur einfachen Übertragung von Papier-Bleistift-Tests auf den Computer werden dabei spezifische Möglichkeiten neuer Technologien genutzt um auch komplexe und bislang nur schwer zu erfassende Merkmale angemessen zu operationalisieren. Beispiele hierfür sind reaktionszeitsensitive Verfahren (z. B. Impliziter Assoziationstest von Greenwald, McGhee und Schwartz, 1998), objektive Persönlichkeitstests (z. B. Objektiver Leistungsmotivationstest von Schmidt-Atzert, 2004) oder adaptive Verfahren (z. B. Lexikon-Wissen Test von Wagner-Menghin, 2004), welche ohne die Rechenleistung eines Computers nicht oder nur schwer zu realisieren wären. Angesichts der zunehmenden Verfügbarkeit von leistungsfähigen Computern sowie der Internetanbindung breiter Bevölkerungsschichten ist computergestützte Diagnostik mittlerweile vielfältig einsetzbar. Eine Repräsentativerhebung (Eimeren & Fries, 2010) weist für Deutschland für das Jahr 2010 bereits über zwei Drittel der Erwachsenen ab 14 Jahren als regelmäßige Internetnutzer aus. Die Erhebung von Informationen über das Internet (Online Forschung) hat im deutschsprachigen Raum eine mittlerweile über 15jährige Geschichte (Welker, 2007). Insbesondere in der psychologischen Forschung sind internetbasierte Erhebungen mittlerweile fest etabliert. Angesichts dieser Entwicklungen sollen in diesem Kapitel die Möglichkeiten und Grenzen internetbasierter Verfahren für die psychologische Diagnostik dargestellt und diskutiert werden.

Psychologische Diagnostik umfasst die systematische Erfassung und Aufbereitung von Informationen über psychologisch relevante Merkmale einer Person mit dem Ziel, Entscheidungen und Handlungsempfehlungen zu begründen, zu kontrollieren und zu optimieren (Jäger & Petermann, 1999). Die Arbeit als Diagnostiker beschränkt sich hierbei nicht nur auf die Auswahl und Applikation geeigneter Testverfahren, sondern ist als zusammenhängender Prozess aufzufassen, der neben der Konkretisierung der Fragestellung und der Auswahl sowie Anwendung diagnostischer Verfahren vor allem auch die Ableitung konkreter Interventionsvorschläge umfasst (Kubinger, 2009). Das Internet kann nun prinzipiell in allen Phasen des diagnostischen Prozess genutzt werden. In der Praxis stellt

jedoch bislang die systematische Erfassung psychologischer Phänomene mit Fragebögen, Leistungstests, Simulationen oder projektiven Verfahren das vornehmliche Anwendungsfeld internetbasierter Diagnostik dar.

Bislang gibt es keine einheitliche Definition von „internetbasierter psychologischer Diagnostik“. Manche Autoren betonen ausschließlich den technologischen Aspekt und subsumieren alle computerbasierten Verfahren darunter, bei denen Testdarbietung und Datenspeicherung über das Internet erfolgt. Dazu zählen dann auch computerbasierte Testungen in Testlaboren, bei denen die Datenspeicherung nicht lokal sondern auf einem entfernten (Internet)server erfolgt (z. B. Meade, Michels & Lautenschlager, 2007). Andere Autoren beziehen sich lediglich auf unkontrollierte Computertestungen, bei denen die Probanden die Internetadresse für einen Online-Test zur Verfügung gestellt bekommen und diesen nach eigenem Ermessen bearbeiten können (z. B. Risko, Quilty & Oakman, 2006). Zusätzlich wird dies noch mit Fragen der Stichprobenqualität vermengt, da internetbasierte Tests häufig einfach online gestellt werden und Teilnehmer ohne explizite Stichprobenziehung gewonnen werden. In diesem Sinne wird internetbasierte Diagnostik implizit manchmal mit *unkontrollierter Computerdiagnostik an anfallenden Stichproben* gleich gesetzt. Dies stellt jedoch lediglich ein Spezialfall in einer ganzen Bandbreite internetbasierter, diagnostischer Zugänge dar. Deshalb müssen zur Beschreibung internetgestützter, psychologischer Diagnostik vier Aspekte unterschieden werden:

- 1) *Medium*: Mithilfe welcher (technologischer) Hilfsmittel erfolgt die Datenerhebung: Persönliche Gespräche, Papier-Bleistift Verfahren, Telefon-, Computer-, PDA-unterstützt etc.?
- 2) *Modus*: Inwiefern erfolgt eine Kontrolle der Durchführungsbedingungen der Datenerhebung (unkontrolliert oder überwacht)?
- 3) *Lokalität*: Wo befindet sich das Testmaterial und wo erfolgt die Datenspeicherung: Direkt vor Ort, z. B. lokal am eigenen Computer, oder räumlich entfernt, z. B. auf einem Internetserver?
- 4) *Selektion*: Wie werden Teilnehmer gewonnen: Erfolgt eine explizite Stichprobenziehung aus einer vorab definierten Population?

All jene Verfahren, die elektronische Informationstechnologien im diagnostischen Prozess nutzen, werden in Abgrenzung zur traditionellen Diagnostik unter dem Begriff der technologiebasierten Diagnostik subsumiert (Jurecka & Hartig, 2007). Hierzu gehören sowohl herkömmliche Computertests, die auf speziell ausgestatteten Testcomputern in entsprechenden Labors durchgeführt werden, als auch internetbasierte Tests, die computergestützt von zuhause oder von unterwegs über das Internet bearbeitet werden können. Darüber hinaus ermöglichen portable Endgeräte wie z. B. Tablet-PCs oder PDAs eine computergestützte Erfassung psychologischer Phänomene unter Alltagsbedingungen. Internetbasierte Diagnostik beschreibt nun jenen Teilbereich, bei denen der diagnostische Prozess über das Internet realisiert wird. Die Teststeuerung erfolgt dabei in der Regel durch einen zentralen, örtlich entfernten Testserver, der alle relevanten Informationen für den Testablauf über das Internet an den jeweiligen Computer des Probanden sendet und dessen Reaktionen (z. B. Antworten auf Testfragen) speichert bzw. gleich auswertet. Dafür wird in der Regel lediglich ein regulärer Computer mit Internetanschluss benötigt. Der Testablauf erfordert häufig keine zusätzliche Software und erfolgt mittels der vorhandenen Funktionalität eines Webbrowsers¹. Internetbasierte Testungen sind dabei hinsichtlich der Methodenvielfalt gegenüber traditionellen Computertests keineswegs eingeschränkt. Bei entsprechender Gestaltung sind die meisten vorliegenden diagnostischen Verfahren auch als internetbasierte Varianten realisierbar, wie beispielsweise reaktionszeitsensitive Tests (z. B. Implizite Einstellungstests; vgl. Nosek, Banaji & Greenwald, 2002) oder Verfahren mit neuartigen Antwortformaten (z. B. Graphische Analogskalen; vgl. Gnambs, 2008) zeigen.

Neben dem Ausmaß der Technologisierung ist das Maß der Kontrollierbarkeit die zweite zentrale Dimension zur Beschreibung computergestützter Verfahren. Nach Bartram (2006, S. 31) können folgende Varianten der Testadministration unterschieden werden, die auch als Grundlage für die Ausarbeitung der „International Guidelines on Computer-Based and Internet-delivered Testing“ (International Test Commission, 2006) dienen:

- *Unkontrollierte Durchführung* („unproctored testing“):
Bei unkontrollierter Testvorgabe erfolgt keine individuelle Identifizierung der Probanden und keine Kontrolle der Testdurchführung, also ob der Test alleine bearbeitet wird oder ob unerlaubte Hilfsmittel Verwendung finden. Bei

¹ Ein Computerprogramm, um Inhalte im Internet betrachten zu können

internetbasierten Testungen handelt es sich um frei zugängliche Testverfahren, die in das Internet gestellt werden und für die über geeignete Rekrutierungsstrategien, z. B. Ausschreibungen in öffentlichen Email-Listen oder Diskussionsforen, Personen zur Teilnahme eingeladen werden.

- *Kontrollierte Durchführung:*

Eine einfache Form der Kontrolle erfolgt, wenn ein Test nicht allgemein, sondern nur einem eingeschränkten Teilnehmerkreis zugänglich gemacht wird. Dies kann erfolgen, indem der Zugang zum Test durch Nutzernamen und Passwort geschützt ist. Dies erlaubt zwar eine rudimentäre Identifizierung der Probanden, doch kann auch hier nicht zweifelsfrei festgestellt werden, ob die Testbearbeitung tatsächlich vom Probanden alleine oder aber mit Unterstützung durch Dritte erfolgt ist.

- *Überwachte bzw. angeleitete Durchführung („proctored testing“):*

Eine eindeutige Identifizierbarkeit der Probanden kann durch einen Testleiter gewährleistet werden, der zu Testbeginn persönlich die Identität des Teilnehmers überprüft und zur Vermeidung von Manipulationsversuchen den gesamten Testverlauf überwacht. Dies kann beispielsweise in regionalen Testzentren (bspw. in einer Filiale eines Unternehmens oder bei der örtlichen Arbeitsagentur) mit offizieller Identitätsprüfung durch neutrale Personen erfolgen (Bartram, 2000; Ployhart, Weekley, Holtz & Kemp, 2003). Das Vorgehen unterscheidet sich hierbei nicht von traditioneller Computerdiagnostik, bei der die Probanden fortlaufend überwacht werden, während sie die Verfahren am Computer bearbeiten. Werden biometrische Erfassungsgeräte wie z. B. Fingerabdruck- oder Irisscanner benutzt, die bereits teilweise in herkömmliche Computersysteme integriert sind, lassen sich auch Varianten (eingeschränkt) überwachter Testdurchführungen realisieren, bei denen Proband und Testleiter räumlich getrennt agieren und die Identitätskontrolle sowie Überwachung der Umgebungsbedingungen über internetbasierte Videoübertragungen erfolgt, die den gesamten Bearbeitungsprozess der Testung transparent und somit nachvollziehbar machen.

Für individualdiagnostische Zwecke dominieren unkontrollierte Internettestungen derzeit vor allem in Form von Selbsttests, z. B. als Entscheidungshilfe zur Studienfachwahl

(Zimmerhofer, Heukamp & Hornke, 2006) oder als Negativ-Screening² im Recruitingprozess (Wottawa & Woike, 2002). Internetbasierte Diagnostik ohne bzw. mit nur eingeschränkter Identifizierbarkeit der Probanden bietet sich derzeit primär an, wenn lediglich anonyme Gruppenaggregate gebildet werden sollen und personenbezogene Einzelleistungen nicht gefragt sind. Ein Beispiel hierfür aus der Personalentwicklung sind internetbasierte 360 Grad-Feedbacks (Penny, 2003). Aufgrund der Weiterentwicklung des Hogrefe-Testsystems (<http://www.hogrefe-testsystem.com>) wie auch des Wiener Testsystems (<http://www.schuhfried.at>) zur Anwendung der inkludierten Verfahren über das Internet ist jedoch speziell in der psychologischen Praxis künftig eher mit einer zunehmenden Anwendung überwachter Internettestungen zu rechnen. Da diese keine kostspieligen Investitionen in eine bestimmte technische Infrastruktur (z. B. eine spezifische Computerhardware) voraussetzt, fördert dies verstärkt On-Demand Testungen auch für Anwenderkreise, die sich bislang diagnostische Verfahren nur selten leisten konnten.

Zusammenfassend folgt daraus folgende Abgrenzung von computer- und internetbasierter, psychologischer Diagnostik:

1. *Computerbasierter Diagnostik* betont vor allem das Durchführungsmedium, also den Computer, als zentrales Element der Datenerhebung. Das Testmaterial kann dabei sowohl lokal wie auch auf einem über das Inter- oder Intranet verbundenen Computer gespeichert sein.
2. *Internetbasierte Diagnostik* hebt darüber hinaus den Ort des Testmaterials und der Datenspeicherung hervor (*Lokalität*). Die Testpräsentation erfolgt von einem entfernten Webserver über das Internet.

Sowohl computerbasierte als auch internetbasierte Diagnostik kann unkontrolliert als auch kontrolliert (Modus) an anfallenden Stichproben wie echten Zufallsstichproben (Selektion) erfolgen.

2 Internetbasierte Erhebungstechniken

Das Internet stellt für psychologisch-diagnostische Verfahren grundsätzlich keine komplett neuen Erhebungstechniken zur Verfügung, die über Testen, Fragen und Beobachten hinausgehen würden. Aufgrund der neuartigen Rahmenbedingungen weisen diese jedoch eine

² Um jene Personen auszufiltern, die trotz der erleichterten Bedingungen unkontrollierter Internet-Testungen extrem schlechte Testleistungen erzielen.

Reihe mediumsspezifischer Besonderheiten auf. Im Folgenden werden die gebräuchlichsten Erhebungsverfahren internetbasierter Forschung (nämlich internetbasierte Tests und Fragebögen, internetbasierte Interviews und Gruppendiskussionen sowie internetbasierte Beobachtungen) kurz skizziert.

2.1 Internetbasierte Tests und Fragebögen

Die Unterscheidung in internetbasierte Tests und Fragebögen ist weitgehend willkürlich. Unter der weit gefassten Begrifflichkeit des Fragebogens werden alle mehr oder weniger standardisierten Befragungsformen – in der kommerziellen Marktforschung ebenso wie in der wissenschaftlichen Forschung - zusammengefasst. Mit Tests wird lediglich jener Teilbereich explizit psychometrischer Verfahren bezeichnet, die genuin psychologische Konstrukte operationalisieren. Diese umfassen jedoch nicht ausschließlich Leistungstests von kognitiven Fähigkeiten sondern Fragebögen ebenso wie objektive und implizite Verfahren zur Erfassung von Persönlichkeitsmerkmalen. Internetbasierte Tests wie Fragebogenverfahren gleichermaßen lassen sich nach Batinic und Bosnjak (2000) durch folgende fünf Kernmerkmale beschreiben:

1) Alokaliät

Die Teilnahme an internetbasierten Untersuchungen steht grundsätzlich jedem, der über einen entsprechenden Computer mit Internetanschluss verfügt, offen – unabhängig vom Ort des Testleiters und Probanden. Durch die weltweite Verfügbarkeit des Internets, auch über geographische Grenzen hinweg, eröffnen sich nun Chancen mit vergleichsweise geringem Aufwand internationale und vor allem auch interkulturelle Studien (Breneman, Geuens, Weijters, Smith & Swinyard, 2005) zu realisieren. Insbesondere global operierende Unternehmen verwenden daher bereits seit einigen Jahren verstärkt internet- bzw. intranetbasierte Diagnostik zur Personalauswahl und –entwicklung (Batinic & Appel, 2009; Thompson & Surface, 2007).

2) Automatisierbarkeit

Der gesamte Ablauf internetbasierter Tests und Befragungen ist vor Untersuchungsbeginn festgelegt und auch programmiert, sodass während des Testverlaufs keinerlei Einbindung des Testleiters notwendig ist. Alle Phasen der Datenerhebung, von der Instruktion, der Testvorgabe inklusive dynamischer Itemauswahl u. ä. bis hin zur automatisierten Speicherung der Nutzerreaktionen (Itemantworten, Reaktionszeiten etc.), erfolgen vollständig automatisiert.

3) *Dokumentierbarkeit*

Durch die automatische Protokollierung relevanter Prozessvariablen, wie z. B. Befragungszeitpunkt, Bearbeitungszeit einzelner Items oder Anzahl und Dauer von Unterbrechungen, kann der gesamte Testverlauf transparent gemacht werden. Derartige non-reaktive Daten stehen post-hoc für Validitätsanalysen zur Verfügung, um die Antwortgüte einzelner Teilnehmerdaten zu evaluieren (z. B. bei ungewöhnlich langen oder kurzen Bearbeitungszeiten).

4) *Flexibilität*

Vorteilhaft erweisen sich internetbasierte Befragungen auch darin, relativ unkompliziert multimediales Reizmaterial (wie Fotos, Animationen, Audio und Videomaterial u. ä.) in den Testverlauf einzubinden. Abseits einfacher Befragungstechniken ermöglichen internetbasierte Testverfahren auch zahlreiche komplexere Testansätze. So erlauben beispielsweise reaktionszeitsensitive Messungen auf Itemebene die valide Messung impliziter Einstellungen und Persönlichkeitsmerkmale (vgl. Bluemke & Friese, 2008; Penke, Eichstaedt & Asendorpf, 2006). Schließlich bieten sich gerade auch für Interneterhebungen adaptive Testformen an, bei denen die einzelnen Items sukzessive in Abhängigkeit der voran gegangenen Antworten vorgegeben werden, um die Testdauer möglichst gering zu halten (vgl. Lesage, Riopel, Raiche & Sodoke, 2008).

5) *Ökonomie*

Das primäre Motiv für den Einsatz internetbasierter Verfahren ist in vielen Fällen ein wirtschaftliches. Bei Einsatz entsprechender Softwareprodukte, z. B. mit den Open Source Anwendungen „dynQuest“ (Rademacher & Lippke, 2007) oder „TestMaker“ (Milbradt, Zimmerhofer & Hornke, 2007), kann grundsätzlich davon ausgegangen werden, dass sich die Testerstellung, die -erprobung wie auch -durchführung deutlich verkürzen lässt.

2.2 Internetbasierte Interviews und Gruppendiskussionen

Mithilfe des Internets können Einzelinterviews wie auch Gruppendiskussionen realisiert werden. Diese lassen sich in synchrone und asynchrone Varianten unterscheiden (Gnambs & Batinic, 2007). Bei synchronen Verfahren treten die Gesprächspartner in eine zeitgleiche Interaktion (vgl. Chaney & Dew, 2003). Synchrone Interviews werden im Internet über Web-Chats oder Videokonferenzanwendungen realisiert. Generell gilt, dass synchrone Verfahren

zeitökonomischer ausfallen, da wie im persönlichen Gespräch auf gestellte Fragen eine unmittelbare Reaktion ohne Zeitverzögerung erfolgt. Jedoch muss konstatiert werden, dass internetbasierte Chat-Interviews in der Regel deutlich länger (fast doppelt so lange) wie vergleichbare persönliche Interviews dauern (Reid & Reid, 2005). Bei synchronen schriftlichen Befragungen werden die Antworten somit stark von der Lese- und Tippgeschwindigkeit der Interviewten beeinflusst. Dies verdeutlicht auch ein Methodenexperiment von Erdogan (2001), die persönliche und online geführte Interviews gegenüberstellte. Im persönlichen Gespräch produzierten die Teilnehmer der Gruppendiskussionen einen bis zu drei Mal längeren Redebeitrag als in den Internetgesprächen. Andererseits meldeten sich die Teilnehmer im Chat häufiger zu Wort. Chat-Interviews führen demnach zwar zu einer höheren Frequenz der Wortmeldungen, bei denen allerdings insgesamt weniger Wörter produziert werden. In qualitativer Hinsicht unterscheiden sich beide Interviewformen jedoch kaum. Chat-Diskussionen wie Face-to-Face Diskussionen führen inhaltlich zu vergleichbaren Ergebnissen, obwohl internetbasierte Diskussionen häufig kontroverser ablaufen und im Chat Beiträge von anderen eher kritisch kommentiert werden.

Im Gegensatz dazu zeichnen sich asynchrone Interviewformen durch zeitversetzte Frage-Antwort-Runden aus; die Gesprächspartner müssen nicht zur selben Zeit am Computer im Internet sein (Turney & Pocknee, 2005). Asynchrone Interviewvarianten werden schriftlich über Email, Mailinglisten oder internetbasierte Diskussionsforen durchgeführt. Bei Email-Befragungen werden den Befragten die Fragen einzeln oder in thematischen Blöcken zur Beantwortung übermittelt. Dabei ist darauf zu achten nicht zu viele Frage auf einmal zu stellen, um die Probanden nicht zu überfordern. Offen zu beantwortende Fragen werden in der Regel anstrengender erlebt als vergleichbare Fragen mit vorgegebenen Antwortalternativen. Um Befragte mit einer zu langen Frageliste nicht zu entmutigen (z. B. durch Ermüdung oder Frustration), ist auf ein ausgewogenes Verhältnis von Fragen und Befragungsepisoden zu achten. E-Mail ist eine Kommunikationsform, die auf eher kurze und in sich geschlossene Texteinheiten aufgebaut ist (Bampton & Cowton, 2002).

Besondere Attraktivität gewinnt das Internet auch aufgrund der bislang kaum vorhandenen Möglichkeit anonyme Kommunikationsszenarien umzusetzen. Anonymisierte online Interaktionen ermöglichen es periphere Hinweisreize auszuschalten, die persönliche Gespräche häufig unbewusst mitbestimmen, und können mithelfen kulturell bedingte

Hierarchien aufgrund z. B. des Geschlechts, Alters oder physischer Charakteristika in den Hintergrund zu drängen (Al-Saggaf & Williamson, 2004). Dies führt zu einer stärkeren Egalisierung der Kommunikationspartner und ermöglicht eine offeneren Gesprächsführung, insbesondere bei Gesprächen über sensible möglicherweise beschämende Themen, über die im persönlichen Dialog nur schwer gesprochen würde - wie beispielsweise im Bereich der Sexualforschung (Davis, Bolding, Hart, Sherr & Elford, 2004).

Die Kommunikation erfolgt bei internetbasierten Interviews derzeit zumeist schriftlich. Diagnostisch problematisch stellt sich hierbei das Fehlen non-verbaler Reize wie Stimmlage, Gestik und Mimik dar, welche die Interpretation von Gesprächsinhalten maßgeblich mitzubestimmen helfen. Die Verwendung von Emoticons, kurzen textlichen oder graphischen Repräsentationen emotionaler Inhalte, kann hierfür nur eingeschränkt Abhilfe schaffen, da diese kaum dieselbe Bandbreite wie die menschliche Emotion aufzuweisen vermögen (Mann & Stewart, 2002). Deshalb können Sarkasmen, Metaphern und ähnlich zweideutige Aussagen, die im persönlichen Gespräch in der Regel einfach als solche zu identifizieren sind, bei Fehlen der non-verbaler Ebene in den schriftlichen Äußerungen leicht als Beleidigung missverstanden werden (Davis et al., 2004) und aufgrund einer fehlerhaften Interpretationen zu Kommunikationsbeeinträchtigungen führen. Mit dem zunehmenden Aufkommen von Internet-Telefonie (insbesondere auch Video-Telefonie) können internetbasierte Interviewsituationen jedoch wieder natürlicher gestaltet werden, sodass eine Abkehr von schriftlichen Interviewformen über Text-Chat hin zu verbalen Dialogen über Videosysteme zu erwarten ist.

2.3 Internetbasierte Beobachtungen

Beobachtung meint das Sammeln von Informationen in einem *nicht-kommunikativen* Prozess, bei dem ein Objekt *aktiv* zum Zentrum der eigenen Aufmerksamkeit gemacht wird (Bortz & Döring, 2006). Dieser Wahrnehmungsprozess bezieht sich dabei primär auf die „sinnliche Erfassung mittels Sehen und Hören“ (Kanning, Hofer & Schulze Willbrenning, 2004, S. 58). Im diagnostischen Prozess stellen Beobachtungen einen, wenn auch manchmal vernachlässigten, elementaren Bestandteil der Informationsgewinnung dar. Besonders Beobachtungen während der Explorationsphase helfen häufig einen ersten Eindruck über Probanden zu gewinnen und unspezifische Hypothesen über zugrundeliegende Problemlagen zu formulieren. Internetbasierte Beobachtungen sind in der Regel nur schwer mit traditionellen Beobachtungsvarianten vergleichbar, welche das gesamte, willkürliche wie

unwillkürliche, Verhalten der Probanden der Beobachtung erschließen. Im Internet sind lediglich jene Verhaltensdaten der Beobachtung zugänglich, die auf explizit gesetzte Aktionen des Nutzers über das Eingabemedium, die Maus oder die Tastatur, zurückzuführen sind. Begleitende, möglicherweise unbewusste Verhaltensweisen wie z. B. Körperhaltung, Blickbewegung oder Gestik bleiben internetbasierten Beobachtungen jedoch in der Regel³ verschlossen.



Abbildung 1. TeilnehmerInnen in Second Life (New Media Consortium, 2007)

Im Internet werden Beobachtungen von Verhalten einzelner Personen sowie von komplexen Interaktionen zwischen Probanden durch graphische Simulationen ermöglicht. In computergenerierten 3D-Welten wie Second Life (<http://www.secondlife.com>) werden am Bildschirm ganze virtuelle Umgebungen visualisiert, in denen sich Probanden frei bewegen

³ Video-basierten Kommunikationsszenarien sind im Rahmen dieser Betrachtungen unberücksichtigt.

und mit anderen Teilnehmern interagieren können (vgl. Abbildung 1). Der Beobachtung sind hierbei nur wenig Grenzen gesetzt: Von Merkmalen der von den Teilnehmern mitgestalteten virtuellen Umgebung über das Aussehen der eigenen Personen in der virtuellen Welt (sogenannter Avatare) bis hin zu Merkmalen des verbalen wie non-verbale Interaktionsverhalten zwischen den Teilnehmern. Für individualdiagnostische Zwecke ist die Anwendung systematischer internetbasierter Beobachtungen zur Erfassung psychologischer Konstrukte bislang allerdings noch wenig verbreitet. Als größtes Hindernis ortet Washburn (2003) das Fehlen geeigneter Softwarelösungen, welche die unkomplizierte Erfassung von Verhaltensdaten ermöglicht. Der Einsatz kommerzieller Produkte, z. B. bestehender Computerspiele, verschafft nur in den wenigsten Fällen Abhilfe, da diese häufig keine oder eine nur unzureichende diagnostisch relevante Variation spezifischer Merkmale der Simulationsumgebung ermöglichen, um die interessierenden psychologischen Konstrukte adäquat zu erheben. Zudem erweist sich die automatische Aufzeichnung des interessierenden Verhaltens bei derartigen Anwendungen nicht immer als unkompliziert. Neuere Ansätze versuchen sich daher in der expliziten Konstruktion derartiger 3-D Welten auf Basis frei verfügbarer Softwarearchitekturen (Frey, Hartig, Ketzler, Zinkernagel & Moosbrugger, 2007), in denen Probanden unter vorgegebenen, den diagnostischen Zielsetzungen entsprechend variierten Aufgabenstellungen frei zu agieren haben. Wie in einem Computerspiel können sie verschiedene Handlungen setzen und mit ihrer virtuellen Umwelt, simulierten wie auch realen Personen, interagieren. Der Gestaltung derartiger Simulationen und den darin zu absolvierenden Aufgaben und Handlungsmöglichkeiten sind wenig Grenzen gesetzt; so lassen sich beispielsweise gesamte Produktionsabläufe in Unternehmen, bei denen die Probanden eine Reihe von Entscheidungen zu treffen haben, ebenso simulieren wie in Anlehnung an Dörners (1989) Lohhausen-Simulation ganze Städte, deren Geschicke man als Bürgermeister zu lenken hat. Aufgrund der Computervorgabe können nicht nur das gesamte Geschehen selbst sondern auch spezifische Verhaltensweisen der Probanden direkt aufgezeichnet werden: z. B. der variierende Abstand zwischen den Teilnehmern während der Lösung einer gemeinschaftlichen Aufgabe. Einen Prototyp derartiger Simulationen stellen Frey, Blunk und Banse (2006) mit Psi-Land vor, einer virtuellen Umgebung zur Untersuchung von Paarinteraktionen. Die Probanden haben in einer virtuellen Spielumgebung in Form einer einsamen Insel als Paar gemeinschaftlich verschiedene Aufgabenstellungen zu lösen, die paarspezifische Verhaltensmuster aktivieren sollen: z. B. Schwimmen zu einer anderen Insel, obwohl nur eine Schwimmweste zur Verfügung steht. Automatisiert wird dabei eine Reihe von Verhaltensdaten erfasst. Beispielsweise werden die Distanz und der Blickkontakt

zwischen den Partnern, als Indikatoren für das Interaktionsverhalten mitprotokolliert, welche nachträglich mit anderen Konstrukten wie Persönlichkeit oder Einstellung in Beziehung gesetzt werden können. Wenn sich derartige Szenarien bislang auch noch am Anfang der Entwicklung befinden, erscheint eine Weiterentwicklung im Sinne internetbasierter Assessment-Center (Eichstaedt, 2005) für die nahe Zukunft vielversprechend, um verschiedene Personen ortsunabhängig gleichzeitig in derartigen Testwelten zusammenzuführen und kognitive Kompetenzen wie auch soziales Interaktionsverhalten simultan zu erfassen (Jude & Wirth, 2007).

3 Umsetzung diagnostischer Aufgaben mithilfe Neuer Medien

Nachfolgend werden ausgewählte Anwendungsgebiete internetbasierter Diagnostik vorgestellt. Diese Darstellung erhebt allerdings keinen Anspruch auf Vollständigkeit, sondern versucht lediglich die Bandbreite internetbasierter psychologischer Diagnostik zu verdeutlichen.

3.1 Exploration und Anamnese

Obleich internetbasiertes Testen bislang den Großteil gängiger Bemühungen darstellt den psychologisch-diagnostischen Prozess auf das Internet auszudehnen, finden sich auch verschiedentlich Versuche, das Internet im Explorationsgespräch zu verankern. Der Dialog wird dabei entweder ausschließlich in textlicher Form über Chat oder Email oder über Audio- und/oder Video mittels Instant-Messenger und Webcams realisiert. Vorteile ergeben sich aufgrund (Rees & Haythornthwaite, 2004):

- *Geringer Kosten*, wenn Proband und Diagnostiker räumlich entfernt sind und für ein persönliches Gespräch weite Anreisen in Kauf genommen werden müssten, z. B. in ländlichen Regionen mit einer geringen Dichte professioneller Hilfsangebote.
- *Schneller Verfügbarkeit* verschiedener Experten mit unterschiedlicher Spezialisierung, deren Unterstützung in Krisensituationen zeitnah in Anspruch genommen werden kann.
- *Automatisierter Gesprächsaufzeichnungen*, die eine detaillierte Auswertung der Gesprächsinhalte auch im Nachhinein ermöglichen.

- Eines *niederschwelligeren Zugangs* zu professioneller Hilfe auch für Personen, die sonst nur schwer in Kontakt mit anderen treten können (z. B. Personen mit Angststörungen).

Angesichts der schneller werdenden Internetverbindungen stellt die Methode der internetbasierten Exploration mit Videoübertragung zumindest für Erstgespräche oder ein grobes Screening der Problemlage in vielen Fällen eine realistische Alternative zum herkömmlichen Vorgehen dar (Barak & Buchanan, 2004). Dabei müssen nicht notwendigerweise qualitative Einschränkungen hinsichtlich des erhobenen Informationsreichtums in Kauf genommen werden; wie verschiedene methodenübergreifende Vergleichsstudien (Hilty, Marks, Urness, Yellowlees & Nesbitt, 2004; Lexcen, Hawk, Herrick & Blank, 2006; Matsuura et al., 2000) nahe legen, führen internetbasierte Explorationsgespräche anhand strukturierter Interviews zu vergleichbaren Erstdiagnosen wie Klassifikationen anhand traditioneller Interviewformen.

Rees und Haythornthwaite (2004) führen jedoch auch verschiedene Hürden an, welche der Durchführung internetbasierter Explorationsgespräche entgegenstehen:

- Insbesondere wenig technikaffine Personenkreise, die nur geringe Erfahrung mit internetbasierten Kommunikationsszenarien besitzen, können durch die ungewohnte Interviewsituation, Unsicherheit bezüglich des Umgangs mit der neuartigen Technologie (z. B. der richtigen Bedienung der Videosoftware) aufweisen, welche zu einer gespannten und gehemmten Gesprächsführung beitragen kann.
- Trotz qualitativ immer leistungsfähigeren Videoanwendungen im Internet muss in der Regel davon ausgegangen werden, dass die informatorische Reichhaltigkeit der non-verbale Kommunikationsebene eingeschränkt ist. Zum einen erscheinen bei kleinformatischen Videos die Konturen zumeist etwas unscharf, wodurch Gesichtsausdrücke manchmal schwierig einwandfrei zu identifizieren sind. So kann beispielsweise Konzentration leicht mit Desinteresse verwechselt werden. Des Weiteren wird selten die gesamte Person übertragen sondern lediglich das Gesicht, wodurch non-verbale Cues wie Körperhaltung oder Gestik ebenfalls höchstens eingeschränkt beobachtet werden können.

- Aufgrund der wenig kontrollierbaren Umgebungsbedingungen der Gesprächssituation (zumeist in den privaten Räumlichkeiten der Probanden) sind Störungen des Interviews durch z. B. Telefon oder andere Personen nur schwer auszuschließen.

Zusammenfassend lässt sich hieraus ableiten, dass das Internet bereits heute eine Reihe vielversprechender Möglichkeiten für die psychologische Praxis bietet, die über die Anwendung psychometrischer Verfahren hinaus gehen: Von der einfachen Sammlung der Vor- und Krankengeschichte mithilfe standardisierter Fragebögen oder Email-Befragungen über die Exploration der Problemlage mit Videointerviews bis zur Durchführung therapeutischer Interventionen über bzw. mithilfe des Internet (vgl. Eichenberg, 2008) eröffnen neue Medien mannigfaltige Anwendungsgebiete.

3.2 Persönlichkeitsdiagnostik

Persönlichkeitstests stellen die bislang populärste Gruppe psychologisch-diagnostischer Verfahren im Internet dar. Nicht zuletzt aufgrund der unkomplizierten technischen Umsetzbarkeit von Selbstbeschreibungsverfahren werden diese seit Mitte der 90er Jahre im Internet realisiert. In vielen Fällen handelte es sich hierbei um etablierte, für die Papier-Bleistift-Vorgabe entwickelte Instrumente, deren Items im Web repliziert wurden. Für den praktischen Einsatz von Persönlichkeitsverfahren im Internet erwies sich jedoch alsbald erschwerend, dass die Anwendung kommerzieller Testverfahren in frei zugänglichen Online-Erhebungen häufig lizenzrechtlichen Bestimmungen zuwider läuft. Herausgeber psychologischer Verfahren untersagen in der Regel den Einsatz ihrer Instrumente im Internet aufgrund von Befürchtungen vor deren unkontrollierten Verbreitung. Aufgrund dieser Problematik wurden mittlerweile im Rahmen des International Personality Item Pool (<http://ipip.ori.org>) eine Reihe frei verfügbarer Persönlichkeitsskalen konstruiert, darunter auch Varianten des NEO-FFI und 16PF, die als „Public Domain“ zur freien Verwendung veröffentlicht wurden und speziell auch in wissenschaftlichen Online-Untersuchungen eingesetzt werden können (Goldberg et al., 2006). Eine ähnliche Initiative (<https://netfiles.uiuc.edu/jrounds/IIP/home.htm>) wurde unlängst auch für berufliche Interessenstests vorgestellt (Armstrong, Allison & Rounds, 2008). Diese Instrumente wurden mittlerweile in zahlreichen Sprachen und unter unterschiedlichen Einsatzbedingungen validiert (vgl. Ehrhart, Roesch, Ehrhart & Kilian, 2008; Mlacic & Goldberg, 2007), wodurch sie sich insbesondere auch für mehrsprachige Forschungsvorhaben anbieten.

Eine Alternative zu Selbstbeschreibungsverfahren, die Traits auch unabhängig von der Fähigkeit zur Introspektion und Einflüssen sozialer Erwünschtheit messen sollen, bietet die objektive Persönlichkeitsdiagnostik. Im Sinne experimentalpsychologischer Verhaltensdiagnostik (Kubinger, 2006) werden dabei von Individuen in einer standardisierten Situation Verhaltensdaten erfasst, um daraus zugrunde liegende Persönlichkeitseigenschaften abzuleiten. Obwohl objektive Persönlichkeitsdiagnostik heute zumeist computerbasiert erfolgt, konnten sich bislang erst wenig internetbasierte Varianten etablieren. Ein in den letzten Jahren zunehmend populärer objektiver Test stellt der Implizite Assoziationstest (IAT) von Greenwald et al. (1998) dar. Ursprünglich in der Sozialpsychologie zur Messung von Einstellungen und Vorurteilen verortet, liegt die Grundintention des Tests in der Messung individueller Unterschiede. Dazu verwendet er keine expliziten Selbstberichte sondern schließt auf Basis von Reaktionszeiten bei Aufgaben, die eine möglichst rasche Entscheidung zwischen zwei einfachen Wahlalternativen erfordert, indirekt auf zugrunde liegende Persönlichkeitsmerkmale. Reaktionszeitmessungen über das Internet sind grundsätzlich keine unkomplizierte Angelegenheit, da die Datenerhebung nicht wie im Labor standardisiert auf explizit eingerichteten Testcomputern erfolgt, die allen Probanden dieselben Bedingungen bieten. Variationen in der Computerausstattung der Teilnehmer und vor allem –auslastung während der Testdurchführung können zusätzliche Fehlervarianz generieren und Reaktionszeitmessungen verzerren. Über verschiedene Korrekturalgorithmen (vgl. Eichstaedt, 2001) können diese Fehler zum Teil zwar reduziert aber nicht gänzlich aufgehoben werden. Wie Methodenexperimente allerdings zeigen, ist dieser zusätzliche technologiebedingte Fehler bei Interneterhebungen in der Regel gering und führt zu nur unwesentlich ungenaueren Daten (Reimers & Stewart, 2007). Insbesondere internetbasierte IATs haben sich in der Vergangenheit als valide Erhebungsinstrumente erwiesen (vgl. Houbjen & Wiers, 2008; Nosek et al., 2002).

3.3 Intelligenz- und Leistungsdiagnostik

Leistungstests zur Erfassung verschiedener Teilbereiche der allgemeinen Intelligenz finden sich erst seit kurzem als internetbasierte Anwendungen. Während Fragebogenverfahren, insbesondere zur Erfassung unterschiedlicher Persönlichkeitsmerkmale, von Beginn an das Forschungsinteresse prägten, wurde der kognitive Bereich weitgehend vernachlässigt. Eine der ersten Anwendungen beschreiben Preckel und Thiemann (2003) in einer Validierungsstudie eines Online-Matrizentests zur Abschätzung intellektueller Hochbegabung. Die Ergebnisse dieser Untersuchung demonstrieren das größte Hemmnis für internetbasierte

Leistungstests, die zu Forschungszwecken an Freiwilligen-Stichproben appliziert werden – große Abbruchquoten aufgrund häufig sehr langer Bearbeitungszeiten (in vorliegendem Fall: beinahe 120min). Von über 3.000 Personen, welche die Testbearbeitung ursprünglich in Angriff nahmen, verblieben aufgrund von vorzeitigen Abbrüchen, fehlenden Werten u. ä. lediglich knapp 200 Personen in der Validierungsstichprobe. Die hohe Dropout-Rate deutet auf einen starken Selbstselektions-Effekt hin; lediglich hoch motivierte oder besonders leistungsstarke Teilnehmer bearbeiteten den Test tatsächlich bis zum Ende. Derartige Selbstselektionsprozesse erschweren nicht nur die Rekrutierung ausreichend großer Stichproben für wissenschaftliche Online-Studien, sondern können aufgrund der resultierenden, sehr spezifischen Stichprobe die Ergebnisse massiv beeinträchtigen – insbesondere bei Experimentaldesigns, wenn sich die Abbruchquoten nicht gleichmäßig auf die einzelnen Versuchsbedingungen verteilen (Gnambs & Strassnig, 2007).

Dieses Problem trifft vornehmlich bei unkontrollierten Internetstudien zu. Dass durchaus auch anspruchsvolle Online-Erhebungen mit umfangreichen Testbatterien über das Internet umgesetzt werden können, wenn die Rahmenbedingungen entsprechend geplant werden, belegt eindrucksvoll eine umfangreiche Zwillingsstudie ($N > 1000$), die als internetbasierte Längsschnittuntersuchung an Kindern im Alter von 10 Jahren realisiert wurde (Haworth et al., 2007). Zwillingsstudien stellen seit jeher ein herausforderndes Unterfangen dar, da es sich um eine sehr selektive Stichprobe handelt, die nur schwer zu erreichen und häufig nur durch örtlich sehr verstreute Erhebungen zu rekrutieren ist. Ausgangspunkt für die Studie von Haworth et al. (2007) stellte eine repräsentative Stichprobe in England geborener Zwillingspaare dar. Die internetbasierte Testbatterie umfasste verschiedene Verfahren zur Erfassung mathematischer, verbaler (Wortverständnis, Wortflüssigkeit und Sprachbeherrschung) sowie kognitiver Fähigkeiten. Aufgrund der sehr umfassenden Testbatterie und der damit verbundenen langen Bearbeitungszeit konnten die Verfahren nicht auf einmal vorgegeben werden. Deshalb konnten die Kinder die Bearbeitung nach freiem Ermessen über mehrere Wochen verteilt vornehmen. Da weder die Kinder selbst noch deren Eltern über hinreichend Vorerfahrung mit Internettestungen verfügten, musste ein entsprechendes Schulungs- sowie fortlaufendes Betreuungsangebot implementiert werden. Neben ausführlichen Bearbeitungshinweisen und Lösungen für möglicherweise auftretende Probleme, die in einem entsprechenden Manual beschrieben wurden, stand zusätzlich ein kostenfreier telefonischer Beratungsservice zur Verfügung. Die Testung selbst wurde von den Eltern beaufsichtigt, indem sich diese mit den bereit gestellten Zugangsdaten auf einer

verschlüsselten Internetseite anmeldeten und dem Kind den Zugriff auf das Testmaterial gewährten. Die eigentliche Testbearbeitung wurde dann von dem jeweiligen Kind allein vorgenommen. Wie nachträgliche Evaluationsgespräche sowie Ergebnisse einer Validierungsstudie, bei der einige Kinder den Test unter Aufsicht eines Untersuchungsleiters zu wiederholen hatten, nahe legen, wurde das beschriebene Vorgehen weitgehend eingehalten. Wenn bei den Probanden ein entsprechendes Commitment erzeugt werden kann, indem die Rahmenbedingungen der Erhebung vertrauenswürdig gestaltet werden, sind auch über das Internet kognitive Leistungstests umzusetzen. Dies ist allerdings mit einem zusätzlichen Aufwand verbunden, der sich insbesondere bei der Testung großer Stichproben aber relativ rasch amortisieren sollte.

Auch in der Praxis beginnen sich internetgestützte Leistungstests bereits zu etablieren. Ein Beispiel hierfür ist der populäre „Test of English as Foreign Language“ (TOEFL) zur Messung der englischen Fremdsprachenkenntnisse, der seit kurzem als eigenständige Internetversion angeboten wird und das Internet zur objektiven und raschen Datenerfassung bei großen Stichproben nutzt (Educational Testing Service, 2008). Der Test kann weltweit in speziellen Testcentern unter Beaufsichtigung absolviert werden. Die Testdaten werden dabei automatisch über das Internet an die Testzentrale weiter geleitet, wo eine standardisierte (semi-)automatisierte Auswertung erfolgt. Neben verschiedenen Subtests zum Les- und Hörverständnis, die aus der Papier-Bleistift Version übernommen wurden, finden sich in der Internetanwendungen auch zahlreiche neue Verfahren auf Basis realitätsnaher Alltagssituationen, bei denen den Probanden kurze Videosequenzen (z. B. ein englischsprachiger Vortrag) vorgespielt werden, auf deren Basis im Anschluss bestimmte Aufgaben zu bearbeiten sind (z. B. eine schriftliche Zusammenfassung des eben Gehörten, die in ein entsprechendes Textfeld einzugeben ist). Erstmals ermöglicht der TOEFL auch die Erfassung der Sprechfertigkeit. Dazu haben die Probanden bestimmte Aufgabenstellungen mündlich zu beantworten. Die Aussagen werden dabei über das Mikrophon am Computer aufgezeichnet und automatisiert über das Internet an die Testzentrale übermittelt, wo professionelle Rater das Gesprochene beurteilen.

3.4 Klinisch-psychologische Diagnostik

Seit einigen Jahren entdeckt auch die klinisch-psychologische Diagnostik das Internet für sich – sowohl zur wissenschaftlichen Grundlagenforschung wie auch in der Praxis zur Einzelfalldiagnostik (siehe Ott & Eichenberg, 2003). Insbesondere reizt viele Forscher die

Aussicht spezifische Subpopulationen im Internet zu erreichen, die auf herkömmlichem Weg nicht oder nur schwer zu rekrutieren wären. Ein Beispiel hierfür stellt Sexomnia dar, eine äußerst seltene Störung bei der Personen im Schlaf sexuelle Handlungen verüben. Herkömmliche Erhebungstechniken sind aufgrund der sehr geringen Prävalenz der Störung und vor allem der fehlenden bzw. eingeschränkten Bereitschaft der Betroffenen (z. B. aus Scham) darüber persönlich Auskunft zu geben, nur schwierig zu realisieren. Anonyme Interneterhebungen ermöglichen nun auch diese Personengruppe zu erreichen und zu untersuchen (Trajanovic, Mangan & Shapiro, 2007).

Internetbasierte Individualdiagnostik im klinischen Bereich ist bislang ipsativen Anwendungen vorbehalten, bei denen Messwerte innerhalb einer Person verglichen werden - beispielsweise anhand einer Prä- / Postmessung inwiefern ein Behandlungserfolg durch eine bestimmte Intervention gegeben erscheint: z. B. in der Psychotherapieforschung (Reynolds & Stiles, 2007) oder neuropsychologischen Diagnostik (Erlanger et al., 2003; Silverstein et al., 2007). Mit WebNeuro (Silverstein et al., 2007) wurde eine entsprechende Testbatterie zur Erfassung aktueller kognitiver und affektiver Funktionen (Wahrnehmung, Aufmerksamkeit, Gedächtnis etc.) entwickelt, die zeitökonomisch und vor allem valide über das Internet zur Forschungszwecken wie auch zur Einzelfalldiagnostik appliziert werden kann. Vielversprechende Ergebnisse werden im klinischen Setting auch reaktionszeitbasierten Verfahren nachgesagt. Bei emotionalen Stroop-Aufgaben gilt es bei einer Reihe affektiv konnotierter oder neutraler Reize, z. B. Wörter oder Gesichter, die in unterschiedlicher Farbe dargeboten werden, die jeweilige Farbe möglichst rasch zu benennen. Eine deutliche verlangsamte Reaktionszeit in bestimmten klinisch relevanten Personengruppen bei emotionalen Reizen deutet auf aktuelle affektive Einflüsse auf die Informationsverarbeitung hin und wird damit erklärt, dass die aktuelle Aufmerksamkeit auf die Bedeutung des zugrundeliegenden Wortes gelenkt wird, was die Nennung des Farbnamens behindert (Krohne & Hock, 2007). Erste Stroop-Anwendungen im Internet liefern jedoch widersprüchlich Ergebnisse. An studentischen Stichproben konnte der erwartete Stroop-Effekt sowohl bei kontrollierter Vorgabe im Labor wie auch bei unkontrollierter Internetvorgabe zu Hause repliziert werden - wenn auch die durchschnittlichen Reaktionszeiten sowohl für neutrale wie auch affektive Worte länger ausfielen (Linnman, Carlbring, Ahman, Andersson & Andersson, 2006). Bei Anwendung an klinisch relevanten Subpopulationen, bei Sozialphobikern (Andersson, Westöö, Johansson & Carlbring, 2006) oder Tinnitus-Patienten (Andersson, Bakhsh, Johansson, Kaldo & Carlbring, 2005), stellte

sich jedoch ein gegenteiliger Effekt ein: Phobiker reagierten auf affektiv relevante Worte schneller als auf neutrale. Eine mögliche Erklärung hierfür liefert die Vigilance-Avoidance-Hypothese von Andersson et al. (2006), welche zeitlich zwei aufeinanderfolgende Prozesse postuliert, die durch die unterschiedlichen Stroop-Verfahren erfasst werden: Traditionelle Stroop-Aufgaben messen die anfängliche Vigilanz auf die widersprüchliche Reizinformation, während internetbasierte Stroop-Aufgaben, bei der die Farbnennung anstelle von raschen mündlichen Nennung durch einen komplizierteren und zeitaufwändigeren Mausclick erfolgt, nachfolgende Vermeidungsstrategien erfasst. Deshalb müssen traditionelle und internetbasierte Stroop-Aufgaben als zwei unabhängige Instrumente betrachtet werden, die unterschiedliche Konstrukte operationalisieren.

3.5 Berufliche Eignungsdiagnostik

Das Personalmanagement in Arbeitsorganisationen wird in den letzten Jahren zunehmend durch Kommunikations- und Informationstechnologien auf der Basis des Inter- bzw. Intranets unterstützt (Anderson, 2003; Hertel & Konradt, 2004). Sogenanntes Electronic Human Resource Management (E-HRM bzw. e-HR, s. Strohmeier, 2002) umfasst dabei sowohl Verfahren der Personalauswahl, der Personalführung und Verwaltung als auch der Personalentwicklung (Hertel & Schroer, 2008). Berufliche Eignungsdiagnostik mit internetbasierten Verfahren ist ein Hauptanwendungsbereich von E-HRM, und entsprechende Verfahren werden sowohl zur Auswahl neuer Mitarbeiter als auch im Rahmen von Personalentwicklungsmaßnahmen eingesetzt, um zum Beispiel als Potenzialanalyse Voraussetzungen für Weiterbildungsprogramme zu erfassen.

In der Diskussion der Vor- und Nachteile internetbasierter Verfahren der Eignungsdiagnostik ist es hilfreich, zwischen der Rekrutierungsfunktion (*E-Recruitment*) und der eigentlichen Diagnosefunktion (*E-Assessment*) der Verfahren zu unterscheiden (Konradt & Sarges, 2003). Im Personalmanagement von Organisationen wurden zunächst die zusätzlichen Möglichkeiten internetbasierter Verfahren für die Rekrutierung, also die Kontaktierung und Information potenzieller Kandidaten erkannt (vgl. Moser, Zempel & Göritz, 2003). So können internetbasierte Verfahren als Bestandteil von Rekrutierungsportalen im Internet potentielle Bewerber deutlich schneller und kostengünstiger erreichen als konventionelle Rekrutierungsmethoden (Anzeigen in Printmedien, etc.). Darüber hinaus ermöglicht die Verbreitung des Internets eine weltweite Rekrutierung, was besonders für global agierende Großkonzerne attraktiv ist (z .B. Gasteiger & Montel, 2005). Interessierte Bewerber können

internetbasierte Verfahren selbst über große Distanzen mit relativ geringem zeitlichem und finanziellem Aufwand bearbeiten. Zudem bieten internetbasierte Verfahren Organisationen die Möglichkeit, sich im Wettbewerb um die kompetentesten Mitarbeiter als modern und innovativ zu präsentieren (sog. „Employer branding“; z. B. Backhaus, 2004; s. a. Konradt & Rack, 2006) und geeignete Kandidaten schneller als andere Organisationen zu binden.

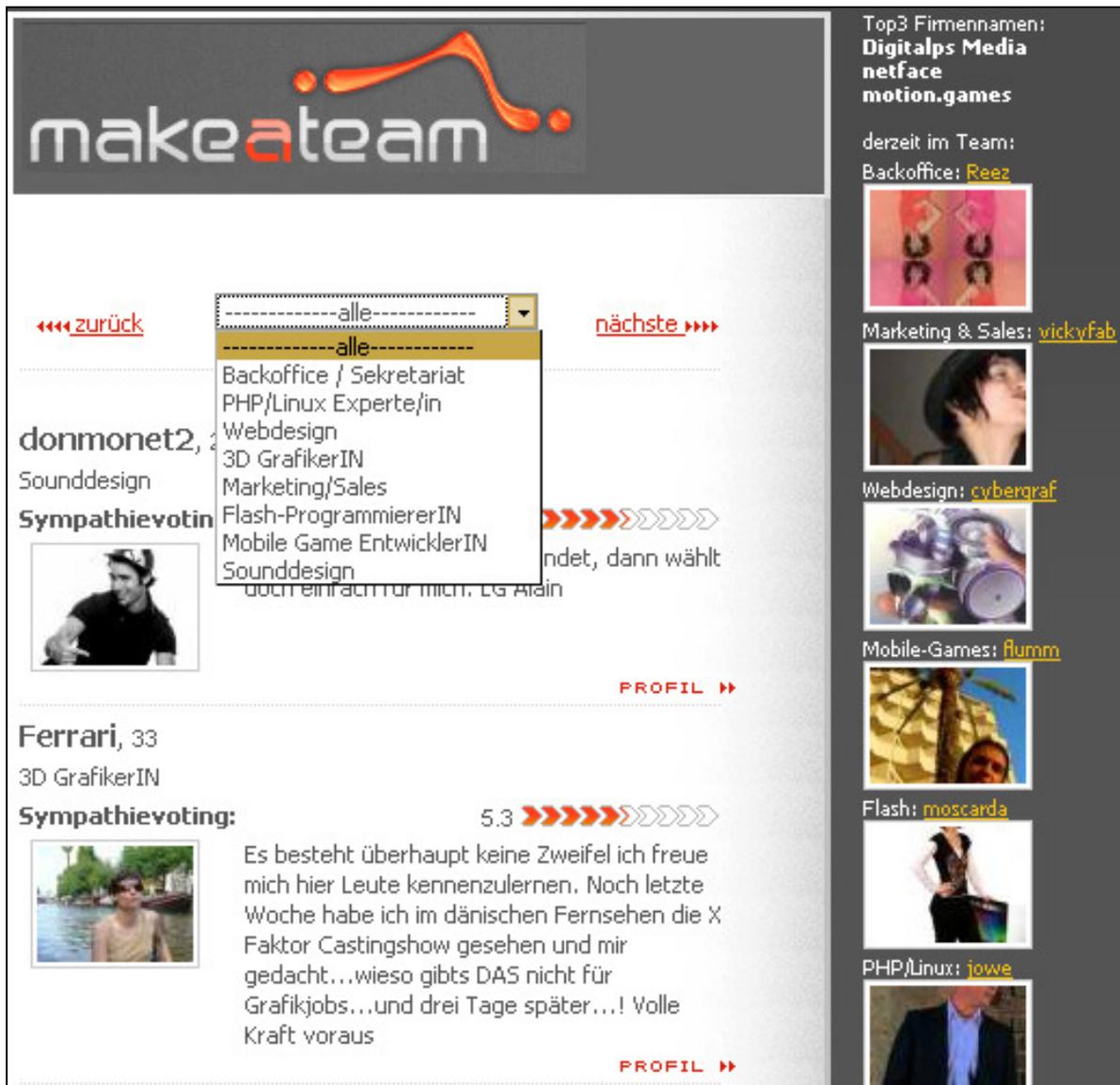


Abbildung 2. Bewerber des Make-A-Team-Castings

Ein aktuelles Beispiel liefert die „Make A Team“-Initiative⁴ eines österreichischen Medienunternehmens, die in der Tradition des Mitmach-Webs (Web 2.0) für den Aufbau eines Tochterunternehmens die Mitarbeiter im Internet unter aktiver Beteiligung der

⁴ <http://www.makeateam.at> (2009-12-20)

Öffentlichkeit auswählt (siehe Abbildung 2). Für diverse berufliche Positionen können sich Interessierte über das Internet durch Veröffentlichung eines selbst zu gestaltenden Profils bewerben. Die Bewerber durchlaufen sodann einen mehrstufigen Bewerbungsprozess, bei dem sowohl eine Expertenjury wie auch die Web-Community (jeder interessierte Internetnutzer) Eignung und Leistung der Bewerber sowohl in praktischen berufsrelevanten Aufgaben, Wissenstests wie auch Persönlichkeitsassessments, die im Laufe mehrerer Wochen online zu absolvieren sind, bewertet werden. Der gesamte Bewerbungsprozess findet öffentlich im Internet statt und stellt gleichermaßen ein Recruitment- wie Marketinginstrument für das neu zu gründende Unternehmen dar, indem die Öffentlichkeit direkt in den Auswahlprozess eingebunden wird.

Im Vergleich dazu finden zusätzliche Möglichkeiten internetbasierter Verfahren für die Diagnosefunktion erst langsam Beachtung. So bestanden internetbasierte Assessment-Verfahren anfangs oft nur aus simplen Übertragungen konventioneller Persönlichkeitsskalen oder biografischer Fragen auf Computer, die mit Computernetzwerken innerhalb einer Organisation (Intranet) oder aber weltweit (Internet) verbunden waren. Zusätzliche Potenziale, die spezifisch für E-Assessment sind, werden erst langsam entwickelt, so zum Beispiel der Einsatz interaktiver Elemente für internetbasierte Interviews oder Assessment Center (Eichstaedt, 2005), in denen computergestützte „virtuelle“ Kooperation und Interaktion zuverlässig dokumentiert und bewertet werden kann. Bei *Karrierejagd durchs Netz*⁵ (Diercks & Weber, 2005) beispielsweise haben die Teilnehmer gleich einem Planspiel mit virtuellen Computerfiguren zu interagieren und verschiedene Aufgabenstellungen zu bearbeiten. Im Rahmen des Spielgeschehens werden gleichsam nebenbei verschiedene biographische Daten sowie Soft Skills (z. B. über interaktive Postkorbaufgaben) erfasst.

Mögliche internetbasierte Verfahren der beruflichen Eignungsdiagnostik sind u. a.:

- Demografische Fragen, inkl. der Zusendung von Dokumentkopien per E-Mail
- Internetbasierte Persönlichkeitsfragebögen und Leistungstests, inkl. der Möglichkeit adaptiven Testens (Chuah, Drasgow & Roberts, 2006; Konradt, Lehmann, Böhm-Rupprecht & Hertel, 2003; Potosky & Bobko, 2004; Tippins et al., 2006)
- Situative Fragen auf der Basis von Simulationen (Video, Multimedia oder Virtual Reality; Aguinis, Henle & Beaty, 2001; Drasgow, Olson-Buchanan & Moberg, 1999)
- Internetbasierte Arbeitsproben (Konradt, Hertel & Joder, 2003)

⁵ <http://www.karrierejagd.de> (2009-12-20)

- Internetbasierte Einstellungsinterviews mittels Web-Cam und Telefon bzw. Instant Messaging Systemen
- Internetbasierte Assessment Center (Eichstaedt, 2005)

4 Gütekriterien internetbasierter Methoden

Zur Beurteilung psychologisch-diagnostischer Verfahren werden in erster Linie die Kriterien der Objektivität, Reliabilität und Validität herangezogen (Amelang & Schmidt-Atzert, 2006). Zusätzlich formulieren gängige Richtlinien, wie jene des Testkuratoriums der Deutschen Psychologenvereinigung (1986) sowie die DIN-Norm 33430 für Verfahren zur berufsbezogenen Eignungsbeurteilungen (Hornke & Winterfeld, 2004; Kersting, 2008), eine Reihe weiterer Qualitätskriterien. Im Folgenden werden internetbasierte Verfahren hinsichtlich ausgewählter Gütekriterien und deren praktischen Konsequenzen diskutiert.

4.1 Objektivität

Im Allgemeinen wird internetbasierten Testverfahren eine hohe Objektivität bescheinigt; insbesondere wird ihnen wie Computertests im Allgemeinen eine hohe Durchführungsobjektivität attestiert. Der gesamte Testverlauf, von der Instruktion über die eigentliche Testapplikation bis hin zum abschließenden Debriefing, erfolgt online, vollkommen automatisiert ohne Eingriff eines Testleiters über den Computer. Dadurch erhalten alle Probanden stets dieselbe Instruktion in derselben Weise; Teilinformationen können nicht von einem Testleiter unabsichtlich „vergessen“ werden. Grundlegende Lösungsprinzipien von Items können exemplarisch demonstriert und durch automatisierte Feedbackschleifen individuell so oft wiederholt werden, bis ein Verständnis der Aufgabenstellung sicher gestellt werden kann (Wagner-Menghin, 2003). Die Übertragung der Aufgaben der Testdurchführung auf den Computer garantiert eine weitgehende Testleiterunabhängigkeit. Das Ausschalten des Testleiters hat allerdings seinen Preis. Bei unkontrollierten Internet-Testungen erfolgt die Testteilnahme in der Regel nicht in eigens dafür eingerichteten Räumlichkeiten (z. B. Testcenter, Computerlabor) sondern im privaten Umfeld des Probanden, wodurch eine Kontrolle der Umgebungsfaktoren wie z. B. Lärmbelastung, Unterbrechungen durch andere Personen, welche das Testergebnis mitunter negativ beeinflussen können, kaum erfolgen kann.

Weiters gelten computer- wie auch internetbasierte Testverfahren als auswertungsobjektiv, da die Datenerfassung (der Itemantworten des Probanden) ebenfalls vollkommen automatisiert erfolgt und übliche Fehler manueller Erhebungen aufgrund von Konzentrationsmängeln, Eingabefehlern oder mehrdeutiger Antwortmarkierungen nicht auftreten. Wurden entsprechende Programmalgorithmen entwickelt, lässt sich sogar der gesamte Auswertungsprozess völlig computerisiert gestalten, sodass dem Testanwender alle relevanten Testkennwerte des Probanden wie Scores, Normen oder Konfidenzintervallen ohne weiteres Zutun berechnet und ausgegeben werden.

Interpretationseindeutigkeit lässt sich durch den Einsatz des Computers nur unwesentlich beeinflussen, da der interpretative Akt der Zusammenführung und Bewertung der vorliegenden Einzelinformationen zur Festlegung eines Maßnahmenkatalogs (z. B. der geeigneten Therapieform) nicht vollständig auf den Computer übertragen werden kann. Unterstützung kann lediglich durch Informationssysteme gegeben werden, die zusammen mit den Testkennwerten des Probanden gleichzeitig auch exemplarische Interpretationsbeispiele für definierte Leistungsgruppen bereit stellen wie sie in der Regel auch in den Manualen der Testverfahren vorzufinden sind. Lediglich für kriteriumsorientierte Entscheidungen, bei der eine Eignung anhand des Erreichens eines vorab definierten Leistungsniveaus als gegeben angesehen wird, zeichnet sich Computerdiagnostik durch eine im Vergleich zum traditionellen Vorgehen höhere Interpretationsobjektivität aus, da das Erreichen des Kriteriums wiederum automatisiert festgestellt werden kann und menschliche Fehler ausgeschlossen werden.

4.2 Reliabilität

Die Messgenauigkeit internetbasierter Verfahren wird zumeist nur unwesentlich durch den Vorgabemodus beeinflusst und erscheint im Allgemeinen vergleichbar mit jener ihrer Papier- und-Bleistift-Pendants. Verschiedene methodenübergreifende Studien konnten sowohl für Persönlichkeits- wie auch klinisch-diagnostische Fragebogenverfahren, die ursprünglich als Papier-Bleistift-Test konzipiert und anschließend über das Internet appliziert wurden, zufriedenstellende Reliabilitätswerte nachweisen (vgl. Tabelle 1), z. B. für Angststörungen (Austin, Carlbring, Richards & Andersson, 2006), Depression (Fortson, Scott, Del Ben & Chen, 2006), psychopathologische Symptome (Jones, Fernyhough, de Wit & Meins, 2008; Vallejo, Jordán, Díaz, Comeche & Ortega, 2007), Persönlichkeit (Aluja, Rossier & Zuckerman, 2007; Meade et al., 2007), Sprachfertigkeit (Chapelle, Jamieson & Hegelheimer,

2003), Stress (Herrero & Meneses, 2006), Alkoholkonsum (Miller et al., 2002) u. v. m. Teilweise wird Internetverfahren sogar eine höhere Messgenauigkeit zugeschrieben. Braunsberger, Wybenga und Gates (2007) erhoben in Form eines Messwiederholungsdesigns Einstellungsratings einmal über Telefoninterviews und einmal in Form einer internetbasierten Befragung. Die Stabilität der Ratings über drei Monate zeigte sich bei der Internet-Erhebung als deutlich höher, was die Autoren auf fehlende Testleitereffekte und eine größere Anonymität der Erhebungssituation zurückführten. Vergleichbare Ergebnisse werden teilweise auch bei der Gegenüberstellung mit Papier-Bleistift-Tests berichtet (vgl. Meade et al. 2007; Ployhart et al., 2003). Ployhart et al. (2003) untersuchten die Auswirkung des Wechsels eines im Rahmen der Bewerberauswahl eingesetzten Papier-Bleistift-Tests zur Erfassung verschiedener Persönlichkeitsfaktoren auf eine internetbasierte Variante. Bei der Internetvorgabe zeigten sich durchwegs deutlich höhere interne Konsistenzen.

Tabelle 1

Reliabilität und Validität von Papier-Bleistift- und internetbasierten Tests

Studie	Stichprobe		Konstrukt	Items	Cronbach's Alpha		Cohens d^a
					Papier-Bleistift	Internet	
Aluja et al. (2007)	3827	Studierende	Sensation Seeking	10	.68	.68	.13
			Aggressivität	10	.64	.65	.00
			Geselligkeit	10	.71	.74	.46
Austin et al. (2006)	110	Patienten	Angststörung	14	.90	.90	.06
Cole, Bedeian & Feild (2006)	4909	Angestellte	Transformationale Führung	20	.96	.96	.16 ^d
Jones et al. (2008)	939	Studierende	Neigung zu Verfolgungswahn	10	.85	.88	.18
Jourbert & Kriek (2009)	2227	Studierende	Persönlichkeit (32 Skalen)	~13	.72 ^b	.74 ^b	.24 ^b
	2109	Manager			.76 ^b	.76 ^b	.17 ^b
Fortson et al. (2006)	185	Studierende	Depression	20	.88	.89	.13
Fouladi et al. (2002)	339	Studierende	Bindungsstärke	20	.96	.93	.15
Meade et al. (2007)	727	Studierende	Persönlichkeit (11 Skalen)	~8	.81 ^b	.83 ^b	.06 ^b
Meyerson & Tryon (2003)	786	Internetnutzer	Sexual Boredom	28	.92	.92	.62
			Satisfaction With Life	5	.87	.88	.19
Miller et al. (2002)	255	Studierende	Screening auf Alkoholmissbrauch	10	.92 ^c	.92 ^c	.06 / .09 ^d
			Durchschnittlicher Alkoholkonsum	1	.86 ^c	.93 ^c	.28 / .28 ^d
Ployhart et al. (2003)	4900	Bewerber	Gewissenhaftigkeit	20	.64	.75	.36
			Verträglichkeit	20	.68	.74	.37
			Neurotizismus	20	.72	.80	.46
Risko et al. (2006)	120	Studierende	Impression Management	20	.70	.73	.11
Vallejo et al. (2007)	185	Studierende	Psychopathologische Symptome	28	.90	.90	.12

Anmerkungen. ^a Papier-Bleistift vs. internetbasierte Vorgabe; ^b Mittel über alle Skalen; ^c Retest-Reliabilität; ^d Latenter Mittelwertsunterschied; ^d

Zeitpunkt 1 und 2

Bislang noch nicht eindeutig geklärt sind die Auswirkungen von Kontextfaktoren, die sich möglicherweise aus der unterschiedlichen Computerausstattung der Probanden ergeben. Während bei Papier-Bleistift Fragebögen aber auch Computertests, die auf entsprechend optimierten Geräten in Computerlabors bearbeitet werden, sicher gestellt werden kann, dass der Test hinsichtlich Schriftgröße, Farbe, Anzahl der Items pro Seite usw. für alle Personen gleich dargestellt wird, kann derartiges bei unkontrollierten Online-Testungen nicht zweifelsfrei garantiert werden, da sich die Computerausstattung zwischen den Teilnehmern in der Regel deutlich unterscheidet – Bildschirmgröße und –auflösung, Geschwindigkeit der Internetverbindung, Computerleistung usw. Derartige Einflussfaktoren können unter Umständen eine zusätzliche Fehlervarianz in den Testergebnissen generieren, die nicht die eigentlich zu erfassende Leistung des Probanden widerspiegelt. Bridgeman, Lennon und Jackenthal (2003) untersuchten in einer experimentellen Versuchsanordnung die Auswirkung divergierender Bildschirmauflösungen auf die Ergebnisse mathematischer und verbaler Leistungstests. Während bei den Mathematikscores keine Verzerrungen offensichtlich wurden, zeigte der Test zum Sprachverständnis, bei dem es Fragen zu verschiedenen Textpassagen zu beantworten galt, einen deutlichen Leistungsunterschied in Abhängigkeit von der Bildschirmauflösung – bei größerem Bildschirm und höherer Auflösung, wenn also mehr Informationen ohne Scrollen auf dem Bildschirm angezeigt werden konnten, erzielten die Probanden signifikant bessere Ergebnisse. Wenn die erzielte Effektgröße von $d < .3$ auch als vergleichsweise gering angesehen werden muss, zeigen die Ergebnisse jedoch deutlich, dass die technische Ausstattung der Probanden einen direkten Einfluss auf die Testleistung nehmen kann. Dieser Problematik tragen auch gängige Richtlinien zur Durchführung computer- und internetbasierter Tests (ICT, 2006) Rechnung und betonen ausdrücklich die Berücksichtigung ergonomischer Überlegungen bei der Präsentation des Testmaterials.

Darunter fallen unter anderem:

- Verwendung geeigneter Auflösung und Farben
- Verwendung eines konsistenten Layouts und Farbschemas
- Einsatz klarer Schriftarten und –größen
- Formulierung eindeutiger Fehlermeldungen
- Integration geeigneter Navigationsmöglichkeiten

Eine Berücksichtigung möglicher Quellen technologiebedingter Verzerrungen erscheint bei Internettestungen somit dringend angeraten. Leider liegen bislang nur wenige Informationen vor, welche Faktoren in welchen Situationen tatsächlich nicht zu vernachlässigende

Verzerrungen generieren und vor allem in welcher Weise diesen am besten zu begegnen sind. Die „International Guidelines on Computer-Based and Internet-Delivered Testing“ (ICT, 2006) und vergleichbare Richtlinien verbleiben hierfür sehr vage und beschränken sich auf unspezifische Hinweise, auf welche Punkte bedacht zu nehmen ist. Ein Grund hierfür kann in einem eklatanten Mangel relevanter Untersuchungsbefunden zu finden sein. Bislang liegen nur wenige Studien vor, die mit ausreichender Power auch kleine Kontexteffekte offen zu legen vermögen. In einer Übersicht über die gegenwärtige Befundlage orten Waters und Pommerich (2007) folgende technologiebedingten Einflussgrößen, die Testleistungen zu beeinträchtigen vermögen (Tabelle 2):

Tabelle 2

*Mögliche Kontexteffekte bei internetbasierten Erhebungen
(adaptiert aus Waters & Pommerich, 2007, S. 16)*

Bildschirmauflösung
Beeinflusst sowohl Größe wie auch Umfang der dargestellten Informationen.
Bildschirmgröße
Beeinflusst die Größe der dargestellten Informationen aber nicht den Umfang der Informationen.
Bildschirmtyp (Röhrenbildschirm vs. LCD)
Beeinflusst das Ausmaß der Bildschirm Spiegelung.
Prozessorleistung
Beeinflusst die Genauigkeit von Zeitmessungen (Eichstaedt, 2001)
Schriftgröße
Beeinflusst den Textumfang, der auf dem Bildschirm angezeigt wird.
Scrollen
Beeinflusst das Ausmaß an Informationen, die auf dem Bildschirm dargestellt werden.
Maustyp
Beeinflusst die Genauigkeit von Reaktionszeitmessungen (Plant, Hammond & Whitehouse, 2003).
Verbindungsgeschwindigkeit
Erzeugt eine Verzögerung beim Laden neuer Informationen und kann zu Teilnahmeabbrüchen oder erhöhter Testangst bei Speed-Tests führen.

Welche Möglichkeiten bieten sich nun möglichen Kontexteffekten zu begegnen? Tippins et al. (2006) schlagen vor, die Bürde der Teststandardisierung zum Teil von den Testentwicklern auf die Probanden zu verlagern. Wenn bekannt ist, dass bestimmte Eigenschaften der Soft- oder Hardware die Testergebnisse negativ beeinflussen können, sind die Probanden über geeignete (oder notwendige) Konfigurationsmöglichkeiten zu instruieren, um ihre Testergebnisse zu maximieren. In der Praxis ist jedoch fraglich, ob (1) ungünstige Soft-/Hardware-Konfigurationen überhaupt vorab bekannt sind und (2) inwiefern die einzelnen Probanden in der Lage sind diese selbständig zu ändern; einerseits kann nicht bei jedem Teilnehmer eine vergleichbar hohe Computerkompetenz vorausgesetzt werden und andererseits besteht unter Umständen gar nicht die Möglichkeit die Computerkonfiguration zu ändern (z. B. in Computerlabors in Schulen und Universitäten). Alternativ schlägt Bennett (2003) eine technische Lösung vor, indem die Testanwendung automatisch das Computersystem des Probanden für die optimale Testdurchführung (z. B. hinsichtlich Schriftgröße oder Bildschirmauflösung) kalibriert. Hierfür wäre jedoch zumeist die Installation zusätzlicher Software auf dem Computer des Probanden von Nöten, weshalb dieses Vorgehen in der Praxis in vielen Fällen als nicht optimal betrachtet werden kann. Schließlich verbleibt noch die Möglichkeit den Test für den kleinsten gemeinsamen Nenner zu optimieren (z. B. für die geringste Bildschirmauflösung). Diese Ansätze sind in der Praxis jedoch technisch sehr anspruchsvoll und ermöglichen dennoch keine vollständige Kontrolle von Kontexteffekten.

4.3 Validität

Die Validität internetbasierter Verfahren wurde (vor allem bei Persönlichkeitsfragebögen) wiederholt über verfahrensvergleichende Studien untersucht, um deren Äquivalenz mit ihren Papier-Bleistift-Vorlagen nachzuweisen. Nach Van de Vijver und Harsveld (1994) kann Methodenäquivalenz in quantitative und qualitative Äquivalenz unterschieden werden. Qualitative (strukturelle) Äquivalenz bezieht sich auf die Konstruktvalidität und untersucht (z. B. mittels Strukturgleichungsmodellen) inwiefern die beiden Präsentationsformate dasselbe psychologische Konstrukt erfassen. Quantitative Äquivalenz zielt hingegen auf den Nachweis vergleichbarer Scores (Mittel, Streuung, Verteilung) und die Anwendbarkeit von Testnormen einer Testversion für eine andere Version ab. Wie zahlreiche methodenübergreifende Studien (vgl. Chuah et al., 2006; Cole et al., 2006; Meade et al., 2007; Meyerson & Tryon, 2003; Salgado & Moscoso, 2003) nahe legen kann für Selbstbeschreibungsverfahren weitgehend qualitative Äquivalenz von Internet- und Papier-

Bleistift-Testungen angenommen werden. In der Regel finden sich keine bzw. nur marginale Unterschiede zwischen beiden Vorgabemodi. Salgado und Moscoso (2003) stellten Internet- und Papier-Bleistift-Varianten eines Fünf-Faktor-Persönlichkeitsfragebogens in Form eines Messwiederholungsdesigns gegenüber. Sie fanden vergleichbare Faktorenstrukturen, Reliabilität und Scores zwischen den Formaten, jedoch (ähnlich wie bei Ployhart et al., 2003) eine etwas größere Varianz in den Ergebnissen der Interneterhebung. Chuah et al. (2006) berichten ähnliche Resultate auf Basis von IRT-Analysen. Sie analysierten die Big Five der Persönlichkeit, die einmal unter Beaufsichtigung als Papier-Bleistift Version, einmal im Computerlabor und einmal unbeaufsichtigt über das Internet vorgegeben wurden. Umfangreiche Strukturanalysen sowie die prädiktive Validität mit externen Kriterien deuteten auf keine substanziellen Unterschiede zwischen den Erhebungsmodalitäten hin. So waren die Faktorenstruktur der verschiedenen Testvorgaben vergleichbar; ebenso fand sich kein Differential Item Functioning (DIF) bezüglich der Präsentationsmedien, die auf beeinträchtigte Testresultate aufgrund der Internetvorgabe hinweisen würden. Selbst quantitative Äquivalenz scheint naheliegend, da die IRT-analytisch erzeugten Fähigkeitsparameter im Durchschnitt zwischen den drei Gruppen vergleichbar waren. Meade et al. (2007) wiederum fanden bei der online Vorgabe von 11 Persönlichkeitsskalen des „Occupational Personality Questionnaires“ zum Teil deutlich bessere Modellfits konfirmatorischer Faktorenanalysen als bei Papier-Bleistift-Vorgabe. Während qualitative Äquivalenz für alle Skalen bestätigt werden konnte, gelang dies für quantitative Äquivalenz nur bei einem Teil. Die Mittelwerte der latenten Konstrukte der verschiedenen Gruppen variierten zum Teil deutlich. Es scheint somit, dass in den meisten Fällen qualitative Äquivalenz von Internet- und Papier-Bleistift-Verfahren angenommen werden kann, wenn diese an vergleichbaren Stichproben appliziert werden; Reliabilität und Faktorenstruktur werden durch den Wechsel des Vorgabemediums nur in den wenigstens Fällen beeinflusst. Auf der anderen Seite berichten Buchanan et al. (2005) im Rahmen ihres Methodenexperiments zur Äquivalenzbeurteilung des „Prospective memory questionnaires“ von einem der wenigen Instrumente, bei dem die internetbasierte Vorgabe zu deutlich divergierenden Faktorenstrukturen führte. Dies verdeutlicht, dass Äquivalenz nicht automatisch angenommen werden darf: “When an existing test is converted from off-line to Web-based format, it is important to establish that it is actually measuring what it is meant to, rather than assuming that because the paper-and-pencil version ‘worked’, so will the on-line version” (Buchanan, 2005, S. 148). Was quantitative Vergleichbarkeit betrifft, ergibt sich jedoch ein deutlich anderes Bild (vgl. Tabelle 1). Die Scoreverteilungen von Fragebögen unterscheiden sich manchmal deutlich bei Internetapplikation (vgl. Aluja et

al., 2007; Ployhart et al., 2003), was vor allem der Anwendung bestehender Normen aus Papier-Bleistift-Erhebung entgegen steht. Es gilt daher für jedes Instrument aufs Neue zu überprüfen, inwiefern psychometrische Äquivalenz angenommen werden kann; insbesondere durch stringent konzipierte Experimentalstudien, die nicht auf Zufallsstichproben unkontrollierter Internettestungen bauen, sondern eine experimentelle Zuweisung zu den Untersuchungsbedingungen vornehmen, um Zufallsvariationen aufgrund von Soziodemographie oder Motivation auszuschließen.

Für Leistungstests können weit weniger eindeutige Schlussfolgerungen gezogen werden, da deren Anwendung im Internet bislang noch weniger verbreitet ist, sodass die Befunde von überwachten Computertestungen herangezogen werden müssen. Im Allgemeinen ist mit dem Wechsel der Modalität bei der Vorgabe von Powertests mit keinem bzw. einem zu vernachlässigenden Methodeneffekt zu rechnen. Williams und McCord (2006) berichten von vergleichbaren Mittelwerten und Varianzen bei Ravens Matrizen tests unter on- und offline Applikation. Dies unterstreichen auch Metaanalysen von Bergstrom (1992) und Kim (1999). Beide berichten bei Computer- und Papier-Bleistift-Vorgabe von weitgehender Äquivalenz; teilweise wird aber von etwas höheren Scores am Computer berichtet. Allerdings ist das Ausmaß der Übereinstimmung von der Art des Tests abhängig. Während bei Mathematik- und Lesetests sowie kognitiven Leistungstests von weitgehender Äquivalenz ausgegangen werden kann, ist dies bei Englischtests und anderen fachspezifischen Tests (z. B. medizinisches oder technisches Wissen) häufig nur bedingt gegeben (Kim, 1999; Wang, Jiao, Young, Brooks & Olson, 2007, 2008). Insbesondere Zeitbegrenzungen führen bei Computertests zu divergierenden Resultaten. Mead und Drasgow (1993) berichten in ihrer Metaanalyse bei kognitiven Leistungstests ohne bzw. sehr großzügiger Zeitbegrenzung von einer durchschnittlichen Korrelation zwischen Computer- und Papier-Bleistift-Vorgabe in Höhe von .97. Speedtests, bei denen es ein bestimmtes Testmaterial möglichst schnell zu bearbeiten gilt, korrelieren jedoch nur zu .72.

Wie experimentelle Vergleichsstudien nahe legen, können diese Befunde auch auf internetbasiertes Erhebungen generalisiert werden (vgl. Ihmke et al., 2009; Potosky & Bobko, 2004). Potosky und Bobko (2004) applizierten einen kognitiven Leistungstest mit und ohne Zeitbegrenzung einmal als Internet- und einmal als Papier-Bleistift-Variante. Für den Speedtest zeigte sich jedoch wie erwartet mit einer modalitätsübergreifenden Korrelation von .60 eine nur ungenügende Übereinstimmung. Ein Grund für die Probleme von computer-

und internetbasierten Speedtests liegt häufig in einer motorischen Komponente. Sind bei einem Verfahren unter Zeitdruck andere Aktionen als einfache Tastendrucke oder Mausklicks vorzunehmen, übt die Vorerfahrung mit dem Computer einen starken Einfluss auf die Testleistung aus. Dies verdeutlicht auch das Scheitern einen Zahlen-Verbindungs-Test zur Erfassung der kognitiven Informationsverarbeitungsgeschwindigkeit auf den Computer zu übertragen (Drapeau, Bastien-Toniazzo, Rous, & Carlier, 2007). Da bei diesem Test zufällig angeordnete Zahlen möglichst schnell nacheinander zu verbinden sind, ist die Geschicklichkeit im Umgang mit der Maus von ausschlaggebender Bedeutung für das Testergebnis. Während entsprechende Markierungen auf einem Blatt Papier in der Regel von den meisten Personen unproblematisch gesetzt werden können, erfordert dieselbe Aufgabe am Computer deutlich kompliziertere Aktionen. Dementsprechend fanden Drapeau et al. (2007) eine nur ungenügende Übereinstimmung zwischen der Computer- und Papier-Bleistift-Version. Für derartige Verfahren erscheint der Einsatz spezieller Eingabemedien von Nöten, die der Handhabung von Papier und Bleistift möglichst ähnlich sind. Mit Tablet-PCs, bei denen mit Stift-ähnlichen Geräten direkt am Bildschirm geschrieben werden kann, stehen derartige Geräte zwar bereits zur Verfügung. Diese sind jedoch in Privathaushalten nur wenig verbreitet, weshalb bei Internettstungen deren Vorhandensein kaum vorausgesetzt werden kann. Speedtests, die über die Wahlalternative zweier Tastendrucke hinausgehen, scheinen daher derzeit weniger als Internetanwendung geeignet.

4.4 Normierung

Die Frage der Normierung betrifft das Vorliegen sowie die Qualität eines Bezugssystems anhand dessen individuelle Testergebnisse eingeordnet werden können (Testkuratorium, 1986). Die Anwendung bestehender Normen, die für Papier-Bleistift-Verfahren entwickelt wurden, bei Internettests setzt äquivalente Testverfahren voraus – sowohl in qualitativer wie auch quantitativer Hinsicht. Da insbesondere quantitative Äquivalenz zwischen Papier-Bleistift- und Internetverfahren häufig nicht nachgewiesen werden kann (s. o.), ist die Anwendung entsprechender Normen nicht möglich. Die Entwicklung eigenständiger Internetnormen steht allerdings für die meisten Verfahren noch aus – nicht zuletzt aufgrund des finanziellen Aufwands, der damit verbunden wäre. Zudem stellt sich für die psychologische Praxis häufig auch die Frage, inwiefern überhaupt normorientierte Entscheidungen von Interesse sind und nicht vielmehr ein kriteriumsorientiertes Vorgehen gefragt ist; im Rahmen der Selektionsdiagnostik (zu Recrutementzwecken beispielsweise)

sind Normen in der Regel weniger relevant, da lediglich die besten Bewerber oder jene, die eine bestimmte Mindestleistung erbracht haben, berücksichtigt werden.

Internetbasierte Diagnostik wird häufig implizit mit Fragen der Repräsentativität verbunden, also inwiefern über das Internet Repräsentativerhebungen möglich wären. Dieser Automatismus ist umso überraschender als derartige Überlegungen in der psychologischen (offline) Forschung traditionell eine nur untergeordnete Rolle einzunehmen scheint oder wie Birnbaum (2000) es treffend formuliert: „Some say that psychological science is based on research with rats, the mentally disturbed, and college students. We study rats because they can be controlled, the disturbed because they need help, and college students because they are available“ (S. 7). Dies mag für viele Fragestellungen und Untersuchungssituationen nicht notwendigerweise ein Negativum darstellen; so setzen experimentelle oder korrelative Studien a priori keine repräsentativen Stichproben voraus (Moser, 1989). Viele psychologische Untersuchungen, beispielsweise zu grundlegenden kognitiven Prozessen wie Aufmerksamkeitsspanne oder Gedächtnis, liefern in der Regel auch auf Basis von Studierenden-Stichproben gültige Ergebnisse (wenn nicht die Generalisierung auf bestimmte Spezialpopulationen wie z. B. sehr junge oder sehr alte Personen von Interesse ist). Erst populationsbeschreibende Aussagen über Einstellungen und Verhalten (wie z. B. 30% der Deutschen beabsichtigen Partei XY zu wählen) sind auf Repräsentativerhebungen angewiesen (Batinic, 2004). Unkontrollierte Interneterhebung auf Basis anfallender Stichproben können kaum als repräsentativ für die Gesamtbevölkerung oder irgendeine andere imaginäre Population erachtet werden; hinsichtlich vieler Merkmale wie z. B. Geschlecht oder ethnische Zugehörigkeit sind sie jedoch weitaus heterogener als traditionelle offline Stichproben und bilden die Bevölkerung - wenn sie auch nicht als repräsentativ für diese betrachtet werden können - weitaus besser ab (Gosling, Vazire, Srivastava & John, 2004). Echte bevölkerungsrepräsentative Erhebungen setzen eine explizite Stichprobenziehung anhand relevanter Merkmale voraus (z. B. in entsprechenden Online-Panels). Im deutschsprachigen Raum sind diese allerdings über das Internet bislang noch kaum zu realisieren, da gewisse Personengruppen (v. a. ältere Frauen) nicht in ausreichendem Umfang im Internet vertreten sind, sodass Repräsentativerhebungen bislang nur für die Altersgruppe der 16 bis 60jährigen (Eimeren & Frees, 2010) realisierbar erscheinen. Mit zunehmender Verbreitung des Internets auch in diesen Bevölkerungsgruppen wird das Internet allerdings auch für Repräsentativerhebungen künftig eine bedeutendere Rolle einnehmen. In Schweden, wo die Internetpenetration deutlich weiter fortgeschritten ist, konnten bereits umfangreiche

epidemiologische Repräsentativerhebungen online durchgeführt werden (Ekman, Dickman, Klint, Weiderpass, & Litton, 2005). Es zeigten sich zwar leichte Unterschiede in der soziodemographischen Zusammensetzung der Internet- und einer parallelen Papier-Bleistift-Stichprobe, derart dass eher höher gebildet an der Internetbefragung teilnahmen. Allerdings waren die entsprechenden Effekte derart gering, dass die Autoren schlossen (S. 108): „as a method, the use of web questionnaires [...] did not introduce bias to any larger extent than paper questionnaires“.

4.5 Fairness

Das Gütekriterium der Fairness kennzeichnet, inwiefern ein Verfahren bestimmte Personengruppen systematisch bevorzugt oder benachteiligt (Amelang & Schmidt-Atzert, 2006). Gelegentlich wird als Argument gegen internetbasierte Testungen angeführt, dass Personen mit einer geringen Erfahrung im Umgang mit Computern und einer erhöhten Computerangst benachteiligt werden (Naglieri et al., 2004). Diese Sichtweise erscheint nach heutigem Kenntnisstand jedoch aus zwei Gründen nur noch bedingt zutreffend:

1. Aktuelle Befunde (Eid, 2004; Shermis & Lombard, 1998; Wang et al., 2007; Williams & McCord, 2006) legen zumeist keinen Zusammenhang mehr von Computerängstlichkeit und –erfahrung mit Leistungen in Computertests nahe.
2. Computer sowie Internet sind im privaten und beruflichen Alltag bereits derart stark verankert, sodass eine zumindest rudimentäre Computererfahrung für die meisten Bevölkerungsgruppen weitgehend als gegeben angesehen werden kann.

Da unkontrollierte Internetttests in der Regel in der vertrauten Umgebung im privaten Umfeld bearbeitet werden, kann unter Umständen das Auftreten genereller Testangst, die durch die ungewohnte, häufig stressinduzierende Testsituation in Testcentern erlebt wird, bei manchen Personen sogar verringert werden. Im Allgemeinen scheint eine geringe Computererfahrung keine systematische Benachteiligung zu bedingen. Im Zweifelsfall ist dieser Umstand jedoch für spezifische Subpopulationen und Situationen gegebenenfalls einer Überprüfung zu unterziehen - Sandene et al. (2005) wie auch Bennett et al. (2008) berichten beispielsweise bei Schülern mit größerer Erfahrung im Umgang mit dem Computer von besseren Schreib- und Mathematikleistungen am Computer.

Insbesondere bei Testungen mit individualdiagnostischer Zielsetzung (z. B. in der Selektionsdiagnostik) ist zu klären, inwiefern unterschiedliche Computerausstattungen der

Probanden zu einer möglichen Ungleichbehandlung beitragen können. Sollten gewisse Computerkonfigurationen (z. B. große Bildschirme oder leistungsstarke Rechner) in bestimmten Situationen bessere Testergebnisse begünstigen (s. o.), würden einzelne Personen lediglich aufgrund ihrer Computerausstattung bevorzugt und ihre tatsächliche Leistung überschätzt. Dies spiegelt sich in Befürchtungen wieder, internetbasierte Erhebungen würden bestimmte soziale Gruppen benachteiligen (Chapman & Webster, 2003), da Personengruppen mit niedrigem sozioökonomischen Status schwieriger Internetzugang haben und zudem leistungsstärkerer (= teurer) Computer häufig in finanzkräftigeren Bevölkerungsschichten vorzufinden sind. Dies mag bei wissenschaftlichen Studien, die mit Ergebnissen auf Gruppenebene arbeiten, möglicherweise eine untergeordnete Rolle spielen, wenn keine systematischen Verzerrungen der Computerausstattung zwischen den untersuchten Gruppen zu erwarten sind, ist aber bei individualdiagnostischen High-Stake-Tests nur schwer zu vertreten. Angesichts aktueller demographischer Befunde, die „Offliner“, also Personen ohne Internetnutzung, typischerweise als weiblich, älter und von geringer Bildung beschreiben (Eimeren & Fries, 2010), schränkt dies die Einsatzmöglichkeit internetbasierter Verfahren für manche Situationen (z. B. zur Bewerberauswahl für bestimmte Niedriglohn-Jobs) erheblich ein. Die Relevanz dieses Problems sollte allerdings mit zunehmender Verbreitung des Internets abnehmen. Darüber hinaus kann internetbasierte Eignungsdiagnostik umgekehrt auch zu einer stärkeren Berücksichtigung von Personen mit eingeschränkter Mobilität führen.

4.6 Verfälschbarkeit

Mit der Einführung von Computertestungen wurde anfänglich die Hoffnung verbunden sozial erwünschtes Antwortverhalten in Selbstbeschreibungsverfahren zu verringern (Fox & Schwartz, 2002), da (1) das Ausschalten des Testleiters einen möglichen sozialen Druck verringern und ein Gefühl von Anonymität erzeugen sowie (2) durch die zumeist fehlende Navigationsmöglichkeiten von Computerfragebögen, bei denen Items nur einzeln oder in kleinen Gruppen nacheinander bearbeitet werden können, die Bildung „richtiger“ Antwortstrategien erschwert werden sollten. Metaanalysen (Dwight & Feigelson, 2000; Richman, Kiesler, Weisband & Drasgow, 1999) bestätigen einen direkten Zusammenhang zwischen Vorgabemodus (Computer versus Papier-Bleistift) und sozialer Erwünschtheit allerdings nicht. Obwohl in älteren Studien noch ein geringeres Ausmaß an Impression Management bei Computertestungen zu beobachten war, verschwindet dieser Zusammenhang mit zunehmender Verbreitung von Computern und entsprechender Vorerfahrung der Versuchspersonen in jüngeren Studien beinahe gänzlich. Nur ein Wechsel

auf ein Computerverfahren führt nicht zwangsläufig zu weniger sozial erwünschten Verzerrungen. Vielmehr moderieren Kontextfaktoren diesen Zusammenhang. Joinson und Paine (2006) identifizierten drei Hauptmerkmale, die zu höherem Self Disclosure, einer größeren Bereitschaft zur Selbstoffenbarung und Bereitwilligkeit dem Gegenüber neue, auch unvorteilhafte Informationen über die eigene Person mitzuteilen, beitragen:

- Je stärker Probanden ihre *Anonymität* gewahrt und keine Zuordnung der Testleistung zur eigenen Person möglich sehen, umso eher ist mit der Preisgabe von sensiblen Informationen zu rechnen. Joinson, Woodley und Reips (2007) berichten beispielsweise von einer deutlich geringeren Bereitschaft das eigene Einkommen anzugeben, wenn sich Versuchspersonen zu Beginn eines internetbasierten Fragebogens mit Name und Passwort zu identifizieren haben gegenüber einer Gruppe, die über anonymisierte Einladungslinks auf den Fragebogen Zugriff hatte.
- Je stärker die *soziale Präsenz* des Testleiters verringert wird, umso größer ist die Offenheit der Probanden, da soziale Hinweisreize wie z. B. das Geschlecht des Interviewers nicht unwillkürlich die Antworten beeinflussen können und der Computer gleichsam als neutralerer Kommunikationspartner agiert.
- Je stärker die Probanden ihre *Privatsphäre* gewahrt sehen und keine Bedenken bezüglich der Sicherheit ihrer Daten hegen, umso eher werden persönliche Informationen preis gegeben. Somit hängt Social Disclosure stark vom Vertrauen der Probanden in den Gesprächspartner ab. Wird befürchtet, dass die eigenen Daten (z. B. aufgrund technischer Mängel oder Datenschutzbedenken) in die falschen Hände gelangen, ist verstärkt mit sozial erwünschten Antworten zu rechnen.

Diesbezügliche Experimentalstudien konnten diese Annahmen weitgehend bestätigen. So geben Teilnehmer in internetbasierten Diskussionsrunden, insbesondere wenn sie auch visuell anonym sind (also keine Videoübertragung erfolgt), spontan signifikant mehr Informationen über die eigene Person preis als in vergleichbaren Präsenzgruppen (Joinson, 2001). Ähnliche Resultate lassen sich bei Bearbeitung standardisierter Selbstbeschreibungsverfahren beobachten. Joinson (1999) ließ Versuchspersonen Skalen zur Erfassung sozial erwünschten Antwortverhaltens im Gruppensetting am Computer im Labor bzw. als Papier-Bleistift Version bearbeiten. Zusätzlich wurde die Teilnahmeanonymität variiert, indem ein Teil der Versuchspersonen ihren Namen auf der ersten Seite des Fragebogens anzugeben hatten. Wie erwartet waren in den anonymen Bedingungen, im Speziellen in der anonymen

Computerbedingung, die geringsten Antwortverzerrungen zu beobachten. Insbesondere bei Individualtestungen, bei denen die Testbearbeitung alleine, ohne gleichzeitige Anwesenheit anderer Personen (z. B. anderer Versuchspersonen oder des Testleiters) erfolgt und eine größere Privatsphäre während der Testdurchführung vorliegt, zeigen sich bei Computertestungen geringere Antwortverzerrungen als bei Papier-Bleistift-Tests (Richman et al. 1999).

In diesem Sinne postuliert die Candor-Hypothese (Buchanan, 2000), dass unkontrollierte Internet-Befragungen im Vergleich zu Labortests (Computer- ebenso wie Papier-Bleistift-Tests) zu weniger sozial verzerrten Antworten führen sollten, da die Testsituation als anonym und weniger wertend erlebt wird. Empirische Überprüfungen dieser Annahme brachten bislang allerdings zwiespältige Resultate. Im Vergleich zu alternativen Erhebungsmethoden berichten Personen bei Interneterhebungen zwar von einem höherem Alkoholkonsum (Booth-Kewley, Larson & Mioshi, 2007), höherem Konsum illegaler Drogen (Wang et al., 2005) und auch einer höheren Anzahl (gleichgeschlechtlicher) Sexualpartner (Ross, Manson, Cooper & Tikkanen, 2005). Aber Selbstbeschreibungsskalen zur Quantifizierung sozialer Erwünschtheit in Persönlichkeitsfragebögen konnten keine vergleichbare Bestätigung erbringen (Fox & Schwartz, 2002; Pettit, 2002; Risko et al., 2006). Ein Grund mag darin zu finden sein, dass das Ausmaß erlebter Anonymität in der Regel nicht explizit miterhoben und stets implizit davon ausgegangen wurde, internetbasierte Befragungen wären anonym, da die Datenerfassung in der privaten Umgebung ohne direkte Anwesenheit des Testleiters erfolgte. Dies ist eine Annahme, die angesichts zunehmender Datenschutzbedenken, ob und inwiefern das eigene Handeln im Internet tatsächlich sicher und anonym sei, stark in Frage zu stellen ist (vgl. Joinson & Paine, 2006).

Generell muss jedoch festgehalten werden, dass sich die bisherige Befundlage (sowohl für Labor- wie auch internetbasierte Computertestungen) bislang als nur eingeschränkt interpretierbar darstellt, da diese beinahe ausschließlich an Freiwilligenstichproben erhoben wurden, für welche die Ergebnisse keine individuellen Konsequenzen zeitigten. Gegenwärtig erscheint es wenig plausibel anzunehmen Probanden würden bei Internettestungen, deren Ergebnisse weitreichende persönliche Folgen haben (z. B. in der Eignungsdiagnostik), weniger versuchen bzw. weniger Schwierigkeiten haben die Testergebnisse in ihrem Sinn zu beeinflussen. Vielversprechender als der Einsatz eines alternativen Präsentationsmediums erscheint ein Paradigmenwechsel weg von Fragebogenverfahren hin zu objektiven Tests,

welche Persönlichkeitsmerkmale anstelle von Selbstauskünften über Leistungskomponenten zu erfassen suchen (vgl. Ziegler, Schmidt-Atzert, Bühner & Krumm, 2007).

Ein wachsendes Problem, das mit der zunehmenden Anwendung internetbasierter Testungen einhergeht, stellt die unkontrollierte Verbreitung von Testinhalten im Internet dar, wenn Probanden Testfragen kopieren, weiterleiten und dadurch Itemlösungen für künftige Teilnehmer a priori publik machen (Chapman & Webster, 2003; Drasgow, Nye, Guo & Tay, 2009). Dadurch wird die gezielte Verfälschung von Testergebnissen, um z. B. bessere Leistungen in internetbasierten Eignungstests im Rahmen der Bewerberauswahl zu simulieren, für Personen mit entsprechender Medienkompetenz⁶ stark vereinfacht. Deshalb werden von Testunternehmen große Anstrengungen unternommen, um die Verbreitung ihrer Items zu verhindern. Einerseits werden internetbasierte Verfahren gegenüber Angriffen aus dem Internet („Hacker“) durch verschiedenen Möglichkeiten der Zugangssicherung und Datenverschlüsselung technisch besonders geschützt, z. B. durch die Implementierung von Internetzugang, Datenbank und Testmaterial auf getrennten Servern (Naglieri et al., 2004). Andererseits durchforsten automatisierte Suchprogramme (sogenannte „web-patrols“) gezielt das Internet nach Diskussionsforen und Webseiten, auf denen derartige Informationen angeboten werden, um den aktiven Tauschhandel unter Probanden zu unterbinden (Burke, 2009). Aus diesem Grund werden internetbasierte Tests häufig auch mit umfangreichen Aufgabenpools bestritten, aus denen ein Proband nicht alle sondern nur ausgewählte Items erhält; beispielsweise im Rahmen adaptiver Testungen, welche unter Umständen sogar eine automatische Itemgenerierung integriert, bei denen vor Testbeginn noch gar keine Items vorliegen, sondern diese erst während der Testung dynamisch für den jeweiligen Probanden erzeugt werden (vgl. Embretson & Yang, 2007). Dadurch verringert sich die Gefahr der Verbreitung ganzer Tests, da jeder Teilnehmer unterschiedliche Itemkombinationen erhält (Drasgow et al., 2009; Tippins et al., 2006). Zudem wird empfohlen Trainingsversionen von internetbasierten Verfahren frei zur Verfügung zu stellen, damit sich interessierte Probanden vorab mit den Materialien vertraut machen können. Dadurch wird unsystematische Varianz reduziert, wenn einzelne Personen Verfahren unkontrolliert mehrmals absolvieren (z. B. unter anderem Namen), und die Gefahr geringer, dass Probanden Items des echten Tests zu Informationszwecken weiterleiten.

⁶ Personen, die in der Lage sind die entsprechenden Informationen im Internet zu finden.

Zur nachträglichen Identifikation von Probanden, denen möglicherweise die Items bereits vorab bekannt waren, wurden verschiedene Strategien entwickelt. Im Rahmen der Bewerberauswahl werden internetbasierte Verfahren in der Regel zu Beginn eines mehrstufigen Diagnoseverfahrens empfohlen, das dann mit konventionellen Verfahren fortgesetzt wird (vgl. Hertel, Konradt & Orlikowski, 2003; Nye, Do, Dragow & Fine, 2008). Ein prototypisches Modell einer solchen Verknüpfung internetbasierter und konventioneller Verfahren ist das *OKP Modell* (Hertel et al., 2003), das die drei Phasen der Orientierung, Kategorisierung und Präzisierung unterscheidet. Innerhalb dieser Phasen wird der Bewerberpool sukzessive anhand vorab festgelegter Kriterien reduziert. In der *Orientierungsphase* wählen sich die Bewerber im Inter- oder Intranet auf den Auswahlseiten ein und beantworten erste demographische Fragen, anhand derer bereits eine erste automatisierbare Vorselektion stattfindet. In der *Kategorisierungsphase* werden die Bewerber dann anhand von Minimalanforderungen der zu besetzenden Position eingestuft, und bei Unterschreitung der Minimalanforderungen abgelehnt. Auch dieser Schritt kann automatisiert realisiert werden. Die verbleibenden Bewerber werden dann per E-Mail aufgefordert, weitere Unterlagen (Zeugnisse, etc.) zuzusenden. In der Phase der *Präzisierung* durchlaufen die Bewerber schließlich spezifische Auswahlverfahren, die vorwiegend in konventioneller Form durchgeführt werden. Unter Umständen kann bereits die Ankündigung eines solchen Vorgehens zu größerer Ehrlichkeit der Probanden beitragen (Naglieri et al., 2004).

Zusätzlich werden unter dem Schlagwort der *data forensic* verschiedene statistische Prozeduren vorgeschlagen, um ungewöhnlicher Antwortmuster zu identifizieren, die auf Itemvorwissen bei einzelnen Probanden hinweisen (vgl. McLeod, Lewis & Thissen, 2003). Die meisten dieser Ansätze basieren auf Item Response Modellen, da diese anhand eines geschätzten Fähigkeitsparameters eine Quantifizierung einzelner Antwortwahrscheinlichkeit ermöglichen und dadurch ungewöhnliche, d. h. unwahrscheinliche Antworten für eine Person, die nicht dem Testmodell entsprechen, objektiv identifizierbar machen. Über Personen-Fit Indizes lassen sich signifikante Häufungen derartig ungewöhnlicher Antworten erfassen. Wenn beispielsweise eine Person viele schwierige Items und vergleichsweise wenig leichte Items löst, kann dies als Indiz für Itemvorwissen herangezogen werden.

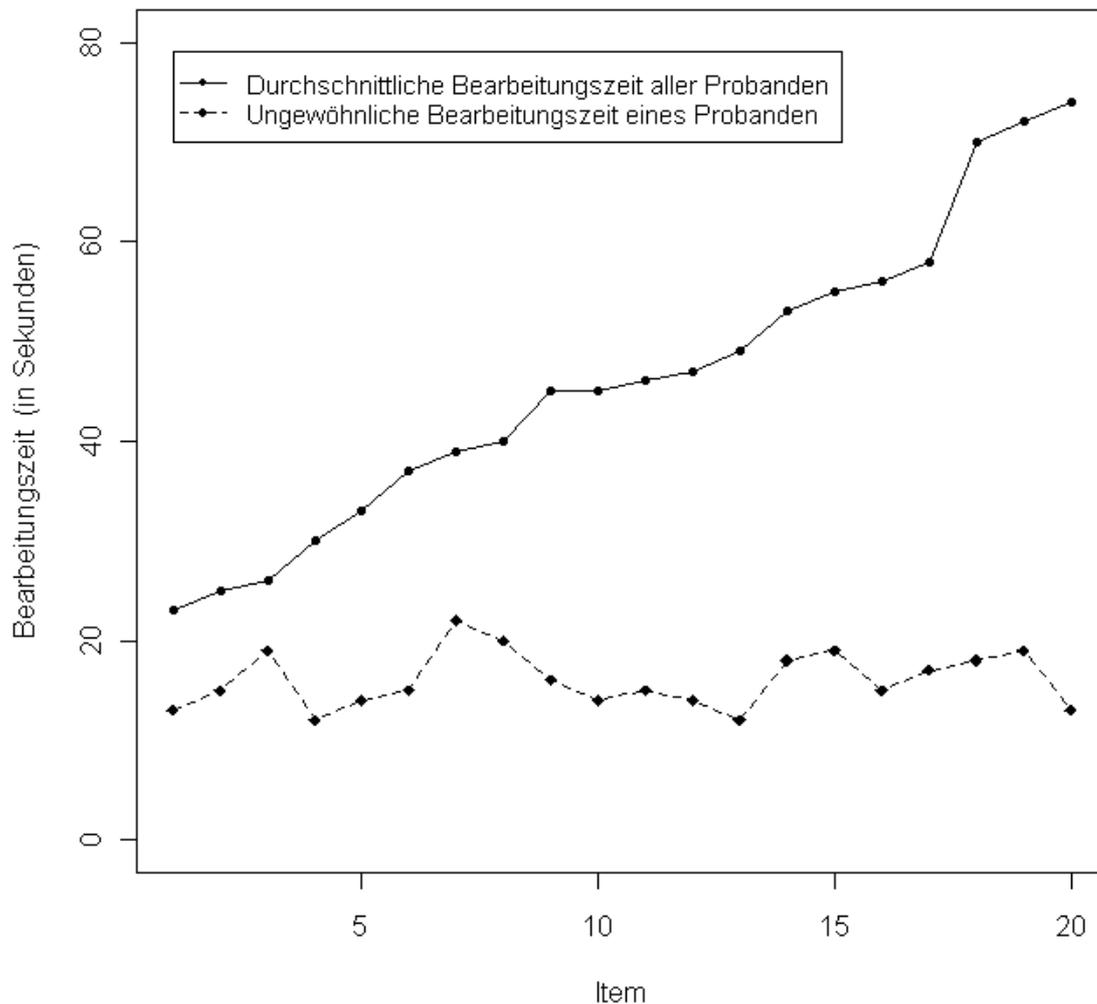


Abbildung 3. Bearbeitungszeit eines 20 Item-Tests als Indiz für Itemvorwissen

Anstelle von Itemantworten werden häufig auch abweichende Bearbeitungszeiten analysiert, da diese am Computer unkompliziert automatisch mitgehoben werden und von Probanden nur schwer absichtlich manipuliert werden können; für einen potentiell betrügerischen Probanden ist es kaum einzuschätzen wie die typische Antwortzeit für ein bestimmtes Item ist, um seine eigene Zeit daran auszurichten. Weist eine Person beispielsweise ungewöhnliche kurze Antwortzeiten vor allem bei vergleichsweise schwierigen Items auf (vgl. Abbildung 3), kann dies einen Indikator für Itemvorwissen darstellen (van der Linden & Guo, 2008). Schließlich wurde auch eine Reihe von Prozeduren entwickelt, um Personen zu identifizieren, die gemeinsam einen Test bearbeiten bzw. von einander abschreiben. Diese Verfahren analysieren wiederum Antwortmuster (Sotaridona, van der Linden, & Meijer, 2006) oder Antwortzeiten (van der Linden, 2009) allerdings von Paaren oder Gruppen von Probanden,

um überzufällige Ähnlichkeiten aufzudecken. All diese Ansätze ermöglichen jedoch nur die Identifizierung von signifikanten Auffälligkeiten und können selten als „Beweis“ für betrügerisches Vorgehen betrachtet werden. In Kombination mit anderen Informationen (z. B. beobachtetes Verhalten bei Realtestungen) helfen sie allerdings verdächtige Probanden zu identifizieren, um deren Leistung nur mit Vorbehalt zu interpretieren.

4.7 Akzeptanz

Akzeptanz wird als das „Ausmaß, in dem subjektive Meinungen, Bewertungen oder gesellschaftspolitische Überzeugungen gegen einen Test angeführt werden“ (Testkuratorium, 1986, S. 360), beschrieben. Diese bezieht sich auf drei Zielgruppen: (a) Probanden, (b) Testanwender (Diagnostiker) sowie (c) Auftraggeber. Liegt die Akzeptanz eines diagnostischen Verfahrens bzw. Methode bei einer dieser Gruppen nicht bzw. in ungenügendem Ausmaß vor, wird deren Anwendung in der Praxis von nur geringem Erfolg gekennzeichnet sein.

Am günstigsten stellt sich die Lage für die Gruppe der Auftraggeber dar. Insbesondere die Aussicht auf ökonomische Einsparungen (sowohl in finanzieller wie auch zeitlicher Hinsicht) verlockt vielerorts die Testapplikation über das Internet vorzunehmen. Dies spiegelt sich eindrucksvoll in Zahlen der kommerziellen Markt- und Meinungsforschung wieder. Innerhalb von nicht einmal 10 Jahren eroberten quantitative Online-Befragungen einen Anteil von über 30 Prozent aller Marktforschungsprojekte (Arbeitskreis deutscher Markt- und Sozialforschungsinstitute, 2008). In manchen Domänen, wie bei Mitarbeiterbefragungen oder ad-hoc Werbemitteltests stellt Internetforschung bereits die dominierende Erhebungsmethodik dar (Wenzel & Hoffmann, 2005). Internetbasierte psychologische Diagnostik prägt derzeit vor allem die berufliche Eignungsdiagnostik. Piotrowski und Armstrong (2006) schätzen den Anteil US-amerikanischer Unternehmen, die derzeit internetbasiertes Bewerberscreening betreiben oder dies in naher Zukunft einführen möchten, auf beinahe ein Drittel. Fallaw und Solomonson (2009) gehen von einem sogar noch höheren Anteil aus: Beinahe 60 Prozent der untersuchten Firmen in den Vereinigten Staaten ermöglichen „remote testing“ (darunter werden internetunterstützte ebenso wie rein telefonische Assessmentvarianten subsumiert) von Jobbewerbern, im europäischen Raum ist dieser Anteil sogar noch deutlich höher. In eingeschränktem Ausmaß spiegelt sich dies auch in der akademischen Fachliteratur wider. Skitka und Sargis (2006) schätzten im Rahmen einer Metaanalyse den Anteil von APA-

Zeitschriften, die in den Jahren 2003/04 zumindest einen Artikel mit internetbasierter Methodik veröffentlicht haben, auf rund 21 Prozent.

Diagnostiker diskutieren vor allem die ethischen Implikationen, die sich aufgrund internetbasierter Testpräsentationen ergeben. Aktuelle ethische Richtlinien zur psychologischen Forschung fordern die vorherige Aufklärung von Teilnehmern wissenschaftlicher Studien: Über Testziel, Testablauf und mögliche Konsequenzen, die aus der Testteilnahme resultieren können (vgl. Deutsche Gesellschaft für Psychologie & Bundesverband deutscher Psychologinnen und Psychologen, 2005). Bei internetbasierten Tests erfolgt dies in der Regel über schriftliche Instruktionen zu Testbeginn. Problematisch erscheint allerdings, dass viele Probanden derartige Informationen höchstens oberflächlich überfliegen aber nur selten genau studieren (Varnhagen et al., 2005; Wogalter, Howe, Sifuentes & Luginbuhl, 1999). Selbst wenn Instruktionen genau gelesen werden, können zahlreiche Einflussfaktoren wie z. B. Lesefähigkeit, Bildung, Alter etc. (vgl. Arscott, Dagnan & Kroese, 1999; Bruzzese & Fisher, 2003) das Verständnis erheblich beeinträchtigen. Bei Face-to-Face Testungen kann der anwesende Diagnostiker den Probanden den Ablauf mündlich genau erklären, auf mögliche Unklarheiten eingehen und durch gezieltes Nachfragen das Verständnis der Instruktion überprüfen. Bei unkontrollierten, internetbasierten Verfahren ist derartige nicht möglich. Ähnliches gilt für automatisierte Ergebnisrückmeldungen an die Probanden. Testergebnisse sollten in der Regel vom Diagnostiker gemeinsam mit dem Probanden diskutiert werden, um auf deren Aussagekraft und Grenzen insbesondere auch in Bezug auf darauf basierende Entscheidungen eingehen zu können. Bei automatisiert generierten Testergebnissen besteht jedoch die Gefahr der Fehl-, vor allem Überinterpretation von Resultaten, wenn Probanden mit unklaren Vorstellungen über die erhobenen Konstrukte und deren Bedeutung eigenständig möglicherweise weitreichende persönliche Entscheidungen darauf basieren. Daher sind schriftliche Testergebnisse mit ausführlichen Interpretationsanleitungen zu versehen, welche die Ausmaß und Grenzen der Generalisierungsfähigkeit der Testergebnisse genau erläutern. Verschärft wird die Problematik unzulässiger Testinterpretationen zudem, wenn Probanden Verfahren bearbeiten, für welche diese Tests gar nicht konstruiert wurden. Prinzipiell eröffnet das Internet auch Personen aus verschiedenen Ländern und Kulturkreisen unkompliziert die Möglichkeit an unkontrollierten Internettestungen aus aller Welt teilzunehmen. Wenn die Konstruktvalidität dieser Tests allerdings nicht vorab im interkulturellen Kontext für Personen

mit z. B. nicht-deutscher Muttersprache überprüft wurde, besteht die Gefahr unzulässiger Schlüsse auf Basis derartiger Testergebnisse (Naglieri et al., 2004).

Auf Seiten der Probanden wird internetbasierten Verfahren im Allgemeinen große Zufriedenheit bescheinigt (vgl. Mangunkusumo et al., 2005; Salgado & Moscoso, 2003; Weber et al., 2003). So attestiert Kubinger (2009, S. 125) der Computardiagnostik, dass sie in der Praxis „außerordentlich gut akzeptiert“ wird. Salgado und Moscoso (2003) berichten beispielsweise, dass Probanden die Internetversion eines Persönlichkeitsfragebogens angenehmer und weniger einschüchternd erlebten und generell die Online-Erhebung einer Papier-Bleistift-Testung vorzogen. Selbst Psychatriepatienten mit wenig Computererfahrung zeigen sich mit Computertests als weitgehend zufrieden und kaum benachteiligt (Weber et al., 2003). Allerdings ist die Akzeptanz von internetbasierter Diagnostik stark situations- wie auch personenabhängig. Bei Mitarbeiterbefragungen ist die Beteiligungsrate stark von der wahrgenommenen Anonymität der Erhebung abhängig. Hegen die Probanden die Befürchtung, dass ihre Antworten unter Umständen mit der eigenen Person in Verbindung gebracht werden könnten, ist mit starkem Non-Response (Antwortverweigerung) zu rechnen. Dies kann der Fall sein, wenn derartige Befragungen unternehmensintern über geschlossene Intranetsysteme, zu denen man erst nach Anmeldung mit Benutzername und Passwort Zugriff erlangt, durchgeführt werden (Thompson & Surface, 2007). In solchen Fällen kann ein Outsourcing an externe Stellen helfen, das Vertrauen in die Anonymität zu erhöhen und die Responserate zu steigern. Zusätzlich können personenspezifische Faktoren die Akzeptanz von internetbasiertem Assessment einschränken. Personen mit einer geringem Computererfahrung beurteilen internetbasierte Verfahren im Recrutementprozess beispielsweise als weniger fair und die Augenscheinvalidität des Selektionsprozesses als gering (Wiechmann & Ryan, 2003). Ähnliches wird auch in Bezug auf videobasierte Online-Interviews bei Bewerbungsgesprächen berichtet. Insbesondere spezifische Persönlichkeitseigenschaften moderieren die wahrgenommene Nützlichkeit und Fairness derartige Selektionsprozeduren (Chapman, Uggerslev, & Webster, 2003). Personen mit hoher Selbstüberwachung (Snyder, 1974), die sehr sensible auf soziale und interpersonale Cues reagieren und entsprechend ihr Verhalten anpassen, sehen sich durch Interviewformen, bei denen die non-verbale Kommunikationsebene fehlt (Telefoninterviews) oder nur eingeschränkt verfügbar ist (Videointerviews), in ihren Selbstdarstellungsfähigkeiten deutlich eingeschränkt und lehnen dementsprechend derartige Selektionsprozesse verstärkt als wenig fair ab.

5 Resümee und Ausblick

Internetbasierte psychologische Diagnostik ist ein vergleichsweise junges Gebiet, das erst durch die rasante Verbreitung moderner Informationstechnologien sowohl im privaten wie auch beruflichen Alltag ermöglicht wurde. Innerhalb nicht einmal einer Dekade hat das Internet einen unverzichtbaren Stellenwert insbesondere für eignungsdiagnostische Zwecke im Rahmen der Personalauswahl wie auch –entwicklung aber auch für allgemeine Zertifizierungsmaßnahmen (wie z. B. den TOEFL-Test) eingenommen. Dieser Wandel in der Art, wie psychologische Tests appliziert werden, erforderte auch die Adaptierung bestehender bzw. Entwicklung neuer Qualitätsstandards, die als Referenzmanual für Testentwickler, Testanbieter und auch Testanwender Leitlinien und Handlungsanweisungen für internetbasierte Diagnostik formulieren. Internationale Richtlinien wurden u. a. von der Internationalen Organisation für Standardisierung (ISO, 2007) und der International Test Commission (ITC, 2006) vorgestellt. Diese fassen allgemeine Empfehlungen für computer- und internetbasiertes psychologisches Assessment zusammen, die länderübergreifend Akzeptanz finden. So umfassen die ISO/IEC 23988:2007 Richtlinien für den Einsatz von Informationstechnologien zur Testadministration, –auswertung und Rückmeldung von Testergebnissen. Allerdings beziehen sie sich nur auf Leistungstest zur Erfassung von Wissen und Fertigkeiten. Andere Formen psychologisch-diagnostischer Verfahren wie Interessens- oder Persönlichkeitstests berücksichtigen sie jedoch nicht. Die Empfehlungen der ITC sind dem gegenüber weiter gefasst und beziehen alle Varianten internetbasierter, psychologisch-diagnostischer Verfahren ein. Sie thematisieren u. a. technische Anforderungen an Hard- und Software, ergonomische Überlegungen zur Gestaltung des graphischen Nutzerinterface, Fragen des Datenschutzes und der Testfairness für Personen mit geringer Medienkompetenz bzw. mit körperlichen Beeinträchtigungen, psychometrische Qualität internetbasierter Tests, Maßnahmen zur Identifizierung von Probanden und der Minimierung von Testverfälschung u. v. m.

Bislang ist internetbasierter Diagnostik primär traditionellen Assessmentvorstellungen verhaftet, bei denen Probanden in einem mehr oder weniger stark kontrollierten Setting mit standardisierten Stimuli konfrontiert werden, um deren Reaktionen zu registrieren und darauf basierende Schlussfolgerungen abzuleiten. Unter dem Gesichtspunkt von psychologischer Diagnostik als einer Form „systematische[r] Datenerhebung, -verarbeitung und –interpretation [zur Vorhersage von] Erleben und Verhalten“ (Kersting, 2008, S. 11), erscheinen im Kontext internetbasierter Diagnostik auch alternative Zugänge vorstellbar, die

andere Informationsquellen abseits explizit evozierter Reaktionen auf Testitems einschließen. Einerseits kann internetbasierte Diagnostik auf ambulatorische Assessmentszenarien erweitert werden, bei denen das natürliche Verhalten einer Person in ihrem Alltag Untersuchungsgegenstand ist (vgl. Ebner-Priemer & Trull, 2009; Klumb, Elfering & Herre, 2009). Insbesondere für klinische aber auch arbeitspsychologische Fragestellungen können Entwicklungen im Bereich gängiger mobiler Endgeräte wie Handys oder PDAs, die viele Personen ständig mit sich führen, eingesetzt werden, um beispielsweise

- standardisierte Leistungstests zu applizieren und Schwankungen über den Tagesverlauf zu registrieren,
- subjektive Befindlichkeiten (z. B. Furcht oder Zufriedenheit) im zeitnahen Längsschnitt und nicht über retrospektive Selbstberichte zu erfassen,
- die Sprechaktivität aufzuzeichnen, um aus dem Verhältnis von Sprechaktivität und Sprechpausen auf das Sozial- und Interaktionsverhalten zu schließen oder
- kontinuierlich psychophysiologische Maße wie Herzrate oder Blutdruck aufzuzeichnen.

Andererseits ermöglicht das Internet bereits heute die Beobachtung von Varianten natürlichen „online“ Verhaltens. Das Internet ist zunehmend nicht nur Gegenstand passiver Informationsrezeption sondern vor allem auch ein sozialer Ort aktiver Partizipation. Personen verbringen zunehmend mehr Zeit „online“, um zu kommunizieren, zu interagieren (z. B. in Multiuser-Online-Spielen) und vor allem auch neue Produkte zu generieren: enzyklopädie-ähnliche Wissensdatenbanken, Fotosammlungen, neue Softwareprodukte etc. Im Gegensatz zu Verhalten in der „offline“ Welt werden im Internet prinzipiell alle Aktionen automatisch protokolliert und hinterlassen so umfangreiche Verhaltensspuren der Nutzer. Zunehmend stellen Personen sogar aktiv Informationen über sich selbst in sozialen Netzwerkdiensten und auf privaten Homepages freiwillig zur Verfügung, sodass sich für diagnostische Zwecke manchmal nicht die Frage stellt wie neue Daten über das Internet *erhoben* sondern vielmehr wie bestehende Informationen über eine Person am besten *gefunden* werden können. Erste Ergebnisse deuten darauf hin, dass diese Datenquellen reichhaltige Informationen über z. B. Persönlichkeitseigenschaften von Probanden liefern können. Häufig reicht die Betrachtung des Profils einer Person in einem sozialen Netzwerkdienst (Evans, Gosling & Carroll, 2008) oder der Besuch einer persönlichen Homepage (Marcus, Machilek & Schütz, 2006; Vazire & Gosling, 2004) aus, um akkurate Persönlichkeitsattribute über diese Person abzuleiten. Korrelationen zwischen Selbsteinschätzungen anhand etablierter Persönlichkeitsinventare und

Fremdeinschätzungen, die auf dem ersten Eindruck einer Homepage basieren, erreichen für z. B. Offenheit für Erfahrungen eine Höhe von bis zu .63 (Vazire & Gosling, 2004). Selbst die Kenntnis von Musikpräferenzen einer Person, die häufig als eine Grundinformation über die eigene Person in internetbasierten Profilen veröffentlicht werden, ermöglicht unbekanntem Urteilern Persönlichkeitsmerkmale konsistent und vor allem auch akkurat abzuleiten (Rentfrow & Gosling, 2006). Neue Ansätze gehen bereits einen Schritt weiter und experimentieren mit automatisierten Anwendungen, um Persönlichkeitsmerkmale einer Person auf Basis schriftlicher Äußerungen z. B. aus Email- oder Blogtexten ohne Zuhilfenahme menschlicher Urteiler zu identifizieren (vgl. Mairesse, Walker, Mehl & Moore, 2007). Dabei werden Textmerkmale wie z. B. Syntax oder Wortschatz mithilfe statistischer Methoden (z. B. regressive Techniken oder neuronale Netze) analysiert, um Persönlichkeitsmerkmale der Autoren vorherzusagen. Linguistische Textmerkmale von Blogbeiträgen erreichen für die Big Five der Persönlichkeit bereits eine Varianzaufklärung von bis zu 65% des interessierenden Traits (Nowson & Oberlander, 2006). Bislang stellen derartige Ansätze noch keine Alternativen zum traditionellen diagnostischen Vorgehen dar, bei denen anhand standardisierter Tests explizit Informationen von Probanden gesammelt werden. Künftig erscheinen vergleichbare Ansätze jedoch vielversprechend, um auf Basis bereits vorliegender Verhaltensspuren erste Grobinformationen über Personen, z. B. mögliche Bewerber im Recruitingprozess, auch hinsichtlich psychologisch relevanter Merkmale wie Persönlichkeit zu sammeln.

6 Literatur

- Aguinis, H., Henle, C. A., & Beaty, J. C. Jr. (2001). Virtual reality technology: A new tool for personnel selection. *International Journal of Selection and Assessment*, 9, 70-83. doi:10.1111/1468-2389.00164
- Aluja, A., Rossier, J., & Zuckerman, M. (2007). Equivalence of paper and pencil vs internet forms of the ZKPQ-50-CC in Spanish and French samples. *Personality and Individual Differences*, 43, 2022-2032. doi:10.1016/j.paid.2007.06.007
- Al-Saggaf, Y., & Williamson, K. (2004). Online communities in Saudi Arabia: Evaluating the impact on culture through online semi-structured interviews. *Forum Qualitative Sozialforschung*, 5(3). Abgerufen von <http://www.qualitative-research.net>
- Amelang, M., & Schmidt-Atzert, L. (2006). *Psychologische Diagnostik und Intervention*. Heidelberg: Springer.
- Anderson, N. (2003). Applicant and recruiter reactions to new technology in selection: A critical review and agenda for future research. *International Journal of Selection and Assessment*, 11, 121-136. doi:10.1111/1468-2389.00235
- Andersson, G., Bakhsh, R., Johansson, L., Kaldo, V., & Carlbring, P. (2005). Stroop facilitation in tinnitus patients: An experiment conducted via the World Wide Web. *Cyberpsychology & Behavior*, 8, 32-38. doi:10.1089/cpb.2005.8.32
- Andersson, G., Westöö, J., Johansson, L., & Carlbring, P. (2006). Cognitive bias via the Internet: A comparison of web-based and standard emotional Stroop tasks in social phobia. *Cognitive Behaviour Therapy*, 35, 55-62. doi:10.1080/16506070500372469
- Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute (ADM). (2008). *Jahresbericht 2008*. Abgerufen von der Website des ADM unter http://www.adm-ev.de/fileadmin/user_upload/PDFS/Jahresbericht_08.pdf
- Armstrong, P. I., Allison, W., & Rounds, J. (2008). Development and initial validation of brief public domain RIASEC marker scales. *Journal of Vocational Behavior*, 73, 287-299. doi:10.1016/j.jvb.2008.06.003
- Arcsott, K., Dagnan, D., & Kroese, B. S. (1999). Assessing the ability of people with a learning disability to give informed consent to treatment. *Psychological Medicine*, 29, 1367-1375. doi:10.1017/S0033291799008715
- Austin, D. W., Carlbring, P. Richards, C. J., & Anderssen, 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. doi:10.1207/s15327574ijt0601_2
- Backhaus, K. B. (2004). An exploration of corporate recruitment descriptions on Monster.com. *Journal of Business Communication*, 41, 115-136. doi:10.1177/0021943603259585
- Bampton, R., & Cowton, C. J. (2002). The e-interview. *Forum Qualitative Sozialforschung*, 3(2). Abgerufen von <http://www.qualitative-research.net>
- Barak, A., & Buchanan, T. (2004). Internet-based psychological testing and assessment. In R. Kraus, J. Zack, & G. Stricker (Hrsg.), *Online Counseling: A Handbook for Mental Health Professionals* (S. 217-239). Boston: Academic Press.
- Bartram, D. (2000). Internet recruitment and selection: Kissing frogs to find princes. *International Journal of Selection and Assessment*, 8, 261-274. doi:10.1111/1468-2389.00155
- Bartram, D. (2006). Testing on the internet: Issues, challenges and opportunities in the field of occupational assessment. In D. Bartram & R. K. Hambleton (Hrsg.), *Computer-Based Testing and the Internet: Issues and Advances* (S. 13-37). New York: Wiley.
- Batinic, B. (2004). Online-Research. In G. Bente, R. Mangold, & P. Vorderer (Hrsg.), *Lehrbuch der Medienpsychologie* (S. 251-270). Göttingen: Hogrefe.
- Batinic, B., & Bosnjak, M. (2000). Fragebogenuntersuchungen im Internet. In B. Batinic (Hrsg.), *Internet für Psychologen* (S. 287-317). Göttingen: Hogrefe.
- Batinic, B., & Appel, M. (2009). Nutzung von Online-Bewerbungen aus Sicht von Bewerbern und Unternehmen. *Zeitschrift für Personalpsychologie*, 8, 14-23. doi:10.1026/1617-6391.8.1.14
- Bennett, R. E. (2003). *Online assessment and the comparability of score meaning* (Bericht 03-05). Abgerufen von der Website des Educational Testing Service unter <http://www.ets.org/Media/Research/pdf/RM-03-05-Bennett.pdf>
- Bennett, R. E., Braswell, J., Oranje, A., Sandene, B., Kaplan, B., & Yan, F. (2008). Does it matter if I take my mathematics test on computer? A second empirical study of mode effects in NAEP. *Journal of Technology, Learning, and Assessment*, 6(9). Abgerufen von <http://www.jtla.org>
- Bergstrom, B. (1992, April). *Ability measure equivalence of computer adaptive and pencil and paper tests: A research synthesis*. Vortrag auf dem Annual Meeting of the American Educational Research Association, San Francisco (USA).

- Birnbaum, M. H. (2000). Decision making in the lab and on the web. In M. H. Birnbaum (Hrsg.), *Psychological experiments on the internet* (S. 3-34). San Diego: Academic Press.
- Birnbaum, M. H. (2004). Human research and data collection via the internet. *Annual Review of Psychology*, *55*, 803-832. doi:10.1146/annurev.psych.55.090902.141601
- Bluemke, M., & Friese, M. (2008). Reliability and validity of the Single-Target IAT (ST-IAT). *European Journal of Social Psychology*, *38*, 977-997. doi:10.1002/ejsp.487
- Booth-Kewley, S., Larson, G., & Miyoshi, D. (2007). Social desirability effects on computerized and paper-and-pencil questionnaires. *Computers in Human Behavior*, *23*, 463-477. doi:10.1016/j.chb.2004.10.020
- Bortz, J., & Döring, N. (2006). *Forschungsmethoden und Evaluation*. Berlin: Springer.
- Braunsberger, K., Wybenga, H., & Gates, R. (2007). A comparison of reliability between telephone and web-based surveys. *Journal of Business Research*, *60*, 758-764. doi:10.1016/j.jbusres.2007.02.015
- Brengman, M., Geuens, M., Weijters, B., Smith, S. M., & Swinyard, W. R. (2005). Segmenting internet shoppers based on their web-usage-related lifestyle: a cross-cultural validation. *Journal of Business Research*, *58*, 79-88. doi:10.1016/S0148-2963(02)00476-9
- Bridgeman, B., Lennon, M. L., & Jackenthal, A. (2003). Effects of screen size, screen resolution, and display rate on computer-based test performance. *Applied Measurement in Education*, *16*, 191-205. doi:10.1207/S15324818AME1603_2
- Bruzzese, J.-M., & Fisher, C. B. (2003). Assessing and enhancing the research consent capacity of children and youth. *Applied Developmental Science*, *7*, 13-26. doi:10.1207/S1532480XADS0701_2
- Buchanan, T. (2000). Potential of the internet for personality research. In M. H. Birnbaum (Hrsg.), *Psychological experiments on the internet* (S. 121-140). San Diego: Academic Press.
- Buchanan, T., Ali, T., Heffernan, T., Ling, J., Parrott, A., Rodgers, J., & Scholey, A. (2005). Nonequivalence of on-line and paper-and-pencil psychological tests: The case of the prospective memory questionnaire. *Behavior Research Methods*, *37*, 148-154.
- Burke, E. (2009). Preserving the integrity of online testing. *Industrial and Organizational Psychology*, *2*, 35-38. doi:10.1111/j.1754-9434.2008.01104.x

- Chaney, M. P., & Dew, B. J. (2003). Online experience of sexually compulsive men who Have sex with men. *Sexual Addiction & Compulsivity, 10*, 259-274.
doi:10.1080/713775414
- Chapelle, C. A., Jamieson, J., & Hegelheimer, V. (2003). Validation of a web-based ESL test. *Language Testing, 20*, 409-439. doi:10.1191/0265532203lt266oa
- Chapman, D.S., Uggerslev, K.L., & Webster, J. (2003). Applicant reactions to face-to-face and technology-mediated interviews: A field investigation. *Journal of Applied Psychology, 88*, 944-953. doi:10.1037/0021-9010.88.5.944
- Chapman, D. S., & Webster, J. (2003). The use of technologies in the recruiting, screening, and the selection processes for job candidates. *International Journal of Selection and Assessment, 11*, 113-120. doi:10.1111/1468-2389.00234
- Chuah, S. C., Drasgow, F., & Roberts, B. W. (2006), Personality assessment: Does the medium matter? *Journal of Research in Personality, 40*, 359-376.
doi:10.1016/j.jrp.2005.01.006
- Cole, M. S., Bedeian, A. G., & Feild, H. S. (2006). The measurement equivalence of web-based and paper-and-pencil measures of transformational leadership: A multinational test. *Organizational Research Methods, 9*, 339-368. doi:10.1177/1094428106287434
- Davis, M., Bolding, G., Hart, G., Sherr, L., & Elford, J. (2004). Reflecting on the experience of interviewing online: perspectives from the Internet and HIV study in London. *AIDS Care, 16*, 944-952. doi:10.1080/09540120412331292499
- Deutsche Gesellschaft für Psychologie (DGPs) & Bundesverband deutscher Psychologinnen und Psychologen (BDP). (2005). *Ethische Richtlinien der DGPs und BDP*. Verfügbar auf der Website des BDP unter http://www.bdp-verband.org/bdp/verband/clips/BDP_Ethische_Richtlinien_2005.pdf
- Diercks, J., & Weber, A. (2005). Recrutainment – Bewerberauswahl mit Recruitment-Spielen. In U. Blötz (Hrsg.), *Planspiele in der beruflichen Bildung* [CD-ROM]. Bielefeld: Bertelsmann.
- Dörner, D. (1989). *Die Logik des Mißlingens*. Hamburg: Rowohlt.
- Drapeau, C. E., Bastien-Toniazzo, M., Rous, C., & Carlier, M. (2007). Nonequivalence of computerized and paper-and-pencil versions of trail making test. *Perceptual and Motor Skills, 104*, 785-791. doi:10.2466/PMS.104.3.785-791
- Drasgow, F., Nye, C. D., Guo, J., & Tay, L. (2009). Cheating on proctored tests: The other side of the unproctored debate. *Industrial and Organizational Psychology, 2*, 46-48.
doi:10.1111/j.1754-9434.2008.01106.x

- Drasgow, F., Olson-Buchanan, J. B., & Moberg, P.J. (1999). Development of an interactive video assessment: Trials and tribulations. In F. Drasgow & J. B. Olson-Buchanan (Hrsg.), *Innovations in computerized assessment* (S. 177-196). Mahwah, NJ: Erlbaum.
- Dwight, S., & Feigelson, M. E. (2000). A quantitative review of the effect of computerized testing on the measurement of social desirability. *Educational and Psychological Measurement, 60*, 340–360. doi:10.1177/00131640021970583
- Ebner-Priemer, U. W. & Trull, T. J. (2009). Ambulatory assessment - An innovative and promising approach for clinical psychology. *European Psychologist, 14*, 109-119. doi:10.1027/1016-9040.14.2.109
- Educational Testing Service (ETS). (2008). *TOEFL iBT Tips: How to prepare for the TOEFL iBT*. Verfügbar auf der Website des ETS unter http://www.ets.org/Media/Tests/TOEFL/pdf/TOEFL_Tips.pdf
- Ehrhart, K. H., Roesch, S. C., Ehrhart, M. G., & Kilian, B. (2008). A test of the factor structure equivalence of the 50-item IPIP Five-factor model measure across gender and ethnic groups. *Journal of Personality Assessment, 90*, 507-516. doi:10.1080/00223890802248869
- Eichenberg, C. (2008). Bedeutung der Medien für klinisch-psychologische Interventionen. In B. Batinic & M. Appel (Hrsg.), *Medienpsychologie* (S. 503-530). Berlin: Springer.
- Eichstaedt, J. (2001). An inaccurate-timing filter for reaction-time measurement by JAVA-applets implementing Internet-based experiments. *Behavior Research Methods, Instruments, & Computers, 33*, 179-186.
- Eichstaedt, J. (2005). Online-Assessment-Center: Ein neues Paradigma. In K.-H. Renner, A. Schütz, & F. Machilek (Hrsg.), *Internet und Persönlichkeit: Differentiell-psychologische und diagnostische Aspekte der Internetnutzung* (S. 315-326). Göttingen: Hogrefe.
- Eid, G. K. (2004). An investigation into the effects and factors influencing computer-based online math problem-solving in primary schools. *Journal of Educational Technology Systems, 33*, 223–240. doi:10.2190/J3Q5-BAA5-2L62-AEY3
- Eimeren, B., & Frees, B. (2010). *ARD / ZDF Onlinestudie 2010*. Abgerufen unter <http://www.ard-zdf-onlinestudie.de>
- Ekman, A., Dickman, P., Klint, A., Weiderpass, E., & Litton, J.-E. (2006). Feasibility of using web-based questionnaires in large population-based epidemiological studies. *European Journal of Epidemiology, 21*, 103-111. doi:10.1007/s10654-005-6030-4

- Embretson, S., & Yang, X. (2007). Automatic item generation and cognitive psychology. In C. R. Rao & S. Sinharay (Hrsg.), *Handbook of Statistics: Psychometrics* (Band 26, S. 747-768). Amsterdam: Elsevier.
- Erdogan, G. (2001). Die Gruppendiskussion als qualitative Datenerhebung im Internet: Ein Online-Offline-Vergleich. *kommunikation@gesellschaft*, 2. Verfügbar unter <http://www.soz.uni-frankfurt.de/K.G/>
- Erlanger, D., Feldman, D., Kutner, K., Kaushik, T., Kroger, H., Festa, J., Barth, J., Freeman, J., & Broshek, D. (2003). Development and validation of a web-based neuropsychological test protocol for sports-related *return-to-play* decision-making. *Archives of Clinical Neuropsychology*, 18, 293-316. doi:10.1016/S0887-6177(02)00138-5
- Evans, D. C., Gosling, S. D., & Carroll, A. (2008, April). *What elements of an online social networking profile predict target-rater agreement in personality impressions?* Vortrag auf der International Conference on Weblogs and Social Media. Seattle (USA). Verfügbar unter http://www.psychster.com/library/EvansGoslingCarroll_ICWSM08.pdf
- Fallow, S. S., & Solomonson, A. L. (2009). *Global assessment trend report* (Bericht 2009). Abgerufen von Previsor Talent Measurement Website unter <http://www.previsor.com>
- Fortson, B., Scott, J., Del Ben, K., & Chen, Y. (2006). Reliability and validity of an internet traumatic stress survey with a college student sample. *Journal of Traumatic Stress*, 19, 709-720. doi:10.1002/jts.20165
- Fouladi, R. T., McCarthy, C. J., & Moller, N. P. (2002). Paper-and-pencil or online? Evaluating mode effects on measures of emotional functioning and attachment. *Assessment*, 9, 204-215. doi:10.1002/jts.20165
- Fox, S., & Schwartz, D. (2002). Social desirability and controllability in computerized and paper-and-pencil personality questionnaires. *Computers in Human Behavior*, 18, 389-410. doi:10.1016/S0747-5632(01)00057-7
- Frey, A., Blunk, H., & Banse, R. (2006). Psi-Land. Paarinteraktionsforschung in einer virtuellen Umgebung. *Zeitschrift für Sozialpsychologie*, 37, 151-159. doi:10.1024/0044-3514.37.3.151
- Frey, A., Hartig, J., Ketzler, A., Zinkernagel, A., & Moosbrugger, H. (2007). The use of virtual environments based on a modification of the computer game Quake III Arena[®] in psychological experimenting. *Computers in Human Behavior*, 23, 2026-2039. doi:10.1016/j.chb.2006.02.010

- Gasteiger, R. M., & Montel, C. (2005). Kulturübergreifende Kompetenzmessung über das Internet. In K.-H. Renner, A. Schütz, & F. Machilek (Hrsg.), *Internet und Persönlichkeit* (S. 327-340). Göttingen: Hogrefe.
- Gnambs, T. (2008). *Graphische Analogskalen in internetbasierten Fragebögen*. München: Akademischer Verlag.
- Gnambs, T., & Batinic, B. (2007). Online Research. In E. Balzer & G. Naderer, *Qualitative Marktforschung in Theorie und Praxis* (S. 343-362). Wiesbaden: Gabler.
- Gnambs, T., & Strassnig, B. (2007). Experimentelle Online-Untersuchungen. In M. Welker & O. Wenzel (Hrsg.), *Online-Forschung 2007 - Grundlagen und Fallstudien* (S. 233-250). Köln: Halem.
- Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, *40*, 84-96.
doi:10.1016/j.jrp.2005.08.007
- Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *American Psychologist*, *59*, 93-104. doi:10.1037/0003-066X.59.2.93
- Greenwald, A., McGhee, D., & Schwartz, J. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*, 1464–1480. doi:10.1037/0022-3514.74.6.1464
- Haworth, C., Harlaar, N., Kovas, Y., Davis, O., Oliver, B., Hayiou-Thomas, M., Frances, J., Busfield, P., McMillan, A., Dale, P., & Plomin, R. (2007). Internet cognitive testing of large samples needed in genetic research. *Twin Research and Human Genetics*, *10*, 554-563. doi:10.1375/twin.10.4.554
- Herrero, J., & Meneses, J. (2006). Short web-based versions of the Perceived Stress (PSS) and Center for Epidemiological Studies-Depression (CESD) Scales: A comparison to pencil and paper responses among internet users. *Computers and Human Behavior*, *22*, 830-846. doi:10.1016/j.chb.2004.03.007
- Hertel, G., & Konradt, U. (2004). *Human Resource Management im Inter- und Intranet*. Göttingen: Hogrefe.
- Hertel, G., Konradt, U., & Orlikowski, B. (2003). Ziele und Strategien von E-Assessment aus Sicht der psychologischen Personalauswahl. In U. Konradt & W. Sarges (Hrsg.), *E-Recruitment und E-Assessment. Rekrutierung, Auswahl und Beurteilung von Personal im Inter- und Intranet* (S. 37-54). Göttingen: Hogrefe.

- Hertel, G., & Schroer, J. (2008). Electronic Human Resource Management (E-HRM): Personalarbeit mit netzbasierten Medien. In B. Batinic & M. Appel (Hrsg.), *Medienpsychologie* (S. 449-476). Berlin: Springer.
- Hilty, D., Marks, S., Urness, D., Yellowlees, P., & Nesbitt, T. (2004). Clinical and educational telepsychiatry applications: A Review. *Canadian Journal of Psychiatry*, *49*, 12-23.
- Hornke, L., & Winterfeld, U. (2004). *Eignungsbeurteilungen auf dem Prüfstand: DIN 33430 zur Qualitätssicherung*. Heidelberg: Spektrum Akademischer Verlag.
- Houbjen, K., & Wiers, R. W. (2008). Measuring implicit alcohol associations via the internet: Validation of web-based implicit association tests. *Behavior Research Methods*, *40*, 1134-1143. doi:10.3758/BRM.40.4.1134
- Ihme, J. M., Lemke, F., Lieder, K., Martin, F., Müller, J. C., & Schmidt, S. (2009). Comparison of ability tests administered online and in the laboratory. *Behavior Research Methods*, *41*, 1183-1189. doi:10.3758/BRM.41.4.1183
- Internationale Organisation für Standardisierung (ISO). (2007). *BS ISO/IEC 23988:2007: Information technology. A code of practice for the use of information technology (IT) in the delivery of assessments*. Genf: ISO.
- International Test Commission (ITC). (2006). International guidelines on computer-based and internet-delivered testing. *International Journal of Testing*, *6*, 143-171. doi:10.1207/s15327574ijt0602_4
- Jäger, R. S., & Petermann, F. (1999). *Psychologische Diagnostik – Ein Lehrbuch*. Weinheim: Beltz PVU.
- Joinson, A. N. (1999). Social desirability, anonymity, and internet based research questionnaires. *Behaviour Research Methods, Instruments & Computers*, *31*, 43-48.
- Joinson, A. N. (2001). Self-disclosure in computer-mediated communication: The role of self-awareness and visual anonymity. *European Journal of Social Psychology*, *31*, 177-192. doi:10.1002/ejsp.36
- Joinson, A. N., & Paine, C. (2006). Self-disclosure, privacy and the internet. In A. Joinson, K. McKenna, T. Postmes, & U.-D. Reips (Hrsg.), *The Oxford Handbook of Internet Psychology* (S. 237-252). Oxford: University Press.
- Joinson, A. N., Woodley, A., & Reips, U.-D. (2007). Personalization, authentication and self-disclosure in self-administered Internet surveys. *Computers in Human Behavior*, *23*, 275-285. doi:10.1016/j.chb.2004.10.012

- Jones, S. R., Fernyhough, C., de-Wit, L., & Meins, E. (2008). A message in the medium? Assessing the reliability of psychopathology e-questionnaires. *Personality and Individual Differences, 44*, 349-359. doi:10.1016/j.paid.2007.08.014
- Joubert, T., & Kriek, H. (2009). Psychometric comparison of paper-and-pencil and online personality assessments in a selection setting. *SA Journal of Industrial Psychology, 35*, 1-11.
- Jude, N., & Wirth, J. (2007). Neue Chancen bei der technologiebasierten Erfassung von Kompetenzen. In J. Hartig & E. Klieme (Hrsg.), *Möglichkeiten und Voraussetzungen technologiebasierter Kompetenzdiagnostik* (S. 49-56). Berlin: Bundesministerium für Bildung und Forschung.
- Jurecka, A., & Hartig, J. (2007). Anwendungsszenarien computer- und netzwerkbasierter Assessments. In J. Hartig & E. Klieme (Hrsg.), *Möglichkeiten und Voraussetzungen technologiebasierter Kompetenzdiagnostik* (S. 69-79). Berlin: Bundesministerium für Bildung und Forschung.
- Kanning, U. P., Hofer, S., & Schulze Willbrenning, B. (2004). *Professionelle Personenbeurteilung*. Göttingen: Hogrefe.
- Kersting, M. (2008). *Qualität in der Diagnostik und Personalauswahl – der DIN-Ansatz*. Göttingen: Hogrefe.
- Kim, J.-P. (1999, Oktober). *Meta-analysis of equivalence of computerized and P&P tests on ability measures*. Vortrag auf dem Annual Meeting of the Mid-Western Educational Research Association, Chicago (USA).
- Klumb, P. L., Elfering, A., & Herre, C. (2009). Ambulatory-assessment studies in industrial and organizational psychology. *European Psychologist, 14*, 120-131. doi:10.1027/1016-9040.14.2.120
- Konradt, U., Hertel, G., & Joder, K. (2003). Web-based assessment of call center agents: Development and validation of a computerized instrument. *International Journal of Selection and Assessment, 11*, 184-193. doi:10.1111/1468-2389.00242
- Konradt, U., Lehmann, K., Böhm-Rupprecht, J., & Hertel, G. (2003). Computer- und internetbasierte Verfahren der Berufseignungsdiagnostik: Ein empirischer Überblick. In U. Konradt & W. Sarges (Hrsg.), *E-Recruitment und E-Assessment. Rekrutierung, Auswahl und Beurteilung von Personal im Inter- und Intranet* (S. 105-124). Göttingen: Hogrefe.

- Konradt, U., & Rack, O. (2006). Personalrekrutierung im Internet. Einfluss der Qualität von Recruiting-Sites auf die Arbeitgeberattraktivität. *Zeitschrift für Personalpsychologie*, 5, 53-59. doi:10.1026/1617-6391.5.2.53
- Konradt, U., & Sarges, W. (2003). *E-Recruitment und E-Assessment*. Göttingen: Verlag für Angewandte Psychologie.
- Krohne, H. W., & Hock, M. (2007). *Psychologische Diagnostik*. Stuttgart: Kohlhammer.
- Kubinger, K. D. (2006). Ein Update der Definition von Objektiven Persönlichkeitstests: Experimentalpsychologische Verhaltensdiagnostik. In T. M. Ortner, R. Proyer & K. D. Kubinger (Hrsg.), *Theorie und Praxis Objektiver Persönlichkeitstests* (S. 38-52). Bern: Huber.
- Kubinger, K. D. (2009). *Psychologische Diagnostik*. Göttingen: Hogrefe.
- Lesage, M., Riopel, M., Raiche, G., & Sodoke, K. (2008). IMS-QTI sub-standards in computerised adaptive testing and interfacing. *International Journal of Advanced Media and Communication*, 2, 115-137. doi:10.1504/IJAMC.2008.018503
- Lexcen, F., Hawk, G., Herrick, S., & Blank, M. (2006). Use of video conferencing for psychiatric and forensic evaluations. *Psychiatric Services*, 57, 713-715. doi:10.1176/appi.ps.57.5.713
- van der Linden, W. J. (2009). A bivariate lognormal response-time model for the detection of collusion between test takers. *Journal of Educational and Behavioral Statistics*, 34, 378-394. doi:10.3102/1076998609332107
- van der Linden, W. J., & Guo, F. (2008). Bayesian procedures for identifying aberrant response-time patterns in adaptive testing. *Psychometrika*, 73, 365-384. doi:10.1007/s11336-007-9046-8
- Linnman, C., Carlbring, P., Åhman, Å., Andersson, H., & Andersson, G. (2006). The Stroop effect on the internet. *Computers in Human Behavior*, 22, 448-455. doi:10.1016/j.chb.2004.09.010
- Mangunkusumo, R., Moorman, P., Van Den Berg-de Ruyter, A., Van Der Lei, J., De Koning, H., & Raat, H. (2005). Internet-administered adolescent health questionnaires compared with a paper version in a randomized study. *Journal of Adolescent Health*, 36, 70.e1. doi:10.1016/j.jadohealth.2004.02.020
- Mairesse, F., Walker, M. A., Mehl, M. R., Moore, R. K. (2007). Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research*, 30, 457-500.

- Mann, C., & Stewart, F. (2002). *Internet communication and qualitative research: A handbook for researching online*. London: Sage.
- Matsuura, S., Hosaka, T., Yukiya, T., Ogushi, Y., Okada, Y., Haruki, Y., & Nakamura, M. (2000). Application of telepsychiatry: A preliminary study. *Psychiatry and Clinical Neurosciences*, *54*, 55-58. doi:10.1046/j.1440-1819.2000.00637.x
- Marcus, B., Machilek, F., & Schütz, A. (2006). Personality in cyberspace: Personal web sites as media for personality expressions and impressions. *Journal of Personality and Social Psychology*, *90*, 1014-1031. doi:10.1037/0022-3514.90.6.1014
- McLeod, L., Lewis, C., & Thissen, D. (2003). A bayesian method for the detection of item preknowledge in computerized adaptive testing. *Applied Psychological Measurement*, *27*, 121-137. doi:10.1177/0146621602250534
- Mead, A. D., & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin*, *114*, 449-458. doi:10.1037/0033-2909.114.3.449
- Meade, A. W., Michels, L. C., & Lautenschlager, G. J. (2007). Are Internet and paper-and-pencil personality tests truly comparable? An experimental design measurement invariance study. *Organizational Research Methods*, *10*, 322-345. doi:10.1177/1094428106289393
- Meyerson, P., & Tryon, W. (2003). Validating internet research: a test of the psychometric equivalence of internet and in-person samples. *Behavior Research Methods, Instruments and Computers*, *35*, 614-620.
- Milbradt, A., Zimmerhofer, A., & Hornke, L. F. (2007). *testMaker - A software for web-based assessments* [Computer software]. Aachen: RWTH Aachen University, Lehrstuhl für Betriebs- und Organisationspsychologie. Verfügbar unter <http://www.global-assess.rwth-aachen.de>
- Miller, E., Neal, D., Roberts, L., Baer, J., Cressler, S., Metrik, J., & Marlatt, G. (2002). Test-retest reliability of alcohol measures: Is there a difference between Internet-based assessment and traditional methods? *Psychology of Addictive Behaviors*, *16*, 56-63. doi:10.1037//0893-164X.16.1.56
- Mlacic, B., & Goldberg, L. R. (2007). An analysis of a cross-cultural personality inventory: the IPIP Big-Five factor markers in Croatia. *Journal of Personality Assessment*, *88*, 168-177. doi:10.1080/00223890701267993
- Moser, K. (1989). Muß psychologische Forschung repräsentativ sein? *Schweizerische Zeitschrift für Psychologie*, *48*, 45-49.

- Moser, K., Zempel, J., & Göritz, A. (2003). Personalmarketing im Internet: Ziele, Strategien, Praktiken und Erfahrungen. In U. Konradt & W. Sarges (Hrsg.), *E-Recruitment und E-Assessment* (S. 19-36). Göttingen: Verlag für Angewandte Psychologie.
- Naglieri, J. A., Drasgow, F., Schmit, M., Handler, L., Prifitera, A., Margolis, A., & Velasquez, R. (2004). Psychological testing on the internet: New problems, old issues. *American Psychologist*, *59*, 150-162. doi:10.1037/0003-066X.59.3.150
- New Media Consortium. (2007). *Avatars among us*. Verfügbar unter <http://www.flickr.com/photos/nmc-campus/851866135/>
- Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research and Practice*, *6*, 101-115. doi:10.1037//1089-2699.6.1.101
- Nowson, S., & Oberlander, J. (2006). Differentiating document type and author personality from linguistic features. *Australian Journal of Intelligent Information Processing Systems*, *9*, 84-88.
- Nye, C. D., Do, B., Drasgow, F., & Fine, S. (2008). Two-step testing in employee selection: Is score inflation a problem? *International Journal of Selection and Assessment*, *16*, 112-120. doi:10.1111/j.1468-2389.2008.00416.x
- Ott, R., & Eichenberg, C. (Hrsg.). (2003). *Klinische Psychologie und Internet*. Göttingen: Hogrefe.
- Penke, L., Eichstaedt, J., & Asendorpf, J. B. (2006). Single-attribute implicit association tests (SA-IAT) for the assessment of unipolar constructs: The case of sociosexuality. *Experimental psychology*, *53*, 283-291. doi:10.1027/1618-3169.53.4.283
- Penny, J. A. (2003). Exploring differential item functioning in 360-degree assessment: Rater source and method of delivery. *Organizational Research Methods*, *6*, 61-79. doi:10.1177/1094428102239426
- Pettit, F. A. (2002). A comparison of world-wide-web and paper-and-pencil personality questionnaires. *Behavior Research Methods, Instruments & Computers*, *34*, 50-54.
- Piotrowski, C., & Armstrong, T. (2006). Current recruitment and selection practices: A national survey of Fortune 1000 Firms. *North American Journal of Psychology*, *8*, 489-496.
- Plant, R. R., Hammond, N., & Whitehouse, T. (2003). How choice of mouse my affect response timing in psychological studies. *Behavior Research Methods, Instruments, and Computers*, *35*, 276-284.

- Ployhart, R. E., Weekley, J. A., Holtz, B. C., & Kemp, C. (2003). Web-based and paper-and-pencil testing of applicants in a proctored setting: Are personality, biodata, and situational judgment tests comparable? *Personnel Psychology, 56*, 733-752. doi:10.1111/j.1744-6570.2003.tb00757.x
- Potosky, D., & Bobko, P. (2004). Selection testing via the internet: Practical considerations and exploratory empirical findings. *Personnel Psychology, 57*, 1003-1034. doi:10.1111/j.1744-6570.2004.00013.x
- Preckel, F., & Thiemann, H. (2003). Online- versus paper-pencil-version of a high potential intelligence test. *Swiss Journal of Psychology, 62*, 131-138. doi:10.1024//1421-0185.62.2.131
- Rademacher, J., & Lippke, S. (2007). Dynamic online surveys and experiments with the free open-source software dynQuest. *Behavior Research Methods, 39*, 415-426.
- Rees, C., & Haythornthwaite, S. (2004). Telepsychology and videoconferencing: Issues, opportunities and guidelines for psychologists. *Australian Psychologist, 39*, 212-219. doi:10.1080/00050060412331295108
- Reid, D. J., & Reid, F. J. M. (2005). Online focus groups: An in-depth comparison of computer-mediated and conventional focus group discussion. *International Journal of Market Research, 47*, 131-162.
- Reimers, S., & Stewart, N. (2007). Adobe Flash as a medium for online experimentation: A test of RT measurement capabilities. *Behavior Research Methods, 39*, 365-370.
- Rentfrow, P. J., & Gosling, S. D. (2006). Message in a ballad: The role of music preferences in interpersonal perception. *Psychological Science, 17*, 236-242. doi:10.1111/j.1467-9280.2006.01691.x
- Reynolds, D. & Stiles, W. (2007). Online data collection for psychotherapy process research. *CyberPsychology & Behavior, 10*, 92-99. doi:10.1089/cpb.2006.9987
- Richman, W. L., Kiesler, S., Weisband, S., & Drasgow, F. (1999). A metaanalytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interview. *Journal of Applied Psychology, 84*, 754-775. doi:10.1037/0021-9010.84.5.754
- Risko, E., Quilty, L., & Oakman, J. (2006). Socially desirable responding on the web: Investigating the Candor hypothesis. *Journal of Personality Assessment, 87*, 269-276. doi:10.1207/s15327752jpa8703_08
- Ross, M. W., Mansson, S. A., Cooper, A., & Tikkanen, R. (2005). Biases in internet sexual health samples: Comparison of an internet sexuality survey and a national sexual

- health survey in Sweden. *Social Science & Medicine*, *61*, 245-252.
doi:10.1016/j.socscimed.2005.01.019
- Salgado, J. F., & Moscoso, S. (2003). Internet-based personality testing: Equivalence of measures and assessees' perceptions and reactions. *International Journal of Selection and Assessment*, *11*, 194–205. doi:10.1111/1468-2389.00243
- Sandene, B., Horkay, N., Bennett, R., Allen, N., Braswell, J., Kaplan, B., & Oranje, A. (2005). *Online assessment in mathematics and writing: Reports from the NAEP technology-based assessment project, research and development Series* (Bericht NCES 2005–457). Verfügbar von der Website des National Center for Education Statistics des U.S. Department of Education unter <http://nces.ed.gov/nationsreportcard/pdf/studies/2005457.pdf>
- Schmidt-Atzert, L. (2004). *Objektiver Leistungsmotivations Test (OLMT)* [Software und Manual]. Mödling: Dr. G. Schuhfried GmbH.
- Shermis, M., & Lombard, D. (1998). Effects of computer-based test administrations on test anxiety and performance. *Computers in Human Behavior*, *14*, 111-123.
doi:10.1016/S0747-5632(97)00035-6
- Silverstein, S., Berten, S., Olson, P., Paul, R., Williams, L., Cooper, N., & Gordon, E. (2007). Development and validation of a world-wide-web-based neurocognitive assessment battery: WebNeuro. *Behavior Research Methods*, *39*, 940-949.
- Skitka, L. J., & Sargis, E. (2006). The internet as psychological laboratory. *Annual Review of Psychology*, *57*, 529-555. doi:10.1146/annurev.psych.57.102904.190048
- Snyder, M. (1974). The self-monitoring of expressive behavior. *Journal of Personality and Social Psychology*, *30*, 526-537. doi:10.1037/h0037039
- Sotaridona, L. S., van der Linden, W. J., & Meijer, R. R. (2006). Detecting answer copying using the kappa statistic. *Applied Psychological Measurement*, *30*, 412-431.
doi:10.1177/0146621606288891
- Strohmeier, S. (2002). eHR: Begriff, Konzept und Praxis. *Information Management & Consulting*, *1*, 6-14.
- Testkuratorium (1986). Mitteilung. *Diagnostica*, *32*, 358-360.
- Thompson, L. F., & Surface, E. A. (2007). Employee surveys administered online. *Organizational Research Methods*, *10*, 241-261. doi:10.1177/1094428106/294696
- Tippins, N., Beaty, J., Drasgow, F., Gibson, W., Pearlman, K., Segall, D., & Shepherd, W. (2006). Unproctored internet testing in employment settings. *Personnel Psychology*, *59*, 189-225. doi:10.1111/j.1744-6570.2006.00909.x

- Trajanovic, N. N., Mangan, M., & Shapiro, C. (2007). Sexual behaviour in sleep. An internet survey. *Social Psychiatry and Psychiatric Epidemiology*, *42*, 1024-1031.
doi:10.1007/s00127-007-0258-0
- Turney, L., & Pocknee, C. (2005). Virtual focus groups: New frontiers in research. *International Journal of Qualitative Methods*, *4*, 32-43.
- Vallejo, M. A., Jordán, C. M., Díaz, M., Comeche, M., & Ortega, J. (2007). Psychological assessment via the internet: A reliability and validity study of online (vs paper-and-pencil) versions of the General Health Questionnaire-28 (GHQ-28) and the Symptoms Check-List-90-Revised (SCL-90-R). *Journal of Medical Internet Research*, *9*(1),
doi:10.2196/jmir.9.1.e2
- Varnhagen, C. K., Gushta, M., Daniels, J., Peters, T. C., Parmar, N., Law, D., Hirsch, R., Takach, B. S., & Johnson, T. (2005). How informed is online informed consent? *Ethics & Behavior*, *15*, 37-48. doi:10.1207/s15327019eb1501_3
- Vazire, S., & Gosling, S. D. (2004). e-Perceptions: Personality impressions based on personal websites. *Journal of Personality and Social Psychology*, *87*, 123-132.
doi:10.1037/0022-3514.87.1.123
- van de Vijver, F., & Hasveld, M. (1994). The incomplete equivalence of the paper-and-pencil and computerized versions of the general aptitude test battery. *Journal of Applied Psychology*, *79*, 852-859. doi:10.1037/0021-9010.79.6.852
- Wagner-Menghin, M. (2003). Computerdiagnostik. In K. Kubinger & R. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 68-81). Basel: Beltz.
- Wagner-Menghin, M. (2004). *Lexikon-Wissen-Test (LEWITE)* [Software und Manual]. Mödling: Dr. G. Schuhfried GmbH.
- Wang, Y.-C., Lee, C.-M., Lew-Ting, C.-Y., Hsiao, C. K., Chen, D.-R., & Chen, W. J. (2005). Survey of substance use among high school students in Taipei: Web-based questionnaire versus paper-and-pencil questionnaire. *Journal of Adolescent Health*, *37*, 289-295. doi:10.1016/j.jadohealth.2005.03.017
- Wang, S., Jiao, H., Young, M., Brooks, T., & Olson, J. (2007). A meta-analysis of testing mode effects in grade K-12 mathematics tests. *Educational and Psychological Measurement*, *67*, 219-238. doi:10.1177/0013164406288166
- Wang, S., Jiao, H., Young, M., Brooks, T., & Olson, J. (2008). Comparability of computer-based and paper-pencil testing in K 12 reading assessments: A meta-analysis of testing mode effects. *Educational and Psychological Measurement*, *68*, 5-24.
doi:10.1177/0013164407305592

- Washburn, D. (2003). The games psychologists play (and the data they provide). *Behavior Research Methods, Instruments, & Computers*, 35, 185-193.
- Waters, S., & Pommerich, M. (2007, April). *Context effects in internet testing: A literature review*. Vortrag auf der XXII. Konferenz der Society of Industrial and Organizational Psychology, New York (USA). Verfügbar auf der Website der Human Resources Research Organization unter <http://www.humrro.org>
- Weber, B., Schneider, B., Fritze, J., Gille, B., Hornung, S., Kühner, T., & Maurer, K. (2003). Acceptance of computerized compared to paper-and-pencil assessment in psychiatric inpatients. *Computers in Human Behavior*, 19, 81-93. doi:10.1016/S0747-5632(02)00012-2
- Welker, M. (2007). Was ist Online-Forschung? Eine Tour d'horizon zu einem erfolgreichen Forschungsfeld. In M. Welker & O. Wenzel (Hrsg.), *Online-Forschung 2007 – Grundlagen und Fallstudien* (S. 19-51). Köln: Halem Verlag.
- Wenzel, O., & Hoffmann, O. (2005). 10 Jahre Online-Forschung: Von Visionen, dem Boden der Tatsachen und letztlich doch erfüllten Erwartungen. *Planung & Analyse*, 1, 24-28.
- Wiechmann, D., & Ryan, A. M. (2003). Reactions to computerized testing in selection contexts. *International Journal of Selection and Assessment*, 11, 215-229. doi:10.1111/1468-2389.00245
- Williams, J. E., & McCord, D. M. (2006). Equivalence of standard and computerized versions of the Raven Progressive Matrices Test. *Computers in Human Behavior*, 22, 791-800. doi:10.1016/j.chb.2004.03.005
- Wogalter, M. S., Howe, J. E., Sifuentes, A. H., & Luginbuhl, J. (1999). On the adequacy of legal documents: Factors that influence informed consent. *Ergonomics*, 42, 593-613. doi:10.1080/001401399185504
- Wottawa, H., & Woike, J. (2002). Internet-Recruiting und –Assessment. *Wirtschaftspsychologie*, 1, 33-38.
- Ziegler, M., Schmidt-Atzert, L., Bühner, M., & Krumm, S. (2007). Fakability of different measurement methods for achievement motivation: Questionnaire, semi-projective, and objective. *Psychology Science*, 49, 291-307.
- Zimmerhofer, A., Heukamp, V., & Hornke, L. F. (2006). Ein Schritt zur fundierten Studienfachwahl - Webbasierte Self-Assessments in der Praxis. *Report Psychologie*, 31, 62-72.