

# Testgütekriterien im Online-Testing im beruflichen Kontext

04.09.13

Verfasser: Dr. Olaf Schröder

## Inhalt

1 Hintergrund.....	2
2 Definition und Arten von Tests .....	3
3 Relevanz von Testgütekriterien .....	4
4 Beschreibung der Testgütekriterien.....	6
4.1 Hauptgütekriterien .....	6
4.2 Nebengütekriterien .....	9
5 Adaptives Testen.....	10
6 Fazit .....	11
Literatur .....	<b>Fehler! Textmarke nicht definiert.</b>

## 1 Hintergrund

Die Nutzung internetbasierter Dienste in Unternehmen ist in nicht einmal 20 Jahren von einer technologischen Innovation zur alltäglichen Selbstverständlichkeit geworden. Internetbasierte Dienste sind heute integrierter Bestandteil aller unternehmerischen Geschäftsbereiche, wie Marketing, Beschaffung, Supply Chain Management oder elektronischem Handel. Auch aus dem Personalwesen sind internetbasierte Dienste nicht mehr wegzudenken. Für diese Form der Nutzung wurde der Begriff des electronic Human Resource Management (eHRM) geprägt<sup>1</sup>. Das Personalwesen kann grob gegliedert werden in Gestaltungs- und Führungsaspekte, wobei die Gestaltungsaspekte neben der Beurteilung und Vergütung die Bedarfsplanung, Gewinnung und Entwicklung von Mitarbeitern umfasst<sup>2</sup>. Internet-Dienste werden auch im Rahmen der Gewinnung und Auswahl von Mitarbeitern (eRecruiting, eSelection<sup>3</sup>) und der Mitarbeiterentwicklung (z.B. eTraining und eLearning<sup>4</sup>, auch eCoaching<sup>5</sup>) eingesetzt.

In diesen Anwendungsfeldern spielen Online-Tests eine immer wichtiger werdende Rolle:

- Im eRecruiting werden sie unter anderem als Self Assessment Tools eingesetzt: Bewerber haben damit im Vorfeld einer Bewerbung die Möglichkeit, den Grad der Übereinstimmung ihrer Qualifikationen und Interessen mit einem z.B. in einer Stellenausschreibung gegebenen Anforderungsprofil zu überprüfen. Die hierdurch mögliche informierte Selbstselektion von Bewerbern führt von vornherein zu besser auf eine Ausschreibung „passenden“ Bewerbungen<sup>6</sup>.
- In der eSelection können Unternehmen den Auswahlprozess durch den Einsatz psychologischer Tests objektiver und transparenter gestalten<sup>7</sup>, zudem prüfen Tests meist nicht nur aktuelles Verhalten bzw. aktuelle Fähigkeiten, sondern auch

---

<sup>1</sup> Vgl. Strohmeier (2007), S. 19

<sup>2</sup> Vgl. Stock-Homburg (2008), S. 16

<sup>3</sup> Vgl. Strohmeier (2007), S. 25

<sup>4</sup> Vgl. Dittler (2003). S. 9f.

<sup>5</sup> Vgl. Geissler (2008), S. 3f.

<sup>6</sup> Vgl. Kirbach / Wattawa (2009), S. 71

<sup>7</sup> Vgl. Hülshager / Maier (2008), S. 108f.

das Potenzial eines Bewerbers, also künftige Leistungsmöglichkeiten<sup>8</sup>. Die Potenzialanalyse ist gerade in einem sich rasch ändernden beruflichen Umfeld von erheblicher Bedeutung. Des Weiteren wird durch Online-Bereitstellung dieser Tests der Auswahlprozess deutlich vereinfacht<sup>9</sup>.

- Im eLearning und eTraining schließlich, also in online bereitgestellten Lernmaterialien, spielen Tests und Assessments ebenfalls eine wichtige Rolle: Als Instrumente der Selbstevaluation oder der Fremdevaluation unterstützen Online-Tests bzw. E-Assessments<sup>10</sup> die Steuerung des Qualifikationsprozesses. Beispielsweise haben Mitarbeiter die Möglichkeit, mit Online-Tests ihr Wissen zu überprüfen und somit ihren Lernfortschritt zu verfolgen und den weiteren Lernprozess zu planen.

## 2 Definition und Arten von Tests

Ein (psychometrischer) Test wird definiert als ein Verfahren zur Messung eines oder mehrerer abstrakter (nicht direkt sinnlich erfahrbare) Merkmale einer Person mit dem Ziel, eine möglichst quantitative Aussage über die Ausprägung des Merkmals bei dieser Person machen zu können<sup>11</sup>.

Tests (und ebenso Online-Tests bzw. allgemeiner technologiebasierte Tests<sup>12</sup>) beziehen sich meist auf ein relativ abstraktes Merkmal oder eine Eigenschaft, der beruflich relevantes Potenzial zugesprochen wird<sup>13</sup>, und die nicht direkt beobachtbar ist. Vielmehr muss auf diese Eigenschaft mittels der Testergebnisse mit größtmöglicher Sicherheit geschlossen werden können. Beispiele dieser zu erschließenden Merkmale sind etwa „Intelligenz“, „Teamfähigkeit“ oder „räumliches Vorstellungsvermögen“. Die durch Tests gemessenen Merkmale können sich auf sehr unterschiedliche Gegenstandsbereiche beziehen. Sehr grob kann zwischen Tests der allgemeinen kogni-

---

<sup>8</sup> Vgl. Schuler (2009), S. 122

<sup>9</sup> Vgl. Geister / Rastetter (2009), S. 11

<sup>10</sup> Vgl. Hornecke / Amelung / Krieger / Rösner (2011), S. 127f.

<sup>11</sup> Vgl. Bühner (2006), S. 23

<sup>12</sup> Jurecka / Hartig (2007, S. 41) unterscheiden zwischen computerbasiertem Assessment (das auch offline am PC durchgeführt werden kann), netzwerkbasierendem Assessment (etwa in einem Local Area Network) und internetbasiertes Assessment. „Online-Testing“ meint in der Regel internetbasiertes Assessment.

<sup>13</sup> Vgl. Schuler (2009), S. 131

tiven Leistungsfähigkeit, wie beispielsweise Intelligenztests, Tests spezifischer kognitiver Fähigkeiten (wie beispielsweise dem räumlichen Vorstellungsvermögen) einschließlich Wissens- und Kenntnistests sowie Tests nicht-kognitiver Merkmale (Persönlichkeits-, Interessen- und Motivationstests) unterschieden werden<sup>14</sup>.

Des Weiteren werden Tests nach ihrem Format unterschieden. Geschlossene bzw. standardisierte Formate bestehen aus Fragen bzw. Aufgaben mit vorgegebenen Antworten<sup>15</sup>. Die Person antwortet durch Auswahl aus einer Menge vorgegebener Alternativen. Bei offenen Antwortformaten hingegen kann die Person ihre Antworten selbständig formulieren. Psychometrische Tests im engeren Sinne haben geschlossene Antwortformate, weil, wie später noch sichtbar wird, bei offenen Antwortformaten nicht alle Testgütekriterien erfüllt werden können.

### 3 Relevanz von Testgütekriterien

Testgütekriterien sind Instrumente zur Qualitätsbeurteilung von Tests<sup>16</sup>. Gültigkeit (Validität), Zuverlässigkeit (Reliabilität) und Objektivität gelten dabei als die drei Hauptgütekriterien von Tests<sup>17</sup>. Darüber hinaus gibt es noch Nebengütekriterien, auf die später eingegangen wird.

Sowohl zur Ermittlung kognitiver Merkmale als auch von Persönlichkeitsmerkmalen gibt es eine Reihe bewährter, empirisch fundierter Testverfahren, deren Gültigkeit auch für das berufliche Umfeld großenteils empirisch belegt worden ist<sup>18</sup>. In der Praxis werden allerdings häufig auch selbst entwickelte Tests eingesetzt<sup>19</sup>, über deren Gültigkeit und Zuverlässigkeit meist keine näheren Informationen vorliegen. In besonderem Maße trifft dies für Tests und Assessments zu, die zur Wissens- und Fähigkeitsüberprüfung im Rahmen von E-Training-Maßnahmen eingesetzt werden.

Ein Verzicht auf die Testgütekriterien der Validität und Reliabilität bzw. ein Einsatz von Tests und Assessments, für die keine Testgütekriterien erhoben wurden und dementsprechend nicht bekannt sind, hat allerdings zur Folge, dass ein wesentlicher

---

<sup>14</sup> Vgl. Hülshager / Maier (2008), S. 109f.

<sup>15</sup> Vgl. Atteslander (2006), S. 136, vgl. Raab-Steiner / Benesch (2008), S. 48

<sup>16</sup> Vgl. Moosbrugger / Kelava (2011), S. 8.

<sup>17</sup> Vgl. Bühner (2006), S. 33f., vgl. Moosbrugger / Kelava (2011), S. 8.

<sup>18</sup> Vgl. Hülshager / Maier (2008), S. 109f.

<sup>19</sup> Vgl. Schuler / Höft (2006), S. 105

Vorteil der Verwendung psychologischer Tests verloren geht. Zwar sind auch diese Tests oft noch relativ objektiv<sup>20</sup> in dem Sinne, dass Testleistungen verschiedener Personen vergleichbar sind, weil alle Personen denselben Test bearbeiten und die Durchführung und Auswertung des Tests sowie die Interpretation des Testergebnisses weitgehend von der Leistung bzw. den Antworten der Person abhängen, die sich dem Test unterzogen hat. Damit weisen sie selbst beim Fehlen von Testgütekriterien noch einen Vorteil gegenüber beispielsweise Einstellungsgesprächen oder Interviews auf, bei denen die Einschätzung des Bewerbers auch von der jeweiligen Person abhängt, die das Einstellungsgespräch führt. Aber wenn zur Gültigkeit und Zuverlässigkeit eines Tests keine näheren Informationen vorliegen, dann lassen sich (außer dem „Augenschein“) keine näheren Angaben darüber machen, was der Test mit dem zu prüfenden Merkmal zu tun hat. Ob beispielsweise ein Test zur Erfassung der räumlichen Auffassungsgabe, dessen Gültigkeit nicht untersucht wurde, wirklich die räumliche Auffassungsgabe misst, oder etwa die Fähigkeit, Produktwissen im Kundengespräch zielführend einzusetzen, darüber können dann nur Vermutungen angestellt werden. Fehlen darüber hinaus Angaben über die Zuverlässigkeit des Tests, so lassen sich keine Angaben darüber machen, ob die erzielten Ergebnisse bei Wiederholung des Tests in ähnlicher Weise ausfallen würden. Das Testergebnis wäre dann ein Zufallsergebnis und wenig aussagekräftig.

Die Ermittlung der Testgütekriterien „Gültigkeit“ (Validität) und „Zuverlässigkeit“ (Reliabilität) ist also notwendig, um präzise Angaben über den diagnostischen Wert des Tests für eine bestimmte Fragestellung wie z.B. die Ermittlung bestimmter Fähigkeiten und Kompetenzen machen zu können. Aber auch weitere Testgütekriterien sind wichtig, wenn der Vorteil des diagnostischen Instruments „Test“ zur Geltung kommen soll. Im Folgenden werden daher die Testgütekriterien näher vorgestellt, und es wird beschrieben, wie diese Gütekriterien bei Online-Tests erreicht werden können.

---

<sup>20</sup> Vgl. Moosbrugger / Kelava (2011), S. 8

## 4 Beschreibung der Testgütekriterien

### 4.1 Hauptgütekriterien

*Objektivität:* Ein Test ist objektiv, wenn seine Durchführung, Auswertung und Interpretation nicht von einer bestimmten Person (z.B. dem bzw. den unternehmensseitig an einem Auswahlprozess Beteiligten) abhängt. Es wird unterschieden zwischen Durchführungs-, Auswertungs- und Interpretationsobjektivität<sup>21</sup>.

- Durchführungsobjektivität liegt vor, wenn das Testergebnis nicht von einer bestimmten Person abhängt, die den Test durchführt. Beim Online-Test ist dieses Kriterium normalerweise erfüllt. Allerdings kann die Durchführungsobjektivität auch hier beeinträchtigt werden. Wird der Test im einstellenden Unternehmen durchgeführt, können Lärm und sonstige Störungen oder auch Reaktionen auf eventuelle Rückfragen von Bewerber zu Bewerber unterschiedlich sein.
- Auswertungsobjektivität bedeutet, dass die Auswertung des Tests unabhängig von einer auswertenden Person ist. Auch dies ist bei Online-Tests meist der Fall. Ein Problem stellt die Auswertungsobjektivität eher bei Tests mit offenen Antwortformaten dar, weil hier die Auswertung z.B. von der Erfahrung der Auswerter bei der Bildung von Kategorien für die freien Äußerungen der Person abhängt<sup>22</sup>.
- Interpretationsobjektivität liegt vor, wenn zwei Personen aus einem Testergebnis dieselben Schlussfolgerungen ziehen<sup>23</sup>. Im berufsbezogenen Kontext sind hier zwei Ebenen zu unterscheiden: Die „nahe“ Interpretation eines Intelligenzwerts oder eines Neurotizismuswerts kann anhand von Normtabellen geschehen. Im Kontext des eRecruiting bzw. der eSelection interessiert jedoch die Frage, wie z.B. ein bestimmter Intelligenzwert im Hinblick auf das Anforderungsprofil einer Stelle zu werten ist, und wie wichtig dieser Wert in Relation zu anderen diagnostischen Ergebnissen ist. Bezüglich dieser Fragen dürfte die Objektivität eher gering sein.

---

<sup>21</sup> Vgl. Bühner (2006), S. 23, vgl. Moosbrugger / Kelava (2011), S. 8f.

<sup>22</sup> Vgl. Raab-Steiner / Benesch (2008), S. 48

<sup>23</sup> vgl. Moosbrugger / Kelava (2011), S. 10

Insgesamt wird Online-Test eine vergleichsweise hohe Objektivität zugesprochen<sup>24</sup>, weil bei Testdurchführung und Auswertung kein menschlicher Testleiter anwesend sein muss, der die Objektivität durch jeweils unterschiedliches Verhalten beeinträchtigen könnte. Testleiter- und Situationseffekte werden minimiert<sup>25</sup>.

*Reliabilität:* Reliabilität bzw. Zuverlässigkeit betrifft die Messgenauigkeit eines Tests. Je geringer die Messfehler sind, desto größer die Reliabilität<sup>26</sup>. Sie wird durch den Reliabilitätskoeffizienten erfasst und liegt zwischen 0 (keine Reliabilität) und 1 (perfekte Reliabilität). Der Reliabilitätskoeffizient eines Tests sollte nicht unter 0,7 liegen. Die Reliabilität wird oft als interne Konsistenz gemessen, das am meisten verbreitete Maß ist Cronbach's Alpha-Koeffizient<sup>27</sup>. Bei Online-Tests kann sich eine Reliabilitätsproblematik vor allem beim Wechsel von der „Papier-und-Bleistift“-Version zur Online-Version ergeben, da durch den Wechsel des Mediums Einflüsse auf die Messgenauigkeit möglich sind<sup>28</sup>. Solche Unterschiede wurden empirisch zum Teil durchaus gefunden, sie scheinen aber eher gering und verringerten sich zudem, wenn die Bedingungen der Testvorgabe bei der Online-Version der Papier-und-Bleistift-Version möglichst ähneln (z.B. Schaffung der Möglichkeit des Zurückblätterns in der Online-Version bzw. PC-Version analog zum Zurückblättern im Papier-und-Bleistift-Test)<sup>29</sup>. Auch gehören Fehlerquellen aus der frühen Zeit computergestützter Tests, die etwa mit einer geringen Ergonomie der Testsoftware und auch der Hardware sowie mit geringer Vertrautheit mancher Nutzer mit Maus und Tastatur einhergingen, heute weitgehend der Vergangenheit an, so dass insgesamt eher von einer Verbesserung der Reliabilität durch Online-Tests ausgegangen werden kann<sup>30</sup>.

*Validität:* Die Validität ist das Ausmaß, in dem ein Test das misst, was er zu messen vorgibt<sup>31</sup>. Grundlegend wird dabei zwischen Inhaltsvalidität, Kriteriumsvalidität und Konstruktvalidität unterschieden.

---

<sup>24</sup> Vgl. Schaper (2009), S. 26

<sup>25</sup> Vgl. Jurecka / Hartig (2007), S. 44

<sup>26</sup> In der klassischen Testtheorie entspricht dies einem hohen Zusammenhang zwischen beobachteten Werten und wahren Werten, vgl. Schnell / Hill / Esser (2005), S. 151.

<sup>27</sup> Vgl. Schnell / Hill / Esser (2005), S. 153

<sup>28</sup> Vgl. Schaper (2009), S. 23-24

<sup>29</sup> Vgl. Schaper (2009), S. 24

<sup>30</sup> Vgl. Ridgeway / McCusker / Pead (2004), S. 21f.

<sup>31</sup> Vgl. Bühner (2006), S. 36, vgl. Moosbrugger / Kelava (2011), S. 13f,

- Inhaltsvalidität (logische bzw. fachliche Validität sowie – hiervon unterschieden<sup>32</sup> - „Augenschein“-Validität) ergibt sich aus logischen bzw. theoretischen und fachlichen Überlegungen. Sie kann nicht numerisch zum Ausdruck gebracht werden und wird daher oftmals nicht genügend ernst genommen<sup>33</sup>. Die theoretische Fundierung eines Tests ist jedoch wichtig.
- Kriteriumsvalidität bezeichnet den Zusammenhang eines Tests mit einem Kriterium, mit dem er korrelieren sollte (z.B. Extraversion und Erfolg als Trainer und Moderator). Kriteriumsvalidität wird also numerisch als Korrelation ausgedrückt. Es werden Vorhersagevalidität, Übereinstimmungsvalidität, retrospektive Validität und inkrementelle Validität unterschieden; letztere spielt gerade im beruflichen Auswahlprozess, aber auch bei der Selbstselektion eine wichtige Rolle, weil der Test in der Regel nicht alleiniges diagnostisches Instrument darstellt, so dass sich oftmals die Frage stellt, welcher zusätzliche Erkenntnisgewinn mit einem Test neben z.B. einem Vorstellungsgespräch verbunden ist.
- Konstruktvalidität bezeichnet Zusammenhänge des Tests mit anderen Variablen, die theoretisch abgeleitet wurden. Eine hohe Konstruktvalidität liegt also vor, wenn theoriekonforme Zusammenhänge zu anderen Tests (bzw. allgemein zu anderen Variablen) bestehen<sup>34</sup>. Es werden konvergente und divergente (diskriminante) Validität unterschieden. Im letzten Fall muss der Test, um theoriekonform und damit konstruktvalide zu sein, gerade eine geringe Übereinstimmung mit bestimmten anderen Variablen aufweisen. Eine Methode zur Bestimmung der Konstruktvalidität ist die faktorielle Validität, bei der mit Hilfe der Faktorenanalyse Zusammenhänge mit anderen Tests, insbesondere gemeinsame und unterschiedliche Konstrukte, identifiziert werden.

Da die Objektivität und die Reliabilität bei Online-Tests tendenziell positiv zu werten sind und die Reliabilität zweier Messungen für die (Kriteriums-) Validität eine Obergrenze darstellt<sup>35</sup>, wird auch die Validität positiv beeinflusst. Empirische Studien haben überwiegend gezeigt, dass die Validität von Online-Tests der entsprechenden

---

<sup>32</sup> vgl. Moosbrugger / Kelava (2011), S. 15

<sup>33</sup> Vgl. Bühner (2006), S. 37

<sup>34</sup> vgl. Moosbrugger / Kelava (2011), S. 17

<sup>35</sup> Vgl. Bühner (2006), S. 42-43



Papier-und-Bleistift-Version nicht unterlegen ist<sup>36</sup>. Generell scheint zu gelten: Je komplexer die Testaufgaben werden, desto geringer scheinen die Einflüsse des Testdarbietungsmediums zu sein.

#### **4.2 Nebengütekriterien**

Zu den Nebengütekriterien psychologischer Tests zählen Normierung, Vergleichbarkeit, Ökonomie und Nützlichkeit<sup>37</sup>.

Normierung bezieht sich auf ein Bezugssystem, vor diesen Hintergrund ein Testwert interpretiert werden kann. Beispielsweise kann ein Intelligenzwert von 110 durch den Vergleich mit einer Norm als „leicht überdurchschnittlich“ gewertet werden. Normen liegen in der Regel für verschiedene Teilstichproben vor.

Vergleichbarkeit ist gegeben, wenn es zu einem Test entweder einen anderen Test mit sehr ähnlichem Gültigkeitsbereich gibt (z.B. zwei verschiedene Tests zur Messung der intrinsischen Arbeitsmotivation), oder wenn es zu einem Test eine oder mehrere Parallelförmigkeiten gibt. Parallelförmigkeiten haben den Vorteil, dass verfälschende Lerneffekte bei Testwiederholung verringert werden. Gerade bei wiederholten Bewerbungen sind solche Effekte leicht möglich.

Die Ökonomie eines Tests wird erhöht, wenn dieser eine kurze Durchführungszeit beansprucht, wenig Material verbraucht, einfach zu handhaben ist und schnell und einfach auszuwerten ist. Gerade Online-Tests sind aus Unternehmenssicht in dieser Hinsicht attraktiv, denn durch automatisierte Auswertung und Rückmeldung werden Arbeitsaufwände reduziert; zudem kann der Bewerber die Ergebnisse, wenn sie automatisch ausgewertet und online bereitgestellt werden, sofort abrufen<sup>38</sup>.

Die Nützlichkeit<sup>39</sup> schließlich ist das am schwierigsten zu messende und zu beurteilende Kriterium. Im Zusammenhang mit der Bewerberauswahl in Unternehmen sollte genau begründet werden, warum welcher Test ausgewählt wird, d.h. welches Merkmal er misst, und welche Bedeutung dieses Merkmal für die berufliche Anforderung hat. Kann eine solche Begründungskette nicht klar und schlüssig gegeben werden,

---

<sup>36</sup> Vgl. Konradt / Lehmann / Böhm-Rupprecht / Hertel (2003), S. 107f.

<sup>37</sup> Vgl. Bühner (2006), S. 43

<sup>38</sup> Vgl. Schaper (2009), S. 27

<sup>39</sup> Vgl. Bühner (2006), S. 44

so sollte auf den entsprechenden Test besser verzichtet werden. Lassen sich aber umgekehrt wünschenswerte Bewerbermerkmale ableiten, für die valide und reliable Tests bereitstehen, dann sollte diese im Interesse einer möglichst treffsicheren Bewerberauswahl auch genutzt werden.

Neben den genannten können noch weitere Gütekriterien angegeben werden, so die Zumutbarkeit, die Unverfälschbarkeit und die Fairness. Zumutbarkeit heißt, dass der Nutzen des Tests in einem angemessenen Verhältnis zur Belastung der Person stehen muss. Die Belastung muss sich nicht nur auf den Aufwand beziehen – dieser wird durch einen Online-Test tendenziell reduziert, wie schon ausgeführt –, sondern sie kann auch z.B. durch als zu persönliche empfundene Fragen, unverständliche Instruktionen oder Merkmale der Testdurchführung (z.B. weite Anreise zu einem Testcenter) beeinträchtigt werden. Unverfälschbarkeit bezieht sich vor allem auf Vermeidung der „sozialen Erwünschtheit“, d.h. der Test wird weniger leicht verfälschbar, wenn seine Intention nicht leicht zu durchschauen ist. Darunter kann allerdings wiederum die „Augenscheinvalidität“ leiden. Testfairness schließlich bezieht sich darauf, dass ein Test nicht zu einer systematischen Benachteiligung bestimmter Personengruppen führen soll. Dies ist der Fall, wenn eine Aufgabe für bestimmte Bevölkerungsgruppen unterschiedlich schwierig ist. Das Thema Fairness wird in der Intelligenzdiagnostik diskutiert. Auch in der Anfangszeit des computergestützten Testens war Fairness ein Thema, da es bezüglich der Vertrautheit des Umgangs mit dem Computer noch beträchtliche Unterschiede gab. Zumindest dieser letztgenannte Punkt spielt heute allerdings keine nennenswerte Rolle mehr.

## 5 Adaptives Testen

Normalerweise sind die Items eines Tests sowie ihre Reihenfolge vorab festgelegt. Dies führt allerdings dann zu Ineffizienz, wenn eine Person viele Aufgaben lösen bzw. nicht lösen kann: Wer auch schwierige Aufgaben löst, für den ist die Vorgabe leichter Aufgaben nicht informativ, weil er diese mit sehr großer Wahrscheinlichkeit ebenfalls lösen wird. Analoge gilt für eine Person, die bereits leichte Aufgaben nicht lösen kann: Hier ist die Vorgabe schwieriger Aufgaben nicht informativ. Ließen sich die Aufgaben an die Leistungsfähigkeit der jeweiligen Person anpassen, so könnte

die Aufgabenvorgabe auf lediglich informative Aufgaben beschränkt werden und somit der Test deutlich ökonomischer gestaltet werden.

So gestaltete Tests werden als adaptive Tests bezeichnet. Die Information, die ein Item für eine bestimmte Testperson liefert, wird dabei dynamisch (zur Testbearbeitungszeit) aus der Wahrscheinlichkeit bestimmt, dass die Person dieses Item lösen wird<sup>40</sup>. Diese Wahrscheinlichkeit wiederum ist eine Funktion der Lösungen der im Test bereits vorangegangenen Aufgaben. Somit kann das jeweils informativste nächste Item ausgewählt werden. Adaptive Tests folgen somit den theoretischen Prinzipien der probabilistischen Testtheorie. Adaptive Tests können sinnvollerweise nur computergestützt erfolgen.

Bezüglich der Testgütekriterien sind die adaptiven Tests mit nicht-adaptiven Tests (mindestens) vergleichbar. Dies haben empirische Untersuchungen gezeigt. Wird der durch die Adaptivität erzielbare Effizienzvorteil nicht zur Reduktion der Items, sondern zur Vorgabe weiterer möglichst informativer Items genutzt, so steigt die (konvergente) Validität bei adaptiven Tests tendenziell höher als bei nichtadaptiven Tests<sup>41</sup>. Allerdings ist der Aufwand zur Erstellung adaptiver Tests meist beträchtlich, weil die Schwierigkeiten (mittlere Lösungswahrscheinlichkeiten) der Aufgaben vorher bekannt sein müssen, also empirisch ermittelt werden müssen.

## 6 Fazit

Testgütekriterien sind bei der Personalauswahl von großer Bedeutung. Insbesondere die Hauptkriterien der Objektivität, der Reliabilität und der Validität erlauben Angaben darüber, inwieweit ein Testergebnis wirklich möglichst ausschließlich von den Merkmalen der Person abhängt, wie zuverlässig es erfasst wird und wie es mit den für eine bestimmte Stellenausschreibung relevanten Merkmalen zusammenhängt. Empirisch überprüfte psychologische Tests weisen gegenüber den meisten anderen Methoden des Recruitments und der Personalauswahl deutlich bessere Gütekriterien auf. Dies gilt umso mehr für Online-Tests, da hier meist noch weitere Fehlerquellen ausgeschaltet werden können. Online-Tests bieten darüber hinaus die – bis heute

---

<sup>40</sup> Vgl. Kubinger (1996), S. 568

<sup>41</sup> Vgl. Frey (2007), S. 287-288

noch nicht sehr intensiv genutzte – Möglichkeit zur Formulierung innovativer, interaktiver Items, die dem konkreten beruflichen Handeln näher angeglichen werden können. Damit verfügt das Online-Testing über das Potenzial zur noch weiteren Erhöhung der Validität. Dies gilt auch für das adaptive Testen, welches gerade vor dem Hintergrund der Testgütekriterien in Zukunft eine noch wichtigere Rolle spielen dürfte.

## Literatur

- Atteslander, Peter (2006): Methoden der empirischen Sozialforschung. 11. Auflage. Berlin: Erich Schmidt Verlag
- Dittler, Ulrich (2003): Einführung – E-Learning in der betrieblichen Aus- und Weiter-bildung. In Dittler, Ulrich (2003): E-Learning. 2. Auflage. München: Oldenbourg Ver-lag, S. 3 – 28
- Frey, Andreas (2011): Adaptives Testen. In Moosbrugger, Helfried / Kelava, Augustin (Hrsg): Testtheorie und Fragebogenkonstruktion. Berlin: Springer Verlag, 2. Auflage. S. 275 - 294
- Geissler, Harald (2008): E-Coaching – eine konzeptionelle Grundlegung. In Geißler, Harald (Hrsg): E-Coaching. Baltmannsweiler: Schneider Verlag Hohengehren, S. 3 – 23
- Geister, Susanne / Rastetter, Daniela (2009): Aktueller Stand zum Thema Online-Tests. In Steiner, Heinke (Hrsg): Online-Assessment. Heidelberg: Springer Verlag, S. 3 – 16
- Hornecke, Jens / Amelung, Mario / Krieger, Katrin / Rösner, Dietmar (2011): Flexibles E-Assessment mit OLAT und ECSpooler. In Rohland, Holger / Kienle, Andrea / Friedrich, Steffen (Hrsg): DeLFI 2011 – Die 9. E-Learning Fachtagung Informatik der Gesellschaft für Informatik e.V. 5. – 8. September 2011, Dresden. Proceedings. Bonn: Gesellschaft für Informatik, S. 127 - 138
- Hülshager, Ute / Maier, Günter (2008): Persönlichkeitseigenschaften, Intelligenz und Erfolg im Beruf. Psychologische Rundschau, 59 (2), S. 108 – 122
- Jurecka, Astrid / Hartig, Johannes (2007): Computer- und netzwerkbasierendes Assessment. In Hartig, Johannes / Klieme, Eckhard (Hrsg): Möglichkeiten und Vorausset-zungen technologiebasierter Kompetenzdiagnostik. Expertise im Auftrag des Bun-desministeriums für Bildung und Forschung. S. 37 – 48 [http://www.bmbf.de/pub/band\\_zwanzig\\_bildungsforschung.pdf](http://www.bmbf.de/pub/band_zwanzig_bildungsforschung.pdf)
- Konradt, Udo / Lehmann, Katrin / Böhm-Rupprecht, Jolanta / Hertel, Guido (2003): Computer- und internetbasierte Verfahren der Berufseignungsdiagnostik: Ein empirischer Überblick. In Konradt, Udo / Sarges, Werner (Hrsg): E-Recruitment und E-Assessment. Göttingen: Hogrefe Verlag, S. 105 – 124
- Kubinger, Klaus D. (1996): Methoden der Psychologischen Diagnostik. In Erdfelder, Edgar / Mausfeld, Rainer / Meiser, Thorsten / Rudinger, Georg (Hrsg): Handbuch quantitative Methoden. Weinheim: Beltz Verlag, S. 567 - 576
- Moosbrugger, Helfried / Kelava, Augustin (2011): Qualitätsanforderungen an einen psychologischen Test (Testgütekriterien). In Moosbrugger, Helfried / Kelava, Augustin (Hrsg): Testtheorie und Fragebogenkonstruktion. Berlin: Springer Verlag, 2. Auflage. S. 7 – 26
- Raab-Steiner, Elisabeth / Benesch, Michael (2008): Der Fragebogen. Von der For-schungsidee zur SPSS-Auswertung. Wien: Facultas Verlags- und Buchhandels AG
- Ridgeway, Jim / McCusker, Sean / Pead, Daniel (2004): Literature Review of E-Assessment. Bristol: Futurelab. [http://dro.dur.ac.uk/1929/1/Ridgeway\\_Literature.pdf](http://dro.dur.ac.uk/1929/1/Ridgeway_Literature.pdf)

Schaper, Niclas (2009): Online-Tests aus diagnostisch-methodischer Sicht. In Steiner, Heinke (Hrsg): Online-Assessment. Heidelberg: Springer Verlag, S. 17 - 36

Schnell, Rainer / Hill, Paul / Esser, Elke (2005): Methoden der empirischen Sozialforschung. 7. Auflage. München: Oldenbourg Verlag

Schuler, Heinz (2009): Auswahl von Mitarbeitern. In von Rosenstiel, Lutz / Regnet, Erika / Domsch, Michael (Hrsg): Führung von Mitarbeitern. 6. Auflage. Stuttgart: Schäffer-Poeschel Verlag, S. 115 – 147

Schuler, Heinz / Höft, Stefan (2006): Konstruktorientierte Verfahren der Personal-auswahl. In Schuler, Heinz (Hrsg): Lehrbuch der Personalpsychologie. 2. Auflage. Göttingen: Hogrefe, S. 101 – 144

Stock-Homburg, Ruth (2008): Personalmanagement. Theorien – Konzepte - Instrumente. Wiesbaden: Gabler Verlag

Strohmeier, Stefan (2007): Research in e-HRM: Review and Implications. Human Resource Management Review, 17, S. 19 - 37