

# Verifizierung von Online-Tests mit kurzen Vor-Ort-Tests: Eine kritische Betrachtung

Anna-Lena Jobmann, Christian Reiß



## Gründe für zweistufiges Testen und Vorgehensweisen

Die Feststellung kognitiver Eignung mit Hilfe von psychologischen Testverfahren ist eine der wichtigsten Methoden der Personalauswahl. Nach wie vor haben kognitive Eignungstests auch in Deutschland eine hohe Vorhersagekraft für späteren Ausbildungs- und Berufserfolg (Goertz, Hülshager & Maier, 2014; Hülshager, Maier & Stumpp, 2007; Stadelmaier, Thiem & Kleinmanns, 2018). Dennoch werden sie immer noch verhältnismäßig weniger häufig eingesetzt als andere, weniger aussagekräftige Verfahren und selbst in ihrer Aussagekraft nicht selten unterschätzt (Benit & Soellner, 2013; Schuler, Hell, Trapmann, Schaar & Boramir, 2007). Neben der Vorhersagekraft überzeugen Eignungstests insbesondere dadurch, dass im Personalauswahlprozess größere Bewerbergruppen frühzeitig fair und objektiv selektiert werden und so Zeit und Kosten gespart werden können.

Das computerbasierte und vor allem Online-Testen hat in den letzten Jahren für die Praxis der Personalauswahl zunehmend an Relevanz gewonnen (Geister & Rastetter, 2008; Tippins, 2015). Vor allem Online-Tests, die an eigenen Geräten und in nicht kontrollierten Situationen bearbeitet werden, bieten aus wirtschaftlicher Sicht große Vorteile: Die/der Bewerber\*in muss nicht eigens für den Eignungstest anreisen und es fallen keine Kosten für Raum, Hardware, Organisation und Testanweisung an. Das geht andererseits auf Kosten der Objektivität, Genauigkeit sowie der Schlussfolgerungen, die aus dem Testergebnis gezogen werden können. Insbesondere durch die Gefahr einer Verfälschung durch die Unterstützung anderer Personen oder nicht erlaubter Hilfsmittel bei der Testbearbeitung lässt das Ergebnis weniger sichere Aussagen über die Fähigkeiten der Teilnehmer\*innen zu als vergleichbare Vor-Ort-Tests (Tippins, Beaty, Drsgow, Gibson, Pearlman, Segall & Spepherd, 2006).

---

## Täuschungsverhalten in Online-Tests ist keine Seltenheit

---

Eine Meta-Analyse mit 49 Studien von Steger, Schroeders & Gnambs (2018) zeigt beispielsweise, dass die Leistungen in Online-Tests höher liegen als bei Vor-Ort-Tests. Insbesondere ist die Testleistung in Online-Tests höher als in Vor-Ort-Tests, wenn diese Aufgaben enthalten, zu denen gut recherchiert werden kann (z.B. Kenntnisse, sprachliche Fähigkeiten). Dieses Ergebnis unterstützt die Vermutung, dass Täuschungsverhalten in Online-Tests keine Seltenheit ist.

Studien zu Täuschungsraten präsentieren ganz unterschiedliche Ergebnisse. Die Raten variieren zwischen 2.5% und 25% Täuschern (Aguado, Vedral, Olea, Ponsoda, Barrada & Abad, 2018; Jensen & Thomsen, 2014; Kantrowitz & Dainis, 2014; Nye, Do, Drasgow und Fine, 2008; Tendeiro, Meijer, Schakel und Majij de Meij, 2013). Das Gesamtbild der bisherigen Studien legt

nahe, dass die Möglichkeit eines Täuschungsverhaltens von den Anwender\*innen nicht ignoriert werden kann. Die Gefahr besteht darin, dass die Vorhersagekraft des Testergebnisses für den späteren Berufserfolg gemindert wird und somit mehr ungeeignete Bewerber\*innen eingestellt werden.

Um dieser Problematik zu entkommen und einen hohen wissenschaftlichen Standard in der Personalauswahl zu gewährleisten, wird beim Einsatz eines Online-Tests für die Vorselektion empfohlen, einen Vor-Ort-Test nachzuschalten (International Test Commission, 2006). Trotz einiger möglicher Gegenmaßnahmen sind die Gefahren von Täuschungsverhalten bei Online-Tests nicht abwendbar (Tippins et al., 2006, Reiß, 2016), sodass der Einsatz eines anschließenden Vor-Ort-Tests die sicherste Möglichkeit darstellt, eine zuverlässige Aussage über die tatsächlichen kognitiven Fähigkeiten der Bewerber\*innen zu erhalten. Dieses Vorgehensweise des zweistufigen Testens hat sich etabliert: Zunächst wird ein Online-Test für den Ausschluss eindeutig ungeeigneter Bewerber verwendet (Negativselektion), anschließend wird ein Vor-Ort-Test zur Auswahl der geeigneten Bewerber\*innen eingesetzt (Positivselektion).

Bei einer Negativselektion steht die Identifizierung eindeutig ungeeigneter Bewerber\*innen im Mittelpunkt. Hier wird dementsprechend die Hürde, die Bewerber\*innen erreichen müssen, um zum Vor-Ort-Test zugelassen zu werden, relativ niedrig gesetzt. Auf Grund der Möglichkeit von Täuschung beim Online-Test wird eine Positiv-Selektion nachgeschaltet, bei der die Eignung der Personen in einem kontrollierten Setting festgestellt werden soll. In diesem Schritt werden die Bewerber\*innen auf Grund ihrer Eignung ausgewählt. Häufig fällt die Entscheidung, wer diesen Auswahlsschritt erfolgreich besteht, auf Basis der Ergebnisse des Vor-Ort-Tests.

---

## **Im Mittelpunkt steht die Feststellung der Eignung der Bewerber\*innen**

---

Im Mittelpunkt beider Schritte steht die Feststellung der (kognitiven) Eignung der Bewerber\*innen. Durch dieses Vorgehen wird der Nutzen eines Täuschungsverhaltens verringert. Es geht jedoch nicht darum, Fehlverhalten zu identifizieren. Wer im ersten Auswahlsschritt betrügt, hat zwar eine höhere Chance, den Vor-Ort-Test zu erreichen, muss sich hier jedoch nochmals beweisen, ohne dass er auf die Hilfe zurückgreifen kann, derer er sich beim Online-Test ggf. bedient hat. Personen, die grundsätzlich nicht geeignet sind, aber im Online-Test auf Grund von Hilfe als geeignet erscheinen, sollten im Vor-Ort-Test nicht bestehen und deswegen nicht zugelassen werden.

Es werden zudem einige zusätzliche Strategien diskutiert, um Täuschungsverhalten zu minimieren (z.B. Kantrowitz & Dainis, 2014), wie die Warnung vor möglichen Konsequenzen, die Zustimmung der Bewerber\*innen zu einem „psychologischen Vertrag“, der Einsatz von ad-

aptiven Tests und der Rückgriff auf Zeitbegrenzungen. Trotz dieser Möglichkeiten wird vom reinen Online-Test auf Grund der Gefahren abgeraten und ein zweistufiges Testen empfohlen. Dabei geben die Richtlinien der International Test Commission (2006) die Empfehlung, die Leistung eines Online-Tests mit denen eines Vor-Ort-Tests zu vergleichen.

Vermutlich basierend auf dieser Richtlinie wurde in einigen Studien der Begriff eines Tests zur Verifikation der Testteilnehmer\*innen begründet (Steger et al., 2018; Tendeiro et al., 2013; Guo & Drasgow, 2010). Die Idee hinter einem Verifizierungstest ist folgende: Liegen die Testergebnisse des Online- und Vor-Ort-(Verifizierungs-) Tests auffällig auseinander, werden die Personen als Täuscher\*innen vom weiteren Verfahren ausgeschlossen. Auch in der Praxis sprechen mehrere Anbieter davon, die Testergebnisse von Online-Tests in einem Vor-Ort-Kurztest zu verifizieren und bei zu großen Abweichungen zwischen den Testleistungen Täuscher\*innen herauszufiltern.

Häufig wird beim Einsatz von Verifizierungs-Tests ein inhaltlich breiterer und längerer Online-Test verwendet, mit dem letztlich auch über den Verbleib im Auswahlverfahren entschieden werden soll. Die im ersten Schritt ausgewählten Personen müssen anschließend vor Ort einen kürzeren Verifizierungs-Test bearbeiten, dessen Ergebnis mit dem Online-Test verglichen wird. Dieser Test dient hauptsächlich dazu, Täuscher\*innen zu identifizieren. Die Verwendung von Verifizierungs-Tests fügt dem zweistufigen Testen eine weitere Komponente hinzu. Der Verbleib im Auswahlverfahren wird nicht nur von der kognitiven Eignung der Bewerber\*in abhängig gemacht, sondern zusätzlich von der Diagnose eines möglichen Täuschungsverhaltens. Aus mehreren im Folgenden diskutierten Gründen ist dieses Vorgehen als kritisch zu betrachten.

Zudem wurden mit den in der Praxis verbreiteten Verfahren bisher wenige wissenschaftliche Studien durchgeführt (Guo & Drasgow, 2010, Tendeiro et al., 2013), sodass eine Diskussion des Vorgehens besonders wichtig ist. Im Folgenden wird die Grundlage der Verifizierungs-Tests diskutiert und in einem Datenbeispiel überprüft. Dieser Beitrag soll eine praxisrelevante Betrachtung der Verifizierungstests unter wissenschaftlichen Standards darstellen.

## Identifikation von Täuscher\*innen

Stellen Sie sich eine Situation vor, in der Person A beim Online-Test Hilfe von Person B erhält, einer sehr schlaun Person, während sie im Vor-Ort-Test ohne diese Hilfe auskommen muss. Person C bearbeitet hingegen sowohl den Online-Test als auch den Vor-Ort-Test ohne Hilfe. Es ist naheliegend anzunehmen, dass Person C eine ähnliche Testleistung in beiden Tests zeigt, während Person A im Online-Test besser abschneidet als im Vor-Ort-Test. Ein Täuschungsversuch ist zunächst nur eine Vermutung, wenn ein Unterschied in der Testleistung beobachtet wird.<sup>1</sup> Es schließt sich die Frage an, ob es möglich ist, bei einem solchen Unterschied Rückschlüsse über das Täuschungsverhalten ziehen zu können. Wie groß genau muss der Unterschied in der Leistung im Online-Test und im Vor-Ort-Test sein, um sicher sagen zu können, dass eine Person beim Online-Test Unterstützung hatte?

<sup>1</sup> Es gibt weitere sinnvolle Annahmen über Umstände, die zu systematischen Unterschieden zweier Testergebnisse einer Person führen können. Diese werden in der Diskussion aufgegriffen.

Eine Grundannahme bezüglich kognitiver Fähigkeiten ist die Stabilität dieser Eigenschaften. Grundlegende kognitive Fähigkeiten lassen sich – vor allem kurzfristig – nur in sehr geringem Maße verändern. Diese Stabilität wird durch eine hohe Korrelation von Testergebnissen bei zweimaliger Teilnahme (an Parallel- oder Re-Tests) in großen Stichproben bestätigt. Jedoch haben Testverfahren niemals eine perfekte Messgenauigkeit, sodass es zu unsystematischen Schwankungen in den Test-Ergebnissen einer Person bei zweimaliger Testteilnahme kommt. Je geringer die Messgenauigkeit, desto größer sind diese Schwankungen. Da die Messgenauigkeit bei kürzeren Tests in der Regel geringer ist, sollten kurze Verifikationstests zu hohen Schwankungen führen und somit diesen Prozess sehr fehleranfällig machen.

Auf Grund der erwarteten Schwankungen zweier Testergebnisse muss ein Vergleich statistisch abgesichert werden, damit dieser nicht willkürlich ist. Eine Methode zum Vergleich von Testergebnissen ist die *kritische Differenz zwischen Beobachtungsvariablen zweier wahrer Werte* heranzuziehen, um diesen Vergleich zu ermöglichen. Voraussetzung ist hier, dass beide Tests das Gleiche messen und die Aufgaben vergleichbare Schwierigkeiten und Trennschärfen aufweisen.

### Eine Veranschaulichung mit simulierten Daten

Die oben genannte Methode wird im Folgenden mit simulierten Daten verglichen. Der Vergleich der beiden Testergebnisse erfolgt einseitig, d.h. als Täuscher\*innen werden nur Personen identifiziert, die im Online-Test besser abschnitten als im Vor-Ort-Test. Dabei betrachten wir (1) die Power der Methoden, d.h. den Anteil der tatsächlichen Täuscher\*innen, die auch als solche identifiziert werden (richtig-negative/unehrliche), und (2) den Alpha-Fehler der Methoden, d.h. den Anteil der Personen, die als Täuscher\*innen identifiziert wurden, obwohl sie keine Täuscher \*innen sind (falsch-negative/unehrliche).

Für den Vergleich wurden Testdaten für zwei Tests unter Idealbedingungen erzeugt. Der Online-Test beinhaltet je 30 Fragen und die Verifizierungs-Tests (ein langer und ein kurzer) 30 bzw. 15 Fragen. Die Messgenauigkeit (Reliabilität) der Testverfahren liegt mit einem Wert von  $r = .90$  (beim langen Verifizierungstest) und  $r = .84$  (beim kurzen Verifizierungstest) in einem für diese Testlänge sehr guten Bereich. Beide Tests messen die gleiche Fähigkeit und sind im strengeren Sinne Paralleltests. Das bedeutet, dass die Items inhaltlich äquivalent, gleich schwierig und trennscharf sind (siehe auch Schmidt-Atzert & Amelang, 2012). Für den kurzen Verifizierungs-Test mit 15 Fragen sind diese parallel zu 15 der 30 Aufgaben des Online-Tests.

Insgesamt haben  $N = 10.000$  fiktive Personen am Online und Verifizierungs-Test teilgenommen. Es wurden zwei Varianten berechnet:

1. In der ersten Simulation haben 75% der simulierten Testteilnehmer\*innen ohne Hilfe beide Tests absolviert und 25% mit Unterstützung.
2. In der zweiten Simulation haben 95% keine Hilfe und 5% Unterstützung oder haben diesen Test von einer fähigeren Person bearbeiten lassen.

Die beiden Varianten lehnen sich an die Studien an, in denen Täuschungsraten ermittelt wurden (s.o.). Eine Annahme, die wir treffen, ist, dass täuschende Personen im Mittel eine um eine Standardabweichung geringere Fähigkeit aufweisen als Personen, die nicht täuschen (eine Standardabweichung wäre in einem dgp-Test gleich 10 Standardwert-Punkte). Bei eher geringer Fähigkeit wird vermutlich eher auf Täuschungshandlungen zurückgegriffen als bei hoher Fähigkeit. Für Täuscher\*innen wurden die Daten im Online-Test so erzeugt, als ob der Test von einer um eine Standardabweichung besseren Person bearbeitet wurde. Diese Personen gilt es im Rahmen einer Verifizierung zu identifizieren. Die Anzahl der richtig Identifizierten (richtig-negativ/unehrlich, *Power*), der falsch Identifizierten (falsch-negative/unehrlich, *alpha-Fehler*), der richtig Nicht-Identifizierten (richtig-positive/ehrlich) und der falsch Nicht-Identifizierten (falsch-positive/ehrlich) wird für die beiden oben genannten Methoden sowie beide Raten von Täuscher\*innen (5% und 25%) verglichen.

Die Ergebnistabellen 1 und 2 zeigen, wie viele der Personen, die in Wahrheit ehrlich oder unehrlich waren, durch die festgesetzten Grenzen als ehrlich oder unehrlich eingestuft wurden.

Für eine Stichprobe mit 25% Täuscher\*innen (Tabelle 1) zeigt sich

- beim längeren Verifizierungstest, dass nur 14 von 25, d.h. 54% aller Täuscher\*innen als solche identifiziert werden (*Power*). Gleichzeitig werden 6 von 75, d.h. 9% der ehrlichen Personen als unehrlich eingestuft (*alpha-Fehler*). Insgesamt werden 18 % aller Testteilnehmer\*innen in diesem Prozess falsch behandelt.
- beim kürzeren Verifizierungstest, dass 9 von 25, d.h. 35% aller Täuscher\*innen als solche identifiziert werden (*Power*). Gleichzeitig werden 3 von 75, d.h. 5% der ehrlichen Personen als unehrlich eingestuft (*alpha-Fehler*). Insgesamt werden 20 % aller Testteilnehmer\*innen in diesem Prozess falsch behandelt.

	Langer Verifizierungstest (30 Fragen)			Kurzer Verifizierungstest (15 Fragen)		
	In Wahrheit...			In Wahrheit...		
	ehrlich	unehrlich		ehrlich	unehrlich	
Durch Verifizierung identifiziert als...	ehrlich	68.65	11.41	71.53	16.36	
	unehrlich	6.35	13.59	3.47	8.64	
		75	25	75	25	100

Tabelle 1: Anteil der richtig-positiven, richtig-negativen, falsch-positiven und falsch-negativen Entscheidungen (in %) für eine Stichprobe mit 25 % Täuscher\*innen,  $\alpha = 5\%$

		Langer Verifizierungstest (30 Fragen) In Wahrheit...			Kurzer Verifizierungstest (15 Fragen) In Wahrheit...		
		ehrllich	unehrlich	100	ehrllich	unehrlich	100
Durch Verifizierung identifiziert als...	ehrllich	86.99	2.11		90.25	3.26	
	unehrlich	8.01	2.89		4.75	1.74	
		95	5	100	75	25	100

Tabelle 2. Anteil der richtig-positiven, richtig-negativen, falsch-positiven und falsch-negativen Entscheidungen (in %) für eine Stichprobe mit 5 % Täuscher\*innen, alpha = 5%

Für eine Stichprobe mit 5% Täuscher\*innen (Tabelle 2) zeigt sich

- beim längeren Verifizierungstest, dass 3 von 5, d.h. 58% aller Täuscher\*innen als solche identifiziert werden (Power). Gleichzeitig werden 8 von 95, d.h. 8% der ehrlichen Personen als unehrlich eingestuft (alpha-Fehler). Insgesamt werden 10 % aller Testteilnehmer\*innen in diesem Prozess falsch behandelt.
- beim kürzeren Verifizierungstest, dass 2 von 5, d.h. 35% aller Täuscher\*innen als solche identifiziert werden (Power). Gleichzeitig werden 5 von 95, d.h. 5% der ehrlichen Personen als unehrlich eingestuft (alpha-Fehler). Insgesamt werden 8 % aller Testteilnehmer\*innen in diesem Prozess falsch behandelt.

### Schlussfolgerungen und weitere kritische Aspekte

Die Simulation veranschaulicht, dass das Vorgehen, Täuscher\*innen mit Hilfe von Verifizierungstests identifizieren zu wollen, erhebliche Risiken mit sich bringt. Zum einen besteht die Gefahr, Personen als Täuscher\*innen zu identifizieren, die keine sind. Davon können 3% bis 8% der Bewerber\*innen betroffen sein. Diesen Bewerber\*innen gegenüber ist das Vorgehen demnach höchst unfair (was es dann auch tatsächlich ist), da ihnen unrechtmäßig ein Täuschungsverhalten vorgeworfen wird.

Darüber hinaus wird man nur einen geringen Anteil der Täuscher\*innen auch als solche erkennen können. Der Anteil der korrekt als solche identifizierten Täuscher\*innen schwankt zwischen 35% und 58%. Impliziert wird damit, dass teilweise mehr als die Hälfte der Täuscher\*innen übersehen werden. Durch die geringere Messgenauigkeit erhöhen sich die Fehleraten, wenn der Verifizierungs-Test kürzer ist als der Online-Test. Insbesondere für den

kürzeren Verifizierungs-Test sind damit erhebliche Fehleinschätzungen bezüglich des Täuschungsverhaltens anzunehmen.

---

## **Bei kürzeren Verifizierungs-Tests sind erhebliche Fehleinschätzungen des Täuschungsverhaltens anzunehmen**

---

Basierend auf den Ergebnissen der kritischen Betrachtung von Verifizierungstests muss der Sinn dieser in Frage gestellt werden. Bewerber\*innen werden mit Leistungstests getestet, weil man sich davon eine Aussage über deren späteren beruflichen Erfolg erhofft. Durch Online-Tests und die damit einhergehende Möglichkeit des Täuschens muss ein Verfahren gefunden werden, mit dem Täuschen wirkungsvoll begegnet werden kann. Ein Verifizierungstest kann das nicht gewährleisten. Stattdessen wird damit der Sinn der Einstellungstests konterkariert. Durch die Einstellungstests sollen geeignete Bewerber\*innen möglichst treffgenau gefunden werden, durch Online-Test und Verifizierungstest wird dieses eigentliche Vorhaben zum Teil zunichtegemacht.

Die dadurch entstehenden Kosten sind hoch. In einem der oben genannten Beispiele (25% Täuscher\*innen, kurzer Verifizierungstest) würden ca. 88 Bewerber\*innen als scheinbar ehrliche potentiell in den nächsten Auswahlsschritt eingeladen. Ca. 17 Personen wären davon Täuscher\*innen, also Bewerber\*innen, die wahrscheinlich kognitiv nicht geeignet sind für die angestrebte Position, bzw. über deren kognitive Eignung man keine valide Aussage treffen kann. Geht man davon aus, dass andere Merkmale und Kompetenzen bei Täuscher\*innen und Nicht-Täuscher\*innen ähnlich verteilt sind, bedeutet das, dass 19 % der Auswahlentscheidungen Fehlentscheidungen sind. Fast jede fünfte so vorgenommene Einstellung wird dann mit einer Person vorgenommen, über deren kognitive Eignung man keine Aussage treffen kann. Würde ein Münzwurf anstatt des Verifizierungstests durchgeführt, wäre man mit 25 % Fehleinstellungen gar nicht so viel schlechter gestellt.

Ein Übersehen der Täuscher\*innen ist bei dem Einsatz eines Verifizierungs-Tests deswegen insbesondere kritisch, da hier die Auswahlentscheidung auf Basis des Online-Tests getroffen wird (s.o.). In eben diesem werden also Bewerber\*innen übersehen, die sich unrechtmäßig einen Vorteil verschafft haben. Im Gegensatz dazu wird bei dem Vorgehen der Negativ- und Positivselektion die Auswahlentscheidung meist auf Basis des Vor-Ort-Tests getroffen.<sup>3</sup> Da der Vor-Ort-Test unter gesicherten Bedingungen stattfindet, besteht hier keine Gefahr, dass Bewerber\*innen sich durch Mithilfe anderer o.ä. unrechtmäßige Vorteile verschaffen.<sup>4</sup>

Des Weiteren könnten Bewerber\*innen, die unrechtmäßig als Täuscher\*in identifiziert wurden, das Verfahren auch rechtlich in Frage stellen. Sollten diese Bewerber\*innen nach dem Beweis über ihr Täuschungsverhalten fragen, ist dieser kaum zu erbringen. Grundsätzlich ist der Ausschluss auf Grund der Annahme eines Täuschungsverhaltens nur auf Basis von Mit-

---

<sup>3</sup> Oder auf Basis einer Kombination des Online- und Vor-Ort-Tests.

<sup>4</sup> Selbstverständlich müssen beim Vor-Ort-Test ebenfalls Sicherheitsmaßnahmen gegen Täuschungsversuche unternommen werden, wie unterschiedliche Testversionen, Aufsicht etc.



telwertsunterschieden zwischen zwei Tests fragwürdig. Die Validität dieser Mittelwertsunterschiede – und damit der Verifizierungstests – wurde bisher nicht ausreichend untersucht, d.h. die Frage: Was sagen diese Unterschiede eigentlich aus? Untersucht wurde bisher, wie häufig solche Mittelwertsunterschiede in Stichproben vorkommen (siehe Kantrowitz & Dainis, 2014), nicht, ob wirklich Täuschungsverhalten zu den Mittelwertsunterschieden geführt hat. Zwar kann man bei Täuscher\*innen von einem systematischen Unterschied ausgehen (d.h. bessere Leistung im Online-Test), aber man kann bei einem auftretenden Unterschied nicht immer zuverlässig auf eine Täuschung rückschließen.

Es könnte weitere systematisch auftretende Gründe für Mittelwertsunterschiede zwischen einem Online- und Vor-Ort-Test geben, die noch nicht ausreichend untersucht wurden. Beispielsweise könnte eine Person in einem Vor-Ort-Test schlechter abschneiden, weil die einer Prüfung ähnliche Situation Ängste hervorruft. Auch naheliegend als Erklärung für eine schlechtere Leistung im Vor-Ort-Test können soziale Ängste sein, wenn die Testung in Gruppen stattfindet. Weitere Eigenschaften der Testsituation könnten potentiell systematische Unterschiede erklären: Beispielsweise können in Online-Tests Pausen häufig selbstbestimmt gewählt werden, während in Vor-Ort-Testungen Pausen vorgegeben werden.

Anzumerken ist zudem, dass die hier verwendeten Beispieldaten eine Idealsituation widerspiegeln, wie sie in der Praxis nicht immer zu finden ist. Für reale Testdaten könnten die Fehlerraten noch höher ausfallen, z.B. wenn die Messgenauigkeiten nicht hoch genug sind oder der Online- und Verifizierungs-Test im eigentlichen Sinne keine Paralleltests sind.

Auf Grund des hier dargestellten empirischen Beispiels und der weiteren inhaltlichen Bedenken bezüglich der Validität und rechtlichen Rahmenbedingungen von Verifizierungs-Tests können wir dieses Vorgehen nicht empfehlen. Nach wie vor empfehlenswert ist das Vorgehen eines zweistufigen Auswahlverfahrens (Negativ- und Positivselektion), das orientiert an den (kognitiven) Anforderungen vorgeht (Tippins et al. 2006).

## Literatur:

- Aguado, D., Vidal, A., Olea, J., Ponsoda, V., Barrada J. R. & Bad, F. J. (2018). Cheating on Unproctored Internet Test Applications: An Analysis of a Verification Test in a Real Personnel Selection Context. *The Spanish Journal of Psychology*, 21, 1-10. DOI: 10.1017/sjp.2018.50
- Benit, N. & Soellner, R. (2013). Scientist-practitioner gap in Deutschland: Eine empirische Studie am Beispiel psychologischer Testverfahren. *Zeitschrift für Arbeits- und Organisationspsychologie*, 57, 145-153. DOI: 10.1026/0932-4089/a00011
- Geister, S. & Rastetter, D. (2009). Aktueller Stand zum Thema Online-Tests. In Heike Steiner (Hrsg.), *Online-Assessment*. Heidelberg: Springer Medizin Verlag. DOI: 10.1007/978-3-540-78919-2
- Goertz, W., Hülshager, U. R. & Maier, G. W. (2014). The Validity of Specific Cognitive Abilities for the Prediction of Training Success in Germany: A Meta-Analysis. *Journal of Personnel Psychology*, 13, 123-133. DOI: 10.1027/1866-5888/a000110
- Guo, J. & Drasgow, F. (2010). Identifying Cheating on Unproctored Internet Tests: The Z-test and the likelihood ratio test. *International Journal of Selection and Assessment*, 18, 351-364. DOI: 10.1111/j.1468-2389.2010.00518.x
- Hülshager, U. R., Maier, G. W., & Stumpp, T. (2007). Validity of General Mental Ability for the Prediction of Job Performance and Training Success in Germany: A meta-analysis. *International Journal of Selection and Assessment*, 15, 3-18. DOI: 10.1111/j.1468-2389.2007.00363.x
- International Test Commission. (2006). International guidelines on computer-based and Internet-delivered testing. *International Journal of Testing*, 6, 143-171. DOI: 10.1207/s15327574ijt0602\_4
- Jensen, C. & Thomsen, J. P. F. (2014). Self-reported cheating in web surveys on political knowledge. *Quality & Quantity*, 48, 3343-3354. DOI 10.1007/s11135-013-9960-z
- Kantrowitz, T. M. & Dainis, A. M. (2014). How Secure are Unproctored Pre-Employment Tests? Analysis of Inconsistent Test Scores. *Journal of Business Psychology*, 29, 605-616. DOI 10.1007/s10869-014-9365-6
- 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
- Reiß, C. (2016). Vor- und Nachteile von Online-Testverfahren. *dgp-Informationen* 57. Jahrgang, Heft 66, 64-69
- Schmidt-Atzert, L. & Amelang, M. (2012). *Psychologische Diagnostik* (5. Auflage). Berlin: Springer. DOI: 10.1007/3-540-37637-2\_40
- Schuler, H., Hell, B., Trapmann, S., Schaar, H. & Boramir, I. (2007). Die Nutzung psychologischer Verfahren der externen Personalauswahl in deutschen Unternehmen: Ein Vergleich über 20 Jahre. *Zeitschrift für Personalpsychologie*, 5, 60-70. DOI: 10.1026/1617-6391.6.2.60
- Stadelmaier, U., Thiem, H. & Kleinmanns, A. (2018). Eignungsdiagnostik für die Feuerwehr: Bewährungskontrolle des dgp-Tests G2 bei der Berufsfeuerwehr München. *dgp-Informationen*, 59(68).
- Steger, D., Schroeders, Ul. & Gnambs, T. (2018). A Meta-Analysis of Test Scores in Proctored and Unproctored Ability Assessments. *European Journal of Psychological Assessment*. Advance online publication. <http://dx.doi.org/10.1027/1015-5759/a000494>
- Tendeiro, J. N., Meijer, R. R., Schakel, L. & Majj de Meij, A. M. (2013). Using Cumulative Sum Statistics to Detect Inconsistencies in Unproctored Internet Testing. *Educational and Psychological Measurement*, 73, 43-161. DOI: 10.1177/0013164412444787.
- Tippins, N. T., Beaty, J., Drasgow, F., Gibson, W. M., Pearlman, K., Segall, D. O. & Shepherd, W. (2006). Unproctored internet testing in employment settings. *Personnel Psychology*, 59, 189-225. DOI: 10.1111/j.1744-6570.2006.00909.x
- Tippins, N. T. (2015). Technology and Assessment in Selection. *Annual Review of Organizational Psychology and Organizational Behavior*, 2, 51-82. DOI: 10.1146/annurev-orgpsych-031413-091317

## Kontakt:

Dipl.-Psych. Dr. Anna-Lena Jobmann  
jobmann@dgp.de

Deutsche Gesellschaft für Personalwesen e.V.  
Kantstr. 153, 10623 Berlin

Dipl.-Psych. Dr. Christian Reiß  
Leiter der Geschäftsstelle Düsseldorf  
reiss@dgp.de

Deutsche Gesellschaft für Personalwesen e.V.  
Hohenzollernstr. 11-13, 40211 Düsseldorf



**Deutsche Gesellschaft  
für Personalwesen e.V.**

Berlin  
Kantstraße 153  
10623 Berlin  
T 030 / 3983718-0  
F 030 / 3983718-29  
E berlin@dgp.de

Düsseldorf  
Hohenzollernstraße 11-13  
40211 Düsseldorf  
T 0211 / 688508-0  
F 0211 / 688508-29  
E duesseldorf@dgp.de

Hannover  
Stammestraße 40 D  
30459 Hannover  
T 0511 / 94393-0  
F 0511 / 94393-43 / 44  
E hannover@dgp.de

Leipzig  
Grassistraße 12  
04107 Leipzig  
T 0341 / 9782-223  
F 0341 / 9782-225  
E leipzig@dgp.de

Stuttgart  
Kernerstraße 43  
70182 Stuttgart  
T 0711 / 9457-6701  
F 0711 / 9457-2765  
E stuttgart@dgp.de