

Wenn keiner



Bianucci, D.E., Rademann, R. & Kersting, M. (2022). Wenn keiner zuschaut... unbeaufsichtigte Leistungstests in der Personalauswahl. *Personalmagazin*, 1/2022, S. 50-53.

Hinweis: Dies ist das akzeptierte Manuskript (Author's Accepted Manuscript) des oben genannten Beitrages. Die Nutzer(innen) dürfen die Inhalte nur zum Zwecke der wissenschaftlichen Forschung ansehen, drucken, kopieren, herunterladen, sowie für Text- und Datamining verwenden. Die Inhalte dürfen weder ganz noch teilweise wörtlich (wieder)veröffentlicht oder für kommerzielle Zwecke verwendet werden. Die Nutzer(innen) müssen sicherstellen, dass die Urheberpersönlichkeitsrechte des Autors sowie gegebenenfalls bestehende Rechte Dritter an den Inhalten oder Teilen der Inhalte nicht verletzt werden.

zuschaut

Die Digitalisierung macht auch vor der Personalauswahl nicht halt. Neu ist dabei unter anderem die Art der Durchführung von Eignungstests. Bislang wurden Bewerbende eingeladen, um unter Aufsicht Leistungstests zu absolvieren. Nun können sie die Tests von zu Hause aus bearbeiten. Dieser Trend zur sogenannten „unbeaufsichtigten Testung“ hat sich durch die coronabedingten Distanzgebote deutlich verstärkt. Das bringt, wie Studien zeigen, einige Vorteile, aber auch viele Nachteile mit sich. Konkrete Tipps zeigen, wie Personaler und Personalerinnen diese Vorgehensweise in der Praxis optimal nutzen können.

Von David Enrico Bianucci, Ralf Rademann und Martin Kersting

● „Einfach und sicher von zu Hause aus“ – was für das Onlinebanking gilt, ist auch für den Eignungstest verlockend. Statt die Bewerbenden zu einem fixen Datum in die einstellende Organisation einzuladen, bearbeiten diese den Eignungstest an einem Ort und Zeitpunkt ihrer Wahl, auf dem eigenen Endgerät. In der Praxis hat sich dafür der Begriff „Onlinetestung“ eingebürgert, der allerdings irreführend ist, da die Testung sehr häufig – auch vor Ort – onlinebasiert ist. Treffender ist der Ausdruck „Testung ohne Aufsicht“. Nach Schätzungen nutzten bereits 2008 sämtliche Unternehmen im US-amerikanischen Fortune 500-Index unbeaufsichtigte Testverfahren zur (Vor-)Selektion von Kandidat(inn)en.

Vorteile – einfach mal zu Hause bleiben

Die Vorteile liegen auf der Hand: Sie sind vor allem organisatorischer und ökonomischer Art. Durch die wegfallende Bindung an Ort und Termin lässt sich Zeit (für die Vorbereitung und Durchführung), Personal und Geld sparen. Die Einladung zur unbeaufsichtigten Testung kann ebenso wie die weitere Kommunikation (Absage oder Einladung zur nächsten Verfahrensstufe) voll automatisiert erfolgen. Die Methode stellt eine Alternative zu anderen Verfahren der Vorauswahl (zum Beispiel nach Schulnoten) dar. Da sich die Verfahren beliebig skalieren lassen, können insgesamt mehr Personen in der Vorauswahl berücksichtigt werden. Ihre Stärken spielen unbeaufsichtigte Tests daher besonders bei großen Kohorten an Bewerbenden aus, also etwa bei der Auswahl von Auszubildenden.

Auch aus der Perspektive der Bewerbenden ist dieses Vorgehen attraktiv. Man spart sich die mühsame Anreise, wenn man das Auswahlverfahren ebenso gut von zu Hause aus erledigen kann. Die soziale Akzeptanz ist dabei hoch: In einer Studie mit 475 Bewerberinnen und Bewerbern haben wir gezeigt, dass die unbeaufsichtigte Testung unter Akzeptanzgesichtspunkten etwas positiver bewertet wird als die beaufsichtigte Testung – insbesondere Frauen fanden es angenehmer, ohne die Anwesenheit anderer Personen getestet zu werden.

Nachteile – es ist nicht alles Gold, was glänzt

Die unbeaufsichtigte Bearbeitung birgt jedoch auch Nachteile, wenn es sich um Leistungstests (nur diese wollen wir hier thematisieren) handelt. Zunächst fehlt mit der Testaufsicht auch die Unterstützung bei inhaltlichen oder technischen Fragen. Dies könnte dazu führen, dass sich die getesteten Personen allein gelassen fühlen.

Wesentlich ist die Frage, ob man den Testergebnissen trauen kann, wenn die Aufsicht fehlt. Mittlerweile liegen zu dem unbeaufsichtigten Testen so viele empirische Studien vor, dass ihre Ergebnisse in einer Metaanalyse zusammengefasst wurden. Laut dieser Metaanalyse, die Diana Steger, Ulrich Schroeders und Tino Gnambs 2020 im „European Journal of Psychological Assessment“ publizierten, fallen die Testergebnisse bei unbeaufsichtigten Tests im Durchschnitt etwas besser aus als bei ihren beaufsichtigten Pendanten. Das könnte daran liegen, dass viele Personen Leistungen vortäuschen, wenn sie unbeaufsichtigt sind. Ein Beweis für das Schummeln sind die besseren Testleistungen allein aber nicht. Denkbar wäre auch, dass die Leistungen zum Beispiel aufgrund reduzierter Testangst ansteigen. Die Ergebnisse anonymer Befragungen legen allerdings nahe, dass mit der Möglichkeit zu täuschen tatsächlich die Täuschungsraten

Mögliche Täuschungshandlungen bei unbeaufsichtigten Leistungstests

- Unzulässiges Vorwissen (Bewerbende haben vorab die Testaufgaben unrechtmäßig erworben, zum Beispiel durch Diebstahl der Aufgaben oder durch eine Vorabbearbeitung unter anderer Identität)
- Nutzung von unerlaubten Hilfsmitteln (zum Beispiel Taschenrechner, Wörterbücher, Suchmaschinen)
- Bearbeitung mit Unterstützung anderer Personen
- Bearbeitung durch andere Personen

steigen. Der Deutsche Hochschullehrerverband berichtet unter Bezug auf eine Studie im Kontext universitärer Prüfungen, dass Studierende in Onlineklausuren mit einer fast doppelt so hohen Wahrscheinlichkeit wie bei Präsenzprüfungen getäuscht hätten.

Anzumerken ist, dass die Testbearbeitung auch dazu genutzt werden kann, sich die Testmaterialien unter Verstoß gegen das Urheberrecht anzueignen (zum Beispiel durch Abfotografieren oder Video Downloader). Damit sind die Tests für bestimmte Einsatzzwecke „verbrannt“. Tests, die für die unbeaufsichtigte Testung genutzt werden, sind nicht sicher und kommen daher für eine beaufsichtigte Testung grundsätzlich nicht mehr in Frage.

Sinkt die Aussagekraft des Tests?

Dass die Testergebnisse bei unbeaufsichtigter Testung tendenziell besser ausfallen, ist per se nicht schlimm – entscheidend ist die Frage, ob die Ergebnisse unbeaufsichtigter Tests aussagekräftig (valide) sind und ob das Vorgehen fair ist. Diese Fragen sind bislang wissenschaftlich nicht ausreichend untersucht. Validitätsstudien mit Daten aus realen Auswahlverfahren gibt es unseres Wissens bislang nur für Persönlichkeitsfragebogen. Hier wirkte sich die unbeaufsichtigte Testung im Durchschnitt über alle Personen betrachtet nicht negativ auf die Aussagekraft aus. Für Leistungstests gibt es bislang nur Simulationsrechnungen – die von negativen Effekten des Täuschungsverhaltens auf die Aussagekraft der Befunde ausgehen. Selbst wenn die Aussagekraft im Durchschnitt gleichbleibt, ist davon auszugehen, dass sich differentielle Effekte zeigen, wenn nicht alle Untergruppen im gleichen Ausmaß täuschen. Studien zeigen, dass Männer eher täuschen als Frauen. Somit könnte sich ein unbeaufsichtigtes Testen zuungunsten bestimmter Gruppen auswirken. Die Ehrlichen wären die Dummen. Die bisherigen Ausführungen beschäftigten sich mit der Frage der so genannten Kriteriumsvalidierung. Das ist unter anderem die Frage, ob sich anhand der Testergebnisse künftiger Erfolg in Ausbildung oder Beruf treffsicher vorhersagen lässt. Eine andere Frage ist die der Konstruktvalidierung – hier geht es darum, ob mit einem unbeaufsichtigt durchgeführten Test überhaupt das Gleiche gemessen wird wie bei einer beaufsichtigten Testung. In einer von der Justus-Liebig-Universität in Gießen in Kooperation mit der Bayer AG durchgeführten Studie haben wir gezeigt, dass dies nur bedingt der Fall ist.

Bei der Bayer AG erhalten Bewerbende für Ausbildungsberufe in der Chemiebranche – sofern sie bestimmte Vorauswahlkri-

terien erfüllen – die Möglichkeit, einen Leistungstest unbeaufsichtigt zu bearbeiten. Sofern sie den Test erfolgreich bestehen, werden sie zur nächsten Auswahlstufe eingeladen. Hier erwartet sie neben einem Interview auch ein erneuter Leistungstest – diesmal unter Aufsicht. Der zweite Test dient der Verifizierung der ohne Aufsicht erzielten Testergebnisse. Diese Prüflogik setzt voraus, dass die beiden Tests dieselben Fähigkeiten erfassen, obwohl der zweite Test keine reine Wiederholung des ersten Tests ist (da man sonst Wiederholungseffekte hätte). Für unsere Studie standen die Daten von 524 Bewerbenden zur Verfügung, die beide Tests absolviert hatten.

Um die Daten besser interpretieren zu können, haben wir die Tests im Rahmen einer Forschungsstudie zusätzlich einer studentischen Stichprobe (N=101) vorgelegt – diesmal wurden beide Tests unter Aufsicht bearbeitet. Im Anschluss daran konnten wir die aus dem Auswahlprozess stammenden Koeffizienten mit denen aus der Forschungsstudie vergleichen. Dabei bestätigte sich nicht nur, dass bei der Testung ohne Aufsicht bessere Ergebnisse erzielt werden, sondern es zeigte sich auch eine deutliche Minderung der korrelativen Zusammenhänge der beiden Tests. Das bedeutet: Wenn beide Tests unter Aufsicht durchgeführt werden, messen sie eher dieselben Fähigkeiten (konvergente Konstruktvalidität). Wird der eine Test ohne Aufsicht, der andere aber mit Aufsicht durchgeführt, messen die beiden Tests tendenziell etwas Verschiedenes. Dies ist in doppelter Hinsicht problematisch: Zum einen misst man bei der Testung ohne Aufsicht nicht unbedingt das, was man messen will (zum Beispiel die kognitive Kompetenz), zum anderen klappt die Verifizierung durch den zweiten Test nicht so gut (weil dieser unter Aufsicht etwas Anderes misst). Aus der Analyse konnten wir konkrete Tipps zur weiteren Optimierung der Tests ableiten.

Ein „How to“ für unbeaufsichtigtes Testen

Unbeaufsichtigte Leistungstests stellen trotz der genannten Nachteile in bestimmten Fällen eine sinnvolle Ergänzung des bestehenden Repertoires an Auswahlmethoden dar. Im Folgenden nennen wir die Punkte, die bei der Auswahl des Aufgabenmaterials, der diagnostischen Strategie, der Durchführung, der Interpretation und der Evaluation zu beachten sind.

1. Auswahl des Aufgabenmaterials

Grundsätzlich ist bei der Auswahl des Aufgabenmaterials zu beachten, dass sich einige Typen von Aufgaben mehr und andere weniger für unbeaufsichtigte Testungen eignen. Insbesondere bei (Wissens-)Aufgaben, deren Lösungen leicht nachzuschlagen sind, lässt sich bei fehlender Aufsicht leicht täuschen. Demgegenüber sind Aufgaben zum logischen Denken vergleichsweise robust. Dieser Hinweis setzt aber nicht den Grundsatz außer Kraft, dass

ein Eignungstest immer auf die Erfassung der Eignungsmerkmale abzielt, die sich aus dem Anforderungsprofil ergeben.

Weitere Fragen sind, welche einzelnen Aufgaben – in Tests „Items“ genannt – den Bewerbenden in welcher Reihenfolge wie lange präsentiert werden. Zur Prävention von Täuschungen ist es günstig, Aufgaben nur kurz zu präsentieren, die Möglichkeit, zu bearbeiteten Items zurückzuspringen zu unterbinden und zumindest das Aussehen der Items und auch die Abfolge der Items zu variieren. Hinsichtlich des Aussehens ist es beispielsweise möglich, eine Additionsaufgabe einmal als „2+8+6“ und beim nächsten Mal als „6+2+8“ zu präsentieren. Solche Variationen setzen aber immer voraus, dass die psychometrische Gleichwertigkeit der Items (zum Beispiel der Schwierigkeitsgrad) und der Item-Abfolgen empirisch geprüft und gewährleistet wird.

Noch günstiger für die Täuschungsprävention ist es, wenn verschiedene Bewerbende verschiedene Items präsentiert bekommen. Dies ist allerdings psychometrisch anspruchsvoll. Es setzt voraus, dass man über viele hundert Items verfügt, die im Rahmen eines prüfbaren Modells wahrscheinlichkeitstheoretischer Überlegungen entwickelt und empirisch kalibriert wurden. Dies ermöglicht das so genannte „adaptive Testen“. Bei Tests, die nach der probabilistischen Testtheorie konstruiert wurden, ist es außerdem mit bestimmten Methoden möglich, Täuschungsverhalten zu detektieren. Derartige Testkonstruktionen sind aber sehr aufwändig und entsprechend selten. Einstellende Organisationen sollten kritisch nachfragen, wenn ihnen derartige Leistungen versprochen werden, die mit Tests, die nach der verbreiteten klassischen Testtheorie konstruiert wurden, nicht erbracht werden können.

2. Diagnostische Strategie

Unbeaufsichtigte Leistungstests eignen sich als Screeningverfahren, sie können eine erste Stufe in einem Auswahlprozess mit einer hohen Anzahl an Bewerbenden darstellen. Sie sind in der Regel eine Ergänzung, aber kein Ersatz für beaufsichtigte Leistungstests.

3. Durchführung

Bei der Durchführung unbeaufsichtigter Tests sollte den Bewerbenden Unterstützung angeboten werden, es sollten Ansprechpersonen für inhaltliche und technische Fragen genannt werden. Die Durchführung kann um zahlreiche Maßnahmen angereichert werden, die zur Täuschungsprävention dienen. Eine einfache, aber effektive Maßnahme besteht darin, die Teilnehmenden darauf hinzuweisen, dass die Organisation sich der Täuschungsmöglichkeit bewusst ist und daher zu einem späteren Zeitpunkt mit den Personen, die das gesetzte Leistungskriterium erfüllen, einen Verifikationstest unter Aufsicht durchführen wird. Dies kann die Täuschungsmotivation senken. Insgesamt ist das Täuschungsrisiko natürlich umso größer, je begehrt das mit dem Testergebnis verbundene Gut ist. Auch das Einfordern einer Erklärung kann wirkungsvoll sein. Hierbei bittet man die Teilnehmenden, einer entsprechenden Aussage zuzustimmen (zum Beispiel: „Ich bestätige, dass ich den Test allein und ohne Hilfsmittel bearbeitet habe. Ich werde die Testmaterialien vertraulich behandeln und nicht weitergeben...“).

Schließlich kann man das Maß der Kontrolle und Beaufsichtigung graduell erhöhen, ohne gleich die Bewerbenden ein-

DAVID ENRICO BIANUCCI und PROF. DR. MARTIN KERSTING arbeiten in der Abteilung Psychologische Diagnostik der Justus-Liebig-Universität Gießen.

RALF RADEMANN ist Head of Vocational Training bei der Bayer AG.

Die Au
Unsere
ffizien
ib – sc
Leicht
Flexibil
Data G
05221

bestellen zu müssen. Neben der gänzlich unbeaufsichtigten Durchführung finden sich auf dem Kontinuum hin zur Testung unter vollständig (digitaler) Beaufsichtigung verschiedene Abstufungen. Die International Test Commission unterscheidet die (1) „beaufsichtigte“ und die (2) „unbeaufsichtigte“ Testung, welche sich wiederum in zwei Substufen aufteilen lassen. Von der (1a) unbeaufsichtigten Bearbeitung eines frei zugänglichen Tests unterscheidet man (1b) die unbeaufsichtigte Bearbeitung eines Tests, der nur über die Vergabe von individuellen Zugangsdaten erreichbar ist. Bei der ersten Stufe (2a) der beaufsichtigten Testung sind die Bewerbenden immer noch am Ort ihrer Wahl. Sie werden aber aus der Ferne überwacht, indem sie beispielsweise gebeten werden, die Webcam ihres Endgeräts freizugeben. So kann sowohl eine Identifizierung (Ausweiskontrolle) als auch eine Überwachung bei der Testbearbeitung durchgeführt werden. Außerdem werden die Bewerbenden gebeten, ihren Bildschirm zu teilen, um die Nutzung unerlaubter Hilfsmittel wie das Öffnen weiterer Browserfenster und eine Aufzeichnung zu verhindern. Selbst eine komplette Überwachung der von den Bewerbenden genutzten Technik (2b) sowie eine Überwachung in Präsenz setzt nicht voraus, dass die Bewerbenden für die Testung in die Organisation kommen. Große Organisationen verfügen häufig über ein dichtes Netz an Niederlassungen, die für eine geografisch gute Abdeckung sorgen. Diese Infrastruktur kann vorteilsbringend genutzt werden, indem Bewerbende zum Ablegen eines Eignungstests in eine der Niederlassungen eingeladen werden. Alternativ kann man Plätze in Testcentern mieten. Diese Art der Administration bringt noch immer organisatorische Vorteile mit sich, minimiert jedoch technische Schwierigkeiten und Täuschungen.

4. Interpretation

Die Ergebnisse von Leistungstests, die ohne Aufsicht bearbeitet wurden, lassen sich nur im negativen Fall eindeutig interpretieren. Die Personen, die es trotz aller Täuschungsmöglichkeiten nicht schaffen, ein gesetztes Mindestkriterium zu erfüllen, müs-

sen nicht zur nächsten Verfahrensstufe eingeladen werden. Dies können unter Umständen viele Personen sein, sodass sich das unbeaufsichtigte Testen lohnt. Bei Personen, die den Test erfolgreich bearbeiten, kann man hingegen nicht ausschließen, dass sie Leistungen vorgetäuscht haben. Daher ist es unerlässlich, die unbeaufsichtigt erbrachten Testleistungen zu verifizieren. Als effektiv hat sich dabei die Verwendung eines anschließenden Tests unter Beaufsichtigung herausgestellt. Diese Testung kann in folgende Auswahlsschritte, etwa Interviews oder Assessment Center, integriert werden. Der Verifikationstest kann kürzer ausfallen als der unbeaufsichtigte Test. Voraussetzung ist jedoch die Parallelität der beiden Tests, sie müssen also nachweislich dieselbe(n) Fähigkeit(en) erfassen ohne identisch zu sein.

5. Evaluation

Ob das unbeaufsichtigte Testen funktioniert, muss regelmäßig evaluiert werden. Dafür eignet sich unter anderem die Analyse verschiedener Test- und Itemkennwerte, beispielsweise die der Itemschwierigkeit. Diese gibt Aufschluss darüber, wie häufig ein Item richtig gelöst wurde. Starke Anstiege der Lösungswahrscheinlichkeit können ein Indiz für Täuschung, etwa durch das Bekanntwerden einzelner Testinhalte, sein. Darüber hinaus muss nachgewiesen werden, dass der Verifizierungstest dasselbe misst wie der unbeaufsichtigte Test und die Tests aussagekräftig sind.

Unbeaufsichtigte Eignungstests – gekommen, um zu bleiben

Aufgrund ihrer in Summe unbestreitbaren Vorteile gehören unbeaufsichtigte Leistungstests in den (digitalisierten) Methodenkoffer der Personalauswahl. Sie eignen sich besonders in Auswahlverfahren mit einem hohen Aufkommen an Bewerbenden zur Vorselektion, sofern ihre Schwachstellen ausreichend berücksichtigt werden. Notwendig sind vor allem Maßnahmen zur Prävention von Täuschungsverhalten, eine zurückhaltende Interpretation sowie die Verifikation der Ergebnisse. ■

in den besten Händen

Die Aufgaben in den Personalabteilungen werden immer komplexer. Unsere Software für die Personalabrechnung unterstützt Sie mit effizienten Workflows und nimmt Ihnen zeitraubende Routineabläufe ab – schnell, komfortabel und korrekt.

- | | | |
|------------------------|--------------|-----------------------|
| • Leicht Bedienbarkeit | • Effizienz | • Zuverlässigkeit |
| • Flexibilität | • Sicherheit | • Hohe Funktionalität |

spdata GmbH & Co. KG
05221 9140 444 | info@spdata.de | www.spdata.de



spdata
software für ihr personal