



Die Kunst der Bewertung

Psychometrische Validierung Central Test

Psychometrische Eigenschaften der Tests

Die regelmäßige Evaluierung der Validität und die Anpassung der Tests sind wesentliche Kriterien zum Nachweis ihrer Effektivität und ihres Nutzens. Central Test widmet diesen Kriterien besondere Aufmerksamkeit und überprüft regelmäßig die psychometrischen Eigenschaften aller verwendeten Instrumente. Diese Überprüfungen implizieren eine Aktualisierung des Inhalts, der statistischen Analysen und der Validitätsstudien. Vor seiner Online-Veröffentlichung wird ein Test nach den vier unten stehenden Kriterien evaluiert.

Zuverlässigkeit oder Reliabilität des Tests

Die Reliabilität bezieht sich auf die Art und Weise, wie ein Test eine gegebene Dimension misst (Stabilität der Ergebnisse, Kohärenz des Maßes). Wenn eine Person den Test erneut bearbeitet, hat sie dasselbe oder ein ganz unterschiedliches Ergebnis? Ein Test, der ähnliche Ergebnisse liefert, wenn eine Person ihn mehrmals absolviert, wird als zuverlässig (reliabel) betrachtet. Es ist jedoch festzuhalten, dass das Maß in den Verhaltenswissenschaften von externen Variablen beeinflusst wird, die sind:

- **Der psychologische oder physische Zustand der Person, die den Test absolviert, zu dem Zeitpunkt, zu dem sie ihn absolviert.** Zum Beispiel können verschiedene Grade von Erschöpfung, Angst oder Motivation die Leistungsfähigkeit des Individuums und somit seine Ergebnisse beeinflussen.

- **Die Umweltfaktoren.** Unterschiede in der Umgebung, in der der Test absolviert wird, wie Licht, Lärm usw., können die Leistungsfähigkeit eines Individuums bei einem Test beeinflussen.

Diese Faktoren werden allgemein als Quellen für „zufällige Messfehler“ bezeichnet. Gäbe es keine zufälligen Messfehler, hätte ein Individuum bei jedem Bearbeiten dasselbe Testergebnis,

sein „wahres Ergebnis“, ungeachtet der Bearbeitungsbedingungen. Ein Indiz für die Zuverlässigkeit des Tests ist der Anteil des Ergebnisses am Test, der nicht durch zufällige Messfehler beeinflusst ist.

Es gibt verschiedene Typen zur Beurteilung der Reliabilität, wobei jeder Typ von verschiedenen Quellen für Messfehler beeinflusst wird. Für die von Central Test entwickelten Tools sind zwei Arten der Reliabilität besonders wichtig:

- **Die interne Konsistenz.** Sie gibt an, in welchem Maße die Items eines Tests dieselbe Dimension messen. Ein erhöhter Koeffizient der internen Konsistenz gibt an, dass die Items des Tests in ihrem Inhalt ähnlich (das heißt homogen) sind. Es ist wichtig anzumerken, dass die Länge des Tests die interne Konsistenz beeinflussen kann. Zum Beispiel kann ein sehr langer Test den Reliabilitätskoeffizienten fehlerhaft erhöhen. Die interne Konsistenz wird im Allgemeinen mit Cronbachs Alpha gemessen, das zwischen 0 (schwach) und 1 (erhöht) variiert.

- **Die Test-Retest-Reliabilität.** Sie gibt den Grad der Wiederholbarkeit der Testergebnisse in der Zeit an. Diese Einschätzung widerspiegelt die Stabilität der Merkmale oder des Konstrukts, die mit dem Test beurteilt werden. Einige Konstrukte sind stabiler als andere. Zum Beispiel ist die Lesefähigkeit eines Individuums über einen gleichen Zeitraum hinweg stabiler als sein Grad der Angst. Für die Konstrukte, die in der Zeit variieren können, kann ein akzeptabler Reliabilitätskoeffizient eines Test-Retests schwächer sein als der der stabilen Konstrukte in der Zeit.

Der akzeptable Reliabilitätswert hängt von der Art des Tests und von der Art des verwendeten Koeffizienten ab. Für die Werkzeuge von Central Test werden die Reliabilitätsmaße in den technischen Dokumenten der Tests bereitgestellt.

Validität des Tests

Die Validität ist das wichtigste Kriterium bei der Auswahl eines Tests. Sie bezieht sich auf die von dem Test *gemessenen Merkmale* und auf die *Genauigkeit*, mit der er diese Merkmale evaluiert werden.

Die Validität gibt den Testergebnissen Sinn. Die Validitätsbeweise geben an, dass eine Verbindung zwischen der Leistungsfähigkeit bei dem Test und einem Leistungsfähigkeitskriterium besteht.

Sie gibt uns Auskunft über den Grad, mit dem es möglich ist, spezifische Schlussfolgerungen oder Prognosen über die Person in Abhängigkeit von ihrem Testergebnis zu erstellen. Mit anderen Worten, sie gibt Auskunft über den Nutzen des Tests. Wenn zum Beispiel ein Test eine prognostische Validierung der Leistungsfähigkeit für eine bestimmte Arbeitsstelle ist, können wir

schlussfolgern, dass eine Person, die ein erhöhtes Testergebnis vorweist, dazu neigt, kompetenter für diese Stelle zu sein, als eine Person, die ein schwaches Ergebnis erzielt hat, sofern alle übrigen Dinge gleich sind.

Es ist wichtig, den Unterschied zwischen der Reliabilität und der Validität zu verstehen. Die Validität sagt aus, in welchem Maße der Test zur Evaluation einer besonderen Situation geeignet ist; die Reliabilität sagt aus, in welchem Maße dieser Test glaubwürdig und stabil ist. Die Validität eines Tests wird in Bezug auf ein besonderes Ziel hergestellt.

Es gibt verschiedene Arten von Validität, die von den Wissenschaftlern definiert werden. Alle können jedoch nach drei Kategorien klassifiziert werden:

- **Die Kriteriumsvalidität.** Sie wird evaluiert durch Überprüfung der Korrelation oder einer anderen statistischen Beziehung zwischen der Leistungsfähigkeit bei dem Test und einem externen Leistungsfähigkeitskriterium. Mit anderen Worten tendieren die Individuen, die ein hohes Testergebnis aufweisen, dahin, leistungsfähiger in Bezug auf dieses externe Kriterium zu sein, als die Individuen, die ein schwaches Testergebnis haben. Wenn das Kriterium gleichzeitig mit dem Absolvieren des Tests erzielt wird, spricht man von konvergenter Validität; wenn das Kriterium später erzielt wird, spricht man von prädikativer Validität.

- **Die Inhaltsvalidität.** Sie wird untersucht, indem man überprüft, ob der Inhalt des Tests für das Konstrukt, das man messen will, repräsentativ ist. Zum Beispiel müssen im Rahmen eines Tests zur Evaluation des kaufmännischen Potentials die Items des Tests zutreffend sein und die Erfordernisse und Qualifikationen für die Stelle des kaufmännischen Angestellten direkt beurteilen. Der Inhalt des Tests muss auch dem Verstehensniveau der Zielpopulation entsprechen.

- **Die Konstruktvalidität.** Sie erfordert den Beweis dafür, dass der Test wirklich das Konstrukt oder das Merkmal misst, das er zu messen vorgibt. Ein Test über das Interesse an der Arbeit zum Beispiel muss das Interesse an der Arbeit messen und nicht den Grad der Motivation. Die Konstruktvalidität wird im Allgemeinen auf zwei verschiedene Arten evaluiert:

- o Indem man Fachleute bittet, die Übereinstimmung zwischen den Testitems (dem Inhalt) und dem evaluierten Konstrukt (**Augenscheinvalidität**) zu beurteilen;
- o Indem man den Test mit anderen Tests anwendet, die zu theoretisch ähnlichen Konstrukten entwickelt wurden, und indem man die Korrelation zwischen beiden (**konvergente Validität**) überprüft oder indem man den Test gleichzeitig mit theoretisch entgegengesetzten Tests einsetzt und ihre Korrelation (**diskriminante Validität**) untersucht.

Die drei Typen der Validität, die Kriteriums-, Inhalts- und Konstruktvalidität, werden je nach der Situation als Validierungsinstrumente verwendet. Im Allgemeinen überschneiden sich diese drei Methoden und, je nach Situation, sind nicht alle angebracht.

Bei Central Test werden die Studien über die Validität unserer Tests intern sowie bei Experten durchgeführt. Die Zusammenfassungen dieser Studien werden in den technischen Dokumenten der Tests bereitgestellt und für die ständige Verbesserung der Tests genutzt.

Soziale Erwünschtheit

Die soziale Erwünschtheit ist eine Verzerrung bei der Evaluation, die der Neigung einer Person entspricht, auf eine Art zu antworten, die von anderen als günstig angesehen wird. Da die Individuen von Natur aus soziale Wesen sind, sind sie geneigt, einen gewissen Grad an sozialer Akzeptanz zu suchen. Die soziale Erwünschtheit ist eine der wichtigsten Verzerrungen in den Persönlichkeitstests. Sie tritt im Kontext der Personalrekrutierung noch deutlicher hervor. Ein Kandidat ist normalerweise immer versucht, so zu antworten, dass er den Personalverantwortlichen positiv beeinflusst.

Wir haben eine Zwei-Punkte-Strategie entwickelt, um das Problem zu lösen. Erstens werden alle Items überprüft, um sicher zu gehen, dass sie nicht auf die soziale Erwünschtheit ansprechen. Zweitens schließen wir in den Tests eine Skala der sozialen Erwünschtheit ein, um das „Mogeln“ zu erkennen und zwar ungeachtet dessen, ob es absichtlich geschieht oder nicht.

Soziolegale Betrachtungen

Schließlich darf ein guter psychometrischer Test nicht gegenüber einer besonderen soziokulturellen Gruppe (positiv oder negativ) verzerrt sein. Er darf keine Diskriminierung in Bezug auf Religion, Geschlecht, ethnische Herkunft oder Kultur der Person zeigen, die den Test absolviert.

Bei der Entwicklung unserer Instrumente werden alle Items untersucht, so dass sie nicht eine bestimmte Population begünstigen. In den Studien zur Validierung werden die Beziehungen zwischen den demographischen Variablen und den Testergebnissen überprüft, um eine potentielle Verzerrung zu vermeiden. Damit erstellen wir sensible aber keine verzerrten Tests.