



WISSENSCHAFTLICHE GRUNDLAGE FÜR ADAPTIVES TESTEN

Was sind Tests?

Tests sind standardisierte Verfahren zur Befragung oder Beobachtung mit dem Ziel, Leistungen oder Persönlichkeitsmerkmale mit Hilfe von numerischen Skalen zu beschreiben. Standardisiert bedeutet, dass die Regeln der Durchführung und Auswertung genau festgehalten sind, so dass Testergebnisse, die zu unterschiedlichen Zeiten und an unterschiedlichen Orten entstanden sind, vollständig vergleichbar sind¹.

Was zeichnet konventionelle Tests aus?

Mit konventionellen Tests, meist Papier-Bleistift-Verfahren, werden vielfach allen Personen die gleichen Aufgaben vorgelegt. Ein solches Verfahren ist in vielen Fällen als ineffizient zu bezeichnen. Schwierige Aufgaben einer Person mit geringen Fähigkeiten vorzulegen, kann als Zeitverschwendung bezeichnet werden, die meist noch mit einer Frustration der Person verbunden ist. Umgekehrt langweilen sich Personen mit sehr grossen Fähigkeiten, wenn sie einfache Aufgaben bearbeiten müssen.

Weshalb computergestützte adaptive Testsysteme?

Computergestützte adaptive Testsysteme passen sich den Fähigkeiten einer Person an, indem die Auswahl der Testaufgaben durch einen Algorithmus so gesteuert wird, dass jeweils die Schwierigkeit einer Aufgabe möglichst nahe bei der geschätzten Fähigkeit liegt, die fortwährend aufgrund des Lösungsverhaltens der Person neu berechnet wird. Sobald sich bei der Berechnung der Fähigkeit der Person nahezu keine Änderungen mehr einstellen, wird der Test abgebrochen und das Testergebnis liegt unmittelbar vor².

Welche Art von Testaufgaben eignen sich für adaptive Tests?

Weil vom Computer umgehend als richtig oder falsch codiert werden müssen, eignen sich für adaptive Tests Aufgaben im Multiple-Choice-Format besonders gut. Es werden aber auch offene Aufgaben eingesetzt, deren Lösungen eine kurze schriftliche Antwort verlangen und vom Computer einwandfrei korrigiert werden können.

¹ Cronbach, Lee J: Essentials of Psychological Testing. New York: HarperCollinsPublisher 1990

² Kubinger, Klaus D.: Adaptives Testen. In Klaus D. Kubinger/Reinhold S. Jäger (Hrsg.), Schlüsselbegriffe der Psychologischen Diagnostik. Weinheim, Basel, Berlin: Beltz 2003, S. 1–9

Testtheoretische Voraussetzungen für adaptives Testen

Die Anwendung von adaptiven Tests verlangt die Nutzung der Vorteile der Item Response Theorie, speziell des Raschmodells³. Das Rasch-Modell geht von einer probabilistischen Beziehung zwischen dem Antwortverhalten einer Person, der Fähigkeit der Person und der Aufgabenschwierigkeit aus. Das bedeutet beispielsweise, dass die Lösungswahrscheinlichkeit für eine Person mit steigender Aufgabenschwierigkeit abnimmt. Bei gleicher Aufgabenschwierigkeit steigt die Lösungswahrscheinlichkeit in Abhängigkeit von der Fähigkeit einer Person.

Bevor die Aufgaben (Items) für ein adaptives Testsystem genutzt werden können, muss sichergestellt werden, dass nur solche Aufgaben eingesetzt werden, für die das Rasch-Modell Gültigkeit hat. Zudem müssen die Schwierigkeitsparameter der Aufgaben kalibriert werden. Die kalibrierten Aufgaben werden in einer Itembank zusammengefasst. Wenn einmal eine Itembank mit kalibrierten Aufgaben erstellt wurde, kann diese während des adaptiven Testens laufend mit neuen Aufgaben erweitert werden.

Der entscheidende Vorteil des Rasch-Modells für das adaptive Testen liegt darin, dass in Kenntnis der kalibrierten Schwierigkeitsparameter der Aufgaben und der Anzahl richtig gelöster Aufgaben für jede Person die Fähigkeit geschätzt werden kann, und zwar unabhängig davon, welche Aufgaben eine Person bearbeitet hat. Das heisst, dass die Fähigkeiten zweier Personen auf der gleichen Skala lokalisiert werden können, obwohl sie eine unterschiedliche Auswahl von Testaufgaben bearbeitet haben.

Wie funktioniert der Algorithmus?

Ein adaptiver Test beginnt für alle Personen mit einer relativ einfachen, zufällig ausgewählten Aufgabe. Nachdem die Aufgabe gelöst wurde, schätzt das System aus dem Schwierigkeitsparameter der Aufgabe und der Lösung (richtig oder falsch) die Fähigkeit der Person (Personenparameter). Danach sucht das System jene Aufgabe, deren Schwierigkeitsparameter am nächsten bei der geschätzten Fähigkeit der Person beziehungsweise dem Personenparameter liegt. Löst beispielsweise eine Person alle Aufgaben von Beginn an richtig, dann schlägt sich dies in der Schätzung ihrer Fähigkeit nieder. Der Personenparameter wird grösser, und dementsprechend weist das System der Person schwierigere Aufgaben zu. Umgekehrt sinkt der Personenparameter, wenn die Person eine Aufgabe falsch löst. Das System weist der Person in diesem Fall Aufgaben zu, deren Schwierigkeitsparameter kleiner sind. Der Test dauert so lange, bis grössere Schwankungen bei der Schätzung der Fähigkeit ausbleiben und das System nur noch Aufgaben zuweist, deren Schwierigkeitsparameter sich kaum mehr von den geschätzten Personenparametern unterscheiden. Die letzte Schätzung des Personenparameters entspricht dem Gesamtwert im Test.

³ Das Rasch-Modell geht auf den dänischen Statistiker Georg Rasch (1901–1980) zurück. Van Linden, Wim. J./Hambleton, Ronald, K.: Handbook of Modern Item Response Theory. New York: Springer, 1997



Genauigkeit des Testergebnisses

Der Vorteil des computergestützten adaptiven Testens liegt darin, dass sie dank den technischen Hilfsmitteln in einfacher Weise standardisiert durchgeführt und ausgewertet werden. Die Zuverlässigkeit des Testergebnisses ist dadurch hoch. Der Test wird zudem erst dann abgebrochen, wenn der Vertrauensbereich des wahren Testergebnisses eine bestimmte Schwelle unterschritten hat. Die Genauigkeit der Messung wird auch deshalb erhöht, weil sich der Test den Fähigkeiten anpasst, was zur Folge hat, dass kaum Aufgaben gelöst werden, die viel zu schwierig oder viel zu einfach sind. Solche Aufgaben führen zu keinen brauchbaren Informationen.

Interpretation der Testergebnisse

Die Anwendung probabilistischer Testmodelle hat den Vorteil, dass die Testergebnisse mit Bezug zu einem Kriterium interpretiert werden können. Vorausgesetzt, der Aufbau des Tests richtet sich nach einem Kompetenzmodell, dann kann für jede Person festgestellt werden, auf welcher Kompetenzstufe sich die Person befindet. Beispiele von Tests, die sich nach einem Kompetenzmodell ausrichten, sind vor allem im Bereich des Fremdsprachenlernens üblich⁴.

Institut für Bildungsevaluation

Das Institut für Bildungsevaluation ist eine Einrichtung mit einem Dienstleistungsangebot für das Bildungswesen sowie für Wirtschaft und Verwaltung. Das Institut verbindet Forschende in einem Netzwerk, das sowohl Grundlagenforschung als auch anwendungsorientierte, projektbezogene Forschung in den Bereichen Bildungsevaluation und Leistungsmessung durchführt und Dienstleistungen für Auftraggeber erbringt.

Zürich, 8. Mai 2007 / PD Dr. Urs Moser

⁴ Ein Beispiel dafür bildet der Gemeinsamen Europäischen Referenzrahmen für Sprachen des Europarats (Common European Framework of Reference for Languages CEF), in dem die Fähigkeiten auf verschiedenen Niveaus von A1 bis zu C2 umschrieben sind. Unterschieden wird zwischen elementarer Sprachverwendung (A2: Key English Test), selbständige Sprachverwendung (B1: Preliminary English; B2: First Certificate) oder kompetenter Sprachverwendung (C1: Certificate in Advanced English; C2: Certificate of Proficiency).