



Universität Zürich
Institut für Bildungsevaluation

Institut für Bildungsevaluation
Assoziiertes Institut
der Universität Zürich

Onlinetests auf dem Bildungsserver?

Erfahrungen mit Test Your ICT-Knowledge auf dem Bildungsserver
«educa.ch»

Urs Moser / Florian Keller
Zürich, Juni 2007



Inhalt

1.	Einleitung.....	3
2.	Test Your ICT Knowledge und HarmoS.....	3
3.	Die Bedeutung von Onlinetests für das nationale Bildungsmonitoring.....	4
4.	Möglichkeiten von Onlinetests auf dem Bildungsserver.....	5
5.	Anforderungen an ein Onlinetestsystem.....	6
6.	Beurteilung der Möglichkeiten des Bildungsservers als Plattform für Onlinetests im Vergleich zu den gestellten Anforderungen.....	8
7.	Fazit.....	9
	Anhang 1: Glossar.....	11
	Anhang 2: Rasch-Modell «mathematisch».....	14



1. Einleitung

Im vorliegenden Papier sind die Erfahrungen zusammengefasst, die mit der Anwendung des Onlinetests «Test Your ICT-Knowledge» auf dem Bildungsserver «educa.ch» in den letzten zwei Jahren gemacht werden konnten. Test Your ICT-Knowledge ist ein Testsystem, das auf einem wissenschaftlich validierten Kompetenzmodell beruht. Im Kompetenzmodell werden auf unterschiedlichen Anforderungsniveaus Standards formuliert, die Schülerinnen und Schüler für einen sinnvollen Umgang mit Informations- und Kommunikationstechnologien am Ende der Primarschule beziehungsweise am Ende der Sekundarstufe I erreichen können. Mit verschiedenen Tests können die Kompetenzen in den drei Handlungsfeldern «Anwendung und Gestaltung, «Austausch und Vermittlung» sowie «Reflexion und Medienkritik» gemessen und die Testergebnisse mit Bezug zum Kompetenzmodell interpretiert werden.

Test Your ICT-Knowledge wurde im Rahmen des Projekts PPP SiN entwickelt und im November 2004 auf einem Server des Zentrums für Online Learning and Training der Universität Zürich (OLAT) aufgeschaltet. Anfangs 2005 wurden die Tests zusätzlich auf den Schweizer Bildungsserver portiert. Eine vollständig überarbeitete zweite Version von Test Your ICT-Knowledge steht seit Dezember 2006 auf dem Server von OLAT und seit Mai 2007 ebenfalls auf dem Bildungsserver zur Verfügung. Ab Oktober 2007 sollen die Tests nur noch auf dem Bildungsserver angeboten werden. Die Tests stehen damit weiterhin allen Schulen beziehungsweise allen Lehrpersonen der Schweiz kostenlos zur Verfügung.

2. Test Your ICT Knowledge und HarmoS

Die wichtigsten Elemente der standardbasierten Reform, wie sie im Rahmen des Projekts «Harmonisierung der obligatorischen Schule» (HarmoS) der Schweizerischen Konferenz der kantonalen Erziehungsdirektoren (EDK) umgesetzt werden sollen, sind [1] Bildungsstandards, die auf einem wissenschaftlich validierten Kompetenzmodell beruhen und die erwartete Kompetenzen klar umschreiben, [2] eine entsprechende Vermittlung und Förderung im Unterricht, [3] Tests, mit denen das Erreichen der Standards überprüft wird sowie [4] eine darauf abgestimmte Feedback-Kultur.

Bei der Entwicklung von Test Your ICT-Knowledge wurde exakt jenes methodische Vorgehen eingeschlagen, das auch bei der Entwicklung von Kompetenzmodellen und Bildungsstandards für das HarmoS-Projekt verfolgt wird. Zuerst wurde ein Kompetenzmodell entwickelt und in verschiedensten wissenschaftlichen und politischen Gremien diskutiert. Aufgrund des Modells wurden Testaufgaben entwickelt und bei einer grossen Stich-



probe kalibriert, das heisst, die Schwierigkeitsparameter der Aufgaben bestimmt. Auf der Basis der Kalibrierung der Aufgaben wurde das Kompetenzmodell validiert. Validierung und Anpassung des Kompetenzmodells erfolgten im Rahmen von Test Your ICT-Knowledge zwei Mal (2004 und 2006). Zumindest für die Deutschschweiz und für die 6. und 9. Klasse der obligatorischen Schulbildung stehen ein empirisch validiertes Kompetenzmodell sowie eine ausgesprochen grosse Itembank mit kalibrierten Testaufgaben zur Erfassung der ICT-Kompetenzen zur Verfügung.

Mit der Entwicklung von Test Your ICT-Knowledge liefert das Institut für Bildungsevaluation mit Unterstützung von PPP-SiN ein sozusagen fertiges Testsystem für den Einsatz in der Schule. Zugleich sind mit der Portierung der Tests auf den Bildungsserver Erfahrungen über den Einsatz von Onlinetests auf einem nationalen Portal zur Bildung gemacht worden, die im Hinblick auf die standardbasierte Reform in der Schweiz genutzt und im vorliegenden Papier kommentiert werden sollen.

3. Die Bedeutung von Onlinetests für das nationale Bildungsmonitoring

Das Projekt HarmoS entwickelt national verbindliche Bildungsstandards. Die grosse Herausforderung für diese standardbasierte Reform wird es sein, dass die nationalen Bildungsstandards zuverlässig und in allen Kantonen in gleicher Weise – anhand eines einheitlichen Massstabes beziehungsweise mit «gleicher Metrik» – überprüft werden.

Im Vergleich zu den meisten gängigen Schulleistungstests können Tests zur Überprüfung von Standards nicht unabhängig voneinander entwickelt werden. Für die Überprüfung von Standards ist es ausserordentlich wichtig, dass die Tests immer auf der gleichen Skala beruhen. Beispiele von Test- oder Monitoringsystemen, die dieses Kriterium erfüllen, sind das National Assessment for Educational Progress (NAEP) oder der Scholastic Aptitude Test (SAT), aber auch Tests, die zur Beurteilung der Fremdsprachenkompetenzen eingesetzt werden und sich am Gemeinsamen Europäischen Referenzrahmen für Sprachen des Europarats GER (Common European Framework of Reference for Languages CEF) orientieren. Auch die PISA-Tests führen zu Ergebnissen, die sich immer wieder auf der gleichen Skala abbilden lassen.

Eine notwendige Voraussetzung dafür, dass mit Tests die Qualität des Bildungssystems im Sinne der standardbasierten Reform gesichert und entwickelt werden kann, liegt deshalb darin, dass Testergebnisse einen empirisch nachgewiesenen Bezug zu den validierten Kompetenzmodellen beziehungsweise zu den Standards haben und dementsprechend auf der gleichen Skala dargestellt werden müssen. Die Nutzung einer einheitlichen Skala verlangt nicht, dass immer identische Tests eingesetzt werden, sondern dass unterschiedliche Tests durch erprobte und kalibrierte Testaufgaben («Link-



Items») miteinander verbunden sind. Die Verbindung verschiedener Tests durch so genannte Link-Items ist anspruchsvoll und verlangt zudem, dass zumindest ein Teil der Testaufgaben vertraulich behandelt wird.

Beides – die Verbindung verschiedener Tests durch kalibrierte Aufgaben zur Sicherung einer einheitlichen Skala als auch die vertrauliche Behandlung von Testaufgaben – ist vor allem dann aufwändig und schwierig zu sicherzustellen, wenn die Tests in gedruckter Version als Paper-and-Pencil-Tests durchgeführt werden. Onlinetests haben bei der Erfüllung dieser Anforderungen gegenüber Paper-and-Pencil-Tests zwei entscheidende Vorteile. Dank einer zentralen Steuerung der Tests beziehungsweise einer Itembank, die eine grosse Menge an kalibrierten Testaufgaben enthält, kann die einheitliche Skala ohne Zusatzaufwand garantiert werden. Eine Itembank kann zudem in einfacher Weise und ebenfalls elektronisch fortlaufend erweitert werden. Gesteuert durch einen Algorithmus können neue Testaufgaben fortlaufend ins bewährte Testsystem integriert und kalibriert werden. Der zweite Vorteil liegt darin, dass die Testaufgaben – ebenfalls ohne zusätzlichen Aufwand – relativ einfach vertraulich behandelt werden können. Die Testaufgaben erscheinen nur am Bildschirm und können im Vergleich zu Paper-and-Pencil-Tests weniger gut kopiert und gestreut werden.

Onlinetest garantieren folglich auf einfache Weise, dass eine Skala konstant bleibt und die Aufgaben vertraulich behandelt werden können. Müssen diese beiden Anforderungen mit Paper-and-Pencil-Tests erfüllt werden, dann steigt der Aufwand deutlich.

4. Möglichkeiten von Onlinetests auf dem Bildungsserver

Die Onlinetests von Test Your ICT-Knowledge laufen seit Mai 2007 auf dem Bildungsserver als statische Tests. Sie ermöglichen den Schülerinnen und Schülern eine individuelle Beurteilung der ICT-Kompetenzen in den drei Handlungsfeldern «Anwenden und Gestalten», «Austausch und Vermittlung» sowie «Reflexion und Medienkritik». Die Beurteilung erfolgt mit Bezug zu drei Kompetenzniveaus, die im Kompetenzmodell umschrieben sind. Ein Kompetenzniveau (entsprechend einem Standard) ist dann erreicht, wenn ein Schüler mindestens 50 Prozent aller Aufgaben eines Niveaus richtig löst. Das Kompetenzmodell ermöglicht eine förderorientierte Beurteilung, die aufzeigt, was die Schülerinnen und Schüler bereits wissen und können und welche Ziele sie als nächstes anzustreben haben.

Alle Institutionen und Lehrpersonen, die bei educanet angemeldet sind, haben Zugang zu den Tests. Die Tests sind persönlich authentifiziert. Die Zugangsberechtigungen werden den Schülerinnen und Schülern von den Lehrpersonen zugewiesen. Danach lösen die Schülerinnen und Schüler die Tests individuell und online am Computer. Am Ende des Tests erhalten die Schülerinnen und Schüler online ein Feedback über ihre vorhandenen ICT-



Kompetenzen. Die individuellen Testergebnisse lassen sich ausdrucken, jedoch nicht das Ergebnis der Klasse als Überblick der Ergebnisse aller Schülerinnen und Schüler. Die Auswertung der Klasse muss von der Lehrperson manuell erstellt werden. Im Vordergrund steht die Interpretation der Testergebnisse mit Kompetenzen gemäss Modell und nicht der soziale Vergleich.

Der Support der User wird vom Schweizer Medieninstitut für Bildung und Kultur (educa.ch) und dem Institut für Bildungsevaluation der Universität Zürich (IBE) gemeinsam geleistet. Das IBE beantwortet Anfragen betreffend die Inhalte der Tests, die psychometrischen Grundlagen und die Interpretation der Testergebnisse. educa.ch beantwortet Anfragen betreffend die Funktionsweise der Test auf dem Bildungsserver. Der Support für die ausgebildeten Kursleiter muss hingegen für die Zukunft neu geregelt werden.

Auch inhaltliche Aktualisierungen werden von educa.ch und vom IBE zurzeit noch gemeinsam geleistet. Kleinere redaktionelle Anpassungen bei den einführenden Texten oder in einzelnen nicht flash-basierten Aufgaben werden nach Absprache mit dem IBE durch educa.ch vorgenommen. Für grössere Anpassungen (Flash-basierte Aufgaben, Entwicklung und Kalibrierung neuer Aufgaben, vollständige Updates) sind jedoch zurzeit noch keine finanziellen Mittel in Aussicht gestellt worden.

5. Anforderungen an ein Onlinetestsystem

Zwischen den bisherigen Möglichkeiten, Onlinetests auf dem Bildungsserver anzubieten und den Anforderungen an ein Onlinetestsystem, das zum einen für das nationale Bildungsmonitoring genutzt werden kann, zum andern aber auch für eine individuelle Zertifizierung, beispielsweise am Ende der Volksschule oder am Ende der Berufsbildung, besteht eine gewisse Diskrepanz. So, wie Onlinetests auf dem Bildungsserver integriert sind, stehen verschiedene Möglichkeiten, die ein modernes Testsystem anzubieten hat, nicht zur Verfügung. Im Folgenden werden deshalb Anforderungen beschrieben, die ein modernes Onlinetestsystem für die Überprüfung von Standards, aber auch für das Bildungsmonitoring oder die individuelle Zertifizierung erfüllen sollte.

1. Onlinetests müssen so organisiert sein, dass sie auf eine Itembank zugreifen, die kalibrierte Testaufgaben enthält. Je flexibler dieser Zugriff ist (zufällige Auswahl von Testaufgaben, gesteuerte Auswahl von Testaufgaben), desto wertvoller wird ein Testsystem. Im Rahmen von HarMoS werden tausende von Testaufgaben kalibriert. Die Testaufgaben zielen auf die Fähigkeiten am Ende der Primarschulstufe sowie am Ende der Sekundarstufe I. Mit Hilfe dieser kalibrierten Aufgaben wäre es möglich, die national festgelegte Skala auch in Zukunft einzusetzen und



einfach zu überprüfen, ob Standards erfüllt wurden beziehungsweise welchem Niveau die geschätzten Kompetenzen entsprechen.

2. Idealerweise werden Onlinetests nicht statisch, sondern dynamisch gestaltet. Bei computergestützten adaptiven Testsystemen passt sich der Test 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. Weil bei diesem Prozess die Lösungen 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. Adaptive Testsysteme sind effizient und kostengünstig, weil Zeit – sowohl für das Testen als auch für die Organisation des Testens – gespart werden kann.
3. Sowohl die Schätzung der Testergebnisse als auch die Kalibrierung von Testaufgaben erfolgt über einen Algorithmus. Bei der Schätzung der Fähigkeiten eines Schülers werden aufgrund der Lösung, die der Schüler bei einer Aufgabe gibt, und aufgrund des Schwierigkeitsgrads der Aufgabe die Fähigkeiten nach dem Raschmodell geschätzt (Anhang 2). Der Algorithmus berechnet die Fähigkeiten und sucht danach jene Aufgabe im System, deren Schwierigkeit den Fähigkeiten am besten entspricht. Dieser Prozess wird so lange fortgesetzt, bis es nahezu keine Schwankungen mehr zwischen Fähigkeit und Aufgabenschwierigkeit gibt. Die Kalibrierung neuer Testaufgaben funktioniert genau umgekehrt. Aufgrund der Fähigkeiten der Schülerinnen und Schüler wird die Aufgabenschwierigkeit immer wieder neu geschätzt. Die Kalibrierung ist ebenfalls dann abgeschlossen, wenn es keine Schwankungen mehr zwischen geschätzter Aufgabenschwierigkeit und Fähigkeit gibt.
4. Onlinetests müssen ständig erweitert werden. Dies ist am einfachsten und ökonomischsten, wenn neue Aufgaben in eine bestehende Itembank aufgenommen und in die Tests integriert werden. Bevor die neuen Aufgaben für die Schätzung der Kompetenzen beziehungsweise für die Ergebnismeldungen genutzt werden, müssen sie direkt im Testsystem kalibriert werden.
5. Onlinetest müssen sicher sein und einen reibungslosen Ablauf ermöglichen. Dazu gehört es beispielsweise, dass bei einem Systemabsturz oder bei einem ungewollten Verlassen der Testumgebung der User wieder dort im Test weiterarbeiten kann, wo der Test unterbrochen wurde. Das heisst, das System speichert den Testverlauf und ermöglicht es, den Test zu unterbrechen.



6. Onlinetest müssen skalierbar sein, das heisst auch bei stark steigender Nutzung über jene Server- und Prozessorleistungen verfügen, die einen uneingeschränkten und störungsfreien Betrieb garantieren. Tests werden in der Regel am Morgen gelöst, meist zwischen Dienstag und Donnerstag. Dies hat zur Folge, dass das ein Onlinetestsystem sehr unterschiedlich ausgelastet ist, die Server- und Prozessorleistung sich aber immer am «worst case» beziehungsweise an einer maximalen gleichzeitigen Nutzung richten muss.
7. Die Ergebnisse von Onlinetests müssen nicht nur in Form einer individuellen Ergebnisrückmeldung genutzt, sondern für verschiedene Ebenen des Bildungssystems (Klasse, Schule, System) aggregiert und dargestellt werden können. Onlinetests haben sowohl für die individuelle Zertifizierung als auch für das Bildungsmonitoring eine grosse Bedeutung, weil sie ökonomisch einsetzbar sind. Dazu müssen die Testergebnisse in einer Datenbank gespeichert sein.
8. Onlinetests werden idealerweise mit einer Lernumgebung verknüpft. Das heisst, dass ein spezifisches Förder- und Lehr-Lern-Angebot gleich mit dem Testsystem verbunden ist. Die Integration von Tests auf dem Bildungsserver macht deshalb Sinn.
9. Um die Testergebnisse für wissenschaftliche Zwecke zu nutzen, sollten im Testsystem weitere Daten erhoben werden. Beispielsweise sind Schultyp, Schuljahr und Kanton sowie das Geschlecht und die Erstsprache der Kinder zusätzlich zu erfassen.

6. Beurteilung der Möglichkeiten des Bildungsservers als Plattform für Onlinetests im Vergleich zu den gestellten Anforderungen

In Anbetracht der nationalen Ausrichtung der Überprüfung von Basiskompetenzen bietet sich der Schweizer Bildungsserver als nationales Portal zur Bildung als geeignete Plattform für Onlinetests grundsätzlich an. Der Bildungsserver orientiert sich an nationalen wie kantonalen Interessen und wird von Bund und Kantonen gemeinsam getragen. Die Erfahrungen zeigen aber, dass der Bildungsserver und educa.ch mit den bestehenden Möglichkeiten weit davon entfernt sind, eine führende Rolle im Rahmen der standardbasierten Reform zu spielen.

Das Problem ist vorab ein technisches und liegt darin begründet, dass die Arbeits- und Kommunikationsplattform des Bildungsservers, educanet², auf einem Content Management System (CMS) basiert, das in erster Linie für Informations- und Kommunikationszwecke, für das Erstellen, Bearbeiten und Austauschen von Text- und Multimediadokumenten, nicht aber für das Durchführen von adaptiven Leistungstests entwickelt worden ist. Eine Realisierung von Onlinetests ist deshalb im Rahmen von educanet² zur Zeit nur beschränkt möglich.

Für die Zukunft sind grundsätzlich die folgenden zwei Szenarien denkbar:

1. Die gesamte Abwicklung der Tests erfolgt extern auf dem Server eines professionellen Testanbieters. Dabei wird educanet² von den Schulen als reines Zugangsportale zur Authentifizierung und Autorisierung der Testkandidaten genutzt.

In der Schweiz betreibt zurzeit der Lehrmittelverlag des Kantons St. Gallen ein Testsystem, das adaptive Tests für die Volksschule anbietet. Es wäre deshalb denkbar, vorab eine Anbindung dieses Testsystems über eine AAI Schnittstelle an den Bildungsserver zu schaffen, so dass vor allem jene Tests zentral zu verwaltet und durchgeführt werden könnten, die im Rahmen eines nationalen Bildungsmonitorings eingesetzt werden sollten.

2. Ein datenbankbasiertes adaptives Testsystem eines professionellen Testanbieters wird in educanet² integriert. Diese Option bedingt technische Abklärungen und es kann nicht vorab garantiert werden, dass alle oben im Abschnitt 5 erwähnten Anforderungen an ein Onlinetestsystem realisiert werden können. Das IBE erklärt sich bereit, den Programmcode eines adaptiven Testsystems eines professionellen Anbieters für technische Abklärungen zur Realisierbarkeit einer Schnittstelle zu educanet² dem Bildungsserver zur Verfügung zu stellen.

7. Fazit

Das Bedürfnis nach Tests, auch nach adaptiven Tests ist in der Schweiz zurzeit gross. In Kenntnis dieser Tatsache können für Test Your ICT Knowledge auf dem Bildungsserver folgende Schlüsse gezogen werden.

1. Obwohl bei der Entwicklung von Test Your ICT Knowledge jenes methodische Vorgehen gewählt wurde, das auch bei der Entwicklung nationaler Bildungsstandards angewendet wird, scheint das Interesse an den Tests und den validierten Kompetenzmodellen gering zu sein. Die EDK ist allerdings bereit, nach Abschluss der Entwicklung nationaler Bildungsstandards für die Fachbereiche Mathematik, Naturwissenschaften sowie Schulsprache und Fremdsprachen auf die Vorarbeiten von Test Your ICT Knowledge aufzubauen.
2. Die Möglichkeiten zur Nutzung des Bildungsservers für Onlinetests sind beschränkt. Zurzeit lassen sich keine weiteren Optionen, wie sie etwa für das adaptive Testen notwendig wären, nutzen. Schon eine Zufallsauswahl von Aufgaben ist zurzeit nicht möglich. Eine nahe liegende Lösung wäre es, den Bildungsserver als nationales Portal für das Testen durch eine Anbindung an ein bereits funktionierendes Testsystem zu nutzen.



3. Für den Fall einer beabsichtigten Nutzung des Bildungsservers als nationales Portal für Onlinetests, die den Anforderungen moderner Testsysteme genügen, müssten grössere finanzielle Ressourcen zur Verfügung gestellt werden.



Anhang 1: Glossar

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 Paper-and-Pencil-Tests, werden häufig 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 die Testaufgaben vom Computer umgehend als richtig oder falsch codiert werden müssen, eignen sich für adaptive Tests Aufgaben im Multiple-Choice-Format besonders gut. Es können aber auch offene Aufgaben eingesetzt werden, 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 (Anhang 2).

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.

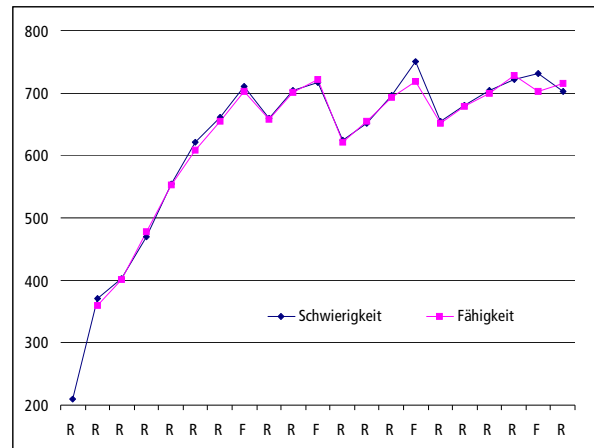
Wie funktioniert der Algorithmus?

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 Perso-

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.

nenparametern unterscheiden. Die letzte Schätzung des Personenparameters entspricht dem Gesamtwert im Test.

Abbildung 1: Graphische Veranschaulichung der Funktionsweise des Algorithmus



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.

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).



Anhang 2: Rasch-Modell «mathematisch»

Der folgende Term entspricht dem Rasch-Modell und stellt die Beziehung zwischen der Lösungswahrscheinlichkeit p , der Aufgabenschwierigkeit δ bei Aufgabe i und der Fähigkeit der Person θ bei Person v mathematisch dar.

$$p(x_{vi}) = \frac{\exp(x_{vi}(\theta_v - \delta_i))}{1 + \exp(\theta_v - \delta_i)}$$

$$p(x_{vi}) = \frac{\exp(0)}{1 + \exp(0)} = \frac{1}{2} = 0.5$$

p = Lösungswahrscheinlichkeit

θ = Fähigkeit/Kompetenz der Person

δ = Schwierigkeit der Aufgabe

Die Lösungswahrscheinlichkeit beträgt $p=0.5$, wenn der Schwierigkeitsparameter δ und der Personenparameter θ identisch sind.