

Marcel RITZ, Aachen & Johanna HEITZER, Aachen

Konstruktion kompetenzorientierter mathematischer E-Tests mit fachdidaktischen und testtheoretischen Kriterien

Um zu untersuchen, wie effizient und präzise ein Test ausgewählte Kompetenzen erhebt, sind bestimmte Kriterien zu beachten. Im Folgenden wird eine Auswahl von diesen aus dem fachdidaktischen und dem testtheoretischen Bereich vorgestellt. Das Wissen und das Beispiel stammen aus den Arbeiten vom Projekt „Test-M“. (Informationen unter www.didaktik.matha.rwth-aachen.de/de/forschung/projekte/testM.html).

In diesem Beitrag werden Kompetenzen schlicht und pragmatisch als fachbezogene Fähigkeiten verstanden; denn in dieser Form kommen sie konkret (manifestiert) in der testtheoretischen Fundierung zum Tragen. Anhand dieser Fähigkeiten soll bei Tests eine relevante Differenzierung innerhalb der Zielgruppe in Bezug auf einen bestimmten Bereich im Bildungsgang möglich sein. Dadurch ist implizit gefordert, dass ein Mindestniveau an zugrundeliegenden Fähigkeiten vorauszusetzen ist.

Fachdidaktische und testtheoretische Kriterien

In diesem Kontext sollen Kriterien helfen, einen effizienten Test zu konstruieren. Dafür wird durch die Kriterien auf Ebene des Tests geprüft, welche Gesamtaussage über die Testperson erhoben werden soll, und auf Ebene der Aufgaben kontrolliert, welchen Anteil diese dazu beitragen.

Zunächst werden hier ausgewählte testtheoretische Kriterien vorgestellt. Hauptsächliches Ziel dieser ist es, durch den Test valide (präzise), reliable (verlässliche) und objektive (eindeutige) Aussagen über die Testteilnehmenden zu gewinnen. Daher beziehen sich die Kriterien überwiegend auf Rahmenbedingungen, die dem Test zugrunde liegen, und auf die Genauigkeit, mit der eine Kompetenz erfasst wird:

Gesamtbearbeitungsdauer: Ab einer bestimmten Bearbeitungszeit sinkt die Konzentration und Motivation der Testpersonen, abhängig von Alter und intrinsischem Interesse. So können beide Faktoren Fehler verursachen, die mit in das Testergebnis eingehen und die Messung verfälschen (Moosbrugger & Kelava, 2012, S.34–35). Auch wenn Studierenden Testzeiten von mehr als 45 Minuten zugemutet werden können, ist für ein fast gänzlich Vermeiden dieser Faktoren die Dauer auf höchstens 20 Minuten zu setzen. Die abgeschätzte Bearbeitungszeit sollte nur vorab mitgeteilt werden, wenn mindestens 80% der Testpersonen diese einhalten können, da sonst zu viele Personen durch gefühlten Zeitdruck flüchtiger arbeiten und dadurch die Messung verfälschen können (Bandalos, 2018, S. 58–59).

Aufgabenschwierigkeitsverteilung: Als Schwierigkeit wird die Wahrscheinlichkeit, mit der eine Person der Zielgruppe die Aufgabe löst, betrachtet. Löst eine Person eine Aufgabe, gibt dies in Zusammenhang mit der Schwierigkeit derselben einen Hinweis auf ihr Kompetenzniveau. Die Verteilung der Aufgabenschwierigkeiten legt fest, auf welchem Niveauspektrum differenziert werden kann. Wenn keine Niveaus fokussiert werden, empfiehlt es sich, eine symmetrische Verteilung von Aufgaben von 27% bis 79% Lösungswahrscheinlichkeit zu nutzen, um eine gleichmäßige Aussagekraft zu erhalten. Aufgaben außerhalb von diesem Prozentbereich differenzieren kaum und sind daher nur in begründeten Fällen zu nutzen (Feldt, 1993, S. 38).

Kleinschrittigkeit der Kompetenzerhebung: Die korrekte Bearbeitung einer Aufgabe zeigt die zugehörige Kompetenz. Umgekehrt kann durch eine falsche Bearbeitung das Fehlen dieser Kompetenz aber nur nachgewiesen werden, wenn dies die einzige relevante Erklärung für den Fehler ist. Für eine exakte Aussage müssen also beide Richtungen genutzt werden. Daher sollte pro zu beantwortender Teilaufgabe möglichst nur eine einzelne Kompetenz abgefragt und eindeutig bewertet werden.

Als nächstes werden ausgewählte fachdidaktische Kriterien vorgestellt. Hauptsächliches Ziel dieser ist es, sicherzustellen, dass ein Test die relevanten fachlichen Inhalte abdeckt und überprüfen kann, ob diese auf unterschiedlichen Stufen kognitiver Auseinandersetzung genutzt werden können:

Lernzieltaxonomien: Hiermit werden unterschiedliche Niveaus beschrieben, auf denen mit Inhalten gearbeitet werden kann. Diese sind in aufsteigender Reihenfolge das Wiedergeben oder Wiedererkennen von Wissen, das Zeigen von Verständnis durch die Formulierung in eigenen Worten, die Anwendung der Inhalte in neuen Situationen, die Analyse der Inhalte durch Erkennen und Zerlegen von Zusammenhängen, die Synthese in der Vernetzung mit anderen Sachverhalten und die Beurteilung des Sachverhalts in Bezug auf einen bestimmten Aspekt (Bloom et al., 1956). Insbesondere bei der Synthese und der Beurteilung benötigt die Bearbeitung nicht mehr nur den konkreten Sachverhalt, sondern auch die Fähigkeit, vom Gegenstand abstrahieren und auf andere Sachverhalte übertragen zu können.

Kriterien für leistungsüberprüfende Aufgaben: Nach Büchter (2006, S. 11–16) sollte sich bei der Konzeption auf die für das Verständnis zentrale Kerne der inhaltsbezogenen Kompetenzen konzentriert werden. Dabei soll die Abfrage einer Kompetenz nicht durch kompetenzunabhängige Aspekte, an denen die korrekte Bearbeitung scheitern könnte, überlagert werden. Auch sollte die Aufgabenstellung die Erwartungen an die Bearbeitung hinsichtlich Darstellung, Genauigkeit und Ausführlichkeit transparent machen, um unnötigen Zeitaufwand zu vermeiden, solange dadurch nicht Hinweise auf den

Lösungsweg gegeben werden. Zudem sind verfahrensorientierte Aufgaben zu vermeiden, da auswendig gelernte Inhalte diese bereits lösen können. Statt diesen sind „fast“ offene Aufgaben für Verstehensorientierung zu nutzen.

Das Spannungsfeld der Kriterien

Bei der Betrachtung der Kriterien beider Seiten fällt auf, dass die testtheoretischen Kriterien alle Kompetenzen als gleich wichtig und austauschbar wahrnehmen und davon ausgehen, dass die Länge und Komplexität der zugehörigen Aufgaben flexibel angepasst werden können. Aus fachdidaktischer Sicht hängt es pro Aufgabe dagegen von dem behandelten Inhalt ab, inwiefern dieser nicht schon inhärent nur mit einer bestimmten Bearbeitungsdauer und/oder mit einem bestimmten Anspruch verbunden ist. Weiter stehen Schwierigkeitsverteilung und Gesamtzeit den Lernzieltaxonomien gegenüber, da das Ideal, zu allen Inhalten auch tieferes Verständnis abzufragen, eine unverhältnismäßige Überzahl an schwierigen Aufgaben oder einen sehr langen Test ergeben würde. Dies führt dazu, dass sowohl die Anzahl von schwierigen oder längeren Aufgaben mit der dadurch beeinflussten Messgenauigkeit – im Allgemeinen und für bestimmte Schwierigkeitsniveaus –, als auch welche Kompetenzen durch die verfügbaren schwierigeren oder längeren Aufgaben erhoben werden, jeweils abgewogen werden muss.

Ebenfalls schwierig zu vereinen ist, dass die fachdidaktisch gewünschten verständnisorientierten Aufgaben Kompetenzen oft verwoben enthalten würden, während aber aus testtheoretischer Sicht solche Kompetenzen nur ungenau gemessen werden können. Umgekehrt sind für die eindeutige Erhebung einzelner Kompetenzen in mehrschrittigen Aufgaben Zwischenergebnisse abzufragen, aber der Anspruch, diese Schritte eigenständig zu erkennen, stark abgeschwächt wird. Bisweilen kann die Bewertung von Teilantworten oder Rechenwegen die Abfrage von Zwischenergebnissen ersetzen. Diese Spannung bedeutet, dass abgewogen werden muss, ob für eine Kompetenz die Messgenauigkeit oder das Nutzen einer komplexeren Aufgabe, die mehrere Anteile in einem umfasst, bevorzugt wird. Bezüglich E-Tests ist hier hinzuzufügen, dass die Möglichkeiten, mit denen Teilantworten beachtet und ausgewertet werden können, ebenfalls beeinflussen, inwiefern sich dieser „Tausch“ von Testeigenschaften lohnt.

Beispiele von Überarbeitungen nach den Kriterien

In der Aufgabe in Abb. 1 sollte zuerst die Scheitelpunktform und anschließend der Scheitelpunkt bestimmt werden. Dadurch waren die beiden Kompetenzen unmittelbar miteinander verbunden. In der Überarbeitung werden beide Kompetenzen nun unabhängig voneinander abgefragt. Aufgrund zeitlicher Kürzung des Tests wurde auf den nichtnormierten Teil verzichtet.

Parabeln

Bestimmen Sie den Scheitelpunkt und die Scheitelpunktform der folgenden Parabeln.
Ergänzen Sie die freien Felder.

$x^2 - 4x + 7$:

Scheitelpunkt (,) Scheitelpunktform · (+)² +

$2x^2 + 6x + 2$:

Scheitelpunkt (,) Scheitelpunktform · (+)² +

Aufgabe 5: Parabeln

Bestimmen Sie den Scheitelpunkt und die Scheitelpunktform der folgenden Parabeln.
Ergänzen Sie die freien Felder.

$(x - 2)^2 + 3$:

Scheitelpunkt: (,)

$x^2 + 4x + 8$:

Scheitelpunktform: (+)² +

Abb. 1: Eine Aufgabe mit unnötigerweise verbundenen Kompetenzen. Rechts neu.

In Abb. 2 wurden die Kompetenzen „Höhe bestimmen“ und „Formel nutzen und berechnen“ in einem Feld abgefragt. In der Überarbeitung hat die Höhe eine eigene Abfrage erhalten, jedoch hat dies den Lösungsweg geebnet.

Flächeninhalte		
Berechnen Sie die Flächeninhalte der folgenden Figuren. Ergänzen Sie die freien Felder.		
Figur	Bild	Flächeninhalt
Gleichschenkliges Dreieck mit Kantenlängen 8, 5, 5		<input type="text"/>

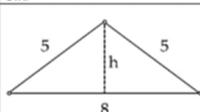
Aufgabe 3: Flächeninhalte		
Berechnen Sie die Flächeninhalte der folgenden Figuren. Geben Sie bei der ersten Figur zusätzlich die Länge der Strecke h an. Ergänzen Sie die freien Felder.		
Figur	Bild	Ergebnis
Gleichschenkliges Dreieck mit Kantenlängen 8cm, 5cm, 5cm		Höhe des Dreiecks: $h =$ <input type="text"/> cm Flächeninhalt des Dreiecks: <input type="text"/> cm ²

Abb. 2: Teil einer Aufgabe mit mehreren Kompetenzen in einer Abfrage. Rechts neu.

Fazit

Die Vereinigung der Kriterien beider Seiten ist für kürzere Reproduktions- und Anwendungsaufgaben mit klar zu bewertenden Antworten gut möglich und erlaubt in diesen die eindeutige Überprüfung von Kompetenzen. Für die aus fachdidaktischer Richtung gewünschte Abfrage von tiefergehendem Verständnis oder mehrschrittigen Aufgaben, müssen die Kriterien aber abgewogen werden, da keine optimale Umsetzung möglich ist. Die Auswertung von Teilantworten kann in diesen Fällen helfen, eine bessere Erfüllung der Kriterien erreichen, hat aber aufgrund des Aufwands und der eingeschränkten Umsetzbarkeit in E-Test-Formaten ebenfalls einen Preis.

Literatur

- Bloom, B. S., Krathwohl, D. R. & Masia, B. B. (1956). *Taxonomy of Educational Objectives*. Longmans.
- Büchter, A. (2006). *Verstehensorientierte Aufgaben als Kern einer neuen Kultur der Leistungsüberprüfung*. http://sinus-transfer.uni-bayreuth.de/fileadmin/MaterialienDB/427/Buechter_Modul_10.pdf
- Bandalos, D. L. (2018). *Measurement Theory and Applications for the Social Sciences*. Guilford Publications.
- Feldt, L. S. (1993). The Relationship Between the Distribution of Item Difficulties and Test Reliability. *Applied Measurement in Education*, 6(1), 37–48.
- Moosbrugger, H. & Kelava, A. (2012). *Testtheorie und Fragebogenkonstruktion* (Springer-Lehrbuch) (2. Aufl.). Springer.