

Christian DÜSI, Mosbach, Guido PINKERNELL, Heidelberg
& Gerhard GÖTZ, Mosbach

Ist der typische Fehler „Überlinearisierung“ systematisch? – Eine Modellierung als latente Variable von Distraktoren mit erhöhtem diagnostischem Potential

Durch die Nutzung von Vor-/Brückenkursen und die häufig damit verbundenen summativen Diagnoseverfahren entsteht vielfach die sog. „Testzeit-Inhalt-Problematik“ (Feldt-Caesar & Bruder, 2014). Im Folgenden wird ein Vorgehen vorgestellt, welches diagnostische Tiefe durch die Nutzung von Distraktoren mit erweitertem diagnostischen Potenzial (Winter, 2011) im Hinblick auf typische/systematische Fehler herstellt. Dieses Testverfahren wurde innerhalb eines Mathematikvorkursangebots mit einer Stichprobe von $N=813$ WiMINT-Studierenden validiert.

Testzeit-Inhalt-Problematik und summatives Testen im Übergang

Viele Hochschulen stellen sich, meist mit fachspezifischen Unterstützungsangeboten, den häufig wahrgenommenen Schwierigkeiten in der Transitionsphase von Schule zu Hochschule (z.B. Bausch et al., 2014). Mit diesen Unterstützungsangeboten entsteht das Desiderat zur theoretischen Reflexion summativer Referenzmodelle im Hinblick auf Wissen und Können am Ende der Sekundarstufen (z.B. Pinkernell, Düsi, & Vogel, 2017). Die Operationalisierung dieser Referenzmodelle führt allerdings notwendigerweise zu einem testtheoretischen Problem, welche als „Testzeit-Inhalt-Problematik“ beschrieben werden kann. Hierbei handelt es sich um die Problematik, dass die vollständige Erfassung der notwendigen Wissens- und Könnensaspekte einer zeitökonomischen Testung diametral entgegensteht.

Eine mögliche Lösung dieser Problematik wurde von Feldt-Caesar und Bruder 2014 vorgestellt. Dieses Testverfahren integriert die Vorteile des elementarisierenden Testens in ein Testverfahren mit komplexeren Aufgaben exemplarisch am Inhaltsbereich des funktionalen Zusammenhangs mit Fokus auf Differential-/Integralrechnung. Dieses Verfahren ermöglicht die Diagnose eines vergleichsweise breiten Inhaltsbereichs mit wenigen Items (wenn die Items der sog. Hauptlinie korrekt beantwortet werden). Die Schleifen, welche durchlaufen werden falls ein Item der Hauptlinie falsch beantwortet wird, elementarisieren die komplexeren Aufgaben fachinhaltlich in (theoretisch) notwendige inhaltliche Einzelschritte um zu analysieren, ob z.B. die fehlende Fähigkeit ein bestimmtes Integral zu berechnen im Bereich begrifflicher Unkenntnis, fehlender Methodenkenntnis o.ä. liegt.

Diagnostische Tiefe an Stelle von diagnostischer Breite

Die diagnostische Breite der Testinstrumente am Übergang zwischen Schule und Hochschule ist weitgehend diskutiert und (in Teilen) durch verschiedene adaptive Testverfahren gelöst. Die Lücke, welche allerdings in diesen Testverfahren übrig zu bleiben scheint, ist die tiefe der Diagnostik. Mit Breite der Diagnostik wird hier die fachinhaltlich möglichst vollständige Testung bezeichnet. Somit erscheint das adaptive Testverfahren von Feldt-Caesar und Bruder als ausgesprochen vielversprechend diese Dimension zu erfüllen. Mit diagnostischer Tiefe ist hier nicht die fachinhaltliche Dimension gemeint, sondern die fachdidaktische Dimension der, ob theoretisch oder empirisch, bekannten fachinhaltsbereichsübergreifenden Fähigkeiten und Probleme. Beispielhaft sei hier die Fähigkeit zu Substitutionsleistungen erwähnt. Diese Fähigkeit wird in arithmetischen Operationen, dem (rechnerischen) Umgang mit Funktionen oder im Substitutionsverfahren der Integralberechnung genutzt.

Um eine derartige diagnostische Tiefe in ein diagnostisch breites Testinstrument zeitökonomisch einzubinden, wurden Distraktoren mit erhöhtem diagnostischem Potenzial genutzt (Düsi & Pinkernell, 2017). Hierbei muss sich der Problematik gestellt werden, dass die mit diesen Distraktoren diagnostizierten Konzepte negativ bestimmt werden müssen, da mit Hilfe der Anzahl ausgewählter Distraktoren gleicher Art über Items hinweg diagnostiziert wird. Auf Grund dieser Tatsache beschränkt sich die Art der diagnostischen Tiefe vorerst auf typische bzw. systematische Fehler.

Typische und systematische Fehler

Von typischen Fehlern spricht man, wenn „Fehler identifiziert werden [können], die von einer Vielzahl der Schülerinnen und Schüler begangen werden“ (Nitsch, 2015). Es handelt sich somit um eine rein empirische Kategorie von Fehlern, welche durch systematische Suche und Kategorisierung der gegebenen Antworten innerhalb offener Fragetypen identifiziert werden können.

Systematische Fehler haben hingegen eine lange Tradition der theoretischen, fachdidaktischen Reflektion. Seemann beschrieb schon 1929, dass „Rechenfehler keine Zufallserscheinungen sondern gesetzmäßige Gebilde“ [sic] seien, die vom „Prinzip des kleinsten Energieaufwands“ determiniert seien (vgl. Radatz, 1980). Im Folgenden wird die Definition von Führer genutzt, die systematische Fehler als „gedankliche Leistungen, die unter ähnlichen Bedingungen weitgehend reproduzierbar sind und auf abweichendem Verständnis, auf kognitiven Konflikten oder auf Interferenzen begrenzt sinnvoller Konzepte beruhen“ (Nitsch, 2015) betrachtet. Im Gegensatz zu typischen

Fehlern handelt es sich hierbei also um eine (fach-)didaktische Kategorie, die erst durch einen fachdidaktisch theoretischen Blick entsteht.

Überlinearisierung als systematischer Fehler

Beispielhaft wird im Folgenden die Erstellung und Evaluation diagnostischer Tiefe durch den typischen/(systematischen) Fehler „Überlinearisierung“ (Malle, 1993) beschrieben. Hierfür wurden vier Single-Choice-Items aus dem Inhaltsbereich der Sek.II-Arithmetik Distraktoren beigefügt, die diesen Fehler abbilden sollen (Beispielhaft Abb. 1).

Aus den Antworten auf diese Items wurden Post-hoc dichotome Variablen erstellt, deren eine „1“ zugeordnet wurde, falls der entsprechende Distraktor, welcher für „Überlinearisierung“ steht, ausgewählt wurde. Im Beispiel von Abb. 1 wäre dies der Distraktor „b) und c)“.

Mit diesen Variablen wurde eine CFA durchgeführt um das theoretisch erwartete Konstrukt, welches als diesen Variablen gemein vermutet wurde, näher zu beschreiben. Die Ladungen der Items befinden sich zum 2. MZP im Bereich zwischen 0.552 und 0.743 und können somit als homogen hoch angesehen werden.

Die Modellkenndaten für dieses Modell zum ersten Messzeitpunkt erscheinen durchweg gut. Zum zweiten Messzeitpunkt verbessern sich die meisten Werte nochmals, sodass insgesamt das Modell als gutes bis sehr gutes Modell angesehen werden kann.

Zur qualitativen Beschreibung des Testinstruments wurde Cronbachs alpha zu jedem MZP berechnet. Zum ersten MZP ergab sich ein Wert von 0.50, zum zweiten MZP ein Wert von 0.54. Diese Werte sind, im Gegensatz zu den Modellkennwerten der CFA, eher schlechte Werte. Es muss allerdings beachtet werden, dass die einzelnen Skalen jeweils nur vier Items enthalten und es sich um ein komplexes Konstrukt mit schwieriger Operationalisierung handelt. Bei derartigen Rahmenbedingungen können auch geringere Cronbachs alpha Werte akzeptiert werden (Schecker, 2014).

Betrachten Sie folgende Rechenregeln (Es gilt: $a, b, c > 0$):

- a) $\sqrt{a \cdot b} = \sqrt{a} + \sqrt{b}$
- b) $\sqrt{a + b} = \sqrt{a} + \sqrt{b}$
- c) $\sqrt{a \cdot b} = \sqrt{a} \cdot \sqrt{b}$
- d) $\sqrt{a + b} = \sqrt{a} \cdot \sqrt{b}$

Entscheiden Sie sich bei jeder Rechenregel, ob sie richtig ist und wählen Sie dann die passende Kombination aus.

Wählen Sie eine Antwort:

- c)
- keine
- b) und c)
- a) und d)
- alle

Abb. 1

Kennwert	1. MZP (N= 602)	2. MZP (N=813)
RMSEA	0.052	0.049
p-Wert (χ^2)	.142	.138
CFI	0.989	0.994
TLI	0.981	0.983

Zusammenfassung der Testbewertung, deskriptive Daten und Ausblick

Durch die vorliegenden Daten zur Modellidentifizierung und der Reliabilitätsanalyse gepaart mit der Diskussion zur Inhaltsvalidität lässt sich sagen, dass das Instrument, insbesondere in den vorhandenen Rahmenbedingungen, gut funktioniert.

Die deskriptiven Daten zur diagnostischen Analyse dieses typischen Fehlers ergaben, dass insgesamt 93 Studierende zum zweiten MZP drei oder vier Mal (von vier möglichen Malen) den „Überlinearisierungs“-Distraktor gewählt haben. Auf Grund probabilistischer Überlegungen zur zufälligen Auswahl einer von fünf Auswahlmöglichkeiten wurden diese Studierenden mit dem Konstrukt „Überlinearisierung“ identifiziert. Der typische Fehler „Überlinearisierung“ kann daher bei 11,4% der Studierenden als systematisch, also als über Aufgaben hinweg beständig, identifiziert werden.

Linearisieren_Gesamtscore_2.MZP		
	Häufigkeit	Prozent
0	371	45,6
1	215	26,4
2	134	16,5
3	63	7,7
4	30	3,7

In der nächsten Vorkursdurchführung wird, neben der Testung weiterer typischer Fehler, dieser Fehler an Hand von Items zur Integralberechnung erweitert und eine längsschnittliche Veränderungen per LCM modelliert.

Literatur

- Bausch, I., Biehler, R., Bruder, R., Fischer, P. R., Hochmuth, R., Koepf, W., ..., Wassong, T. (Hrsg.). (2014). *Mathematische Vor- und Brückenkurse: Konzepte, Probleme und Perspektiven*. Wiesbaden: Springer Spektrum.
- Düsi, C., Pinkernell, G. (2017). Distraktorenerstellung im Zusammenhang mit einem Testverfahren zu Mathematikvorkursen. *Beiträge zum Mathematikunterricht 2017*.
- Feldt-Caesar, N., Bruder, R. (2014). Diagnose von Grundwissen und Grundkönnen durch ein adaptives Testverfahren. *Beiträge zum Mathematikunterricht 2014*.
- Malle, G. (1993). *Didaktische Probleme der elementaren Algebra*. Wiesbaden: Vieweg.
- Nitsch, R. (2015). *Diagnose von Lernschwierigkeiten im Bereich funktionaler Zusammenhänge*. Wiesbaden: Springer Fachmedien Wiesbaden.
- Pinkernell, G., Düsi, C., Vogel, M. (2017). Aspekte des Wissens und Könnens der elementaren Algebra. *Beiträge zum Mathematikunterricht 2017*.
- Radatz, H. (1979). *Fehleranalysen im Mathematikunterricht*. Braunschweig, Wiesbaden: Vieweg.
- Schecker, H. (2014). Überprüfung der Konsistenz von Itemgruppen mit Cronbach's Alpha. Springer Spektrum: Berlin, Germany.
- Winter, K. (2011). Entwicklung von Item-Distraktoren mit diagnostischem Potential zur individuellen Defizit- und Fehleranalyse: didaktische Überlegungen, empirische Untersuchungen und konzeptionelle Entwicklung für ein internetbasiertes Mathematik-Self-Assessment. Münster: WTM, Verl. für wiss. Texte und Medien.