Boris GIRNAT, Universität Siegen, DE

Empirische Validierung eines Stufenmodells zum algebraischen Denken in der Sekundarstufe I mithilfe kognitiver Diagnostik

1. Ein Stufenmodell des algebraischen Denkens nach Aké u. a.

Traditionell wird Algebra im Mathematikunterricht der Sekundarstufe I verortet. Die Debatte um die „frühe Algebra“ hat jedoch gezeigt, dass die symbolische Form der Algebra nur eine (späte) Stufe des algebraischen Denkens ist, das sich bereits in der Primarstufe entwickelt. Auch wenn es schwierig ist, eine Definition des algebraischen Denkens anzugeben, sieht man es als das typischste Element an, dass Schüler eine Fähigkeit zu Verallgemeinerungen haben. Als weitere Merkmale werden genannt: „Umgehen mit Operationen als Objekten und ihren Umkehrungen; Herstellen von Beziehungen zwischen Zahlen, Mengen und Relationen (relationales Denken); Verallgemeinern; Umgehen mit Unbekannten; Umgehen mit Veränderungen; Nutzen von (symbolischen) Repräsentationen“ (Fritzlar und Karpinski-Siebold, 2012, S. 261). Die symbolischen Repräsentationen stehen in dieser Aufzählung am Ende. Aké u. a. (2013) haben ein sechsstufiges Modell vorgeschlagen, das algebraische Fähigkeiten von der Vorschule bis hin zur universitären Strukturalgebra beschreibt. Hier werden nur die drei ersten Stufen vorgestellt, da in ihnen der Übergang von der nicht-symbolischen zur symbolischen Algebra stattfindet und die höheren Stufen nicht für die Sekundarstufe I relevant sind, d. h. die dritte Stufe (L2) beschreibt den normativen Zielzustand – die symbolische Algebra –, die in der Sekundarstufe I erreicht werden soll. Die vorangegangenen Stufen (L0 und L1) umfassen zwar Verallgemeinerungen und andere Aspekte des algebraischen Denkens, erreichen aber nicht die symbolische Ebene. Aké u. a. beschreiben die Stufen wie folgt:

L0: „Extensive objects, expressed by natural, numerical, iconic or gestural language, are involved. Symbols that refer to an unknown value can also intervene, but that value is obtained as a result of operations on particular objects“ (Aké u. a., 2013, S. 3).

L1: „Intensive objects, whose generality is explicitly recognized by natural, numerical, iconic or gestural languages, are involved. Symbols that refer to the recognized intensive objects are used, but there is no operation with those objects“ (Aké u. a., 2013, S. 4).

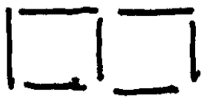
L2: „Indeterminate or variables expressed in literal-symbolic language to refer the intensive objects recognized are involved, but they are linked to the spatial or temporal information of the context“ (Aké u. a., 2013, S. 5).

2. Testaufgaben zur Validierung des Stufenmodells

Die Validierung des Stufenmodells greift den Gedanken auf, dass auf Stufe L0 nur Zahlenwerte zu konkret vorliegenden Sachverhalten ermittelt werden können, auf Stufe L1 eine Verallgemeinerung auf nicht konkret vorliegende

Sachverhalte möglich ist, aber stets konkrete Zahlenwerte berechnet werden und dass erst auf Stufe L2 ein symbolischer Ausdruck erstellt werden kann, der die Berechnung beliebiger Werte erlaubt. Nach dieser Maßgabe sind drei Textaufgaben konstruiert, die jeweils nacheinander Antworten auf diesen drei Stufen verlangen. Als Beispiel wird die erste Aufgabe angegeben. Die zweite handelt von einer Taxifahrt mit Grundgebühr und Kilometergeld, die dritte vom Abbrennen einer Kerze. Allen drei Aufgaben liegt ein linearer Zusammenhang der Art $m{\cdot}x{+}n$ zugrunde.

Mit Zündhölzern werden Ketten gelegt. Bei jedem Schritt wächst die Kette. Trage in die Tabelle ein, aus wie vielen Zündhölzern die Kette im jeweiligen Schritt besteht. Trage zuletzt den Term ein, mit dem man die Anzahl der Zündhölzer allgemein berechnen kann.

Schritt	1	2	3	10	x
Kette				ohne Bild	ohne Bild
Anzahl Zündhölzer	4	**L0**	**L0**	**L1**	**L2**

3. DINA – ein kognitiv-diagnostisches Modell

Kognitiv-diagnostische Modelle (CDMs) sind ein Zugang zur Kompetenzmodellierung (vgl. Rupp u. a., 2010). Anders als metrische Modelle, wie sie z. B. bei PISA, TIMSS und VERA benutzt werden, gehen sie nicht davon aus, dass Kompetenzen kontinuierlich verteilt sind und sich durch metrische Messwerte darstellen lassen, vielmehr haben sie eine dichotome Grundannahme: Eine Person besitzt eine Kompetenz oder sie besitzt sie nicht. Dementsprechend werden den Personen keine Messwerte zugewiesen, sondern Kompetenzprofile, die beschreiben, welche Kompetenzen eine Person – mit einer gewissen Wahrscheinlichkeit – hat oder nicht hat. CDMs eignen sich also besonders für eng umrissene Kompetenzen, bei denen eine graduelle Abstufung unplausibel ist. Die hier betrachteten Kompetenzen – algebraische Aufgaben auf den Stufen L0, L1 oder L2 lösen zu können – scheinen diese Annahme zu erfüllen. Um diese Kompetenzen zu analysieren, wird das DINA-Modell benutzt, das von drei Grundannahmen ausgeht (vgl. Rupp u. a., 2010, S. 96–101): 1) Jeder Aufgabe werden diejenigen Kompetenzen zugewiesen, die zum Lösen der Aufgabe notwendig sind; 2) eine Person muss über *alle* Kompetenzen verfügen, die einer Aufgabe zugeordnet sind, um die Aufgabe zu lösen; 3) ob eine Person über die Kompetenzen des Modells verfügt, lässt sich nur mit einer gewissen Wahrscheinlichkeit ermitteln. Dass also eine Person eine Aufgabe löst, deren Kompetenzen sie nicht besitzt, oder dass sie eine Aufgabe nicht löst, obwohl sie über alle zugehörigen

Kompetenzen verfügt, ist wegen des Punktes 3 keine Modellverletzung. Allerdings verschlechtert jeder solche Fall die Passung des Modells. Bei „zu vielen“ solcher Fälle muss das Modell zurückgewiesen werden. Dafür existieren Gütekriterien und statistische Tests (vgl. George u. a., 2016, S. 9–11). Hier wird das DINA-Modell wie folgt spezifiziert: Es werden fünf Kompetenzen angenommen. Die ersten drei Kompetenzen bestehen darin, die Antworten der drei Testaufgaben *jeweils* auf L0-Niveau beantworten zu können. Diese Kompetenzen umfassen also das Verständnis der *jeweiligen* Sachsituation *und* der Fähigkeit, die L0-Fragen zu beantworten. Die vierte Kompetenz wird so spezifiziert, dass die L1-Fragen *aller drei* Testaufgaben beantwortet werden können, *sofern* die vorangegangenen L0-Fragen beantwortet worden sind (entsprechend der Hierarchie des Stufenmodells). Analog bezieht sich die fünfte Kompetenz auf die L2-Antworten *aller drei* Aufgaben. Das Kompetenzprofil (1,1,0,1,0) gibt also an, dass ein Schüler die L0-Fragen der ersten beiden Aufgaben beantworten konnte und in diesen Fällen die Verallgemeinerung auf L1-Niveau geschafft hat, nicht aber die auf L2.

4. Ergebnisse der Studie

Die angegebenen Aufgaben wurden im Januar 2016 in einem Algebratest mit 639 Schülern der 8. und 9. Klasse im dreigliedrigen Schulsystem des Kantons Aargau (Schweiz) eingesetzt. Dabei wurden nur Schulen des mittleren und höheren Niveaus berücksichtigt, da Algebra im unteren Niveau (I) kein zentraler Gegenstand des Mathematikunterrichts ist. Innerhalb der beiden betrachteten Niveaus (II und III) ist die Studie repräsentativ. Das oben beschriebene DINA-Modell hat ausgezeichnete Kennwerte (max(χ^2) der Items $p = 0{,}72$; mittlerer RMSEA der Items 0,026; SRMSR 0,030; MADcor 0,024; mittlerer Diskriminationsparameter 0,553, vgl. George u. a., 2016, S. 9–11, zur Interpretation der Werte). Alle Berechnungen wurden mit dem CDM-Paket unter R durchgeführt (George u. a., 2016). Zusätzlich wurden ein Test benutzt, der die Spezifikation des DINA-Modells empirisch überprüft (vgl. de la Torre, 2008). Diesen Test hat das Modell bestanden. Insgesamt kann das Stufenmodell von Aké u. a. als empirisch validiert angesehen werden: Die drei Stufen lassen sich im DINA-Modell modellieren und die hierarchische Abhängigkeit, die das Stufenmodell annimmt, kann als bestätigt gelten. Für die normative Zielvorgabe der Sekundarstufe I ist es wichtig, wie viele Schülerinnen und Schüler die symbolische Stufe der Algebra (L2) erreichen. Trotz der Positivauswahl durch den Ausschluss des Schulniveaus I wird sie nur von 63% der Schülerinnen und Schüler erreicht. Die folgende Tabelle gibt Auskunft, wie die Beherrschung der symbolischen Stufe der Algebra (L2) mit drei Kovariaten (Geschlecht, Jahrgangsstufe und Schulniveau) zusammenhängt (mit dem üblichen Signifikanzniveau von $p = 0{,}05$ im χ^2-Test und Cramérs V als Effektstärke, wobei 0,1 für einen schwachen, 0,3 für einen mittleren und 0,5 für einen starken Effekt steht, vgl. Liebetrau, 1983):

	L0 oder L1	L2	signifikant?	Cramérs V
männlich	38%	62%	nein	0,03
weiblich	35%	65%		
8. Klasse	35%	65%	nein	0,04
9. Klasse	38%	62%		
Niveau II	68%	32%	ja	0,41
Niveau III	24%	76%		

Es gibt keinen signifikanten Geschlechterunterschied. Bemerkenswert ist der signifikante und effektstarke Unterschied zwischen dem mittleren (II) und dem höheren Schulniveau (III): Vergleicht man die 32% und 76% der Schülerinnen und Schüler, die jeweils die symbolische Stufe der Algebra erreichen, so kann man davon sprechen, dass die beiden Schulniveaus in verschiedenen Welten liegen – wohlgemerkt: das untere Niveau ist nicht berücksichtigt. Ebenso interessant ist das Fehlen eines signifikanten Unterschiedes zwischen der 8. und 9. Klasse. Das heißt: Es lässt sich kein Lernzuwachs in der Kompetenz feststellen, einen außermathematischen Sachverhalt durch einen linearen Term symbolisch zu darzustellen.

Literatur

Aké, L., Godino, J., Gonzato, M., und Wilhelmi, M. (2013). Proto-Algebraic Levels of Mathematical Thinking. In Lindmeier, A. M., & Heinze, A. (Hrsg.), *Proceedings of the 37th Conference of the International Group for the Psychology of Mathematics Education*, *Band 2* (S. 1–8). Kiel: PME.

de la Torre, J. (2008). An Empirically Based Method of Q-Matrix Validation for the DINA Model: Development and Applications. *Journal of Educational Measurement*, *45*, 343–362.

Fritzlar, T., und Karpinski-Siebold, N. (2012). Algebraisches Denken und mathematische Begabungen im Grundschulalter In Ludwig, M., und Kleine, M. (Hrsg.), *Beiträge zum Mathematikunterricht* (S. 261 – 264). Münster: WTM-Verlag.

George. A. C., Robitzsch, A., Kiefer, T., Groß, J., und Ünlü, A. (2016). The R Package CDM for Cognitive Diagnosis Models. *Journal of Statistical Software, 74(2)*, 1–24.

Liebetrau, A. M. (1983). Measures of association. *Quantitative Applications in the Social Sciences, 32*, S. 15–16, Newbury Park: Sage Publications.

Rupp, A., Templin J., Henson R. (2010). *Diagnostic Measurement: Theory, Methods, and Applications.* The Guilford Press, New York.