

Rationale Aufgabenanalyse zur Untersuchung von fachbezogenen Aspekten der Unterrichtsqualität bei TEDS-Validierung

Einbettung in das Forschungsvorhaben von TEDS-Validierung

Die Studien TEDS-Unterricht und TEDS-Validierung betrachten die Beziehungen zwischen der fachlichen Kompetenz von Mathematiklehrkräften und den Lernzuwächsen der von ihnen unterrichteten Schülerinnen und Schüler unter Bezug auf die Unterrichtsqualität als Mediatorvariable. In Validitätsüberprüfungen des Projekts wird untersucht, ob ähnliche strukturelle Korrelationen über verschiedene deutsche Bundesländer hinweg ermittelt werden können. Da instrumentell auf die dort standardmäßig durchgeführten Tests zur Messung des Lernzuwachses zurückgegriffen wurde, ist es notwendig, die konzeptionelle Äquivalenz der verschiedenen Schülerleistungstests zu prüfen. Vor allem aber wird untersucht, inwiefern die von TEDS-Unterricht identifizierten Dimensionen der Unterrichtsqualität auch mit einem alternativen Instrument identifiziert werden können.

Ausgewählte theoretische und methodische Überlegungen

Die beiden skizzierten Zielsetzungen werden mittels einer rationalen Aufgabenanalyse mit Fokus auf mathematisch-kognitiven Aspekten untersucht. Aufgrund der hohen Bedeutung für den Mathematikunterricht können Aufgaben theoretisch als Indikatoren für kognitive Subdimensionen der Unterrichtsqualität betrachtet werden (Bromme et al., 1990), wofür auch empirische Nachweise gefunden werden konnten (Neubrand et al., 2011). Dies gilt erst recht für die Bestimmung der konzeptuellen Äquivalenz von Schülerleistungstests (van den Ham et al., 2014). In der rationalen Aufgabenanalyse wird sich theoretisch und empirisch auf Vorarbeiten u.a. von COACTIV im Hinblick auf Unterrichtsqualität (u.a. Jordan et al., 2006), von PISA zur Bestimmung der mathematischen Komplexität von Testitems (Turner et al., 2015) und des IPN hinsichtlich der konzeptionellen Äquivalenz (u.a. van den Ham et al., 2014) bezogen.

Das aus dem bestehenden Analyseschema von COACTIV weiterentwickelte Klassifikationssystem erfasst sogenannte Oberflächenmerkmale von Aufgaben, wie z.B. externe Verweise auf Differenzierung. Die Einordnung der inhaltlichen Kompetenzen folgt der Struktur der deutschen Bildungsstandards und erlaubt eine klare Zuweisung zu einer curricularen Stufe. Der Fokus liegt

jedoch auf den kognitiven Eigenschaften der Items, insbesondere auf den allgemeinen mathematischen Kompetenzen.

Für das Aufgabenanalysesystem wurden hauptsächlich ordinalskalierte Operationalisierungen entwickelt. Es stellt eine Unterscheidung zwischen vier Komplexitätsstufen sicher, die unabhängig vom Alter der Schüler(innen), dem Schultyp oder dem mathematischen Inhalt klar unterscheidbar und erreichbar sind. Damit in der Stichprobe alle Stufen angemessen oft vorkommen können, werden soweit möglich Eigenschaften verwendet, von denen anzunehmen ist, dass sie im normalen Regelunterricht entsprechend vorzufinden sind.

In der Studie TEDS-Validierung haben rund 100 Mathematiklehrkräfte an den Lehrerprofessionstests teilgenommen. Eine Untergruppe von 38 Personen wurde während des Unterrichts direkt von externen Rater(inne)n beobachtet. Aus diesen Unterrichtsstunden wurden bei 29 Lehrkräften alle Aufgaben eingesammelt, was zusammen rund 2700 Aufgaben ergeben hat.

Operationalisierung am Beispiel der Modellierungskomplexität

Die Weiterentwicklung der Kategorien aus den bestehenden Analyse-schemata von COACTIV und PISA wurde nach den Prinzipien der qualitativen Textanalyse von Mayring (2010) durchgeführt. Exemplarisch wird im Folgenden die Überarbeitung der Modellierungskomplexität vorgestellt, die notwendig erschien, da kein empirischer Zusammenhang zu Lehrkompetenzen bzw. Leistungszuwächsen hergestellt werden konnte (Neubrand et al. 2011). Auch die Analysen nach der von Turner et al. gebildeten Kategorie (2015) konnten keinen Nachweis erbringen, inwiefern die ermittelte Modellierungskomplexität schwierigkeiterklärend ist. Dabei war zunächst der Begriff der *Modellierungskomplexität* von Mathematikaufgaben festzulegen. Vor dem Hintergrund verschiedener Definitionen (Greefrath & Vorhölter 2016), aber auch in Abgrenzung zu anderen Analyse-kategorien (wie z.B. Problemlösen oder Sprachliche Komplexität) wurde dabei als Wesenskern von Modellierungstätigkeiten die Übersetzungsprozesse zwischen den Ebenen Realität und Mathematik identifiziert und der Begriff hierauf beschränkt. Übersetzungsprozesse umfassen Aktivitäten wie die Entwicklung eines Realmodells (Vereinfachen, Annahmen treffen, Strukturieren, Recherchieren von Informationen), das Mathematisieren sowie das Interpretieren und Validieren der Resultate bzw. des Modells. Weitere innermathematische Aktivitäten spielen eine bedeutende Rolle im Modellierungsprozess, werden aber gesondert in den entsprechenden Kategorien des Klassifikationsschemas berücksichtigt.

Die Ausarbeitung der ordinalskalierten Operationalisierung der Komplexitätsstufen für Modellieren war eingebettet in den oben skizzierten Anforderungsrahmen des gesamten Analyseschemas. So wurde bspw. die höchste Stufung nicht „echten“ Modellierungsaufgaben vorbehalten, da angenommen wird, dass diese im Unterricht selten vorkommen (Greefrath & Vorhölter 2016). Veränderte Ausformulierungen der Komplexitätsstufen wurden auf Aufgaben aus diversen Schulbüchern, auf Musteraufgaben aus Schülerleistungstests und auf Sammlungen von spezifischen Modellierungsaufgaben angewandt (u.a. Siller et al. 2018). Anhand dieser Aufgaben wurden einzelne Merkmale herausgearbeitet, die einen gewissen Grad der Modellierungskomplexität kennzeichnen. In „Rücklaufschleifen“ (Mayring 2010) wurden die aufgestellten Operationalisierungen immer wieder erprobt und revidiert. Die Einholung von Expertenmeinungen im Rahmen der Arbeitsgruppe Mathematikdidaktik der Universität Hamburg und darüber hinaus, diente zudem als qualitätssichernde Maßnahme.

Beispielhaft sei ein Anpassungsprozess skizziert: Die Vergabe der Codestufe 0 wurde zunächst bei COACTIV nur für rein innermathematische Aufgaben und Aufgaben mit Einheiten ohne weiteren Realitätsbezug vergeben, da ein realitätsbezogener Kontext für eine Modellierungsaufgabe vorauszusetzen ist. In Anlehnung an Turner et al. (2015) werden bei TEDS-Validierung jedoch auch Aufgaben *mit* realitätsbezogenem Kontext Code 0 zugewiesen, wenn die Beziehung zwischen der außermathematischen Situation und dem mathematischen Model für die Lösung der Aufgabe nicht relevant ist (ebd.). Anders als bei Turner et al. wird dies jedoch sehr streng interpretiert: Ist nämlich denkbar, dass der realitätsbezogene Kontext in bei der Aufgabenbearbeitung helfen oder verwirren könnte, wird trotzdem Code 1 vergeben.

Vor der Analyse der Aufgaben aus dem Datensatz wurde anhand eines Pools von 100 Aufgaben aus unterschiedlichen Quellen die Güte des neu entwickelten Klassifikationsschemas anhand der Übereinstimmung von vier Rater(innen) überprüft. Als Reliabilitätsmaße wurden die mittlere prozentuale Übereinstimmung (PÜ) und Kendalls Konkordanzkoeffizient W herangezogen (Wirtz & Caspar 2002), deren Werte der nachstehenden Tabelle zu entnehmen sind.

| | PÜ (vor Disk.) | PÜ (nach Disk.) | Kendalls W (vor Disk.) | Kendalls W (nach Disk.) |
|--------------------------|-------------------|--------------------|---------------------------|----------------------------|
| Modellierungskomplexität | 79,7 % | 96 % | 0,9079 | 0,9845 |

Zusammenfassende Überlegungen zur Validität und Reliabilität

In der erneuten Arbeit an dem COACTIV-Klassifikationsschema haben sich die Stärken des Instrumentes gezeigt, d.h. auch wenn dieses in Teilen angepasst wurde, so können doch weite Teile übernommen werden. Im Hinblick auf die Modellierungskomplexität wurden die vorhandenen Operationalisierungen in einer Kombination aus deduktiver Ableitung aus der Literatur und induktiver Analyse von konkreten Aufgaben überarbeitet, dabei von anderen Analysekatégorien abgegrenzt und hierzu Expertenmeinungen eingeholt. Aufgrund des beschriebenen Vorgehens kann Inhaltsvalidität angenommen werden. Die Vorabüberprüfungen der Interraterreliabilität zeigen zumindest zufriedenstellende Übereinstimmungswerte. Insgesamt ist eine Operationalisierung entstanden, die es erlaubt, den angenommenen Zusammenhang zu Unterrichtsqualität, Leistungszuwächsen und Lehrerkompetenzen zu erklären.

Literatur (Auswahl)

- Bromme, R., Seeger, F. & Steinbring H. (1990). *Aufgaben als Anforderungen an Lehrer und Schüler* (IDM-Untersuchungen zum Mathematikunterricht, Bd. 14). Köln: Aulis.
- Greefrath, G. & Vorhölter, K. (2016). *Teaching and Learning Mathematical Modelling. Approaches and Developments from German Speaking Countries*. Switzerland: Springer.
- Jordan, A., Ross, N., Krauss, S., Baumert, J., Blum, W., Neubrand, M. et al. (2006). *Klassifikationsschema für Mathematikaufgaben: Dokumentation der Aufgabenklassifikation im COACTIV-Projekt*. Berlin: Max-Planck-Institut für Bildungsforschung.
- Mayring, P. (2010). *Qualitative Inhaltsanalyse. Grundlagen und Techniken* (11. Auflage). Weinheim: Beltz.
- Neubrand, M., Jordan, A., Krauss, S., Blum, W. & Löwen, K. (2011). Aufgaben im COACTIV-Projekt: Einblicke in das Potenzial für kognitive Aktivierung im Mathematikunterricht. In M. Kunter et al. (Hrsg.), *Professionelle Kompetenz von Lehrkräften* (S. 115-132). Münster: Waxmann.
- Siller, H.-S., Greefrath, G. & Blum, W. (Hrsg.) (2018). *Neue Materialien für einen realitätsbezogenen Mathematikunterricht 4. 25 Jahre ISTRON-Gruppe - eine Best-of-Auswahl aus der ISTRON-Schriftenreihe*. Wiesbaden: Springer Spektrum.
- Turner, R., Blum, W. & Niss, M. (2015). Using competencies to explain mathematical item demand: A work in progress. In K. Stacey & R. Turner (Hrsg.), *Assessing mathematical literacy. The PISA experience* (85-115). Springer: Cham.
- Van den Ham, A.-K., Nissen, A., Ehmke, T., Sälzer, C., & Roppelt, A. (2014). Mathematische Kompetenz in PISA, IQB-Ländervergleich und NEPS – Drei Studien, gleiches Konstrukt? *Unterrichtswissenschaft*, 42(4), 321–341.
- Wirtz, M. A. & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität. Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen*. Göttingen: Hogrefe.