

Dominik LEISS, Anika BÜRGERMEISTER, Birgit HARKS, Eckhard KLIE-
ME, Katrin RAKOCZY, alle Frankfurt am Main; Werner BLUM, Kassel

Consequences of Classroom Assessment - Vorstellung des Projekts CoCa¹

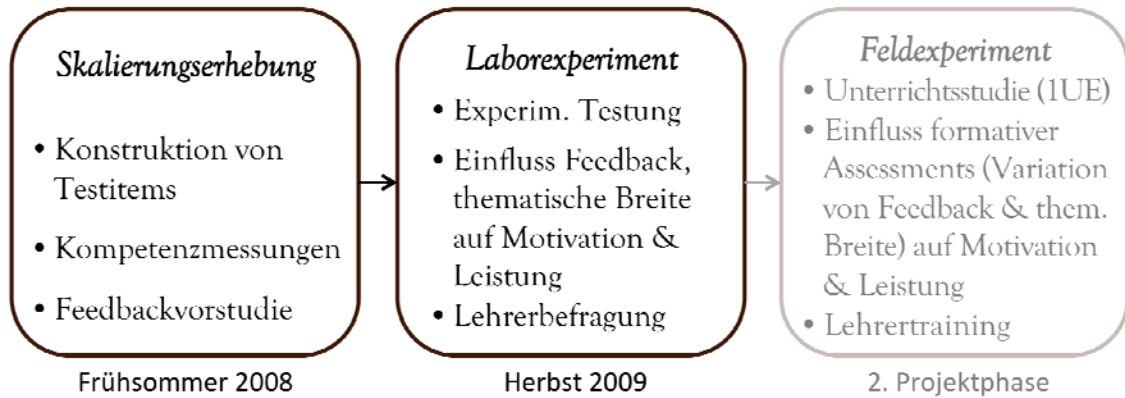
Anknüpfend an die aktuelle Diskussion um den Einsatz formativer und summativer Leistungsmessungen im Unterricht (vgl. Shepard, 2000) steht im Projekt COCA die Untersuchung von Nutzung und Folgen unterrichtsnaher (formativer) bzw. inhaltsbreiter (summativer) Tests sowie darauf basierender Feedbackvarianten im Vordergrund des Interesses. Die zentralen Fragestellungen des Projekts sind dementsprechend u.a.

- I. Eignen sich aktuelle psychometrische Verfahren (mehrdimensionale IRT-Modelle mit qualitativ unterschiedlichen Niveaus), um relativ kurzfristige Lernvorgänge (z.B. eine Unterrichtseinheit mit ca. vier bis sechs Wochen) zu erfassen?
- II. Können dabei mathematische Kompetenzen, wie sie etwa bei den Bildungsstandards Mathematik (vgl. Leiss & Blum 2006) beschrieben werden, psychometrisch getrennt werden?
- III. Welche Folgen haben die Durchführung von unterrichtsnahen bzw. inhaltsbreiten Tests sowie unterschiedliche Rückmeldeformen (prozessbezogen, sozial-vergleichend, kriterial – genauer siehe weiter unten) der Testergebnisse auf das Leistungsverhalten sowie auf kognitive und motivationale Zielvariablen?
- IV. Welche Voraussetzungen – beispielweise Einstellungen, diagnostische und andere Kompetenzen – auf Seiten der Lehrenden sind bestimmend für die Gestaltung der Leistungsbewertung und welche Folgen hat diese wiederum für Motivation und Kognition der Lernenden?

Design

COCA fokussiert seine Untersuchungen auf Realschulklassen bzw. B-Kurse des Jahrgangs 9, auf technische Kompetenzen [TK] und Modellierungskompetenzen [MK] sowie auf die beiden Inhaltsbereiche Pythagoras und Lineare Gleichungssysteme. Das zugrunde liegende Design für die auf drei Projektphasen mit je zwei Jahren ausgerichteten Studien sieht dabei folgendermaßen aus:

¹ Hierbei handelt es sich um ein von der DFG im Rahmen des Schwerpunktprogramms "Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen" gefördertes Projekt (Projektleiter: Prof. Dr. E. Klieme)



Im Folgenden wird kurz die abgeschlossene Skalierungserhebung dargestellt; über die folgenden Studien wird an anderer Stelle berichtet.

Skalierungserhebung

Das primäre Ziel der Skalierungserhebung bestand zum einen darin, ausreichend Items für die vorgesehenen Inhaltsgebiete und die entsprechenden Kompetenzen für die anstehenden Labor- und Unterrichtstudien zu entwickeln und zu skalieren, und zum anderen darin, bei der Aufgabenpilotierung erste Feedbackvarianten empirisch zu erproben bzw. zu analysieren. Dementsprechend stand am Anfang dieser Studie die Konstruktion einer hinreichend großen Zahl (131) von Items. Dabei wurden neben Aufgaben, welche die Kompetenzen in ihrer Gänze verlangen, auch solche Items konstruiert, die das Vorwissen zu den Themengebieten systematisch erfassen oder einzelne Teilkompetenzen des Modellierens abtesten, und es wurden Parallelaufgaben entwickelt, bei denen systematische Variationen zur Bestimmung von schwierigkeitsbestimmenden Merkmalen vorgenommen wurden. Die Items wurden in einem Multimatrixdesign an ca. 1500 hessischen Schülern pilotiert. Dabei erfüllten 92% die psychometrischen Anforderungen an gute Testaufgaben, so dass prinzipiell ausreichend Items aus beiden Kompetenz- und Inhaltsbereichen vorhanden waren. Dabei zeigte eine zweidimensionale Skalierung (mit Einfachladung) des Tests mit den beiden Dimensionen *Technisches Arbeiten* und *Modellieren* zwar eine relativ hohe Korrelation (.87) zwischen diesen Dimensionen, allerdings hatte dieses Modell signifikant bessere Fitwerte (AIC/ BIC/ CAIC) als ein ein-dimensionales Modell, das nur von einem allgemeinen Konstrukt *mathematische Fähigkeit* ausgeht. Insofern handelt es sich bei Modellierungskompetenz und Technischer Kompetenz nicht nur um theoretisch sondern auch um psychometrisch differenzierbare Personenfähigkeiten, so dass mit diesen Tests nun Instrumente vorliegen, die es erlauben, getrennt nach den beiden Kompetenzen sensible Leistungs(fortschritts)messungen im schulischen Leistungsspektrum vorzunehmen (vgl. Abb. 1).

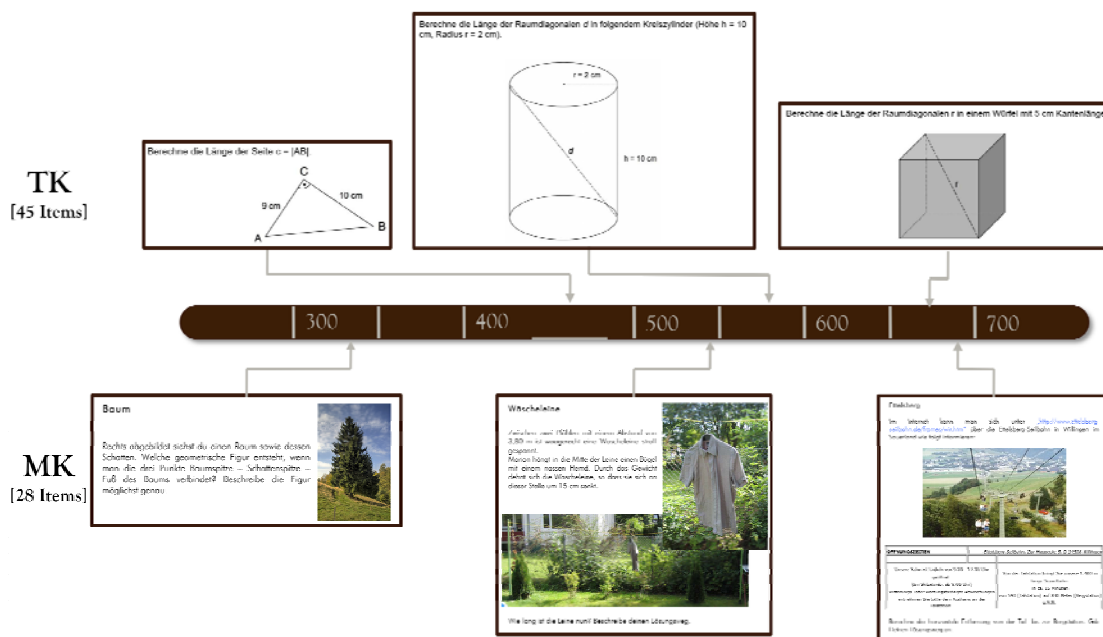


Abb. 1: Kompetenz- und inhaltspezifisches Aufgabenspektrum am Beispiel Pythagoras

Zudem zeigte sich eine Fülle an interessanten Einzelergebnissen wie z.B. dass bei der Aufgabenvariation die Aufgabenvariante mit einem ausführlicheren Text bei Modellierungsaufgaben durchaus mit höheren Lösungsquoten verbunden sein kann als die Aufgabenvariante mit weniger Text oder dass insbesondere Aufgaben, welche die Teilkompetenz „Validieren“ abzufragen versuchen, einen relativ hohen Schwierigkeitsgrad aufweisen.

Parallel zur Testung der Schüler wurde ein Großteil der beteiligten Lehrpersonen mithilfe eines Fragebogens zu verschiedenen Aspekten ihres Mathematikunterrichts, aber auch zur Leistungsstärke der Einzelschüler befragt. Ergebnisse dieser Befragungen waren u.a., dass

- eine verbale Rückmeldekultur, verbunden mit verschiedenen lehrer- und notenzentrierten Bewertungsformen, im Unterricht dominieren,
- partizipative Formen (Selbst- oder Peer-Evaluation) insgesamt selten sind und sich vergleichsweise häufig bei Lehrern mit guten diagnostischen Kenntnissen (laut Selbsteinschätzung) finden,
- lehrer- und notenzentrierte Rückmeldungen mit niedriger Motivation und eine aus Schülersicht ipsative Bezugsnormorientierung der Lehrperson mit einer hohen Motivation der Schüler einhergehen,
- hohe diagnostische Kompetenz der Lehrpersonen mit besseren Testleistungen der Schüler verbunden ist.

Fünf Monate nach der Skalierungserhebung erhielten die Schüler von 14 Klassen eine individuelle Rückmeldung, wobei die drei folgenden Formen des Feedbacks eingesetzt wurden.²

1. Prozessbezogenes Feedback

Hierbei werden den Schülern anhand ihrer Aufgabenlösungen ihre individuelle Stärken und Schwächen erläutert.

2. Sozial-vergleichendes Feedback

Bei dieser Variante wird den Lernenden rückgemeldet, wie leistungsstark sie im Vergleich zur durchschnittlichen Klassenleistung in den beiden untersuchten Kompetenzen TK und MK sind.

3. Kriteriales Feedback

Bei diesem Feedback wird die Schülerleistung mit normativ gesetzten Lernzielen (für Realschüler/innen der Jahrgangsstufe 9) verglichen. Die individuelle Schülerleistung und das Lernziel werden dabei anhand eines Kompetenzstufenmodells – getrennt für technische Kompetenz und Modellierungskompetenz – dargestellt.

Im Anschluss an das Feedback mussten die Schüler/innen einen Fragebogen zur emotionalen und motivationalen Wirkung des Feedbacks ausfüllen. Im Verhältnis der drei Feedbackarten wurde dabei das kriteriale Feedback von den Lernenden vergleichsweise positiv beurteilt. Unter der Bedingung, dass zwischen Testung und Feedback mehrere Monate liegen – was beispielsweise bei vielen Large Scale Erhebungen der Fall ist –, scheint also eine kriteriumsbezogene Rückmeldung den größten motivationalen Effekt zu haben. Der Einfluss unterschiedlicher Feedbackformen auf Motivation und Leistung bei zeitnaher Rückmeldung soll in den folgenden Experimenten untersucht werden.

Literatur

Shepard, L. (2000). The role of assessment in a learning culture. *Educational researcher*, 29 (7), 4-14.

Leiss, D. & Blum, W. (2006): Beschreibung zentraler mathematischer Kompetenzen. In: Blum, W. et al. (Hg.). *Bildungsstandards Mathematik: konkret*. Berlin: Cornelsen Scriptor, 33-50.

² Die drei Rückmeldebedingungen wurden hinsichtlich der durchschnittlichen Testleistung der Schüler/innen parallelisiert. Die parallelisierte Zuordnung erfolgte über die Klassen hinweg.