

Kompetenzzuwachsmessung bei Lehrkräftefortbildungen durch retrospektive Selbsteinschätzung am Beispiel von „Stochastik kompakt“

Der neue Kernlehrplan für die Sekundarstufe II des Gymnasiums, der ab 2014 in Kraft getreten ist, schreibt die Behandlung des Themas Stochastik inklusive Hypothesentestens verbindlich vor. Um den neu entstandenen Weiterbildungsbedarf zu decken, entwickelte das Deutsche Zentrum für Lehrerbildung Mathematik (DZLM) die 4-tägige Fortbildungsreihe „Stochastik kompakt“. Im folgenden Artikel werden ausgewählte Ergebnisse zum Kompetenzzuwachs der Lehrkräfte, die den Kurs besucht haben, vorgestellt.

1. Kursdesign

Der Kurs wurde anhand der Designprinzipien des DZLMs (Barzel & Selter, 2015), einer Interpretation der Bildungsstandards und mit Hilfe von Ergebnissen stochastischer Forschung (Biehler, Ben-Zvi, Bakker, & Makar, 2013; Biehler & Eichler, 2015; Biehler, Eichler, Engel, & Warmuth, 2010) entwickelt. Der Kurs mit dem Namen „Stochastik kompakt“ wurde 2013 zum ersten Mal und je ein weiteres Mal in den Jahren 2014 und 2015 durchgeführt. Die letzten beiden Fortbildungstage beschäftigten sich immer mit dem Hypothesentesten und dabei unter anderem mit einem unterrichtlichen Einstieg in das Testen mit dem P-Wert Konzept, Fehlvorstellungen beim Testen und Hypothesentesten anhand von authentischen Beispielen. Dieses Vorgehen wurde Best@Kontext genannt (Oesterhaus & Biehler, 2014). Insgesamt konnten mit den drei Durchführungen der Fortbildung ca. 270 Lehrkräfte aus ganz NRW erreicht werden.

2. Forschungsmethode und Design

In den drei Fortbildungsjahrgängen wurden unterschiedliche Fragebögen eingesetzt. Erhoben wurden unter anderem die Einstellung zum Lehren und Lernen von Stochastik, der Technologieeinsatz und die Stages of Concern der Lehrkräfte (Biehler & Nieszporek, eingereicht). Darüber hinaus wurden auch Daten zur Veränderungen der fachlichen Kompetenz im Bereich des Hypothesentestens erhoben. Hierfür konnte aber kein Wissenstest eingesetzt werden, stattdessen musste auf eine andere Art der Befragung zurückgegriffen werden. Gründe dafür waren einerseits die Ermangelung eines geeigneten und validen Testes. Andererseits sahen wir es für die Teilnehmerinnen und Teilnehmer nicht als zumutbar an, sich einem umfangreichen Wissenstest zu unterziehen. Es wurde stattdessen ein Fragebogen mit verschiedenen

Items zur Selbsteinschätzung der Kompetenzentwicklung eingesetzt, wie das auch in anderen vergleichbaren Studien praktiziert wird. Diese Selbsteinschätzung wurde immer mithilfe eines retrospektiven Posttest erhoben. Durch diesen Ansatz fand die Einschätzung der Kompetenz anhand einer Skala statt, und es konnten Pretest- und ähnliche Effekte ausgeschlossen werden (Lam & Bengo, 2003).

Der Fragebogen war dabei wie folgt aufgebaut. Es wurden zu verschiedenen inhaltlichen Kategorien des Hypothesentestens je zwei Frageitems gestellt. Eines der Items bezog sich dabei immer auf den Zeitpunkt vor dem Besuch der Fortbildung und das andere Item auf die Kompetenzeinschätzung nach der Fortbildung. Die Lehrkräfte beantworteten die Items mit Hilfe einer Likert-Skala von „Trifft nicht zu“ (=1) „bis Trifft zu“ (=6). Zudem bestand auch die Möglichkeit, keine Angaben zu machen (siehe Abbildung 1).

Ich kenne didaktische Vorteile des Einsatzes von Simulation beim Hypothesentesten.							
	Trifft nicht zu				Trifft zu		k.A.
Vor der Veranstaltung	<input type="radio"/>						
Nach der Veranstaltung	<input type="radio"/>						

Abbildung 1: Beispiel Frage, die zu allen Messzeitpunkten eingesetzt worden ist

Die Art und Anzahl der Kategorien wurden immer an den Inhalt der Fortbildung angepasst und variierten daher in den Jahren 2013 bis 2015. Allerdings kamen 19 Kategorien in allen drei Erhebungsjahren vor. Für die Analyse wurden nur diese Items berücksichtigt, um einen Vergleich zwischen den Kohorten zu ermöglichen und um Antworten auf die folgenden Forschungsfragen zu finden:

1. Schätzen die Lehrkräfte ihre Kompetenz nach der Fortbildung besser ein?
2. Gibt es zwischen den Kategorien unterschiedliche Zuwächse?
3. Lassen sich Unterschiede zwischen den Jahrgängen feststellen?

3. Ausgewählte Resultate

Ca. 68% der Teilnehmer füllten den Fragebogen zur Kompetenzentwicklung aus. Allerdings wurde dabei nicht immer zu jedem Item eine Selbsteinschätzung abgegeben. Es konnten daher zwischen 165 und 179 Teilnehmerdaten, abhängig vom jeweiligen Item, erhoben werden. Im Durchschnitt entspricht dies 175 Antworten pro Item. Davon stammen 52 aus dem Jahr 2013, 58 aus dem Jahr 2014 und 65 aus 2015. Im Folgenden werden drei Kategorien genauer betrachtet. Wie man der Tabelle 1 entnehmen kann, wurde die Kompetenz zum Zeitpunkt vor der Fortbildung in allen drei Jahrgängen eher gering eingeschätzt. Die Unterschiede zwischen den Messungen sind bei auf-

gelisteten Items in keinem Fall signifikant ($p \geq .168$). Somit kann davon ausgegangen werden, dass die Teilnehmerinnen und Teilnehmer mit einem ähnlichen Kompetenzlevel in die Fortbildung gestartet sind.

Vorher Einschätzung (Mittelwerte)	2013	2014	2015
Mittelwert aller Items	2,76	2,61	2,27
Ich kenne didaktische Vorteile des Einsatzes von Simulation beim Hypothesentesten	1,45	2,10	2,00
Ich kann einen Hypothesentest zu einem festen Signifikanzniveau konstruieren und durchführen.	3,20	3,03	2,76
Ich kann einen Hypothesentest mit P-Werten durchführen.	2,04	2,14	2,02

Tabelle 1: Mittelwerte ausgewählter Items, unterschieden nach Durchgängen

Die Frage nach dem Hypothesentesten mit P-Werten wurde neben drei anderen Items in der Vorher-Einschätzung signifikant schlechter bewertet als der Durchschnitt ($p < .001$). Es handelte sich dabei um Themen, die in der Schule selten unterrichtet werden und daher neu für die Lehrkräfte waren. Im Vergleich dazu steht z.B. das Testen mit festen Signifikanzniveau, welches erkennbar besser abschnitt als das Testen mit dem P-Wert ($p < .001$).

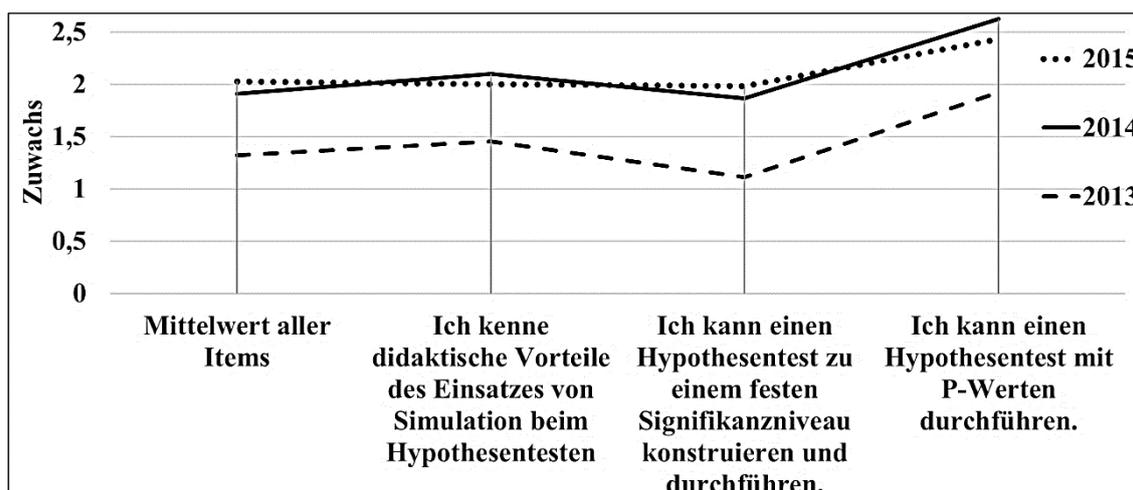


Abbildung 2: Zuwächse ausgewählter Items, unterschieden nach Durchgängen

Betrachtet man nun die Kompetenzzuwächse der Kategorien, zeigt sich zunächst, dass sich die Teilnehmerinnen und Teilnehmer nach dem Besuch der Fortbildung immer besser als vorher einschätzen ($p < .001$). Es zeigt sich auch ein deutlicher Unterschied bei der Kompetenzzunahme beim Testen mit dem P-Wert gegenüber dem mit festen Signifikanzniveau ($p < .001$). Dadurch war die Differenz der beiden Kompetenzlevel nach der Fortbildung zwar geringer, aber immer noch signifikant ($p < .001$). Im Vergleich zu 2013 fiel in den Durchgängen 2014 und 2015 die Zunahme erheblich größer ausgefallen ($p = .003$). Nicht nur bei fachlichen, sondern auch bei didaktischen Themen wie dem Item „didaktische Vorteile des Einsatzes von Simulation beim Hypothesentesten“ waren Unterschiede erkennbar. Die Daten der Kurse von 2014 und 2015 waren sehr ähnlich ($p \geq .104$). Die Ungleichheit zwischen den

2013 und den späteren Jahrgängen kann auf die Verbesserung der Fortbildung zwischen 2013 und 2014 zurückgeführt werden. (Ausgewertet wurden die Daten mithilfe von verschiedenen t-Tests und ANOVAs.)

4. Ausblick

Es ist geplant, die Ergebnisse der Kompetenzzuwächse mit weiteren Ergebnissen, z.B. mit den der Stages of Concern in Verbindung zu bringen und Zusammenhänge offenzulegen. Für kommende Fortbildungen ist zudem vorgesehen, zusätzliche zur retrospektiven Kompetenzeinschätzung eine Vorher-Einschätzung in einem Pretest zu erheben. Auch sollen mögliche Ergebnisse durch eine Interviewstudie verifiziert werden.

5. Danksagung

Wir bedanken uns bei Janina Niemann (geborene Oesterhaus) für die Erhebung und Bereitstellung der Daten der Jahrgänge 2013 und 2014, bei Ruben Loest für die Durchführung der Befragung im Jahr 2015 und bei Hauke Friedrich für die Aufbereitung der Daten.

Literaturverzeichnis

- Barzel, B., & Selter, C. (2015). Die DZLM-Gestaltungsprinzipien für Fortbildungen. *Journal für Mathematik-Didaktik*, 36(2), 259-284.
- Biehler, R., Ben-Zvi, D., Bakker, A., & Makar, K. (2013). Technology for Enhancing Statistical Reasoning at the School Level. In M. A. Clements, A. J. Bishop, C. Keitel, J. Kilpatrick, & F. K. S. Leung (Eds.), *Third International Handbook of Mathematics Education* (pp. 643-689): Springer New York.
- Biehler, R., & Eichler, A. (2015). Leitidee Daten und Zufall. In W. Blum, S. Vogel, C. Drüke-Noe, & A. Roppelt (Eds.), *Bildungsstandards aktuell: Mathematik für die Sekundarstufe II* (pp. 70-80). Braunschweig: Diesterweg.
- Biehler, R., Eichler, A., Engel, J., & Warmuth, E. (2010). Leitidee Daten und Zufall für die Sekundarstufe II – Kompetenzprofile für die Bildungsstandards aus Sicht der Stochastik und ihrer Didaktik. Retrieved from http://www.stochastik-in-der-schule.de/Dokumente/Leitidee_Daten_und_Zufall_SekII.pdf.
- Lam, T. C., & Bengo, P. (2003). A comparison of three retrospective self-reporting methods of measuring change in instructional practice. *American Journal of Evaluation*, 24(1), 65-80.
- Oesterhaus, J., & Biehler, R. (2014). Designing and implementing an alternative teaching concept within a continuous professional development course for German secondary school teachers In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, Arizona, USA*. Voorburg, The Netherlands: International Statistical Institute, iase-web.org.