



Oehen Annabel¹, Wilhelm Markus¹, Furrer Florian², Lohse-Bossenz Hendrik³, Tardent Josiane² & Gut-Glanzmann Christoph²

¹ Pädagogische Hochschule Luzern (IFNMG), ² Pädagogische Hochschule Zürich und ³ Pädagogische Hochschule Heidelberg

*Kontakt: annabel.oehen@phlu.ch

Hintergrund

Projektübersicht

Die Forschung hinsichtlich der Handlungskompetenzen von (angehenden) Lehrpersonen in der Handlungskette Planen-Unterrichten-Reflektieren (Lehrzyklus PUR) wurde in den vergangenen Jahren intensiv vorangetrieben, wobei sich gerade in den naturwissenschaftlichen Fachdidaktiken nur wenige empirische Studien mit diesem Thema befassen. Studien, die alle drei Handlungskompetenzen gleichzeitig in Beziehung bringen, finden sich kaum. Noch gänzlich unbekannt sind die Wirkweisen einer Abfolge zweier (PURPUR) oder mehrerer Lehrzyklen in Bezug auf den Lerneffekt durch die Wiederholung des Lehrzyklus und in Bezug auf den Transfer der Handlungskompetenzen in andere fachliche Unterrichtsthemen. Mit dem Dissertationsvorhaben, das ein Teil des SNF-Projekts PURPUR ist, soll die Unterrichtsqualität analysiert werden.

Stand der Forschung

- > Wichtiges Element zur Sicherstellung von gutem Unterricht ist insbesondere für Lehramtsstudierende die Vor- und Nachbereitung [1], [2]
- > Ältere [3], [4] wie auch aktuellere Studien [5] belegen die Schwierigkeit, die Unterrichtsqualität hinsichtlich der Lernwirksamkeit aufgrund von Unterrichtsbeobachtungen (Unterrichtsvideos) einzuschätzen
- > Die Unterrichtsqualität hängt vom Professionswissen [6], [7] und in geringerem Mass von anderen Dispositionen der professionellen Kompetenz [8] ab
- > Viele Lektionen der gleichen Lehrperson sind nötig, um eine zuverlässige Einschätzung vornehmen zu können [5]
- > Lerneffekte von Lehramtsstudierenden über mehrere Lehrzyklen hinweg sind kaum erforscht [9]

Fragestellung

Fragestellungen des Dissertationsvorhabens

1. Welchen Einfluss haben Aspekte der professionellen Kompetenz (kontextspezifisches Fach- und fachdidaktisches Wissen, Überzeugungen, Motivation, Enthusiasmus, Selbstregulation sowie Aspekte der Selbstwirksamkeitserwartung) auf die **fachdidaktische Unterrichtsqualität**?
2. Welche Lerneffekte (zwischen zwei Unterrichtseinheiten) zeigen angehende Lehrpersonen bei der **Unterrichtsdurchführung** in zwei verschiedenen Fachkontexten beziehungsweise zwei verschiedenen Testzeitpunkten hinsichtlich der **fachdidaktischen Unterrichtsqualität**?
3. Welche Zusammenhänge existieren zwischen dem **Unterricht** und den Leistungen der Schülerinnen und Schüler?

Methode

Stichprobe

- > Angehende Lehrpersonen Sekundarstufe 1 (Schwerpunkt Natur und Technik)
- > Praktikum im 6. bzw. 7. Semester
- > 3 Pädagogische Hochschulen der Schweiz
- > N=69, in Kodierung einbezogen: n=61 (28 m / 33 w)



Untersuchung zur Lösegeschwindigkeit von Vitamintabletten [10]

Kodierprozess

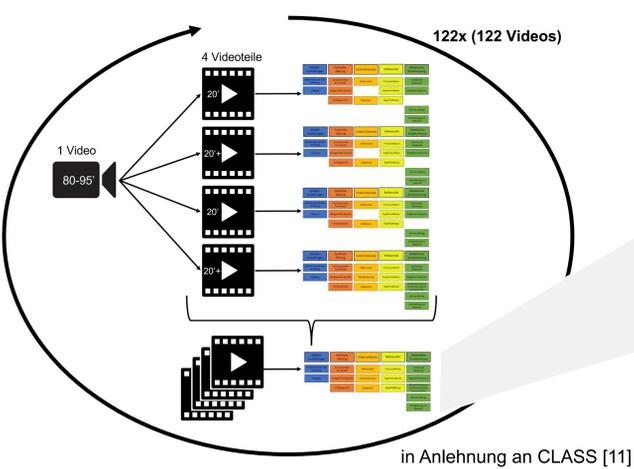
- > m = 7 Raterinnen an Kodierung beteiligt (Schulung sowie Objektivierung)
- > Um das Mass der Übereinstimmung beurteilen zu können, wird eine substantielle Menge von N = 96 Videoteilen von den m = 7 Raterinnen doppelt eingeschätzt.
- > Alle N = 488 Videoteile kodiert
- > Kodierverfahren ist holistisch und hochinferent

Daten

- > Durchführung von zwei Doppellektionen (à 90 Minuten) zu EXPH mit Fokus einmal Biologie und einmal Chemie → 2 x 90 min Unterrichtsvideos
- > Professionswissen sowie demografische Daten mit Fragebogen erhoben
- > Leistung der Schülerinnen und Schüler mit NAW-Test erhoben

Qualitätsraster für U

Vorgehensweise zur Auswertung der Videos



Qualitätsraster zur Analyse von Unterricht zum experimentellen Handeln

Schüler-vorstellungen	Fachliche Klärung	Unterrichtsziele	Reflexivität	Didaktische Strukturierung
Aktivierung und oder Erhebung	Konzeptionelle Korrektheit	Zielkonzepte	Prozessreflexion	Emotionale Situierung
Umgang	Sachgerechte Sprache	Zielorientierung	Ergebnisreflexion	Kognitiver Anspruch
	Schülergerecht	Zielklarheit	Begriffsbildung	Sequenzierung und Strukturierung des Lernprozesses
				Zeit am Auftrag
				Handhabung von Material

in Anlehnung an [7], [12], [13], [14]

Allgemeiner Aufbau von Indikator

Indikator nicht behandelt		Indikator behandelt			
99 (nicht messbar)	77 (kein EXPH)	0 (falsch)	1	2	3 (Norm)
- Nicht EXPH	- Nicht EXPH	- EXPH	- EXPH	- EXPH	- EXPH
- Indikator nicht thematisiert	- Indikator thematisiert	- Indikator wird falsch thematisiert	- Indikator wird genügend thematisiert	- Indikator wird gut thematisiert	- Indikator wird sehr gut thematisiert

Resultate

Hauptkodiererin mit jeweils einer weiteren Kodiererin (Ratername)

Ratername	*Gew. Kappa
BUM	0.932
SSC	0.903
FUG	0.915
ULK	0.889
GUC	0.884
HUJ	0.874

Gewichtetes Kappa (squared)

Subjekte: 20

Rater: 2

* Mittelwert von jeweils 16 Videoteilen

Unterscheidung von «nicht experimentellem Handeln» (99 und 77) und «experimentellem Handeln» (0-3)

Itemname	Cohen's Kappa	z	p-Wert	Übereinstimmung
SAK	0.589	5.77	7.76E-09	84%
SUM	1	9.8	0	100%
FKO	0.796	7.96	1.78E-15	96%
FSA	1	9.8	0	100%
FSU	0.564	5.68	1.35E-08	78%
UOR	NaN	NaN	NaN	100%
UKL	0.84	8.24	2.22E-16	97%
DEM	0.756	7.45	9.59E-14	93%
DKO	0.795	7.96	1.78E-15	99%
DSE	NaN	NaN	NaN	100%
DZE	NaN	NaN	NaN	100%
DMA	0.85	8.43	0	96%
RPR	0.7	6.88	5.98E-12	92%
RER	0.973	9.54	0	99%
RBE	0.714	7	2.63E-12	92%

Cohen's Kappa (oneway, absolute, single)

Subjekte: 96

Rater: 7

Unterscheidung von «nicht experimentellem Handeln» (99), «nicht EXPH» mit Indikator (77) und «EXPH» mit Indikator (0-3)

Itemname	Cohen's Kappa	z	p-Wert	Übereinstimmung
SAK	0.614	7.37	1.65E-13	84%
SUM	1	12.5	0	100%
FKO	0.751	9.17	0	95%
FSA	0.66	9.22	0	98%
FSU	0.572	6.93	4.09E-12	76%
UOR	NaN	NaN	NaN	100%
UKL	0.84	8.24	2.22E-16	97%
DEM	0.743	8.67	0	91%
DKO	0.795	7.96	1.78E-15	99%
DSE	NaN	NaN	NaN	100%
DZE	NaN	NaN	NaN	100%
DMA	0.816	8.82	0	95%
RPR	0.709	7.17	7.35E-13	92%
RER	0.974	9.92	0	99%
RBE	0.776	9.14	0	92%

Cohen's Kappa (oneway, absolute, single)

Subjekte: 96

Rater: 7

Unterscheidung von Abstufung des gezeigten Indikators (0-3)

Itemname	Subjekte	ICC	P-Wert	95%-Confidence Interval
SAK	64	0.775	1.17E-14	0.656 < ICC < 0.857
SUM	89	0.478	8.67E-07	0.301 < ICC < 0.623
FKO	83	0.594	1.12E-09	0.435 < ICC < 0.717
FSA	93	0.682	1.46E-14	0.557 < ICC < 0.777
FSU	42	0.549	5.93E-05	0.3 < ICC < 0.729
UOR	96	0.779	1.80E-21	0.687 < ICC < 0.847
UKL	9	0.789	0.00211	0.349 < ICC < 0.947
DEM	14	0.887	2.68E-06	0.693 < ICC < 0.962
DKO	93	0.516	4.43E-08	0.351 < ICC < 0.651
DSE	96	0.499	9.09E-08	0.333 < ICC < 0.635
DZE	96	0.83	2.77E-26	0.755 < ICC < 0.883
DMA	78	0.674	3.95E-12	0.533 < ICC < 0.779
RPR	12	0.434	0.0627	-0.135 < ICC < 0.794
RER	25	0.751	3.56E-06	0.516 < ICC < 0.881
RBE	13	0.837	5.59E-05	0.562 < ICC < 0.947

ICC (Model: oneway, Type: consistency, single)

Rater: 7

Interrater-Reliabilität

Interpretation

Die einzelnen Kodierinnen haben alle eine sehr gute (> .80) [15] Übereinstimmung mit der Hauptkodiererin. Die Aufteilung in «nicht EXPH» und «EXPH» gelingt auf Itemebene gut bis sehr gut. Auch die beiden Items «Aktivierung und oder Erhebung» sowie «Schülergerecht» weisen eine gute Übereinstimmung auf (> .50), da diese schwierig zu erfassen sind [15]. Wird «nicht EXPH» weiter aufgeteilt, um zu schauen, ob der Indikator thematisiert wird oder nicht, zeigt sich ein ähnliches Bild. Bei der Abstufung («falsch» = 0 bis «sehr gut» = 3) des Indikators mit «EXPH» ist die Übereinstimmung bei vielen Items gut [15]. Allgemein lässt sich sagen, dass die Interrater-Reliabilität, besonders in Bezug auf das holistische sowie hochinferente Kodierverfahren sehr hoch ist und somit eine Kodierung unabhängig der Person möglich ist.

Referenzen

[1] Staub, F. C., & Kreis, A. (2013). Fachspezifisches Unterrichtstraining in der Aus- und Weiterbildung von Lehrpersonen. *Journal für Lehrerinnenbildung*, 2, 8-13. [2] Stender, A., Brückmann, M., & Neumann, K. (2015). Vom Professionswissen zum kompetenten Handeln im Unterricht: Die Rolle der Unterrichtsplanung. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 33(1), 121-133. [3] Hugener, I., Pauli, C., & Reusser, K. (2007). Inszenierungsmuster, kognitive Aktivierung und Leistung im Mathematikunterricht. In D. Lemmermöhle, M. Rothgang, S. Bögeholz, M. Hasselhorn, & R. Watermann (Eds.), *Professionelles Lehren. Erfolgreich Lehren*, 109-122. Münster: Waxmann. [4] Kurzer, M., & Baumert, J. (2006). Who is the expert? Construct and criteria validity of student and teacher ratings of instruction. *Learning Environments Research*, 9(3), 291-293. [5] Prætorius, A. K., Pauli, C., Reusser, K., Rakoczy, K., & Klein, E. (2014). One lesson is all you need? Stability of instructional quality across lessons. *Learning and Instruction*, 31, 2-12. [6] Anholter, S. (2017). Förderung des fachspezifischen Professionswissens von Chemieamtstudierenden. Berlin: Logos. [7] Tardent, J. (2020). Unterrichtsplanungen von angehenden Lehrpersonen zum experimentellen Handeln. Heidelberg: Pädagogische Hochschule Heidelberg. Verfügbar unter: <https://opus.ph-heidelberg.de/handle/document/3333/8> [8] Kurzer, M., Baumert, J., Blum, W., Klumpp, U., Kraus, S., & Neubrand, M. (Eds.). (2011). *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTU*. Münster: Waxmann. [9] Anholter, S., & Rappert, D. (2018). Experimentell-fachdidaktisches Wissen und Handeln von Chemie Lehramtsstudierenden. In C. Maier (Ed.), *Authentizität und Lernen - das Fach in der Fachdidaktik*. Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Berlin 2018, 316-318. Regensburg: Universität Regensburg [10] Harmon, M., Smith, T. A., Martin, M. D., Kelly, D. L., Beaton, A. E., Mullis, I. V. S., Gonzalez, E. J., Orwood, G. (1999). *Performance assessment in IEA's Third International Mathematics and Science Study*. Chestnut Hill: Boston College. [11] Pianta, R. C., Haure, B. K., & Mintz, J. (2012). Classroom Assessment Scoring System (CLASS) Secondary Manual. Charlottesville: University of Virginia. [12] Kattmann, U., Duit, R., Gropengieser, H., & Komorek, M. (1997). Das Modell der Didaktischen Rekonstruktion - Ein Rahmen für naturwissenschaftsdidaktische Forschung und Entwicklung. *Zeitschrift für Didaktik der Naturwissenschaften*, 3, 3-18. [13] Wetzfel, H. (2015). Die Qualität von Unterrichtsplanungen zu hypothesengeleitetem Experimentieren analysieren. Vorstellung eines hoch inferenten Ratinginstruments und erster Ergebnisse. Gehalten auf dem 568. Kongress 2015: Qualitäts- und Bildungsdiskurs, St. Gallen. [14] Berlin, J. (2012). Das Experiment als Lerngelegenheit. Vom interkulturellen Vergleich des Physikunterrichts zu Merkmalen seiner Qualität. Berlin: Logos. [15] Witt, M., Caspar, F. (2002). Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratinglisten. Göttingen: Hogrefe.