

1 Entwicklung pädagogisch-soziologischer Diagnosekompetenz von Lehramtsstudierenden

Hendrik Baumbach & Wilfried Hansmann

Im Forschungsprojekt „Professionalisierung und Diagnosekompetenz – Kompetenzentwicklung und -förderung im Lehramtsstudium“ haben wir untersucht, wie Lehramtsstudierende anhand von audio- bzw. videographierten Unterrichtsszenen (Fallvignetten) Geschehnisse in schulischen und unterrichtlichen Kontexten beschreiben und analysieren. Insbesondere sind wir der Frage nachgegangen, inwieweit sie dabei über pädagogisch-soziologische Diagnosekompetenz verfügen und diese im Semesterverlauf erwerben bzw. verbessern können. Die Untersuchungen in den Experimentalgruppen wurden von Überlegungen der Problemlöseforschung geleitet (algorithmische Verfahren vs. Heuristiken, vgl. Funke & Zumbach 2006) sowie von der Hypothese, dass sich ungünstige Lernvoraussetzungen durch entsprechend aufwendig gestaltete Strategie-Instruktionen mit mehr Übungszeit und direkter Anleitung bis zu einem gewissen Grad kompensieren lassen (Bannert 2007: 109f.).

Nachdem in den vorhergehenden Beiträgen das Erkenntnisinteresse, die Forschungsfragen und das Forschungsdesign sowie die Durchführung der Untersuchung vorgestellt wurden, folgt nun die Erläuterung der Forschungsergebnisse aus messtheoretischer Perspektive. Die Einordnung der Ergebnisse in die aktuellen bildungs- und professionstheoretischen Diskurse zur Lehrer(aus)bildung wird im Kapitel III.2 vorgenommen.

Zunächst beschreiben wir die Ausprägungen pädagogisch-soziologischer Diagnosekompetenz der Probanden in Abhängigkeit von den verschiedenen Messzeitpunkten (Kap. 1.1). Daran schließen sich Darstellungen der Messergebnisse nach Kompetenzniveaus (Kap. 1.2), Fachsemestern (Kap. 1.3) und Geschlechtern (Kap. 1.4) an. Ein weiteres Kapitel ist den Ergebnissen von Probandengruppen gewidmet, die zu Beginn der Untersuchung zu den Leistungsschwächeren gehörten (Kap. 1.5). Um Auswirkungspotentiale förderdiagnostischen Feedbacks auf die Kompetenzentwicklung der Studierenden im Detail zu verdeutlichen, werden schließlich die Entwicklungsprozesse von zwei Studierenden miteinander kontrastiert (Kap. 1.6).

1.1 Die Kompetenzstände der Studierenden im Überblick

Die Testergebnisse geben Auskunft zu der Frage, über welche pädagogisch-soziologischen Diagnosekompetenzen Adressaten der Ersten Lehrerbildungsphase zu einem bestimmten Zeitpunkt ihres Studiums verfügen. Am Forschungsprojekt waren jeweils drei Experimental- (EG) und Kontrollgruppen (KG) beteiligt. Entscheidendes Kriterium für die Trennung zwischen EG und KG war die

Einführung und Erarbeitung bzw. Übung eines algorithmischen Diagnoseverfahrens, das anhand von schulischen und unterrichtlichen Fallbeispielen im Seminarkontext der Experimentalgruppen durchgeführt wurde. In den Kontrollgruppen fehlten ähnliche Übungsangebote. Alle Seminarteilnehmer der Experimental- und der Kontrollgruppen bearbeiteten jeweils einen Einstiegstest (EIN) zu Beginn und einen Abschlusstest (ABS) am Ende des jeweiligen Semesters bzw. Untersuchungszeitraums. Der in jeder Testung maximal erreichbare Kompetenzwert K_i wurde mit 100 normiert; erwartet wurden minimal 20 und maximal 80 Punkte (vgl. Hansmann et al. 2012; Baumbach 2012).

1.1.1 Die Kompetenzstände zu Beginn der Untersuchung

Der erste Test zur Ermittlung der Kompetenzstände wurde von den Studierenden zu Beginn des Wintersemesters 2010/11 in der zweiten bzw. dritten Semesterwoche bearbeitet. Bis zu diesem Zeitpunkt waren keinerlei Fallübungen o.ä. durchgeführt worden. Die Auswertung der Tests ergab, dass die Probanden über die verschiedenen Experimental- und Kontrollgruppen hinweg auf Skala 1 einen K_i -Mittelwert zwischen 21.84 [2.85] (EG2) und 32.91 [2.80] (EG3) erreicht hatten¹, d.h. sie befanden sich in der Ersttestung auf dem unteren Kompetenzniveau der Skala 1 („Beschreiben von Situationsvariablen, Praktiken und Effekten“, vgl. Dirks 2012a). Ein K_i -Wert von etwa 20 (Beginn des ersten Kompetenzniveaus) konnte erreicht werden, auch wenn nur eine gering ausgeprägte Diagnosekompetenz vorlag, da jedwede Bearbeitung der Testfrage zufällige Item-Treffer generiert; lagen Probanden mit einem K_i -Wert deutlich unterhalb von 20, ist zu vermuten, dass sie die Testaufgabe nicht in ernsthafter Weise bearbeitet haben. Insgesamt gelang es den Lehramtsstudierenden zum ersten Testzeitpunkt – wenn auch in begrenztem Maße – die in der Fallvignette explizit vorhandenen Informationen bei der Bearbeitung der Testaufgabe einzubeziehen; allerdings fanden die Situationsvariablen, d.h. die Voraussetzungen und Bedingungen der Situation, das Akteurshandeln sowie die aufgetretenen Effekte in den Analysen der Studierenden nur vereinzelt Berücksichtigung.

Auf Skala 2 des durchgeführten Tests (Prozesse des Deutenden Verstehens und Verstehenden Erklärens, vgl. ebd.) wurde ein K_i -Mittelwert gemessen, der zwischen 25.80 [2.74] (KG3) und 32.63 [2.52] (EG1) variierte und bei fünf der sechs Testgruppen leicht oberhalb der K_i -Mittelwerte auf Skala 1 lag. Unter Berücksichtigung der ermittelten Standardfehler ließ sich nur in EG2 ein auffälliger Unterschied zwischen den K_i -Mittelwerten der Ersttestung auf den beiden Subskalen feststellen. Die Lehramtsstudierenden erreichten somit auf der Subskala 2 ebenfalls insgesamt ein unteres Kompetenzniveau. Selbst wenn weniger Informationen aus dem Transkript entnommen wurden (Skala 1), gelang es den Probanden, ihre Analysen durch Inferenzen anzureichern. Die Kürze des Transkripts, das neben einer Lehrerin nur zwei Schüler jeweils einmal zu Wort

1 Standardfehler sind in eckigen Klammern angegeben.

kommen ließ, und die Offenheit in der Rahmung der geschilderten Situation eröffnete den Probanden Freiräume für ihre individuellen Interpretationen bzw. Reflexionen (vgl. Deutendes Verstehen und Verstehendes Erklären; zu den fallbezogenen Interpretationspotentialen s. Dirks 2012b).

Zusammenfassend ergab die Testung zu Semesterbeginn, dass die Probanden auf beiden Subskalen ein unteres Kompetenzniveau erreichten. Für die Zweittestung erwarteten wir eine Steigerung der diagnostischen Kompetenz insbesondere in den Experimentalgruppen, wobei allerdings EG3 im Unterschied zu EG1 und 2 zwar in das algorithmische Diagnoseverfahren eingeführt, dieses aber danach nicht weiter geübt wurde.

1.1.2 Die Kompetenzstände in der Zweittestung nach EG/KG auf Skala 1

Die nachfolgende Abbildung 1 zeigt in der Übersicht die Kompetenzstände der Studierenden im Abschlusstest in Form der K_1 -Werte als Boxplots auf Skala 1 („Beschreiben von Situationsvariablen, Praktiken und Effekten“):

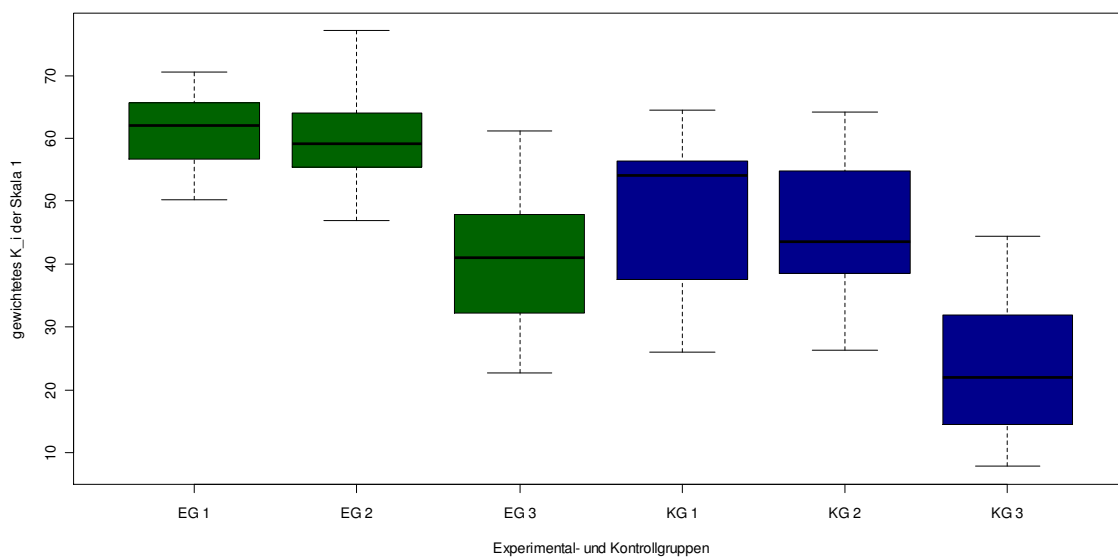


Abb. 1: Ergebnisse der Abschlusstestung nach Experimental- und Kontrollgruppen auf Skala 1; EG1-3 (Experimentalgruppen), KG1-3 (Kontrollgruppen)

Unter Berücksichtigung der ermittelten Standardfehler werden von links nach rechts drei Niveaunklassen erkennbar, denen sich die einzelnen Testgruppen zuordnen lassen:

- Einen K_1 -Mittelwert von etwa 60 erreichen EG1 (60.96) und EG2 (59.62).
- Ein K_1 -Mittelwert zwischen 40 und 50 ist in EG3 (41.49), KG1 (47.85) und KG2 (45.26) gemessen worden.
- KG1 erreicht einen deutlich niedrigeren K_1 -Mittelwert von nur 23.43.

Zunächst fällt auf, dass die Kontrollgruppe KG1, die sich weder mit unterrichtlichen bzw. schulischen Situationen noch mit der Bearbeitung von

Fallvignetten oder mit pädagogischer Diagnostik beschäftigt hatte, auf dem Niveau der Ersttestung verblieb. Diesen Probanden gelang es nicht, die fallimmanenten Informationsstrukturen der Situation, der Praktiken der Akteure und deren Effekte genauer zu beschreiben und somit dem Transkript einen höheren Anteil an Informationen zu entnehmen als bei der Ersttestung.

Im Gegensatz zum Einstiegstest erzielten die Kontrollgruppen KG1 und KG2 im Abschlusstest einen deutlich höheren K_i -Mittelwert. Die Probanden der beiden Gruppen hatten zwar nicht das algorithmische Diagnoseverfahren kennengelernt und geübt, sich aber stattdessen im Seminarkontext auf der Basis ihres Vorwissens und verschiedener Theorieangebote mit unterrichtlichen und schulischen Situationen, dem Lehrerhandeln sowie der Lehrerrolle und den einzelnen Lehrertypen auseinandergesetzt. Gegenstand ihrer Lehrveranstaltungen war u.a. die Thematisierung von Unterrichtsqualität. Den Probanden dieser Testgruppen gelang es auch ohne den Einsatz von Algorithmen, annähernd die Hälfte der erwarteten Informationen aus der Fallschilderung zu entnehmen und diese beschreibend analytisch zu bearbeiten. Offensichtlich hat die Seminararbeit in diesen Kontrollgruppen, die auf den Erwerb von Kenntnissen zu bestimmten Unterrichtssituationen und Handlungsbedingungen (Klassenstufe, -größe, Schulart, Arbeitsmaterialien, Unterrichtsmethodik, Schüler-Lehrer-Interaktion etc.) ausgerichtet war, die Bearbeitung der Aufgabe im Abschlusstest befördert, sodass die zugehörigen Items erfüllt wurden. Anscheinend können sich die Erfassung und Beschreibung schulischer und unterrichtlicher Situationen, d.h. die auf Skala 1 gemessenen Diagnosekompetenzen, demnach auch durch eine detaillierte Betrachtung und Besprechung von typischen Situationsvariablen des Lehrer-/Schülerhandelns verbessern, wenngleich nicht in dem Maße wie in den Experimentalgruppen.

Denn signifikant besser schnitten auf Skala 1 innerhalb des durchgeführten Abschlusstests im Mittelwert diejenigen Gruppen ab, die das algorithmische Diagnoseverfahren erlernt und sukzessive anhand von verschiedenen Beispielen geübt hatten. Diesen Probanden gelang eine umfassende Beschreibung der in der Fallvignette dargelegten Situation, d.h. ihren Analysen lagen umfangreichere Informationen hinsichtlich der situativen Variablen, dem Handeln der Akteure und deren Wirkungen zugrunde. Hinzuweisen ist ferner darauf, dass die K_i -Werte der Probanden von EG1 und EG2 homogener erscheinen und oberes und unteres Quartil näher zusammen liegen als in den übrigen Testgruppen.

Für EG3 zeigte der Abschlusstest auf der ersten Subskala nur eine mittlere Verbesserung des K_i -Mittelwerts, obwohl die Probanden in das algorithmische Diagnoseverfahren eingeführt worden waren. Offensichtlich gelang es ihnen nur in begrenztem Umfang, diesbezügliche Wissensbestände in der Zweittestung zu reaktivieren und anzuwenden. Ihre Kompetenzsteigerung auf Skala 1 lag lediglich auf dem Niveau von KG1 und KG2. Als eine Hauptursache mag hier die fehlende Übung des Verfahrens anhand unterschiedlicher Fallvignetten zu Buche schlagen.

Zusammenfassend kann festgestellt werden, dass die Experimentalgruppen EG1 und EG2 im Abschlusstest ihren K_1 -Mittelwert auf der Skala 1 deutlich steigerten; in Einzelfällen verdoppelten bis verdreifachten Probanden sogar ihre Testergebnisse, was einem mittleren Effekt von 0.47 (EG1) bzw. 0.39 (EG2) auf der Cohen-Skala entspricht. Geringe positive Effekte sind demgegenüber in den Kontrollgruppen KG1 und KG2 mit Werten von 0.09 bzw. 0.12 zu beobachten; die Probanden hatten das Diagnoseverfahren nicht kennengelernt und nicht geübt. Keinerlei Effekte waren in EG3 und KG3 messbar.

Unser Fazit lautet: Die bestmögliche Entwicklung pädagogisch-soziologischer Diagnosekompetenz auf der ersten Subskala kann offensichtlich zurückgeführt werden auf die Verknüpfung von theoriegeleiteter Einführung des algorithmischen Diagnoseverfahrens mit deren praktischer Erprobung und Anwendung anhand mehrerer Fallvignetten unter Berücksichtigung berufsfeldspezifischer Kernprobleme.

1.1.3 Die Kompetenzstände in der Zweittestung nach EG/KG auf Skala 2

Die zweite Abbildung veranschaulicht die beobachteten K_1 -Werte über die einzelnen Probandengruppen auf der Skala 2, welche die diagnostischen Kompetenzen des Deutenden Verstehens und Verstehenden Erklärens umfasst. Auf dieser Subskala wurde gemessen, inwieweit die Lehramtsstudierenden fehlende Informationen inferieren und dabei angemessene Brücken- bzw. Aggregationshypothesen mit Bezug auf die im Transskript appräsentierten Handlungs-/Wirkungszusammenhänge bilden konnten (vgl. Dirks 2012a).

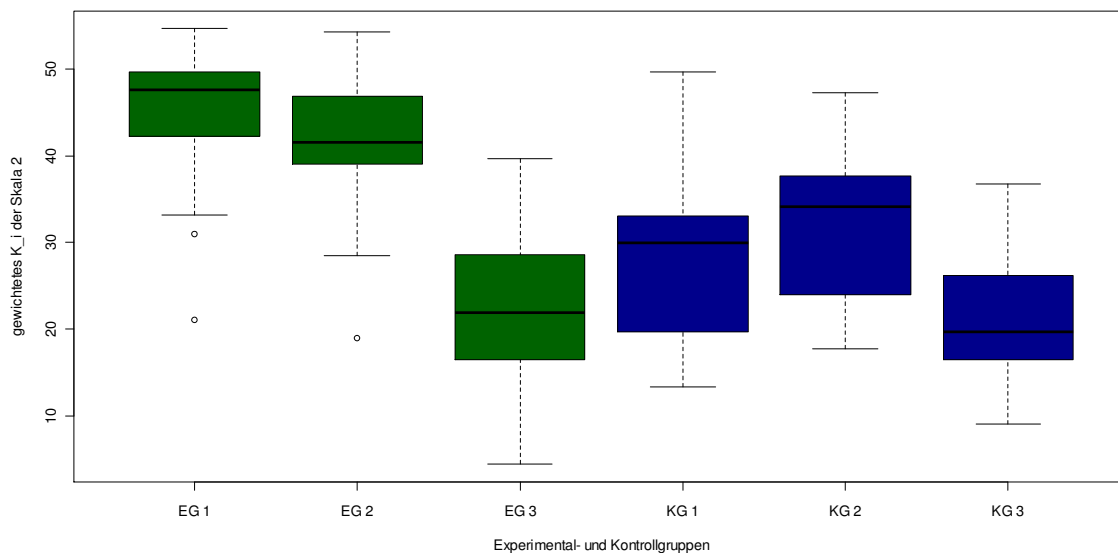


Abb. 2: Ergebnisse der Abschlusstestung nach Experimental- und Kontrollgruppen auf Skala 2; EG1-3 und KG1-3

Auf den ersten Blick liegen die auf Skala 2 beobachteten diagnostischen Fähigkeiten der einzelnen Probandengruppen deutlich näher beieinander als jene auf der Skala 1. Dies zeigt sich anhand der relativen Ranking-Platzierung der einzelnen Gruppen, wonach – abgesehen von einer Vertauschung von KG1 und

KG2 – die übrigen Testgruppen der Reihenfolge auf Skala 1 folgen. Jedoch muss auf zahlreiche z.T. kleine und angesichts der geringen Stichprobengröße hypothetisch zu formulierende Unterschiede hingewiesen werden. Signifikant ist zunächst die Abweichung der K_i -Werte in der absoluten Sichtweise zwischen den beiden Skalen:

- Während die EG1 und EG2 auf Skala 1 noch einen K_i -Wert von etwa 60 erreichen, bleiben die Werte der Probanden auf Skala 2 bei etwa 40 bis 50 Punkten stehen. In Form einer linearen Transformation folgen die übrigen Gruppen.
- Dabei wird die Abstufung zwischen EG3, KG1 und KG2 auf der einen sowie KG3 auf der anderen Seite relativiert.
- Insbesondere zeigt die in EG3 durchgeführte Vorstellung des algorithmischen Diagnoseverfahrens keinerlei Effekt auf Skala 2; ihr K_i -Mittelwert liegt nicht signifikant über dem der KG3.

Rätsel geben des weiteren der beobachtete Unterschied zwischen den KG1 und KG2 auf sowie die drei Ausreißer in den Gruppen EG1 und EG2, die ihre gute Beschreibung der Situationsvariablen, des Akteurshandelns und der beobachteten Effekte (Steps I bis III auf Skala 1) nicht in deren Verstehens- und Erklärenszusammenhänge (Skala 2) übertragen konnten.

Eine Ursache für diese Befunde liegt möglicherweise im unterschiedlichen Umfang der in beiden Testungen eingesetzten Fallvignetten (vgl. Dirks 2012b). Die Ersttestung wurde im Vergleich zum Abschlusstest mit einem sehr viel kürzeren Transkript durchgeführt; die hier vorliegende überschaubare Informationsmenge (vgl. Item-Pool der ersten Subskala, Dirks et al. 2012) hat mutmaßlich zur Bildung von Inferenzen angeregt, die der zweiten Subskala zugeordnet sind. Demgegenüber enthielt die in der Zweittestung eingesetzte Fallvignette (vgl. Anhang, ebd.) umfangreichere Informationen, deren sorgfältige Ermittlung dazu führte, dass Inferenzen (Skala 2) allein aus Gründen einer begrenzten Bearbeitungszeit in deutlich geringerem Umfang gebildet werden konnten.

Denkbar ist jedoch auch, dass Probanden davon ausgegangen sind, es sei für die Analyse der Fallvignette ausreichend, die Situation zu beschreiben. Sollte diese Vermutung zutreffen, haben sich die Probanden möglicherweise dazu verleiten lassen, sich auf einzelne Details zu kaprizieren und dabei die Aggregationslogik (Skala 2) zu vernachlässigen. Zudem könnte die leicht negative Entwicklung des K_i -Mittelwertes in EG3 und KG3 dem veränderten Verhältnis der Item-Anzahl auf den jeweiligen Subskalen in der ersten und zweiten Testung geschuldet sein. Schließlich ist der geringe Kompetenzzuwachs auf Skala 2 in den Experimentalgruppen 1 und 2 im Vergleich zu Skala 1 möglicherweise auch auf Schwerpunktsetzungen der Fallarbeit innerhalb der Steps I bis III (vgl. erste Subskala) zurückzuführen.

Innerhalb der EG1 und EG2 wurden, wie Abbildung 2 zeigt, insgesamt drei Probanden als Ausreißer nach unten beobachtet. Jan² (EG1) erreichte in beiden Testungen mit 22.91 (EIN) und 21.02 (ABS) den gleichen K_i -Wert; nicht viel anders sah es bei Julia (EG2) mit den Werten 19.96 (EIN) und 18.91 (ABS) aus. In beiden Fällen konnte keine Verbesserung der diagnostischen Fähigkeiten auf Skala 2 gemessen werden. Tobias (EG1) steigerte sich immerhin von 16.33 (EIN) auf 30.97 (ABS). Werden diese drei Probanden isoliert betrachtet, so korrelieren die Differenzen der K_i -Werte im Einstiegs- und Abschlusstest im Vergleich der beiden Subskalen nur mit 0.17³, d.h. ihre K_i -Werte verändern sich zwischen Erst- und Zweittestung auf Skala 2 anders als auf Skala 1. Auf Skala 1 erreichen alle drei Probanden im Abschlusstest einen K_i -Wert zwischen 50 und 60 und verbessern sich damit deutlich gegenüber dem Einstiegstest. Dass sich diese Tendenz nicht auf Skala 2 überträgt, mag der begrenzten Bearbeitungszeit des Tests geschuldet sein. Während alle drei Probanden bei der in der Analyse zuerst vorgenommenen Beschreibung der Situation (Steps I–III) noch deutlich Zuwächse in ihrem K_i -Wert aufweisen, könnte die Bearbeitungszeit für die Steps IV und V bereits sehr knapp geworden sein, sodass viele Brücken- und Aggregationshypothesen bis zum Ende der Bearbeitungszeit ungenannt blieben. Insbesondere die dem algorithmischen Diagnoseverfahren immanente schrittweise Fallanalyse kann bei begrenzter Bearbeitungszeit gerade innerhalb der Experimentalgruppen zu einer Unterschätzung der tatsächlichen Kompetenzwerte auf Skala 2 führen. Die nachgewiesene deutliche Verbesserung in den Steps I bis III zeigt sich bei den Probanden der Experimentalgruppen in der erhöhten Anzahl von Item-Treffern, die im offenen Antwortformat zumeist mehr Bearbeitungszeit verbraucht haben dürften als bei den Testpersonen in den Kontrollgruppen. Demnach stand den Probanden der Experimentalgruppe möglicherweise sogar etwas weniger Zeit zur Verfügung, Items innerhalb der Steps IV und V zu erfüllen.

Eine weitere Ausdifferenzierung der Forschungsbefunde für die Skala 2 soll nun mit Blick auf die Kompetenzniveaus erfolgen.

1.2 Darstellung der Ergebnisse nach Kompetenzniveaus

Außerhalb der bisher für die einzelnen Testgruppen vorgestellten quantitativen Befunde in Form von gewichteten K_i -Werten lassen sich die gemessenen Kompetenzentwicklungen auch anhand von Kompetenzniveaus beschreiben. Den folgenden Ausführungen liegen die tabellarischen Übersichten der Kompetenzniveaus (vgl. Dirks 2012b) zugrunde. In der Tabelle 1 sind die Befunde unterteilt nach den einzelnen Subskalen als gewichtete K_i -Mittelwerte dargestellt.

2 Um die Anonymität der Probanden zu gewährleisten, wurden die Vornamen maskiert.

3 Die Korrelationen zwischen Skala 1 und Skala 2 im Einstiegstest beträgt für die gesamte Testkohorte 0.17, im Abschlusstest beträgt sie 0.69.

1.2.1 Befunde unterteilt nach einzelnen Subskalen

Den Probanden der Testgruppen EG1 und EG2 gelang in der Anwendung des algorithmischen Diagnoseverfahrens eine deutliche Steigerung ihrer Fähigkeiten. Während sie der Fallvignette zu Semesterbeginn nur Informationen mit einfacher Verknüpfungsstruktur entnommen und in der Regel nur einzelne Dimensionen der im Fall appäsentierten sozialen Realität analysiert hatten, wurde in der abschließenden Testung ein mittleres bis oberes Kompetenzniveau erreicht. Die Probanden erkannten in der deutlich längeren Fallvignette auch solche Informationen, die nur schwer im Transkript aufzufinden sind, und erzielten eine insgesamt stärker ausdifferenzierte Diagnose als in der Ersttestung.

		max	EG1	EG2	EG3	KG1	KG2	KG3
K_i-Mittelwert Skala 1	EIN	100	29.92 [2.15]	21.84 [2.85]	32.91 [2.80]	30.22 [4.34]	25.09 [4.07]	24.67 [3.96]
	ABS	100	60.96 [1.31]	59.62 [1.84]	41.49 [1.89]	47.85 [3.34]	45.26 [3.50]	23.43 [3.20]
	DIF		+ 31.03	+ 37.78	+ 8.58	+ 17.64	+ 20.17	- 1.24
Kompetenzniveau Skala 1	EIN		I	0 - I	I	I	I	I
	ABS		II - III	II - III	I - II	II	II	0 - I
	DIF		1 - 2 Niv.	2 Niv.	0 - 1 Niv.	1 Niv.	1 Niv.	0 Niv.
K_i-Mittelwert Skala 2	EIN	100	32.63 [2.52]	28.74 [2.85]	28.49 [1.98]	31.34 [3.84]	29.71 [3.36]	25.80 [2.74]
	ABS	100	44.84 [1.89]	40.54 [2.17]	21.63 [1.62]	29.45 [2.93]	32.41 [2.55]	21.35 [2.28]
	DIF		+ 12.22	+ 11.80	- 6.52	- 1.89	+ 2.70	- 4.45
Kompetenzniveau Skala 2	EIN		I	I	I	I	I	I
	ABS		II	I - II	I	I	I	0 - I
	DIF		1 Niv.	0 - 1 Niv.	0 Niv.	0 Niv.	0 Niv.	0 Niv.

Tab. 1: EG1-EG3 und KG1-KG3; Standardfehler in Klammern

Die Probanden der KG1 und KG2 erreichten gleichermaßen eine Steigerung um ein Kompetenzniveau auf einer mittleren Ebene der ersten Subskala. Auch bei ihnen scheint sich die Fähigkeit, Informationen aus der Fallvignette zu entnehmen und beschreibend wiederzugeben, zwischen den beiden Erhebungen verbessert zu haben. Die Vermittlung didaktischer Modelle und Anleitungen in den Kursen EG3, KG1, KG2 hat aber nicht zu einer Steigerung der Diagnosekompetenz geführt, wie sie in den Experimentalgruppen 1 und 2 festgestellt werden konnte. Auf der zweiten Subskala verringern sich dann allerdings die auf Skala 1 für die Kontrollgruppen beobachteten Effekte. Die Probanden verbleiben auf dem unteren Kompetenzniveau, während die Probanden der EG1 und EG2 eine signifikante Steigerung bis zum mittleren Kompetenzniveau erreichen. Diese Probanden haben im ersten Test ohne Anwendung des algorithmischen Verfahrens nur einzelne fehlende Informationen in angemessener Weise inferiert und in einen aggregationslogischen Zusammenhang gebracht. In der zweiten Erhebung fällt auf, dass insbesondere solche naheliegenden Inferenzen häufiger gebildet wurden, wenn

diese sich z.B. auf die Schulform oder Klassenstufe beziehen und sich durch einen detailliert ausdifferenzierten kausal-funktionalen Rückbezug auf die im Transkript explizit enthaltenen Informationen auszeichnen. In der abschließenden Testung gelingt den Probanden also eine stärkere Verknüpfung von Bedingungen der Situation, dem expliziten Handeln der Akteure und dessen unmittelbaren Wirkungen mit Inferenzen der Steps IV und V.

1.2.2 Befunde nach vektoriellen Skalen

Die Kompetenzentwicklung der Probanden lässt sich schließlich noch gemäß des Ansatzes der *vektoriellen Skalen* (Hagemann 2002: 47ff.) illustrieren, indem beide Subskalen orthogonal aufeinander bezogen werden. Die Veränderung zwischen erster und zweiter Testung wird als Pfeil markiert. Die idealtypische Entwicklung der pädagogisch-soziologischen Diagnosekompetenz – hier in der Zusammensetzung aus beiden Subskalen – vollzieht sich entlang der Hauptdiagonale, also gleichmäßig auf den beiden Subskalen. Aus diesem Grund soll nicht die Differenz zwischen Einstiegs- und Abschlusstest die Kompetenzveränderung bestimmen – hier ausgedrückt in der Länge des Pfeils –, sondern wiederum ein Niveaumodell, dessen Niveaus senkrecht zur Hauptdiagonale verlaufen (vgl. Baumbach 2012). Das bedeutet, dass die gemessene Veränderung in den beiden Erhebungen als besonders günstig im Sinne des Kompetenzmodells zu bezeichnen ist, wenn sie auf beiden Subskalen in gleichem Ausmaß stattfindet. Auf eine ausformulierte Beschreibung dieser Niveaus ist verzichtet worden, da mit der Unterteilung nach Teilkompetenzen, die sich in den beiden Subskalen widerspiegelt, keine weitere Erläuterung dieser Niveaus vorgenommen werden kann, die nicht schon in den skalenbezogenen Tabellen geleistet wurde (vgl. Dirks 2012b).

Die nachstehende Abbildung 3 zeigt die Entwicklung der pädagogisch-soziologischen Diagnosekompetenz aller Probanden der EG1 in einem kompensatorisch angelegten Niveauschema. Die Unterteilung in nunmehr sieben Niveaus (N1–N7) resultierte aus der Einteilung der beiden Subskalen in fünf Abschnitte (KN0, KN1, KN2, KN3, KN3+).

Die Probanden erreichen in der ersten Erhebung zu Beginn des Semesters überwiegend das zweite oder dritte Niveau und steigern sich in der abschließenden Testung, nachdem sie das algorithmische Diagnoseverfahren im Seminar kennengelernt und geübt haben. Der Mittelwert verdeutlicht, dass die Kompetenzentwicklung in dieser Testgruppe an die idealtypische Entwicklungslinie, d.h. an die Hauptdiagonale angenähert ist, wobei die Veränderung auf Skala 1 größer ausfällt. Einzelne Probanden der EG1 verbessern sich – immer bezogen auf die anteiligen, gewichteten K_i -Werte – auf der zweiten Subskala nicht oder verschlechtern sich sogar leicht, wie einzelne waagerechte oder leicht nach rechts hin absinkende Pfeile verdeutlichen. Letztere kompensieren dies jedoch mit unverkennbar gesteigerten Fähigkeiten auf der ersten Subskala, sodass sie insgesamt ihre diagnostische Kompetenz um mindestens ein Niveau verbessern. In

der Summe erreichen die Probanden in der Zweittestung das obere vierte oder das fünfte Niveau:

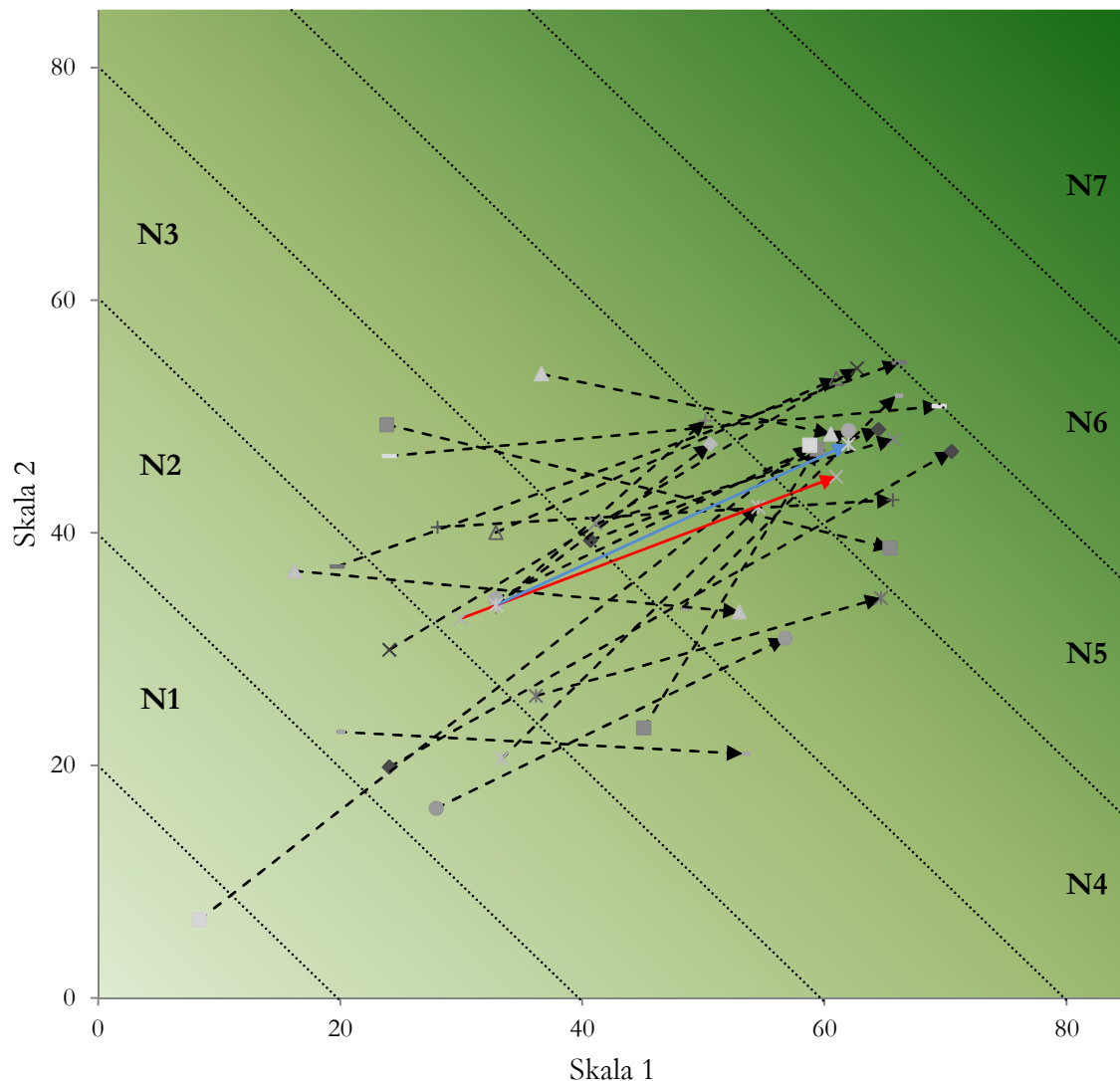


Abb. 3: K_T -Werte der Probanden aus EG1 von der Erst- zur Zweittestung in sieben Kompetenzniveaus (N1 - N7) auf beiden Subskalen; Mittelwert in rot, Median in blau

Werden die Befunde aus Abbildung 1 in dieses Schema mit sieben Niveaus übertragen, dann erreichen alle Testgruppen in der Fallanalyse zu Beginn des Tests das zweite Niveau, wobei EG1 und KG1 einschließlich der Fehlerrechnung dem oberen Bereich hin zum dritten Niveau zugehörig sind. Im abschließenden Test steigern sich die beiden Experimentalgruppen, in denen das algorithmische Diagnoseverfahren eingeführt und geübt worden ist, bis auf N5 (EG1) bzw. N4-5 (EG2). Dies entspricht einem Kompetenzzuwachs von zwei bis drei Niveaus. Die EG3, KG1 und KG2 verbessern sich nur geringfügig um maximal ein Niveau, während für die KG3 keine signifikante Veränderung des Kompetenzniveaus messbar ist.

Bei der Interpretation dieser geringfügigen Kompetenzsteigerungen müssen auch die Sensibilisierung der Probanden durch die Cover-Story (vgl. Hansmann et al.

2012) und die nicht auszuschließenden, auf das Testformat bezogenen Übungseffekte berücksichtigt werden. Die gemessenen Zuwächse innerhalb der Kompetenzniveaus verstehen sich damit als eine obere Schranke der tatsächlichen Fähigkeitszunahmen. Um Übungseffekte in EG1 und EG2 soweit wie möglich auszuschließen bzw. zu minimieren, wurden die Probanden mit ganz unterschiedlichen Fallvignetten konfrontiert; dies betrifft v.a. die Länge der Vignetten, deren Bezug zum Unterrichtsfach, ‚best‘- und ‚worst-practise‘-Beispiele sowie die Fokussierung auf das Lehrer- oder Schülerhandeln. Das im Einzelnen gewählte Antwortformat und die Fragestellung wurden in den Übungsfällen ebenfalls variiert und auf diese Weise Übereinstimmungen mit den Erhebungsinstrumenten weitestgehend vermieden. Forschungsbefunde belegen aber, dass die Schwierigkeit eines Tests und auch die Schwierigkeit einer Fallvignette u.a. vom Vorwissen und von den Vorerfahrungen der Probanden mit einem bestimmten Test- und Textformat abhängig ist (vgl. Dirks 2012b). Demnach kann nicht ausgeschlossen werden, dass die Arbeit mit Fallvignetten in EG1 und EG2 den zugeordneten Probanden kleinere Vorteile in der abschließenden Testung verschafft hat.

1.3 Ergebnisse nach Fachsemestern

Dass sich die diagnostischen Fähigkeiten und Fertigkeiten von Lehramtsstudierenden innerhalb ihres Studiums sukzessive verbessern, ist eine naheliegende Hypothese. Zu ihrer Überprüfung wurde die Stichprobe in eine Gruppe der Erst- bis Drittsemester und eine zweite der höheren Semester disjunkt untergliedert. Diese Aufteilung orientiert sich an der modularen Studienstruktur des Erziehungs- und Gesellschaftswissenschaftlichen Studiums für das Lehramt (EGL) an der Philipps-Universität Marburg. Die geltende Studienordnung verlangt von den Studierenden innerhalb der ersten vier Semester die Absolvierung des Einführungsmoduls EGL 1 („Einführung in theoretische Grundlagen des Berufsfeldes“) und des Moduls EGL 2 („Rezeption und Reflexion von Schule“) im Umfang von insgesamt 10 Semesterwochenstunden. Mit dem Modul EGL 2 ist ein fünfwöchiges Schulpraktikum in einem Block innerhalb der vorlesungsfreien Zeit verbunden, welches in einem vierstündigen Seminar vorbereitet und mit einem weiteren Seminar begleitet wird (vgl. EGL-M). In unserer Untersuchung ist bewusst auf eine Untergliederung nach Semesterzahl verzichtet worden, da nahezu alle Probanden der Teststichprobe bereits weitere Seminarveranstaltungen aus den übrigen Modulen des EGL-Studiums besucht hatten bzw. besuchten.

Die folgende Tabelle 2 zeigt zusammenfassend in Bezug zu den zwei Subskalen die beiden nach Fachsemestern unterschiedenen Gruppen jeweils für alle Testgruppen (EG u. KG) und dann gemeinsam für EG1 und EG2. Sowohl auf Skala 1 als auch auf Skala 2 erreichen die Probanden unabhängig von ihrem Fachsemester einen etwa gleichen Kompetenzwert in der ersten Testung. Das bedeutet zunächst einmal, dass die untersuchten Bereiche der diagnostischen Kompetenz nicht oder nur unwesentlich eher beiläufig im Rahmen der Marburger Lehrerbildung von den

Studierenden erworben werden. Sind konkrete Veranstaltungsangebote zur Ausbildung schulbezogener diagnostischer Fähigkeiten und Fertigkeiten vorhanden, in denen die theoretische Einführung mit unmittelbarer Anwendung auf exemplarische Fälle verbunden wird, dann – so der isolierte Blick auf EG1 und EG2 – steigern die Studierenden im niedrigen und im höheren Semester ihren Kompetenzwert deutlich. Besonders hervorzuheben ist, dass das algorithmische Diagnoseverfahren in den EG1 und EG2 unabhängig vom Fachsemester auf beiden Subskalen zu einer gleichwertigen Verbesserung geführt hat. So ist für die Probanden der EG1 und EG2 zusammengenommen auf Skala 1 ein Anstieg des Kompetenzwertes um 32.56 Punkte bei den Studierenden des ersten bis dritten Fachsemesters und um 34.52 Punkte für Probanden höherer Fachsemester messbar. Auf Skala 2 bleiben die Zuwächse von 12.47 Punkten bzw. 11.58 Punkten auch deutlich über den Werten aller Testgruppen, jedoch abermals reichlich geringer als auf der ersten Subskala, wie die Effekte nach der Cohen-Skala dokumentieren.

				Fachsem.	Fachsem.	Fachsem.	Fachsem.
				1-3	1-3	> 3	> 3
				EG u. KG	EG1 u. EG2	EG u. KG	EG1 u. EG2
				N=44	N=9	N=62	N=29
Skala 1	K _i -Mittelwert	EIN	100	30.93 [2.26]	25.24 [3.28]	26.52 [1.68]	26.64 [2.21]
		ABS	100	41.55 [2.20]	57.80 [2.30]	51.54 [1.82]	61.15 [1.22]
		DIF		+ 10.62	+ 32.56	+ 25.02	+ 34.52
	σ	EIN		15.00	9.84	13.19	11.91
		ABS		14.61	6.89	14.37	6,55
		Cohen's d		+ 0.05	+ 0.45	+ 0.13	+ 0.37
Skala 2	K _i -Mittelwert	EIN	100	28.90 [1.74]	29.73 [4.16]	30.06 [1.47]	31.24 [2.15]
		ABS	100	26.87 [1.83]	42.58 [3.77]	35.01 [1.55]	43.02 [1.54]
		DIF		- 2.04	+ 12.85	+ 4.95	+ 11.78
	σ	EIN		11.56	12.47	11.58	11.58
		ABS		12.12	11.30	12.19	8.30
		Cohen's d		- 0.01	+ 0.09	+ 0,04	+ 0.12

Tab. 2: Kompetenzentwicklung nach Fachsemestern in zwei Gruppen (Erst- bis Drittsemester und Lehramtsstudierende ab dem vierten Semester); Standardfehler in Klammern

Anzumerken ist ferner, dass die beobachteten Kompetenzzuwächse in der Unterteilung nach Fachsemestern auf die Zusammensetzung der einzelnen Testgruppen zurückzuführen sind. Nach Tabelle 1 verbesserten zwar alle getesteten Lehramtsstudierenden höherer Fachsemester ihre K_i-Werte im Mittelwert um 25.02 Punkte auf Skala 1 positiv, wohingegen die Probanden im ersten bis dritten Semester nur eine Steigerung von 10.62 Punkten erreichten. Sieht man jedoch den größeren Anteil von Probanden aus EG1 und EG2 in der Gruppe der

Studierenden im höheren Fachsemester,⁴ dann relativiert sich die Beobachtung, dass sich die Gruppe der fortgeschrittenen Lehramtsstudierenden im Allgemeinen stärker verbessert als die Gruppe der Studienanfänger.

Insgesamt ist verifizierbar, dass die Fachsemesterzahl der Lehramtsstudierenden keinen nachweisbaren Effekt auf die selbstständige oder angeleitete Ausbildung pädagogisch-soziologischer Diagnosekompetenz hat.

1.4 Ergebnisse nach Geschlechtern

Die Erprobung des algorithmischen Diagnoseverfahrens sollte auch darüber Auskunft geben, welchen Einfluss Genderspezifika auf die Entwicklung pädagogisch-soziologischer Diagnosekompetenz bei Lehramtsstudierenden haben. Tabelle 3 differenziert daher nach Geschlechtern und vergleicht die gemessenen Kompetenzwerte K_i der Probanden, die das Analyseverfahren nach dem Ansatz von Dirks & Hansmann kennengelernt und geübt haben (EG1 u. EG2), mit denen aller Testgruppen (EG u. KG) im Mittelwert.

				weiblich	weiblich	männlich	männlich
				EG u. KG	EG1 u. EG2	EG u. KG	EG1 u. EG2
				N=66	N=20	N=41	N=18
Skala 1	K_i-Mittelwert	EIN	100	29.95 [1.60]	24.66 [2.21]	26.00 [2.41]	28.13 [3.02]
		ABS	100	46.81 [1.91]	60.76 [1.45]	47.25 [2.30]	59.91 [1.67]
	σ	DIF		+ 16.86	+ 36.10	+ 22.26	+ 31.78
		EIN		12.97	9.88	15.46	12.82
	Cohen's d	ABS		15.49	6.48	14.75	7.08
				+ 0.08	+ 0.52	+ 0.10	+ 0.30
Skala 2	K_i-Mittelwert	EIN	100	30.63 [1.54]	32.63 [3.19]	27.73 [1.49]	28.94 [1.81]
		ABS	100	31,13 [1.50]	41.88 [1.97]	31.88 [2.20]	44.08 [2.17]
	σ	DIF		+ 0.50	+ 9.24	+ 4.15	+ 15.13
		EIN		12.49	14.29	9.56	7.70
	Cohen's d	ABS		12.17	8.80	14.09	9.19
				+/- 0.00	+ 0.07	+ 0.03	+ 0.21

Tab. 3: Vergleich der Kompetenzentwicklung nach Geschlechtern; Standardfehler in Klammern

Die aus den Prätest-Ergebnissen abgeleitete Vermutung, dass Lehramtsstudentinnen unter Zuhilfenahme des algorithmischen Vorgehens eine in der Fallvignette vorliegende Situation eingehender erfassen können als ihre männlichen Kommilitonen, lässt sich anhand der ermittelten Befunde kaum bestätigen. Vielmehr gelingt es im Kompetenzbereich *Informationen ermitteln* (Skala 1, Steps I-III) Probandinnen und Probanden der EG1 und EG2 gleichermaßen, die

4 Der Testgruppe der im Fachsemester 1 bis 3 studierenden Probanden gehörten nur etwa 20% der Probanden der EG1 und EG2 an; in der Testgruppe der höheren Fachsemester liegt ihr Anteil dagegen mit rund 47% deutlich höher.

Bedingungen der Situation, das Akteurshandeln sowie die unmittelbaren Wirkungen aus der Fallvignette zu entnehmen und zu interpretieren. Die Studentinnen dieser beiden Testgruppen steigern ihren K_1 -Wert um 36.10 Punkte; die Studenten verbessern sich um eine geringfügig niedrigere Punkteanzahl (31.78).

Hinsichtlich der Teilkompetenzen im zweiten Skalenbereich, *Deutendes Verstehen* (Step IV) und *Verstehendes Erklären* (Step V), scheint für die männlichen Probanden zunächst ein größerer Effekt nach Cohen sichtbar. In den EG1 und EG2 verzeichnen sie einen Kompetenzanstieg um 15.13 Punkte; die weiblichen Probanden erzielen dagegen einen etwas kleineren Zugewinn von 9.24 Punkten. Diese kleineren Differenzen der K_1 -Zuwächse begründen sich mit den geringfügigen Unterschieden beider Teilpopulationen in der ersten Testung; sie sind unter Einbezug der Standardfehler allenfalls als marginal zu bezeichnen. Die auf Skala 2 beobachteten Unterschiede auf der Cohen-Skala der Effektstärke von +0.07 für die weiblichen und +0.21 für die männlichen Probanden unterliegen zusätzlich noch dem Einfluss differierender Standardabweichungen in der weiblichen und männlichen Gruppe. Mit dem Fokus auf die Werte aller weiblichen und männlichen Testteilnehmer sind auf Skala 1 und Skala 2 fast umgekehrte Entwicklungen der Standardabweichungen aufgetreten, die im Rahmen der durchgeführten Feldstudie bislang lediglich mit deren zufälliger Zusammensetzung erklärt werden können. Gleichwohl beeinflusst diese die Aussagekraft der Effektstärke.

1.5 Ergebnisse der Leistungsschwachen in der EG1

Die dem Diagnosekompetenzmodell immanente Struktur mit einzelnen algorithmischen Steps erlaubt nicht nur eine theoriegeleitete, möglichst optimale Erfassung des zu diagnostizierenden Falls, sondern die verschiedenen Steps verstehen sich zugleich als plausible Abfolge von Arbeitsschritten. Dabei sind mit jedem Step konkrete, fallunabhängige Fragekomplexe verbunden, die auch von den Probanden in der abschließenden Erhebung mehr oder weniger stark berücksichtigt worden sind. Insbesondere die EG1 und EG2, in denen der Algorithmus systematisch eingeführt und anhand exemplarischer Beispiele erprobt wurde, weisen einen vergleichsweise größeren Diagnosekompetenzzuwachs auf. Vor diesem Hintergrund interessierte uns, inwieweit jene Probanden von der algorithmischen Strukturierung profitieren, die im Einstiegsfall mit ihrem intuitiven, heuristischen Vorgehen fallspezifische Dimensionen nur grob oder unzureichend erfasst haben, also zu den Leistungsschwächeren zählten.

Um diese Frage beantworten zu können, ist eine Schlussgruppe definiert worden, zu der diejenigen 40% der Probanden aus EG1 gehören, für welche im Einstiegstest die geringsten K_1 -Werte (getrennt nach den beiden untersuchten Subskalen) ermittelt worden sind. Diese Schlussgruppe wird im Mittelwert mit der gesamten EG1 verglichen. Die folgende Tabelle 4 zeigt die Befunde.

Tatsächlich bestätigen die gemessenen K_1 -Werte des Haupttests, dass sich die Schlussgruppe mit einem Zuwachs von 44.14 Punkten zwischen Einstiegs- und

Abschlusstestung stärker verbessert hat als die Gruppe EG1 insgesamt, welche einen großen, aber geringeren Zuwachs von 35.05 Punkten erreichte. Insbesondere gelingt es einzelnen Probanden der ursprünglichen Schlussgruppe auf Skala 1 in der zweiten Testung im arithmetischen Mittel sogar oberhalb des Mittelwertes der gesamten Experimentalgruppe abzuschließen. Dieser Befund legt nahe, dass sich einzelne Probanden der ursprünglichen Schlussgruppe in den mittleren und oberen Leistungsbereich der gesamten Experimentalgruppe verbessert hatten.

		max	Skala 1	Skala 2
K_i-Mittelwert der Schlussgruppe EG1	EIN	100	15.39 [1.62]	19.35 [1.28]
	ABS	100	59.53 [2.05]	39.74 [2.90]
	DIF		+ 44.14	+ 20.40
K_i-Mittelwert der EG1	EIN	100	26.88 [1.84]	30.89 [1.89]
	ABS	100	61.93 [1.09]	42.92 [1.45]
	DIF		+ 35.05	+ 12.03

Tab. 4: Vergleich der K_i-Mittelwerte der Schlussgruppe der EG1 mit dieser Testgruppe insgesamt; Standardfehler in Klammern

Auf der zweiten Subskala ist analog hierzu eine Annäherung der Schlussgruppe an den Gesamtmittelwert der EG1 und ein größerer gemittelter Zuwachs des K_i-Wertes der Schlussgruppe zu konstatieren. Innerhalb der ersten Testung weicht der Mittelwert der Schlussgruppe vom Mittelwert der gesamten EG1 noch negativ mit 11.54 Punkten ab. Nach Einführung und Anwendung des algorithmischen Diagnoseverfahrens im Seminkontext verringerte sich in der zweiten, nach etwa drei Monaten durchgeführten Testung diese Differenz auf 3.18 Punkte. Im Vergleich der beiden Subskalen zeigt sich, dass sowohl in der Schlussgruppe als auch in der EG1 insgesamt die Verbesserungen auf Skala 1 annähernd doppelt so groß ausfallen wie auf Skala 2. Insofern hat sich anscheinend der Effekt anteilig auf beide in den Subskalen abgebildeten Kompetenzbereiche übertragen. Die erhobenen Daten stützen zwar die Vermutung, dass leistungsschwächere Lehramtsstudierende durch den Einsatz des algorithmischen Diagnoseverfahrens bessere Ergebnisse erzielen, jedoch ist diese Hypothese durch größere Versuchsgruppen eingehender zu verifizieren.

1.6 Der Vergleich zweier Probanden der EG1

Die Entwicklung der Diagnosekompetenz von Studierenden, die sich an unserem Forschungsprojekt beteiligt haben, ist in den verschiedenen Lerngruppen, aber auch innerhalb der Lerngruppen selbst unterschiedlich verlaufen. Neben den oben bislang im Mittelwert dargestellten Resultate erscheint die Fokussierung von individuellen Kompetenzentwicklungen beachtenswert. Im Besonderen ist zu zeigen, dass auch vergleichbare Kompetenzwerte durchaus verschiedene Blickwinkel auf die ausgewählten Fallvignetten zulassen.

Für diesen Fallvergleich sind Datensätze zweier männlicher Probanden der EG1 mit vergleichbaren Kompetenzentwicklungen ausgewählt worden. Als Daten liegen ihre Kompetenzergebnisse aus dem Einstiegs- und Abschlusstest vor sowie verschriftlichte Selbsteinschätzungen ihrer persönlichen Kompetenzentwicklung im Kontext des Seminars und der dort absolvierten Übungen. Beide Studierende befanden sich im Untersuchungszeitraum im dritten Fachsemester und hatten im vorhergehenden Sommersemester das Modul EGL 2 („Rezeption und Reflexion von Schule“) bereits absolviert. Proband 1 – wir wollen ihn Lukas nennen – war zum Erhebungszeitraum 23 Jahre alt und studierte die Fächer Mathematik und Sport, während Proband 2 – Daniel, ebenfalls 23 Jahre alt – für Chemie und Sport eingeschrieben war. Die allgemeine Hochschulreife haben beide im Bundesland Hessen erworben; sie hatten das im Rahmen der Datenerhebung als EG1 einbezogene Seminar „Fallarbeit und Forschendes Lernen“ innerhalb des Moduls EGL 3 mit erster Priorität gewählt.

1.6.1 Fallarbeit und Kompetenzstand im Einstiegstest

Lukas erreicht in der ersten Testung auf Skala 1 einen Kompetenzwert K_i von 32.94, auf Skala 2 von 33.81 Punkten. Beide K_i -Werte liegen nahe zu den Mittelwerten der EG1 und etwa auf der Hauptdiagonale in der Veranschaulichung als vektorielle Skalen. Die pädagogisch-soziologische Kompetenz entspricht dem unteren Niveau 2. Dieser Proband konnte in seiner Analyse neun von 36 Items identifizieren, wobei er sich in seinen Ausführungen besonders auf das Konfliktlösungshandeln der Lehrerin konzentriert und dieses auch bewertet hat (vgl. Dirks 2012b, Cluster-Übersicht <Handy-Fall>, Cluster C). Einen zweiten Schwerpunkt seiner Analyse bildet die Handlungswahl von Schüler B, der sich nach dem Verlust des Handys unmittelbar an die Lehrerin gewandt hatte (ebd., Cluster H).

Für Daniel ist innerhalb der Ersttestung ein ähnlicher K_i -Wert auf beiden Skalen gemessen worden: Mit insgesamt acht berücksichtigten Items erreicht er auf Skala 1 32.83, auf Skala 2 vergleichbare 33.63 Punkte. Folglich entsprechen seine Kompetenzwerte auf beiden Skalen ebenfalls nahezu dem Mittelwert der EG1. Wie bei Lukas übertragen sich diese Werte in das Modell der vektoriellen Skalen mit dem Kompetenzniveau 2. Doch während Lukas die Handlungswahl von Schüler B aus der Fallvignette besonders thematisiert hat, fehlt diese Betrachtungsweise bei Daniel vollkommen. Die Konfliktlösestrategien der Lehrkraft in der Fallvignette behandelt er nur bewertend, indem er auf einen möglichen Fehlschluss hinweist (Cluster I); diese Interpretation der Vorgänge stellt er in den Zusammenhang mit der unzureichenden Informiertheit der Lehrkraft in der gegebenen Situation (Cluster E). Auch findet sich – im Gegensatz zu Lukas' Analyse – ein Hinweis auf Konfliktlösungsalternativen (Cluster J).

1.6.2 Der Kompetenzstand nach der Hälfte des Seminars (Eigen-evaluation)

Die Seminargruppe, an der Lukas und Daniel im Wintersemester 2010/11 teilnahmen, erhielt nach Ablauf der Hälfte der Seminarveranstaltungen den Auftrag, in schriftlicher Hausarbeit während der Weihnachtsferien ihre persönliche Kompetenzentwicklung einzuschätzen. Beide Studierende reichten ausführliche Exposés ein, die hier in Auszügen wiedergegeben werden:

Lukas verwendet ein vereinfachtes Verfahren zur Selbsteinschätzung seiner Kompetenzentwicklung (vgl. Hansmann 2012), indem er drei Kompetenzbereiche unterscheidet: Informationsermittlung, textbezogenes Interpretieren sowie Reflektieren und Bewerten. Hinsichtlich der Bewertung seiner Leistungen in den genannten Bereichen greift er auf eine fünfteilige Stufung zurück, die vom Seminarleiter vorgeschlagen worden war. In jedem Bereich ordnet er seine Leistungen auf der höchsten Stufe ein und begründet diese Einschätzung folgendermaßen:

An dieser Stelle möchte ich noch hinzufügen, dass diese sehr gute Benotung meinerseits nicht anmaßend zu verstehen ist. Ich habe mir in den Ferien einfach viel Zeit für die Bearbeitung genommen, weshalb ich auch jetzt noch fest davon überzeugt bin, alles aus dem Unterrichtsskript zu den jeweiligen Algorithmen herausgefiltert und thematisiert zu haben.

In einer Zusammenfassung seiner Kompetenzentwicklung vom Beginn des Seminars bis zu den Weihnachtsferien gelangt er zu dem Urteil:

Ich [habe] seit Semesterbeginn definitiv Fortschritte gemacht. Ich muss gestehen, dass ich ein wenig verwundert war, als wir in unserer ersten Seminarsitzung dazu aufgefordert wurden, eine Unterrichtsszene zu analysieren. „Wie macht man das? Welche Aspekte spielen dabei eine Rolle?“ Ich griff auf Vorgehen aus meiner Schulzeit zurück, was, denke ich, zwar in die richtige Richtung ging, aber bei weitem zu oberflächlich war.

In Sachen Unterrichtsanalyse mithilfe der Algorithmen fühle ich mich mittlerweile versierter, auch wenn ich, zugegebenermaßen, bei ihrer Verwendung noch meine Notizen zu Rate ziehen muss. Dieser Makel ist jedoch nur eine Frage von auswendig zu lernendem **Wissen**, welches ich mir noch anzueignen gedenke. Ihre Verwendung, das **Können** [Herv. i.O.], fällt mir zusehends leichter, da ich dererlei analytisches Vorgehen einem Schreiben ins Blaue hinein vorziehe.

Bei jeder Unterrichtsanalyse gilt es, vereinfacht ausgedrückt, dem durch die Algorithmen gespannten Roten Faden nachzulaufen und unterwegs alles Wichtige aus dem Transkript einzusammeln.

Natürlich muss ich dieses Vorgehen noch üben, um behaupten zu können, obige Kompetenzbereiche in Gänze erlernt zu haben.

Auch Daniel setzt das o.g. verkürzte Verfahren zur Selbsteinschätzung ein und sieht – ebenso wie Lukas – die eigene Kompetenzentwicklung in den Bereichen

„Informationsermittlung“ sowie „Reflektieren & Bewerten“ auf der höchsten Kompetenzstufe angesiedelt; dagegen hat er im Bereich „Textbezogenes Interpretieren“ nach seinem Eindruck lediglich die vorletzte Stufe erreicht. Sein Kommentar lautet zusammenfassend:

Seit unserer ersten Task zu Semesterbeginn kann man eindeutig sagen, dass ich mich bezüglich der Unterrichtsanalyse verbessert habe. Ohne das Wissen, das ich mir während des Seminars aneignen durfte, ging ich deshalb zu unwissenschaftlich an eine solche Analyse heran. Ohne einen „Fahrplan“, wie die einzelnen Algorithmen für mich einen darstellen, verlor ich mich daher in Kleinigkeiten, ohne das Ganze im Blick zu behalten.

Bezüglich der Unterrichtsanalyse mithilfe der Algorithmen fühle ich mich deshalb mittlerweile vertrauter. Von **Wissen** ist denke ich aber nur zu sprechen, wenn man den oben erwähnten „Fahrplan“ absolut verinnerlicht hat. Ich denke, dass ich nach der Klausurvorbereitung von Wissen sprechen kann.

Ihre Verwendung, das **Können** [Herv. i.O.], ist mir dagegen mittlerweile sehr vertraut und mit jeder weiteren Analyse kann ich dieses vorhandene Können noch verfeinern.

1.6.3 Der Kompetenzstand am Ende des Seminars (Abschlusstest)

Im Abschlusstest entspricht die Analyse von Lukas unter Einbezug beider Skalen dem oberen vierten Kompetenzniveau. Insgesamt wurden 32 Items als erfüllt betrachtet, was sich in einem K_T -Wert von 50.55 Punkten auf Skala 1 und einem Kompetenzwert von 47.65 Punkten auf Skala 2 ausdrückt. Von den 13 verschiedenen inhaltlichen Clustern (vgl. Dirks 2012b, Cluster-Übersicht Fallvignette „Beginn einer Erdkundestunde“) behandelt Lukas zwölf; unberücksichtigt bleibt – wie auch bei Daniel – die Verbindung der vielen aufgetretenen Disziplinstörungen der Schülerinnen und Schüler mit ihrem Alter bzw. der Schulform (Cluster E). Die Disziplinierungsstrategien der Lehrperson erscheinen dagegen als besonderer Schwerpunkt der Analyse, genauso wie das Klassenklima und das Verhältnis zwischen Lehrperson und Lerngruppe (Cluster I). Analytisch erfasst werden von beiden Probanden auch die inhaltliche und methodische Gestaltung des Unterrichtsausschnittes in der Fallvignette (Cluster K). Ebenfalls nachgegangen wird der Frage, inwieweit die räumlichen Gegebenheiten im Klassenzimmer, die didaktische Gestaltung des Lernangebots und der unmittelbare Lernerfolg für die Bewertung der Qualität des Unterrichts bedeutsam sind (Cluster L).

Die gemessenen Kompetenzwerte für Daniel stimmen auch im Abschlusstest nahezu vollständig mit denen von Lukas überein. Daniel erreicht auf Skala 1 einen K_T -Wert von 50.13 bzw. auf Skala 2 von 49.68 Punkten. Insgesamt sind 34 Items als erfüllt angesehen worden. Im Vergleich zu Lukas fallen ähnliche, aber auch unterschiedliche inhaltliche Schwerpunktsetzungen bei der Analyse der Fallvignette auf. Bei beiden Probanden spielen die Vorbereitung der Lehrkraft, vor allem in

Bezug auf den Einsatz der Tafel, ebenso wie geschlechtsspezifische Akteurspraktiken und -zuschreibungen eine untergeordnete Rolle. Im Gegensatz zu Daniel befasst sich Lukas allerdings eingehend mit der Kenntnis des Lehrers über das Wissen einzelner Schülerinnen und Schüler bzw. der Lerngruppe (Cluster D). Daniel widmet sich jedoch verstärkt der Frage nach der Unterrichtsqualität (Cluster L). Von acht vorgesehenen Items in diesem Cluster finden sich in der Analyse von Daniel sieben. Hier fehlt einzig der Hinweis auf lerneffektivere Ansätze zum Wissensaufbau, der sich aus einer methodischen Reflexion des Lehrerhandelns ergibt (Item 96).

1.6.4 Fazit

Für den Fallvergleich wurden zwei Kandidaten ausgewählt, die ausführliche Selbsteinschätzungen mit breit diversifizierten Handlungsstrukturvariablen vorgelegt haben. Aufgrund ihres theoretisch repräsentativen Charakters können beide Fälle einen detaillierten Einblick geben, wie das gesamte Forschungsvorhaben von den Studierenden aufgenommen wurde. Die hier vollzogene minimale Fallkontrastierung hatte nicht den Zweck, weitere Facetten des Kompetenzerwerbs zu verdeutlichen, die sicherlich interessant gewesen wären, aber leider aus Gründen der Forschungsökonomie nicht näher untersucht werden konnten.

Der Vergleich der zwei Messungen am Beginn und zum Ende des Semesters mit den zugehörigen Eigenevaluationen der beiden ausgewählten Probanden zur Semestermitte gestattet aber einen exemplarischen Blick auf das Verhältnis von Selbst- und Fremdeinschätzung: Beide Probanden weisen im Rahmen der Reflexion ihres Lernfortschrittes darauf hin, dass sie ihre fallanalytischen Fähigkeiten verbessert haben; die abschließende Kompetenzmessung bestätigt diesen Eindruck. Die Bemerkung von Daniel, die intuitive Herangehensweise bei der Fallanalyse sei „unwissenschaftlich“, lässt sich aus den Befunden nicht abschließend bewerten, doch zeugt seine Testantwort im Einstiegstest von fehlenden Antwortstrukturen bzw. Erfahrungen mit dem Genre des Unterrichtstranskripts. Gleichfalls fällt auf, dass er in seiner Analyse mehrfach zwischen Situations- und Akteursbeschreibung auf der einen und Brückenhypothesen auf der anderen Seite hin und her springt. Dies hat zur Folge, dass die vereinzelt vorgenommenen Bewertungen (besonders des Lehrerhandelns) durchweg nicht auf einer tieferen und breiteren Erfassung der Situationsbedingungen beruhen und somit als unsicher gelten müssen. In der abschließenden zweiten Testung erstellt Daniel nicht nur zuerst eine facettenreiche Beschreibung der Situation, der Akteure und ihres Handelns, bevor er diese zu verstehen und zu erklären versuchte, sondern er strukturiert seine Analyse auch nach den einzelnen algorithmischen Steps. Innerhalb seiner Eigenevaluation zur Mitte des Semesters hat Lukas darauf hingewiesen, dass für ihn das algorithmische Verfahren einen ‚Fahrplan‘ zur Analyse der Fallvignetten darstelle, der erst die Grundlage dafür bietet, die Aggregationslogik der Situation stetig im Blick zu behalten; dementsprechend scheint er dann auch bei der Analyse des Abschlussfalls die einzelnen mit den Steps verknüpften Leitfragen behandelt zu haben. Beiden

Probanden ist es gelungen, die dem algorithmischen Verfahren zugrundeliegenden Strukturelemente, die sie in ihrer Zwischenevaluation als noch nicht vollständig verfügbare Wissensbestände betrachtet haben, bei der späteren Zweittestung zum Abschluss der Seminarveranstaltung zu aktivieren und auf die Fallvignette anzuwenden.

1.7 Zusammenfassung

Überblicken wir die in diesem Kapitel vorgestellten Ergebnisse des Forschungsprojekts „Professionalisierung und Diagnosekompetenz – Kompetenzentwicklung und -förderung im Lehramtsstudium“, kann zusammenfassend Folgendes festgestellt werden: Die Kompetenzstände der Studierenden erreichen zu Beginn der Untersuchung in den verschiedenen Experimental- und Kontrollgruppen durchweg auf Skala 1 („Beschreiben von Situationsvariablen, Praktiken und Effekten“) einen K_1 -Mittelwert zwischen 21.84 [2.85] (EG2) und 32.91 [2.80] (EG3)⁵; ihre Diagnosekompetenz ist somit in der Ersttestung auf dem unteren Kompetenzniveau der Skala 1 angesiedelt. Demnach gelingt es den Probanden zu diesem Testzeitpunkt nur in begrenztem Maße, die in der Fallvignette explizit enthaltenen Informationen bei der Bearbeitung der Testaufgabe einzubeziehen. Auf Skala 2 (Prozesse des „Deutenden Verstehens und Verstehenden Erklärens“) wurde ein K_1 -Mittelwert gemessen, der zwischen 25.80 [2.74] (KG3) und 32.63 [2.52] (EG1) variierte und bei fünf der sechs Testgruppen leicht oberhalb der K_1 -Mittelwerte auf Skala 1 lag. Auf der zweiten Subskala wurde damit ebenfalls ein unteres Kompetenzniveau erreicht.

Am Ende des Untersuchungszeitraums konnte auf Skala 1 bei denjenigen Gruppen ein Kompetenzzuwachs festgestellt werden, die das algorithmische Diagnoseverfahren erlernt und sukzessive anhand von verschiedenen Beispielen geübt hatten. Die Probanden dieser Gruppen (EG 1 und EG2) waren in der Lage, die in der Fallvignette enthaltenen Informationen detailgetreu zu beschreiben.

Auch auf Skala 2 erreichten die Probanden der EG1 und EG2 eine signifikante Steigerung ihres Kompetenzniveaus, während die anderen Testgruppen auf dem unteren Kompetenzniveau verblieben. Hatten die Probanden in den Experimentalgruppen 1 und 2 im ersten Test ohne Kenntnis des algorithmischen Verfahrens nur einzelne fehlende Informationen plausibel ergänzt und zur Hypothesenbildung genutzt, wurden in der zweiten Erhebung naheliegende Inferenzen z.B. zur Schulform oder Klassenstufe deutlich häufiger und mit stärker ausdifferenziertem, argumentativem Rückbezug auf die im Transkript explizit enthaltenen Informationen formuliert. Zudem können sich offenbar Probanden mit zunächst schwächeren Analyseleistungen unter Anwendung des algorithmischen Verfahrens besonders verbessern; Voraussetzung hierfür sind jedoch Übungsmöglichkeiten, die bspw. in EG3 nicht angeboten wurden, so dass hier auch keine Verbesserungen konstatiert werden konnten. Eine

5 Standardfehler sind in eckigen Klammern angegeben.

erfolgsversprechende Förderung diagnostischer Kompetenzen erfordert somit eine gezielte Schulung, entweder im Grund- oder im Hauptstudium.

In welchem Umfang sich diese Untersuchungsergebnisse im Hinblick auf die Durchführung von Lehrveranstaltungen im Lehramtsstudium verallgemeinern lassen, muss aber aufgrund der Besonderheiten des Hessischen und Marburger Erziehungs- und Gesellschaftswissenschaftlichen Lehrerstudiums (Modulare Struktur, Curriculum, Größe der Lerngruppen, keine Unterscheidung in Grund- und Hauptstudium etc.) und auch aufgrund der geringen Stichprobengröße offen bleiben.

Literatur

- Bannert, Maria (2007). Metakognition beim Lernen mit Hypermedien. Münster: Waxmann.
- Baumbach, Hendrik (2012). Das Messmodell – Messverfahren zur Analyse fallbasierter Diagnosekompetenzen. In: Wilfried Hansmann; Una Dirks & Hendrik Baumbach (Hrsg.), Professionalisierung und Diagnosekompetenz – Kompetenzentwicklung und -förderung im Lehramtsstudium (Kap. II. 6, Online-Schriftenreihe der Philipps-Universität Marburg: Professionalisierung und Diagnosekompetenz). [URL: <http://archiv.ub.uni-marburg.de/es/2012/0022/>].
- Dirks, Una (2012a). Pädagogisch soziologische Diagnosekompetenz im Modell. In: Wilfried Hansmann, Una Dirks & Hendrik Baumbach (Hrsg.), Professionalisierung und Diagnosekompetenz. Kompetenzentwicklung und -förderung im Lehramtsstudium. (Kap. II. 5, Online-Schriftenreihe der Philipps-Universität Marburg: Professionalisierung und Diagnosekompetenz). [URL: <http://archiv.ub.uni-marburg.de/es/2012/0018/>].
- Dirks, Una (2012b) Prä- und Posttest-Kongruenzen – Gemeinsamkeiten und Differenzen in der fallspezifischen Anforderungsstruktur. In: Wilfried Hansmann; Una Dirks & Hendrik Baumbach (Hrsg.), Professionalisierung und Diagnosekompetenz – Kompetenzentwicklung und -förderung im Lehramtsstudium. (Kap. II.8, Online-Schriftenreihe der Philipps-Universität Marburg: Professionalisierung und Diagnosekompetenz). [URL: <http://archiv.ub.uni-marburg.de/es/2012/0023/>].
- Dirks, Una, Wilfried Hansmann & Hendrik Baumbach (2012). Item Pools der administrierten Fälle [31 Seiten]. In: Wilfried Hansmann, Una Dirks & Hendrik Baumbach (Hrsg.), Professionalisierung und Diagnosekompetenz – Kompetenzentwicklung und -förderung im Lehramtsstudium. (Kap. II.7, Online-Schriftenreihe der Philipps-Universität Marburg: Professionalisierung und Diagnosekompetenz). [URL: <http://archiv.ub.uni-marburg.de/es/2012/0021/>].

- Dirks, Una & Wilfried Hansmann (2013). Die Ergebnisse aus bildungstheoretischer Perspektive. In: Wilfried Hansmann; Una Dirks & Hendrik Baumbach (Hrsg.), Professionalisierung und Diagnosekompetenz – Kompetenzentwicklung und -förderung im Lehramtsstudium (Kap. III.2, Online-Schriftenreihe der Philipps-Universität Marburg: Professionalisierung und Diagnosekompetenz). [URL: http://archiv.ub.uni-marburg.de/opus/schriftenreihen_ebene2.php?sr_id=30&la=de].
- EGL-M: Studium der Erziehungs- und Gesellschaftswissenschaften für das Lehramt“ in der für den Testzeitraum gültigen Version vom WS 2006/07: <http://www.uni-marburg.de/fb21/schulpaed/studium/studieninformationenallgemein/eglm/infoeglm12.pdf>
- Funke, Joachim & Jörg Zumbach (2006). Problemlösen. In: Heinz Mandl & Helmut Felix Friedrich (Hrsg.). Handbuch Lernstrategien, S. 206-220. Göttingen: Hogrefe.
- Hagemann, Birte (2002). Leistungsmessung bei schriftlichen mathematischen Problemlösungen in Abhängigkeiten vom Bewertungsmodell, Diss. phil. Univ. Duisburg.
- Hansmann, Wilfried (2012). Fallbasierte Seminararbeit im algorithmisierten Kurzverfahren. In: Wilfried Hansmann, Una Dirks & Hendrik Baumbach (Hrsg.), Professionalisierung und Diagnosekompetenz – Kompetenzentwicklung und -förderung im Lehramtsstudium. (Kap. IV. 2, Online-Schriftenreihe der Philipps-Universität Marburg: Professionalisierung und Diagnosekompetenz). [URL: <http://archiv.ub.uni-marburg.de/es/2012/0020/>].
- Hansmann, Wilfried, Una Dirks & Hendrik Baumbach (2012). Datenerhebung und Probandengruppen. In: Wilfried Hansmann; Una Dirks & Hendrik Baumbach (Hrsg.), Professionalisierung und Diagnosekompetenz – Kompetenzentwicklung und -förderung im Lehramtsstudium (Kap. IV. 2, Online-Schriftenreihe der Philipps-Universität Marburg: Professionalisierung und Diagnosekompetenz). [URL: <http://archiv.ub.uni-marburg.de/es/2012/0017/>].

Zitation

Baumbach, Hendrik & Wilfried Hansmann (2012). Entwicklung pädagogisch-soziologischer Diagnosekompetenz von Lehramtsstudierenden [22 Seiten]. In: Wilfried Hansmann, Una Dirks & Hendrik Baumbach (Hrsg.), Professionalisierung und Diagnosekompetenz – Kompetenzentwicklung und -förderung im Lehramtsstudium. (Online-Schriftenreihe der Philipps-Universität Marburg: Professionalisierung und Diagnosekompetenz. Kap. III. 1). [URL: http://archiv.ub.uni-marburg.de/opus/schriftenreihen_ebene2.php?sr_id=30&la=de].