

Buholzer, Alois; Zulliger, Sandra

## Die Entwicklung von diagnostischen Fähigkeiten bei Studierenden der pädagogischen Hochschulen im Laufe ihrer Ausbildung

*Beiträge zur Lehrerinnen- und Lehrerbildung 31 (2013) 2, S. 186-197*



Empfohlene Zitierung/ Suggested Citation:

Buholzer, Alois; Zulliger, Sandra: Die Entwicklung von diagnostischen Fähigkeiten bei Studierenden der pädagogischen Hochschulen im Laufe ihrer Ausbildung - In: Beiträge zur Lehrerinnen- und Lehrerbildung 31 (2013) 2, S. 186-197 - URN: urn:nbn:de:0111-pedocs-138464

in Kooperation mit / in cooperation with:



<http://www.bzl-online.ch>

### Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen. Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

### Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

### Kontakt / Contact:

peDOCS  
Deutsches Institut für Internationale Pädagogische Forschung (DIPF)  
Informationszentrum (IZ) Bildung  
E-Mail: [pedocs@dipf.de](mailto:pedocs@dipf.de)  
Internet: [www.pedocs.de](http://www.pedocs.de)

## **Die Entwicklung von diagnostischen Fähigkeiten bei Studierenden der pädagogischen Hochschulen im Laufe ihrer Ausbildung**

Alois Buholzer und Sandra Zulliger

**Zusammenfassung** Die Diagnosefähigkeit von Lehrpersonen umfasst neben der Beurteilungsgüte vor allem die Fähigkeit, Diagnosen didaktisch adaptiv zu nutzen und lernfördernde Rückmeldungen daraus abzuleiten. Wie sich diagnostische Fähigkeiten bei angehenden Lehrpersonen und im Berufsalltag entwickeln, ist noch weitgehend unerforscht. Die vorliegende Längsschnittstudie mit PH-Studierenden untersucht die diagnostischen Fähigkeiten bezüglich formaler und inhaltlicher Aspekte im Fach Deutsch und diskutiert sie kritisch. Als Beurteilungsmaterial dienten videografierte Reportagen von Schülerinnen und Schülern sowie deren Vorbereitungsnotizen. Es konnte gezeigt werden, dass die Studierenden im Laufe der Ausbildung lernen, den Lernstand von Schülerinnen und Schülern differenzierter zu erfassen, und es ihnen auch immer besser gelingt, adäquate Formen der Weiterarbeit zu entwerfen. Bei anderen inhaltlichen und formalen Aspekten, z.B. der Fähigkeit zur Beurteilung des Wortschatzes oder des Textaufbaus, konnten während des Studiums hingegen keine signifikanten Veränderungen nachgewiesen werden. Die Diskussion der Ergebnisse im Kontext der Lehrerinnen- und Lehrerausbildung und früherer Forschungsarbeiten hebt die Wichtigkeit einer empirischen Überprüfung von Ausbildungszielen wie beispielsweise Diagnosekompetenz hervor und fordert weitere, vertiefte Analysen.

**Schlagwörter** diagnostische Fähigkeit – Kompetenzentwicklung – Lehrerinnen- und Lehrerbildung

### **The Development of Diagnostic Skills in Pre-Service Teachers During the Course of Their Study**

**Abstract** Apart from the general quality of a given assessment, diagnostic skills of teachers particularly include the ability to make use of diagnoses in an instructionally adaptive way and to derive feedback which promotes the pupils' learning. However, it is still largely unexplored how diagnostic skills in pre-service teachers and in everyday teaching practice actually develop. The longitudinal study to be presented in this article investigates and critically discusses the diagnostic skills of teacher students from the University of Teacher Education Lucerne with respect to formal and content-related aspects in the subject German. Videoed school student reports and their preparatory notes served as assessment materials. The results show that students learn to assess the pupils' learning progress in a more differentiated way, and that they increasingly succeed in designing adequate forms of further work in the course of their study. No significant changes could be found concerning other content-related and formal aspects, e.g. the ability to assess vocabulary or text structure. The discussion of the findings in the context of teacher education

and previous research emphasizes the importance of an empirical examination of educational objectives such as diagnostic competence, and calls for further in-depth analyses.

**Keywords** diagnostic ability – competence development – teacher education

## 1 Einleitung

Diagnostische Fähigkeiten gehören zum *Grundinventar* des erfolgreichen Unterrichtens. Lehrpersonen beobachten kontinuierlich das Lernverhalten ihrer Schülerinnen und Schüler, ziehen Rückschlüsse auf den erworbenen Wissensstand, erschliessen mit Prüfungen, ob die anvisierten Lernziele erreicht werden, erteilen Noten und geben Prognosen für die künftige Schullaufbahn ab. Ihr diagnostisches Handeln ist vielfältig und das Spektrum möglicher Diagnoseaufgaben erstreckt sich «von den formellen Beurteilungsaufgaben (Notengebung) des Lehrers, die im Idealfall auf methodisch kontrolliert erhobenen und sorgfältig reflektiert verarbeiteten Informationen beruhen, bis zu den im Rahmen des unterrichtlichen Handelns unter Zeit- und Handlungsdruck kurzfristig zu erbringenden Urteilsleistungen» (Schrader, 1989, S. 46). Für die Erfüllung dieser breit angelegten Aufgaben sind die Lehrpersonen überwiegend auf sich allein gestellt. Es liegen zwar standardisierte Messinstrumente wie «Klassenscockpit» (vgl. [www.klassenscockpit.ch](http://www.klassenscockpit.ch)) oder «Stellwerk» (vgl. [www.stellwerk-check.ch](http://www.stellwerk-check.ch)) vor, welche die Möglichkeit bieten, die Leistungen der eigenen Klasse mit einer repräsentativen Stichprobe zu vergleichen. Dennoch bestimmen die Lehrpersonen weitgehend über die Prüfungsinhalte und die Art und Weise ihrer Überprüfung. Zudem entwickeln sie das Prüfinstrumentarium meistens selbst und legen auch die Beurteilungskriterien fest, anhand derer sie nach der Korrektur eine Einstufung der Leistungen ihrer Schülerinnen und Schüler vornehmen. Die einzelne Lehrperson entscheidet somit über Erfolg oder Misserfolg der Prüfung.

Lehrpersonen haben einen grossen Handlungs- und Gestaltungsspielraum bei der Diagnose und Beurteilung von Leistungen der Schülerinnen und Schüler. Dies erstaunt, werden doch diagnostische Analysen nicht nur durchgeführt, um Hinweise für die individuelle Lernförderung zu gewinnen, sondern auch, um bestimmte Berechtigungen (z.B. hinsichtlich der künftigen Schullaufbahn) zu vergeben. Da solche weitreichenden Entscheidungen letztlich auf den subjektiv gefärbten Beurteilungsprozessen und -ergebnissen der Lehrpersonen basieren, sind Ausbildung und Reflexion der eigenen diagnostischen Fähigkeiten im Laufe des Studiums für angehende Lehrpersonen umso wichtiger. Im vorliegenden Beitrag werden nach einer theoretischen Einführung Resultate einer Studie zur Entwicklung der diagnostischen Fähigkeiten während der Ausbildung vorgestellt. Im Fokus der Studie stehen dabei die Fragen, wie angehende Lehrpersonen eine Schülerleistung beurteilen, was sie der Schülerin oder dem Schüler rückmelden und wie sie mit ihr oder ihm weiterarbeiten würden.

Die empirische Unterrichtsforschung geht heute davon aus, dass es sich bei der Diagnosekompetenz *nicht um eine einheitliche Grundkompetenz* handelt, welche sich über die verschiedenen Fächer und diagnostischen Aufgaben hinweg generalisieren lässt. Diese Auffassung wird durch aktuelle Forschungsstudien bestätigt, die nur geringe Interkorrelationen zwischen den einzelnen Komponenten der diagnostischen Kompetenz nachweisen (Brunner, Anders, Hachfeld & Krauss, 2011; Spinath, 2005). Spinath (2005) konnte z.B. in ihrer Studie keine Korrelationen zwischen den Akkuratheitskomponenten innerhalb eines Merkmals und über die Merkmale hinweg feststellen und geht damit von einer hohen Merkmalspezifität aus. Sie legt nahe, den Begriff der diagnostischen Kompetenz nicht mehr zu verwenden und ihn durch «diagnostische Fähigkeiten» zu ersetzen. In neueren Arbeiten wird diese Anregung aufgenommen. So wird in der COACTIV-Studie der Begriff der diagnostischen Fähigkeiten genannt, die dem fachdidaktischen und dem pädagogisch-psychologischen Wissen zugeordnet werden (Anders, Kunter, Brunner, Krauss & Baumert, 2010; Brunner et al., 2011). Diagnostische Fähigkeiten sind somit spezifisch und stark mit den Fachinhalten gekoppelt, was im Bereich der Ausbildung von Lehrpersonen für eine enge Anbindung an fachwissenschaftliche und fachdidaktische Inhalte spricht.

## **2 Diagnostizieren als Grundlage für individuelle Lernförderung und Lernberatung**

Die Urteilsgüte steht häufig im Zentrum der Diskussionen um die diagnostischen Fähigkeiten von Lehrpersonen. Diagnosen spielen im Hinblick auf das Unterrichtsgeschehen jedoch eine weit umfassendere Rolle. Sie sind eingebunden in den Planungsprozess von individuellen Förderangeboten oder wichtig für die Differenzierung des Unterrichts. Diagnosen werden so als Teil von didaktischen Überlegungen verstanden, um die Lernangebote adaptiv an die Schülerinnen und Schüler anzupassen (Horstkemper, 2006; Buholzer, 2006). Unterstrichen wird die enge Verbindung von diagnostischen und didaktischen Überlegungen durch das Ergebnis von Weinert und Helmke (1987), wonach Lehrpersonendiagnosen erst dann wirksam werden, wenn sie mit unterrichtlichen Massnahmen (starke Strukturierung des Unterrichtsangebots) zusammenhängen. Teil davon ist eine angemessene Rückmeldung der diagnostischen Erkenntnisse an die betreffende Schülerin oder den betreffenden Schüler. Insbesondere die Kommunikation von diagnostischen Urteilen kann für Lehrpersonen jedoch eine grosse Herausforderung darstellen, «weil immer die Gefahr besteht, dass die Lernmotivation eines Schülers oder einer Schülerin, die eine wesentliche Ressource ihres Bildungsprozesses darstellt, durch unpassende Äusserungen untergraben wird» (Herzog, 2013, S. 6). Gleichzeitig ist bekannt, dass systematische Rückmeldungen zum Lernstand und zur Lernentwicklung eines der wirksamsten Instrumente zur Förderung von Lernen darstellen. Dazu gehören sowohl die Analyse des Lernstands zusammen mit einer einzelnen Schülerin oder einem einzelnen Schüler als auch Überlegungen dazu, wie das Lernen verbessert werden könnte und welche Schritte anstehen, damit die Einzelne,

der Einzelne oder die Klasse als Ganze weiter vorankommen. Diagnosen bilden so die Grundlage für Lernberatungen, um über Fehler zu sprechen und die aufgezeigten Ressourcen weiter zu vertiefen. Ressourcen bilden einen wichtigen Teil von Diagnosen: nicht nur aus Motivationsgründen, sondern auch als Anknüpfungspunkt für ein nachfolgendes Lernangebot.

### 2.1 Unterschiedliche Qualitäten von diagnostischen Fähigkeiten

Diagnostische Fähigkeiten können bei Lehrpersonen *mehr oder weniger gut ausgebildet* sein; dies legen verschiedene Studien zur Genauigkeit von Urteilsleistungen nahe. Die Genauigkeit von Urteilen wird empirisch häufig durch einen Vergleich der Vorhersagewerte von Lehrpersonen mit den tatsächlich gemessenen Leistungen der Schülerinnen und Schüler erfasst. Aus diesen Studien können zwei Hauptergebnisse herausgelesen werden: Erstens gelingt den Lehrerinnen und Lehrern der soziale Vergleich recht gut. Dies zeigt sich an den treffgenauen Leistungsrangfolgen, die Lehrpersonen von ihren Klassen erstellen können. Zweitens geben die Lehrpersonen in Bezug auf die Beurteilung der absoluten Leistung einer Schülerin oder eines Schülers eher ungenaue Schätzungen ab (Schrader & Helmke, 1990; Spinath, 2004). Vergleicht man beispielsweise die erteilten Noten mit objektiven Testergebnissen, so stellte Ingenkamp bereits 1971 fest, dass für objektiv gleiche Rechenleistungen unterschiedliche Noten vergeben werden. Auch in neueren Studien erweisen sich Notensetzungen (auch zwischen verschiedenen Klassen) als nicht vergleichbar. In der MARKUS-Studie konnten mässige Korrelationen von  $r = .30$  zwischen den Klassen aufgezeigt werden (Jäger, 2009). In der Meta-Analyse von Hoge und Coladarci (1989) zeigt sich eine mittlere Korrelation von  $r = .66$  zwischen den Lehrpersoneneinschätzungen und den tatsächlichen Leistungen der Schülerinnen und Schüler. Auf der Basis von nach Hoge und Coladarci (1989) publizierten Studien haben Südkamp, Kaiser und Möller (2012) eine weitere Meta-Analyse durchgeführt. Sie fanden ebenfalls eine mittlere Korrelation von  $r = .63$  zwischen den Lehrpersonenurteilen und den Leistungen der Schülerinnen und Schüler in Standardtests.

### 2.2 Diagnostische Fertigkeiten und Fähigkeiten erwerben

Diagnostische Kompetenzen bestehen aus Fertigkeiten und Fähigkeiten, die in der Ausbildung vermittelt und später im Schulalltag verfeinert und ausgebaut werden. Wie andere professionelle Kompetenzen auch werden sie durch jahrelange praktische Erfahrung und deren theoriegeleitete Reflexion ausgebildet. Die Wissensgrundlage für den Aufbau von Kompetenzen wird in der Ausbildung gelegt «und in der Form von Reflexions- und Gestaltungswissen in die Praxis mitgenommen, sodass auf der Basis verallgemeinerungsfähigen wissenschaftlichen Wissens über individuelles Fallverstehen eine (dann zunehmend) professionelle Praxisbewältigung angegangen werden kann» (Lersch, 2006, S. 177). Es kann daher davon ausgegangen werden, dass der Grossteil der diagnostischen Kompetenzen im Rahmen der Berufstätigkeit und nicht in der Grundausbildung erworben wird (Hascher, 2008). Wie sich die einzelnen Kompetenzbereiche im Laufe der Berufsbiografie ausbilden, ist jedoch unklar. In dieser Hin-

sicht besteht sowohl theoretischer als auch empirischer Forschungsbedarf (Baumert & Kunter, 2006). Kaum untersucht ist auch die Entwicklung der Kompetenzen im Laufe der Ausbildung zur Lehrperson (Baer, Dörr, Guldemann, Kocher, Larcher, Müller & Wyss, 2008).

Diagnostische Kompetenz erschöpft sich nicht darin, einzelne Diagnoseinstrumente zu kennen und diese anzuwenden. Die diagnostizierende Lehrperson benötigt Metawissen, um angepasste Methoden und Instrumente auswählen, die Ergebnisse einordnen und schliesslich richtig interpretieren zu können. Das Wissen über Entwicklungsverläufe und Störungsbilder sowie deren Ursachen, Kenntnisse über Möglichkeiten der Intervention und Prävention wie auch das Wissen über die personellen und organisatorischen Zuständigkeiten sind auf das Engste miteinander verknüpft (Hascher, 2008). Hascher (2003) schlägt vor, dass Studierende nebst eigenen Lernerfahrungen im Unterricht, in Werkstätten und in Projekten insbesondere in der berufspraktischen Tätigkeit Diagnosen erstellen und diese zusammen mit der Praktikumsleiterin oder dem Praktikumsleiter evaluieren sollten. Wichtig scheint dabei die Reflexion der subjektiven Wahrnehmungs- und Beurteilungsprozesse zu sein: Woraus erschliesst sich, ob eine Schülerin den Auftrag verstanden statt? Woran erkennt man, dass ein Schüler bei einer Stillarbeit nicht mehr weiterkommt?

Für die Ausbildung werden von Oser (1997) in Bezug auf die pädagogische Diagnostik zwei Standardgruppen beschrieben. Unter dem ersten Standard «Schülerunterstützende Beobachtung und Diagnose» werden Ausbildungsziele genannt, welche den Umgang mit kritischen Entwicklungen und auftretenden Problemen betreffen. Der Umgang bezieht sich dabei sowohl auf einzelne Schülerinnen und Schüler wie auch auf das Geschehen im Klassenzimmer. Die zweite Standardgruppe umfasst die Ausbildungsziele zum Bereich «Leistungsmessung». Die Curricula von pädagogischen Hochschulen in der Schweiz orientieren sich unter anderem an diesen Standards von Oser (1997) und konkretisieren ihre Ziele in Kompetenzbeschreibungen. Diese Beschreibungen, die festhalten, was die PH-Studierenden zu unterschiedlichen Zeitpunkten und am Ende ihrer Ausbildung können und wissen sollen, bilden für die Ausgestaltung der verschiedenen Lehrangebote (Module) wie auch für die Gesamtausbildung einen wichtigen Orientierungsrahmen. Die Zielvorstellungen sind dahin gehend formuliert, dass deren Umsetzung nicht nur deklariert, sondern auch empirisch überprüft werden kann (Alleman-Ghionda & Terhart, 2006).

An der PH Luzern beispielsweise tragen Ausbildungsmodule der Fachdidaktik zusammen mit den Modulen der Bildungs- und Sozialwissenschaften dazu bei, bei den angehenden Lehrpersonen diagnostische Fähigkeiten aufzubauen. Im Referenzrahmen der Ausbildung wird die Diagnose- und Beurteilungskompetenz in Anlehnung an Ingenkamp und Lissmann (2008) folgendermassen umschrieben: «Um Lernprozesse möglichst optimal anzuregen und zu begleiten und die Prozesse und Leistungen der Schülerinnen und Schüler zu beurteilen, müssen Lehrpersonen das Verständnis ihrer

Schülerinnen und Schüler erkunden, ihre Lern- und Denkwege verstehen und Lernergebnisse analysieren» (PH Luzern, 2012, S. 12). Die diagnostische Kompetenz wird als Voraussetzung für die Adaptivität des Unterrichts angesehen (Beck, Baer, Guldemann, Bischoff, Brühwiler, Müller, Niedermann, Rogalla & Vogt, 2008), da die Fähigkeit, den Lernstand und die Leistungsfähigkeit der Schülerinnen und Schüler unter Berücksichtigung unterschiedlicher Bedingungen einschätzen zu können, eine Voraussetzung für die Auswahl von Aufgaben und einer angepassten Form der Unterstützung darstellt. Ob es aber in der Lehrerinnen- und Lehrerbildung gelingt, die oben diskutierten Ziele und Standards zu erreichen, wurde bisher wenig überprüft. Die nachfolgend vorgestellte Längsschnittuntersuchung (2003 bis 2007) geht der Frage nach, wie sich die diagnostischen Fähigkeiten von angehenden Lehrpersonen im Fach Deutsch während ihres Studiums verändern.

### **3 Studie zur Entwicklung von diagnostischen Fähigkeiten von PH-Studierenden**

Wie vorangehend ausgeführt, gibt es nur wenige gesicherte empirische Erkenntnisse dazu, wie sich die diagnostischen Fähigkeiten von Lehrpersonen entwickeln. Zudem handelt es sich bei den diagnostischen Fähigkeiten nicht um ein einheitliches Konstrukt. Eine umfassende Erhebung der diagnostischen Fähigkeiten ist deshalb nicht möglich. In der vorzustellenden Studie werden sie als Fähigkeit aufgefasst, eine mündliche Deutschaufgabe unter Beachtung der Ziele der Aufgabenstellung angemessen einzuschätzen, auf dieser Grundlage eine Rückmeldung zu geben sowie einen Vorschlag zur Weiterarbeit zu unterbreiten. Die Ergebnisse beziehen sich somit nur auf das Fach Deutsch und auf die Arbeitsergebnisse einzelner Schülerinnen oder Schüler.

#### **3.1 Design der Längsschnittstudie**

Im Rahmen der Qualitätsüberprüfung der Lehre an der damals neu gegründeten PHZ Luzern<sup>1</sup> wurde von 2003 bis 2007 eine Längsschnittstudie durchgeführt. Die diagnostischen Fähigkeiten der Studierenden im Fach Deutsch wurden am Anfang (2003), in der Mitte (2005) und am Ende ihres Studiums (2006/2007) mit der immer gleichen experimentellen Methode erfasst. Die Studie begann gleichzeitig mit der Aufnahme des Lehrbetriebs an der PHZ Luzern im Jahr 2003. Für die Bestimmung der Fortschritte während der Ausbildung wurde mit einem echten Längsschnitt gerechnet. Für den Vergleich von Anfang und Mitte des Studiums konnte mit 60 Studierenden der Sekundarstufe I gerechnet werden und für den Vergleich von Anfang und Ende des Studiums mit 28 Studierenden. Im Folgenden werden die Resultate für die Studierenden der Sekundarstufe I detailliert aufgezeigt. Ebenfalls untersucht, aber in diesem Beitrag nicht weiter berücksichtigt wurden auch die Studierenden aus den Studiengängen Kindergarten/Unterstufe sowie Primarstufe.

---

<sup>1</sup> Am 1. August 2013 wurde aus der PHZ Luzern die Pädagogische Hochschule Luzern (PHLU).

Zur Erfassung ihrer diagnostischen Fähigkeiten wurde den Studierenden eine Videoaufnahme einer Schülerin oder eines Schülers gezeigt, die oder der eine selbst verfasste zweiminütige Radioreportage zu einem Strandbad vorstellt. Die den Schülerinnen und Schülern vorgelegte Aufgabe, die den Orientierungsarbeiten zum Fach Deutsch entnommen wurde, überprüft zum mündlichen Sprachgebrauch das Lernziel «Über Ereignisse und Beobachtungen in der Standardsprache berichten». Den Studierenden wurden am Anfang ihres Studiums, d.h. zum ersten Messzeitpunkt, zwei Radioreportagen – eine von Schülerin/Schüler A und eine von Schülerin/Schüler B – gezeigt. Zu Arbeitsergebnis A (und zu einem weiteren Arbeitsergebnis, das jedoch nicht ausgewertet wurde) wurde bei den Studierenden in der Mitte des Studiums, d.h. zum zweiten Messzeitpunkt, eine (erneute) Beurteilung eingeholt. Arbeitsergebnis B wiederum wurde den Studierenden (zusammen mit einem weiteren, nicht ausgewerteten Arbeitsergebnis) am Ende des Studiums nochmals vorgelegt. Durch die Variation der Arbeitsergebnisse sollte der Wiedererkennungseffekt über die drei Messzeitpunkte hinweg verringert werden; trotzdem dürfte die Vergleichbarkeit durch das zweimalige Vorlegen des gleichen Arbeitsergebnisses gewährleistet sein.

Zur gezeigten Radioreportage wurde den Studierenden ein Fragebogen<sup>2</sup> mit vier offenen Fragen ausgeteilt. Als Zusatzinformationen enthielt der Fragebogen den Arbeitsauftrag an die Schülerinnen und Schüler mitsamt den Lernzielen, die Beurteilungskriterien sowie die Notizen der betreffenden Schülerin oder des betreffenden Schülers. Die Studierenden wurden danach gefragt, was sie auf der Basis der Notizen der Schülerin bzw. des Schülers und der beobachteten Videosequenz in Bezug auf den Lernstand feststellen, was sie rückmelden und wie sie mit ihr oder ihm weiterarbeiten würden.

Als Vergleichsmaßstab zu den Angaben der Studierenden wurde ein Expertenurteil erstellt. Die Expertenrunde bestand aus je einer Fachperson der Zielstufe und der Fachdidaktik Deutsch sowie zwei Fachpersonen der Bildungs- und Sozialwissenschaften. Die Expertinnen und Experten beurteilten das Arbeitsergebnis unter den gleichen Bedingungen wie die Studierenden. Die einzelnen Experteneinschätzungen wurden zu einem Gesamturteil zusammengefasst, welches als Grundlage für die Erstellung eines Kategoriensystems mit inhaltlichen und formalen Kategorien diente. Die inhaltlichen Kategorien nehmen Bezug auf die Beurteilung des verwendeten Wortschatzes, des Aufbaus des Arbeitsergebnisses, der Stimmführung und der Aussprache sowie auf den Inhalt der Weiterarbeit. Mit den formalen Kategorien wird erfasst, ob der Lernstand differenziert dargestellt wurde, die Beurteilungen sich auf Tatsachenbeschreibungen beziehen und die Rückmeldungen sowie die vorgeschlagene Weiterarbeit einen Bezug zur Aufgabenstellung aufweisen. Bei den inhaltlichen Kategorien variiert die Anzahl der erfassten Kriterien je nach beurteiltem Arbeitsergebnis. Mithilfe des Kategoriensystems wurde eine quantitative Inhaltsanalyse durchgeführt. Die offenen Antworten der Studierenden wurden dahin gehend ausgewertet, ob der jeweilige im Kategoriensystem festgehaltene Aspekt ganz, teilweise oder gar nicht vorhanden war. Dafür wurde eine dreistufige

---

<sup>2</sup> Der Fragebogen kann beim Autor und bei der Autorin dieses Beitrags bezogen werden.



## Die Entwicklung von diagnostischen Fähigkeiten bei PH-Studierenden

Skala verwendet (0 = nicht vorhanden; 1 = teilweise vorhanden; 2 = vorhanden). Zur Bestimmung der Fortschritte wurden mit dem Wilcoxon-Test die jeweiligen Werte der einzelnen Messzeitpunkte miteinander verglichen. Das Signifikanzniveau wurde nach Bonferroni-Holm korrigiert und das Ausgangsniveau für ein signifikantes Ergebnis auf 0.05 festgelegt.

### 3.2 Fortschritte bei den einzelnen diagnostischen Fähigkeiten

Durch die Auswertung der Ergebnisse kann insbesondere aufgezeigt werden, welchen Aspekten die Studierenden bei der Beurteilung des Arbeitsergebnisses Beachtung geschenkt haben und welchen nicht bzw. auf welche Bereiche sie während ihres Studiums sensibilisiert wurden. In Tabelle 1 werden die Werte der Studierenden der Sekundarstufe I und die signifikanten Fortschritte in den einzelnen Kategorien über die drei Messzeitpunkte hinweg aufgeführt.

Betrachtet man das Gesamtergebnis aller Kategorien, so zeigen sich bei den Studierenden der Sekundarstufe I keine signifikanten Unterschiede zwischen dem ersten und

Tabelle 1: Studierende der Sekundarstufe I: Veränderungen über die drei Messzeitpunkte hinweg

		Mittelwerte und Standardabweichungen								t <sub>1</sub> - t <sub>3</sub>
		t <sub>1</sub>				t <sub>2</sub>		t <sub>3</sub>		
		A		B		A	B	M	SD	
		M	SD	M	SD	M	SD	M	SD	
Formale Kategorien	Differenzierung Lernstand	1.2	0.5	0.9	0.5	1.5	0.4	1.3	0.3	p = 0.004
	Interpretation Lernstand	0.8	0.3	0.8	0.4	1.1	0.4	1.1	0.4	n.s.
	Fokus Rückmeldung	1.6	0.5	1.2	0.7	1.3	0.6	1.6	0.6	n.s.
	Fokus Weiterarbeit	1.2	0.5	1.2	0.4	1.0	0.4	1.4	0.4	n.s.
<b>Formale Kategorien (gesamt)</b>		<b>1.2</b>	<b>0.3</b>	<b>1.0</b>	<b>0.3</b>	<b>1.2</b>	<b>0.3</b>	<b>1.4</b>	<b>0.3</b>	<b>p = 0.001</b>
Inhaltliche Kategorien	Beurteilung Wortschatz	0.6	0.4	0.7	0.3	0.6	0.4	0.6	0.3	n.s.
	Beurteilung Aufbau	0.9	0.8	0.7	0.5	1.0	0.7	0.8	0.6	n.s.
	Beurteilung Stimmführung	0.8	0.4	0.6	0.4	0.8	0.4	0.9	0.3	n.s.
	Beurteilung Aussprache	0.4	0.4	0.1	0.2	0.6	0.4	0.4	0.4	p = 0.003
	Angemessenheit Weiterarbeit	0.4	0.5	0.2	0.3	0.4	0.3	0.8	0.5	p = 0.000
<b>Inhaltliche Kategorien (gesamt)</b>		<b>0.6</b>	<b>0.3</b>	<b>0.5</b>	<b>0.2</b>	<b>0.7</b>	<b>0.3</b>	<b>0.7</b>	<b>0.2</b>	<b>p = 0.001</b>
<b>Gesamt</b>		<b>0.9</b>	<b>0.2</b>	<b>0.7</b>	<b>0.2</b>	<b>0.9</b>	<b>0.2</b>	<b>1.0</b>	<b>0.2</b>	<b>p = 0.000</b>

Anmerkungen: M = Mittelwert; SD = Standardabweichung.

Beurteilung von Schülerin/Schüler A: n = 60; Beurteilung von Schülerin/Schüler B: n = 28.

Die Mittelwerte können zwischen 0 und 2 variieren (0 = nicht vorhanden; 1 = teilweise vorhanden; 2 = vorhanden).

Gruppenvergleiche: Wilcoxon-Test für abhängige Stichproben, signifikante Unterschiede (p < 0.05) nach der Bonferroni-Holm-Korrektur.

dem zweiten, jedoch zwischen dem ersten und dem dritten Messzeitpunkt. Die Gesamtmittelwerte der drei Messzeitpunkte variieren lediglich zwischen 0.7 und 1.0 und streuen nur sehr geringfügig um den jeweiligen Mittelwert. Die Mittelwerte der inhaltlichen Kategorien liegen im Vergleich zu den formalen Kategorien wesentlich tiefer und somit weiter weg vom Expertenurteil. Da auf eine Kontrollgruppe verzichtet wurde, sind die unterschiedlichen Werte (und Entwicklungen) zwischen den Kategorien von besonderem Interesse für die Auswertung.

Bei den formalen Aspekten der Beurteilung zeigen sich deutliche Fortschritte der Studierenden bei der differenzierten Darstellung des Lernstands («Differenzierung Lernstand»). Eine differenzierte Darstellung des Lernstands liegt dann vor, wenn mindestens fünf inhaltlich verschiedene Aspekte präzise beschrieben werden (was und wie etwas ist) und der Situationsbezug (z.B. Videoaufnahme) berücksichtigt wurde. Sowohl zwischen dem Anfang und der Mitte als auch zwischen dem Anfang und dem Ende des Studiums machen die Studierenden hier hochsignifikante Fortschritte in der Beurteilung. Die Studierenden haben somit gelernt, genauer hinzuschauen, differenziertere Einschätzungen vorzunehmen und den Lernstand präziser zu beurteilen. Inwieweit die Studierenden ihre Beurteilungen belegt und interpretiert haben, wurde unter der formalen Kategorie «Interpretation Lernstand» zusammengefasst. Die Studierenden der Sekundarstufe I weisen bereits in der Mitte des Studiums einen Fortschritt auf, nicht jedoch am Ende des Studiums. Mit der Kategorie «Fokus Rückmeldung» wurde erfasst, ob sich die Rückmeldung, die die Studierenden der betreffenden Schülerin oder dem betreffenden Schüler geben würden, auf die Aufgabenstellung bezieht. Bei dieser Kategorie ist zwischen dem Anfang und der Mitte des Studiums ein signifikanter Rückgang von 1.6 auf 1.3 zu beobachten, während sich zwischen dem Anfang und dem Ende des Studiums keine Unterschiede zeigen. Diese Abnahme in der Mitte des Studiums ist anhand der vorliegenden Daten schwer zu erklären. Es scheint, dass sich die Studierenden in der Mitte ihrer Ausbildung freiere Rückmeldungen zutrauen und erst auf das Ende des Studiums hin lernen, ihre Rückmeldungen auf die Aufgabenstellung zu fokussieren.

Bei den einzelnen inhaltlichen Kategorien zeigen sich keine signifikanten Fortschritte (mit Ausnahme der Beurteilung der Aussprache und der Angemessenheit der Weiterarbeit). Obwohl die Ausgangswerte relativ tief liegen und somit eine Annäherung an das Expertenurteil möglich wäre, wurden keine Fortschritte erzielt. Ein möglicher Grund dafür könnte darin bestehen, dass sich die Studierenden im Rahmen der Ausbildung nicht mit diesen Lerninhalten befasst haben oder dass das Gelernte nicht auf die vorgelegte Aufgabenstellung transferiert werden konnte. Ein weiterer Grund für das Ausbleiben von Fortschritten im inhaltlichen Bereich könnte in der Ausgestaltung des Forschungsdesigns liegen. Beim Expertenurteil haben Personen aus verschiedenen Fachbereichen mitgewirkt, wodurch insbesondere bei den inhaltlichen Kategorien eine Annäherung an das kumulierte Expertenurteil für die Studierenden schwieriger wird.

In Bezug auf die Vorschläge zur Weiterarbeit kann festgehalten werden, dass unter formalen Gesichtspunkten bei den Rückmeldungen zur Weiterarbeit («Fokus Weiterarbeit») während der Ausbildung keine Fortschritte erzielt wurden. Hier wurde beurteilt, ob die Vorschläge zur Weiterarbeit innerhalb der Aufgabenstellung liegen. Unter inhaltlichen Gesichtspunkten hingegen sind deutliche Fortschritte zwischen dem ersten und dem letzten Messzeitpunkt zu verzeichnen. Das heisst, es gelingt den Studierenden immer besser, aus den diagnostischen Urteilen eine adäquate Weiterarbeit abzuleiten, d.h. in diesem Fall, eine nochmalige Diagnose oder eine Wiederholung der Übung auf der Grundlage eines Analogbeispiels, eines Trainings einer Teilfertigkeit und/oder der Veränderung der Lernsituation vorzuschlagen.

#### **4 Schlussfolgerungen**

Aus den Studienergebnissen geht hervor, dass die Studierenden deutliche Fortschritte bei der differenzierten Darstellung des Lernstands insgesamt und insbesondere in Bezug auf die Einschätzung der Aussprache erzielt haben. Die Studierenden sind darüber hinaus zunehmend in der Lage, aus der Beurteilung Schlüsse für die weitere Arbeit zu ziehen. Dieser Punkt ist nicht unwesentlich, weiss man doch aus der empirischen Unterrichtsforschung, dass die Fähigkeit von Lehrpersonen, ein Schülermerkmal treffend zu beurteilen, bei den Schülerinnen und Schülern allein noch nicht zu einem Lernfortschritt führt, sondern erst in Verbindung mit einer daran anschliessenden Strukturierung des Angebots (Weinert & Helmke, 1987). Es wird aber auch ersichtlich, dass sich – in Übereinstimmung mit anderen Ergebnissen von Studien zur Diagnosekompetenz (z.B. Spinath, 2005) – nicht alle Facetten der Diagnosekompetenz in gleichem Masse weiterentwickeln. So zeigen sich bei den inhaltlichen Aspekten Wortschatz, Aufbau des Arbeitsergebnisses und Stimmführung während des Studiums keine signifikanten Fortschritte in der Beurteilung. Offenbar setzen sich Beurteilungsprozesse aus unterschiedlichen Subprozessen zusammen, die im Laufe der Ausbildung unterschiedlich stark aufgegriffen und entwickelt werden. Die unterschiedlichen Annäherungen könnten ausserdem auf die Tatsache zurückgeführt werden, dass es sich beim Vergleichsmassstab um eine normative Setzung, gebildet durch die Einschätzungen der Expertinnen und Experten, handelt.

Mit dieser – im Rahmen der Qualitätsüberprüfung durchgeführten – Untersuchung konnten zwar erste Hinweise zum Aufbau von diagnostischen Fähigkeiten gewonnen werden, für Ableitungen von Implikationen für die Aus- und Weiterbildung ist es jedoch momentan noch zu früh – dafür sind weitere Studien erforderlich. Wir wissen erst wenig darüber, wie die diagnostischen Urteile zustande gekommen sind und welche Kognitionen im Beurteilungsprozess leitend waren. Wie Studien belegen, hängen diagnostische Urteile auch mit Kontextfaktoren zusammen; beispielsweise fällt es Lehrpersonen in Klassen mit ausgeprägter Heterogenität leichter, akkurate Diagnosen zu stellen (Karing, 2009). Auch der Zweck der Beurteilung unterstützt oder mildert

Urteilsverzerrungen (Krolak-Schwerdt, Böhmer & Gräsel, 2009). Diesen Kontextbedingungen bzw. der ökologischen Validität müsste bei einer revidierten Studiendurchführung mehr Beachtung geschenkt werden. Dazu gehört auch eine differenziertere Erfassung der Voraussetzungen bei den Studierenden (z.B. Erfahrungshintergrund und Vorwissen). Zudem müsste mit einem Kontrollgruppendesign überprüft werden, wie sich die Kompetenzentwicklung der PH-Studierenden im Vergleich zu Studierenden ohne entsprechende Fachinputs entwickelt. Nur so können kausale Rückschlüsse auf Lerninhalte und Studienaufbau gezogen werden.

## Literatur

- Allemann-Ghionda, C. & Terhart, E.** (2006). Kompetenzen und Kompetenzentwicklung von Lehrerinnen und Lehrern: Ausbildung und Beruf. *Zeitschrift für Pädagogik*, 51. Beiheft, 7–11.
- Anders, Y., Kunter, M., Brunner, M., Krauss, S. & Baumert, J.** (2010). Diagnostische Fähigkeiten von Mathematiklehrkräften und ihre Auswirkungen auf die Leistungen ihrer Schülerinnen und Schüler. *Psychologie in Erziehung und Unterricht*, 57 (3), 175–193.
- Baer, M., Dörr, G., Guldemann, T., Kocher, M., Larcher, S., Müller, P. & Wyss, C.** (2008). Wirkt Lehrerbildung? – Kompetenzaufbau und Standarderreichung in der berufswissenschaftlichen Ausbildung an drei Pädagogischen Hochschulen in der Schweiz und in Deutschland. *Empirische Pädagogik*, 22 (3), 259–273.
- Baumert, J. & Kunter, M.** (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9 (4), 469–520.
- Beck, E., Baer, M., Guldemann, T., Bischoff, S., Brühwiler, C., Müller, P., Niedermann, R., Rogalla, M. & Vogt, F.** (2008). *Adaptive Lehrkompetenz. Analyse von Struktur, Veränderbarkeit und Wirkung handlungssteuernden Lehrwissens*. Münster: Waxmann.
- Brunner, M., Anders, Y., Hachfeld, A. & Krauss, S.** (2011). Diagnostische Fähigkeiten von Mathematiklehrkräften. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 215–234). Münster: Waxmann.
- Buholzer, A.** (2006). *Förderdiagnostisches Sehen, Denken und Handeln*. Donauwörth: Auer.
- Hascher, T.** (2003). Diagnose als Voraussetzung für gelingende Lernprozesse. *Journal für Lehrerinnen- und Lehrerbildung*, 3 (2), 25–30.
- Hascher, T.** (2008). Diagnostische Kompetenzen im Lehrberuf. In C. Kraler & M. Schratz (Hrsg.), *Wissen erwerben, Kompetenzen entwickeln. Modelle zur kompetenzorientierten Lehrerbildung* (S. 71–86). Münster: Waxmann.
- Herzog, W.** (2013). Beurteilungskultur. *Profi-L*, Heft 1, 4–6.
- Hoge, R.D. & Coladarsi, T.** (1989). Teacher-Based Judgments of Academic Achievement: A Review of Literature. *Review of Educational Research*, 59 (3), 297–313.
- Horstkemper, M.** (2006). Fördern heisst diagnostizieren. Pädagogische Diagnostik als wichtige Voraussetzung für individuellen Lernerfolg. *Friedrich Jahresheft XXIV: Diagnostizieren und Fördern*, 4–7.
- Ingenkamp, K.** (Hrsg.). (1971). *Die Fragwürdigkeit der Zensurenggebung. Texte und Untersuchungsberichte*. Weinheim: Beltz.
- Ingenkamp, K. & Lissmann, U.** (2008). *Lehrbuch der Pädagogischen Diagnostik*. Weinheim: Beltz.
- Jäger, R.S.** (2009). Diagnostische Kompetenz und Urteilsbildung als Element von Lehrprofessionalität. In O. Zlatkin-Troitschanskaia (Hrsg.), *Lehrprofessionalität: Bedingungen, Genese, Wirkungen und ihre Messung* (S. 105–116). Weinheim: Beltz.
- Karing, C.** (2009). Diagnostische Kompetenz von Grundschul- und Gymnasiallehrkräften im Leistungsbereich und im Bereich Interessen. *Zeitschrift für Pädagogische Psychologie*, 23 (3–4), 197–209.

## Die Entwicklung von diagnostischen Fähigkeiten bei PH-Studierenden

- Krolak-Schwerdt, S., Böhmer, M. & Gräsel, C.** (2009). Verarbeitung von schülerbezogener Information als zielgeleiteter Prozess. *Zeitschrift für Pädagogische Psychologie*, 23 (3–4), 175–186.
- Lersch, R.** (2006). Lehrerbildung im Urteil der Auszubildenden. Eine empirische Studie zu beiden Phasen der Lehrerausbildung. *Zeitschrift für Pädagogik*, 51. Beiheft, 164–181.
- Oser, F.** (1997). Standards in der Lehrerbildung. Teil I: Berufliche Kompetenzen, die hohen Qualitätsmerkmalen entsprechen. *Beiträge zur Lehrerbildung*, 15 (1), 26–37.
- PH Luzern.** (2012). *Referenzrahmen der PH Luzern*. Luzern: PH Luzern.
- Schrader, F.-W.** (1989). *Diagnostische Kompetenz von Lehrern und ihre Bedeutsamkeit für die Gestaltung des Unterrichts*. Frankfurt am Main: Lang.
- Schrader, F.-W. & Helmke, A.** (1990). Lassen sich Lehrer bei der Leistungsbeurteilung von sachfremden Gesichtspunkten leiten? Eine Untersuchung zu Determinanten diagnostischer Lehrerurteile. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 22 (4), 312–324.
- Spinath, B.** (2004). Diagnostische Kompetenzen von Lehrerinnen und Lehrern. *Lernende Schule*, Nr. 26, 16–17.
- Spinath, B.** (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie*, 19 (1–2), 85–95.
- Südkamp, A., Kaiser, J. & Möller, J.** (2012). Accuracy of Teachers' Judgments of Students' Academic Achievement: A Meta-Analysis. *Journal of Educational Psychology*, 104 (3), 743–762.
- Weinert, F.E. & Helmke, A.** (1987). Schulleistungen – Leistungen der Schule oder der Kinder? *Bild der Wissenschaft*, 24, 62–73.

### Autor und Autorin

**Alois Buholzer**, Prof. Dr., PH Luzern, Institut für Schule und Heterogenität (ISH),  
alois.buholzer@phlu.ch

**Sandra Zulliger**, lic. phil., PH Luzern, Institut für Schule und Heterogenität (ISH),  
sandra.zulliger@phlu.ch

