

# Die Entwicklung der Diagnosekompetenz bei PH-Studierenden

**Eine empirische Studie**

**Institut für Schule und Heterogenität (ISH)**

**Alois Buholzer  
Sandra Zulliger**

# **Die Entwicklung der Diagnosekompetenz bei PH-Studierenden**

## **Eine empirische Studie**

Alois Buholzer  
Sandra Zulliger

In Zusammenarbeit mit der Q-Gruppe Luzern

Dezember 2010

## **Dank**

Ganz herzlich danken wir Michael Zutavern, der das Projekt von Anfang an unterstützt und interessiert mitverfolgt hat. Danken möchten wir auch Hanni Lötscher, die uns in verschiedenen Phasen des Projekts mit ihrem Fachwissen im Hintergrund beratend zur Seite stand. Des Weiteren danken wir Marianne Müller und Eveline Gutzwiller-Helfenfinger herzlich für die Unterstützung bei der Auswertung der Daten.

Ohne die Mithilfe der Studierenden ist ein solches Projekt nicht zu bewerkstelligen. Den an der Studie teilnehmenden Studierenden danken wir herzlich für Ihre Geduld, die aufgewendete Zeit und ihre Bereitschaft, ihre Kompetenzen auf die Probe zu stellen.

Bitte wie folgt zitieren:

Buholzer, A. & Zulliger, S. (2010). Die Entwicklung der Diagnosekompetenz bei PH-Studierenden. Eine empirische Studie. Luzern: Forschungsbericht Nr. 26 der Pädagogischen Hochschule Zentralschweiz, Hochschule Luzern.

## Abstract

Die vorliegende, an der Pädagogischen Hochschule Zentralschweiz, Luzern (PHZ LU) durchgeführte Längsschnittuntersuchung (2003 bis 2007) geht der Frage nach, wie sich die Diagnosekompetenz der PH-Studierenden während ihres Studiums entwickelt und ob allenfalls der Studiengang (Kindergarten/Primarstufe – Sekundarstufe I), das Fach Deutsch oder die Zugehörigkeit zum regulären oder dem Aufbaustudiengang einen Einfluss auf dessen Verlauf haben.

Aus der Literatur und bisherigen Studien lässt sich wenig Aufschluss darüber gewinnen, wie sich professionelle Kompetenzen von Lehrpersonen insgesamt (und die Diagnosekompetenz im Besonderen) während der Ausbildung aufbauen. Im Zentrum der Diagnosekompetenz steht die Genauigkeit der Beurteilung. Unter schulpädagogischer Perspektive gehören zur Diagnosekompetenz neben treffgenauen Urteilen ebenso die Rückmeldung und die Vorschläge zur Weiterarbeit.

Für die vorliegende Untersuchung wurde dieser Aspekt berücksichtigt und die diagnostische Kompetenz eingegrenzt als die Fähigkeit, (a) eine mündliche Sprachproduktion unter Beachtung der Ziele in der Aufgabenstellung angemessen einzuschätzen, (b) rückzumelden und (c) einen passenden Vorschlag zur Weiterarbeit ableiten zu können. Die auf diese drei Aspekte fokussierte Diagnosekompetenz wurde insgesamt drei Mal während des Studiums erfasst. Die 133 untersuchten PH-Studierenden beantworteten vier offene Fragen zu einer videographierten zweiminütige Radioreportage einer Primarschülerin, eines Primarschülers der 5. Klasse. Die Antworten der Studierenden wurden kategorisiert und mit einem Expertenurteil verglichen.

Der Längsschnitt bestätigt die Erwartungen hinsichtlich einer Zunahme der diagnostischen Kompetenz während des Studiums. Die deutlichsten Fortschritte sind bei der eigentlichen Urteilsleistung zu erkennen: die Aussagen über den Lernstand sind am Ende des Studiums signifikant differenzierter und präziser als noch am Anfang. Bei den inhaltlichen Kriterien kann die Sprache (z.B. die korrekte Aussprache der Standardsprache) im Laufe des Studiums immer besser beurteilt werden. In anderen Bereichen zeigen sich jedoch keine Fortschritte, so etwa in der Beurteilung des sprachlichen Aufbaus der Sprachproduktion. Bei wiederum anderen Bereichen lassen sich teilweise Fortschritte (kurz- und/oder langfristig und je nach Studiengang) feststellen, so beispielsweise bei der Beurteilung des verwendeten Wortschatzes oder bei der Stimmführung.

Die Verläufe der beiden untersuchten Studiengänge Kindergarten/Primarstufe und Sekundarstufe unterscheiden sich voneinander. Für die untersuchten Studierenden der Kindergarten/Unterstufe (KU) und Primarstufe (PS) kann im Verlauf des Studiums generell von einem signifikanten Zuwachs der Diagnosekompetenz ausgegangen werden. Die Fortschritte zeigen sich sowohl zwischen der ersten und zweiten Messung (in der Mitte des Studiums) wie auch zwischen der ersten und dritten Messung (am Ende des Studiums). Es lassen sich also kurzfristige wie längerfristige Fortschritte bei den Studierenden nachweisen.

Für die Studierenden der Sekundarstufe I lässt sich ein ähnlicher Verlauf beschreiben. Es fällt jedoch auf, dass die Studierenden der Sek I - verglichen mit den Studierenden von KU/PS - generell mit höheren Werten ihr Studium beginnen. Sie kommen bereits zu Beginn des Studiums den Anforderungen des Expertenurteils näher. Auch bei den Sek-Studierenden kann im Verlauf des Studiums generell von einem signifikanten Zuwachs der Diagnosekompetenz ausgegangen werden. Die Fortschritte zeigen sich allerdings nicht kurzfristig (zwischen der ersten und zweiten Messung), sondern längerfristig zwischen der ersten und dritten Messung (am Ende des Studiums).

# Inhaltsverzeichnis

<b>1. Einleitung</b> .....	<b>6</b>
<b>2. Theoretische Überlegungen und Forschungsstand zur pädagogischen Diagnostik</b> .....	<b>7</b>
<b>2.1. Bedeutung und Begrifflichkeiten der Pädagogischen Diagnostik</b> .....	<b>7</b>
<b>2.2. Forschungsergebnisse zu Lehrerurteilen</b> .....	<b>11</b>
2.2.1. Studien zur Genauigkeit der Leistungsbeurteilung.....	11
2.2.2. Studien zum Beurteilungsprozess.....	14
<b>2.3. Schulpädagogische Überlegungen</b> .....	<b>16</b>
2.3.1. Verbindung von Diagnoseprozessen und Kommunikation .....	16
2.3.2. Verbindung von Diagnostik und Unterricht .....	17
<b>2.4. Konsequenzen für die Ausbildung von Lehrpersonen</b> .....	<b>20</b>
2.4.1. Kompetenzen und Kompetenzaufbau.....	20
2.4.2. Implikationen für die Aus- und Weiterbildung von Lehrkräften.....	22
<b>3.1. Eingrenzung des Forschungsthemas</b> .....	<b>27</b>
<b>3.2. Fragestellung</b> .....	<b>27</b>
<b>3.3. Forschungsdesign</b> .....	<b>28</b>
<b>3.4. Stichprobe</b> .....	<b>28</b>
<b>3.5. Untersuchungsinstrument</b> .....	<b>29</b>
<b>3.6. Kategoriensystem</b> .....	<b>30</b>
<b>3.7. Auswertung</b> .....	<b>31</b>
3.7.1. Inhaltsanalyse .....	31
3.7.2. Statistische Auswertung .....	31
3.7.3. Interpretation der Ergebnisse.....	32
<b>4. Ergebnisse</b> .....	<b>33</b>
<b>4.1. Kindergarten/Unterstufe und Primarstufe</b> .....	<b>33</b>
4.1.1. Gesamtbeurteilung.....	33
4.1.2. Formale Kategorien.....	34
4.1.3. Inhaltliche Kategorien .....	34
<b>4.2. Sekundarstufe I</b> .....	<b>35</b>
4.2.1. Gesamtbeurteilung.....	36
4.2.2. Formale Kategorien.....	36
4.2.3. Inhaltliche Kategorien .....	37
4.2.4. Subgruppenvergleiche .....	38
<b>4.3. Zusammenfassung der Ergebnisse</b> .....	<b>38</b>
<b>5. Diskussion</b> .....	<b>40</b>



**6. Literatur.....43**

**Anhang .....47**

**Fragebogen.....47**

# 1. Einleitung

Das Grundgerüst der Lehrerinnen- und Lehrerbildung – insbesondere an den Hochschulen in der Schweiz und in Österreich – basiert auf Kompetenzbeschreibungen. In definierter, differenzierter und möglichst präziser Weise sind die Ziele und die erwarteten Ergebnisse des Unternehmens „Lehrerinnen- und Lehrerbildung“ ausgeführt. Die Beschreibung, was die PH-Studierenden zu unterschiedlichen Abschnitten und am Ende ihrer Ausbildung können und wissen sollen, bildet daher für die Ausgestaltung der verschiedenen Lehrangebote (Module) wie auch für die Gesamtausbildung einen wichtigen Richtpunkt. Diese Zielvorstellungen sind dahingehend formuliert, dass deren Umsetzung nicht nur deklariert, sondern auch empirisch überprüft werden kann (vgl. Allemann-Ghionda & Terhart, 2006).

Der vorliegenden Arbeit liegt eine Forschungsstudie zur Entwicklung einer zentralen professionellen Kompetenz – der Diagnosekompetenz – zu Grunde. Diagnostizieren, d.h. Schülerinnen und Schüler<sup>1</sup> in ihrem Lernen zu unterstützen, den Lernstand zu beurteilen und Hilfen bei Lernschwierigkeiten anzubieten, gehört zum Inventar unterrichtlichen Handelns. Insbesondere wenn die Heterogenität einer Lerngruppe anerkannt und als Grundlage für die Unterrichtsgestaltung angenommen wird, ist eine profunde pädagogische Diagnostik unabdingbar. Entsprechend spielt die Diagnosekompetenz von Lehrerinnen und Lehrern eine bedeutsame Rolle für erfolgreiches Unterrichten.

Obwohl dieser Kompetenz eine hohe Bedeutung zugemessen wird, war sie lange nicht Gegenstand der Forschung. Eine Kehrtwendung und in der Folge auch neue Massnahmen und Impulse gingen von den PISA Erhebungen 2000 aus (vgl. Helmke, Hosenfeld & Schrader, 2004, S. 1). Die in der PISA Erhebung 2000 untersuchten deutschen Klassen- bzw. Deutschlehrkräfte hatten weitgehend Mühe, Schülerinnen und Schüler mit schwachen Leseleistungen in ihren Klassen zu erkennen. Die Diagnosekompetenz der Lehrpersonen ist seither in der erziehungswissenschaftlichen Diskussion stark präsent (vgl. Artelt, Stanat, Schneider & Schiefele, 2001, S. 119).

An der PHZ Luzern gehören die Ausbildung und die Förderung der Diagnosekompetenz der Studierenden zum Ausbildungsprogramm. Mit der vorliegenden Studie soll aufgezeigt werden, wie sich die Diagnosekompetenz der PH-Studierenden während der Ausbildung entwickelt. Rückschlüsse einzelner Ausbildungselemente, seien es im Pflicht- oder Wahlprogramm angebotene Ausbildungselemente oder die praktische Ausbildung an den Kooperationschulen, auf die Kompetenzentwicklung der Studierenden sind dabei nicht eindeutig möglich. Vielmehr geht es darum, die Entwicklung der Diagnosekompetenz über die ganze Ausbildungsspanne hinweg aufzuzeigen und daraus weitere Erkenntnisse zu dem wenig erforschten Kompetenzbereich zu gewinnen und allenfalls auch Hinweise für die Optimierung der Lehrerinnen- und Lehrerbildung zu erhalten.

Im ersten Teil des Berichts werden der theoretische Hintergrund und der aktuelle Forschungsstand zur pädagogischen Diagnostik aufgezeigt. Die Anlage und die Ergebnisse der zwischen 2003 und 2007 durchgeführten Längsschnittuntersuchung werden im darauf folgenden Kapitel vorgelegt. In einem letzten Teil werden die Ergebnisse der Untersuchung diskutiert.

---

<sup>1</sup> Aus Gründen der besseren Lesbarkeit wird auf die durchgehende Verwendung der männlichen und weiblichen Schreibweise verzichtet. Die männliche Form bezieht sich in solchen Fällen auf beide Geschlechter.

## 2. Theoretische Überlegungen und Forschungsstand zur pädagogischen Diagnostik

Im ersten Teil der theoretischen Fundierung der Arbeit werden die Begrifflichkeiten und die Bedeutung der pädagogischen Diagnostik einleitend und in einer Gesamtübersicht dargelegt. Einzelne, in den darauffolgenden Kapiteln behandelte Themen, werden dabei bereits angeschnitten. Studien zum Prozess und der Genauigkeit von Lehrerurteilen bilden das zweite Kapitel. Die Einbettung der Pädagogischen Diagnostik in das unterrichtliche Handeln erfolgt im dritten Kapitel. Der Vermittlung von Kompetenzen (an Pädagogischen Hochschulen) ist das vierte Kapitel des Theorieteils gewidmet.

### 2.1. Bedeutung und Begrifflichkeiten der Pädagogischen Diagnostik

In der Diskussion um einen gelingenden Umgang mit Heterogenität (vgl. Grossenbacher 2010), in aktuellen Modellen der Unterrichtsgestaltung (vgl. Beck, Baer, Guldimann, Bischoff, Brühwiler, Müller, Niedermann, Rogalla & Vogt 2008) oder in Fragen der individuellen Förderung (vgl. Fischer, 2009) wird häufig auf die Kompetenz des Diagnostizierens verwiesen. Allgemein wird dieser Kompetenz eine Schlüsselstellung zur Förderung von Lernerfolg von Schülerinnen und Schülern zugesprochen.

Dementsprechend wird gefordert, dass Lehrpersonen durch Aus- und Weiterbildung diese Kompetenz aufbauen. So führt Girmes (2006) in ihrem Tableau der professionellen Kompetenzen als erste von insgesamt 6 Kompetenzen, die diagnostische Kompetenz auf. Und Oser (1997a) fasst in seiner Übersicht unter dem Titel „Schülerunterstützende Beobachtung und Diagnose“ Ausbildungsziele zusammen, welche die Erfassung von Entwicklungen und auftretenden Problemen betreffen (mehr dazu im Kapitel zu den Konsequenzen für die Ausbildung von Lehrpersonen, Seite 20).

Die Pädagogische Diagnostik lässt sich vom Begriff der Diagnostik her erschliessen. Diagnostizieren heisst „Durchforschung“ im Sinne von „Unterscheidung“ oder „Entscheidung“, gebildet aus „dia“ („durch“) und „gnostik“ („Erkenntnis“ oder „Urteil“). Diagnostizieren bedeutet somit auseinanderhalten, erkennen oder unterscheiden und beinhaltet insbesondere eine Urteilsbildung. Die pädagogische Diagnostik bezieht sich auf den Kontext des Lernens und Leistens von Schülerinnen und Schülern – es ist eine Diagnostik, „welche Lernvorgänge einschliesslich ihrer Voraussetzungen, Rahmenbedingungen und Ergebnisse untersucht, um aktuelle Lernvorgänge zu optimieren oder Lernerfolge für die Gesellschaft zu dokumentieren“ (Ingenkamp, 1997, S. 16).

Angelegt in der Konzeption von Ingenkamp sind zwei grundsätzlich verschiedene Ausrichtungen der pädagogischen Diagnostik: die Selektionsdiagnostik und die Förder- bzw. Lernprozessdiagnostik (vgl. Buholzer, 2010). Erstere zielt darauf ab, mit verschiedenen Formen der Leistungsmessung festzustellen, wie gut die Schülerinnen und Schüler den vermittelten Lerninhalt beherrschen. Die Funktion der Diagnostik besteht in diesem Sinne in der abschliessenden Beurteilung und Dokumentation der Lernleistungen von Schülerinnen und Schülern. Ihre Ergebnisse dienen der Vorbereitung für Entscheide der Bildungslaufbahn mit längerfristigen und folgenreichen Wirkungen. Eine förderorientierte Diagnostik hingegen beabsichtigt die Analyse von Lernprozessen mit dem Ziel, diese bestmöglich durch individuelle Förderung zu stützen und weiterzubringen. Die Ausrichtung einer förderorientierten Diagnostik steht somit im Dienst gezielter (und begründeter) pädagogischer Förderung (vgl. Horstkemper, 2006). Die förderdiagnostische Analyse um-

fasst die jeweiligen Lernbereiche, die Entwicklung von Förderperspektiven und die Frage, mit welchem Angebot die Lernentwicklung bestmöglich unterstützt werden kann.

Im pädagogischen Kontext kann allgemein unter Diagnostik das „Insgesamt von Erkenntnisbemühungen im Dienste aktueller pädagogischer Entscheidungen“ (Klauer, 1978, zit. nach Horstkemper, 2004, S. 204) verstanden werden. Das Erkenntnisinteresse und der dahinter stehende Zweck, das Treffen von pädagogischen Entscheidungen, sind nach Klauer die beiden Elemente, welche zusammen die pädagogische Diagnostik ausmachen. Das Feld der pädagogischen Diagnostik geht somit über die herkömmliche Leistungsmessung und -beurteilung hinaus und umfasst nach Lukesch (1998) die folgenden Bereiche einer „pädagogisch-psychologischen Diagnostik“:

- Lernergebnisse (Schulleistungen)
- Lernvoraussetzungen (z.B. Vorwissen, sprachlicher Entwicklungsstand, kognitive Bedingungen oder Interesse)
- Fähigkeiten für Schuleignung (Schuleingangsdiagnose) und Schullaufbahn
- Spezialbegabungen (kognitive Hochbegabung, musikalisches Potenzial, gestalterische Fähigkeiten, motorische Fähigkeiten, kreative Fähigkeiten)
- persönlichkeitspezifische Vorbedingungen für schulisches Lernen (motivationale, emotionale, begabungsmässige Merkmale)
- Teilleistungsschwächen
- Behinderungen in Bezug auf körperliche oder kognitive Merkmale
- Gruppenprozesse in der Schulklasse
- Verhalten der Lehrperson (z.B. Instruktionsverhalten, -qualität, Pygmalion-Effekt)
- Lernumfeld (Familien-, Peergruppen-, Medieneinflüsse) (vgl. Lukesch, 1998, S. 35).

Entsprechend sollen Lehrpersonen über die Kompetenzen verfügen, pädagogische Diagnosen in vielerlei Hinsicht zu erstellen. Nach Schrader handelt es sich dabei um die Fähigkeit, „die zur Optimierung pädagogischer Entscheidungen erforderlichen Diagnoseaufgaben sach- und zielgerecht auszuführen. Das Spektrum möglicher Diagnoseaufgaben reicht dabei 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).

Diagnoseprozesse, welchem Zwecke sie auch dienen, sind so zu gestalten, dass Leistungen von Schülerinnen und Schülern korrekt gemessen und beurteilt werden. Es sind also möglichst präzise Diagnosen zu erstellen, die Aufschluss darüber geben, welche vermittelten Lernziele beherrscht und welche noch nicht beherrscht werden. Um genaue Lerndiagnosen bilden zu können, ist die Einhaltung der drei Gütekriterien Validität (Zielbezug), Reliabilität (ausreichende Zuverlässigkeit) und Objektivität von hoher Bedeutung. Dies impliziert eine Ausrichtung des Prüfverfahrens auf die im vorangehenden Unterricht bearbeiteten Lernziele. Die Reliabilität wird dann erreicht, wenn bei einer nochmaligen Durchführung ähnliche Resultate gewonnen werden und möglichst alles, was die Zufälligkeit von Prüfungsergebnissen erhöht, vermieden wird. Das Kriterium der Objektivität wird in der schulischen Leistungsmessung dann umgesetzt, wenn die Leistungsfeststellung und -beurteilung möglichst unabhängig von subjektiver Beeinflussung z.B. durch Verzerrungen oder Stereotypen der Lehrkraft erfolgt.

Im Kern einer Diagnoseleistung liegt also die sogenannte Urteilsgenauigkeit. Die Urteilsgenauigkeit kann je nach Person (Diagnostiker) und dem zu beurteilenden Merkmal unterschiedlich ausgeprägt sein. Um die Urteilsgenauigkeit auszuweisen wird häufig auf das Mass der Akkuratheit

verwiesen. Dieses Mass beschreibt die Abweichung der subjektiven Beurteilung vom wahren respektive objektiven Wert. Grundsätzlich kann die Beurteilung bei einer oder mehreren oder bei allen zu beurteilenden Personen sowie bei einer, mehreren oder allen Dimensionen akkurat oder abweichend sein.

In Anlehnung an Cronbach unterscheiden Schrader und Helmke (1987) für den Schulkontext drei Urteils-komponenten respektive Indikatoren der Akkuratheit. Die Niveauelemente beschreibt die Tendenz von Lehrpersonen, die Merkmalsausprägung der Klasse im Vergleich zu den tatsächlichen Schülerleistungen zu über- oder unterschätzen. Die Differenzierungskomponente umschreibt die Tendenz der Lehrpersonen, die Streuung des eingeschätzten Merkmals im Vergleich zur tatsächlichen Leistung zu über- oder unterschätzen. Analog beschreibt die Vergleichskomponente die Tendenz, inwiefern Lehrpersonen die relative Merkmalsausprägung der Schüler untereinander korrekt einzuschätzen vermögen (vgl. Schrader & Helmke, 1989).

Diagnoseprozesse können in Teilprozesse differenziert werden. Heidemeier (2005) unterscheidet drei aufeinanderfolgende Teilprozesse: Beobachtung (observation), Beurteilung (retrieval/judgment) und Kommunikation (siehe Abbildung 1). Der Teilprozess Kommunikation wurde in Ergänzung zum ursprünglichen Modell von Landy und Farr hinzugefügt und wird v.a. durch den Beurteilungszweck (z.B. Förder- vs. Selektionsfunktion) beeinflusst. Gemäss dem Modell wirken sich die folgenden Faktoren auf die Beurteilungsleistung aus: (a) Eigenheiten des Instruments und der eingesetzten Skalen (Bezugsnormen), (b) situative Rahmenbedingungen der Leistungsbeurteilung, (c) Eigenschaften der untersuchten Personen und ihrer beruflichen Position.

Der erste Teilprozess, das Beobachten, wird durch die gebotenen Gelegenheiten zum Beobachten und durch die Motivation des Beobachters, allgemeiner zusammengefasst durch die berufliche Position (position characteristics), beeinflusst.

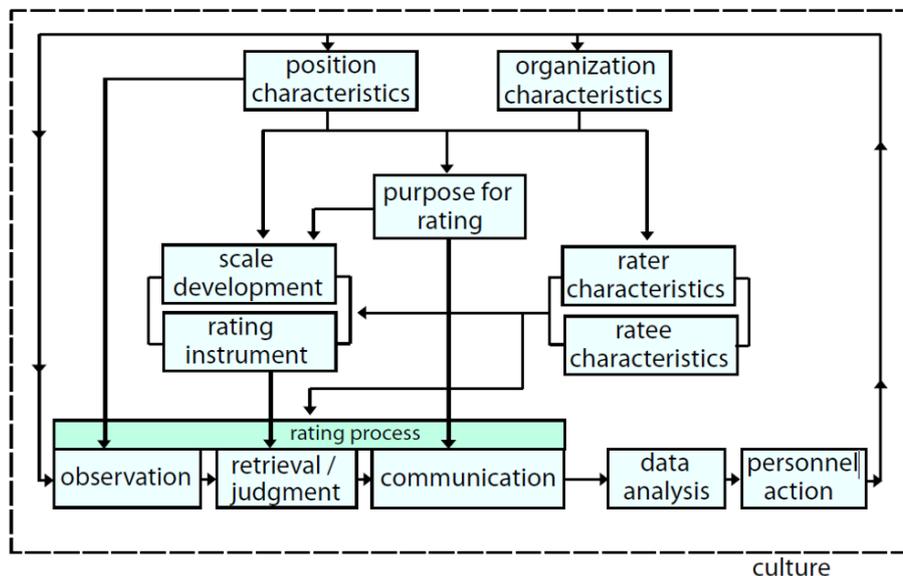


Abbildung 1: Process model of performance rating. A modification and extension of the model suggested by Landy & Farr 1980 (Heidemeier, 2005, S. 66)

Die durchgeführten Beobachtungen werden in einem zweiten Teilprozess beurteilt, wobei hier das Beurteilungsinstrument mit seinen Eigenheiten und die eingesetzten Skalen einen grossen Einfluss ausüben. Das Beurteilungsinstrument selbst wird indirekt durch den Beurteilungszweck (purpose for rating), die Charakteristiken des Beurteilers und der Beurteilten (rater characte-

ristics, rate characteristics) und durch den organisatorischen Rahmen bestimmt. Der dritte Teil des Beurteilungsprozesses, die Kommunikation des Beurteilungsergebnisses nach Aussen, wird auch stark durch den Beurteilungszweck bestimmt. Die Person des Beurteilers beeinflusst den Beurteilungsprozess als Ganzes. Wie in den weiter unten folgenden Abschnitten aufgezeigt wird, beeinflussen bestimmte Merkmale, wie z.B. die kognitive Komplexität, den Beurteilungsprozess. Die beiden Output-Elemente des Beurteilungsprozesses, die Datenanalyse und die Personalentscheidung (personnel action) können in Bezug auf die Schule mit dem Selektionsprozess gleichgesetzt werden. Als übergeordnetes und den ganzen Prozess umfassender Beeinflussungsfaktor gilt die Kultur (vgl. Heidemeier, 2005, S. 66).

Wie ein Merkmal beurteilt wird, hängt von der verwendeten Beurteilungsnorm ab. Üblicherweise unterscheidet man zwischen der Lernzielnorm (Orientierung an fachlichen Anforderungen), der Sozialnorm (Orientierung an der sozialen Gruppe, zum Beispiel die Schulklasse) oder der Individualnorm (Orientierung am individuellen Lernfortschritt) (vgl. Ingenkamp, 1997, S. 24). Forschungsbefunde legen nahe, dass die Wahl der Bezugsnorm eng mit dem Lehr- und Lernverständnis der Lehrperson zusammenhängen. Lehrpersonen, die sich v.a. auf die Sozialnorm berufen, weisen häufig ein Lernverständnis auf, das auf ein Begabungskonzept mit zeitlich überdauernden Faktoren rekurriert (vgl. Sacher, 2001, S. 48f.). Lehrpersonen achten gemäss eigenen Angaben besonders auf die Lernziel- und Individualnorm, die Sozialnorm wird als weniger wichtig eingeschätzt (vgl. Roos, 2001, S. 188).

Vereinzelt liegen Studien vor, welche die Effekte der verschiedenen Bezugsnormorientierungen auf das Selbstkonzept und Lernverhalten der Schülerinnen und Schüler untersucht haben (vgl. zusammenfassend Sacher, 2001, S. 41-50). Inwiefern sich die Verwendung der verschiedenen Bezugsnormen auf die Akkuratheit von Beurteilungen auswirkt, ist nicht geklärt. Aufgrund von Leistungsvergleichen zwischen Schulklassen und den Studien von Schrader und Helmke (1987) darf angenommen werden, dass sich die Wahl der Lernzielnorm als Bezugsnorm positiv auf die Beurteilungsakkuratheit auswirkt.

Allgemein wird davon ausgegangen, dass kriterienorientierte Vorgehensweisen einem ganzheitlich/intuitiven Vorgehen überlegen sind (vgl. Baurmann, 2002). Die empirischen Ergebnisse kommen hier jedoch zu unterschiedlichen Ergebnissen. Arendt ermittelt in seiner Studie, dass sich das globale Urteil („Gesamtzugriff“) nicht von einem detaillierten Urteil („Detailzugriff“) unterscheidet (vgl. 1996, S. 6). Viele fachdidaktischen Hilfsmittel verweisen jedoch auf die Notwendigkeit einer kriterienorientierten Beurteilung (vgl. zum Beispiel Jurt Betschart, Hofstetter & Vogel Wiederkehr, 2000).

Bei der Anwendung des geschilderten Beurteilungsprozess kann es zu Fehleinschätzungen kommen. Diese entstehen u.a. dadurch, dass sich das Individuum, eventuell unbewusst, bestimmte Bezugspunkte setzt. Bei der Beurteilung von Schülerleistungen können beispielsweise Vorinformationen über den Schüler, Quervergleiche zu anderen Leistungsbereichen oder Vergleiche mit den Leistungen der Bezugsgruppe (z.B. Schulklasse) das Urteil verfälschen. Bei Urteilsleistungen ist daher immer von Interesse, a) auf welche Informationen der Urteiler zurückgreift und b) an welchen Kriterien er den beobachteten Sachverhalt misst und c) zu welchem Schluss respektive Urteil der Vergleich führt.

## 2.2. Forschungsergebnisse zu Lehrerurteilen

Aufgrund der hohen Bedeutung von Lerndiagnosen bezüglich der Laufbahneempfehlungen wie auch bezüglich der individuellen Lernförderung müsste von den Lehrpersonen eine hohe Kompetenz erwartet werden. Gemäss empirischen Daten ist jedoch von verbreiteten Defiziten in diesem Bereich auszugehen (vgl. zum Beispiel von der Groben, 2003). Diese Erkenntnis ist nicht neu, hat doch Ingenkamp schon 1969 auf diesen Missstand hingewiesen. Besonders für formale Beurteilungen, die mit der Erteilung von Qualifikationen einher gehen, ist ein faires Urteil und somit der Einbezug von Gütekriterien relevant (vgl. Helmke, 2009, S. 126). Inwieweit Fehler in eine Beurteilung eingeflossen sind, kann an der Güte eines Verfahrens abgelesen werden. Die Güte ist am grössten, wenn eine hohe Objektivität, Zuverlässigkeit und Gültigkeit nachgewiesen werden kann.

Es müsste aufgrund der hohen Relevanz erwartet werden, dass Lehrpersonen ihre Urteile gemäss den allgemeinen Gütekriterien von Objektivität, Validität und Reliabilität bilden würden. Verschiedene Untersuchungen zur Notengebung von Lehrpersonen zeigen jedoch, dass die Notenurteile nicht den grundlegenden diagnostischen Gütekriterien entsprechen (vgl. zusammenfassend Ingenkamp, 1997 und Sacher, 2001 oder beispielhaft in Moser, Keller & Tresch, 2003).

Aktuelle Forschungsarbeiten versuchen diesem Kompetenzmangel auf die Spur zu kommen. Gemäss Schrader (2009) konzentrieren sich die Forschungen (a) auf Genauigkeit von Leistungsbeurteilung und (b) auf den Prozess der Urteilsbildung. Die folgenden beiden Unterkapitel nehmen bei der Darstellung der aktuellsten Forschungsergebnisse die Einteilung in diese beiden Richtungen auf.

### 2.2.1. Studien zur Genauigkeit der Leistungsbeurteilung

Die Überprüfung der Genauigkeit von Urteilen ist die gemeinsame Fragestellung einer ersten Gruppe von Forschungsstudien. Dabei wird die Urteilsgenauigkeit als Indikator für die Diagnosekompetenz der urteilenden Person (Lehrkraft) aufgefasst. Die Urteile zu Schüler- und Aufgabenmerkmalen werden mit Ergebnissen verglichen, die aus Testverfahren oder anderen wissenschaftlichen Methoden ermittelt worden sind. Stimmen die Ergebnisse der Lehrerurteile mit den Testergebnissen überein, kann von einer hohen Urteilsgenauigkeit ausgegangen werden.

Lorenz und Artelt (2009) untersuchten die Diagnosekompetenz von 127 Grundschullehrkräften anhand von Längsschnittdaten für die Fächer Deutsch und Mathematik. Die Lehrerurteile wurden mit den gleichzeitig erhobenen Testleistungen der jeweiligen Schulklassen verglichen. Die Ergebnisse zur Güte von Diagnoseurteilen (gemessen mit der Rangkomponente) zeigen, dass Werte zwischen den tatsächlich gemessenen Schülerleistungen und den Lehrerurteilen je nach Bereich und Zeitpunkt zwischen  $r=.54$  und  $r=.66$  zusammenhängen. An den Standardabweichungen kann abgelesen werden, dass es zwischen den Lehrkräften grosse Unterschiede gibt. Neben den interindividuellen Unterschieden zeigen sich jedoch auch intraindividuelle Unterschiede. Die Diagnoseleistungen erfolgen zudem stark fachbezogen und nicht fachübergreifend. Lehrkräfte, die den Wortschatz ihrer Schülerinnen und Schüler gut diagnostizieren können, sind auch in der Lage die Lesekompetenz gut zu beurteilen, nicht aber mathematische Kompetenzen. Die Umkehrung gilt auch, wer über eine hohe diagnostische Kompetenz im mathematischen Bereich verfügt, ist nicht notwendigerweise auch ein guter Diagnostiker im sprachlichen Bereich. Die Stabilitäten der diagnostischen Urteile sind über die beiden Messzeitpunkte mittelhoch, im Bereich Mathematik liegen sie etwas tiefer als im sprachlichen Bereich (Textverstehen und Wortschatz).

Karing (2009) untersuchte diagnostische Urteile von Grundschullehrkräften und Gymnasiallehrkräften zu Schülerleistungen (Arithmetik, Wortschatz, Textverstehen) sowie Fachinteresse in Deutsch. Sie verglich die Güte der Diagnoseleistung von Lehrkräften ebenfalls anhand der Rangordnungskomponente. Die Ergebnisse verdeutlichen, dass Leistungen einfacher zu diagnostizieren sind als nicht-kognitive Merkmale wie das Fachinteresse. Erwartungsgemäss schätzen Grundschullehrkräfte, welche die Schülerinnen und Schüler in mehreren Fächern unterrichten, die Schülerleistungen in Arithmetik, Wortschatz und Textverstehen sowie das Fachinteresse in Deutsch akkurater ein als Gymnasiallehrkräfte. Weitere Aspekte wie die Zusammenarbeit mit Eltern und mit Lehrerkollegen hängen jedoch nicht mit der Diagnoseleistung zusammen. Sowohl bei der Gruppe der Grundschullehrpersonen als auch bei den Gymnasiallehrkräften bestehen beträchtliche Unterschiede zwischen den einzelnen Lehrpersonen. Als bedeutungsvoll für die Genauigkeit der Diagnoseleistung erweist sich die Leistungsheterogenität, denn es besteht ein Zusammenhang zwischen der Leistungsstreuung und der Diagnosekompetenz. Eine hohe Leistungsstreuung der Klasse (sowohl in der Grundschule wie im Gymnasium) hängt positiv mit der Urteilsgüte der Lehrkräfte zusammen. Man kann also annehmen, dass sich eine höhere Heterogenität der Merkmale urteilserleichternd auswirkt.

Schrader wies bereits 1998 nach, dass Lehrpersonen in Klassen, in denen die Leistungen der Schülerinnen und Schüler sehr heterogen ausfallen, eine bessere diagnostische Kompetenz entwickeln. Er stellte des Weiteren fest, dass sich Lehrpersonen in ihrer diagnostischen Kompetenz nicht darin unterscheiden, (1) ob sie in grossen oder kleinen Klassen unterrichten, (2) wie viel Unterrichtszeit sie in der Klasse unterrichten, (3) wie gross der Anteil der fremdsprachigen Kinder (4) oder wie leistungsfähig die Klasse ist. Erwartungswidrig ist bestimmt die Tatsache, dass in kleineren Klassen die diagnostische Kompetenz nicht zunimmt, obwohl hier der Vorteil bestünde, sich intensiver mit einzelnen Kindern zu beschäftigen (vgl. 1989, S. 121-124). Wild und Rost bestätigen diesen Sachverhalt: Bei kleineren Schulklassen wirkte sich eine längere Kontaktdauer nicht auf die Urteilsgenauigkeit aus. Die oftmals geforderten kleineren Klassen ermöglichen somit keine genaueren Diagnosen durch die Lehrpersonen. Es wird jedoch vermutet, dass bei grösseren Klassen die Lehrpersonen mehr Aufwand betreiben, um zu den gleichen Ergebnissen zu gelangen (vgl. Wild & Rost, 1995, S. 86f.).

Von der Groeben spricht mit Blick auf Analysen bei PISA von einem Kompetenzmangel (2003, S. 4f.). In den Auswertungen zeigt sich, dass die meisten der „schwachen“ Leserinnen und Leser von den Lehrpersonen gar nicht erkannt worden sind. Nur etwa 11% der Jugendlichen, welche im PISA-Test unterhalb der Kompetenzstufe I (das niedrigste Niveau von insgesamt 5 Stufen) lagen, wurden von den Lehrpersonen als „schwache“ Leserinnen und Leser diagnostiziert. Knapp 4% der Schülerinnen und Schüler, welche das Kompetenzniveau I erreichten, wurden von den Lehrpersonen als schwach im Lesen eingestuft (vgl. Artelt et al., 2001, S. 119f.).

Tabelle 1: Anteil von Schülerinnen und Schülern innerhalb der Kompetenzstufen, die von ihren Lehrerinnen und Lehrern als "schwache Leser" identifiziert wurden (in %) (Artelt et al., 2001, S. 119)

Diagnose der Lehrkraft	Leistung im PISA-Test		
	Unter Kompetenzstufe I	Kompetenzstufe I	Über Kompetenzstufe I
Schwache Leser	11.4	3.7	2.8
Nicht schwache Leser	88.6	96.3	97.2

Ebenso ernüchternd sind die Befunde am anderen Ende der Leistungsheterogenität. Im Gegensatz zur Identifikation der Gruppe hochintelligenter Schüler und Schülerinnen mit überdurchschnittlichen Schulleistungen erkennen Lehrpersonen Schüler und Schülerinnen mit überdurchschnittli-

chen Intelligenzwerten, aber nur durchschnittlichen Schulleistungen mässig gut und solche mit unterdurchschnittlichen Schulleistungen überhaupt nicht (vgl. Rost & Hanse, 1997). Den Lehrpersonen fällt es somit schwer, die Leistungen und auch das Potential der Schülerinnen und Schüler richtig einzuschätzen. In Bezug auf die *Beurteilung der absoluten Leistung* eines Schülers / einer Schülerin (Niveauelemente) geben Lehrpersonen also eher ungenaue Schätzungen ab.

McElvany, Schröder, Hachfeld, Baumert, Richter, Schnotz, Horz und Ullrich (2009) orientieren sich ebenfalls an der Urteilsgenauigkeit, sie verwenden jedoch für ihre Untersuchung Lernmaterialien zum integrativen Lesen von Texten mit instruktionalen Bildern. Lehrpersonen werden u.a. aufgefordert, den Schwierigkeitsgrad von Lernaufgaben zu bestimmen. Die Aufgabenschwierigkeit wird als Indikator genommen, wie gut Lehrkräfte in der Lage sind, Anforderungen einzuschätzen. Die Ergebnisse zeigen, dass Lehrpersonen – gemessen an der Rangordnung – die Aufgabenschwierigkeiten mässig genau einschätzen können. Die tatsächlich erbrachten Schülerleistungen werden von den Lehrpersonen treffend in der richtigen Rangfolge gebracht, der Tendenz nach, werden sie leicht unterschätzt. Mit der Berufserfahrung und dem fachdidaktischen Wissen steht die Diagnosekompetenz lediglich in einem schwachen Zusammenhang. Dieser Umstand deutet darauf hin, dass noch wenig geklärt ist, wie Wissen und Erfahrungseinflüsse auf die Diagnosekompetenz wirken. Auffallend ist auch, dass die verschiedenen Dimensionen zur Einschätzung der Aufgabenschwierigkeit und zur Diagnose der effektiv erbrachten Schülerleistungen wenig zusammenhängen.

Diese neueren Befunde stützen im Wesentlichen auch ältere Untersuchungen zur Diagnosekompetenz. So wurden beispielsweise 2002 in der Schulstudie „SALVE“ die unterrichtlichen und individuellen Bedingungen der Lern- und Motivationsentwicklung bei Schülerinnen und Schülern der 5. und 6. Klassenstufe untersucht. Die Lehrpersonen mussten angeben, wie viele Aufgaben die Schülerinnen und Schüler aus einem vorgegebenen Leistungstest lösen können. Anschließend wurde mit einem Test die Schülerleistung gemessen. Die Differenz zwischen dem prognostizierten Wert und dem tatsächlich gemessenen Wert dient dann als Indikator für die Fähigkeit, zur adäquaten Einschätzung des Lernstandes der Schülerinnen und Schüler. Die Vergleiche zeigen, dass die Lehrkräfte durchschnittlich von einem Anteil von 58.7% der Schüler ausgingen, welche die zu beurteilenden 12 Aufgaben lösen können. Die Bandbreite der Einschätzungen reicht dabei von 35.7% bis zu einem Maximum von 82.8%. Die empirischen Lösungshäufigkeiten der Klassen liegen unter den Einschätzungen der Lehrpersonen. Im Minimum werden die 12 Aufgaben von 11.9% der Schüler gelöst, das Maximum liegt bei 60.6%. Im Durchschnitt überschätzen die Lehrpersonen ihre Klassen um 18%, das Maximum liegt hier bei 45.5% (vgl. Hosenfeld, Helmke & Schrader, 2002, S. 72f.).

Lehrpersonen überschätzen – im Gegensatz zur Studie von McElvany et al. – also ihre Schülerinnen und Schüler bezüglich ihrer Leistungen. Nach Hosenfeld et al. könnte der Grund darin liegen, „... dass Lehrpersonen eher die Kompetenz einschätzen – also die prinzipielle Fähigkeit ihrer Schüler, eine Aufgabe zu lösen – und nicht die Performanz, also die tatsächliche Lösung der Testsituation. Dies würde darauf hin deuten, dass die Lehrkräfte leistungsmindernde Faktoren wie das Vergessen bereits durchgenommenen und früher beherrschten Stoffs; die Begrenzung der Bearbeitungszeit in einem Test, Flüchtighkeitsfehler, mangelnde Anstrengung, Aufregung und Leistungsangst usw.) nicht hinreichend in Rechnung stellen“ (Hosenfeld et al. 2002, 79). Unter umgekehrten Vorzeichen werden die Aufmerksamkeit, das Verständnis und das Interesse von Schülerinnen und Schülern beurteilt. Bei diesen Merkmalen liegen die Werte der Lehrpersonen tiefer als jene der Schülerinnen und Schüler. Inwiefern unterschiedliche Wahrnehmungsperspektiven oder selbstwertdienliche Einflüsse der Schülerinnen und Schüler verantwortlich sind, kann mit der Studie nicht erklärt werden.

### 2.2.2. Studien zum Beurteilungsprozess

Eine zweite Gruppe von Studien befasst sich mit dem Prozess der Urteilsbildung. Dabei werden Einflussfaktoren auf den Beurteilungsprozess und die dabei auftretenden Urteilsverzerrungen und Beurteilungsfehler geprüft.

*Verarbeitungsziel:* Die Studie von Krolak-Schwerdt, Böhmer und Gräsel (2009) geht davon aus, dass der Urteilsprozess und die Qualität des diagnostischen Urteils vom Verarbeitungsziel abhängen (vgl. dazu auch das Modell von Heidemeier, 2005, S. 9). Die Untersuchung bezieht sich auf Experten (erfahrene Lehrpersonen) und Laien (Studierende der Naturwissenschaften). Ihnen wurden leicht und schwer kategorisierbare Fallbeschreibungen von Schülerinnen und Schülern vorgelegt, unter der Zielvorgabe, sich entweder einen Eindruck über den Schüler, die Schülerin zu bilden (Eindrucksbildung) oder die weitere Entwicklung des Schülers / der Schülerin vorherzusagen (Vorhersageziel). Dabei wird untersucht, wie die Informationen der Fallbeschreibungen verarbeitet werden: kategoriengeleitet oder merkmalsgeleitet. Kategoriengeleitete Verarbeitungsprozesse laufen oftmals automatisch mit geringem kognitivem Aufwand ab. Damit wird schnelles Reagieren möglich, die resultierenden Urteile sind jedoch durch die verwendeten sozialen Kategorien verzerrt. Merkmalsgeleitete Verarbeitung bedeutet hingegen „... dass die Aufmerksamkeit auf die spezifischen Merkmale der Person gelegt wird, ohne dass soziale Kategorien eine grosse Rolle spielen. Es findet eine eingehende Beschäftigung mit den gegebenen Personinformationen mit hohem kognitivem Aufwand statt. Nachfolgende Abrufprozesse sind daher durch die für die Person spezifischen, individuellen Merkmale gekennzeichnet“ (Krolak-Schwerdt et al., 2009, S. 176). Welche Kategorie aktiviert ist, bleibt also nicht folgenlos, denn sie bilden den Hintergrund auf der die neuen Informationen wahrgenommen, enkodiert, interpretiert und im Gedächtnis gespeichert werden.

Die Ergebnisse zeigen, dass erfahrene Lehrpersonen unter dem Ziel der Vorhersage individuelle Schülermerkmale fokussieren und reproduzieren, unter dem Eindrucksziel verwenden sie eine verfügbare soziale Kategorie (z.B. Managersohn). In diese Richtung weisen auch die Befundmuster für die Reproduktionen. Bei Laien hatte das Verarbeitungsziel (Eindruck vs. Vorhersage) keinen Einfluss auf die Verarbeitung der Schülermerkmale. Diese Ergebnisse bestätigen die Vermutung, wonach erfahrene Lehrpersonen (Experten) Informationen je nach Verarbeitungsziel nutzen. Sie weisen damit eine höhere Flexibilität im Denken auf als Laien. Damit können insbesondere Rückschlüsse auf die Qualität von informellen Diagnoseleistungen – wie sie oftmals im Schulalltag vorkommen – gezogen werden.

*Ankereffekte:* Die Studie von Dünnebier, Gräsel und Krolak-Schwerdt (2009) basiert auf dem gleichen theoretischen Ansatz der sozialen Kognition, wie die Studie von Krolak-Schwerdt, Böhmer und Gräsel (2009). In ihrer Studie prüfen sie, ob Ankereffekte einen Einfluss auf die Urteilsbildung und die Qualität des Urteils haben. In der Studie wurde eine vermeintliche Benotung einer vorgegebenen Leistung mit einer 2 (hoher Anker) oder einer 4 (niedriger Anker) verwendet. Untersucht werden wiederum die Urteile von Novizen (Lehramtstudierende) und von Experten (erfahrene Lehrpersonen) in Bezug auf zwei verschiedene Verarbeitungsziele: (a) erste Eindrucksbildung und (b) folgenschwere Empfehlung für eine Schullaufbahn. Die Ergebnisse bestätigen die Hypothese, wonach Lehrpersonen sich bei der ersten Eindrucksbildung von Ankerinformationen beeinflussen lassen. Müssen jedoch längerfristige Prognosen erstellt werden (in Bezug auf die Schullaufbahn) sind Ankerinformationen für das Urteil bedeutungslos. Bei Novizen hingegen spielen Ankerinformationen unabhängig vom Verarbeitungsziel eine Rolle. Diese Aussagen gelten für das Fach Deutsch, für das Fach Mathematik konnte das Ergebnis nicht bestätigt werden: hier verzerrt der Anker unter allen Bedingungen das Urteil, d.h. ist unabhängig von der Expertise und/oder dem Verarbeitungsziel. Zusammenfassend schreiben die Autoren: „Die gerin-

gere Anfälligkeit der Lehrer-Experten unter der Bedingung der Prognose gibt Hinweise darauf, dass schwerwiegende und folgenreiche Diagnosen, wie es Schullaufbahneempfehlungen darstellen, unter Berücksichtigung und Abwägung aller relevanten Personmerkmale, also eher merkmalsbasiert, getroffen werden und damit weniger anfällig für Fehlentscheidungen oder Urteilsverzerrungen sind“ (Dünnebier et al., 2009, S. 193). Damit wird auch die These unterstützt, dass mit zunehmender Expertise die Informationsverarbeitung besser der Situation angepasst werden kann und somit die Flexibilität im Denken erhöht wird.

Südkamp und Möller (2009) untersuchten mit der Methode des simulierten Klassenraums, wie die Einschätzung von Fähigkeiten hypothetischer Schüler durch Informationen zum Leistungsniveau der Klasse beeinflusst wird. In ihrer Studie gingen sie der Frage nach, ob sich bei direkten (Schätzung des Anteils korrekter Antworten) und indirekten Leistungseinschätzungen (Notenangaben) Referenzgruppeneffekte zeigen. Diese als „Big-Fish-Little-Pond Effekt“ bezeichnete Einflussgrösse wurde u.a. für das akademische Selbstkonzept nachgewiesen. Die Ergebnisse der Studie von Südkamp und Möller (2009) belegen Referenzgruppeneffekte bei indirekten Einschätzungen (Noten), nicht aber bei direkten Einschätzungen von Schülerleistungen. In Klassen mit niedrigem Leistungsstand fielen die Beurteilungen bei identischem Leistungsniveau günstiger aus als in Klassen mit hohem Leistungsniveau.

*Soziodemographische Merkmale:* Weitere Belege zu Verzerrungen der Beurteilung auf Grund bestimmter soziodemografischer Merkmale können an Forschungsergebnissen zur Beschulung von Migrantenkindern (vgl. Kronig, Haeberlin & Eckhart, 2000, S. 146; Stallmann, 1999) oder zur Beschulung von Kindern aus bildungsfernen Schichten abgelesen werden (vgl. Coradi Vellacott & Wolter, 2005, S. 26). Entsprechende Belege finden sich auch bei Empfehlungen von Primarlehrpersonen beim Übergang in die Sekundarstufe I (vgl. Ditton, Krüsken & Schauenberg, 2005, S. 298f.). Angelsächsische Studien nennen des Weiteren die physische Attraktivität von Schülerinnen und Schülern, ihr Geschlecht, ihr Benehmen sowie ihre ethnische Zugehörigkeit als Einflussfaktoren für die Schülereinschätzung durch Lehrpersonen (vgl. Dusek & Joseph, 1985). Die aufgezählten leistungsunabhängigen Kriterien wirken sich dabei besonders bei durchschnittlichen Schülerinnen und Schülern aus. Bei Schülerinnen und Schülern mit weit vom Mittelfeld entfernten Leistungen ist die Wahrscheinlichkeit einer leistungsgerechten Zuweisung am höchsten (vgl. SKBF 2006, S. 65).

Die aktuelle empirische Befundlage kann mit Schrader (2009) wie folgt zusammengefasst werden: Zur *Reliabilität und Stabilität* von Diagnoseleistungen bei Lehrpersonen ist wenig bekannt, erste Anhaltspunkte zur Stabilität finden sich in der Studie von Lorenz und Artelt (2009), die von einer mittleren Stabilität bei Lehrerurteilen berichten. Was *Zusammenhänge zwischen Komponenten und Merkmalsbereichen* anbetrifft, muss aufgrund der Studienergebnisse eher von einer hohen Merkmalspezifität ausgegangen werden. Dies zeigt sich u.a. auch aufgrund der Fachbezogenheit: Wer im Fach Deutsch ein guter Diagnostiker ist muss deshalb nicht auch in den Naturwissenschaften über die entsprechende Kompetenz verfügen. Es muss daher im Anschluss an Spinath bezweifelt werden, ob man überhaupt von einem einheitlichen Konstrukt der Diagnosekompetenz ausgehen kann. Wie die *Berufserfahrung* mit Diagnoseleistungen zusammenhängt, ist nicht eindeutig geklärt. Teilweise können Effekte bei erfahrenen Lehrkräften nachgewiesen werden, teilweise bestätigen sie sich nicht. Diese Befundlage ist auch Hinweis dafür, die Berufserfahrung von Lehrkräften noch differenzierter und in Zusammenhang mit interferierenden Variablen in den Studien abzubilden. Schliesslich gibt es bis anhin kaum Studien, welche die *Wirkungen diagnostischer Kompetenz* auf das Leistungsverhalten der Schülerinnen und Schüler belegen. Bis jetzt liegt nur die Studie von Schrader und Helmke vor, welche die Funktion von hoher Diagnosekompetenz auf die Schülerleistungen aufgezeigt hat: Bei Schülerinnen und Schülern mit günstigen kognitiven Lernvoraussetzungen wirkt eine hohe Diagnosekompetenz nur schwach mode-

rierend. „Schüler mit ungünstigen kognitiven Eingangsvoraussetzungen profitieren dagegen von hoher Unterrichtsstrukturierung erst dann stärker, wenn gleichzeitig auch die Diagnosekompetenz des Lehrers hoch ist“ (Schrader, 1989, S. 228-229). (weitere Ausführungen dazu in Kapitel 2.3.2, Seite 17)

Insgesamt kann mit Schrader jedoch festgehalten werden, dass die empirische Forschungsbasis zur diagnostischen Kompetenz von Lehrpersonen und ihren Bedingungen und Folgen bislang noch recht schmal ist: „Für ein umfassendes Verständnis diagnostischer Kompetenz ist es erforderlich, Aufschluss über die konzeptuellen und methodischen Grundlagen dieses Merkmals, seine Erscheinungsformen und Entstehungsbedingungen zu gewinnen und die Wirkungskette von der Nutzung dieser Kompetenz für diagnostische Urteile und Entscheidungen bis hin zu den davon abhängigen Handlungen und deren Folgen aufzuhellen“ (Schrader, 2009, S. 238).

Insbesondere wissen wir wenig darüber, wie die Diagnosekompetenz im Lauf der Lehrerinnen- und Lehrerausbildung erworben wird. Das vorliegende Projekt setzt sich mit dieser Thematik auseinander.

## 2.3. Schulpädagogische Überlegungen

Unter schulpädagogischer Perspektive wird das Diagnostizieren von Lehrkräften nicht nur zum Beurteilungsprozess und der Güte des Urteils diskutiert. Einbezogen werden Fragen des Rückmeldens an den Schüler / die Schülerin und Überlegungen zur nachfolgenden pädagogischen Intervention.

### 2.3.1. Verbindung von Diagnoseprozessen und Kommunikation

Schulpädagogische Überlegungen gehen davon aus, dass die Aussagekraft von Diagnosen erhöht wird, wenn die Erstellung von Diagnosen in Kooperation mit Lehrerkollegen erfolgt. Durch den Informationsaustausch können diagnostische Aussagen breiter abgestützt werden. Es fließen mehr diagnostische Angaben und unterschiedliche Sichtweisen und diagnostische Modelle in die Urteilsbildung und die nachfolgende Förderplanung ein. Blinde Flecken und Einseitigkeiten können so vermieden werden, was zu treffgenaueren und präziseren Lerndiagnosen führt. Kritische Rückmeldungen zum eigenen Urteil helfen zudem, die Offenheit und die Suche nach weiterer Differenzierung und Ergänzung systematisch weiterzuführen (vgl. Horstkemper, 2006). Zudem werden durch die Gegenüberstellung von Fremd- und Selbstperspektive Diagnoseinformationen validiert. Auf die Vorteile, welche eine kollegiale Praxis mit sich bringt, verweist Staub mit dem fachspezifisch-pädagogischen Coaching. Im Rahmen dieses Coachings lernen Lehrerinnen und Lehrer on the job, ihren Unterricht an Kernperspektiven zu orientieren und zu reflektieren. Dabei spielen insbesondere diagnostische Fragen wie die folgenden eine gewichtige Rolle: Welches lernzielrelevante Vorwissen kann bei welchen Schülern vorausgesetzt werden? Erwarten wir von (bestimmten) Schülern bestimmte Verstehensschwierigkeiten? (vgl. Staub, 2006, S. 139)

Ausserdem ist es bedeutsam, Schülerinnen und Schüler aktiv an der Aufklärung und Optimierung ihrer Lernprozesse zu beteiligen. Zu diesem Zweck ist nach Formen des Feedbacks zu suchen, die es den Lernenden erlauben, relevante Informationen für ihren weiteren Lernprozess zu entnehmen. Exemplarisch kann dieser Ansatz mit der dialogischen Diagnostik von Ruf (2003) aufgezeigt werden. Dieser Zugang ermöglicht es, das eigene Lernen und das Lernen anderer unter die Lupe zu nehmen und daraus die entsprechenden Konsequenzen abzuleiten. Eine spezifische Form sind Gespräche, welche begleitend zum Lernprozess durchgeführt werden. Diese Gesprächsform,



manchmal auch als „(klinisches) Interview“ bezeichnet, erlaubt Einblicke in die Lernprozesse des Schülers oder der Schülerin und lässt damit die metakognitiven Fähigkeiten sichtbar werden. Ein Lehr-Lern-Dialog findet im Rahmen einer individuellen Lernbegleitung statt. Die Lehrperson begleitet den Lernprozess, indem sie einerseits den Lösungsvollzug beobachtet und andererseits spezifische Fragen zu den festgehaltenen Sachverhalten stellt. Der Schüler kann beispielsweise aufgefordert werden, die Aufgabe, die er lösen soll, in eigenen Worten zu schildern, oder er soll aufzeigen, mit welchen Strategien er die Problemlösung vorantreiben will. Leichte Veränderungen in der Aufgabenstellung (was wäre, wenn) oder ein Wechsel der Repräsentationsebenen können weitere Erkenntnisse liefern.

Besonders deutlich kommt der Dialog zum Zug, wenn das Lernen durch bestimmte Bedingungen beeinträchtigt, belastet oder behindert wird. In diesem Fall sind Förderkonferenzen sinnvoll, um mit gemeinsamen Absprachen einen kooperativen Diagnoseprozess in Gang zu setzen. Mit dem „Roundtable“ stellt Steppacher ein Konzept der Zusammenarbeit gleichberechtigter Partnerinnen und Partner vor. Der runde Tisch verfolgt das Ziel, Förderbedürfnisse von Schülerinnen und Schülern zu analysieren, Umsetzungsmassnahmen zu bestimmen und zu evaluieren (vgl. Steppacher, 2004, S. 20).

### **2.3.2. Verbindung von Diagnostik und Unterricht**

Diagnostische Urteile alleine sind nutzlos, erst in Verbindung mit einer entsprechenden Unterrichtsgestaltung oder einer individuellen Massnahmenplanung entfalten sie ihre Wirkung (vgl. Schrader & Helmke, 1987). Charakteristisch für die pädagogische Diagnostik ist daher der Umstand ihrer engen Verknüpfung mit dem unterrichtlichen Handeln. Für Horstkemper bilden zutreffende Diagnosen die Basis für pädagogisches Handeln und geben Hinweise für die Unterrichtsgestaltung: „Wenn die Lernausgangslage in einer Lerngruppe nicht richtig eingeschätzt wird, dann kann der Unterricht nicht optimal auf die Adressaten hin konzipiert werden“ (Horstkemper, 2006, S. 6). Diese enge Verbindung fusst auf einer pragmatischen und einer konzeptionellen Grundlage. Aus pragmatischen Gründen finden diagnostische Prozesse im schulischen Kontext in der Regel während des regulären Unterrichts statt. Die Durchführung von Lernstandsanalysen, Gespräche über Lernprozesse oder Kontrollen zur Zielerreichung am Ende eines Lernzyklus gehören zum Inventar unterrichtlichen Handelns. Darüber hinaus finden informelle Diagnosen ohnehin im laufenden Unterricht statt. Diese erfolgen dann, wenn Störungen auftreten, die Lernenden nicht mehr weiterkommen oder generell im Unterrichtsverlauf eine mangelnde Passung zwischen Anforderungsstruktur und Lernvoraussetzungen festgestellt wird. Unterricht kann denn auch über weite Strecken als eine Abfolge einfacher Diagnosen verstanden werden: Lehrerinnen und Lehrer nehmen Verhaltensweisen wahr, interpretieren sie (meist aufgrund bewährter Alltagstheorien), versuchen die Situation oder das Verhalten zu verstehen und handeln entsprechend. Vom diagnostischen Urteil hängt der weitere Verlauf des Unterrichts, die Reaktion auf eine bestimmte Verhaltensweise oder die Bereitstellung von Hilfestellungen und zusätzlichen Erklärungen ab. Aus konzeptioneller Sicht ist die Trennung von Diagnose und Förderung zu überwinden, denn pädagogische Diagnostik steht immer im Dienste der Unterrichtsgestaltung. Erst durch die Verbindung von Diagnostik und Förderung können die entsprechenden Wirkungen in Bezug auf den Lernerfolg erzielt werden. In diesem Sinne ist auch die Aussage von von der Groeben (2003) zu verstehen, die Diagnostik als „didaktische Herausforderung“ bezeichnet.

Pädagogische Diagnostik und Unterricht gründen häufig auf ähnlichen (fach-)didaktischen Stufenmodellen, was zu einer weiteren Annäherung der beiden Teilbereiche pädagogischen Handelns führt. Insbesondere wenn es um die diagnostische Erfassung von Lernständen zum Beispiel im Bereich des Schriftspracherwerbs oder des Zahlbegriffserwerbs geht, werden in der Regel Stu-

fenmodelle berücksichtigt, die auch für die Unterrichtsgestaltung Gültigkeit haben. Beispielsweise kann das Modell von Valtin (1996), das die Leseentwicklung stark mit dem Erwerb des Schreibens verschränkt und die Entwicklung in Stufen aufzuzeigen vermag, sowohl der Lernstandserfassung dienen wie gleichzeitig als didaktisches Modell zur Unterrichtsplanung und -gestaltung einbezogen werden.

Zwei diagnostische Ansätze legen den Schwerpunkt auf die Verbindung von Diagnoseprozess und pädagogischen Massnahmen:

Die *strukturbezogene Diagnostik* verbindet unterrichtliche Prozesse mit diagnostischen Fragen. Mit Hilfe eines Lernstrukturgitters werden personbezogene Lernprozesse in Abhängigkeit von aufgabenbezogenen Merkmalen untersucht (vgl. Kutzer, 1999). Das Lernstrukturgitter weist die beiden Dimensionen Komplexität und Niveau auf und kann den stufenartigen Verlauf des Lernprozesses in der Abhängigkeit der beiden erwähnten Dimensionen aufzeigen. Die Dimension Komplexität umfasst die objektive Sachstruktur, welche einen Sachverhalt, einen Lerninhalt oder ein Phänomen bestimmt. Die Sachstruktur gliedert sich in verschiedene Komplexitätsstufen und reicht von basalen bis hin zu komplexen Strukturen. Die Dimension Niveau beschreibt die stufenweise Verinnerlichung objektiver Sachverhalte auf der Seite des Lernenden. Die Verinnerlichung nimmt zu, wenn die konkreten Handlungen und das anschauliche Lernen immer stärker durch abstrakte Lernprozesse abgelöst werden. Die Dimension der Aneignung umfasst mit anderen Worten das subjektive Entwicklungs-, Begriffs- oder Abstraktionsniveau des Lernenden. Für die Diagnostik ergeben sich dadurch drei Untersuchungsschwerpunkte:

- „Die Frage nach der Beherrschung des Lerngegenstandes: Ausgehend von einer Beschreibung der Elemente einer Sachstruktur soll der Stand der aktuellen Leistung eruiert werden mit dem Ziel, die nächste Entwicklungsstufe zu finden.
- Die Analyse der Bedürfnisstruktur: Als Ausgangspunkt für motivierendes Handeln wird die Frage nach dem Bezug des Schülers zum Lerngegenstand gestellt.
- Die Inangangsetzung von Aneignungsprozessen: Ausgehend vom entsprechenden Entwicklungsstand, also von den Lernvoraussetzungen, werden Lernprozesse beschrieben“ (Bundschuh, 1994, S. 97).

Das Ergebnis der Auseinandersetzung mit dem Lerngegenstand kann an Schülerarbeiten und -produktionen abgelesen werden. Nebst Angaben über den erfolgten Lernprozess enthält die Analyse je nach Aufgabenstellung Informationen über den Stand der verschiedenen Entwicklungsbereiche und die psychische Befindlichkeit des Schülers oder Schülerin. Die Palette der zu analysierenden Schülerarbeiten ist gross und reicht von Klassenarbeiten, Hausaufgaben, Werkgestalten und bildnerischen Erzeugnissen bis hin zu Präsentationen und Projektergebnissen. In der Praxis zeigt sich, dass solche Analysen für die Entwicklung von diagnostischen Arbeitshypothesen eine grosse Bedeutung besitzen.

Von besonderem Interesse für diagnostische Überlegungen sind dabei Fehler, die bei der Erarbeitung eines Lerngegenstandes entstehen. Es handelt sich dabei um subjektive Lösungsstrategien, die dem Schüler zur gegebenen Zeit zur Verfügung stehen und von denen er annimmt, dass sie richtig sind: „Fehler sind deshalb nicht nur falsch, sie sind zugleich Ergebnis geistiger Prozesse des Kindes und Indiz für seine ‚innere Realität‘, seine geistige Arbeit, seine Vorstellungen und sein Vorwissen“ (Wiater, 2004, S. 6).

Auf den analogen Umstand weist auch der aus der Sonderpädagogik stammende Begriff der *Förderdiagnostik* hin. Förderdiagnostik sucht selbstredend einen engen Bezug von Diagnostik und Förderprozessen (vgl. Buholzer, 2006; 2010). So betonen etwa Eberwein und Knauer, dass Förderdiagnostik immer auf Lernprozesse und damit auf die Veränderung einer schulischen Prob-

lemsituation hin gerichtet sei. Durch die Erkenntnisse der Förderdiagnostik wird der Rahmen für die spätere Förderung abgesteckt und macht Strukturen für das weitere pädagogische Handeln sichtbar. Im lernprozessbegleitenden Teil der Förderdiagnostik besteht die Aufgabe darin, „... die Auswirkungen der vom Lehrer getroffenen pädagogischen Massnahmen, sein Förderkonzept im Hinblick auf die Zielsetzungen zu kontrollieren und gegebenenfalls zu korrigieren bzw. zu revidieren und neue Hypothesen zu formulieren“ (Eberwein & Knauer, 2003, S. 10). Förderdiagnostische Angaben liefern die Voraussetzung, den Unterricht oder die Förderung an die Lernbedürfnisse der Schülerinnen und Schüler anzupassen. Der Förderdiagnostik geht es also nicht darum, Grundlagen für Zuweisungsentscheide, sondern vielmehr konkrete Hinweise zur Gestaltung einer bedürfnisorientierten pädagogischen Arbeit zu liefern (vgl. Mand, 1998, S. 43). Darin eingeschlossen sind sowohl Informationen für die Förder- und Unterrichtsgestaltung wie auch für die beratende Tätigkeit (vgl. Mutzeck, 2004, S. 15).

Wie sinnvoll eine enge Verknüpfung von Diagnostik und Förderung ist, lässt sich empirisch untermauern. So stellt in der Studie von Schrader (1989) die Diagnosekompetenz für sich genommen kein hinreichendes Merkmal dar, die Leistung von Schülerinnen und Schülern zu verbessern. Die diagnostisch kompetente Lehrperson ist nach dem Ergebnis dieser Studie nicht gleichzeitig auch die erfolgreiche Lehrperson. Die Diagnosekompetenz besitzt hingegen eine ausgeprägte moderierende Wirkung. Die Lernentwicklung bei Schülerinnen und Schülern ist dann besonders hoch, wenn zutreffende Lerndiagnosen mit unterrichtlichem Handeln verknüpft werden. Erfolgversprechend sind dabei Strukturierungshilfen und individuelle fachliche Unterstützung. „Diese moderierende Wirkung diagnostischer Kompetenz wurde dahingehend interpretiert, dass diagnostische Kompetenz die Wirksamkeit bestimmter Massnahmen und Verhaltensweisen optimiert“ (Schrader, 1989, S. 228). Dies heisst zusammenfassend, dass die diagnostische Kompetenz nur dann wirksam werden kann, „...wenn der Lehrer seine Diagnosefähigkeit in ein geeignetes Unterrichtsverhalten umsetzt“ (Schrader, 1989, S. 165).

Hinweise auf Wirkungen können auch Studien zum adaptiven Unterricht entnommen werden. In einem adaptiven Unterricht ist es Aufgabe der Lehrkraft, „unter bestmöglicher Berücksichtigung der inhaltlichen Ziele des Unterrichts (Sachkompetenz), der Wissens- und Lernvoraussetzungen der Schülerinnen und Schüler und der situativen Aspekte des Lernens (diagnostische Kompetenz) [und] der optimalen Orchestrierung der Lernsituation (Klassenmanagement) den Unterricht so zu gestalten, dass möglichst viele Schülerinnen und Schüler optimal lernen und verstehen“ (Beck et al., 2008, S. 224). Gemäss der Studie von Beck et al. (2008) wirkt sich die adaptive Lehrkompetenz teilweise positiv auf den Lernfortschritt von Schülerinnen und Schülern in heterogen zusammengesetzten Schulklassen aus. Der Einfluss der adaptiven Lehrkompetenz kann auf die in der Untersuchung verwendete Unterrichtseinheit mit klar festgelegten Lernzielen und beschränkter Lernzeit nachgewiesen werden. Ein Transfer auf unspezifischere naturwissenschaftliche Kompetenzen in heterogenen Schulklassen konnte jedoch nicht belegt werden (vgl. 2008, S. 125-129).

## 2.4. Konsequenzen für die Ausbildung von Lehrpersonen

### 2.4.1. Kompetenzen und Kompetenzaufbau

Die erfolgreiche Anwendung einer akkuraten und lernförderlichen Diagnostik erfordert von Lehrpersonen entsprechende Kompetenzen. Kompetenzen sind Fähigkeiten und Fertigkeiten und bilden somit die Voraussetzung, damit Lernprozesse professionell initiiert und begleitet werden können. Im Unterrichtsalltag sind die einzelnen Kompetenzen einer Lehrperson nur schwer voneinander abzugrenzen, da sie meist im Verbund zum Einsatz kommen. Vielmehr ist die Rede von Kompetenzbündeln oder Kompetenzprofilen (vgl. Oser, Curcio & Düggele, 2007, S. 14).

Etymologisch kann Competentia als „Zusammentreffen“ oder „Zuständigkeit“ übersetzt werden. Competere heisst demzufolge „zusammenlangen, zusammentreffen, entsprechen oder zukommen“. Übertragen bedeutet dies, wenn die Anforderungen an eine Situation mit dem Repertoire von Fähigkeiten einer Person „zusammentreffen“, ist die Person kompetent, das Problem oder die Aufgabe zu bewältigen.

Allgemein bezeichnet Kompetenz nach Meyer und Klapper (1996) die Fähigkeit, eine Leistung in immer wieder neuen Situationen zielbezogen, effektiv und unter Beachtung ethischer Regeln zu erbringen. Besitzt also eine Person Kompetenz, „... so kann sie etwas, ist handlungsfähig und übernimmt für sich und andere Verantwortung. Sie besitzt die Kompetenz, so tätig zu werden, dass sie eine Absicht, ein Ziel oder einen Zweck unter Beachtung von Handlungsprinzipien, Werten, Normen und Regeln, mit Bezug auf konkrete, die jeweilige Handlungssituation bestimmende Bedingungen, zu erreichen vermag“ (1996, S. 31).

Nach Criblez, Oelkers, Reusser et al. (2009) bündeln Kompetenzen verschiedene Fähigkeiten und Fertigkeiten, um komplexe Situationen (wie sie im Unterricht immer wieder vorzufinden sind) erfolgreich zu bewältigen und zu gestalten. Dies wird durch das Mobilisieren von Ressourcen möglich. Zu diesen Ressourcen zählen wir:

- Wissen (z.B. Kenntnisse, die in einer konkreten Situation abgerufen werden können)
- Fertigkeiten (z.B. Verfahrensweisen und Techniken, Routinen)
- Fähigkeiten (z.B. Empathie, Solidarität, Kommunikationsfähigkeit usw.)
- Ressourcen des Umfeldes (z.B. Infrastruktur, persönliches Netzwerk).

Ziel kompetenten Handelns ist es, „anstehende Aufgaben oder Probleme zielorientiert und verantwortungsbewusst zu lösen, die Lösungen zu reflektieren und zu bewerten und das eigene Repertoire an Handlungsmustern weiterzuentwickeln“ (Frey, 2006, S. 31). Wie oben ausgeführt, bilden Kompetenzen, die Disposition, schulische Situationen erfolgreich zu bewältigen. Kompetenzen selber lassen sich daher auch nicht direkt beobachten, sichtbar ist nur die auf Grundlage der Kompetenz realisierte Performanz.

Der Aufbau von Kompetenzen darf wohl nicht mit der naiven Vorstellungen einer linearen Anwendung zuvor erworbenen theoretischen Wissens auf nachfolgende Probleme in der beruflichen Praxis erklärt werden. Vielmehr wird davon ausgegangen, dass (theoretisches) Wissen und (professionelles) Wissen zusammenfallen: „Dem zufolge verlangt die Ausbildung professioneller Kompetenzen (=berufliches Können) im Prinzip die synchrone Kombination (...) von systematischen und situierten Lernprozessen, über die sich erst die erforderlichen unterschiedlichen Lerntransfers (vertikal, horizontal, lateral) realisieren lassen (...). (Lersch, 2006, S. 43). Gängige Modelle zum Kompetenzaufbau gehen von Kompetenzstufen oder -niveaus aus. So schlagen zum Beispiel Meyer und Klapper ein Stufungsmodell vor, das von einer zunehmenden Selbständigkeit



des Denkens, Fühlens und Handelns ausgeht. Diese zunehmende Selbständigkeit drückt sich in folgenden Niveaus aus:

1. „naiv ganzheitliches Nachvollziehen einer Handlung
2. Ausführen einer Handlung nach Vorgabe
3. Ausführung einer Handlung nach Einsicht
4. selbständige Prozesssteuerung und ihre didaktische Reflexion“ (Meyer & Klapper, 2006, S. 91).

Eine Lehrperson auf der höchsten Stufe übernimmt gewissermassen gemeinsam mit dem Lernenden Verantwortung für den Lehr-Lern-Prozess. Sie initiieren selbständig Lehr- und Lernprozesse, steuern ihn nach (selbstgewählten) Kriterien und reflektieren anhand der gesetzten Ziele den Erfolg und Misserfolg.

Über welche Kompetenzen Lehrpersonen zu verfügen haben, darüber besteht Uneinigkeit (vgl. Allemann-Ghionda & Terhart, 2006). Exemplarisch soll hier in gebotener Kürze auf das Tableau der Kompetenzen (Girmes, 2006), die Kompetenzklassen von Frey (2006) und die adaptive Lehrkompetenz (Beck et al., 2008) hingewiesen werden. Im Tableau der Lehrkompetenzen geht Girmes für die Basis-Ausbildung von Lehrerinnen und Lehrern von folgenden Kompetenzen aus: (1) diagnostische Kompetenz, (2) institutionelle Kompetenz, (3) curriculare Kompetenz, (4) methodische Kompetenz, (5) personell-kulturelle Kompetenz, (6) reflexive und evaluative Kompetenz (Girmes, 2006, S. 27). Diese sechs Kompetenzbereiche sind dann wiederum mit insgesamt zehn zugehörigen Aufgaben verknüpft, die im Rahmen der Lehrerinnen- und Lehrerausbildung bearbeitet werden müssen. Der Kompetenzaufbau erfolgt in Verbindung und Relation von theoretischer Klärung und praktischer Erprobung und deren Reflexion.

Frey führt vier Kompetenzklassen auf: Fachkompetenz, Methodenkompetenz, Sozialkompetenz und Personalkompetenz. Die Fachkompetenzklasse umfasst fachspezifische respektive disziplinorientierte Fähigkeitskonzepte. Die Methodenkompetenzklasse ihrerseits betrifft Fähigkeitskonzepte, um innerhalb des definierten Sachbereichs denk- und handlungsfähig zu sein, zum Beispiel, die Fähigkeit, einen Sachverhalt zu analysieren, einen Lernprozess zu strukturieren, über Zusammenhänge zu reflektieren. Die Sozialkompetenzklasse umfasst u.a. Fähigkeiten je nach Situation und Aufgabe in Kooperation mit anderen ein anvisiertes Ziel verantwortungsvoll zu lösen, Konflikte zu lösen und mit anderen zu kommunizieren (2006, S. 33). Die Personalkompetenzklasse umfasst „... Fähigkeitskonzepte, Einstellungen oder Eigenschaften, die benötigt werden, um für sich selbst verantwortlich und motiviert zu handeln“ (2006, S. 33).

Diese breit angelegten beruflichen Kompetenzen werden teilweise auch eingeschränkt, zum Beispiel auf den Bereich des konkreten Unterrichts, wie das Beispiel der adaptiven Lehrkompetenz zeigt. Hinter dieser Lehrkompetenz stehen die Sachkompetenz, die Diagnosekompetenz, die didaktische Kompetenz und die Klassenführungskompetenz (vgl. Beck et al., 2008):

Bezüglich der Diagnosekompetenz, ist unklar, wie sich diese zusammensetzt und wie sie aufgebaut wird. Die Annahme, dass die diagnostische Kompetenz eine allgemeine, fächerübergreifende Fähigkeit bildet, ist dabei auf Grund bisheriger Studien (siehe Kapitel 2.2) nicht haltbar. Zur Entwicklung der diagnostischen Kompetenz ist dabei, wie in den vorangehenden Kapiteln aufgezeigt, genauso wenig bekannt. Wie dargelegt, liegt die Vermutung jedoch nahe, dass das fachdidaktische Wissen mit der Leistungsdiagnostik in Zusammenhang steht (vgl. Baumert & Kunter, 2006, 481ff.). So wird die diagnostische Kompetenz als Kompetenzfacette zusammen mit dem Erklärungswissen unter dem fachdidaktischen Wissen eingeordnet. Das fachdidaktische Wissen

kann zu einer der fünf Kompetenzbereiche des Professionswissens gezählt werden (vgl. Rauin, 2007, S. 61).

Wann bei der Schlüsselqualifikation der pädagogischen Diagnose von Kompetenz gesprochen werden kann, umschreibt Girmes (2006) wie folgt: „Die ausgebildete Fähigkeit (...) Wissen über Lernprozesse, Entwicklungszusammenhänge und auf das Verfügen über Wege und Methoden, dieses Wissen für die einlassende Teilnahme, aber auch für die erkennende Diagnose der Möglichkeiten und Hemmnisse bei Adressaten zu nutzen. Diese Fähigkeit nenne ich diagnostische Kompetenz, wenn sie die pädagogische Kommunikation vorzubereiten und angemessen zu gestalten hilft (...).“ (Girmes, 2006, S. 15).

Im Gegensatz dazu schlägt Spinath auf Grund ihrer Untersuchungsergebnisse gar vor, auf den Begriff der diagnostischen Kompetenz zu verzichten. Mit ihrer Untersuchung widerlegt sie das Vorhandensein eines Konstrukts diagnostischer Kompetenz. Von der diagnostischen Kompetenz einer Lehrperson kann gar nicht die Rede sein, da die Akkuratheit einer Einschätzung je nach Schülermerkmal variiert. „Die Befunde sprechen insofern gegen die Annahme eines Konstrukts diagnostischer Kompetenzen als a) verschiedene Akkuratheitskomponenten innerhalb eines Merkmals und b) gleiche Komponenten über Merkmale hinweg entweder keine bedeutsamen oder nur punktuell positive Korrelationen aufweisen“ (Spinath, 2005, S. 93). Der Begriff der diagnostischen Kompetenzen wird als unscharf und irreführend kritisiert und konnte bislang als übergreifende Fähigkeit empirisch nicht nachgewiesen werden (vgl. Spinath, 2004, S. 17).

Allgemein kann gesagt werden, dass die Lehrpersonen professionelle Kompetenzen durch jahrelange praktische Erfahrung und deren theoriegeleiteter Reflexion ausbilden. Die Wissensgrundlage für die Ausbildung der Kompetenzen wurde dabei 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). Wie sich die einzelnen Kompetenzbereiche im Laufe der Berufsbiographie ausbilden, ist jedoch unklar. In dieser Hinsicht besteht sowohl theoretischer als auch empirischer Forschungsbedarf (vgl. Baumert & Kunter, 2006, S. 505). Kaum untersucht ist auch die Entwicklung der Kompetenzen im Laufe des Lehrstudiums (vgl. Baer, Dörr, Guldemann, Kocher, Larcher, Müller & Wyss, 2008, S. 259). Erfahrene Lehrpersonen wissen dabei in der Regel nicht, wie und warum sie etwas tun, welche handlungsleitenden Kognitionen ihrem Tun zu Grunde liegen. (vgl. Reusser & Messner, 2002, S. 295)

#### **2.4.2. Implikationen für die Aus- und Weiterbildung von Lehrkräften**

Aus den Erfordernissen an eine hohe Diagnosekompetenz sind Konsequenzen für die Aus- und Weiterbildung von Lehrerinnen und Lehrern abzuleiten. Lehrerinnen und Lehrer müssen – unabhängig ihres beruflichen Tätigkeitsfeldes – im Aufbau von Basisqualifikationen im diagnostischen Bereich angeleitet und unterstützt werden. Darüber hinaus sind spezifische Qualifikationen für eine funktionale Ausdifferenzierung erforderlich. Dazu zählen wir beispielsweise Fachpersonen für individuelle Förderung (vgl. Landeskompetenzzentrum für Individuelle Förderung), Schulische Heilpädagoginnen und Heilpädagogen oder Lehrpersonen mit weiteren spezifischen Funktionen. Für diese Fachpersonen wurden je spezifische diagnostische Konzepte (z.B. die Förderdiagnostik) ausgearbeitet und entsprechende Hilfsmittel bereitgestellt. Diese spezifischen Qualifikationen werden in den weiteren Ausführungen nicht weiterverfolgt, vielmehr steht die Aneignung von Basisqualifikationen im Vordergrund. Im Zentrum steht dabei die Urteilsbildung, welche sich ganz allgemein darin messen lässt, wie gross ihr jeweiliger Beitrag zur Lösung des

Entscheidungsproblems ist. Ein weiteres Kriterium der Urteilsbildung ist die Genauigkeit. Ziel der diagnostischen Kompetenz ist es, ein möglichst gutes Urteil über die Schülermerkmale und ihre Ausprägung bilden zu können, um auf dieser Grundlage, die notwendigen pädagogischen Entscheidungen zu treffen (siehe auch Kapitel 2.2.1 zur Genauigkeit von Lehrerurteilen).

Auf welche Aspekte im Bereich der Diagnosekompetenz zu achten ist, findet sich in der Zusammenstellung der Standards zur Lehrerinnen- und Lehrerbildung von Oser (vgl. 1997a, b). Er fasst in seiner Übersicht unter dem Titel „Schülerunterstützende Beobachtung und Diagnose“ Ausbildungsziele zusammen, welche den Umgang mit kritischen Entwicklungen und auftretenden Problemen betreffen. Der Umgang bezieht sich sowohl auf einzelne Schülerinnen und Schülern wie auch auf das Geschehen im Klassenzimmer. Lehrpersonen sollen in der Lage sein:

- „Zu diagnostizieren, welche Ursachen Misserfolg, Aggression, Ängste, Blockierungen etc. haben und darauf zu reagieren.
- Den entwicklungspsychologischen Stand der Schülerinnen und Schüler in verschiedenen Bereichen (Intelligenz, Sprache, Moral, soziales Verhalten usw.) zu diagnostizieren.
- Nachahmungsprozesse unter den Schülerinnen und Schülern zu beobachten.
- Unterschiedliche Gefährdungen (z.B. Gewalt, Drogen, Selbstmord usw.) in jedem Alter, das ich unterrichte, festzustellen und entsprechend einzugreifen.
- Die Ablösung vom Elternhaus zu verstehen und auf unterschiedliche Ablösungsformen zu reagieren.
- Wie man spezifische Lernschwierigkeiten diagnostizieren kann“ (Oser, 1997a, S. 32).

Kretschmann (2004) hat Erfordernisse und Elemente einer Diagnostik-Ausbildung für Lehrerinnen und Lehrer zusammengestellt. Die Diagnosekompetenz soll zu einer Optimierung pädagogischer Angebote beitragen. Dies ist dann der Fall, wenn folgende Bedingungen erfüllt sind:

- „Die Ausführenden - Lehrerinnen und Lehrer - verfügen über hinreichende Modelle über Ursachen und Verläufe der Entwicklungsprozesse ihrer Klientel; der regulären Entwicklung und alterstypischer Störungen und Gefährdungen.
- Den Ausführenden stehen Diagnoseinstrumente zur Verfügung, die für das Arbeitsfeld und die zu tätigen Aufgaben entwickelt worden sind. So sollten Lehrerinnen und Lehrer, um Lernfortschritte oder Entwicklungsprobleme zu diagnostizieren, nicht darauf angewiesen sein, sich bei Instrumenten zu bedienen, die für die Psychiatrie oder für die psychologische Kindertherapie entwickelt wurden.
- Die Ausführenden sind in der Lage, Diagnoseinstrumente kompetent zu handhaben.
- Die Lehrerinnen und Lehrer verfügen über Förderkompetenz. Sie wissen, welche pädagogischen und welche Fördermassnahmen auf eine diagnostizierte Konstellation folgen müssen und sie sind in der Lage, sie auch auszuführen.
- In den Schulen besteht Zeit und Raum, Präventions- und Förderangebote hinreichend oft, hinreichend intensiv und hinreichend lange vorzuhalten und
- dort, wo die Möglichkeiten von Schule nicht ausreichen, bestehen Möglichkeiten interinstitutioneller Kooperation, in Form von niederschwelliger, rasch realisierbarer und unbürokratischer Zusammenarbeit mit Bereichen wie Jugendhilfe oder anderen unterstützenden Diensten“ (Kretschmann, 2004, S. 125f.).

Nach Buholzer fusst die diagnostische Kompetenz einer Lehrperson auf einem Wissen in viererlei Hinsicht (vgl. Buholzer, 2010). Vor diesem Hintergrund ist zu überlegen, von welchen Voraussetzungen die Herausbildung einer hohen Diagnosekompetenz auszugehen hat:

*Reflexionsfähigkeit.* Wie auch die konkrete Diagnosesituation gestaltet ist, Lehrkräfte sind nie unbeteiligte, objektive Diagnostiker. Vielmehr sind sie Interaktionspartner, die durch ihre Handlungen an diagnostischen Prozessen mitbeteiligt sind. Aus diesem Umstand ergibt sich die Anforderung, sich mit dem Diagnoseprozess reflexiv auseinanderzusetzen. Im Mittelpunkt stehen dabei Fragen, wie die Routinen der diagnostischen Urteilsbildung gestaltet sind und von welchen Grundannahmen die Analysen ausgehen. Der gezielte Einbezug einer Fremdperspektive ist bei der Reflexion von großer Bedeutung, so können blinde Flecken, Urteilstendenzen und mögliche Verzerrungen aufgedeckt und einer Bearbeitung zugänglich gemacht werden. Es liegt auf der Hand, dass durch dieses Vorgehen auch eine Konfrontation mit der eigenen Arbeit stattfindet, denn misslingende Lernprozesse können auch mit den Grenzen eigener methodisch-didaktischer und kommunikativer Fähigkeiten zusammenhängen. Pädagogisches Handeln impliziert daher immer auch die Fähigkeit, sich als Teil des Bedingungsgefüges zu sehen und die sich daraus ergebenden Spannungen reflexiv zu verarbeiten (vgl. Horstkemper, 2006).

*Lerntheoretische und (fach-)didaktische Kenntnisse.* Um die gewonnenen Informationen richtig deuten zu können, sind abgesehen von der Reflexionsfähigkeit auch spezifische Kenntnisse über Lernprozesse und die dahinterliegenden Bedingungen erforderlich. Hierfür sind Modelle von Bedeutung, die Auskunft geben über Ursachen und Verläufe von Lern- und Entwicklungsprozessen sowie der regulären Entwicklung und alterstypischer Störungen und Gefährdungen. Die Analyse des Bedingungsgefüges ist insbesondere für die nachfolgenden Förder- und Unterrichtsprozesse von Bedeutung.

*Wissen über Methoden und Instrumente.* Zur Erfassung der Lernprozesse sind eine vertiefte Kenntnis über das methodische Vorgehen und geeignete diagnostische Instrumente notwendig. Eingeschlossen sind auch Kenntnisse über die Grenzen und die Möglichkeiten der gewählten Methode sowie über die zu erwartenden Probleme bei der Durchführung. Nicht alle Lehrkräfte müssen das Gleiche wissen – wichtig ist jedoch, dass im Schulteam die notwendigen Informationen vorliegen. Für förderorientierte Diagnosezwecke sind nicht nur klassische Lernstandsanalysen zu verwenden, sondern auch Aufzeichnungen von Lehrkräften zum Beispiel in Lehrertagebüchern, Portfolios, Fehleranalysen, Screenings etc.

*Pädagogisch-didaktisches Wissen.* Für die Interpretation und die Fortführung von Lerndiagnosen sind Kenntnisse über das Auslösen und Begleiten von Lernprozessen erforderlich. Zentral dabei ist die Frage, wie ein diagnostizierter Zustand in einen neuen und besseren Zustand überführt werden kann – wobei hier nicht einseitig an die Veränderung der Person gedacht wird, sondern auch an die Modifikation der Bedingungen und der Situation. Lehrkräfte wissen also, welche pädagogischen Konzepte und Fördermassnahmen auf eine diagnostizierte Konstellation folgen müssen. Damit allerdings diagnostische Erkenntnisse ihre Wirkung entfalten können, muss auch der Unterricht so gestaltet sein, dass er auf diagnostische Erkenntnisse überhaupt eingehen kann. Ein Unterricht, in dem alle im Gleichschritt zum selben Zeitpunkt das Gleiche lernen ist unter dieser Perspektive so nicht mehr möglich.

Mit den Ausführungen von Oser (1997a, b), Kretschmann (2004) und Buholzer (2010) wird deutlich, dass die Diagnosekompetenz weit über die (herkömmliche) Leistungsbeurteilung hinausgeht und auf einer Vielzahl von Voraussetzungen beruht. Diese anspruchsvolle Kompetenz kann nur eingebettet in die grundlegenden entwicklungspsychologischen und (fach-) didaktischen Modelle, gelernt und umgesetzt werden. Um diesen Ansprüchen zu genügen, müssen Lehrpersonen das Diagnostizieren lernen (vgl. Hascher, 2003, S. 29). Die Ausbildung der Diagnosefähigkeit des

gegenstandsbezogenen Entwicklungsstandes der Schülerinnen und Schüler wird dabei neben Kompetenzen zur Planung, Gestaltung und Evaluation des Unterrichts in der fachdidaktischen Ausbildung angesiedelt (vgl. Reusser & Messner, 2002, S. 292).

Zum Aufbau von diagnostischer Kompetenz sollen die Studierenden der Lehrerbildung nebst eigenen Lernerfahrungen im Unterricht, in Werkstätten und Projekten insbesondere in der berufspraktischen Tätigkeit Lerndiagnosen erstellen und zusammen mit der Praktikumsleiterin / dem Praktikumsleiter evaluieren. Wichtig erscheint dabei die Reflexion der subjektiven Wahrnehmungs- und Beurteilungsprozesse zu sein: Woraus erschliesse ich, ob eine Schülerin den Auftrag verstanden statt. Woran erkenne ich, dass ein Schüler bei einer Stillarbeit nicht mehr weiterkommt? Diagnosekompetenz erfordert „einen Prozess des Umdenkens und des Lernens, sowohl von Schüler/innen als auch von Lehrpersonen. Studierende der Lehrerbildung sollten früh mit den Möglichkeiten einer ‚mehrperspektivischen Diagnose‘ ... vertraut gemacht werden“ (Hascher, 2003, S. 30).

Günther & Brunstein (2003) gehen noch weiter und bieten für Studierende der Lehrerbildung ein einwöchiges Psychodiagnostisches Praktikum an. Vorbedingung für das Praktikum ist der erfolgreiche Abschluss von Diagnostikvorlesung und -seminar. Während des Praktikums arbeiten zwei bis drei Studierende mit einer Klasse zusammen. „Im Vordergrund der einwöchigen Praktikumstätigkeit steht die fallbezogene Anwendung und Reflexion diagnostischer Methoden zur Erfassung und Beurteilung ausgewählter Merkmale des Schülerverhaltens im kognitiven, motivationalen und sozialen Bereich. Die Studierenden haben die Aufgabe, eine Informationssammlung zu ausgewählten Verhaltensbereichen ‚ihrer‘ BeobachtungsschülerInnen anzulegen, aus denen nachfolgend eine Zusammenfassung der dabei gewonnenen diagnostischen Informationen anzulegen ist“ (2003, S. 23). Zum Einsatz kommen unterschiedliche Methoden wie Kurzzeitbeobachtungen und Situationsbeobachtungen in den Pausen, Informationsgespräche mit Lehrpersonen und Mitschülerinnen und -schülern sowie ein psychodiagnostisches Explorationsgespräch mit dem Schüler / der Schülerin. Je nach Fragestellung können noch weitere methodische Verfahrensweisen zum Einsatz gelangen. In einem Praktikumsbericht werden die Ergebnisse dokumentiert und die diagnostische Vorgehensweise kritisch reflektiert (vgl. Günther & Brunstein, 2003, S. 20-24).

Das Grundgerüst der Ausbildung an der PHZ Luzern wird mit Hilfe von Handlungskompetenzen umschrieben. Diese Kompetenzen bilden den Referenzrahmen für die Planung der Ausbildungsmodule<sup>2</sup> von Lehrpersonen des Kindergartens, der Primarschule und der Sekundarstufe I. Sie sind auf einer Konkretheitsebene definiert, die einen überprüfbaren Lernfortschritt der Studierenden für ein bis zwei Kompetenzen in jedem Ausbildungsmodul zulassen. Insgesamt sind es mehr als 20 Handlungskompetenzen, die während der Ausbildung aufgebaut werden. Welche Handlungskompetenzen sich für eine Grundausbildung eignen, muss anhand von vier Kriterien begründet werden können, die sich auf die Definition von Standards nach Oser (vgl. 1997a, b) beziehen:

- Theoretische Begründungen: Lässt sich der Standard aus einer fachbezogenen Theorie oder einer Theorie des Lernens, der Entwicklung oder der Bildung von Schülerinnen und Schülern ableiten? Möglich für Begründungen sind auch Bezüge zu Theorien über das Lehrerhandeln.

---

<sup>2</sup> Die folgenden Beschreibungen zu den Zielen und Inhalten des Studiums beziehen sich auf den Ausbildungsgang, wie er von den an der Untersuchung beteiligten Studierenden durchlaufen wurde (Stand 2004). Mittlerweile wurden die Handlungskompetenzen weiter ausdifferenziert und auch die Studienangebote teilweise geändert.

- Empirisches Wissen: Belegen empirische Studien die Wirksamkeit dieser Handlungskompetenz auf das Lernen von Schülerinnen und Schülern?
- Schulisches Erfahrungswissen: Wird diese Handlungskompetenz von Lehrpersonen als im Schulalltag bedeutsam bezeichnet? Gibt es dazu Erfahrungsberichte, Fallstudien, Beobachtungen?
- Unterschiedliche Qualitäten und Entwicklungsstufen: Kann die Handlungskompetenz in Stufen gelernt werden? Gibt es ein Wissen über Entwicklungsverläufe im Lehrerlernen/ Lehrerbio-graphie für diese Handlungskompetenz? Weiss man etwas über unterschiedlich gute Umsetzungen der Kompetenz (z.B. Experte – Novize)?

Im vorliegenden Projekt wird die Handlungskompetenz: „Ich kann die Lernvoraussetzungen der Kinder und Jugendlichen, ihre Interaktionen und die Einflüsse von Lernsituationen und von am Lernprozess beteiligten Systemen differenziert erfassen und für meinen Unterricht nutzen“ ausgewählt.

Wir fassen diese Fähigkeit in der Folge als diagnostische Kompetenz auf. Es ist unbestritten, dass diese Kompetenz für die Planung und Gestaltung eines lernwirksamen Unterrichts unerlässlich ist. Die diagnostische Kompetenz darf wohl neben der Sachkompetenz, der Klassenführungs-kompetenz und der didaktischen Kompetenz als eine der vier zentralen Kompetenzen, über welche eine Lehrpersonen verfügen muss, bezeichnet werden (vgl. Beck et al. 2008).

Im Rahmen der Ausbildung liefern Inhalte der Bildungs- und Sozialwissenschaften, der Fachdidaktiken wie auch die reflektierten Erfahrungen in der Schulpraxis ihre entsprechenden Beiträge zum Aufbau dieser Handlungskompetenz. Allgemeine Grundlagen werden bereits im stufenübergreifenden Grundjahr gelegt. Spezifische diagnostische Fähigkeiten werden ab dem 3. Semester im Rahmen der Stufenausbildungen erworben.

### 3. Anlage der empirischen Untersuchung

Das Vorgehen bei der eigenen empirischen Untersuchung wird im folgenden Kapitel beschrieben.

#### 3.1. Eingrenzung des Forschungsthemas

Mit dem vorliegenden Forschungsprojekt soll der Frage nachgegangen werden, inwiefern es in der Ausbildung gelingt, die Studierenden in ihrer diagnostischen Kompetenz weiter zu bringen. Dies wäre dann der Fall, wenn ein Kompetenzzuwachs im Laufe des Studiums bei der untersuchten Stichprobe zu verzeichnen ist. Die diagnostische Kompetenz als Ganzes zu erfassen ist kaum möglich. Für die Untersuchung wird die diagnostische Kompetenz eingeschränkt als die Fähigkeit:

- a) den Lernstand von zwei Schülern in einem ausgewählten Lernbereich anhand einer bestimmten Aufgabenstellung präzise einzuschätzen
- b) die beiden Schüler in einer Rückmeldung darüber zu informieren und
- c) auf der Grundlage des festgestellten Lernstandes eine angemessene Weiterarbeit zu skizzieren.

Die einzelnen Elemente der Untersuchung, die Fragestellung, das Forschungsdesign, die Stichprobe, das Untersuchungsinstrument, das Kategoriensystem und das Vorgehen bei der Auswertung werden im Folgenden genauer beschrieben.

#### 3.2. Fragestellung

Die vormals erwähnte Hauptfragestellung lässt sich in weitere Unterfragen aufteilen.

##### Hauptfragestellung

Wie entwickelt sich die Diagnosekompetenz der PH-Studierenden im Verlaufe ihrer Ausbildung?

Wie im vorangehenden Kapitel ausgeführt bezieht sich die Diagnosekompetenz auf die Aspekte (a) Genauigkeit des Urteils, (b) Einbettung in eine Rückmeldung, (c) Angemessene pädagogische Intervention

##### Unterfragen

- Studiengang: Zeigen sich zwischen den Studierenden der einzelnen Studiengänge Unterschiede bezüglich ihrer Diagnosekompetenz?
- Fach Deutsch: Zeigen sich bei den Sek I Studierenden bezüglich ihrer Diagnosekompetenz Unterschiede zwischen denjenigen, die Deutsch als eines ihrer Studienfächer gewählt haben im Vergleich zu denjenigen, die Deutsch nicht als Studienfach haben?
- Reguläres / Aufbau-Studium: Zeigen sich bei den Sek I Studierenden bezüglich ihrer Diagnosekompetenz Unterschiede zwischen denjenigen des regulären Studiengangs und des Aufbaustudiums?

### 3.3. Forschungsdesign

Um die vorangehend aufgezeigten Fragestellungen zu beantworten, muss die Diagnosekompetenz der gleichen Studierenden mehrmals während ihrer Ausbildung gemessen werden. Das für die Untersuchung gewählte Paneldesign beinhaltet die Messung der immer gleichen Variablen bei der jeweils identischen Stichprobe zu mehreren Zeitpunkten. Die Veränderungsdynamik auf der individuellen Ebene kann so nachvollzogen werden. Ein Paneldesign bietet der konstanten Untersuchungseinheit wegen den Vorteil der Reduzierung der Stichprobenfehler für die einzelnen Messungen. Der Wegfall einzelner Probanden bei den jeweiligen Messzeitpunkten führt jedoch zu einem Schrumpfen der Stichprobe über den gesamten Untersuchungszeitraum hinweg (vgl. Diekmann, 1999, S. 266ff.). Die Panelmortalität konnte im vorliegenden Fall durch die Anbindung der Studierenden an die PHZ Luzern tief gehalten werden.

In Abbildung 2 ist das Forschungsdesign mit den einzelnen Panelwellen aufgeführt. Für den echten Längsschnitt wird die Kompetenzmessung bei den Studierenden der Diplomstudiengänge Kindergarten/Unterstufe, Primarstufe, Sekundarstufe I insgesamt drei Mal durchgeführt, zu Beginn, in der Mitte und am Ende ihres Studiums. Das Erhebungsinstrument kommt an den jeweiligen drei Messzeitpunkten zum Einsatz. Da das Studium für die Lehrpersonen der Sekundarstufe I länger dauert als dasjenige für die Kindergarten/Unterstufe und für die Primarstufe, wird die dritte Messung für die Studiengänge KU / PS und Sek I zeitlich versetzt durchgeführt. Für die Berechnung des Längsschnitts werden lediglich die Vergleichswerte der Schülerin A und der Schülerin B mit einbezogen.

	<b>1. Messung</b> 2003 1. Semester	<b>2. Messung</b> 2005 4. Semester	<b>3. Messung</b> 2006 6. Semester
<b>Studierende</b> <b>KU und PS</b>	Beurteilung Schülerin / Schüler <b>A und B</b>	Beurteilung Schülerin / Schüler <b>A und D</b>	Beurteilung Schülerin / Schüler <b>B und F</b>
			<b>3. Messung</b> 2007 9. Semester
<b>Studierende</b> <b>Sek I</b>	Beurteilung Schülerin / Schüler <b>A und B</b>	Beurteilung Schülerin / Schüler <b>A und D</b>	Beurteilung Schülerin / Schüler <b>B und G</b>

Abbildung 2: Forschungsdesign

### 3.4. Stichprobe

Von den insgesamt 211 für das Studienjahr 2003/2004 eingeschriebenen Studierenden konnte bei einer repräsentativen Stichprobe von 121 Studierenden mindestens einmal die diagnostische Kompetenz erfasst werden.

Nach Studiengang aufgeteilt, ergibt sich die folgende Verteilung:

6 Studierende der Kindergarten/Unterstufe

34 Studierende der Primarstufe

56 Studierende des regulären Studiengangs Sekundarstufe I

25 Studierende des Aufbaustudiengangs Sekundarstufe I

Die 25 Studierenden des Aufbaustudiengangs Sekundarstufe I wurden dabei nur zum ersten und zweiten Messzeitpunkt befragt.

Über die drei Messzeitpunkte hinweg mussten Ausfälle in Kauf genommen werden, die durch Studienabbrüche, Beurlaubungen, Wiederholungen des Studienjahrs oder Nichterscheinen an den jeweiligen Terminen zustande kamen.

### 3.5. Untersuchungsinstrument

In der vorliegenden Untersuchung wird die diagnostische Kompetenz eingeschränkt als die Fähigkeit, (a) eine mündliche Sprachproduktion unter Beachtung der Ziele in der Aufgabenstellung angemessen einzuschätzen, (b) rückzumelden und (c) einen angemessenen Vorschlag zur Weiterarbeit ableiten zu können.

Die Studierenden diagnostizieren anhand einer videographierten Schülerproduktion den Lernstand einer Primarschülerin, eines Primarschülers der 5. Klasse. Bei der Schülerproduktion handelt es sich um eine etwa zweiminütige Radioreportage über ein Strandbad, welche die Schülerinnen und Schüler anhand einer detaillierten Lernaufgabe erstellt haben. Zufällig ausgewählte Schülerinnen und Schüler wurden während der Präsentation ihrer Radioreportage einzeln videographiert. Es wird angenommen, dass diese Schülerproduktion aus dem Fach Deutsch auch für die Sek I Studierenden ohne Studienfach Deutsch verständlich und lösbar ist. Die für die Untersuchung ausgewählten Schülerproduktionen unterscheiden sich nicht bezüglich ihres Schwierigkeitsgrads.

Zur Beurteilung der jeweiligen Schülerproduktion füllen die PH-Studierenden einen Fragebogen mit den vier folgenden offenen Fragen aus.

- 1.) Was stellen Sie fest in Bezug auf den Lernstand der Schülerin X?
  - a) aufgrund der Notizen
  - b) aufgrund der beobachteten Videosequenz
- 2.) Was würden Sie dieser Schülerin rückmelden?
- 3.) Wie würden Sie mit dieser Schülerin weiterarbeiten?

Des Weiteren stehen den Studierenden der ausführliche Schüler-Arbeitsauftrag mit Lernziel, Beurteilungskriterien und einer Illustration des Strandbades und die Notizen der Schülerin / des Schülers zur Verfügung (siehe Anhang Seite 47). Die Einschätzungen der Studierenden werden anschliessend mit den Kriterien eines Expertenurteils verglichen. Zur Erstellung des Expertenurteils wurde eine PHZ-interne Fachjury gebildet, zusammengesetzt aus Vertretern der Zielstufe, der Fachdidaktik Deutsch sowie der Bildungs- und Sozialwissenschaften. Die Expertenrunde wurde aufgefordert, unter den gleichen Bedingungen die jeweiligen Schülerproduktionen einzuschätzen. Das Expertenurteil bildet die Grundlage für die Erstellung des Kategoriensystems.

### 3.6. Kategoriensystem

Die mit dem Untersuchungsinstrument erhobene diagnostische Kompetenz (Genauigkeit, Rückmeldung, Weiterarbeit) kann in einem ersten Schritt in eine formale und eine inhaltliche Dimension aufgeteilt werden.

Die formale Dimension setzt sich dabei aus den folgenden Kategorien zusammen:

- **Differenzierung Lernstand:** Präzise Beschreibung von inhaltlichen Aspekten unter Berücksichtigung des Situationsbezugs (Kamera usw.).
- **Interpretation Lernstand:** Belegte und sorgfältig/differenziert formulierte Beurteilungen von Tatsachenbeschreibungen.
- **Fokus Rückmeldung:** Die Rückmeldungen beziehen sich auf die Aufgabenstellung.
- **Fokus Weiterarbeit:** Die Vorschläge zur Weiterarbeit liegen innerhalb der Aufgabenstellung. Es werden Förderziele mit den jeweils geeigneten Massnahmen aufgezeigt.

Die inhaltliche Dimension setzt sich aus sechs Kategorien zusammen, wobei sich die ersten vier Kategorien je nach der zu beurteilenden Schülerproduktion A oder B unterscheiden:

- **Beurteilung Wortschatz:** Aspekte zum verwendeten Wortschatz, unterschiedlich je nach Schülerproduktion A oder B.
- **Beurteilung Aufbau:** Aspekte zum Aufbau der Radioreportage, unterschiedlich je nach Schülerproduktion A oder B.
- **Beurteilung Stimmführung:** Aspekte zur Stimmführung, unterschiedlich je nach Schülerproduktion A oder B.
- **Beurteilung Sprache:** Aspekte zur (Aus-)sprache, unterschiedlich je nach Schülerproduktion A oder B.
- **Angemessenheit Weiterarbeit:** Die vorgeschlagene Weiterarbeit besteht aus einer Diagnosefortsetzung, einer Wiederholung der Übung an einem Analogbeispiel, dem Training einer Teilfertigkeit und/oder der Veränderung der Lernsituation.

Der folgende Auszug aus dem Kategoriensystem zeigt beispielhaft die Operationalisierung der Kategorie Beurteilung Wortschatz auf:

Tabelle 2: Kategoriensystem: Bsp. Beurteilung des Wortschatzes von Schülerin A

Beurteilung Wortschatz	- Hat drei verschiedene Szenen beschrieben (gemäss Aufgabenstellung)
	- Eine / zwei Personen (Herr Lehmann / Herr Meier) steht /stehen im Zentrum (gemäss Aufgabenstellung)
	- abwechslungsreiche und treffende Wortwahl (Adjektive und Verben)
	- phantasievolle und kreative Beschreibung der Strandszene
	- hat die Stimmung nicht beschrieben oder nur rudimentär/knapp (gemäss Aufgabenstellung)
	- macht eine Pointe („Plöffler trauen nie“) / Witzige Formulierung

Die Anzahl der auf der inhaltlichen Ebene erfassten Kriterien hängt von der jeweils zu beurteilenden Schülerproduktion ab. So werden beispielsweise bei der Einschätzung der Stimmführung bei der Schülerin A sechs Aspekte unterschieden, während bei der Schülerin B dieses Merkmal mit sieben Aspekten (Stimmführung am Anfang gut jedoch Abfall gegen Schluss, verkrampt sich, verständliche Artikulation, Pausen eher zufällig, häufige Stockungen, klammert sich an No-

tizen, angenehme Lautstärke) erfasst wird. Die Anzahl der Kriterien pro Schülerproduktion wurde von der Expertenrunde vorgegeben und entspricht den jeweils für diese Dimension als relevant angesehenen Punkten.

### 3.7. Auswertung

#### 3.7.1. Inhaltsanalyse

Die Generierung der quantitativen Daten aus den offenen Antworten der Studierenden erfolgt mit einer quantitativen Inhaltsanalyse. Dabei wird der Textcorpus hinsichtlich bestimmter Aspekte quantifiziert. Mit der Inhaltsanalyse werden einzelne Merkmale von Texten erfasst, indem die Textstellen in Kategorien, die Operationalisierungen der interessierenden Merkmale darstellen, eingeordnet werden. Die Häufigkeiten in den einzelnen Kategorien geben Auskunft über die Merkmalsausprägungen des untersuchten Textes (vgl. Früh, 2001, S. 39f.).

Im konkreten Fall heisst dies für die Auswertung der vier offenen Fragen, dass die Antworttexte nach dem bereits im vorangehenden Kapitel aufgezeigten Kategoriensystem kategorisiert werden. Es wird dabei beurteilt, inwieweit die Einschätzungen des jeweiligen Probanden mit den von den Experten vorgegebenen Aspekten übereinstimmen. Dabei wird erfasst, ob der jeweilige Aspekt ganz, teilweise oder gar nicht in der Beurteilung der oder des Studierenden vorkommt. Um diese jeweiligen Kategorisierungen quantifizieren zu können, wurden Punkte verteilt (Aspekt vorhanden = 2 Punkte, teilweise vorhanden = 1 Punkt, nicht vorhanden = 0 Punkte).

#### 3.7.2. Statistische Auswertung

Mittelwertvergleiche erlauben einen Vergleich der einzelnen Messzeitpunkte miteinander. Um mit abhängigen Stichproben rechnen zu können, werden jeweils die an zwei Zeitpunkten gemessenen identischen Schülerproduktionen miteinander verglichen. Wie in Abbildung 2 aufgeführt, wurden die Beurteilungen der Radioreportagen von Schülerin A und Schüler B jeweils zwei Mal erfasst. Schülerin A beim 1. und 2. Messzeitpunkt und Schülerin B beim 1. und 3. Messzeitpunkt. Die Veränderungen zwischen den zwei Messzeitpunkten in der Beurteilung der jeweiligen Schülerproduktion werden mit dem Wilcoxon-Test (nichtparametrischer Test für abhängige Stichproben) berechnet.

Um die Unterfrage nach dem Einfluss des Vorhandenseins des Studienfachs Deutsch auf die Diagnoseleistung der Studierenden der Sek I beurteilen zu können, wurde eine neue Variable mit den Differenzen der Werte zwischen den jeweiligen Messzeitpunkten berechnet. Ausgehend von dieser Variablen wurden zwischen den gebildeten Gruppen Mann-Whitney-U-Tests durchgeführt. Der Einfluss des Faktors Deutsch/ Nicht Deutsch als Studienfach kann so berechnet werden. Das Gleiche gilt für die Berechnung des Einflusses des Studiengangs. Mittels des gleichen Analyseverfahrens soll untersucht werden, ob sich die Studierenden im regulären Masterstudiengang Sekundarstufe I von denjenigen im Masterstudiengang Sekundarstufe I für Personen mit einem Lehrdiplom unterscheiden. Wo es möglich war, das Fach Deutsch und den Studiengang zu vier Gruppen zu kombinieren, wurden die Unterschiede zwischen den Gruppen mit dem nonparametrischen Kruskal-Wallis-Test eruiert. Damit kann aufgezeigt werden, ob eine Wechselwirkung zwischen dem Fach und dem Studiengang besteht. Es wurde somit ein eher konservatives Auswertungsverfahren gewählt.

Um den durch das Anwenden multipler Tests erhöhten  $\alpha$ -Fehler gering zu halten, wird mit der Bonferroni-Holm Korrektur das  $\alpha$ -Niveau adjustiert. Das Ausgangsniveau für ein signifikantes Ergebnis wird dabei auf 0.05 festgelegt.

### 3.7.3. Interpretation der Ergebnisse

Die Ergebnisse der Untersuchung wurden dahingehend interpretiert, ob ein Unterschied zwischen den jeweiligen Messzeitpunkten feststellbar ist. Das Expertenurteil dient dabei als Näherungswert an eine optimale Beurteilung einer Schülerproduktion. Die absolute Abweichung der Diagnoseleistungen der Studierenden von dem Urteil der Experten wird nicht berechnet, da es sich beim Expertenurteil um einen Zusammenzug von drei Expertenurteilen handelt. Besonders bei den inhaltlichen Kategorien kommt dieser quantitative Aspekt zum Tragen. Bei den Kategorien zur Beurteilung des Wortschatzes müssen z.B. bei Schülerin B sechs verschiedene Punkte erwähnt werden, um die volle Punktzahl, d.h. das Expertenurteil, erreichen zu können. In der für die Bearbeitung des Auftrags gegebenen Zeit ist das Zusammentragen aller dieser sechs Punkte nur schwer machbar. Wie im kommenden Ergebnisteil ersichtlich, liegen die Werte für die inhaltlichen Kategorien entsprechend weit vom absolut erreichbaren Mittelwert von 2.0 entfernt. Die Abweichung vom Expertenurteil wird deswegen bei der Interpretation der Ergebnisse nicht berücksichtigt. Stattdessen interessieren die Unterschiede der Werte zwischen den jeweiligen Messzeitpunkten.

Es wurde die Null-Hypothese getestet, dass es keinen statistisch bedeutsamen Unterschied zwischen den Mittelwerten der einzelnen Gruppen gibt. Falls die Nullhypothese verworfen wird, kann von einem Unterschied ausgegangen werden. Das Signifikanzniveau für die Annahme eines Unterschieds wurde auf  $p \leq 0.05$  festgelegt. Ausgehend von diesem  $\alpha$ -Alpha Niveau wird die bereits erwähnte Bonferroni Holm Korrektur durchgeführt.

Das Signifikanzniveau gibt die Wahrscheinlichkeit des Fehlers an, den wir begehen, wenn die Null-Hypothese (es besteht kein Zusammenhang bzw. kein Unterschied) verworfen wird.

## 4. Ergebnisse

Die Ergebnisse werden auf Grund der unterschiedlichen Befragungszeitpunkte einzeln nach Studiengang dargestellt. Weitere Aufteilungen und Subgruppenvergleiche erfolgten bei den Studierenden der Sekundarstufe I. Vergleiche konnten zwischen den Studierenden des regulären Studiengangs und des Aufbaustudiums, dem Geschlecht und den Studierenden mit und ohne Studienfach Deutsch gezogen werden.

Des Weiteren wurde nach statistischen Zusammenhängen zwischen der diagnostischen Aussage, der Rückmeldung und der Weiterarbeit gesucht, welche jedoch keine brauchbaren Ergebnisse brachte und im Folgenden nicht weiter ausgeführt wird.

### 4.1. Kindergarten/Unterstufe und Primarstufe

Die Messung und die Auswertung der Ergebnisse der Studierenden der dreijährigen Diplomstudiengänge Kindergarten/Unterstufe und Primarstufe wurden zusammen vorgenommen. Für die einzelnen Messzeitpunkte konnte die folgende Anzahl gültiger Messungen erhoben werden:

1. Messzeitpunkt (2003; 1./2. Semester): 33 gültige Messungen
2. Messzeitpunkt (2005; 3./4. Semester): 30 gültige Messungen
3. Messzeitpunkt (2006; 5./6. Semester): 19 gültige Messungen

Für die Auswertung wurde eine abhängige Stichprobe angestrebt. Um eine Veränderung in der Beurteilung der Schülerproduktionen feststellen zu können, wurden nur die Werte von Studierenden in die Auswertung einbezogen, zu denen Werte zu beiden Messzeitpunkten vorliegen und somit eine Veränderung aufzeigbar ist.

Für die Beurteilung der Schülerin A konnte bei den Studierenden der Kindergarten/Unterstufe und Primarstufe eine Vergleichsgruppe von 27 Personen zusammengestellt werden. Für die Schülerin B besteht eine Vergleichsgruppe von 18 Personen.

Das Video von Schülerin A wurde zum 1. und 2. Messzeitpunkt gezeigt, d.h. am Anfang und in der Mitte des dreijährigen Studiums. Die beim 1. und 3. Messzeitpunkt mit der gleichen Aufgabenstellung vorgelegte Schülerarbeit B zeigt im Vergleich zur Schülerarbeit A keine Unterschiede im Schwierigkeitsgrad.

#### 4.1.1. Gesamtbeurteilung

Zu den einzelnen Kompetenzbereichen wurden Mittelwerte berechnet und miteinander verglichen. Die Gesamtmittelwerte und Testresultate für den Studiengang Kindergarten/Primarstufe (KU/PS) sind in Tabelle 3 aufgeführt. Signifikante Fortschritte zeigen sich bereits nach dem ersten Studienjahr wie die Werte für die Schülerproduktion A zeigen. Erwartungsgemäss sind auch für den längeren Zeitraum vom Anfang bis zum Ende des Studiums ( $t_1 - t_3$ ) signifikante Fortschritte nachweisbar (korrigiert nach Bonferroni-Holm).

Tabelle 3: KU/PS, Gesamtbeurteilung: Veränderungen über die 3 Messzeitpunkte hinweg

	Mittelwerte und Standardabweichungen								Vergleich	
	t1				t2		t3		t1 – t2	t1 – t3
	A		B		A		B		A	B
	M	SD	M	SD	M	SD	M	SD		
Gesamt	0.8	0.2	0.7	0.2	1.0	0.2	1.1	0.3	p = 0.000	p = 0.001

Beurteilung von Schülerin A: n = 27; Beurteilung von Schülerin B: n = 18

Die Mittelwerte können zwischen 0 und 2 variieren.

Gruppenvergleiche: Wilcoxon-Test für abhängige Stichproben (korrigiert nach Bonferroni-Holm)

#### 4.1.2. Formale Kategorien

Die Gesamtergebnisse werden im Folgenden genauer nach formalen und inhaltlichen Kategorien betrachtet. Über alle formalen Kategorien hinweg kann bereits in der Mitte des Studiums und ebenso am Ende des Studiums ein signifikanter Fortschritt gegenüber dem Studienbeginn nachgewiesen werden (siehe Tabelle 4). Im Detail zeigen sich bei der differenzierten und präzisen Darstellung des Lernstands bereits signifikante Fortschritte nach dem ersten Studienjahr und auch am Schluss des Studiums. Signifikant verbessert hat sich gegen Schluss des Studiums auch die Interpretation des Lernstands, d.h. es kommen weniger unbelegte Beurteilungen vor und die Interpretationen werden vorsichtiger geäußert und mit Tatsachenbeschreibungen unterlegt.

Tabelle 4: KU/PS, formale Kategorien: Veränderungen über die 3 Messzeitpunkte hinweg

		Mittelwerte und Standardabweichungen								Vergleich	
		t1				t2		t3		t1 – t2	t1 – t3
		A		B		A		B		A	B
		M	SD	M	SD	M	SD	M	SD		
Formale Kategorien	Differenzierung Lernstand	1.0	0.6	0.7	0.5	1.4	0.3	1.4	0.2	p = 0.000	p = 0.001
	Interpretation Lernstand	0.9	0.3	0.7	0.2	1.2	0.5	1.4	0.3	n.s.	p = 0.0
	Fokus Rückmeldung	1.7	0.4	1.6	0.5	1.6	0.4	1.4	0.6	n.s.	n.s.
	Fokus Weiterarbeit	1.0	0.5	1.2	0.5	1.1	0.4	1.9	0.7	n.s.	n.s.
Formale Kategorien (gesamt)		1.2	0.3	1.1	0.3	1.3	0.2	1.5	0.3	p = 0.004	p = 0.001

Beurteilung von Schülerin A: n = 27; Beurteilung von Schülerin B: n = 18

Die Mittelwerte können zwischen 0 und 2 variieren.

Gruppenvergleiche: Wilcoxon-Test für abhängige Stichproben (korrigiert nach Bonferroni-Holm)

#### 4.1.3. Inhaltliche Kategorien

Bei den inhaltlichen wie auch den vormalig aufgeführten formalen Kategorien sind die Gesamtwerte vom ersten zum zweiten wie auch vom ersten zum dritten Messzeitpunkt signifikant (siehe Tabelle 5). Die für die Schülerproduktionen A und B unterschiedlich ausgeprägten Kategorien zur Beurteilung des Wortschatzes, des Aufbaus, der Stimmführung und der Sprache kommen hier zum Tragen. Bei der Beurteilung des Wortschatzes zeigt sich nach dem ersten Jahr eine Verbesserung, jedoch nicht am Ende des Studiums, was auf die unterschiedlichen Ausgangsniveaus für die Beurteilung von Schülerin A (m = 0.3) und Schülerin B (m = 0.7) zurück geführt werden kann. Für die Beurteilung der Schülerproduktion A mit dem relativ geringen Anfangswert von 0.3 ist dabei eine Angleichung an das Expertenurteil von 2.0 wesentlich wahrscheinlicher als für

die Schülerproduktion B, welche bereits zum ersten Messzeitpunkt relativ nah beim Höchstwert liegt. Das Gleiche lässt sich für die Beurteilung der Stimmführung sagen, wo das unterschiedliche Kategoriensystem auch als Ursache der ungleichen Ausgangsniveaus vermutet werden kann.

Die Werte für die Beurteilung der Sprache liegen nach dem ersten Jahr und auch am Studienabschluss signifikant höher als im ersten Studienjahr.

Was die Angemessenheit der Weiterarbeit (nach dem diagnostischen Urteil) anbetrifft lassen sich signifikante Fortschritte bei den KU/PS-Studierenden zwischen dem ersten und zweiten Messzeitpunkt (in der Mitte des Studiums) nachweisen, nicht aber zwischen dem zweiten und dritten Messzeitpunkt.

Tabelle 5: KU/PS, inhaltliche Kategorien: Veränderungen über die 3 Messzeitpunkte hinweg

		Mittelwerte und Standardabweichungen								Vergleich	
		t1				t2		t3		t1 – t2	t1 – t3
		A		B		A		B		A	B
		M	SD	M	SD	M	SD	M	SD		
Inhaltliche Kategorien	Beurteilung Wortschatz	0.3	0.3	0.7	0.4	0.7	0.5	1.0	0.5	p = 0.001	n.s.
	Beurteilung Aufbau	0.8	0.8	0.6	0.5	1.2	0.8	1.0	0.5	n.s.	n.s.
	Beurteilung Stimmführung	0.9	0.4	0.5	0.2	1.0	0.4	0.9	0.3	n.s.	p = 0.003
	Beurteilung Sprache	0.2	0.3	0.2	0.5	0.6	0.4	0.6	0.4	p = 0.001	p = 0.006
	Angemessenheit Weiterarbeit	0.3	0.4	0.2	0.4	0.5	0.4	0.7	0.5	p = 0.006	n.s.
Inhaltliche Kategorien (gesamt)		0.5	0.2	0.4	0.2	0.8	0.3	0.8	0.3	p = 0.000	p = 0.001

Beurteilung von Schülerin A: n = 27; Beurteilung von Schülerin B: n = 18

Die Mittelwerte können zwischen 0 und 2 variieren.

Gruppenvergleiche: Wilcoxon-Test für abhängige Stichproben (korrigiert nach Bonferroni-Holm)

## 4.2. Sekundarstufe I

Im Folgenden werden die Ergebnisse der Studierenden der Sekundarstufe I aufgezeigt.

Für die drei Messzeitpunkte ergeben sich die folgende Anzahl gültiger Messungen:

1. Messzeitpunkt (2003; 1./ 2.Semester): 81 gültige Messungen
2. Messzeitpunkt (2005; 3./ 4. Semester): 64 gültige Messungen
3. Messzeitpunkt (2007; 9. Semester): 30 gültige Messungen (keine L-Studierenden mehr dabei)

Für die Beurteilung der Schülerin A konnte aus dem 1. und 2. Messzeitpunkt eine Vergleichsgruppe von 60 Personen zusammengestellt werden. Bei der Beurteilung von Schülerin B vom 1. und 3. Messzeitpunkt ergibt sich eine Vergleichsgruppengrösse von 28 Personen.

### 4.2.1. Gesamtbeurteilung

Tabelle 6: Sek I, Gesamtbeurteilung: Veränderungen über die 3 Messzeitpunkte hinweg

	Mittelwerte und Standardabweichungen								Vergleich	
	t1				t2		t3		t1 – t2	t1 – t3
	A		B		A		B		A	B
	M	SD	M	SD	M	SD	M	SD		
Gesamt	0.9	0.2	0.7	0.2	0.9	0.2	1.0	0.2	n.s.	p = 0.000

Beurteilung von Schülerin A: n = 60; Beurteilung von Schülerin B: n = 28

Die Mittelwerte können zwischen 0 und 2 variieren.

Gruppenvergleiche: Wilcoxon-Test für abhängige Stichproben (korrigiert nach Bonferroni-Holm)

Betrachtet man das Gesamtergebnis aller Kategorien, zeigen sich bei den Studierenden der Sekundarstufe I keine signifikanten Unterschiede zwischen dem ersten und dem zweiten, jedoch zwischen dem ersten und dritten Messzeitpunkt. Die Gesamtmittelwerte der drei Messzeitpunkte variieren lediglich zwischen 0.7 und 1.0 und streuen nur sehr geringfügig um den jeweiligen Mittelwert.

### 4.2.2. Formale Kategorien

Über alle formalen Kategorien hinweg lässt sich zwischen Studienbeginn (1. Messzeitpunkt) und Studienende (3. Messzeitpunkt) eine signifikante Zunahme an diagnostischer Kompetenz feststellen. Zwischen dem ersten und zweiten Messzeitpunkt ist die Zunahme noch nicht signifikant, obwohl einzelne Unterkategorien signifikante Resultate vorweisen. Wie sich in der folgenden genaueren Auswertung zeigt, sind die Unterschiede bei den Unterkategorien unterschiedlicher Richtung und heben sich so im Gesamturteil wieder auf.

Ein signifikanter Fortschritt kann nach dem ersten Jahr für die differenzierte Darstellung und Interpretation des Lernstands aufgezeigt werden. Eine negative, signifikante Veränderung zeigt sich beim Fokus der Rückmeldung, deren Gesamtmittelwert von 1.6 bereits von Anfang an Nahe beim Expertenurteil von 2.0. Die Leistungsabnahme ist damit zu erklären, dass die Mittelwerte auf einem relativ hohen Niveau anfangen und sich zum zweiten Messzeitpunkt wieder zur Mitte hin bewegen (regression-to-the-mean).

Bei der Beurteilung des Radiobeitrags von Schülerin B liegen 4 Jahre zwischen den beiden Messzeitpunkten. Signifikante Unterschiede zum ersten Messzeitpunkt werden hier häufiger erwartet. Insgesamt kann für die formalen Kategorien ein signifikanter Fortschritt aufgezeigt werden. Bei den Unterkategorien zeigt sich lediglich bei der differenzierten Darstellung des Lernstands ein Fortschritt. Eine signifikant negative Entwicklung kann am Ende des Studiums für keine der Unterkategorien festgestellt werden. Lediglich der Bezug der Rückmeldung zeigt eine geringfügige, jedoch nicht signifikante Abnahme der Werte auf.

Tabelle 7: Sek I, formale Kategorien: Veränderungen über die 3 Messzeitpunkte hinweg

		Mittelwerte und Standardabweichungen								Vergleich	
		t1				t2		t3		t1 – t2	t1 – t3
		A		B		A		B		A	B
		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 = 0.000	p = 0.004
	Interpretation Lernstand	0.8	0.3	0.8	0.4	1.1	0.4	1.1	0.4	p = 0.003	n.s.
	Fokus Rückmeldung	1.6	0.5	1.2	0.7	1.3	0.6	1.6	0.6	p = 0.002	n.s.
	Fokus Weiterarbeit	1.2	0.5	1.2	0.4	1.0	0.4	1.4	0.4	n.s.	n.s.
Formale Kategorien (gesamt)		1.2	0.3	1.0	0.3	1.2	0.3	1.4	0.3	n.s.	p = 0.001

Beurteilung von Schülerin A: n = 60; Beurteilung von Schülerin B: n = 28

Die Mittelwerte können zwischen 0 und 2 variieren.

Gruppenvergleiche: Wilcoxon-Test für abhängige Stichproben (korrigiert nach Bonferroni-Holm)

### 4.2.3. Inhaltliche Kategorien

Die Mittelwerte der inhaltlichen Kategorien liegen im Vergleich zu den formalen Kategorien wesentlich tiefer und somit weiter weg vom Expertenurteil. Die Werte schwanken hier zwischen 0.4 und 0.9 (siehe Tabelle 8), während bei den formalen Kategorien alle Werte über 0.8 liegen.

Über alle Kategorien hinweg lässt sich bei der inhaltlichen Beurteilung der Schülerproduktion nach dem ersten Jahr noch kein Fortschritt nachweisen. Lediglich die Unterkategorie, welche die Beurteilung der Sprache betrifft, wird bereits nach einem Jahr adäquater eingeschätzt.

Über die gesamte Studiendauer lässt sich über alle inhaltlichen Kategorien eine schwache (aber dennoch signifikante) Zunahme der inhaltlichen Urteilsgüte feststellen. Die Fortschritte sind dabei bei der Beurteilung der Sprache und in den Vorschlägen für eine angemessene Weiterarbeit (aufgrund der diagnostizierten Lernausgangslage) zu beobachten.

Tabelle 8: Sek I, inhaltliche Kategorien: Veränderungen über die 3 Messzeitpunkte hinweg

		Mittelwerte und Standardabweichungen								Vergleich	
		t1				t2		t3		t1 – t2	t1 – t3
		A		B		A		B		A	B
		M	SD	M	SD	M	SD	M	SD		
Inhaltliche Kategorien	Beurteilung Wortschatz	0.6	0.4	0.7	0.3	0.6	0.4	0.6	0.3	n.s.	n.s.
	Beurteilung Aufbau	0.9	0.8	0.7	0.5	1.0	0.7	0.8	0.6	n.s.	n.s.
	Beurteilung Stimmführung	0.8	0.4	0.6	0.4	0.8	0.4	0.9	0.3	n.s.	n.s.
	Beurteilung Sprache	0.4	0.4	0.1	0.2	0.6	0.4	0.4	0.4	p = 0.000	p = 0.003
	Angemessenheit Weiterarbeit	0.4	0.5	0.2	0.3	0.4	0.3	0.8	0.5	n.s.	p = 0.000
Inhaltliche Kategorien (gesamt)		0.6	0.3	0.5	0.2	0.7	0.3	0.7	0.2	n.s.	p = 0.001

Beurteilung von Schülerin A: n = 60; Beurteilung von Schülerin B: n = 28

Die Mittelwerte können zwischen 0 und 2 variieren.

Gruppenvergleiche: Wilcoxon-Test für abhängige Stichproben (korrigiert nach Bonferroni-Holm)

#### 4.2.4. Subgruppenvergleiche

Bei den Studierenden der Sekundarstufe I konnten zusätzlich Subgruppen gebildet werden. Es kann unterschieden werden zwischen dem Geschlecht der Studierenden, zwischen Studierenden mit dem Studienfach Deutsch und ohne Studienfach Deutsch und zwischen Studierenden im regulären Studiengang und dem Aufbaustudiengang. Die Unterteilung nach Studienfach Deutsch ist deshalb interessant, weil es sich um eine Schülerproduktion aus dem Fach Deutsch handelt und hier aufgezeigt werden soll, ob die zukünftigen Deutsch-Lehrpersonen besser abschneiden als diejenigen ohne Deutsch. Das Gleiche gilt für die Studierenden des Aufbaustudiengangs, von denen auf Grund ihrer bereits gesammelten Erfahrung als Lehrperson bessere Leistungen erwartet werden. Für die einzelnen Subgruppen konnte im vorliegenden Forschungsdesign keine signifikanten Unterschiede nachgewiesen werden. Es zeigen sich jedoch Hinweise, dass bei einer grösseren Stichprobe und der Anwendung von weniger konservativen Berechnungsmethoden, sich durchaus Unterschiede aufweisen lassen bezüglich dem Studienfach und dem Studiengang. Keinerlei Hinweise zeigen sich bezüglich dem Faktor Geschlecht.

### 4.3. Zusammenfassung der Ergebnisse

In der vorliegenden Untersuchung wird die diagnostische Kompetenz aufgefasst als die Fähigkeit, (a) eine mündliche Sprachproduktion unter Beachtung der Ziele in der Aufgabenstellung angemessen einzuschätzen, (b) rückzumelden und (c) einen angemessenen Vorschlag zur Weiterarbeit ableiten zu können. Die Ergebnisse der Untersuchung werden dahingehend interpretiert, ob ein Unterschied zwischen den jeweiligen Messzeitpunkten feststellbar ist. Das Expertenurteil dient dabei als Näherungswert an eine optimale Beurteilung einer Schülerproduktion.

Für **Studierende von Kindergarten/Unterstufe und Primarstufe** kann im Verlauf des Studiums generell von einem signifikanten Zuwachs der Diagnosekompetenz ausgegangen werden. Die Fortschritte zeigen sich sowohl zwischen der ersten und zweiten Messung (in der Mitte des Studiums) wie auch zwischen der ersten und dritten Messung (am Ende des Studiums). Es lassen sich also kurzfristige wie längerfristige Fortschritte bei den Studierenden nachweisen.

Die Veränderung der Diagnosekompetenz kann nach formalen und inhaltlichen Gesichtspunkten aufgeschlüsselt werden. Unter den formalen Gesichtspunkten fallen Aspekte wie der Differenzierungsgrad der diagnostischen Aussagen oder die Fokussierung der Rückmeldung. Die inhaltlichen Gesichtspunkte meinen die eigentliche Beurteilung verschiedener Diagnostikbereiche, also zum Beispiel die Beurteilung der Stimmführung oder der Gliederung der geforderten Schülerproduktion (Radioreportage). Dabei wird ein Vergleich zwischen dem Urteil der Studierenden und dem Expertenurteil vorgenommen.

Über alle formalen Kategorien hinweg kann bei den Studierenden von Kindergarten/Unterstufe und Primarstufe sowohl nach einem Jahr wie auch am Ende des Studiums ein signifikanter Fortschritt gegenüber dem Studienbeginn nachgewiesen werden. Bei der differenzierten und präzisen Darstellung des Lernstands zeigen sich dabei bereits Fortschritte nach dem ersten Studienjahr und auch am Schluss des Studiums. Gegen Ende des Studiums hat sich auch die Interpretation des Lernstands verbessert, d.h. es kommen weniger unbelegte Beurteilungen vor und die Interpretationen werden vorsichtiger geäußert und mit Tatsachenbeschreibungen unterlegt.

Bei den inhaltlichen Kategorien sind die Gesamtwerte bei den Studierenden von Kindergarten/Unterstufe und Primarstufe vom ersten zum zweiten wie auch vom ersten zum dritten Messzeitpunkt signifikant. Die für die Schülerproduktionen A und B unterschiedlich ausgeprägten Kategorien zur Beurteilung des Wortschatzes, des Aufbaus, der Stimmführung und der Sprache

kommen auch hier zum Tragen. Bei der Beurteilung des Wortschatzes zeigt sich zwar nach dem ersten Jahr eine Verbesserung, jedoch nicht am Ende des Studiums, was auf die unterschiedlichen Ausgangsniveaus für die Beurteilung von Schülerin A ( $m = 0.3$ ) und Schülerin B ( $m = 0.7$ ) zurückgeführt werden kann. Für die Beurteilung der Schülerproduktion A mit dem relativ geringen Anfangswert von 0.3 ist dabei eine Angleichung an das Expertenurteil von 2.0 wesentlich wahrscheinlicher als für die Schülerproduktion B, welche bereits zum ersten Messzeitpunkt relativ nah beim Höchstwert liegt. Das Gleiche lässt sich für die Beurteilung der Stimmführung sagen, wo das unterschiedliche Kategoriensystem auch als Ursache der ungleichen Ausgangsniveaus vermutet werden kann.

Für **Studierende der Sekundarstufe I** lässt sich ein ähnlicher Verlauf beschreiben. Es fällt jedoch auf, dass die Studierenden der Sek I - verglichen mit den Studierenden von KU/PS - generell mit höheren Werten ihr Studium beginnen. Sie kommen damit bereits zu Beginn des Studiums den Anforderungen des Expertenurteils näher. Auch bei den Sek-Studierenden kann im Verlauf des Studiums generell von einem signifikanten Zuwachs der Diagnosekompetenz ausgegangen werden. Die Fortschritte zeigen sich allerdings nicht kurzfristig (zwischen der ersten und zweiten Messung), sondern längerfristig zwischen der ersten und dritten Messung (am Ende des Studiums).

Über alle formalen Kategorien hinweg lässt sich zwischen Studienbeginn (1. Messzeitpunkt) und Studienende (3. Messzeitpunkt) eine signifikante Zunahme an diagnostischer Kompetenz feststellen. Zwischen dem ersten und zweiten Messzeitpunkt ist die Zunahme noch nicht signifikant, obwohl einzelne Unterkategorien signifikante Resultate vorweisen. Ein signifikanter Fortschritt kann nach dem ersten Jahr für die differenzierte Darstellung und Interpretation des Lernstands aufgezeigt werden. Eine negative, signifikante Veränderung zeigt sich allerdings beim Fokus der Rückmeldung. Die Rückmeldungen orientieren sich weniger an den Vorgaben der Lernaufgaben oder beziehen sich nicht auf die Aufgabenstellung. Dieser Rückgang wird allerdings am Ende wieder aufgeholt, sodass am Ende der Ausbildung wieder das Anfangsniveau erreicht wird.

Die Mittelwerte der inhaltlichen Kategorien liegen im Vergleich zu den formalen Kategorien wesentlich tiefer und somit weiter weg vom Expertenurteil. Über alle Kategorien lässt sich bei der inhaltlichen Beurteilung der Schülerproduktion nach dem ersten Jahr noch kein Fortschritt nachweisen. Lediglich die Unterkategorie, welche die Beurteilung der Sprache betrifft, wird bereits nach einem Jahr adäquater eingeschätzt. Über die gesamte Studiendauer lässt sich eine schwache (aber dennoch signifikante) Zunahme der inhaltlichen Urteilsgüte feststellen. Die Fortschritte sind bei der Beurteilung der Sprache und in den Vorschlägen für eine angemessene Weiterarbeit zu beobachten.

## 5. Diskussion

Gegenstand der vorliegenden Untersuchung ist der Aufbau der Diagnosekompetenz im Fach Deutsch bei Studierenden an einer Pädagogischen Hochschule. Der Längsschnitt bestätigt die Erwartungen hinsichtlich einer Zunahme der diagnostischen Kompetenz während des Studiums. Diese Ergebnisse wurden nicht aufgrund von Selbstdeklarationen oder -beurteilungen ermittelt, sondern beruhen auf der Messung von diagnostischen Aussagen im Vergleich zu einem Expertenurteil. Sie unterstützen jedoch Befunde zur Diagnosekompetenz von PH-Studierenden, die durch Selbsteinschätzung gewonnen wurden (vgl. Baer, Dörr, Fraefel et al., 2005). Die Studierenden lernen vor allem genauer hinzuschauen, differenziertere Einschätzungen vorzunehmen und den Lernstand präziser zu beurteilen. Trotz den positiven Befunden, was die grundsätzliche Veränderung bzw. Zunahme der Diagnosekompetenz anbetrifft, muss berücksichtigt werden, dass die Zunahmen während der Ausbildung zwar über alle untersuchten Bereiche hinweg signifikant, jedoch relativ klein sind.

Die deutlichsten Fortschritte sind bei der eigentlichen Urteilsleistung zu erkennen: die Aussagen über den Lernstand sind am Ende des Studiums signifikant differenzierter und präziser als noch am Anfang. Bei den inhaltlichen Kriterien kann die Sprache (z.B. die korrekte Aussprache der Standardsprache) im Laufe des Studiums immer besser beurteilt werden. In anderen Bereichen zeigen sich jedoch keine Fortschritte, so etwa in der Beurteilung des sprachlichen Aufbaus der geforderten Sprachproduktion. Bei wiederum anderen Bereichen lassen sich teilweise Fortschritte (kurz- und/oder langfristig und je nach Studiengang) feststellen, so beispielsweise bei der Beurteilung des verwendeten Wortschatzes oder bei der Stimmführung.

Die Verläufe der beiden untersuchten Studiengänge Kindergarten/Primarstufe und Sekundarstufe unterscheiden sich voneinander. Hinsichtlich der formalen und inhaltlichen Kriterien lassen sich im 4.5 Jahre dauernden Studiengang für die Sekundarstufe die Veränderungen nur längerfristig nachweisen, also zwischen dem ersten und dritten Messzeitpunkt. Im kürzeren Studiengang der Kindergarten/Primarstufe können die Veränderungen sowohl kurzfristig (d.h. zwischen dem ersten und zweiten Messzeitpunkt) wie auch längerfristig (zwischen dem ersten und dritten Messzeitpunkt) nachgewiesen werden. Vermutlich spielen für diesen Befund sowohl die längere Studiendauer wie auch die generell höheren Eingangsvoraussetzungen bei den Sek-Studierenden eine Rolle. Die längere Studiendauer ermöglicht, dass sich die Studierenden zu Beginn des Studiums verstärkt mit Grundlagen (des Fachs Deutsch) befassen können, ohne dass gleich konkrete Diagnoseleistungen erwartet werden. Analog zeigen sich bei den Auswertungen grössere Veränderungen erst beim zweiten Messzeitpunkt. Andererseits ist es schwieriger bei höheren Eingangsvoraussetzungen die Leistungen bei jedem Messzeitpunkt signifikant zu verbessern. Die Studierenden der Sekundarstufe sind somit trotz des intensiveren fachdidaktischen Vorlaufs am Schluss des Studiums nicht besser als die Studierenden der Kindergarten/Primarstufe. Weshalb sich die Eingangsvoraussetzungen der Studierenden der Primarstufe von denjenigen der Sekundarstufe I unterscheiden, konnte mit vorliegender Studie nicht eruiert werden.

Der Diagnosekompetenz wird für einen lernförderlichen Unterricht eine hohe Bedeutung zugemessen. Insbesondere für einen adaptiven Lehrstil, der sich vor allem für heterogene Lerngruppen eignet, stellt die Fähigkeit angemessene Lerndiagnosen zu stellen, eine wichtige Voraussetzung dar (vgl. z.B. Beck et al., 2008). Unter schulpädagogischer Perspektive wird das Diagnostizieren nicht nur im Zusammenhang mit der Urteilsgüte diskutiert. Aufgegriffen wird vielmehr der Bezug zum Rückmelden und Überlegungen zur nachfolgenden pädagogischen Intervention (vgl. Horstkemper, 2006; Buholzer, 2010). Diesem Zugang fühlt sich die vorliegende Untersuchung verpflichtet. Neben der Messung der Urteilsgüte wurden auch die Rückmeldungen an den Schüler/die Schülerin und die Vorschläge zur Weiterarbeit erfasst. Die Ergebnisse belegen, dass auch

in diesen Bereichen teilweise Fortschritte erzielt wurden. Unter formalen Gesichtspunkten findet bei den Primarstudierenden keine Fokussierung zur Rückmeldung statt und auch bei den Vorschlägen zur Weiterarbeit zeigen sich nur kurzfristige Effekte. Ein ähnliches Bild zeigt sich bei den Studierenden der Sekundarstufe I. Auch hier sind nur bedingt Hinweise zu finden, dass die Fokussierung von Rückmeldungen und Weiterarbeit zunimmt. Was die Güte von Vorschlägen zur Weiterarbeit anbetrifft, kann längerfristig von einer schwachen Zunahme berichtet werden. Was die engeren Verbindungen anbetrifft zwischen diagnostischer Aussage, Rückmeldung und anschliessender Intervention konnten keine Zusammenhänge ermittelt werden. Offenbar fällt es den Studierenden schwer, für einen diagnostizierten Sachverhalt eine zugeschnittene Förderung zu konzipieren. Dies wäre – mit Blick auf die Studie von Helmke und Schrader (vgl. 1987) besonders wichtig, um Lernerfolg bei den Schülern zu erzielen. Alleine gute Diagnosen zu erstellen, macht noch nicht den guten Lehrer / die gute Lehrerin aus, erst in der Kombination von Diagnose und entsprechender Strukturierung im Unterricht werden Schülerinnen und Schüler in ihrem Lernen auch tatsächlich gefördert. Dieser Schnittstelle zwischen dem diagnostischen Urteil und den gefolgerten Fördermassnahmen gilt es künftig besondere Aufmerksamkeit zu schenken. Auffallend ist, dass sich Studien im Bereich der pädagogischen Diagnostik sehr detailliert auf das diagnostische Urteil beziehen (vgl. zum Beispiel Lorenz & Artelt 2009 oder Karing 2009), den Anschluss, respektive Massnahmenplanungen jedoch ausblenden. Relevant für das Berufsfeld wären gerade solche wissenschaftsgestützte Aussagen über jene Schnittstelle.

Ein weiterer interessanter Aspekt ist die Tatsache, dass es für die Diagnoseleistung (im engeren Sinne) wie auch für die Rückmeldung und die Vorschläge zur Weiterarbeit mit dem Kind, nicht entscheidend ist, ob es sich um Studierende mit Unterrichtserfahrung (Lehrpersonen im Aufbaustudium) oder um reguläre Studierende lediglich mit Praktikumserfahrung handelt. Es zeigen sich zu keinem Zeitpunkt signifikante Unterschiede zwischen den beiden Studierendengruppen. Damit wird ein Ergebnis von Schrader (1989) gestützt, wonach weder die allgemeine Berufserfahrung noch die jahrgangsspezifische Unterrichtserfahrung mit der Ausprägung der Diagnosekompetenz zusammenhängen (vgl. 1989, 125f.). Auch McElvany et al. (2009) wiesen nur einen schwachen Zusammenhang zwischen Berufserfahrung und Diagnosekompetenz nach. Offenbar ist gemäss diesen und auch vorliegender Studien der gemessene Erfahrungshintergrund nicht das entscheidende Kriterium für das Erbringen einer akkuraten Diagnoseleistung. Mit Schrader (2009) muss jedoch angemerkt werden, dass die Anzahl Berufsjahre, die als Indiz für die Berufserfahrungen gelten, ein sehr grobes Unterscheidungsmerkmal darstellen. Dasselbe gilt auch für die Unterscheidung von Novize und Experte (vgl. Dünnebier et al., 2009) oder die Differenzierung von Gymnasial- und Grundschullehrern (vgl. Karing, 2009). Wollte man den effektiven Erfahrungshintergrund von Lehrpersonen mit einbeziehen, müsste hier genauer erfasst werden, um welche Art von Berufserfahrung es sich handelt. Gemäss den vorliegenden Befunden ist das Studium für den Aufbau der Diagnosekompetenz offenbar wichtiger als die Lehrererfahrung. Der mit der Studie eingeschlagene Weg, den Kompetenzaufbau im Rahmen der Lehrerinnen- und Lehrerbildung abzubilden, soll deshalb weiterverfolgt werden.

Erwartungswidrig ist das Ergebnis, wonach es offenbar keine Rolle spielt, ob die Studierenden der Sekundarstufe I das Fach Deutsch gewählt haben oder nicht. Ein Student mit Schwerpunkt Naturwissenschaften beispielsweise unterscheidet sich also nicht von Studierenden mit Studienfächern im sprachlichen Bereich. Dieses Ergebnis muss vorsichtig interpretiert werden, denn der Grund für diese Nicht-Unterscheidung könnte auf die kleinen Stichproben zurückgeführt werden, was auch für die Unterscheidung zwischen Personen mit und ohne Lehrererfahrung gilt. Andererseits liegt die Vermutung Nahe, dass fachspezifische Bereiche nur wenig relevant sind, Schülerarbeiten zu diagnostizieren. Dieser Befund widerspricht jedoch Annahmen, wonach die Diagnosekompetenz eng mit dem Fachinhalt zusammenhänge, wie dies etwa Lorenz und Artelt (2009)

nachgewiesen haben. Auch andere Studien widerlegen die Annahme eines übergreifenden Konstrukts der Diagnosekompetenz (z.B. Spinath, 2005).

Der vorliegende Forschungsansatz versucht die Diagnosekompetenz anhand einer Schülerproduktion zu messen und nicht über eine Selbsteinschätzung (durch die Studierenden selbst) oder durch Fremdeinschätzung (z.B. die Praxislehrperson) zu ermitteln. Diese Verfahren werden zwar angewendet (z.B. Baer et al., 2005), müssen jedoch hinsichtlich der Validität kritisch hinterfragt werden. Mit Selbstdeklarationen und Vergleichen mit Fremdeinschätzungen können nur Teilaspekte der tatsächlichen Kompetenzentwicklung aufgedeckt werden. Mit dem vorliegenden methodischen Verfahren kann die Validität der Ergebnisse erhöht werden, dennoch muss im Nachhinein kritisch angemerkt werden, dass die Studienergebnisse auf einem Vergleich mit einem Expertenurteil beruhen. Das Expertenurteil setzt sich aus dem Urteil verschiedener Experten mit theoretischer und praktischer Ausrichtung zusammen und stellt als solches einen Kompromiss dar. Es handelt sich also nicht um einen objektiven oder neutralen Wert, wie er beispielsweise durch ein normiertes Testverfahren ermittelt werden könnte. Die Annäherung der Studierenden an das Expertenurteil muss daher auch als eine relative Annäherung betrachtet werden. Zudem kumulieren im Expertenurteil die Urteile von drei Personen, was sich besonders bei den inhaltlichen Kategorien bemerkbar macht.

Ausbildungsinstitutionen haben Kompetenzen beschrieben, die von PH-Studierenden im Laufe der Ausbildung erreicht werden sollen. Um Rückschlüsse auf die Effektivität der Lehrerinnen- und Lehrerausbildung zu machen, ist es erforderlich, diese wie auch andere Kompetenzen regelmässig zu überprüfen. Dies bedingt einerseits klare und überprüfbare Zielvorgaben seitens der Ausbildungsinstitution und andererseits ein Instrumentarium, um zuverlässige und valide Informationen zu erhalten.

Mit dieser Studie können zwar erste Hinweise zum Kompetenzaufbau gewonnen werden, für Ableitungen von Implikationen für die Aus- und Weiterbildung ist es momentan noch zu früh – dafür sind weitere Studien erforderlich. Wir wissen wenig darüber, wie die diagnostischen Urteile zustande gekommen sind. Welche (impliziten wie expliziten) Bezugsnormen und welche Kognitionen im Beurteilungsprozess leitend waren. Wie Studien belegen, hängen diagnostische Urteile auch mit Kontextfaktoren zusammen, beispielsweise fällt es leichter akkurate Diagnosen in Klassen mit ausgeprägter Heterogenität zu stellen (Karing, 2009). Auch der Zweck der Beurteilung unterstützt oder mildert Urteilsverzerrungen (vgl. Krolak-Schwerdt et al., 2009). Diesen Kontextbedingungen, bzw. der ökologischen Validität, müssten bei einer revidierten Studiendurchführung mehr Beachtung geschenkt werden. Dazu gehört auch eine differenziertere Erfassung der Voraussetzungen bei den Studierenden (z.B. Erfahrungshintergrund und Vorwissen).

Vor diesem Hintergrund und den gewonnenen Erfahrungen soll die Methode zur Erfassung des Aufbaus der Diagnosekompetenz von unterrichtsnahen Schülerprodukten weiter verfeinert werden. Denkbar wäre eine verstärkte Konzentration auf objektivere Vergleichsgrössen und der Einbezug einer grösseren Stichprobe, um die Studienergebnisse noch stärker abstützen zu können.

## 6. Literatur

- Allemann-Ghionda, Chr. & Terhart, E. (2006). Kompetenzen und Kompetenzentwicklung von Lehrerinnen und Lehrern: Ausbildung und Beruf. *Zeitschrift für Pädagogik*, 51. Beiheft, 7-11.
- Arendt, M. (1996). Beurteilung mündlicher Leistungen. Eine Untersuchung (Teil 2). *Praxis Fremdsprachenunterricht*, (4), 3-8.
- Artelt, C.; Stanat, P.; Schneider, W. & Schiefele, U. (2001). Lesekompetenz: Testkonzeption und Ergebnisse. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann & M. Weiss (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 69-137). Opladen: Leske + Budrich.
- Baer, M., Dörr, G., Fraefel, U. et al. (2005). Standarderreichung beim Erwerb von Unterrichtskompetenz in der Lehrerinnen- und Lehrerbildung. Analyse der Wirksamkeit der berufsfeldorientierten Ausbildung. Zürich: PHZH.
- 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, Chr., Müller, P., Niedermann, R., Rogalla, M. & Vogt, F. (2008). *Adaptive Lehrkompetenz*. Münster: Waxmann.
- Baurmann, J. (2002). *Schreiben, Überarbeiten, Beurteilen*. Seelze: Kallmeyer.
- Buholzer, A. (2006). *Förderdiagnostisches Sehen, Denken und Handeln*. Donauwörth: Auer.
- Buholzer, A. (2010). Lernprozesse förderorientiert diagnostizieren. In A. Buholzer & A. Kummer Wyss (Hrsg.), *Alle Gleich – alle unterschiedlich. Zum Umgang mit Heterogenität in Schule und Unterricht* (S. 97-108). Seelze & Zug: Kallmeyer & Klett.
- Bundschuh, K. (1994). *Praxiskonzepte der Förderdiagnostik. Möglichkeiten der Anwendung in der sonder- oder heilpädagogischen Praxis*. Bad Heilbrunn: Klinkhardt.
- Coradi Vellacott, M. & Wolter, S. C. (2005). *Chancengerechtigkeit im schweizerischen Bildungswesen. Trendbericht SKBF Nr. 9*. Aarau: Albdruk.
- Criblez, L., Oelkers, J., Reusser, K., Berner, E., Halbheer, U. & Huber, Chr. (2009). *Bildungsstandards*. Seelze: Kallmeyer.
- Diekmann, A. (1999). *Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen* (5., durchgesehene Auflage). Reinbek: Rowohlt Taschenbuch Verlag.
- Ditton, H., Krüsken, J. & Schauenberg, M. (2005). Bildungsungleichheit – der Beitrag von Familie und Schule. *Zeitschrift für Erziehungswissenschaft*, 8(2), 285-304.
- Ditton, H. & Merz, D. (2000). *Qualität von Schule und Unterricht. Kurzbericht über erste Ergebnisse einer Untersuchung an bayerischen Schulen*. <http://www.quassu.net/Bericht1.pdf> (besucht am 8. August 2010)
- Dünnebier K., Gräsel C. & Krolak-Schwerdt S. (2009). Urteilsverzerrungen in der schulischen Leistungsbeurteilung. *Psychologie. German Journal of Educational Psychology. Themenheft: Diagnostische Kompetenz von Lehrkräften*, 23(3-4), 237-245.
- Dusek, J. B. & Joseph, G. (1985). The bases of teacher expectancies. In J. B. Dusek (Ed.), *Teacher expectancies* (pp. 229-250). Hillsdale, New Jersey, London: Erlbaum.
- Eberwein, H. & Knauer, S. (2003). Einführung und Problemstellung. In H. Eberwein & S. Knauer (Hrsg.), *Handbuch Lernprozesse verstehen. Wege einer neuen (sonder-)pädagogischen Diagnostik* (2., neu ausgestattete Auflage) (S. 7-12). Weinheim, Basel, Berlin: Beltz.
- Fischer, Ch. (2009). Individuelle Förderung besonders begabter Kinder mit Lern- und Leistungsschwierigkeiten. In Ch. Fischer, U. Westphal & Ch. Fischer-Ontrup (Hrsg.), *Individuelle Förderung: Lernschwierigkeiten als schulische Herausforderung* (S. 178-191). Münster: LIT.

- Frei, B. (2003). *Pädagogische Autorität. Eine empirische Untersuchung bei Schülerinnen, Schülern und Lehrpersonen der 5., 6. und 8. Schulklasse*. Münster: Waxmann.
- Frey, A. (2006). Methoden und Instrumente zur Diagnose beruflicher Kompetenzen von Lehrkräften – eine erste Standortbestimmung zu bereits publizierten Instrumenten. *Zeitschrift für Pädagogik*, 51. Beiheft, 30-46.
- Früh, W. (2001). *Inhaltsanalyse. Theorie und Praxis* (5., überarbeitete Auflage). Konstanz: UVK Medien.
- Girmes, R. (2006). Lehrerprofessionalität in einer demokratischen Gesellschaft. Über Kompetenzen und Standards in einer erziehungswissenschaftlich fundierten Lehrerbildung. *Zeitschrift für Pädagogik*, 51. Beiheft, 14-29.
- Groeben, von der, A. (2003). Verstehen lernen. Diagnostik als didaktische Herausforderung. *Pädagogik*, 55(4), 6-9.
- Grossenbacher, S. (2010). Kompetenz und Professionalität entwickeln. In A. Buholzer & A. Kummer Wyss (Hrsg.), *Alle gleich – alle unterschiedlich! Zum Umgang mit Heterogenität in Schule und Unterricht* (S. 162-168). Seelze: Kallmeyer.
- Günther, G. & Brunstein, J.C. (2003). Erfahrungen aus dem psychodiagnostischen Praktikum. *Journal für Lehrerinnen- und Lehrerbildung*, 3(2), 20-24.
- Hascher, T. (2003). Diagnose als Voraussetzung für gelingende Lernprozesse. *Journal für Lehrerinnen- und Lehrerbildung*, 3(2), 25-30.
- Heidemeier, H. (2005). *Self and supervisor ratings of job-performance: Meta-analysis and a process model of rater convergence*. [http://www.opus.ub.uni-erlangen.de/opus/volltexte/2005/170/pdf/Heidemeier\\_Diss\\_Mai\\_2005.pdf](http://www.opus.ub.uni-erlangen.de/opus/volltexte/2005/170/pdf/Heidemeier_Diss_Mai_2005.pdf) (besucht am 20. Januar 2010)
- Helmke, A. (2009). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts*. Seelze: Kallmeyer.
- Helmke, A., Hosenfeld, I. & Schrader, F.-W. (2004). Vergleichsarbeiten als Instrument zur Verbesserung der Diagnosekompetenz von Lehrkräften. In R. Arnold & C. Griesse (Hrsg.), *Schulleitung und Schulentwicklung* (S. 119-144). Hohengehren: Schneider-Verlag.
- Horstkemper, M. (2006). Fördern heisst diagnostizieren. Pädagogische Diagnostik als wichtige Voraussetzung für individuellen Lernerfolg. *Friedrich Jahresheft XXIV: Diagnostizieren und Fördern*, 4-7.
- Horstkemper, M. (2004). Diagnosekompetenz als Teil pädagogischer Professionalität. In *Neue Sammlung* 44(2), 201-214.
- Hosenfeld, I., Helmke, A. & Schrader, F.-W. (2002). Diagnostische Kompetenz: Unterrichts- und lernrelevante Schülermerkmale und deren Einschätzung durch Lehrkräfte in der Unterrichtsstudie SALVE. In M. Prenzel & J. Doll (Hrsg.), *Bildungsqualität von Schule: Schulische und ausserschulische Bedingungen mathematischer, naturwissenschaftlicher und überfachlicher Kompetenzen*. *Zeitschrift für Pädagogik*, 45. Beiheft, 65-82.
- Ingenkamp, K. (1997). *Lehrbuch der Pädagogischen Diagnostik*. Weinheim: Beltz.
- Jurt Betschart, J., Hofstetter, M. Th & Vogel Wiederkehr, S. (2000). *Beurteilen und Fördern im Deutschunterricht*. Band 1. Zürich: sabe.
- Karing, C. (2009). Diagnostische Kompetenz von Grundschul- und Gymnasiallehrkräften im Leistungsbereich und im Bereich Interessen. *Psychologie. German Journal of Educational Psychology. Themenheft: Diagnostische Kompetenz von Lehrkräften*, 23(3-4), 197-209.
- Kretschmann, R. (2004). Diagnostikausbildung – für alle Lehrerinnen und Lehrer? In W. Mutzeck & P. Jogschies (Hrsg.), *Neue Entwicklungen in der Förderdiagnostik. Grundlagen und praktische Umsetzungen* (S. 123-137). Weinheim: Beltz.
- Krolak-Schwerdt, S., Böhmer, M. & Gräsel C. (2009). Verarbeitung von schülerbezogener Information als zielgeleiteter Prozess: Der Lehrer als „flexibler Denker“. *Psychologie. German Journal of Educational Psychology. Themenheft: Diagnostische Kompetenz von Lehrkräften*, 23(3-4), 175-186.
- Kronig, W., Haeberlin, U. & Eckhart, M. (2000). *Immigrantenkinder und schulische Selektion. Pädagogische Visionen, theoretische Erklärungen und empirische Untersuchungen zur Wirkung integrierender und separierender Schulformen in den Grundschuljahren*. Bern: Paul Haupt.

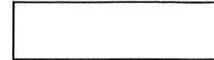
- Kutzer, R. (1999). Überlegungen zur Unterrichtsorganisation im Sinne strukturorientierten Lernens. In H. Probst (Hg.), *Mit Behinderungen muss gerechnet werden* (S. 15-69). Solms Oberbiel: Jarick.
- Landy, F. J. & Farr, J. L. (1980). Performance rating. *Psychological Bulletin*, 87, 72-104
- Langfeldt, H.P. & Tent, L. (1999). *Pädagogisch-psychologische Diagnostik, Band 2. Anwendungsbereiche und Praxisfelder*. Göttingen: Hogrefe.
- Lersch, R. (2006). Lehrerbildung im Urteil der Auszubildenden. Eine empirische Studie zu beiden Phasen der Lehrerausbildung. *Zeitschrift für Pädagogik*, 51. Beiheft, 164-181.
- Lorenz, Ch. & Artelt, C. (2009). Fachspezifität und Stabilität diagnostischer Kompetenz von Grundschullehrkräften in den Fächern Deutsch und Mathematik. *Psychologie. German Journal of Educational Psychology. Themenheft: Diagnostische Kompetenz von Lehrkräften*, 23(3-4), 211-222.
- Lukesch, H. (1998). *Einführung in die pädagogisch-psychologische Diagnostik*. Regensburg: S. Roderer Verlag.
- Mand, J. (1998). Förderdiagnostik als Lernprozessdiagnostik. Begründung, Methodenprobleme und diagnostischer Alltag. In H. Eberwein & S. Knauer (Hrsg.), *Handbuch Lernprozesse verstehen* (S. 39-53). Weinheim: Beltz.
- McElvany, N., Schröder, S., Hachfeld, A., Baumert, J., Richter, T., Schnotz, W., Horz, H. & Ullrich, M. (2009). Diagnostische Fähigkeiten von Lehrkräften bei der Einschätzung von Schülerleistungen und Aufgabenschwierigkeiten bei Lernmedien mit instruktionalen Bildern. *Psychologie. German Journal of Educational Psychology. Themenheft: Diagnostische Kompetenz von Lehrkräften*, 23(3-4), 223-235.
- Meyer, H. & Klapper, A. (2006). Unterrichtsstandards für ein kompetenzorientiertes Lernen und Lehren. In R. Hinz & B. Schumacher (Hrsg.), *Auf den Anfang kommt es an* (89-108). Wiesbaden: Verlag für Sozialwissenschaften.
- Moser, U., Keller, F. & Tresch, S. (2003). *Schullaufbahn und Leistung. Bildungsverlauf und Lernerfolg von Zürcher Schülerinnen und Schüler am Ende der 3. Volksschulklasse*. Bern: h.e.p. Verlag.
- Mutzeck, W. (2004). Grundlegende Aspekte der Diagnostik in der Förderdiagnostik. In W. Mutzeck & P. Jogschies (Hrsg.), *Neue Entwicklungen in der Förderdiagnostik. Grundlagen und praktische Umsetzungen* (S. 10-20). Weinheim: Beltz.
- Oser, F. (1997a). Standards in der Lehrerbildung. Teil I: Berufliche Kompetenzen, die hohen Qualitätsmerkmalen entsprechen. *Beiträge zur Lehrerbildung*, 15 (1), 26-37.
- Oser, F. (1997b). Standards in der Lehrerbildung Teil II: Wie werden Standards in der schweizerischen Lehrerbildung erworben? Erste empirische Ergebnisse. *Beiträge zur Lehrerbildung*, 15 (2), 210-228.
- Oser, F.; Curcio, G.-P. & Düggele, A. (2007). Kompetenzmessung in der Lehrerbildung als Notwendigkeit – Fragen und Zugänge. *Beiträge zur Lehrerbildung*, 25 (1), 14-25.
- Rauin, U. (2007). *Im Studium wenig engagiert – im Beruf schnell überfordert. Studierverhalten und Karrieren im Lehrerberuf – Kann man Risiken schon im Studium prognostizieren?* [http://www.forschung-frankfurt.uni-frankfurt.de/2007/Forschung\\_Frankfurt\\_2007/3-07/Im\\_Studium\\_wenig\\_\\_\\_12\\_.pdf](http://www.forschung-frankfurt.uni-frankfurt.de/2007/Forschung_Frankfurt_2007/3-07/Im_Studium_wenig___12_.pdf) (besucht am 28.05.2009)
- Reusser, K. & Messner, H. (2002). Das Curriculum der Lehrerinnen- und Lehrerbildung – ein vernachlässigtes Thema. *Beiträge zur Lehrerbildung*, 20(3), 282-299.
- Roos, M. (2001). *Ganzheitliches Beurteilen und Fördern in der Primarschule*. Chur: Verlag Rüegger.
- Rost, D. H. & Hanse, P. (1997). Wer nichts leistet, ist nicht begabt? *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 29, 167-177.
- Ruf, U. (2003). Lerndiagnostik und Leistungsbewertung in der Dialogischen Didaktik. *Pädagogik*, 55(4), 10-16.
- Sacher, W. (2001). *Leistungen entwickeln, überprüfen und beurteilen. Grundlagen, Hilfen und Denkanstöße für alle Schularten* (3., erweiterte und überarbeitete Auflage). Bad Heilbrunn: Klinkhardt.
- Schrader, F.-W. (2009). Anmerkungen zum Themenschwerpunkt Diagnostische Kompetenz von Lehrkräften. *Psychologie. German Journal of Educational Psychology. Themenheft: Diagnostische Kompetenz von Lehrkräften*, 23(3-4), 237-245.

- Schrader, F.-W. & Helmke A. (1987). Diagnostische Kompetenz von Lehrern: Komponenten und Wirkungen. *Empirische Pädagogik. Zeitschrift zu Theorie und Praxis erziehungswissenschaftlicher Forschung* I(1), 27-52.
- Schrader, F.-W. (1989). *Diagnostische Kompetenz von Lehrern und ihre Bedeutsamkeit für die Gestaltung des Unterrichts*. Frankfurt: 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* (4), 312-324.
- Schweizerische Koordinationsstelle für Bildungsforschung (SKBF) (2006). *Bildungsbericht Schweiz 2006*. Suhr: Druckerei AG.
- Spinath, B. (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie*, 9(1/ 2), 85-95.
- Spinath, B. (2004). Diagnostische Kompetenzen von Lehrerinnen und Lehrern. *Lernende Schule*, 26, 16-17.
- Staub, F.C. (2006). Wenn der Coach kommt ... Diagnose- und Unterrichtskompetenz stärken durch neue Beratungsformen. *Friedrich Jahresheft XXIV: Diagnostizieren und Fördern*, 138-140
- Stallmann, M. (1999). Soziale Herkunft und Oberschulübergänge in einer Berliner Schülergeneration. Eine Logit-Analyse von Schülerbögen. *Zeitschrift für Pädagogik*, 36(2), 214-258.
- Steppacher, J. (2004). Förderdiagnostik in der Schulischen Heilpädagogik. Eine kooperative und interdisziplinäre Aufgabe. *Schweizerische Zeitschrift für Heilpädagogik*, 10(4), 18-23.
- Südkamp, A. & Möller, J. (2009). Referenzgruppeneffekte im Simulierten Klassenraum: Direkte und indirekte Einschätzungen von Schülerleistungen. *Psychologie. German Journal of Educational Psychology. Themenheft: Diagnostische Kompetenz von Lehrkräften*, 23(3-4), 161-174.
- Valtin, R. (1996). Zur Entstehung von Lern-Behinderungen durch falsche Lehr-/Lernkonzepte beim Schriftspracherwerb. In H. Eberwein (Hg.), *Handbuch Lernen und Lern-Behinderungen* (S. 369-387). Weinheim: Beltz.
- Wiater, W. (2004). Fehler = Helfer. *Lernchancen*, 39, 4-7.
- Wild, K.-P. & Rost, D. H. (1995). Klassengröße und Genauigkeit von Schülerbeurteilungen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 27(1), 78-90.



## Anhang

### Fragebogen



## Auftrag

Mündlicher Sprachgebrauch



### AUFGABE 1

## Strandbad live - eine Radioreportage

Du machst für eine Radiostation eine Reportage über das Geschehen an diesem berühmten Strand. Beschreibe einen Rundgang durch das Strandbad und schildere, was du siehst und erlebst. Notiere dir dabei einige Stichworte.

- Mach eine kurze Einleitung, so dass die Zuhörenden wissen, worum es geht.
- Schildere mindestens drei Szenen und rücke dabei eine Person ins Zentrum.
- Schliesse die Reportage mit einem passenden Schluss.

Sprich die Reportage in Standardsprache auf Tonband, so dass sie später als Klassenradiosendung gehört werden kann.

### LERNZIEL:

Über Ereignisse und Beobachtungen in der Standardsprache berichten

### Achte darauf:

1. Du stellst die Atmosphäre und Stimmung in den einzelnen Szenen anschaulich dar.
2. Die Einleitung führt die Zuhörenden in die Situation hinein, der Schluss rundet ab.
3. Du sprichst ausdrucksvoll und flüssig.
4. Deine Aussprache in der Standardsprache ist klar und deutlich.

ISH  
Alois Buholzer  
Sandra Zulliger

Pädagogische Hochschule  
Zentralschweiz · Luzern



Schwerpunkt: Mündlicher Sprachgebrauch (und ums Baden)

Mündlicher Sprachgebrauch

© PHZ ISH

2

Notizen der Schülerin A

guten Morgen im Radio Pilatus,  
mit Ursula Seu!

In dem Strandbad Baldeg ist  
heute <sup>sehr</sup> viel los, es hat sehr viele  
Leute hier. Hier ist ja Peter Meier  
immer eine Stuppe im Mund.  
Ja ja so ist er schon <sup>immer gewesen</sup> hier ist seine  
Frau Marta Müller ihr ist war-  
scheinlich langweilig wie es aus-  
sieht und hier ist aber ein kömmischer

Mann, den kenne ich doch aber  
dass ist Andi Lehmann. Er ist schon  
immer ein Plöfer gewesen. <sup>mit seinen Stucke</sup>  
<sup>von Peter</sup> Meier wollte von dem Sprungturm  
springen mit der Stuppe nein nein  
Peter Meier dass geht nicht. Andi  
Lehmann wollte zurück schwimmen  
aber er traute nicht, Plöfer trauen  
nie. Noch ein schöner Morgen mit  
Radio Pilatus.



ISH  
Alois Buholzer  
Sandra Zulliger



1. Was stellen Sie fest in Bezug auf den Lernstand der Schülerin A?

a) aufgrund der Notizen

b) aufgrund der beobachteten Videosequenz

ISH  
Alois Buholzer  
Sandra Zulliger



2. Was würden Sie dieser Schülerin rückmelden?

3. Wie würden Sie mit dieser Schülerin weiterarbeiten?

Notizen der Schülerin B

- \* Guten Tag miteinander " wir sind heute in der badie Lido  
~~hier ist es wunder schön das wasser ist 23° und die luft 31°~~
  - \* Heute rede ich mit Stefan Lütli er ist schpizer, Tuhner.  
 er turnt immer vor dem Schwimmen er sagte  
 es sei gesund, die Orientierung
  - \* ich rede heute noch mal mit henn Schreiber  
 er ist rätungs schwimmer er erzelt uns von seinen  
 von der badie Lido und von tagen: it  
 so ... ich beobachte die Leute...
- Der Bistro ist über füllt mit Leuten und  
 sie haben sehr viel stres,  
 und das wars Tüs

ISH  
Alois Buholzer  
Sandra Zulliger



1. Was stellen Sie fest in Bezug auf den Lernstand der Schülerin B?

a) aufgrund der Notizen

b) aufgrund der beobachteten Videosequenz



ISH  
Alois Buholzer  
Sandra Zulliger



2. Was würden Sie dieser Schülerin rückmelden?

3. Wie würden Sie mit dieser Schülerin weiterarbeiten?