

AFRA STURM

Beurteilen und Kommentieren von Texten als fachdidaktisches Wissen

Abstract

Schülertexte angemessen beurteilen und auf dieser Basis für die SchülerInnen ein schriftliches Feedback formulieren zu können, das die SchülerInnen beim Überarbeiten ihrer Texte wirksam unterstützt, setzt seitens der Lehrperson hohes fachdidaktisches Wissen voraus. In diesem Beitrag wird dargelegt, wie fachdidaktisches Wissen im SNF-Projekt „Novizinnen und Novizen im Schreibunterricht (NoviS)“ konzipiert und bei Lehramtsstudierenden Ende Ausbildung sowie Ende des ersten Berufsjahrs erhoben wurde. Die Ergebnisse weisen darauf hin, dass weniger als ein Drittel der (angehenden) Lehrpersonen angemessene Textbeurteilungen verfassen kann. Ähnliches gilt für Kommentare. Zudem zeigt sich im Verlauf des ersten Berufsjahrs keine signifikante Veränderung des fachdidaktischen Wissens.

Einleitung

Bei der Gestaltung von Lehr-Lern-Prozessen kommt der Expertise von Lehrpersonen eine hohe Bedeutung zu: Allgemein wird den Lehrpersonen der größte Einfluss auf die Schülerleistungen attestiert (Hattie 2012). Dabei wird dem fachdidaktischen Wissen die höhere prädiktive Kraft im Hinblick auf Schülerleistungen zugesprochen als dem fachlichen Wissen, so vor allem in der Domäne Mathematik (Kleickmann et al. 2013).

In der Domäne Schreiben liegen erst wenige Studien zur Expertise von Lehrpersonen vor, die hauptsächlich danach fragen, inwiefern die Lehrpersonen nachweislich wirksame Förderansätze anwenden. Die Frage, wie fachliches oder fachdidaktisches Wissen zu konzipieren ist, spielt in diesen Studien eine untergeordnete Rolle.

In diesem Beitrag wird eine Teilstudie aus dem SNF-Projekt „Novizinnen und Novizen im Schreibunterricht (NoviS)“ vorgestellt, in der u. a. fachliches und fachdidaktisches Wissen als Teil des Professionswissens in der Domäne Schreiben erfasst wurde. Dabei werden in erster Linie die Konzipierung der Instrumente zur Erfassung des fachdidaktischen Wissens vorgestellt und ausgewählte Ergebnisse präsentiert.

1 | Fachdidaktisches Wissen und lerntheoretische Aspekte

Bisherige Studien zur Expertise von Lehrpersonen in der Domäne Schreiben erfassten in erster Linie Professionswissen mittels Selbstauskünften.¹ So erfragten Gilbert/Graham (2010) bei Lehrpersonen der Klassen 4–6 insbesondere den Einsatz nachweislich effektiver Schreibinstruktionen wie bspw. die explizite Vermittlung von Schreibstrategien, das Vermitteln von basalen Schreibfähigkeiten (Rechtschreibung und Handschrift), kooperatives Schreiben oder das Nachahmen von Modellen bzw. Mustern. Des Weiteren legten sie den Lehrpersonen eine Auswahl zu möglichen Adaptionen an schwächere SchülerInnen (zusätzliche Schreibzeit, zusätzliches Vermitteln verschiedener Fähigkeiten usw.) sowie eine Auswahl an Schreibaufgaben vor, die vom Ausfüllen von Arbeitsblättern bis zum Verfassen komplexerer Texte wie Erzählungen, Forschungsjournale oder -berichte im Sachunterricht, Zeitungsartikel etc. reichten.

Dass Selbstauskünfte mit Vorsicht zu interpretieren sind, zeigt sich u. a. darin, dass zwischen den Angaben der befragten Lehrpersonen zu den im eigenen Unterricht eingesetzten wirksamen Schreibinstruktionen und den in der Befragung gewählten Schreibaufgaben ein Missverhältnis feststellbar ist: So setzen die Lehrpersonen überwiegend Aufgaben ein, die das Vermitteln von Schreibstrategien gerade nicht bedingen, an erster Stelle Arbeitsblätter, die lediglich Kurzantworten erfordern. Außerdem gaben fast alle Lehrpersonen an, die im Fragebogen aufgeführten effektiven Schreibinstruktionen mehrmals im Jahr anzuwenden, während sie gleichzeitig lediglich ca. 15 Min. pro Tag für die Instruktion aufwenden.² Die nachweislich wirksamen Schreibinstruktionen erfordern Gilbert/Graham (2010) zufolge jedoch mehr Instruktionszeit.

Nicht auszuschließen ist, dass ein solches Antwortverhalten auf soziale Erwünschtheit zurückzuführen ist, wie Gilbert/Graham (2010) selbst anmerken. Darüber hinaus muss auch damit gerechnet werden, dass die befragten Lehrpersonen die zur Auswahl stehenden Praktiken anders verstehen, zumal die einzelnen Praktiken im Fragebogen nicht erläutert wurden.

Kiuhara/Graham/Hawken (2009) befragten Lehrpersonen der Klassen 9–12 ebenfalls zu effektiven Schreibinstruktionen sowie Adaptionen an schwächere SchülerInnen, zusätzlich auch, wie sie die Schreibfähigkeiten ihrer SchülerInnen evaluieren. Dabei gaben ca. 98 % der Sprach-Lehrpersonen an, dass sie Kriterienraster bzw. holistische Skalen einsetzen, sowie weitere 67 %, dass sie professionell beurteilen. Allerdings führen Kiuhara et al. (2009) nicht aus, was unter professionellem Beurteilen zu verstehen ist. Kritisch merken sie an, dass sie nicht erfragt haben, wie häufig evaluiert wird: Entsprechend sei damit zu rechnen, dass die Lehrpersonen – ähnlich wie in Bezug auf Adaptionen an schwache SchülerInnen – eher selten evaluieren würden.

Sowohl Gilbert/Graham (2010) als auch Kiuhara et al. (2009) begründen die Auswahl der erfragten Aspekte damit, dass sie wichtige Pfeiler von Schulreformen darstellen. Des Weiteren erwähnen sie, dass die Wirksamkeit von Schulreformen nicht zuletzt von den Fähigkeiten der Lehrpersonen abhängen. Eine genauere Diskussion zur Frage, wie die Expertise von Lehrpersonen in der Domäne Schreiben konzeptionell zu fassen ist, führen sie jedoch nicht.

Obwohl nicht davon ausgegangen werden kann, dass Selbstauskünfte die tatsächliche Unterrichtspraxis abbilden, stellen sie dennoch ein wichtiges Hilfsmittel dar, um allfällige Inkonsistenzen aufzudecken (Pajares 1992), wie nicht zuletzt obige Diskussion gezeigt hat. Um zuverlässigere Befunde zu erhalten, ist es jedoch unerlässlich, dass Selbstauskünfte mit zusätzlichen Instrumenten bzw. methodischen Zugängen ergänzt werden. Zur Diskussion stehen direkte sowie indirekte Verfahren: Zu ersteren zählen insbesondere Wissenstests wie

¹ Zur Expertise zählen einerseits das Professionswissen, das u. a. fachliches sowie fachdidaktisches Wissen umfasst, andererseits Überzeugungen (vgl. bspw. Kunter/Pohlmann 2009). Zu Überzeugungen in der Domäne Schreiben vgl. Graham et al. (2001).

² Zum Vergleich: Die kantonalen Stundentafeln in der Schweiz sehen für den Deutschunterricht in 4.–6. Klassen ca. 5 Lektionen pro Woche vor. Es kann davon ausgegangen werden, dass an Schweizer Schulen die Unterrichtszeit – Instruktion durch die Lehrperson sowie Schreibzeit für die SchülerInnen – in der Domäne Schreiben zwischen 1–2 Lektionen beträgt und die eigentliche Instruktionszeit damit deutlich tiefer liegt. Vermutlich dürfte die durchschnittliche Instruktionszeit bei Gilbert / Graham (2010) aber auch den fächerübergreifenden Schreibunterricht einschließen.

im Mathematik-Projekt COACTIV (Krauss et al. 2008),³ zu letzteren werden videobasierte Beobachtungsverfahren (Hugener et al. 2006) oder advokatorische Messinstrumente wie Filmvignetten gerechnet (Oser/Heinzer/Salzmann 2010).

Im Projekt NovIS wurden u. a. Wissenstests eingesetzt (vgl. Übersicht in Kap. 2, Tabelle 1): Wird das Professionswissen von Lehrpersonen in Form von Wissenstests erhoben, stellt sich zunächst die Frage, über welches fachdidaktische Wissen Lehrpersonen in der Domäne Schreiben verfügen sollten und wie dieses erfasst werden kann.

1.1 | Fachdidaktisches Wissen als Teil des Professionswissens

Das Professionswissen von Lehrpersonen wird in der Nachfolge von Shulman (1986) hauptsächlich in die folgenden drei Formen differenziert:

- a | fachliches Wissen als Wissen in der jeweiligen Unterrichtsdisziplin
- b | fachdidaktisches Wissen als Wissen über die fachbezogene Gestaltung von Lehr-Lern-Prozessen
- c | curriculares Wissen als Wissen über Lehrplaninhalte und passende Lernmaterialien

In der Wissenstaxonomie von Shulman (1986, 9) wird fachdidaktisches Wissen als „pedagogical content knowledge“ bezeichnet, wobei es in erster Linie um die Lehrbarkeit der zu vermittelnden Inhalte geht. Gemeint ist damit die Art und Weise, wie Inhalte präsentiert und formuliert bzw. erklärt werden, sodass sie für die SchülerInnen versteh- und lernbar sind. Das schließt in der Domäne Schreiben seitens Lehrpersonen ein Wissen über nachweislich wirksame Vermittlungsarten sowie ein Wissen über zielführende instruktionale Adaptionen in Bezug auf schwächere SchülerInnen ein, wie dies auch in Gilbert/Graham (2010) sowie Kiuhara et al. (2009) der Fall ist.

Insgesamt ist fachdidaktisches Wissen bei Shulman (1986) noch sehr allgemein gefasst: Eine fachbezogene Spezifizierung ist entsprechend unumgänglich. In der Domäne Mathematik wird fachdidaktisches Wissen ausgehend von einer lerntheoretischen Auffassung näher bestimmt, die in dieser Domäne allgemein akzeptiert sein dürfte und die davon ausgeht, dass SchülerInnen „im Fachunterricht im Kontext von fachlich gehaltvollen Lernumgebungen ›kognitiv angeregt‹ werden“ sollen (Krauss et al. 2008, 232). Entsprechend dieser konstruktivistischen Auffassung wird mathematikdidaktisches Wissen in drei Wissenskomponenten aufgefächert (Krauss et al. 2008, 234):

- | b1 Wissen über das Verständlichmachen von mathematischen Inhalten
- | b2 Wissen über mathematikbezogene Schülerkognitionen
- | b3 Wissen über das kognitive Potenzial von Mathematikaufgaben

In der Domäne Schreiben ist im deutschsprachigen Raum seit den 80er-Jahren der prozessorientierte Ansatz stark vertreten (Fix 2006) und stellt auch in der kompetenzorientierten Schreibdidaktik ein zentrales Element dar (Becker-Mrotzek/Böttcher 2012). Der prozessorientierte Ansatz geht mit der lerntheoretischen Auffassung eines eher gemäßigten Konstruktivismus einher:

Sprachunterricht versteht sich in diesem Sinne nicht als Instruktion, sondern er soll die Lerner dabei unterstützen, Wissen in einem aktiven und selbstständigen Konstruktionsprozess aufzubauen. Dieser kognitive Begründungsansatz liefert Argumente für einen Unterricht, der den Schwerpunkt auf den Verlauf der individuellen Aneignung verlagert. (Fix 2006, 119)

Damit grenzt sich dieser Ansatz von (direkter) Instruktion ab, die auf einer stärker lenkenden Rolle der Lehrperson beruht und die u. a. darin besteht, dass die Lehrperson ein schrittweises Vorgehen modelliert und dabei Phasen mit angeleitetem und zunehmend selbstständigerem Üben einplant. Die bereits zitierten Studien von Gilbert/Graham (2010) sowie Kiuhara et al.

³ COACTIV steht für „Cognitive Activation in the Classroom“.

(2009) stehen stellvertretend für direkte Instruktion, zumal sie bei den instruktionalen Praktiken jene zur Auswahl vorgeben, die sich in verschiedenen Meta-Analysen im Vergleich zu anderen Förderansätzen positiver auf die Schülerleistungen auswirkten, zuvorderst die explizite Vermittlung von Schreibstrategien, die auf direkter Instruktion basiert. Dieser Ansatz wurde im deutschsprachigen Raum auf empirischer Basis erstmals von Glaser/Keßler/Brunstein (2009) umgesetzt.

Der fachbezogenen Gestaltung von Lehr-Lern-Prozessen in der Domäne Schreiben liegen damit verschiedene lerntheoretische Auffassungen zugrunde. Auf diesem Hintergrund ist zunächst unklar, wie analog zu COACTIV (Krauss et al. 2008) in der Domäne Schreiben fachdidaktisches Wissen ausdifferenziert werden kann. So dürften das „Verständlichmachen von Schreibprozessen“ (vgl. b1) sowie insbesondere das lernförderliche Potenzial von Schreibaufgaben (vgl. b3) je nach Ansatz sehr unterschiedlich definiert und eingeschätzt werden.

Dennoch kann nicht zuletzt auf empirischer Basis davon ausgegangen werden, dass es sich bei der direkten Instruktion und dem (gemäßigten) Konstruktivismus durchaus um komplementäre Ansätze handelt (Lipowsky 2009, 80). So gilt etwa, dass Lernende mit günstigen Voraussetzungen eher von offenen bzw. prozessbezogenen Ansätzen, Lernende mit ungünstigen Voraussetzungen dagegen eher von direkter Instruktion profitieren. Schreibenlernen ist, wie Harris/Graham (1994, 245) betonen, zu komplex, als dass ein einzelner instruktionaler Ansatz alles abdecken könnte. Fachdidaktisches Wissen wäre entsprechend so zu konzipieren und zu erfassen, dass dem Rechnung getragen wird.

Hinzu kommt, dass anders als im Fach Mathematik prozessbezogenes um textproduktbezogenes fachdidaktisches Wissen zu ergänzen wäre: Sowohl in der prozessorientierten Schreibdidaktik wie auch in Förderansätzen, die der direkten Instruktion zuzuordnen sind, spielt Genre- bzw. Textsortenwissen eine zentrale Rolle. Gemeint ist auch hier die Art und Weise, wie Genre- bzw. Textsortenwissen zu vermitteln sind, sodass es für die SchülerInnen versteh- und lernbar ist. Anders formuliert: Das „Verständlichmachen von schreibbezogenen Inhalten“ (b1) wäre entsprechend in das Verständlichmachen von Schreibprozessen und Textsortenmerkmalen zu differenzieren.

Wirksame Förderung als Element von b1 schließt auch formatives Beurteilen ein, das der Steuerung des Lernprozesses auf ein bestimmtes Lernziel hin dient (Hattie/Timperley 2007). Gerade auch in der Domäne Schreiben zeigt formatives Beurteilen nachweislich positive Effekte auf die Schreibleistung der SchülerInnen (Graham/Harris/Hebert 2011). Um Adaptationen an schwächere SchülerInnen zielführend vornehmen zu können, müssen Lehrpersonen nicht zuletzt in der Lage sein, den Lernstand der SchülerInnen angemessen einschätzen zu können, und zwar sowohl aus einer Prozess- wie auch Produktperspektive.

Aus schreibdidaktischer Sicht kommt dem Beurteilen von Textprodukten ein hohes Gewicht zu (Jost et al. 2011; Kiuvara et al. 2009). Aus diesen Gründen wurde das Beurteilen von Textprodukten im Projekt NoviS zum fachdidaktischen Wissen gerechnet, auch wenn aus empirischer Perspektive noch nicht geklärt ist, ob diagnostische Kompetenz einen Aspekt des fachdidaktischen Wissens, eine Kombination aus fachlichem und fachdidaktischem Wissen oder vielmehr eine tatsächliche abgrenzbare Fähigkeit darstellt (Kunter/Pohlmann 2009, 265).

Da fachdidaktisches Wissen – wie zu Beginn dieses Kapitels erwähnt – auch die Art und Weise umfasst, wie Inhalte präsentiert und formuliert werden, stellt sich die Frage danach, wie Lehrpersonen ihre Einschätzung von Texten den Lernenden gegenüber kommunizieren. Schriftliche Kommentare von Lehrpersonen können denn auch auf eine lange Tradition im Schreibunterricht zurückblicken. Aus diesen Gründen wurde im Projekt NoviS auch das schriftliche Kommentieren von Lehrpersonen dem fachdidaktischen Wissen zugeordnet.

Im Hinblick auf die schreibbezogenen Schülerkognitionen (in Analogie zu b2 oben) sei angemerkt, dass sich die Merkmale schwacher SchreiberInnen sehr allgemein beschreiben lassen: Sie planen kaum, generieren weniger Inhalte, messen Schreibstrategien keinen Wert bei, brechen den Schreibprozess frühzeitig ab, verfügen oftmals auch über weniger gut ausgebildete basale Schreibfähigkeiten sowie geringeres Textsortenwissen usw. (vgl. bspw. Troia 2006). Entsprechend wurde Wissen darüber im Projekt NoviS zum fachlichen Wissen gerech-

net (Sturm et al. i. Dr.). Schwierigkeiten im Textproduktionsprozess können sich denn auch bei ganz verschiedenen Aufgaben zeigen.

Je nach Fragestellung und allenfalls auch aus pragmatischen Gründen muss ein Schwerpunkt gelegt werden. Im Projekt NoviS lag der Fokus beim fachdidaktischen Wissen auf dem formativen Beurteilen, und zwar lediglich bezogen auf Textprodukte. Der Schreibprozess dagegen wurde bei der Erfassung von fachlichem Wissen berücksichtigt (Sturm et al. i. Dr.). Werden wie im Projekt NoviS nicht alle Facetten des fachdidaktischen Wissens, das hier im Zentrum steht, erhoben, gilt es, dies bei der Interpretation der Ergebnisse zu berücksichtigen.

1.2 | Forschungsstand zum formativen Beurteilen in der Domäne Schreiben

Die Meta-Analyse von Graham et al. (2011, 17) weist für formatives Beurteilen in der Domäne Schreiben einen moderaten Effekt von $d = .77$ auf die Schreibleistung aus. Allerdings umfassen die darin versammelten Studien sowohl Feedback von Erwachsenen (Lehrpersonen und Eltern) wie auch von Peers. Ähnlich weisen Graham et al. (2012) nur für Feedback durch Erwachsene einen Effekt von $d = .80$ aus. Entgegen der Erwartung liegen jedoch erst wenige Studien zum Beurteilen von Textprodukten mittels Kriterienraster vor. Ein möglicher Grund dafür kann darin gesehen werden, dass diese Form des Beurteilens häufiger summativ angelegt sein dürfte. In ihrer Meta-Analyse halten Koster et al. (2015, 268) fest, dass Beurteilen bzw. Feedback durch verschiedene Akteure erfolgen und auch produkt- oder prozessbezogen sein kann: Um besser einschätzen zu können, wie eine bestimmte Art von Feedback wirke, brauche es zusätzliche Forschung.

Betrachtet man die wenigen Studien zum lehrerseitigen kriterienorientierten Beurteilen von Textprodukten isoliert, ist ein positiver Effekt auf die Schreibleistung nicht nachweisbar (vgl. ausführlicher Sturm 2014). Das kann an der Qualität der verwendeten Raster liegen – eingesetzt wurde das Raster „6+1“, das die Kriterien Ideen, Organisation, Stil, Wortwahl, Satzflüssigkeit und Konventionen (Rechtschreibung und Grammatik) vorsieht –, aber auch an der Qualität der Beurteilung selbst. Dem hält Collopy (2008) entgegen, dass ein solches Raster die Lehrpersonen u. a. dabei unterstützen würde, den Schwerpunkt nicht nur auf sprachformale Aspekte zu legen. Das setzt allerdings voraus, dass die Lehrpersonen über ein adäquates Verständnis der einzelnen Kriterien verfügen.

Schwächere SchülerInnen fokussieren beim Überarbeiten ihrer Texte hauptsächlich Oberflächenaspekte wie Rechtschreibung oder Handschrift und nehmen Revisionen auf der lokalen Ebene vor, das heißt auf Wort- und Satzebene (Steendam et al. 2014). Sollen auch schwächere SchülerInnen zunehmend in der Lage sein, einen Text auf der globalen Ebene bzw. auf der eigentlichen Textebene zu überarbeiten, müssen sie auch sogenannte höhere Aspekte wie Inhalt oder Struktur wahrnehmen können.⁴ Entsprechend setzt dies voraus, dass auch Lehrpersonen globale Aspekte erkennen und angemessen einschätzen. Koster et al. (2015) etwa betonen, dass Lehrpersonen in den Niederlanden nicht genügend auf den Schreibunterricht vorbereitet werden und dass die verfügbaren Lehrmittel die Lehrpersonen nicht ausreichend unterstützen würden, gerade auch im Hinblick auf das formative Beurteilen.

Da Graham et al. (2011) in ihre Meta-Analyse nur gerade eine Studie zu schriftlichen Kommentaren aufnehmen konnten, die zudem nur einen geringen positiven Effekt auf Schülerleistungen ausübte, können sie zu dieser Form des Beurteilens, die sie als sehr weit verbreitet ansehen, keine Aussagen ableiten. Parr/Timperley (2010) weisen in ihrer Studie einen mittleren Zusammenhang zwischen der Qualität des Kommentars und den Schreibleistungen der SchülerInnen nach und zeigen dabei gleichzeitig auf, dass eine Weiterbildung dazu füh-

⁴ Oberflächenmerkmale werden häufig als „lower-order concerns“ (= LOCs) und globale Aspekte als „higher-order concerns“ (= HOCs) bezeichnet (Steendam et al. 2014, 909).

ren kann, dass Lehrpersonen stärker globale Aspekte fokussieren und die lokale Ebene etwas zurückstellen.

Kommentare von Lehrpersonen sind zudem häufig, wie Cho/MacArthur (2010) zusammenfassend festhalten, nicht im jeweiligen Schülertext verankert, ambig und manchmal auch widersprüchlich. Ambiges und nicht im Schülertext verankertes Feedback kann auch auf fehlendes fachdidaktisches Wissen zurückgeführt werden: In solchen Fällen tendieren Lehrpersonen dazu, in ihrem Feedback lediglich allgemeine Prinzipien zu formulieren, die für die SchülerInnen nicht operationalisierbar sind (Myhill/Jones/Watson 2013).⁵ Dazu zählen etwa Hinweise, dass die Satzanfänge variiert werden sollten oder dass das Hinzufügen von Adjektiven einen Text spannender mache. Dagegen können Lehrpersonen mit hohem fachdidaktischem Wissen genauer aufzeigen, welche Wirkung im Text bspw. eine Veränderung der Wortfolge haben kann.

Nicht zuletzt haben Kommentare, die direktiv formuliert sind, im Vergleich zu dialogisch angelegten Kommentaren einen geringeren Effekt auf die Überarbeitung der Texte (MacArthur 2015, 275). Des Weiteren stellen Bouwer/van den Bergh (2014) fest, dass Lehrpersonen zu direktiven Kommentaren neigen und dass das Feedback (inkl. Kommentar) weniger eine Frage der Qualität des Schülertextes sei, sondern eher eine Frage des Feedback-Stils.

Wie Bouwer et al. (2015) in ihrer Studie zudem zeigen, unterscheiden sich Lehrpersonen mit einer mindestens fünfjährigen Berufserfahrung hinsichtlich Qualität der Textbeurteilung nicht von Lehramtsstudierenden. Erstere beurteilen jedoch Texte präziser und vor allem konsistenter über verschiedene Aufgaben hinweg.

2 | Methode

Das Projekt „Novizinnen und Novizen im Schreibunterricht (NoviS)“ (2013–2015), das vom Schweizerischen Nationalfonds (SNF) gefördert wurde, ging im Rahmen zweier Teilstudien zwei Hauptfragen nach:

- | *Teilstudie A:* Welches fachliche und fachdidaktische Wissen und welche Überzeugungen bringen angehende Primar-Lehrpersonen für den Schreibunterricht mit und wie verändern sich diese im Verlauf des ersten Berufsjahrs?
- | *Teilstudie B:* Welche Unterrichtspraxen wenden Novizen und Novizinnen in ihrem Schreibunterricht an und wie begründen sie diese? Lassen sich zwischen der Expertise (Wissen und Überzeugungen) und dem Unterricht Zusammenhänge darstellen?

In diesem Beitrag wird ein Ausschnitt aus der Teilstudie A fokussiert, und zwar das fachdidaktische Wissen (vgl. Tabelle 1). Analog zu Krauss/Blum (2012) wurde in NoviS von der Hypothese ausgegangen, dass die Lehrerfahrung keine Auswirkung auf das Ausmaß an fachdidaktischem Wissen hat.

Teilstudie A	
N = 99 (t ₁ : Ende Ausbildung) / 51 (t ₂ : Ende erstes Berufsjahr)	
Instrument	Erfasst
A1 Offene Aufgabe zum eigenen Schreibprozess	
A2 Offene Aufgabe zum Schreibprozess eines schwachen Primarschülers	fachliches Wissen
A3 Beurteilung eines Schülertexts	
A4 Verfassen eines Kommentars an den Schüler zu seinem Text	fachdidaktisches Wissen

⁵ Myhill et al. (2013) sprechen genau genommen von fachdidaktischem Grammatikwissen, meinen damit aber nicht traditionelles Grammatikwissen, sondern sprachliches Wissen im Kontext der Textproduktion.

A5	Selbstwirksamkeitsskala zum eigenen Schreiben	
A6	Selbstwirksamkeitsskala zum Schreibunterricht	Überzeugungen
A7	Skala zu schreiblerntheoretischen Orientierungen	

Tabelle 1: Überblick Erhebungsinstrumente Teilstudie A und Stichprobe

Die Befragung wurde online durchgeführt: Auf diese Weise konnten die Probanden innerhalb eines festgesetzten zeitlichen Rahmens von ca. drei Wochen die Befragung von zuhause aus durchführen.⁶ Eingesetzt wurde dazu das Online-Tool Toss, das einen Texteditor enthält und von der Pädagogischen Hochschule der Fachhochschule Nordwestschweiz (FHNW) zur Schreibkompetenzabklärung bei erstsemestrigen Studierenden verwendet wird. Für die Befragung im Rahmen von NoviS wurde zusätzlich ein Online-Fragebogen integriert.

2.1 | Stichprobe

Befragt wurden Studierende der Pädagogischen Hochschule der FHNW, die den Studiengang Primarstufe absolvierten. Der erste Messzeitpunkt (t_1) fand im Mai 2013 im Abschlusssemester statt, der zweite Messzeitpunkt im Juni/Juli 2014 am Ende des ersten Berufsjahres (t_2).

An der ersten Befragung nahmen 116 Studierende teil. Davon schlossen 99 die Befragung ab. Die gesamte Rücklaufquote beträgt 46 %. Zu t_2 beteiligten sich noch 54 Lehrpersonen, wovon 3 die Befragung vorzeitig beendeten. Für jede Befragung mit einer Dauer von ca. eineinhalb Stunden wurde den Probanden jeweils ein Honorar von 50 CHF ausbezahlt.

In der Schweiz machen die Frauen die Mehrzahl der Lehrpersonen aus, insbesondere auch auf der Primarstufe. Dies spiegelt sich im Sample wider, in welchem 81 % (t_1) bzw. 89 % (t_2) weiblichen Geschlechts sind. Die Studierenden waren zu t_1 zwischen 21 und 51 Jahre alt, wobei das Durchschnittsalter bei 25.7 Jahren ($SD = 5.13$) lag. Mit Ausnahme einer Lehrperson sprechen alle Befragten Deutsch als Erstsprache. Der schulische bzw. berufliche Hintergrund der Probanden zu t_1 – vor Abschluss ihrer Lehramtsausbildung – ist unterschiedlich: Als höchste bislang abgeschlossene Ausbildung verfügen 80.2 % über eine Matura, 7.8 % über einen Hochschulabschluss, 1.7 % über eine höhere Fach- bzw. Berufsbildung, 5.2 % über eine Berufslehre und 5.2 % über ein anderweitiges Lehrdiplom.

2.2 | Instrument und Durchführung

Fachdidaktisches Wissen wurde mit zwei Aufgaben erfasst: Abb. 1 zeigt die Aufgabe zur Beurteilung eines Schülertextes und Abb. 2 den konkreten Schülertext. Insgesamt erhielten die Probanden lediglich minimale Kontextangaben. So wurde darauf verzichtet, genauer auszuführen, welche Instruktion mit der Schreibaufgabe verbunden war oder welche Schreibleistungen der Schüler davor erbrachte, um auszuschließen, dass die Textbeurteilung durch Einstellungen zur Instruktion oder zum Schüler beeinflusst wird.

Tim, ein 4.-Klässler mit Schweizerdeutsch als Muttersprache, hatte die Aufgabe, eine spannende Geschichte zu schreiben. Dazu hatte er eine Lektion Zeit.

Bitte beurteilen Sie den Text und begründen Sie Ihre Einschätzung des Textes so, dass Ihre Beurteilung z. B. für andere Lehrpersonen nachvollziehbar ist. Nennen Sie insbesondere auch die Kriterien, welche für Sie wichtig sind.

Abb. 1: Aufgabe A3, einen Schülertext beurteilen

⁶ Der Zugang erfolgte zwar über die Reflexionsseminare im Rahmen der berufspraktischen Studien, welche einen obligatorischen Bestandteil der Ausbildung zur Primarlehrperson bilden: Die Befragung selbst konnte aber nicht während des Seminars erfolgen.

Des Weiteren wurde den Probanden kein bestimmtes Kriterienraster vorgegeben: Auf diese Weise kann erfasst werden, welche Kriterien die Probanden zur Textbeurteilung heranziehen, insbesondere ob sie in der Lage sind, von sich aus globale Aspekte einzubeziehen und auch adäquat einzuschätzen. Dazu muss der Schülertext aber auch entsprechende Merkmale aufweisen: Der Schülertext stammt aus dem Projekt MyMoment (vgl. Furger/Schneider 2011) und wurde so ausgewählt, dass er auf der lokalen Ebene – v. a. Rechtschreibung und Grammatik – sowie auf der globalen bzw. eigentlichen Textebene – Kohärenz, Themenentfaltung etc. – Schwierigkeiten aufzeigt. Damit kann bspw. eruiert werden, ob die Probanden bei ihrer Beurteilung allenfalls einen einseitigen Schwerpunkt setzen.

Der Junge und das Monster

Der Junge malt ein Bild.

Er ist gleich fertig aber plötzlich springt das gemalte Monster raus und sagte: „Warum hab kine Arme.“

Der Junge erschrickt: „Hö warum Lebst du? Dan sagt das Monster: „Ich darf doch Leben es ist doch eine freie Welt.“ Der Junge ist verwirrt und er sagt: „Du bist doch nur eine Zeichnung.“ Das Monster sagt: „Wir können doch Freunden sein?“ Der Jung antwortet komisch: „Ja das finde ich gut.“ Und dann sind sie Freunde und bleiben für immer und ewig. Und wen sie noch nicht Gestorben sind dann Leben sie noch heute.

Abb. 2: Schülertext zu Aufgabe A3

Im Anschluss an die Textbeurteilung wurden die Probanden aufgefordert, zuhanden desselben Schülers einen Kommentar zu verfassen (vgl. Abb. 3). Zwar muss im Online-Tool Toss jeweils eine Aufgabe abgeschlossen werden, bevor eine weitere Aufgabe gelöst werden kann, die Probanden konnten jedoch die zuvor abgeschlossene Aufgabe ohne Weiteres einsehen.

Sie haben soeben die Geschichte von Tim, einem 4.-Klässler mit Schweizerdeutsch als Muttersprache, beurteilt.

Schreiben Sie Tim nun einen Kommentar zu seinem Text und geben Sie ihm dabei auch Tipps, wie er die Geschichte noch besser schreiben könnte.

Abb. 3: Aufgabe A4, einen schriftlichen Kommentar an den Schüler verfassen

Um einen Einblick zu erhalten, ob die Probanden ihren schriftlichen Kommentar eher direktiv oder dialogisch formulieren, wurden sie gebeten, Tipps zu formulieren, wie der Schüler seinen Text besser schreiben könnte. Der Kommentar dient damit einer formativen Beurteilung und setzt voraus, dass der Schüler auch die Gelegenheit erhält, seinen Text zu überarbeiten.

2.3 | Auswertung

Die Auswertungskategorien – vgl. den Überblick in Tabelle 2 unten – wurden basierend auf dem Forschungsstand, wie er in Kapitel 1.2 dargelegt wurde, festgelegt. Die Kategorien „Ebene“, „Wertung“ und „Textbezug“ haben eine deskriptive Funktion und konnten mit ihren Ausprägungen für die Textbeurteilung und den Kommentar identisch definiert werden. Die Qualität wurde für die beiden Testaufgaben mit spezifischen Kategorien erfasst: mit „Angemessenheit der Ebene“ für die Textbeurteilungen und mit „Tipps“ und „Form“ für die Kommentare.

Textbeurteilung und Kommentare	
<i>Kategorie</i>	<i>Ausprägungen</i>
<i>Ebene</i>	<i>lokal</i> : Fokus eher auf Wort-/Satzebene (v. a. Orthografie, Grammatik, Wortschatz, Umfang) <i>global</i> : Fokus eher auf Textebene (v. a. Gesamtidee, Kohärenz, Themenentfaltung, Erzählmuster, Aufbau) <i>gemischt</i> : Fokus gleichermaßen auf lokaler und globaler Ebene
<i>Wertung</i>	<i>negativ</i> : v. a. nicht Gelungenes im Fokus <i>positiv</i> : v. a. Gelungenes im Fokus <i>gemischt</i> : Gelungenes + nicht Gelungenes gleichermaßen im Fokus
<i>Textbezug</i>	<i>vorhanden vs. nicht vorhanden</i>
Qualität Textbeurteilungen	
<i>Angemessenheit lokale Ebene</i>	<i>erfüllt vs. nicht erfüllt</i>
<i>Angemessenheit globale Ebene</i>	<i>erfüllt vs. nicht erfüllt</i>
Qualität Kommentare	
<i>Tipps</i>	<i>verankert vs. nicht verankert</i>
<i>Form</i>	<i>direktiv</i> : Belehrung, Anweisung <i>dialogisch</i> : subjektiv und fragend <i>gemischt</i> : gleichermaßen direktiv + dialogisch formuliert

Tabelle 2: Überblick Auswertungskategorien

Im Folgenden werden die deskriptiven Kriterien und die Qualitätskriterien i. e. S. separat ausgeführt. Daran anschließend werden das Vorgehen und die Interrater-Reliabilität beschrieben.

Kodierregeln deskriptive Kategorien

Da sowohl die Beurteilungen wie auch die Kommentare im Textumfang stark variierten, erfolgte die Kodierung nicht auf der Basis der Rohzahlen von Aussagen, die einer Kategorie zuzuordnen sind, sondern basierend auf der überwiegenden Anzahl bestimmter Aussagentypen. So erhielt ein Text – Beurteilung oder Kommentar – bezogen auf die Kategorie „Ebene“ die Ausprägung „lokal“, wenn die überwiegende Anzahl der entsprechenden Aussagen lokal waren, das heißt, wenn der Text mindestens zwei lokale Stellen mehr enthielt als globale (und umgekehrt). Mit „gemischt“ wurde kodiert, wenn sich lokale und globale Aspekte die Waage hielten.

Gleiches gilt für die Kategorie „Wertung“: Wenn ein Text mindestens zwei negative Aussagen mehr enthielt als positive, erhielt er die Ausprägung „negativ“ (und umgekehrt); wenn der Unterschied geringer war, wurde „gemischt“ zugewiesen. Zusätzlich wurde festgelegt, dass in einem Text auch Stellen als negative Aussagen identifiziert wurden, die nicht Gelungenes implizit formulierten („Ich würde dem Jungen sagen, dass er, auch wenn er in direkter Rede schreibt, trotzdem auf Hochdeutsch schreiben soll“, G04-1132, t₁).

In Bezug auf die Kategorie „Textbezug“ enthielt das Manual die Anweisung, dass mindestens zwei Textbezüge vorliegen sollten, wobei mindestens ein Textbezug der globalen Ebene zuzuordnen sein sollte. In solchen Fällen erhielt ein Text die Ausprägung „vorhanden“, ansonsten „nicht vorhanden“.

Wie in Kap. 1.2 dargelegt wurde, erzielt ein Feedback, das die globale Ebene fokussiert, die bessere Wirkung, da so bei der Überarbeitung entsprechend auch die Textqualität und nicht nur Oberflächenmerkmale im Vordergrund stehen (ein eher globaler Fokus schließt

nicht aus, dass die Beurteilung oder der Kommentar auch lokale Aspekte aufführt). Damit kann bei den Ausprägungen die Rangfolge *global* > *gemischt* > *lokal* angenommen werden. Bei der Kategorie „Wertung“ kann jedoch keine Rangfolge theoretisch abgeleitet werden. Dagegen gilt für Feedback mit Textbezug, dass diese für die SchülerInnen besser nachvollziehbar sind.

Kodierregeln Qualität

Um einschätzen zu können, ob es sich bei den Textbeurteilungen um angemessene Beurteilungen handelt, wurde der Schülertext acht ExpertInnen zur Beurteilung vorgelegt.⁷ Zum einen erhielten die ExpertInnen zunächst die gleiche Aufgabe wie die Probanden, zum anderen sollten sie den Schülertext zusätzlich mit einem vorgegebenen Raster evaluieren, das sowohl lokale als auch globale Aspekte enthält. Während Oser et al. (2010) ein Expertenurteil im Konsensverfahren generierten, wurde im Projekt NovIS aus der Beurteilung mit dem vorgegebenen Raster der Einfachheit halber der Durchschnittswert aller ExpertInnen für die weitere Auswertung zugrunde gelegt und mit ihren offenen Antworten abgeglichen.

Die Qualität der Textbeurteilungen wurde bezogen auf die lokale und globale Ebene separat eingeschätzt: Für die globale Ebene wurde festgelegt, dass Beurteilungen mit „erfüllt“ zu kodieren sind, wenn der Schülertext in Übereinstimmung mit den ExpertInnen auf globaler Ebene insgesamt als nicht ganz erreicht, in Bezug auf einzelne Kriterien wie Themenentfaltung als nicht erreicht, hinsichtlich Erzählmuster sowie Gliederung in Erzählanlass, Ereignisse und Erzählabschluss als erreicht eingeschätzt wurde. Da nicht damit gerechnet werden konnte, dass die Textbeurteilungen der Probanden alle relevanten Kriterien enthalten, wurden die Textbeurteilungen als erreicht eingeschätzt, wenn mindestens drei globale Kriterien adäquat ausgeführt waren. Ähnliches galt für die lokale Ebene: Hinsichtlich der Kriterien Wortwahl und Rechtschreibung wurde eine Beurteilung mit „erfüllt“ kodiert, wenn der Schülertext als erreicht, hinsichtlich Satzbau als nicht ganz erreicht eingestuft wurde. Da die Probanden mehrheitlich kein Gesamturteil formulierten – sei es in Form einer Ziffernote oder umschreibend –, wurde dieser Aspekt bei der Auswertung nicht weiter berücksichtigt.

Die Qualität der Kommentare bestimmt sich im Wesentlichen durch Tipps sowie die Art und Weise, wie das Feedback und die Tipps den SchülerInnen kommuniziert werden. Da allgemeine Tipps, die sich nicht direkt auf den Schülertext beziehen, für die SchülerInnen nicht operationalisierbar sind (Myhill et al. 2013), erhielt ein solcher Kommentar – analog zu Kommentaren ohne Tipps – bei der Kategorie „Tipps“ die Ausprägung „nicht verankert“ zugewiesen. Auch Tipps, die zwar einen Bezug zum Schülertext aufweisen, dennoch nicht nachvollziehbar sind, wurden als „nicht verankert“ kodiert. Hinsichtlich der Kategorie „Form“ wurde basierend auf Bouwer/van den Bergh (2014) zwischen dialogisch und direktiv formulierten Kommentaren unterschieden: Kommentare, die mehrheitlich in Bezug auf allgemeine Normen sowie belehrend formuliert waren, erhielten die Ausprägung „direktiv“ („weiter sollst du alle Verben blau unterstreichen und schauen, dass du möglichst viele verschiedene Verben brauchst“). Waren Bewertungen dagegen mehrheitlich subjektiv formuliert und wurde der Schüler v. a. durch Fragestellungen zum Überarbeiten angeregt, wurde „dialogisch“ kodiert. Hielten sich dialogische und direktive Formulierungen die Waage, wurde mit „gemischt“ kodiert. Da sich dialogische Kommentare positiver auf die Überarbeitung der Texte auswirken (MacArthur 2015), wurde die Rangfolge *dialogisch* > *gemischt* > *direktiv* festgelegt.

Vorgehen

Der erste Kodierschritt bestand darin, festzulegen, welches die zu kodierenden thematischen Einheiten sind. Erst in einem zweiten Schritt wurden die Kategorien mit den entsprechenden Ausprägungen kodiert, wobei die Qualität zuletzt kodiert wurde.

⁷ Die ExpertInnen arbeiten am Zentrum Lesen der Pädagogischen Hochschule FHNW oder stammen aus dessen Umfeld.

Zu den Auswertungskategorien wurde für beide Aufgaben je ein Manual erstellt, das sowohl das Vorgehen genau festlegt als auch die einzelnen Kategorien mit ihren Ausprägungen detailliert beschreibt. Dabei wurde jede Kategorie und Ausprägung mit Ankerbeispielen unterlegt. Zusätzlich wurden in Kodierregeln auch Hinweise formuliert, wie in Zweifelsfällen zu kodieren ist.

Das folgende Beispiel illustriert, wie die thematischen Einheiten geklammert wurden. Festgelegt wurde etwa in Bezug auf Rechtschreibung, dass jeder Regelbereich als thematische Einheit zu kodieren ist. Wurde ein Aspekt bei der Textbeurteilung mehrmals aufgegriffen, wurde er ungeachtet der Wiederholung kodiert.

[Die Geschichte ist etwas kurz geworden,] [aber einfallsreich]. [Die Frage welche das Monster am Anfang stellt, gefällt mir gut und ist lustig.] [Den Abschlusssatz finde ich toll, da er dies aus Märchen übernommen hat. Es entsteht eine besondere Mischung. Einerseits die Zeichnung welche sich zum Monster verwandelt ist eine Phantasiegeschichte und andererseits der Schluss mit der Märchenfloskel.] [Der erste Teil der Geschichte ist etwas kurz und kann etwas ausgeschmückt werden.]

[Im Bezug auf die Rechtschreibung: Ich glaube, da haben sich einige Flüchtigkeitsfehler eingeschlichen und bei einem so kurzen Text und der Zeitdauer sollten dies nicht so oft passieren.] [Die Verdoppelungen sollten mit Tim auch noch geübt werden.] [Spannend ist die Geschichte nicht so sehr, da es keinen Höhepunkt oder plötzliche Wendung hat.]

[...]

Abb. 4: Textbeurteilung von G01-1093, t₁

Interrater-Reliabilität

Sowohl von den Textbeurteilungen wie auch von den Kommentaren wurden je insgesamt 70 zufällig ausgewählte Texte aus beiden Messzeitpunkten von zwei unabhängigen RaterInnen kodiert. Als Maß für die Reliabilität wurde Cohen's Kappa gewählt (Cohen 1968). Für die Textbeurteilung konnten für alle Kategorien befriedigende (Werte über 0.6) bis sehr gute Werte erreicht werden, für den Kommentar lag der Wert für den Textbezug leicht unter 0.6 bei 0.586. Da die beiden Rater lediglich in zwei von 70 Fällen zu einer unterschiedlichen Einschätzung gelangten, wurde dieser Wert als befriedigend erachtet.

3 | Ergebnisse

Im Folgenden werden die Vorkommen der deskriptiven Kategorien in den Textbeurteilungen und Kommentaren näher beschrieben sowie Gemeinsamkeiten und Unterschiede in ihrer Verteilung ausgeführt. Daran anschließend werden die Ergebnisse zur Qualität der Textbeurteilungen und der Kommentare berichtet und allfällige Veränderungen im Verlauf des ersten Berufsjahrs dargestellt.

Textbeurteilung und Kommentar – Gemeinsamkeiten, Unterschiede

Durchschnittlich umfassen die Textbeurteilungen zu t₁ rund 962 (*SD* = 497.93) und zu t₂ rund 908 Zeichen (*SD* = 492.50). Die Kommentare fallen im Vergleich zu den Textbeurteilungen mit 631 zu t₁ (*SD* = 314.73) und 596 zu t₂ (*SD* = 238.34) um einen Drittel kürzer aus. Der Umfang der Texte variiert beträchtlich, worauf auch die hohen Standardabweichungen deuten: So umfasst die kürzeste Textbeurteilung 163 und die längste Textbeurteilung 2'739 Zeichen, der kürzeste Kommentar 116 und der längste Kommentar 2'318 Zeichen.

Tabelle 3 umfasst die Kategorien, die der Auswertung beider Texte zugrunde lagen. Sie ist so zu lesen, dass bspw. 30,3 % der Probanden zu t_1 bei der Beurteilung des Schülertextes eher einen Fokus auf die globale Ebene gelegt haben.

	Textbeurteilung		Kommentar	
	t_1 (N = 99)	t_2 (N = 49)	t_1 (N = 99)	t_2 (N = 49)
<i>Ebene</i>				
Fokus eher global	30,3 %	34,7 %	37,4 %	34,7 %
Fokus gemischt	31,3 %	38,8 %	55,6 %	55,1 %
Fokus eher lokal	38,4 %	26,5 %	7,1 %	10,2 %
<i>Wertung</i>				
Fokus eher positive Aspekte	27,3 %	28,6 %	16,2 %	6,1 %
Fokus gemischt	41,4 %	32,7 %	80,8 %	87,8 %
Fokus eher negative Aspekte	31,3 %	38,8 %	3 %	6,1 %
<i>Textbezug</i>				
vorhanden	28,3 %	32,7 %	22,2 %	10,2 %

Tabelle 3: Textbeurteilung und Kommentar: prozentuale Verteilung der deskriptiven Kategorien

Insgesamt zeigt Tabelle 3, welche Schwerpunkte Probanden am Ende ihrer Ausbildung (t_1) und am Ende des ersten Berufsjahr (t_2) legen: Tendenziell überwiegen in Bezug auf Ebene und Wertung die gemischten Formen. Insbesondere dominiert der gemischte Fokus bei den Kommentaren. Zudem wird deutlich, dass die Beurteilungen und Kommentare eher selten einen Bezug zum konkreten Schülertext aufweisen.

Textbeurteilungen mit Textbezug unterscheiden sich in Bezug auf ihre Textlänge von jenen ohne Textbezug (t_1 : $M = 1301$ bzw. 829 ; t_2 : $M = 1187$ bzw. 794). Der Unterschied ist zu beiden Messzeitpunkten signifikant (t_1 : $t(97) = -4.68$, $p = .000$; t_2 : $t(47) = -2.84$, $p = .002$). Für Kommentare mit Textbezug gilt dies nur zu t_1 (t_1 : $M = 755$ bzw. 595 , $t(97) = -2.14$, $p = .035$; t_2 : $M = 693$ bzw. 593 , $t(47) = -.89$, $p = .375$).

Interkorrelationen zwischen Textbeurteilung und Kommentar können nur wenige nachgewiesen werden: so für Ebene zu t_1 ($r_s = .21$, $p = .035$) und für Textbezug zu t_2 ($Chi^2 = 3.89$, $N = 48$, $df = 1$, $p = .003$). Die Zusammenhänge sind insgesamt gering. Ein qualitativer Vergleich zeigt jedoch, dass Textbeurteilung und Kommentar inhaltlich durchaus korrespondieren können: Im folgenden Beispiel nimmt die Probandin im Kommentar u. a. die Beobachtung aus der Textbeurteilung auf, dass der Konflikt schnell gelöst werde und dadurch keine Spannung entstehe. Dies erfolgt im Kommentar jedoch implizit, indem sie fragt, ob sich der Junge in der Geschichte tatsächlich so schnell von seinem Schock erholt habe.

Textbeurteilung	Kommentar
Bei dieser Geschichte wird keine Spannung aufgebaut, da der Konflikt sehr schnell gelöst und kaum erklärt wird. Es fehlen Gefühle und Details in der Geschichte, damit Spannung entsteht. Zudem ist der Titel nicht attraktiv. Die Idee mit dem Monster finde ich gut, jedoch wird dieses Element nicht richtig genutzt. Man versteht, was der Schüler sagen will, dennoch wirken die Sätze eher plump. Die Rechtschreibung und die Grammatik finde ich für sein Alter nicht schlecht und da das	Ich mag die Idee mit dem Monster, welches aus dem Bild springt, sehr. Da hattest du eine sehr gute Idee! Aus meiner Sicht hättest du beschreiben dürfen wie beispielsweise das Monster aussieht, damit sich der Leser eine Vorstellung von deinem Monster machen kann. Und meinst du der Junge konnte sich so schnell von seinem Schock erholen? Da darf man ruhig noch mehr dazu erfinden. Oder wie reagieren denn seine Eltern und Freunde von der Schule, wenn er plötzlich ein Monster als Freund hat?

Thema eine spannende Geschichte schreiben war, würde ich dieses Kriterium weniger gewichten.

Tabelle 4: Textbeurteilung und Kommentar im Vergleich, G02-1106, t₁

Dieses Beispiel zeigt auch Unterschiede auf: So betont die Probandin im Kommentar das Positive viel stärker; negative Aspekte bringt sie implizit ins Spiel. In der Textbeurteilung fokussiert sie negative Aspekte, während sie im Kommentar in erster Linie die lokalen Aspekte weglässt. Damit korrespondieren Textbeurteilung und Kommentar v.a. hinsichtlich der globalen Ebene. Aus einer qualitativen Perspektive fällt zudem auf, dass im Kommentar kein Bezug zu den typischen Merkmalen narrativer Texte hergestellt wird, sondern in Bezug auf die fehlende Spannung ohne weitere Begründung ergänzt wird, dass „man dazu ruhig mehr erfinden dürfe“.

Zur Qualität der Textbeurteilungen

Tabelle 5 zeigt, dass die Textbeurteilungen der Probanden weder am Ende ihrer Ausbildung noch nach dem ersten Berufsjahr mit den Einschätzungen von Experten übereinstimmen, und zwar sowohl in Bezug auf die lokale als auch globale Ebene.

	Textbeurteilung	
	t ₁ (N = 99)	t ₂ (N = 49)
Angemessenheit lokale Ebene: erfüllt	29,3 %	28,6 %
Angemessenheit globale Ebene: erfüllt	21,2 %	32,7 %

Tabelle 5: Übereinstimmung der Textbeurteilung mit Expertenurteil

Textbeurteilungen, die in Bezug auf die globale Ebene angemessen sind, weisen zu beiden Messzeitpunkten eine signifikant höhere Anzahl Zeichen auf als global nicht angemessene Textbeurteilungen (t₁: t(97) = -4.80, p = .000; t₂: t(47) = -3.11, p = .003). Dagegen unterscheiden sich lokal angemessene Textbeurteilungen in ihrem Textumfang nicht signifikant von solchen, die auf lokaler Ebene nicht angemessen sind, auch wenn sich eine starke Tendenz zu t₂ zeigt (t₁: t(97) = -.99, p = .324; t₂: t(47) = -1.98, p = .053).

Das folgende Beispiel steht exemplarisch für eine Textbeurteilung, die mit dem Expertenurteil weder auf der lokalen noch der globalen Ebene übereinstimmt: Zum einen wird die orthografische Leistung des Schülers von dem Probanden nicht korrekt eingeschätzt – hier zeigt sich auch fehlendes fachliches Wissen –, zum anderen werden in Bezug auf die globale Ebene von dem Probanden weder die relevanten Kriterien herangezogen noch die genannten Kriterien angemessen eingeschätzt. Darüber hinaus fällt auf, dass die Textbeurteilung eher die lokale Ebene sowie negative Aspekte fokussiert und keinen direkten Bezug zum Schülertext aufweist.

Den Test würde ich wohl als eher knapp genügend beurteilen in seiner Gesamtbeurteilung.

Bei der Rechtschreibung würde ich im Minimum die phonetische korrekte Schreibweise erwarten, was nur teilweise von Tim erreicht wird. Der Umfang der Arbeit ist für eine ganze Lektion Schreibzeit sehr knapp. Die Geschichte hat keinerlei unerwartete Handlungsstränge oder Überraschungen bereit. Der Wortschatz befindet sich im normalen bis eher im spärlicheren Bereich. Auch fehlt in den Sätzen immer mal wieder ein Wort oder eine Präposition. Der Satzsatz ist ein abgekupferter Satzsatz, wie ihn viele Kinder benutzen. Aber halt auch verbraucht und alles andere als interessant und kreativ.

Insgesamt viele Dinge, die Tim noch besser machen könnte bzw. noch lernen sollte bei seinem schriftlichen Ausdruck.

Abb. 5: Textbeurteilung, G01-1094, t_1 , keine Übereinstimmung mit Expertenurteil

Dem gegenüber zeigt Abb. 6 einen Ausschnitt aus einer Textbeurteilung durch eine Probandin, die mit der Expertenbeurteilung übereinstimmt. Dieses Beispiel illustriert darüber hinaus, wie globale Aspekte zwar durchaus angemessen beurteilt werden, dabei aber die Kriterien oftmals nicht mit dem Fachausdruck benannt werden.

[...] Inhalt: Der Junge, der malt, hat keine Persönlichkeit. Er könnte einen Namen haben und ein bisschen beschrieben werden (wo ist er? wie alt ist er? wie sieht er aus?). Der Plot finde ich einfallsreich und eigentlich spannend. Jedoch ist er ein wenig plakativ geschrieben, die Verhandlung darüber, dass sie Freunde sein könnten, geschieht sehr schnell. An und für sich passiert wenig/ gibt es wenig Handlung. Der Junge malt, das Monster kommt heraus und sie sprechen miteinander, fertig. [...]

Abb. 6: Auszug Textbeurteilung, G02-1107, t_1 , in Übereinstimmung mit Expertenurteil

Zur Qualität der Kommentare

Die Qualität der Kommentare wurde mittels ihrer Form und ihrer Verankerung bestimmt. Tabelle 6 illustriert, dass die wenigsten Kommentare konkrete im Text verankerte Tipps oder Überarbeitungsvorschläge enthalten. Zudem sind die wenigsten Kommentare eher dialogisch formuliert: Entweder sind sie eher direktiv oder gemischt formuliert.

	Kommentar	
	t_1 (N = 99)	t_2 (N = 49)
<i>Tipps</i>		
verankert	9,1 %	6,1 %
<i>Form</i>		
dialogisch	7,1 %	6,1 %
gemischt	58,6 %	51 %
direktiv	34,3 %	42,9 %

Tabelle 6: Übersicht Kommentar zu Tipps und Form

Kommentare mit im Schülertext verankerten Tipps weisen zu t_2 eine signifikant höhere Anzahl Zeichen auf als Kommentare ohne textuell verankerte Tipps (t_1 : $t(97) = -1.39$, $p = .165$; t_2 : $t(47) = -2.17$, $p = .035$). Hinsichtlich Form korrelieren Kommentare jedoch zu beiden Messzeitpunkten signifikant mit der Anzahl Zeichen (t_1 : $r_s = .24$, $p = .018$; t_2 : $r_s = .59$, $p = .000$).

Das folgende Beispiel in Abb. 7 zeigt einen Kommentar, der zu ungefähr gleichen Teilen direktive und dialogische Anteile enthält, also von mittlerer Qualität ist. Oftmals betreffen in solch gemischten Kommentaren die Anweisungen (direktive Anteile) eher die lokale Ebene, Fragen oder Leserreaktionen (dialogische Anteile) dagegen die globale Ebene. Zudem sind die Tipps nicht nachvollziehbar im Text des Schülers verankert: So ist nicht klar, weshalb das Monster beschrieben werden (ansatzweise ist dies im Schülertext erfüllt: das Monster hat keine Arme, es wird lebendig) oder weshalb der Schüler andere Wörter für „sagen“ kennen sollte.

Lieber Tim

Du hast eine sehr spannende Geschichte geschrieben! Ich finde deine Idee, das Monster leben zu lassen, sehr kreativ.

Deine Geschichte ist strukturiert, man kann eine kurze Einleitung und einen Schluss finden.

Hat der Junge denn keinen Namen? Versuche das nächste Mal der Person und auch dem Monster einen Namen zu geben. Du könntest auch noch beschreiben, wie das Monster aussieht. Vielleicht erleben die beiden noch ein Abenteuer?

Versuche das nächste Mal deinen Text nochmals durchzulesen, damit du weniger Schreibfehler machst. Bestimmt erkennst du viele davon beim zweiten Mal Lesen! Gib Acht auf die Gross-/Kleinschreibung (leben oder Leben?). Du hast die direkte Rede sehr gut eingesetzt und ich kann sehen, dass du diese beherrschst! Kennst du noch andere Wörter für „sagen“?

Abb. 7: Gemischt formulierter Kommentar, Tipps ohne angemessene Verankerung im Text, G06-1163, t₁

Das Beispiel in Abb. 8 steht exemplarisch für einen qualitativ nicht angemessenen Kommentar, der vorwiegend direktiv ist und die Hinweise nicht am Text des Schülers verortet. Der Kommentar formuliert entsprechend allgemeine Prinzipien, die so für den Schüler kaum operationalisierbar sind:

Tim, nimm dir unbedingt noch die Zeit, um den Text durchzulesen und jedes Wort zu kontrollieren. Achte darauf, dass du keine Buchstaben in Wörtern vergisst.

Das nächste Mal wünsche ich mir, dass du eine etwas spannendere Geschichte schreibst. Dazu kannst du beispielsweise mehr Adjektive gebrauchen. Adjektive beschreiben etwas besser und genauer, somit auch spannender.

Ebenfalls sollst du darauf achten, dass ein Leser die Geschichte versteht und du der Reihe nach alles, was wichtig ist, aufschreibst.

Abb. 8 : Direktiv formulierter Kommentar, Tipps ohne Verankerung im Text, G03-1118, t₁

Veränderungen im Verlauf des ersten Berufsjahrs

In Bezug auf die einzelnen Kriterien zeigen sich weder bei den Textbeurteilungen noch bei den Kommentaren signifikante Veränderungen nach einem Jahr Berufserfahrung ab Ende Ausbildung, wie Tabelle 7 illustriert:

Textbeurteilung	
Ebene ^a	$Z = -2.19, N = 49, p = .095$
Angemessenheit lokale Ebene ^b	$Chi^2 = .08, N = 49, df = 1, p = .824$
Angemessenheit globale Ebene ^b	$Chi^2 = 2.17, N = 49, df = 1, p = .454$
Kommentar	
Ebene ^a	$Z = -.18, N = 49, p = .857$
Tipps ^b	$Chi^2 = 7.16, N = 49, df = 1, p = .219$
Form ^a	$Z = -.33, N = 49, p = .741$

Tabelle 7: Veränderungen von t₁ zu t₂; a: Wilcoxon-Test, b: McNemar-Test

Auch hinsichtlich Textmenge zeigen sich weder bei den Textbeurteilungen noch bei den Kommentaren signifikante Veränderungen ($t(49) = 1.28, p = .205$ bzw. $t(49) = 1.41, p = .164$).

4 | Diskussion

Die Frage, über welches fachdidaktische Wissen (angehende) Lehrpersonen verfügen, kann aufgrund der vorliegenden Daten zunächst dahingehend beantwortet werden, dass die Probanden am Ende ihrer Ausbildung sowie am Ende des ersten Berufsjahrs mehrheitlich nicht in der Lage sind, einen Schülertext adäquat einzuschätzen, zumindest nicht, wenn sie den Text ohne vorgegebenes Raster beurteilen sollen. Das schließt nicht aus, dass sie – wie in der Studie von Bouwer et al. (2015) – mit einem guten Beurteilungsraster durchaus in der Lage wären, Schülertexte angemessen zu beurteilen. Allerdings stellt sich die Frage, ob sie über ausreichendes Wissen verfügen, um qualitativ gute von qualitativ schlechten Beurteilungsrastern unterscheiden und entsprechend auswählen zu können. Nicht zuletzt ist unklar, inwiefern sie über ausreichendes Textsortenwissen verfügen.

In Übereinstimmung mit Bouwer et al. (2015) zeigte sich kein Effekt durch zunehmende Berufserfahrung, weder bei den Textbeurteilungen noch bei den Kommentaren. Das konnte nicht nur in Bezug auf die Qualitätskriterien festgestellt werden, sondern auch in Bezug auf die in den Textbeurteilungen und Kommentaren angesprochenen Aspekte wie globale vs. lokale Ebene usw.

Bemerkenswert ist, dass Kommentare und Textbeurteilungen nur bedingt miteinander korrespondieren: So zeigt sich in den Kommentaren eine deutlich stärkere Tendenz, Positives hervorzuheben – ungeachtet der tatsächlichen Qualität des Schülertextes – sowie lokale und globale Aspekte gleichermaßen zu berücksichtigen. Insgesamt kann dies übereinstimmend mit Bouwer/van den Bergh (2014) als eine Frage des Feedback-Stils interpretiert werden.

Da insbesondere die Kommentare ihre Hinweise nur selten am Schülertext verorten, formulieren sie oft allgemeine Prinzipien, die für den Schüler kaum nachvollziehbar sein dürften. Da die Kommentare zudem überwiegend direktive Elemente enthalten, werden dem Schüler in erster Linie isolierte normative Setzungen vermittelt. Wie in Kapitel 1.2 ausgeführt wurde, wirken sich vage und direktiv formulierte Kommentare weniger günstig auf Überarbeitungen aus als dialogisch formulierte Kommentare.

Die hier präsentierten Ergebnisse stellen nur einen Ausschnitt dar: Nicht erfasst wurde etwa fachdidaktisches Wissen über nachweislich wirksame Vermittlungsarten. Darüber hinaus wird noch zu prüfen sein, inwiefern fachliches und fachdidaktisches Wissen zusammenhängen. Vor allem aber wird im Rahmen der Teilstudie B des Projekts NovIS genauer aufzuzeigen sein, welche Rolle fachliches und fachdidaktisches Wissen im Schreibunterricht von Lehrpersonen im ersten Berufsjahr spielen.

An Koster et al. (2015) anschließend kann basierend auf den hier präsentierten Ergebnissen festgehalten werden, dass zum einen die Ausbildung kritisch hinterfragt werden muss, zum anderen den Lehrmitteln und auch den Weiterbildungsangeboten eine große Bedeutung zukommt. Dabei sollten Lehrmittel nicht nur Beurteilungsraster anbieten: Wünschbar wäre auch, dass sie Muster enthalten, die den Lehrpersonen aufzeigen, wie die Beurteilungen an Schülertexten verortet werden können (vgl. auch Sturm 2014). Weiterbildungen sollten darüber hinaus den Lehrpersonen aufzeigen, wie Kriterien mit Blick auf die Schreibaufgabe und die damit verbundenen Lernziele zu gewichten sind und welche didaktischen Folgerungen aus Textbeurteilungen abgeleitet werden können.

Literatur

- Becker-Mrotzek, Michael / Böttcher, Ingrid (Hg.) (2012): Schreibkompetenz entwickeln und beurteilen. 4., überarbeitete Auflage. Berlin: Cornelsen Scriptor.
- Bouwer, Renske / Béguin, Anton / Sanders, Ted / van den Bergh, Huub (2015): Effect of genre on the generalizability of writing scores. In: *Language Testing*, 32, H. 1, S. 83–100.
- Bouwer, Renske / van den Bergh, Huub (2014): Effectiveness of teacher feedback. EARLI SIG Conference on Writing Research, Amsterdam.
- Cho, Kwangsu / MacArthur, Charles (2010): Student revision with peer and expert reviewing. In: *Learning and Instruction*, 20, H. 4, S. 328–338.
- Cohen, Jacob (1968): Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. In: *Psychological Bulletin*, 70, H. 4, S. 213–220.
- Collopy, Rahel M.B. (2008): Professional development and student growth in writing. In: *Journal of Research in Childhood Education*, 23, H. 2, S. 163–178.
- Fix, Martin (2006): *Texte schreiben: Schreibprozesse im Deutschunterricht*. Stuttgart: UTB.
- Furger, Julienne / Schneider, Hansjakob (2011): Schreiben in und mit digitalen Medien: die Web-Plattform my-Moment. *Leseforum*. Online unter <http://leseforum.ch/fokusartikel1_2011_2.cfm> (letzter Aufruf 5.5.2016).
- Gilbert, Jennifer / Graham, Steve (2010): Teaching writing to elementary students in grades 4–6: A national survey. In: *The Elementary School Journal*, 110, H. 4, S. 494–518.
- Glaser, Cornelia / Keßler, Christina / Brunstein, Joachim C. (2009): Förderung selbstregulierten Schreibens bei Viertklässlern. Effekte auf strategiebezogene, holistische und subjektive Maße der Schreibkompetenz. In: *Zeitschrift für Pädagogische Psychologie*, 23, H. 1, S. 5–18.
- Graham, Steve / Harris, Karen R. / Fink, Barbara / MacArthur, Charles (2001): Teacher efficacy in writing: A construct validation with primary grade teachers. In: *Scientific Studies of Reading*, 5, H. 2, S. 177–202.
- Graham, Steve / Harris, Karen R. / Hebert, Michael (2011): *Informing writing. The benefits of formative assessment*. New York: Carnegie Corporation of New York.
- Graham, Steve / McKeown, Debra / Kihara, Sharlene / Harris, Karen R. (2012): A meta-analysis of writing instruction for students in the elementary grades. In: *Journal of Educational Psychology*, 10, H. 4, S. 879–896.
- Harris, Karen R. / Graham, Steve (1994): Constructivism: Principles, paradigms, and integration. In: *The Journal of Special Education*, 28, H. 3, S. 233–247.
- Hattie, John (2012): *Visible learning for teachers: Maximizing impact on learning*. London, New York: Routledge Chapman/Hall.
- Hattie, John / Timperley, Helen (2007): The power of feedback. In: *Review of Educational Research*, 77, H. 1, S. 81–112.
- Hugener, Isabell / Rakoczy, Katrin / Pauli, Christine / Reusser, Kurt (2006): Videobasierte Unterrichtsforschung: Integration verschiedener Methoden der Videoanalyse für eine differenzierte Sicht auf Lehr-Lernprozesse. In: Rahm, Sybille / Mammes, Ines / Schratz, Michael (Hg.): *Schulpädagogische Forschung. Unterrichtsforschung – Perspektiven innovativer Ansätze*. Innsbruck: Studienverlag, S. 41–53.
- Jost, Jörg / Lehnen, Katrin / Rezat, Sara / Schindler, Kirsten (2011): Schriftliches Beurteilen lernen. In: Bräuer, Gerd/Schindler, Kirsten (Hrsg.): *Schreibarrangements für Schule, Hochschule, Beruf*. Freiburg: Fillibach, S. 221–239.
- Kihara, Sharlene A. / Graham, Steve / Hawken, L. (2009): Teaching writing to high school students: A national survey. In: *Journal of Educational Psychology*, 101, H. 1, S. 136–160.
- Kleickmann, Thilo / Richter, Dirk / Kunter, Mareike / Elsner, Jürgen / Besser, Michael / Krauss, Stefan / Baumert, Jürgen (2013): Teachers' content knowledge and pedagogical content knowledge the role of structural differences in teacher education. In: *Journal of Teacher Education*, 64, H. 1, S. 90–106.
- Koster, Monica / Tribushinina, Elena / de Jong, Peter F. / van den Bergh, Huub (2015): Teaching children to write: A meta-analysis of writing intervention research. In: *Journal of Writing Research*, 7, H. 2, S. 249–274.
- Krauss, Stefan / Blum, Werner (2012): The conceptualisation and measurement of pedagogical content knowledge and content knowledge in the CoACTIV study and their impact on student learning. In: *Journal of Education*, H. 56, S. 45–66.
- Krauss, Stefan / Neubrand, Michael / Blum, Werner / Baumert, Jürgen / Brunner, Martin / Kunter, Mareike / Jordan, Alexander (2008): *Die Untersuchung des professionellen Wissens deutscher Mathematik-*

- Lehrerinnen und -Lehrer im Rahmen der CoACTIV-Studie. In: Journal für Mathematik-Didaktik, 29, H. 3–4, S. 223–258.
- Kunter, Mareike / Pohlmann, Britta (2009): Lehrer. In: Wild, Elke / Möller, Jens (Hg.): Pädagogische Psychologie. Heidelberg: Springer, S. 261–282.
- Lipowsky, Frank (2009): Unterricht. In: Wild, Elke / Möller, Jens (Hg.): Pädagogische Psychologie. Heidelberg: Springer, S. 73–101.
- MacArthur, Charles A. (2015): Instruction in evaluation and revision. In: MacArthur, Charles A. / Graham, Steve / Fitzgerald, Jill (Hrsg.): Handbook of writing research. Second Edition. New York: Guilford Press, S. 272–287.
- Myhill, Debra / Jones, Susan / Watson, Annabel (2013): *Grammar matters: How teachers' grammatical knowledge impacts on the teaching of writing*. In: Teaching and Teacher Education, H. 36, S. 77–91.
- Oser, Fritz / Heinzer, Sarah / Salzmann, Patrizia (2010): Die Messung der Qualität von professionellen Kompetenzprofilen von Lehrpersonen mit Hilfe der Einschätzung von Filmvignetten. In: Unterrichtswissenschaft, 38, H. 1, S. 5–28.
- Pajares, Frank (1992): Teachers' beliefs and educational research: Cleaning up a messy construct. In: Review of Educational Research, 62, H. 3, S. 307–332.
- Parr, Judy M. / Timperley, Helen S. (2010): Feedback to writing, assessment for teaching and learning and student progress. In: Assessing Writing, 15, H. 2, S. 68–85.
- Shulman, Lee S. (1986): Those who understand: Knowledge growth in teaching. In: Educational Researcher, 15, H. 2, S. 4–14.
- Steendam, Elke Van / Rijlaarsdam, Gert / van den Bergh, Huub / Sercu, Lies (2014): The mediating effect of instruction on pair composition in L2 revision and writing. In: Instructional Science, 42, H. 6, S. 905–927.
- Sturm, Afra (2014): Experten- und Novizen-Feedback in der Domäne Schreiben. In: Leseforum. Online unter <http://www.leseforum.ch/myUploadData/files/2014_3_Sturm.pdf> (letzter Aufruf 5. 5. 2016).
- Sturm, Afra / Schneider, Hansjakob / Lindauer, Nadja / Sommer, Tim (i. Dr.): Schreibbezogenes Fachwissen bei Lehrpersonen im ersten Berufsjahr. In: Krelle, Michael / Senn, Werner (Hg.): Qualitäten von Deutschunterricht – Empirische Unterrichtsforschung im Fach Deutsch. Stuttgart: Fillibach bei Klett, S. 139–161.
- Troia, Gary A. (2006): Writing Instruction for Students with Learning Disabilities. In: MacArthur, Charles A. / Graham, Steve / Fitzgerald, Jill (Hg.): Handbook of writing research. New York: Guilford Press. S. 324–336.

Prof. Dr. Afra Sturm

Zentrum Lesen
 Fachhochschule Nordwestschweiz
 afra.sturm@fhnw.ch