

Förderung der Überarbeitung von Textzusammenfassungen sachunterrichtlicher Fachtexte durch Self-Assessment und Peer Feedback mit Prompts

Veronika Barkela & Miriam Leuchter

1. Einleitung

Die Vermittlung pädagogischen Fachwissens während der universitären Lehramtsausbildung hat für den Sachunterricht eine große Bedeutung. Pädagogisches Fachwissen wird als Amalgam von Fachwissen und allgemeinem pädagogischen Wissen verstanden (Shulman 1987) und beschreibt das Wissen, wie fachliche Inhalte pädagogisch sinnvoll aufbereitet und Lernenden zugänglich gemacht werden können (Blömeke, Gustafsson & Shavelson 2015; Kunter, Klusmann, Baumert, Richter, Voss & Hachfeld 2013). Die Verfügbarkeit pädagogischen Fachwissens ist das zentrale Kriterium für professionelle Kompetenz von Lehrpersonen (Baumert, Kunter, Blum, Brunner, Voss, Jordan, Klusmann, Krauss, Neubrand & Tsai 2010). Mehrere Studien zeigen, dass pädagogisches Fachwissen starken Einfluss auf die Lehrqualität von Lehrpersonen und den Lernerfolg von Schüler/innen hat (Kulgemeyer & Riese 2018; Meschede, Fiebranz, Möller & Steffensky 2017; Park, Jan, Chen & Jung 2011; Hill, Rowan & Ball 2005). Jedoch stellt der Erwerb pädagogischen Fachwissens Studierende und Lehrpersonen vor große Herausforderungen (Hume & Berry 2011). Der Erwerb pädagogischen Fachwissens mit Hilfe von Fachliteratur ist die Grundlage, die in der universitären Lehramtsausbildung gelegt werden kann. Deshalb ist eine frühe Auseinandersetzung mit Fachliteratur zu Aspekten des pädagogischen Fachwissens in der sachunterrichtlichen Lehramtsausbildung sinnvoll. Es zeigt sich jedoch, dass Studierende Schwierigkeiten haben, Fachtexte zu erschließen, relevante Aspekte zu identifizieren und diese in eigene Worte zu fassen (Friend 2001).

Eine effektive Lernstrategie diesbezüglich ist das Schreiben von Textzusammenfassungen. Es fördert eine intensive Auseinandersetzung mit dem Textgegenstand und der eigenen Schreibkompetenz (Lenhard, Baier, Lenhard, Hoffmann & Schneider 2013; Perin, Lauterbach, Raufman & Kalamkarian 2017; Friend 2001). Das Verdichten von Textaussagen erfordert, die eigene Wortwahl und

Satzkonstruktion auf den Prüfstand zu stellen und fördert so ein detailliertes Verständnis des Textinhalts (Westby, Culatta, Lawrence & Hall-Kenyon 2010). Deshalb wird es als gängige Schreibaufgabe für Studierende unterer Semester eingesetzt (Mok & Chan 2016).

Unser Ziel ist, ein intelligentes tutorielles System zu entwickeln, welches Studierenden bei der Zusammenfassung von Texten hilft und sie zur Selbstevaluierung des eigenen Textverständnisses und der Textproduktion anregt. In einem ersten Schritt haben wir für die vorliegende Pilotstudie die Möglichkeiten zur Förderung der Überarbeitung von Textzusammenfassungen sachunterrichtlicher Fachtexte durch Self-Assessment und Peer Feedback mit Prompts untersucht.

Das Schreiben von Textzusammenfassungen erfordert mehrere kognitive Prozesse (Perin et al. 2017, Westby et al. 2010). Vorwissen muss aktiviert werden (Kellogg & Raulerson 2007), neue Informationen müssen in das eigene Wissensnetzwerk eingeflochten werden, und ein mentales Modell des zu schreibenden Textes muss entworfen werden (Becker-Mrotzek, Grabowski, Jost, Knopp & Linnemann 2014; Schnotz 2006). Schließlich wird dieses mentale Modell während der Textproduktion in eigene Worte übersetzt und niedergeschrieben (Hayes 2012). Schreibende, denen die Koordinierung dieser Prozesse besser gelingt, schreiben Zusammenfassungen, die durch eine präzise Darstellung der Kerninhalte geprägt sind, die wenige Wörter oder Satzteile aus dem Originaltext beinhalten und in denen unwichtige Aspekte und Redundanzen vermieden werden (Westby et al. 2010).

Kognitionspsychologisch wird das Schreiben (von Textzusammenfassungen) in die Subprozesse Planen, Generieren und Überarbeiten unterteilt, die durch individuelle Voraussetzungen, wie z.B. Metakognition, Langzeitgedächtnis und Lesefähigkeit beeinflusst werden (Hayes 2012). Eine bessere Koordination dieser Subprozesse gelingt mit Hilfe von Schreibstrategien, die unter anderem das Überarbeiten von Texten fokussieren (Kieft, Rijlaarsdam, Galbraith & van den Bergh 2007).

2. Textüberarbeitung

Studien zeigen, dass Schreibende, die mehr Zeit in die Überarbeitung ihrer Texte investieren, eine höhere Textqualität erreichen (Butler & Britt 2010). Das Überarbeiten eines Textentwurfs beinhaltet das Erkennen und Diagnostizieren von Problemen sowie die Auswahl adäquater Strategien zu deren Behebung

(Hayes 2012). Dabei zeigen sich abweichende Überarbeitungsstrategien zwischen erfahrenen und unerfahrenen Schreibenden.

Erfahrene Schreibende verstehen das Überarbeiten als rekursiven Prozess (Sommers 1980) und überdenken vor allem Struktur, Gedankenorganisation und Argumentationslinien ihres Textes (Roscoe, Snow, Allen & McNamara 2015). Sie legen ihren Entwürfen eine angemessenere mentale Repräsentation zu Grunde als unerfahrene Schreibende (Kellogg & Whiteford 2012) und berücksichtigen während des Überarbeitens potentielle Adressaten (Becker-Mrotzek et al. 2014).

Im Gegensatz dazu nehmen unerfahrene Schreibende das Überarbeiten von Texten weniger als Möglichkeit wahr, Texte zu optimieren, sondern sehen es eher als zusätzliche Aufgabe (Hoogreveen & van Gelderen 2013). Sie überarbeiten ihre Texte eher auf der Oberflächenstruktur (Grammatik, Zeichensetzung, Rechtschreibung – Roscoe et al. 2015) und erkennen seltener Probleme in der Tiefenstruktur (Argumentation, Kohärenz, Textaufbau – Butler & Britt a.a.O.). Im Besonderen nehmen sie bei der Überarbeitung ihrer Texte oberflächliche Wortwiederholungen wahr und formulieren sie um. Konzeptuelle Wiederholungen fallen ihnen jedoch selten auf und werden kaum überarbeitet (Sommers a.a.O.). Allerdings zeigen mehrere Studien, dass unerfahrene Schreibende nach Strategietrainings ihre Texte besser überarbeiten konnten (Butler & Britt a.a.O., Roscoe et al. 2015). Formatives Feedback, in Form von Self-Assessment und Peer Feedback, kann dabei eine wirksame Förderstrategie sein (Graham, Hebert & Harris 2015).

3. Self-Assessment

Unter Self-Assessment wird die Fähigkeit eines Lernenden verstanden, den eigenen Lernprozess oder die eigene Arbeit und Leistung selbstreguliert anhand von Kriterien zu beurteilen (Panadero & Alonso-Tapia 2013). Self-Assessments unterstützen eine adäquate Selbsteinschätzung, wenn Aufgaben und Bearbeitungskriterien präzise und konkret formuliert sind, metakognitive Strategien adressiert werden und eine angemessene Unterweisung in die Bedeutung von Self-Assessment stattfindet (Boud, Lawson & Thompson 2013).

Mehrere Studien zeigen, dass mit Hilfe von Lernunterstützungen, wie Strategietrainings und angeleiteten Self-Assessments, der Gebrauch von angemessenen Schreibstrategien vermittelt und intensiviert werden kann (MacArthur, Phi-

lippakos & Ianetta 2015). Boud et al. (2013) beobachteten, dass Studierende, die vier Semester lang regelmäßige Self-Assessments bzgl. ihres Schreibens durchführten, anfänglich dazu tendierten sich zu überschätzen, die Einschätzungen sich jedoch nach einiger Zeit an eine Experteneinschätzung angleichen.

4. Peer Feedback

Peer Feedback findet statt, wenn Personen einer Statusgruppe sich gegenseitig beim Lernen unterstützen und durch Lehren selbst lernen (Topping, Smith, Swanson & Elliot 2000). Es ermöglicht Personen beim Verfassen von Texten, Rückmeldung zu ihrem Text zu erhalten und zeigt auch der Feedback gebenden Person Aspekte auf, wie die eigene Arbeit verbessert werden könnte (Panadero & Brown 2016). Peer Feedback muss jedoch den Herausforderungen Rechnung tragen, die Studierende beim Überarbeiten von Text(zusammenfassung)en haben (vgl. Kap. 2). Deshalb ist, analog zum Self-Assessment, Peer Feedback dann effektiv, wenn die entsprechenden Schreibstrategien angesprochen werden und die Bedeutung des Peer-Feedbacks deutlich gemacht wird (Hoogeveen & van Gelderen 2013, Falchikov 2006). Der Einsatz von Prompts hat sich dabei bewährt und als unterstützend erwiesen (Gan & Hattie 2014).

5. Prompts zur Unterstützung von Self-Assessment und Peer Feedback

Der Einsatz von Prompts verfolgt das Ziel, die Aufmerksamkeit von Studierenden auf bestimmte Aspekte des Lernprozesses zu lenken (Bannert, Sonnenberg, Mengelkamp & Pieger 2015). Besonders regen sie Lernende zur Reflektion und zum Einsatz von Lernstrategien an (Proske, Narciss & McNamara 2012). Mit Hilfe von Prompts können kognitive und metakognitive Prozesse stimuliert sowie Studierende zur intensiven Auseinandersetzung mit Arbeitsaufträgen motiviert werden. Prompts unterstützen Studierende, sich an Lernkonzepte und Vorgehensweisen zu erinnern und diese einzusetzen (Bannert & Reimann 2012). Aufgrund dieser Studien können wir schlussfolgern, dass Prompts eine effektive Unterstützungsform für den Einsatz von Self-Assessment und Peer Feedback sein können.

6. Die Studie

Ziel dieser Pilotstudie war es zu untersuchen, ob Studierende mit Hilfe eines von uns entwickelten Schreibtrainings 1) bessere Textqualität erreichen und 2) die

Überarbeitung der eigenen Texte anhand angeleiteten Self-Assessments oder zusätzlichen Peer Feedbacks besser gelingt.

Folgende Forschungsfragen sollen bearbeitet werden:

1. Verbessert sich die Qualität von Textzusammenfassungen sachunterrichtlicher Fachtexte nach einem Schreibtraining?
2. Wie schätzen Studierende ihre Textzusammenfassungen im Vergleich zu einer Expertenbewertung ein?
3. Wie bewerten Peers im Vergleich zu Experten die Textzusammenfassungen ihrer Kommiliton/innen?
4. Wie unterscheiden sich Self-Assessment und Peer Feedback voneinander?

6.1 Studiendesign und Vorgehen

Die Studie fand in einem Prätest – Treatment – Posttest Design mit zwei Treatmentsitzungen statt. Insgesamt umfasste das Schreibtraining vier Sitzungen à 90 Minuten, welche in Lehrveranstaltungen integriert waren. Die erste Sitzung (Prätest) fand zu Beginn des Semesters statt. Nach einer kurzen Information zu dem Projekt wurden die Studierenden gebeten, eine Textzusammenfassung über einen vorgegebenen Fachtext zu erstellen. Nach sechs Wochen fanden im Abstand von jeweils zwei Wochen drei weitere Sitzungen statt. Während der zweiten und dritten Sitzung wurden die Studierenden jeweils mit einer zehnminütigen Einführung auf wichtige Aspekte einer Textzusammenfassung hingewiesen und aufgefordert, eine Textzusammenfassung zu erstellen. Dabei erhielten sie Unterstützung durch mit Prompts angeleitetem Self-Assessment vor und nach der Erstellung der Textzusammenfassung und konnten die Textzusammenfassung noch einmal überarbeiten. Anschließend wurde Peer Feedback von Studierenden (den Sitznachbarn) gegeben, und die Texte konnten ein letztes Mal überarbeitet werden. Die letzte Sitzung (Posttest) war wie die erste Sitzung gestaltet: Studierende fassten einen Text ohne Unterstützung zusammen.

Als Aspekte von Textqualität haben wir Inhalt, Plagiatsvermeidung und Redundanzvermeidung erhoben.

6.2 Textmaterial

Als Textmaterial kamen sachunterrichtliche Fachtexte mit ähnlicher Länge zum Einsatz (451 bis 638 Wörter). Die Fachtexte behandelten die Themen Untersuchungsformen (Grygier & Hartinger 2009), Denk- und Arbeitsweisen (Möller,

Bohrmann, Hirschmann, Wilke & Wyssen 2013), Variablenkontrollstrategie (Sodian & Mayer 2013) und Conceptual Change (Möller & Steffensky 2010). Die Komplexität der Texte in dieser Reihenfolge wurde mittels Lesbarkeitsindex (LIX) nach Björnsson (Lenhard & Lenhard 2014-2017) beurteilt und anhand eines Expertenratings mit vier Experten verifiziert, was die angestrebte ansteigende Textkomplexität bestätigte. Björnssons Formel errechnet die Lesbarkeit eines Textes anhand der Summe aus der durchschnittlichen Satzlänge und dem prozentualen Anteil langer Wörter (mehr als sechs Buchstaben). Texte, die mit einem LIX ab 50 bewertet werden, gelten als Sachliteratur, ein LIX ab 60 beschreibt Fachliteratur. Alle Texte wurden als komplexe bis sehr komplexe Sach- bzw. Fachtexte eingestuft (LIX 58,3-71,7) und wurden leicht vorstrukturiert den Studierenden präsentiert.

6.3 Schreibumgebung

Eine internetfähige Schreibumgebung wurde in LimeSurvey programmiert. Während des Treatments werden sieben Schritte bearbeitet (siehe Tabelle 1).

Tabelle 1: 7 Schritte zur Erarbeitung einer Zusammenfassung

Instrument	Beispiel
1. Self-Assessment: Vorwissen aktivieren	<ul style="list-style-type: none"> • Haben Sie zu diesem Thema schon einmal Wissen erworben?
2. Self-Assessment: Neue Informationen verarbeiten	<ul style="list-style-type: none"> • Können Sie die Gedankengänge des Textes an bis zu drei eigenen Beispielen festmachen?
3. Zusammenfassen	<ul style="list-style-type: none"> • 1. Entwurf der Textzusammenfassung
4. Self-Assessment: Zusammenfassung überprüfen	<ul style="list-style-type: none"> • Relevanz: Wurden Ihrer Meinung nach die Kernaspekte des Textabschnittes identifiziert?
5. Self-Assessment: Weiterführende Fragen formulieren	<ul style="list-style-type: none"> • Welche weiterführenden Fragen stellen Sie sich zu dem Thema?
6. Peer Feedback geben anhand der gleichen Reflektionsprompts wie in 4.)	<ul style="list-style-type: none"> • Redundanz: Wurden die Aussagen des Textabschnittes eindeutig und präzise formuliert?
7. Peer- Feedback einarbeiten	<ul style="list-style-type: none"> • 2. Überarbeitung anhand der Anregungen des Peers

Die Studierenden arbeiteten die Arbeitsanweisungen nacheinander ab. Der Fachtext war sichtbar, wenn Studierende ihre Zusammenfassung erstellten oder über-

arbeiteten. Auch während der beiden Feedbackphasen (Self und Peer) waren der Originaltext sowie die bisher geschriebene Zusammenfassung sichtbar. Für das Peer Feedback tauschten die Teilnehmenden die Computer.

6.4 Maße

Die Textqualität wurde anhand der Aspekte Inhalt (Identifizieren relevanter Aspekte), Vermeidung Redundanz (wiederholende Textstellen) – mit vierstufiger Likertskala (Max. = 4) – und Vermeidung Plagiat (Übernahme Originaltext) – mit dreistufiger Likertskala (Max. = 3) – gemessen. Dazu wurden die Textzusammenfassungen mit einer eigens erstellten und in einem Expertenrating verifizierten „Optimallösung“ mittels der Sprachverarbeitungssoftware „conText“ (Lenhard et al. 2013) verglichen. Die Sprachverarbeitungssoftware identifiziert und sortiert Wörter anhand ihres Kontextes. Dafür benötigt sie einen großen Textkorpus, der für diese Studie aus 201.288 unterschiedlichen, aussagekräftigen Wörtern besteht.

Die Software vergleicht die inhaltliche Abdeckung zwischen Textzusammenfassung der Studierenden und Originaltext mit der inhaltlichen Abdeckung zwischen Optimallösung und Originaltext. Die Software identifiziert ebenfalls Redundanzen in der Zusammenfassung anhand einer Ähnlichkeitsberechnung der einzelnen Sätze zueinander. Diese beiden Aspekte werden von dem Programm prozentual ausgegeben und für unsere Zwecke in eine vierstufige Likertskala zusammengefasst. Die Software markiert Satzteile, die aus drei oder mehr Wörtern in derselben Reihenfolge wie im Originaltext bestehen, als Plagiate. Sodann errechnet sie den prozentualen Anteil plagiierter Satzteile und gibt sie als eine dreistufige Likertskala aus, wobei ab 20% plagiierter Stellen eine geringe Plagiatsvermeidung angenommen wird.

Self-Assessment und Peer Feedback erfolgten über ein vierstufiges Bewertungsschema. Unterschiede zwischen Expertenbewertung, Self-Assessment und Peer Feedback ermittelten wir, indem wir zu jedem Messzeitpunkt (t2, t3) die Differenzen berechneten. Je stärker sich die Werte Null näherten, desto mehr stimmten Experten, Self-Assessment und Peer Feedback überein. Wurde die Expertenbewertung mit den anderen beiden Maßen verglichen, markierten positive Werte eine höhere Expertenbewertung als Self-Assessment und Peer Feedback, negative Werte hingegen eine geringere Expertenbewertung. Wurden Self-Assessment und Peer Feedback verglichen, markierten positive Werte ein höhe-

res Peer Feedback als Self-Assessment und negative Werte entsprechend ein geringeres Peer Feedback.

6.5 Stichprobe

Achtzehn Bachelorstudierende des Faches Grundschulpädagogik (14 Frauen und 4 Männer, $M_{\text{Alter}}=22,3$ Jahre, $SD=0,8$) nahmen an der Studie teil und bearbeiteten alle vier Sitzungen. 15 der 18 Teilnehmer hatten das Abitur mit Leistungskurs Deutsch abgeschlossen. Durchschnittlich lasen die Studierenden in den letzten 3 Monaten 20-50 Seiten wissenschaftliche oder belletristische Literatur ($M_{\text{wissenschaftliche Literatur}}=1,9$; $SD=1,1$; $M_{\text{belletristische Literatur}}=2,3$; $SD=1,0$; $Max.=4$). Aufgrund der kleinen Stichprobe berichten wir im Folgenden Mittelwerte.

7. Ergebnisse

7.1 Verbesserung der Textqualität von t1 zu t4

In diesem Kapitel berichten wir Werte, die mit „conText“ (Lenhard et al. 2013) ermittelt wurden. Bezüglich Inhalt verschlechterte sich die Leistung durchschnittlich in jeder Sitzung ($M_{t1}=3,94$; $SD=0,24$; $M_{t2}=3,76$; $SD=0,31$; $M_{t3}=3,71$; $SD=0,47$; $M_{t4}=2,89$; $SD=0,56$).

Bezüglich Vermeidung Redundanz ($M_{t1}=2,61$; $SD=0,78$; $M_{t2}=3,59$; $SD=0,51$; $M_{t3}=3,41$; $SD=0,51$; $M_{t4}=2,78$; $SD=1,06$) verbesserte sich die Leistung von Sitzung 1 zu Sitzung 2 (mit Self-Assessment und Peer Feedback). Die Leistungen in Sitzung 3 und 4 nahmen wieder kontinuierlich ab, blieben aber über dem Niveau von Sitzung 1.

Die Leistungen bezüglich Vermeidung Plagiat ($M_{t1}=2,50$; $SD=0,79$; $M_{t2}=2,65$; $SD=0,61$; $M_{t3}=2,24$; $SD=0,90$; $M_{t4}=2,72$; $SD=0,67$) zeigten ebenfalls keine klare Tendenz über die vier Sitzungen.

7.2 Self-Assessment vs. Expertenbewertung für die beiden Treatmentzeitpunkte (t2 und t3).

In Inhalt ($M_{t2}=0,38$; $SD_{t2}=0,89$; $M_{t3}=0,18$; $SD_{t3}=0,64$) und Vermeidung Redundanz ($M_{t2}=0,59$; $SD_{t2}=0,80$; $M_{t3}=0,06$; $SD_{t3}=0,71$) schätzten sich die Studierenden im Vergleich zur Expertenbewertung durchschnittlich geringer ein, die Differenz zwischen Self-Assessment und Expertenbewertung ist in Sitzung 3 jeweils geringer als in Sitzung 2.

Bezüglich Vermeidung Plagiat ($M_{t2}=0,56$; $SD_{t2}=0,73$; $M_{t3}=-0,06$; $SD_{t3}=0,90$) schätzten sich die Studierenden in Sitzung 2 im Vergleich zur Expertenbewertung durchschnittlich geringer ein. In Sitzung 3 schätzten sich die Studierenden im Mittel jedoch leicht höher ein.

Tabelle 2: Übersicht der Einschätzungsdifferenzen von Self-Assessment, Peer-Assessment und Expertenbewertung

	Inhalt		Vermeidung Redundanz		Vermeidung Plagiat	
	$M_{t2} (SD_{t2})$	$M_{t3} (SD_{t3})$	$M_{t2} (SD_{t2})$	$M_{t3} (SD_{t3})$	$M_{t2} (SD_{t2})$	$M_{t3} (SD_{t3})$
Self. vs. Experte	0,38 (0,89)	0,18 (0,64)	0,59 (0,80)	0,06 (0,71)	0,56 (0,73)	-0,06 (0,90)
Peer vs. Experte	0,19 (0,51)	0,06 (0,58)	0,25 (0,68)	0,03 (0,67)	-0,13 (0,89)	-0,35 (1,00)
Self vs. Peer	-0,13 (0,89)	0,03 (0,51)	0,59 (0,89)	0,06 (0,71)	1,56 (0,77)	1,32 (0,47)

Anmerkung: Werte = 0 bedeuten, beide Einschätzungen stimmen überein. Werte > 0 bedeuten, Experte schätzt höher ein als Self oder Peer; Peer schätzt höher ein als Self. Werte < 0 bedeuten, Self oder Peer schätzt höher ein als Experte; Self schätzt höher ein als Peer.

7.3 Peer Feedback vs. Expertenbewertung für die beiden Treatmentzeitpunkte (t2 und t3).

Die Aspekte Inhalt ($M_{t2}=0,19$; $SD_{t2}=0,51$; $M_{t3}=0,06$; $SD_{t3}=0,58$) und Vermeidung Redundanz ($M_{t2}=0,25$; $SD_{t2}=0,68$; $M_{t3}=0,03$; $SD_{t3}=0,67$) bewerteten die Peers durchschnittlich etwas geringer als die Experten. Auch hier zeigt sich der Trend, dass die Bewertungen in Sitzung 3 durchschnittlich ähnlicher zur Expertenbewertung sind, als in Sitzung 2.

Bezüglich Vermeidung Plagiat ($M_{t2}=-0,13$; $SD_{t2}=0,89$; $M_{t3}=-0,35$; $SD_{t3}=1,00$) bewerteten die Peers die Textzusammenfassungen zu beiden Messzeitpunkten höher als die Expertenbewertung, in der dritten Sitzung noch ausgeprägter.

7.4 Self-Assessment vs. Peer Feedback für die beiden Treatmentzeitpunkte (t2 und t3).

Bei Inhalt ($M_{t2}=0,38$; $SD_{t2}=0,62$; $M_{t3}=0,03$; $SD_{t3}=0,51$) und Vermeidung Redundanz ($M_{t2}=0,59$; $SD_{t2}=0,89$; $M_{t3}=0,06$; $SD_{t3}=0,71$) bewerteten die Peers durchschnittlich etwas höher als die Studierenden sich selbst, der Unterschied nahm von Sitzung 2 zu Sitzung 3 jedoch ab.

Peers schätzten die Plagiatsvermeidung ihrer Kommilitoninnen und Kommilitonen jedoch deutlich höher ein als diese selbst ($M_{12}=1,56$; $SD_{12}=0,77$; $M_{13}=1,32$; $SD_{13}=0,47$)

Nachdem die Peers die Zusammenfassungen ihrer Kommiliton/innen anhand der gleichen Reflektionsprompts wie im Self-Assessment (vgl. Tabelle 1) bewertet hatten, wurden sie aufgefordert, in einem Textfeld elaborierte Rückmeldung zu schreiben („*Welche Anregungen können Sie Ihrer Partnerin/ Ihrem Partner geben?*“). Die Rückmeldungen bestehen häufig aus einem Lob (z.B. „*Hast du super gemacht!*“ oder „*Weiter so!*“), Hinweisen auf Rechtschreibfehler oder Interpunktion (z.B. „*Achte auf die Rechtschreibung im 1. Satz*“). Hinweise zur Organisation oder Struktur der Textzusammenfassung konnten ebenfalls identifiziert werden (z.B. „*Es wäre besser, den zweiten Satz in zwei separate Sätze zu trennen, denn dann wäre er leichter zu verstehen.*“), jedoch deutlich seltener. Hinweise zum Inhalt sind meist unkonkret (z.B. „*Evtl. mehr ausführen.*“) und werden kaum gegeben. Allerdings ist die Inhaltsabdeckung in den Messzeitpunkten zwei und drei relativ hoch, dies könnte ein Grund dafür sein, dass Studierende eine Rückmeldung zum Inhaltsaspekt eventuell als nicht notwendig erachteten.

8. Diskussion

Ziel dieser Pilotstudie war, erste Eindrücke zur Textqualität von Studierenden zu erhalten und Self-Assessment, Peer Feedback und Expertenbeurteilungen miteinander in Beziehung zu setzen.

8.1 Textqualität

Eine allgemeine Verbesserung der Textqualität in den drei Aspekten Inhalt, Vermeidung Redundanz und Vermeidung Plagiat konnte nicht beobachtet werden. Vielmehr verschlechterten sich die Studierenden bei der Erfassung relevanten Inhalts im Verlauf des Trainings. Gemäß der guten Ausgangswerte haben die Studierenden anscheinend kaum Probleme, relevante Inhalte aus Fachtexten zu extrahieren, deren Komplexität unter einem LIX von 70 (Lenhard & Lenhard 2014-2017) eingestuft wird. Bei komplexeren Fachtexten ($LIX > 70$) scheint es Studierenden aber schwerer zu fallen, alle relevanten Inhaltsaspekte in ihre Textzusammenfassungen aufzunehmen. Die Studierenden könnten sich jedoch im Verlauf des Schreibtrainings weniger angestrengt haben. Wir schließen dar-

aus, dass die verwendeten Übungstexte zu leicht waren und werden sie deshalb durch komplexere sachunterrichtliche Fachtexte ($LIX > 70$) ersetzen, zusätzlich wird bei jedem Text die Anstrengungsbereitschaft erhoben werden.

Bezüglich der Vermeidung redundanter Textstellen beobachteten wir eine Verbesserung der Leistung von Sitzung 1 zu Sitzung 2. Wir vermuten einen Zusammenhang mit dem Einsatz der Prompts, die in Sitzung 2 zum ersten Mal eingesetzt wurden und die Studierenden zum Self-Assessment und Peer Feedback angeregt haben. Im weiteren Verlauf nahm die Leistung wieder leicht ab. Das kann ein Indikator dafür sein, dass die Prompts die Studierenden beim Erkennen von Redundanzen nicht ausreichend unterstützten. Es fällt den Studierenden v.a. schwer, strukturelle Redundanzen zu identifizieren (Roscoe et al. 2015), deshalb werden wir Prompts formulieren, die mehr auf das Erkennen solcher strukturellen Redundanzen abzielen.

Die Studierenden konnten zuletzt besser auf den Originaltext in ihrer Zusammenfassung verzichten. Eine herausfordernde Aufgabe bei Textzusammenfassungen ist das Umformulieren in eigene Worte (Friend 2001). Bestehende Forschung weist darauf hin, dass Studierende, denen das Vermeiden von Plagiaten gelingt, den Textgegenstand auch besser erfasst haben (Becker-Mrotzek et al. 2014). Diese Interpretation scheint uns für unsere Studie jedoch nur teilweise zulässig, da es den Studierenden in Sitzung 4 jeweils weniger gut gelang, relevante Inhalte zu identifizieren. Dieser Zusammenhang könnte erforscht werden, indem beispielsweise komplexe Inhaltsfragen zum Text beantwortet werden müssen.

8.2 Vergleich Self-Assessment Peer Feedback und Expertenbewertung

Beim Vergleich von Self-Assessment mit Peer Feedback und Expertenbewertung lässt sich für unsere Studie schlussfolgern, dass die Studierenden sich selbst bzgl. Inhalt und Vermeidung Redundanz eher schlechter als die Experten einschätzten. Im Gegensatz dazu schätzten Peers die Textzusammenfassungen eher besser ein als die Studierenden. Ähnliche Bewertungstendenzen wurden auch in anderen Studien erkannt (Falchikov 2006).

Die Resultate weisen jedoch auch darauf hin, dass es den Studierenden schwer zu fallen scheint, Plagiate zu identifizieren und ihre Mitstudierenden darauf hinzuweisen. Das oberflächliche Peer Feedback könnte unter anderem daran liegen,

dass die Studierenden nur kurz darin eingewiesen wurden. Die Prompts konnten sie nicht ausreichend unterstützen, substantielles Feedback zu geben.

Dies bestätigt Forschungsergebnisse zum eher oberflächlichen Peer Feedback und zu den Schwierigkeiten, strukturelle Probleme in den Texten zu identifizieren (Hoogeveen & van Gelderen 2013, Sommers 1980). In Bezug auf Self-Assessment und Peer Feedback werden wir deshalb Prompts formulieren, die das Erkennen von Plagiaten besser unterstützen.

8.3 Limitationen und Schlussfolgerungen

Wir möchten auf zwei einschränkende Aspekte in dieser Studie hinweisen. Erstens ist die Stichprobe sehr klein. Zweitens ist der Interventionszeitraum mit nur zwei Sitzungen sehr kurz. Effekte sind dadurch bei einer so komplexen Aktivität wie dem Schreiben nicht offenzulegen.

Aus unseren Ergebnissen können wir schließen, dass ein Schreibtraining Potential birgt, Studierende bei der Zusammenfassung von sachunterrichtlichen Fachtexten zu unterstützen. Da regelmäßige Übung eine große Bedeutung für die Entwicklung von Schreibkompetenz hat (Bannert & Reimann 2012, Kellogg & Raulerson 2007), wird das Schreibtraining auf sechs Sitzungen erweitert. Außerdem werden sachunterrichtliche Fachtexte mit höherer Komplexität ausgewählt. Hinsichtlich von Self-Assessment und Peer Feedback sollen Prompts ausdifferenziert werden, um den Blick auf strukturelle, verständnisorientierte Überarbeitungen im Text zu lenken (Proske et al. 2012). Zudem soll den Studierenden der Sinn von Self-Assessment und Peer Feedback erläutert werden (Panadero & Alonso-Tapia 2013, Hoogeveen & van Gelderen 2013).

Darüber hinaus sollen die Studierenden individuelles und ortsunabhängiges Feedback bekommen können. Deshalb soll die Möglichkeit in Betracht gezogen werden, den Studierenden automatisiertes Feedback durch ein intelligentes, tutorielles System zu geben, welches auf Latenter Semantischer Analyse (Lenhard 2008) basiert. Dazu wird die Sprachverarbeitungssoftware „conText“ (Lenhard et al. 2013) weiterentwickelt, welche in der Sachunterrichtsausbildung eingesetzt werden soll. Durch eine solche Bearbeitung von Fachtexten zu formativem Assessment, Denk- und Arbeitsweisen und Scaffolding im Sachunterricht können die Studierenden sich intensiv mit diesen Theorien befassen, ein nachhaltiges Verständnis dafür entwickeln und so ihr pädagogisches Fachwissen erweitern und vertiefen.

Literatur

- Bannert, M. & Reimann, P. (2012): Supporting self-regulated hypermedia learning through prompts. In: *Instructional Science*, 40, 193-211.
- Bannert, M.; Sonnenberg, C.; Mengelkamp, C. & Pieger, E. (2015): Short- and long-term effects of students' self-directed metacognitive prompts on navigation behavior and learning performance. In: *Computers in Human Behavior*, 52, 293-306.
- Baumert, J.; Kunter, M.; Blum, W.; Brunner, M.; Voss, T.; Jordan, A.; Klusmann, U.; Krauss, S.; Neubrand, M. & Tsai, Y.M. (2010): Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. In: *American Educational Research Journal*, 47, 1, 133-180.
- Becker-Mrotzek, M.; Grabowski, J.; Jost, J.; Knopp, M. & Linnemann, M. (2014): Adressatenorientierung und Kohärenzherstellung im Text: Zum Zusammenhang kognitiver und sprachlich realisierter Teilkomponenten von Schreibkompetenz. In: *Halbjahresschrift für die Didaktik der deutschen Sprache und Literatur*, 19, 21-43.
- Blömeke, S.; Gustafsson, J.E. & Shavelson, R.J. (2015): Beyond dichotomies competence viewed as a continuum. In: *Journal of Psychology*, 223, 1, 3-13.
- Boud, D.; Lawson, R. & Thompson, D.G. (2013): Does student engagement in self-assessment calibrate their judgement over time? In: *Assessment & Evaluation in Higher Education*, 38, 941-956.
- Butler, J.A. & Britt, M.A. (2010): Investigating instruction for improving revision of argumentative essays. In: *Written Communication*, 28, 70-96.
- Falchikov, N. (2006): Peer Feedback Marking: Developing Peer Assessment. In: *Innovations in Education and Training International*, 32, 175-187.
- Friend, R. (2001): Effects of strategy instruction on summary writing of college students. In: *Contemporary Educational Psychology*, 26, 3-24.
- Gan, M.J.S. & Hattie, J. (2014): Prompting secondary students' use of criteria, feedback specificity and feedback levels during an investigative task. In: *Instructional Science*, 42, 861-878.
- Graham, S., Hebert, M. & Harris, K.R. (2015): Formative assessment and writing: A Meta-Analysis. In: *The Elementary School Journal*, 115, 523-547.
- Grygier, P. & Hartinger, A. (2009): *Gute Aufgaben Sachunterricht: Naturwissenschaftliche Phänomene begreifen; 48 gute Aufgaben für die Klassen 1 bis 4*. Berlin.
- Hayes, J.R. (2012): Modeling and remodeling writing. In: *Written Communication*, 29, 369-388.
- Hill, H.C.; Rowan, B. & Ball, D.L. (2005): Effects of teachers' mathematical knowledge for teaching on student achievement. In: *American Educational Research Journal*, 42, 2, 371-406.
- Hoogeveen, M. & van Gelderen, A. (2013): What works in writing with peer response? A review of intervention studies with children and adolescents. In: *Educational Psychology Review*, 25, 473-502.
- Hume, A. & Berry, A. (2011): Constructing CoRes – a strategy for building PCK in pre-service science teacher education. In: *Research in Science Education*, 41, 3, 341-355.
- Kellogg, R.T. & Raulerson, B.A. (2007): Improving the writing skills of college students. In: *Psychonomic Bulletin & Review*, 14, 237-242.

- Kellogg, R.T. & Whiteford, A.P. (2012): The development of writing expertise. In: Grigorenko, E.L.; Mambrino, E. & Preiss, D.D. (Eds.): *Writing: a mosaic of new perspectives*. New York, 109-124.
- Kieft, M.; Rijlaarsdam, G.; Galbraith, D. & van den Bergh, H. (2007): The effects of adapting a writing course to students' writing strategies. In: *British Journal of Educational Psychology*, 77, 565-578.
- Kulgemeyer, C. & Riese, J. (2018): From professional knowledge to professional performance: The impact of CK and PCK on teaching quality in explaining situations. In: *Journal of Research in Science Teaching*, 55, 10, 1393-1418.
- Kunter, M.; Klusmann, U.; Baumert, J.; Richter, D.; Voss, T. & Hachfeld, A. (2013): Professional competence of teachers: Effects on instructional quality and student development. In: *Journal of Educational Psychology*, 105, 3, 805-820.
- Lenhard, W. & Lenhard, A. (2014-2017): Berechnung des Lesbarkeitsindex LIX nach Björnson. Bibergau.
- Lenhard, W. (2008): Bridging the gap to natural language: A review on intelligent tutoring systems based on latent semantic analysis. URL: <https://nbn-resolving.org/urn:nbn:de:bvb:20-opus-27980> [17.12.2018].
- Lenhard, W.; Baier, H.; Lenhard, A.; Hoffmann, J. & Schneider, W. (2013): conText – Förderung des Leseverständnisses durch das Arbeiten mit Texten. Göttingen.
- MacArthur, C.A.; Philippakos, Z.A. & Ianetta, M. (2015): Self-regulated strategy instruction in college developmental writing. In: *Journal of Educational Psychology*, 107, 855-867.
- Meschede, N.; Fiebranz, A.; Möller, K. & Steffensky, M. (2017): Teachers' professional vision, pedagogical content knowledge and beliefs: On its relation and differences between pre-service and in-service teachers. In: *Teaching and Teacher Education*, 66, 158-170.
- Mok, W.S.Y. & Chan, W.W.L. (2016): How do tests and summary writing tasks enhance long-term retention of students with different levels of test anxiety? In: *Instructional Science*, 44, 567-581.
- Möller, K. & Steffensky, M. (2010): Naturwissenschaftliches Lernen im Unterricht mit 4- bis 8-jährigen Kindern. Kompetenzbereiche frühen naturwissenschaftlichen Lernens. In: Leuchter, M. (Hrsg.): *Didaktik für die ersten Bildungsjahre. Unterricht mit 4- bis 8-jährigen Kindern*. Zug, 163-178.
- Möller, K.; Bohrmann, M.; Hirschmann, A.; Wilke, T. & Wyssen, H.-P. (2013): *Spiralcurriculum Magnetismus: Naturwissenschaftlich arbeiten und denken lernen*. Seelze.
- Panadero, E. & Alonso-Tapia, J. (2013): Self-assessment: theoretical and practical connotations: when it happens, how is it acquired and what to do to develop it in our students. In: *Electronic Journal of Research in Educational Psychology*, 11, 551-579.
- Panadero, E. & Brown, G.T.L. (2016): Teachers' reasons for using peer assessment: positive experience predicts use. In: *European Journal of Psychology of Education*, 32, 133-156.
- Park, S.; Jang, J.Y.; Chen, Y.C. & Jung, J. (2011): Is pedagogical content knowledge (PCK) necessary for reformed science teaching? Evidence from an empirical study. In: *Research in Science Education*, 41, 2, 245-260.

- Perin, D.; Lauterbach, M.; Raufma, J. & Kalamkarian, H.S. (2017): Text-based writing of low-skilled postsecondary students: Relation to comprehension, self-efficacy and teacher judgments. In: *Reading and Writing*, 30, 887-915.
- Proske, A.; Narciss, S. & McNamara, D.S. (2012): Computer-based scaffolding to facilitate students' development of expertise in academic writing. In: *Journal of Research in Reading*, 35, 136-152.
- Roscoe, R.D.; Snow, E.L.; Allen, L.K. & McNamara, D.S. (2015): Automated detection of essay revising patterns: Applications for intelligent feedback in a writing tutor. In: *Technology, Instruction, Cognition and Learning*, 10, 59-79.
- Schnotz, W. (2006). Was geschieht im Kopf des Lesers? Mentale Konstruktionsprozesse beim Textverstehen aus der Sicht der Psychologie und der kognitiven Linguistik. In: Blühdorn, H.; Breindl, E. & Waßner, U.H. (Hrsg.): *Text – Verstehen. Grammatik und darüber hinaus*. Berlin, 222-238.
- Shulman, L.S. (1987): Knowledge and teaching: Foundations of the new reform. In: *Harvard Educational Review*, 57, 1, 1-22.
- Sodian, B. & Mayer, D. (2013): Entwicklung des wissenschaftlichen Denkens im Vor- und Grundschulalter. In: Stamm, M. & Edelman, D. (Hrsg.): *Handbuch frühkindliche Bildungsforschung*. Wiesbaden.
- Sommers, N. (1980): Revision strategies of student writers and experienced adult writers. In: *College Composition and Communication*, 31, 378-388.
- Topping, K.J.; Smith, E.F.; Swanson, I. & Elliot, A. (2000): Formative peer assessment of academic writing between postgraduate students. In: *Assessment & Evaluation in Higher Education*, 25, 149-169.
- Westby, C.; Culatta, B.; Lawrence, B. & Hall-Kenyon, K. (2010): Summarizing expository texts. In: *Topics in Language Disorders*, 30, 275-287.