

Anastasia Drackert (Bochum), Darja Felberg (Bochum)

Ausbildung der TestentwicklungsKompetenz angehender LehrerInnen und LinguistInnen durch Forschendes Lernen

The article describes a research-based project within the seminar *‘Testing, Assessing and Evaluating L2 Competence in Russian for teaching and research purposes’*, whose objective was to train future teachers of L2 Russian and Slavists in research-based learning and improve their test development literacy. During the semester the students developed and piloted a C-test for placement purposes at the Department of Slavic Studies, which can also be used to assess general language proficiency in Russian. The article presents details of the ten-step-project. Subsequently, the perceived learning outcomes of the seminar participants are summarized and discussed in relation to research-based learning.

В этой статье описывается проект, реализованный в рамках семинара «*Оценка уровня владения русским языком для образовательных и научных целей*» на факультете славистики Пурского университета Бохума. Целью проекта было развитие компетенции будущих учителей и славистов в разработке языковых тестов с помощью методики исследовательского обучения. В течение семестра студенты должны были разработать и опробовать вступительный С-тест для факультета славистики, который также можно использовать для определения уровня общего владения русским языком. В статье представлена подробная информация о десяти этапах проекта. В заключение приводятся результаты обучения с точки зрения участников семинара и анализируются перспективы исследовательского обучения.

1. Einführung

Will man die sprachlichen Fähigkeiten von Lernenden feststellen, ist die Anwendung von Tests und zumeist auch deren Entwicklung unabdingbar. Insbesondere im Fremdsprachenunterricht stellt das Testen und Evaluieren fremdsprachlicher Kenntnisse und Fähigkeiten einen entscheidenden Bestandteil dar, um den Prozess des Fremdsprachenerwerbs der Lernenden gezielt voranzubringen. Eine Lehrkraft für Russisch als Fremdsprache beschäftigt sich bis zu 50% ihrer Arbeitszeit mit der Beurteilung von Leistungen (Drackert & Stadler 2017). Aus diesem Grund müssen (zukünftige) Lehrende im Umgang mit der Entwicklung und Anwendung von Tests kompetent geschult sein.

Auch in der linguistisch orientierten Spracherwerbsforschung spielen Tests eine entscheidende Rolle. Inwieweit z.B. bilinguale LernerInnen eine neue Fremdsprache schneller lernen als monolinguale, oder inwiefern bei HerkunftssprecherInnen auf bestimmten sprachlichen Ebenen Sprachverlust stattfindet, wird auch vor allem anhand von Sprachtests oder Testaufgaben beurteilt. Da Forschungsfragen in der (Angewandten) Sprachwissenschaft ebenso anhand von Tests beantwortet werden, ist ihre Qualität sowie die Kompetenz, gute Tests zu entwickeln, auch für LinguistInnen von hoher Relevanz (Drackert 2016).

Im Lehramtsstudium, aber auch im Laufe eines philologischen Studiums gibt es mit wenigen Ausnahmen kaum Veranstaltungen zur Ausbildung der Testentwicklungskompetenz (Drackert & Stadler 2017; Hasselgreen, Carlsen & Helness 2004; Popham 2011; Vogt 2010), daher verfügen die beiden Zielgruppen demnach nicht über das nötige *Know-how* und müssen Schülerleistungen bewerten oder Testinstrumente entwickeln, ohne in der Testentwicklung geschult zu sein (Gunn & Gilmore 2014). Besonders her-

ausfordernd ist die Situation in den weniger gelernten und gelehrteten Sprachen wie Russisch als Fremdsprache, für die es im Vergleich zu Englisch als Fremdsprache kaum Literatur zum Testen und nur sehr wenige Quellen für *best practice* Beispiele guter oder bereits pilotierter Tests gibt.

Der Prozess der Testerstellung verläuft periodisch und ist mit viel Aufwand verbunden, daher sollte eine Ausbildung in diesem Bereich entsprechend umfangreich sein. Essentiell ist hierbei, dass die Aneignung der Testkompetenz durch die Erfahrung in der selbstständigen Testentwicklung erfolgt. Die bloße Auseinandersetzung mit der theoretischen Materie schult die Studierenden nur unzureichend.

2. Testentwicklungskompetenz und deren Ausbildung

Testentwicklungskompetenz ist ein Teil des größeren Konstrukt *language assessment literacy* (LAL). Vogt definiert den Begriff LAL als Konstrukt, welches mehrere Komponenten umfasst: „solides Wissen über Mechanismen und Prozesse der Beurteilung“, „Anspruch an eigene Professionalität zur Durchführung qualitativ hochwertiger Leistungsmessung und -beurteilung“ und die „Fähigkeit zur kritischen Evaluation von Tests und zur Auswahl und Erstellung angemessener Formen der Leistungsmessung und -beurteilung“ (Vogt 2010: 268). Je nach Zielgruppe (z.B. ForscherInnen oder LehrerInnen) können unterschiedliche Komponenten der Entwicklungskompetenz mehr oder weniger wichtig sein (Malone 2013). Unabhängig von der Zielgruppe stellt die Testentwicklungskompetenz, also die **Erstellung** angemessener Formen der Leistungsmessung und -beurteilung, eine bedeutende Komponente der Beurteilungskompetenz dar.

2.1 Testentwicklungsprozess

Der Prozess der Testerstellung fängt mit der Entscheidung an, einen Test zu einem bestimmten Zweck anzubieten. Es muss entschieden werden, **wozu – also zu welchem Ziel – der Test angewendet wird** und welche Entscheidungen anhand der Interpretation der Testergebnisse zu treffen sind. Je nach Testzweck werden die Ergebnisse unterschiedlich gehandhabt, z. B. wenn Lernende in Niveaus eingestuft, deren sprachliche Kompetenzen bestimmt, ihre Leistung und ihr Fortschritt evaluiert oder ihr Förderbedarf aufgezeigt werden sollen.

Im nächsten Schritt ist festzustellen, **was der Test messen** und welchen Kontext der Sprachverwendung der Test widerspiegeln soll. Fremdsprachenkenntnisse werden als Fertigkeiten (z.B. Hörverstehen), sprachliche Mittel (Grammatik, Lexik) oder als eine Kombination aus beidem gemessen (C-Test, *Elicited Imitation Test*).

Als Nächstes erfolgt der eigentliche Prozess der **Testerstellung**. Zum einen müssen ein Test und seine Aufgaben derart spezifiziert sein, dass sie die Sprachfähigkeiten der Lernenden dem gegebenen Kontext entsprechend messen. Zum anderen ist das konkrete Wissen über die Bewertung einzelner Kompetenzen sowie über die Besonderheiten einzelner Aufgabenformate erforderlich. Es müssen zudem der Rahmen des Tests (Ort, Zeit, TeilnehmerInnen), die Testanweisungen, die Zeiteinteilung, die Gewichtung der einzelnen Testbestandteile, der Input und seine Sprache, die erwarteten Antworten und die Bewertungsmethoden bestimmt werden (vgl. Eberharter, Kremmel & Zehentner 2018). An dieser Stelle sollen die Praktikabilität und die zur Verfügung stehenden Ressourcen in einem bestimmten Lernkontext in Erwägung gezogen werden. Diese können vorhandene Möglichkeiten eingrenzen (z.B. Nichtvorhandensein von Computern für adaptive Tests) oder erweitern.

Der Testerstellung folgt die **Pilotierung des Tests**. Grundsätzlich gilt: Je wichtiger die Entscheidungen, die anhand der Testergebnisse getroffen werden, desto aufwändiger sollte die Erprobung sein (Drackert 2016). Während der Erprobung erhalten TesterstellerInnen wichtige Informationen und Rückmeldungen zum Test, u.a. über den Schwierigkeitsgrad und die Plausibilität einzelner Antworten oder Formulierungen. Sobald die Testergebnisse vorliegen, ist es sinnvoll, diese einer Datenanalyse zu unterziehen. Um die Reliabilität und Validität eines Tests und seiner Ergebnisse zu untersuchen, sollten TestentwicklerInnen zumindest die am häufigsten verwendeten Ansätze der Datenanalyse verstehen und interpretieren können. Eine korrekte **Auswertung der Ergebnisse** dient der Überarbeitung und somit **Verbesserung des Tests** für zukünftige Durchführungen (Douglas 2010).

Die Komplexität des Testentwicklungsprozesses kann im Rahmen dieses Beitrags nicht in all ihrer Detailliertheit erläutert werden. Hier sei auf Handbücher für unterschiedliche Kontexte (z.B. das Handbuch zur Entwicklung und Durchführung von Sprachtests von ALTE 2012) verwiesen. Hervorzuheben ist aber die Tatsache, dass die Logik einzelner Schritte erst im Prozess der Testerstellung begriffen werden kann. Dies kann z.B. durch Forschendes Lernen erreicht werden.

2.2 Forschendes Lernen

Durch den sogenannten Bologna-Prozess, einen Hochschulreformprozess, der die Etablierung einheitlicher Standards auf europäischer Ebene zum Ziel hatte, werden europäische Hochschulen gefordert, ihre Studiengänge an Beschäftigungsfähigkeit (*employability*) und Kompetenzentwicklung auszurichten (Mieg 2017; Fichten 2017). Um diesen Anforderungen ent-

gegenzukommen, bietet sich Forschendes Lernen, das 1970 erstmals offiziell als Begriff in der Hochschulpolitik auftrat, als eine geeignete Lehr- und Lernform an der Hochschule an (Reinmann 2015; Mieg 2017).

Beim Forschenden Lernen steht das Vermitteln bereits bestehender Erkenntnisse im Hintergrund, es zeichnet sich vielmehr durch die „aktive Teilhabe der Studierenden am Prozess der Wissensgewinnung“ aus (Pasternack 2017: 38f). Die Studierenden stehen im Mittelpunkt. Sie sind die EntscheidungsträgerInnen hinsichtlich der Organisation und Umsetzung von Forschungsvorhaben, die Lehrkraft unterstützt sie und gibt Feedback (Mieg 2017; Wiemer 2017). Dies bedeutet aber nicht, dass „Studierende sich an den Fronten der Forschung bewegen sollen“, d.h. sie sollen keine „neuen“ Erkenntnisse erarbeiten (Pasternack 2017: 38f). Dennoch ist Forschendes Lernen vor allem durch Selbstständigkeit, Theoriebezug und Reflexion seitens der Studierenden gekennzeichnet (Fichten 2017). Im engeren Sinne zeichnet sich Forschendes Lernen

vor anderen Lernformen dadurch aus, dass die Lernenden den Prozess eines Forschungsvorhabens [...] in seinen wesentlichen Phasen – von der Entwicklung der Fragen und Hypothesen über die Wahl und Ausführung der Methoden bis zur Prüfung und Darstellung der Ergebnisse in selbstständiger Arbeit oder in aktiver Mitarbeit in einem übergreifenden Projekt – (mit)gestalten, erfahren und reflektieren (Huber 2009: 11).

Die Abfolge der Phasen erfolgt jedoch in der Praxis nicht unbedingt linear. Die Studierenden sollten diese Phasen „möglichst vollständig durchlaufen, wenigstens im Blick haben“, was allerdings sehr stark vom Kontext abhängig ist und sich nicht immer realisieren lässt (Huber 2014: 23).

Dadurch, dass Forschendes Lernen Studierenden ermöglicht, Wissen eigenverantwortlich zu organisieren und zu reflektieren, findet *deep level learning*, d.h. nachhaltiges Lernen statt (Huber 2009). Lernen geht über Rezeption und Reproduktion hinaus (Huber 2009). Inhalte werden von den

Studierenden durch eigenständige, aktive, produktive Forschung effektiver erlernt. Außerdem fördert Forschendes Lernen den Erwerb von (Schlüssel-)Kompetenzen, die es den Studierenden erlauben, selbstständig neue Erkenntnisse zu generieren (Wulf 2017; Hofer 2013), z.B. „Kommunikations- und Teamfähigkeit, Präsentationstechniken, der Umgang mit modernen Informationstechnologien, interkulturelle Kompetenzen und Fremdsprachenkenntnisse, die Fähigkeit, Wissen und Informationen zu verdichten und zu strukturieren“ (Wissenschaftsrat 2000: 21f), „Kompetenzen zum Zeit- und Aufgabenmanagement, Diskussionsfähigkeit und Teamarbeit“ (Geeb, Krauß-Leichert & Verch 2009: 164), um nur einige zu nennen.

Die Bandbreite an Lehr- und Lernmöglichkeiten, die durch „Variation in der Umsetzung [Forschenden Lernens] und durch Kombination mit anderen Formen forschungsnahen Lernens“ entstehen können, ist grenzenlos (Reinmann 2015: 121). Dies kann im Kontext des Lehrens und Lernens an Hochschulen genutzt werden, um auf die Heterogenität der Studierendenzugangspopulation einzugehen und sie gewinnbringend im Forschungsprozess einzusetzen (Reinmann 2015).

3. Projekt

Das Testentwicklungsprojekt wurde im Rahmen des zum ersten Mal stattgefundenen Seminars ‚*Prüfen, Testen, Evaluieren der Sprachkompetenzen im Russischen in Lehre und Forschung*‘ im Sommersemester 2017 am Seminar für Slavistik der Ruhr-Universität Bochum durchgeführt. Diese im linguistischen Modul verankerte Veranstaltung hatte zum Ziel, angehende SlawistInnen und RussischlehrerInnen in die Grundlagen des Testens, Prüfens und Evaluierens einzuführen und behandelte folgende Fragen: Welche Funktionen hat der Einsatz von Tests und Testaufgaben im Unterricht

und in der (Sprachlehr- bzw. -lern-)Forschung? Was sind die Kriterien eines guten Tests? Welche Aufgaben und Testtypen sind für das Messen einzelner Kompetenzen geeignet? Wie testet man die soziopragmatische Kompetenz im Russischen? Mit welchen Methoden kann man Vokabelkenntnisse im Russischen testen? Wie können Filmabschnitte oder Bildergeschichten für die Evaluation der Russischkenntnisse eingesetzt werden? Im Laufe des Seminars hatten die Studierenden die Möglichkeit, selbst an mehreren Bewertungsverfahren teilzunehmen und an einem gemeinsam geplanten Testentwicklungsprojekt mitzuarbeiten, der im Fokus des vorliegenden Beitrags steht. Bei dem Projekt handelte es sich um die Entwicklung eines C-Tests zu Einstufungszwecken am Seminar für Slavistik, der aber auch generell zur Ermittlung der allgemeinen Sprachkompetenz benutzt werden kann. Die erste Autorin hat das Seminar geleitet, die zweite Autorin war im Seminar als studentische Hilfskraft eingesetzt.

3.1 C-Test: Format und Anwendung

Der C-Test ist ein praktikables und reliables Verfahren zur globalen Ermittlung der allgemeinen Kompetenz in Fremd-, Zweit- und Erstsprachen (Baur & Spettmann 2007; Eckes & Grotjahn 2006). Wie andere *cloze tests* basieren auch C-Tests auf dem Prinzip der reduzierten Redundanz (Klein-Braley 1997). Ein C-Test besteht in der Regel aus fünf bis acht Texten unterschiedlicher Thematik und Schwierigkeit. Beginnend mit dem zweiten Wort des zweiten Satzes wird die zweite Hälfte jedes zweiten Wortes getilgt. Dabei bleiben Wörter aus einem einzigen Buchstaben, Eigennamen und Zahlen unberücksichtigt. Bei Wörtern mit einer ungeraden Anzahl von Buchstaben wird die Hälfte des Wortes plus ein Buchstabe getilgt. Der erste und der letzte Satz bleiben unverändert und bilden den Kontext für die Lösungsfindung. In der Regel bestehen C-Test-Texte aus 60

bis 80 Wörtern, 20 oder 25 von ihnen sind getilgt¹ (siehe Anhang 1 für ein Beispiel eines C-Test-Textes). Die Aufgabe der TestteilnehmerInnen besteht darin, die Lücken des C-Tests zu vervollständigen. Die übliche Zeit für die Lösung eines C-Test-Textes beträgt 5 Minuten.

Die Vervollständigung einer Lücke erfordert die Rekonstruktion der Bedeutung eines getilgten Wortes sowie dessen Form. Sie greift auf unterschiedliche Kompetenzen (z.B. Lesen) und Kenntnisse zurück (z.B. Grammatik, Orthographie) und erfordert den Einsatz unterschiedlicher Strategien. Zahlreiche Studien lieferten Hinweise bezüglich der Eignung dieses Testformats für die Ermittlung der allgemeinen Sprachkompetenz (u.a. Drackert & Timukova 2019) und zeigten vielfältige Anwendungsbereiche des C-Tests, z.B. für Einstufungszwecke auf (Norris 2018; Mozgalina & Ryshina-Pankova 2015; Steurer 1996).

In dem Seminar fiel die Wahl des Testformats auf den C-Test, weil dieser sowohl in der Praxis als auch in der Forschung breit eingesetzt wird und nicht nur für Philologen allgemein, sondern auch für die LehramtskandidatInnen von hoher Relevanz sein kann.

3.2 TeilnehmerInnen

An dem Seminar nahmen elf Bachelor- und Master-Studierende des Seminars für Slavistik teil, acht von ihnen im Studiengang *Master of Education*. Es handelte sich um eine sehr heterogene Teilnehmergruppe, in der FremdsprachenlernerInnen, HerkunftssprecherInnen und MuttersprachlerInnen des Russischen vertreten waren. Mit anderen Worten: Es gab im Kurs SeminarteilnehmerInnen, die erst seit vier Semestern Russisch gelernt hatten,

¹ Siehe Grotjahn (2002) für weitere Konstruktionsregeln.

und Studierende, die in einem russischsprachigen Land bereits ein Studium abgeschlossen hatten.

3.3 Schritte des Projekts

Das Testestellungsprojekt erstreckte sich über das ganze Semester. Der Zeitplan des Projekts, integriert in den Semesterplan, befindet sich in Anhang 2. Das Projekt bestand aus zehn Schritten:

1. Kennenlernen des C-Test-Formats aus LernerInnenperspektive
2. Kennenlernen der Theorie zum C-Test
3. Bestimmung des Zwecks des zu entwickelnden C-Tests und Beschreibung des Lernkontextes
4. Auswahl der Texte & Erstellung der Lückentexte
5. Rückmeldung zu den erstellten Texten durch
 - a. Seminarleitung
 - b. LektorInnen am Seminar für Slavistik
6. Auswahl der Texte für die Pilotierung
7. Pilotierung
8. Testkorrektur, Dateneingabe und Datenauswertung
9. Interpretation der Ergebnisse
10. Reflexion

3.3.1 Kennenlernen des C-Test-Formats aus LernerInnenperspektive

Bereits in der ersten Einheit wurden die SeminarteilnehmerInnen aufgefordert, einen kurzen C-Test in der russischen Sprache zu lösen und ihre

Erfahrung als TestteilnehmerIn zu reflektieren. Unter anderem wurden folgende Fragen besprochen: Was misst der Test? Welche Kenntnisse und Kompetenzen braucht man, um den Test lösen zu können? Wozu könnte man so einen Test benutzen? Wie viel Zeit braucht man, um einen C-Test lösen zu können? Diese Fragen wurden erst einmal ohne Einbezug der Forschungsliteratur beantwortet, was die Studierenden zur Reflexion aufforderte. Viel wichtiger ist jedoch, dass sie beim späteren Lesen der Literatur zum C-Test auf die eigene Reflexion zurückgreifen und Vergleiche anstellen konnten.

Zudem haben die TeilnehmerInnen jeweils einen C-Test korrigiert, die Punktwerte für alle TeilnehmerInnen wurden in eine Excel-Tabelle eingegeben, die resultierende Verteilung der Testergebnisse wurde durch die Seminarleitung visualisiert und mit der allgemeinen Sprachkompetenz der SeminarteilnehmerInnen, auch im Zusammenhang mit ihrer Lernbiographie, verglichen. So konnten sie evidenzbasiert die Verlässlichkeit des Testformats persönlich beurteilen.

3.3.2 Kennenlernen der Theorie zum C-Test

In der dritten Sitzung wurden einige theoretische Aspekte wie das C-Test-Konstrukt, Gütekriterien für Tests sowie Vor- und Nachteile des C-Test-Formats, Anwendungsbereiche von C-Tests und die C-Test-Konstruktion in Anlehnung an die gelesene Literatur (Grotjahn 2002; Steurer 1996) besprochen. Da die Konstruktion des C-Tests den nächsten eigenständigen Arbeitsschritt der Studierenden darstellte (siehe 3.3.4), wurde darauf besonders detailliert eingegangen.

3.3.3 Bestimmung des Zwecks des zu entwickelnden Tests und Beschreibung des Lernkontextes

Gleich im Anschluss wurde das Projektvorhaben im Detail präsentiert: Im Laufe des Semesters sollte ein C-Test zu Einstufungszwecken am Seminar für Slavistik entwickelt werden. Alle Studierenden waren mit dem Lernkontext des Seminars für Slavistik vertraut, wenn auch in unterschiedlichem Maße. Insbesondere wurde auf die Ziele des Russischunterrichts eingegangen. Die Herausforderung am Seminar für Slavistik wie auch in anderen fremdsprachlichen Philologien besteht darin, den drei Studienzweigen bzw. Studiengängen, also Linguistik, Literatur- und Kulturwissenschaft, sowie der Lehramtsausbildung gerecht zu werden. Die Auswahl der Texte in Bezug auf die Thematik und Textsorten sollte diese Heterogenität widerspiegeln. Zudem wurden die Niveaus bestehender Kurse und die zu erreichenden Kompetenzniveaus aufeinander aufbauender Kurse thematisiert.

3.3.4 Auswahl der Texte & Erstellung der Lückentexte

Im nächsten Schritt sollte jede(r) Studierende zwei für einen C-Test geeignete Texte auswählen und diese in das C-Test-Format übertragen. Zur Unterstützung bekamen die Studierenden ein C-Test-Beispiel in der Form, in der sie ihre Texte einreichen sollten (Anhang 1).

Die Aufgabenstellung lautete wie folgt:

Suchen Sie zwei Texte unterschiedlicher Schwierigkeitsstufen (GeR² A1–C2) und entwickeln Sie daraus zwei C-Test-Texte mit je 25 Lücken nach den Konstruktionsregeln von Grotjahn 2002. Schicken Sie diese als Word-Datei mit Lösungen an die studentische Hilfskraft bis zum 21.05.2017.

Der Schwierigkeitsgrad der Texte (bzw. das angestrebte Niveau), die die Studierenden aussuchen sollten, wurde nach ihrer Sprachbeherrschung des Russischen bestimmt. Das heißt, FremdsprachenlernerInnen sollten Texte für die Niveaus A1–A2, HerkunftssprecherInnen für die Niveaus B1–B2 und MuttersprachlerInnen für die Niveaus C1–C2 aussuchen. Insgesamt entstanden 22 Texte, die die Sprachbeherrschung auf unterschiedlichen Niveaus überprüfen sollten.

3.3.5 Rückmeldung zu den erstellten Texten

In dieser Phase fand die Begutachtung der erstellten C-Test-Texte statt. Diese erfolgte einerseits durch die Seminarleitung, andererseits durch die LektorInnen am Seminar für Slavistik.

Rückmeldung durch die Seminarleitung

Bei der Begutachtung durch die Seminarleitung wurde auf verschiedene Kriterien geachtet. Die Texte sollten für das Russischlernen am Seminar für Slavistik thematisch angemessen sein. Zudem sollten alle formalen Konstruktionsregeln beachtet werden.

² Die unterschiedlichen Abkürzungen GeR (zumeist in Deutschland verwendet) und GERS (in Österreich üblich) wurden beibehalten.

Das thematische Kriterium in den eingereichten Texten war weitgehend erfüllt. Trotz der genauen Vorgaben ergaben sich jedoch systematische Abweichungen von den Konstruktionsregeln:

- Nicht alle erstellten C-Test-Texte hatten genau 25 Lücken;
- Das Tilgungsprinzip (halbe Anzahl der Buchstaben/plus eins) wurde nicht durchgehend berücksichtigt;
- Die Regel bezüglich des ersten und letzten Satzes (erster und letzter Satz bleiben unverändert) wurde nicht durchgängig eingehalten;
- Die Lösungen wurden nicht in einer Wortliste präsentiert;³
- Die Lückenstriche waren von unterschiedlicher Länge;
- Die Einreichung der C-Test-Texte erfolgte nicht immer als Word-Datei.

Auf der Grundlage dieser Verbesserungsbereiche wurden individuelle Rückmeldungen zu den einzelnen Tests aller SeminarteilnehmerInnen erstellt. Die Erfahrung der ersten Phase der Testerstellung wurde zudem im Seminar besprochen. Darüber hinaus bekamen die Studierenden eine allgemeine Rückmeldung in Form einer Checkliste (siehe Anhang 4) sowie eine individuelle Rückmeldung zu ihren C-Test-Texten. Dabei wurde die Wichtigkeit der Regeleinhaltung diskutiert. Die erstellten C-Test-Texte sollten die Studierenden bearbeiten und erneut einreichen.

Rückmeldung durch die LektorInnen am Seminar für Slavistik

Gleichzeitig wurden die C-Test-Texte durch drei LektorInnen am Seminar für Slavistik evaluiert. Die Beteiligung der LektorInnen war essentiell, da sie am besten mit dem Lernkontext vertraut sind und letztendlich die curricularen EntscheidungsträgerInnen darstellen. Bei der Evaluation sollten

³ Diese Übersicht ermöglicht es allerdings sofort zu sehen, ob einige Wörter mehrmals getilgt wurden wie im Beispiel in Anhang 3.

sie vor allem die Relevanz des Themas für das Curriculum, die in den Texten verwendete Sprache und die Schwierigkeit einzelner Lücken beurteilen. Außerdem wurde die Schwierigkeit einzelner Texte diskutiert.

3.3.6 Auswahl der Texte für die Pilotierung

Das Ziel des nächsten Schrittes war es, aus den 22 C-Test-Texten zehn für die Pilotierung auszusuchen. Dieser Schritt wurde ohne Beteiligung der Studierenden gemacht, sondern erfolgte durch die LektorInnen sowie durch die Erstautorin. Jede(r) LektorIn sollte die von den Studierenden bearbeiteten und eingereichten C-Test-Texte nochmals durchlesen und angeben, ob und in welchem Maße die einzelnen C-Test-Texte für den eventuellen Einsatz im Rahmen des Einstufungstests geeignet sind. Am Ende wurden neun Texte unterschiedlicher Thematik und unterschiedlicher Schwierigkeitsstufe ausgewählt (Anhang 4), kleinere Verbesserungen und Umformulierungen vorgenommen und einige Titel der C-Test-Texte modifiziert, sodass sie den Inhalten der Texte besser entsprachen. Die Texte wurden in die Reihenfolge ihrer eingeschätzten Schwierigkeit gebracht. Die Studierenden waren in dieser Phase zwar nicht beteiligt, erhielten allerdings einen Einblick in die Entscheidungen und Diskussionen bezüglich der Textauswahl.

3.3.7 Pilotierung

Die Datenerhebung fand in der 7. und 8. Woche des Semesters statt. Die Studierenden wurden zu zweit unterschiedlichen Russischkursen zugeteilt. Sie sollten die LektorInnen der jeweiligen Kurse kontaktieren und einen

Termin für die Pilotierung des Tests in den Sprachkursen vereinbaren. Aus Sicherheitsgründen erhielten die Studierenden nur so viele Exemplare von Tests, wie es Russischlernende im zugewiesenen Kurs gab, sowie eine Kopie für die Lehrkraft.

Zur Unterstützung bekamen die Studierenden eine Checkliste bzw. die Beschreibung des Ablaufs für die Pilotierung (siehe Anhang 5). Da die meisten Studierenden keine Erfahrung mit Datenerhebungen hatten, wurde der Ablauf der Pilotierung im Seminar simuliert.

Während der Pilotierung wurden die LektorInnen gebeten, die C-Tests zu lösen und zu kommentieren, v.a. in Bezug auf Tippfehler und die Schwierigkeit der Lücken, und ggf. Umformulierungen vorzuschlagen.

In der Pilotierung wurden insgesamt Daten von 57 RussischlernerInnen aus acht unterschiedlichen Kursen am Seminar für Slavistik gewonnen.

3.3.8 Testkorrektur, Dateneingabe und Datenauswertung

Die gemeinsame Korrektur der ausgefüllten C-Tests fand im Rahmen einer Seminarsitzung statt. Jede(r) Studierende sollte jeweils einen Text aus dem C-Test-Set, d.h. insgesamt 57 C-Test-Texte nach dem vorgegebenen Antwortschlüssel korrigieren.

Zwei Seminarteilnehmerinnen gaben die Daten in eine Excel-Tabelle nach dem vorgegebenen Kodierungsschema ein. Dabei handelte es sich um die Sprachlernbiographie der TestteilnehmerInnen, ihre Testergebnisse in jedem einzelnen C-Test-Text sowie die Einschätzungen der RussischlektorenInnen bezüglich ihrer Russischkenntnisse (siehe 3.3.6). Es wurden

sprachbiographische Informationen, wie z.B. die Dauer des Russischunterrichts, Geburtsort, Muttersprache(n) und Länge des Auslandsaufenthalts im russischsprachigen Ausland erhoben.

Nach der Datenauswertung sollten die Studierenden sich überlegen, welche Informationen sie aus den Daten gewinnen können und möchten. In der Sitzung wurden gemeinsam mit den Studierenden u.a. folgende Fragen formuliert:

- Wie schwierig sind die einzelnen C-Test-Texte?
- Wie reliabel sind einzelne C-Test-Texte?
- Korrelieren die Ergebnisse im Test mit dem Geburtsort/der Länge des Auslandsaufenthalts/der Länge des Russischunterrichts?
- Welche fünf Texte eignen sich am besten für die Endversion des C-Tests?

Die eingegebenen Daten wurden von der Seminarleiterin auf ihre Vollständigkeit geprüft, in der Statistiksoftware SPSS analysiert und für die Interpretation der Ergebnisse aufbereitet.

3.3.9 Interpretation der Ergebnisse

In der vorletzten Phase des Projekts sollten die Studierenden Antworten auf die u.a. von ihnen selbst formulierten Fragen aus den Daten heraus eigenständig gewinnen und diskutieren. Dabei wurden sie in die wichtigsten statistischen Konzepte eingeführt und lösten konkrete Aufgaben, die zu bestimmten Schlussfolgerungen führen sollten.

Für die Zusammensetzung eines C-Tests ist es u.a. wichtig, dass Texte unterschiedlicher Schwierigkeitsstufen inkludiert sind. Dafür erhielten die Studierenden Tabelle 1 und sollten die pilotierten Texte in der Reihenfolge ihrer Schwierigkeit (1 – der einfachste, 9 – der schwierigste) ordnen. Durch den Datensatz konnten Texte mit ähnlichem Schwierigkeitsgrad ermittelt und aus der Endversion des Tests ausgeschlossen werden. Dazu wurde neben der Thematik (Relevanz des Inhalts für das Curriculum) auch die Reliabilität einzelner C-Test-Texte mit den Studierenden analysiert.

Für ihre Testentwicklungskompetenz generell war allerdings die Erkenntnis wichtiger, dass die tatsächliche Schwierigkeit eines C-Test-Textes (oder eines Items generell) nur zum Teil vorhergesagt werden kann. Für die Pilotierung wurden die C-Test-Texte von den Studierenden und LektorInnen nach ihrer eingeschätzten Schwierigkeit platziert. Die Ergebnisse aus der Pilotierung zeigten allerdings, dass z.B. Text 5 und nicht Text 2 der zweitleichteste Text war oder dass Text 4 für die RussischlernerInnen viel schwieriger war als zu Beginn vermutet. Dies bedeutet, dass die tatsächliche Schwierigkeit von Test(item)s nur empirisch ermittelt werden kann. Daher ist die Pilotierung von Aufgaben, auch wenn nur in geringem Maße, notwendig.

	<i>N</i>	<i>Min</i>	<i>Max</i>	<i>M</i>	<i>SD</i>	Cronbachs Korrigierte Alpha, Item- Skala- Korrelation lassen	wenn Item wegge- lassen
Text 1	57	5	24	15,40	5,98	,905	,985
Text 2	57	0	25	14,42	7,23	,922	,984
Text 3	57	0	25	10,16	7,92	,951	,983
Text 4	57	0	25	09,25	6,98	,948	,983

Text 5	57	1	24	15,07	7,10	,892	,985
Text 6	57	0	25	09,61	7,87	,942	,983
Text 7	57	1	25	11,51	7,21	,964	,982
Text 8	57	1	25	09,54	7,81	,946	,983
Text 9	57	0	25	07,75	7,64	,933	,983
Gesamtergebnis	57	11	221	102,51	62,60		

Tab. 1 Schwierigkeit einzelner C-Test-Texte und ihre Reliabilität

Neben der Schwierigkeit und Reliabilität einzelner Texte konnten aus den erhobenen Daten Hinweise zur Validität des C-Test-Textes gewonnen werden. Je länger man eine Fremdsprache lernt, je mehr Zeit man im Ausland verbringt oder die Sprache in der Familie spricht (HerkunftssprecherInnen), desto besser fallen die Ergebnisse in einem C-Test in der Regel aus. Wenn dies nicht der Fall ist, kann es u.U. sein, dass der Test keine validen Aussagen über die allgemeine Sprachkompetenz treffen kann. Vor diesem Hintergrund sollten die SeminarteilnehmerInnen der Frage nachgehen, ob die Ergebnisse des entwickelten C-Tests auch die Lernbiographie der RussischlernerInnen widerspiegelt. Dazu analysierten die Studierenden die Boxplots und Liniendiagramme verschiedener Variablen, u.a. den Lernstatus (Abb.1), die Dauer des Auslandsaufenthalts (Abb.2) und den Russischkurs der TestteilnehmerInnen (Abb. 3).

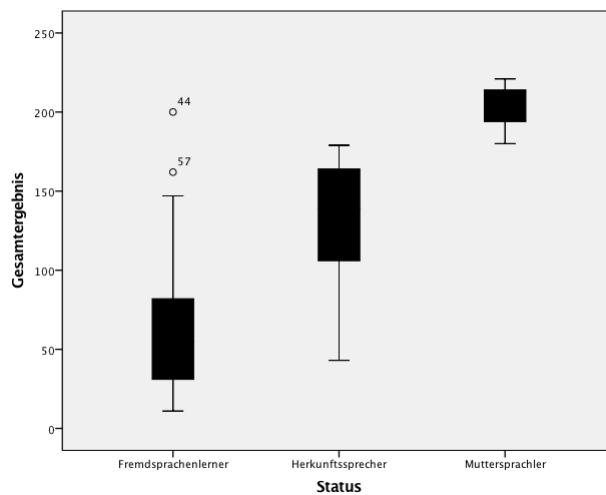


Abb. 1 Erreichte Punkte im C-Test nach Lernstatus

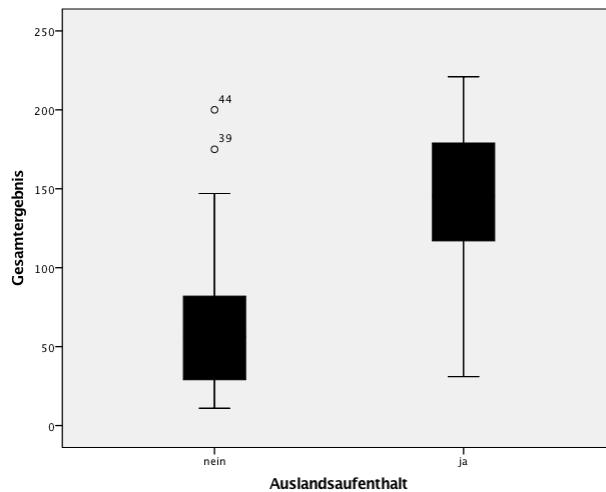


Abb. 2 Erreichte Punkte im C-Test nach Auslandsaufenthalt im russischsprachigen Ausland

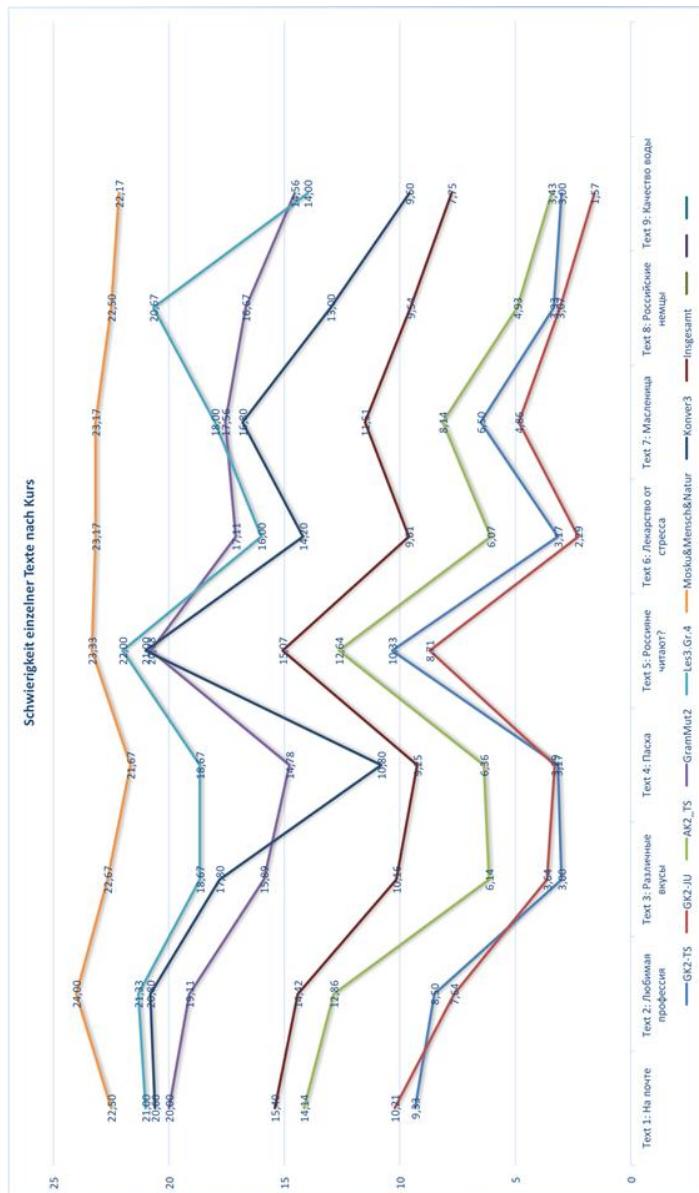


Abb. 3 Erreichte Punkte im C-Test nach Kurs

Anastasia Drackert und Katrin Bente Karl (Hg.), Didaktik der slawischen Sprachen

Beiträge zum 2. Arbeitskreis in Innsbruck (19.02.–20.02.2018)

© 2019 innsbruck university press, ISBN 978-3-903187-80-1, DOI 10.15203/3187-80-1

Anhand der unterschiedlichen Analysen sollte jede(r) SeminarteilnehmerIn eine eigene Version des Einstufungstests vorschlagen und mit den besprochenen Erkenntnissen aus der Datenerhebung begründen. Die Entscheidung bez. der endgültigen Zusammensetzung des C-Tests wurde nach dem Abschluss des Seminars getroffen.

3.3.10 Reflexion

Einen bedeutenden Teil des Projekts bildet die abschließende Reflexion des Arbeitsprozesses und des erstellten Tests durch die SeminarteilnehmerInnen. Zum Ende des Projekts bzw. des Seminars wurden die Studierenden gebeten, über das durchgeführte Forschungsprojekt zu reflektieren. Die Studierenden bekamen folgende Fragen und sollten ihre Gedanken zu den einzelnen Punkten aufschreiben:

- Was haben Sie durch das gemeinsame Seminarprojekt lernen können?
- Was waren die „AHA“-Momente im Forschungsprozess?
- Was haben Sie sich bei einem Testentwicklungsprozess anders vorgestellt? // Wie haben sich Ihre Vorstellungen vom Testentwicklungsprozess verändert?
- Hat das Projekt Ihr Interesse am Testen/Prüfen geweckt?
- Hat das Projekt Ihr Interesse am Forschen geweckt?
- Wenn ich in der Zukunft einen Test selbst entwickeln sollte, so würde ich es machen...
- Forschendes Lernen im Seminar hat mir geholfen ...
- Für meine zukünftige Arbeit nehme ich Folgendes mit ...

Dies geschah anonym, die Studierenden konnten sich dabei entscheiden, wie viele Fragen sie beantworten. Insgesamt sind zehn schriftliche Rückmeldungen erfolgt. Diese wurden abgetippt, analysiert und zu vier wesentlichen Punkten zusammengefasst.

Testentwicklungskompetenz und Interesse am Testen

Die Rückmeldungen der Studierenden machen deutlich, dass ihr Interesse am Prüfen und Testen durch das Seminarprojekt geweckt wurde. Außerdem hat das Projekt ihnen gezeigt, wie viele Schritte tatsächlich bei einer Testerstellung und -durchführung zu beachten sind, wie folgende Zitate deutlich machen:

Durch das gemeinsame Seminarprojekt habe ich gelernt, was ich beachten muss bei der Erstellung eines Tests (Niveau, Lernbiographie etc. des Lernenden), und wie es ist, einen Test in einer Gruppe selbstständig durchzuführen. Bei der Erstellung gab es deutlich mehr Dinge zu beachten als ich ursprünglich dachte, und das hat das Interesse in mir geweckt, einen weiteren Test zu erstellen mit dem Hintergrundwissen, welches ich jetzt habe⁴

Ich habe gelernt, was man bei der Erstellung eines Tests im Vorfeld beachten muss (angemessenes Thema, Niveaustufen, Probandensuche, Kontakt mit Lehrenden).

Ein Test kann nicht beliebig ausgesucht werden. Testen benötigt Zeit, Geduld und Systematik.

Ich habe durch das Seminarprojekt viel über das Entwickeln und Pilotieren von Tests lernen können. Mich hat dabei sehr interessiert, welche Schritte ich beim Erstellen eines Tests befolgen und auch welche Gütekriterien ich einhalten muss. Besonders die Wichtigkeit der Pilotierung von Tests hat mich überrascht! [...]

Da man als Studierender immer eine andere Perspektive zum Thema Testen hat, konnte ich mich hier in die Rolle eines Prüfers versetzen und schauen, wie das

⁴ Alle Zitate sind wortwörtlich wiedergegeben.

Prüfen eigentlich funktioniert. Vor allem ist mir auch klargeworden, dass man anhand von C-Tests sehr viel über die Kenntnisse eines Teilnehmers sagen kann. Bei der Bearbeitung soll man auf viele Sachen achten: die Länge eines Textes, Schwierigkeitsgrad, Wörter, die sich wiederholen, dürfen nicht zweimal getilgt werden usw. Was die Thematik angeht, sollen die Texte auch unterschiedlich und nicht spezifisch sein.

Die Aussagen zeigen auf, dass die Studierenden tatsächlich in ihrer Testentwicklungskompetenz geschult wurden. Diese Zitate veranschaulichen ebenfalls, dass die Lerninhalte verinnerlicht wurden (vgl. *deep level learning*). Erreicht wurde dies nicht durch das Lesen der Fachliteratur oder das Auswendiglernen wichtiger Punkte, sondern dadurch, dass die Studierenden selbstständig den Forschungsprozess durchlaufen konnten und somit verstanden haben, warum bestimmte Aspekte der Testerstellung eine Rolle spielen, wie z.B. die Pilotierung oder die Gegebenheiten einer Lerngruppe.

Interesse an Forschung

Durch das Forschende Lernen im Projekt konnte ebenso aufgezeigt werden, was bei Forschungsprozessen allgemein bedacht werden muss. Dass die Studierenden der Forschung positiv gegenüberstehen, zeigen die folgenden Zitate:

Das Projekt hat mir aufgezeigt, wie man an solche Forschungsprojekte herangehen soll, was für Stufen durchlaufen werden sollten, um von einer Forschungs-idee zu einem Forschungsergebnis zu kommen.

Mir ist aufgefallen, dass aus Forschungsergebnissen auch andere interessante Schlüsse gezogen werden können bzw. auch andere Fragen beantwortet werden können, als vorher geplant war. Das Projekt hat mir aufgezeigt, dass es zeitaufwändig ist, Forschung zu betreiben und deshalb würde ich mir in Zukunft genug Zeit einplanen für weitere Forschungsprojekte.

Was ich persönlich lernen konnte, war, wie umfangreich und vor allem aufwendig eine solche Empirie ist.

Forschendes Lernen hat mir geholfen einen praktischen Einblick in die Arbeit mit Tests zu erhalten. Es war viel aufschlussreicher, an der Erarbeitung eines Tests beteiligt zu sein als nur darüber zu reden, wie man theoretisch einen Test entwickeln sollte.

Zugegebenermaßen bin ich eher daran interessiert, die Ergebnisse der Forschungen für mich zu nutzen als selbst zu forschen, aber durch das Projekt konnte ich viel darüber lernen, auf welche Weise ich Tests entwickeln kann, um später aussagekräftige Ergebnisse zu erhalten! Insofern hat sich mein Interesse am Forschen definitiv erhöht.

Auch hier wird deutlich, dass die aktive Mitarbeit am Forschungsprojekt dazu beigetragen hat, dass die Studierenden die Inhalte nachhaltiger lernen als durch die bloße Auseinandersetzung mit der Theorie. Das praktische Arbeiten im Seminar wurde als gewinnbringend empfunden und das Interesse für zukünftige Forschungsvorhaben wurde geweckt bzw. bekräftigt.

Mehrwert des Projekts für die Tätigkeit als LehrerIn

Die Reflexion verdeutlicht, dass die Studierenden aus dem Forschungsprojekt einen Nutzen für die zukünftige Ausübung der Lehrtätigkeit bezüglich des Testens ziehen konnten. Dies zeigen folgende Zitate:

Für meine zukünftige Arbeit nehme ich Folgendes mit:

- Wie erstelle ich einen Test?
- Worauf muss ich besonders achten?
- Wie führe ich einen Test durch?
- Wie fasse ich die Ergebnisse zusammen?
- Welche Forschungsfragen beantwortet dieser?

Для меня этот проект был очень интересен и нов. Этот проект заставил задуматься, сколько сил прикладывает преподаватель, составляя и проводя тесты или контрольные работы: начиная с отбора правильных текстов и заканчивая их анализом. Работа над проектом «заставила» меня также просмотреть литературу, которая раньше не была интересна, например, как точно изучается русский как иностранный, какие критерии характерны для каждой ступени изучения.

Ich werde auf jeden Fall versuchen, die Tests gemäß den Gütekriterien zu erstellen. Auch möchte ich versuchen, die Tests zu pilotieren und zu testen, bevor ich sie für meine Arbeiten benutze.

Forschendes Lernen im Seminar hat mir geholfen, die Sprachkompetenz besser einschätzen zu können und die Vor- und Nachteile von unterschiedlichen Testformaten erkennen zu können. Ich denke, dass mir das Wissen über die Testformate (sowohl allgemeine Sprachkompetenz als auch Teilkompetenz) im Lehrerberuf sehr nützlich sein werden.

Den Studierenden war es möglich, Parallelen zwischen den Seminarinhalten und dem zukünftigen Lehrerberuf zu ziehen. Dies demonstriert, dass ein Seminar dieser Art die Studierenden zumindest ein Stück weit auf ihre Lehrtätigkeit vorbereitet.

Förderung von Schlüsselkompetenzen

Des Weiteren zeigen die Reflexionen der SeminarteilnehmerInnen, dass die sogenannten Schlüsselkompetenzen durch das Forschungsprojekt gefördert wurden. Es wird die „Arbeit im Team“ genannt sowie „sich selbst und das Vorhaben einer Gruppe Studierender zu präsentieren“, die „Auseinandersetzung mit den theoretischen Grundlagen“ wie auch das Kennenlernen „einiger statistischer Grundlagen“. Wie oben bereits erwähnt (siehe 2.2), begünstigte Forschendes Lernen die Aneignung für den Erkenntnisgewinn wichtiger Schlüsselkompetenzen.

4. Diskussion und Ausblick

Die abschließenden Äußerungen der Studierenden aus der Reflexionsphase des Seminarprojekts (siehe 3.10) deuten darauf hin, dass die Studierenden durch das Forschende Lernen einen Lernertrag in mindestens vier Bereichen wahrnahmen: 1) Testentwicklungskompetenz und Interesse am

Testen, 2) Interesse an Forschung, 3) Mehrwert des Projekts für die zukünftige Lehrtätigkeit und 4) Förderung von Schlüsselkompetenzen.

Forschendes Lernen als geeignetes Format: Möglichkeiten

Zu den wichtigsten Erkenntnissen der Studierenden zählt die Notwendigkeit der Pilotierung von Tests vor ihrem Einsatz im schulischen, akademischen oder wissenschaftlichen Kontext. Zum Beispiel konnten die Studierenden bereits bei der Testauswertung selbst feststellen, dass einige C-Test-Texte einen anderen Schwierigkeitsgrad hatten als zuvor vermutet, was sicherlich nicht nur an dem Text selbst sondern auch an den Lücken lag.

Ebenso konnten sie bei der Analyse der Ergebnisse feststellen, dass zu einigen Lücken weitere Antwortmöglichkeiten denkbar sind und in die Antwortschlüssel aufgenommen werden mussten. Durch das Forschende Lernen haben sich die Studierenden dieses Wissen selbst erarbeitet und nicht als Tatsachen vermittelt bekommen. Diese aktive Teilhabe an der Wissensgewinnung, u.a. durch die Reflexion einzelner Schritte, führte zu nachhaltigem Lernen (Huber 2009), was in den Äußerungen der Studierenden zum Ausdruck gebracht wird. Sicherlich wäre es interessant, die Auswirkung des Forschenden Lernens auf die Aneignung der Testentwicklungscompetenz longitudinal zu untersuchen. Im Sinne des nachhaltigen Lernens könnte untersucht werden, welche erworbenen Kompetenzen die SeminarteilnehmerInnen in der Schule oder in der Forschung tatsächlich einsetzen.

Die im Laufe des Projekts gewonnenen Erkenntnisse haben die Studierenden dazu verleitet, weitere Fragen zu stellen: Wie sieht es mit anderen Testformaten aus? Was macht einen Text schwierig? Was mache ich als Lehrkraft oder WissenschaftlerIn, wenn die SchülerInnen bzw. ProbandInnen in Klassenarbeiten bzw. linguistischen Experimenten abweichende Antworten geben, die allerdings auch richtig sind? Diese Fragen wurden

dann in Bezug auf weitere für die einzelnen TeilnehmerInnen relevante Kontexte diskutiert, sodass ein Transfer stattgefunden hat.

Einige Kompetenzen, Erkenntnisse und Erfahrungen, die z.B. durch die Arbeit mit Tabellen bzw. Graphiken erworben wurden, sind zudem für den Erwerb weiterer allgemeiner Kenntnisse im Bereich Statistik relevant. Die statistischen Daten und ihre Analyse, sei es in Form von Tabellen oder Graphiken, waren für die Studierenden bereits deshalb zugänglicher, weil sie verstanden, wie die Zahlen zu Stande gekommen sind und was sie im Kontext dieser Studie bedeuten. Dieses Wissen über den Umgang mit solchen Daten können die Studierenden nun auch auf weitere Kontexte übertragen. Zudem behaupteten die Studierenden, dass sie die aus dem Projekt erworbenen Kompetenzen und ihr Fachwissen auf das Verständnis und die Durchführung empirischer Untersuchungen allgemein übertragen können.

Besonders erfreulich ist, dass die Studierenden ihren eigenen Angaben nach sich für empirische Forschung begeistern ließen. Trotz der Erkenntnis, dass diese aufwendig ist, haben sich mehrere Studierende dazu entschieden, kleinere empirische Untersuchungen im Rahmen ihrer Hausarbeiten durchzuführen.

Zu betonen ist, dass es keine negativen Rückmeldungen bezüglich des Projekts seitens der Studierenden gab, was für eine positive Erfahrung der Studierenden mit Forschendem Lernen spricht. Zu bedenken ist jedoch die Möglichkeit von gewissen Antworttendenzen im Sinne der 'sozialen Erwünschtheit', die bewusst positive Antworten auslöst, um gewissen Erwartungen oder Vorstellungen der Dozentin zu entsprechen.

Grenzen des Forschenden Lernens im Seminar

Beim Forschenden Lernen im engeren Sinne sollen die Studierenden die EntscheidungsträgerInnen hinsichtlich der Organisation und Umsetzung von Forschungsvorhaben sein, die Lehrkraft hingegen soll sie unterstützen

und Feedback geben (Mieg 2017). Im beschriebenen Seminarprojekt war jedoch die Kursleiterin für viele Entscheidungen zuständig. Dies ist der Tatsache geschuldet, dass Testentwicklung und Validierung einen sehr aufwendigen Prozess darstellen. Dieser kann nur dann von Studierenden allein bewältigt werden, wenn sie bereits Erfahrungen in der Testerstellung oder Validierung, Ideen für konkrete Projekte und den Zugang zum Feld bzw. Lernsetting haben, in dem die Pilotierung durchgeführt werden kann. Zudem ist solch eine selbstständige Arbeit nur innerhalb eines Seminars möglich, das ausschließlich dem Projekt gewidmet ist (Schön, Zimmermann & Johnson 2012). Da in diesem Seminar neben der Förderung der Testentwicklungskompetenz weitere Lernziele verfolgt wurden, war uneingeschränktes selbstständiges Arbeiten nur teilweise möglich (z.B. in der Pilotierungsphase).

Eine Herausforderung im Projekt für die Seminarleiterin bestand in der großen Heterogenität der SeminarteilnehmerInnen. Da die Gruppe aus Studierenden zweier unterschiedlicher Studiengänge – Slawische Philologie und Russisch im Rahmen des Master of Education – bestand, hatten die Studierenden sicherlich unterschiedliche Erwartungen an das Projekt. Die Auseinandersetzung mit dem C-Test-Format war für die LinguistInnen wahrscheinlich von höherer Relevanz als für die M.Ed.-Studierenden. Diese hätten idealerweise Testaufgaben entwickelt, die sie z.B. im Rahmen von Klassenarbeiten einsetzen könnten. Hier spielte die Praktikabilität in Bezug auf die Pilotierung eine zentrale Rolle, da eine enge Zusammenarbeit mit Schulen nötig gewesen wäre. Zudem wären u.a. die Auseinandersetzung mit dem Lehrplan für das Fach Russisch oder die Berücksichtigung des schulischen Zeitplans in Bezug auf die Terminierung der Klassenarbeiten erforderlich, was aus zeitlichen Gründen sowie aufgrund der linguistischen Ausrichtung des Seminars nicht im Rahmen von 13 SWS hätte geleistet werden können. Nichtsdestotrotz, auch wenn der Schwerpunkt im Seminar nicht auf den *classroom tests* lag, konnten die zukünftigen LehrerInnen einen erfolgreichen Transfer auf diese leisten.

Optimierung der Arbeit

Forschendes Lernen ist nicht nur für Studierende, sondern auch für HochschullehrerInnen ein aufwendiges Unterfangen. Mindestens folgende zwei Aspekte bedürfen einer Optimierung für zukünftige Projekte dieser Art. Zum einen könnte den Studierenden eine Checkliste zur Testerstellung vorgegeben werden, anhand derer sie nicht nur die Tests erstellen, sondern vor der Einreichung auch selbst korrigieren könnten (siehe Anhang 3). Zum anderen könnte die individuelle Rückmeldung (Schritt 4a) durch Peer-Korrektur ersetzt werden. Dadurch wird den Studierenden mehr Verantwortung übertragen, was autonomes Lernen stärker fördert (Black & Harrison 2001). Je nach Gruppengröße könnte man die Studierenden von Seminarbeginn an in allen Phasen zu zweit arbeiten lassen. Die Umsetzung dieser Verbesserungen könnte den Zeitaufwand für die Seminarleitung verringern und gleichzeitig den Lernprozess der Studierenden fördern.

Das Integrieren eines Forschungsprojekts als Teil eines Seminars ist ein Kompromiss zwischen dem Bestreben, den Studierenden ein breites Spektrum an Themen anzubieten und gleichzeitig vertieft am Beispiel eines Themas Kompetenzen aktiv auszubilden. Das vorgestellte Modell kann im Rahmen anderer Philologien implementiert und für weitere Themenbereiche ausgeweitet werden.

Literatur

- ALTE (2012). *Handbuch zur Entwicklung und Durchführung von Sprachtests. Zur Verwendung mit dem GER*. Frankfurt a. M.: telc GmbH.
- Baur, R. & Spettmann, M. (2007). Screening – Diagnose – Förderung: Der C-Test im Bereich DaZ. In B. Ahrenholz (Hrsg.), *Deutsch als Zweitsprache*.

- sprache: Voraussetzungen und Konzepte für die Förderung von Kindern und Jugendlichen mit Migrationshintergrund.* (95–110). Freiburg i.Br: Fillibach.
- Black, P. & Harrison, C. (2001). Self- and peer-assessment taking responsibility: the science students role in formative assessment. *The School Science Review*, 83 (302), 43–49.
- Douglas, D. (2010). *Understanding Language Testing*. London: Hodder Education.
- Drackert, A. (2016). *Validating Language Proficiency Assessments in Second Language Acquisition Research. Applying an Argument-Based Approach*. Frankfurt a. M.: Peter Lang GmbH.
- Drackert, A. & Stadler, W. (2017). Leistungsbeurteilungskompetenz von Russischlehrkräften in Deutschland, Österreich, der Schweiz und Südtirol (DACHS): Zwischen Status Quo und aktuellen Bedürfnissen. *Zeitschrift für Fremdsprachenforschung*, 28 (2), 233–258.
- Drackert, A. & Timukova, A. (2019). What does the analysis of C-test gaps tell us about the construct of a C-test? A comparison of foreign and heritage language learners' performance. *Language Testing*. <https://doi.org/10.1177/0265532219861042>.
- Eberharter, K., Kremmel, B. & Zehentner, M. (2018). Die Erstellung von Aufgaben: der Testentwicklungszyklus. In B. Hinger & W. Stadler (Hrsg.) *Testen und Bewerten fremdsprachlicher Kompetenzen. Eine Einführung* (57–68). Tübingen: Narr Studienbücher.
- Eckes, T. & Grotjahn, R. (2006). A closer look at the construct validity of C-tests. *Language Testing*, 23, 290–325.
- Fichten, W. (2017). Forschendes Lernen in der Lehramtsausbildung. In H. A. Mieg & J. Lehmann (Hrsg.), *Forschendes Lernen: Wie die Universität und die Fachhochschule erneuert werden kann.* (155–164). Campus Verlag.
- Geeb, F., Krauß-Leichert, U. & Verch, U. (2009). Forschendes Lernen als Kooperationsprojekt an Fachhochschulen im Bereich Information

- und Technik. In L. Huber, J. Hellmer & F. Schneider (Hrsg.), *Forschendes Lernen im Studium. Aktuelle Konzepte und Erfahrungen* (157–168). Bielefeld: Univeristätsverlag Webler.
- Grotjahn, R. (2002). Konstruktion und Einsatz von C-Tests: Ein Leitfaden für die Praxis. In R. Grotjahn (Hrsg.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen*. Band 4 (211–225). Bochum: AKS-Verlag.
- Hofer, R. (2013). Forschendes Lernen in der Lehrerinnen- und Lehrerbildung: Widersprüchliche Anforderungen zwischen Forschung und Profession. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 31 (3), 310–320.
- Huber, L. (2014). Forschungsbasiertes, Forschungsorientiertes, Forschendes Lernen: Alles dasselbe? Ein Plädoyer für eine Verständigung über Begriffe und Unterscheidungen im Feld forschungsnahen Lehrens und Lernens. *Hochschulforschung*, 1+2, 22–29.
- Huber, L. (2009). Warum Forschendes Lernen nötig und möglich ist. In L. Huber, J. Hellmer & F. Schneider (Hrsg.), *Forschendes Lernen im Studium. Aktuelle Konzepte und Erfahrungen* (9–35). Bielefeld: Universitätsverlag Webler.
- Klein-Braley, C. (1997). C-Tests in the context of reduced redundancy testing: An appraisal. *Language Testing*, 14, 47–84.
- Malone, M. (2013). The essentials of assessment literacy: Contrasts between testers and users. *Language Testing*, 30, 329–344.
- Mieg, H. A. (2017). Einleitung: Forschendes Lernen – erste Bilanz. In H.A. Mieg & J. Lehmann (Hrsg.), *Forschendes Lernen: Wie die Universität und die Fachhochschule erneuert werden kann* (15–31). Campus Verlag.
- Mozgalina A. & Ryshina-Pankova M. (2015). Meeting the Challenges of Curriculum Construction and Change: Revision and Validity Evaluation of a Placement Test. *The Modern Language Journal*, 99 (2), 1–22.
- Norris, J. (2018). *Developing C-tests for estimating proficiency in foreign language research*. Frankfurt a. M.: Peter Lang GmbH.

- Pasternack, P. (2017). Konzepte und Fallstudien: Was die Hochschulfor- schung zum Forschenden Lernen weiß. In H. A. Mieg & J. Lehmann (Hrsg.), *Forschendes Lernen: Wie die Universität und die Fachhochschule erneuert werden kann* (37–44). Campus Verlag.
- Reinmann, G. (2015). Heterogenität und forschendes Lernen: Hochschul- didaktische Möglichkeiten und Grenzen. In B. Klages, M. Bonillo, S. Reinders & A. Bohmeyer (Hrsg.), *Gestaltungsraum Hochschul- lehre. Potenziale nicht-traditionell Studierender nutzen* (121–137). Opladen; Berlin; Toronto: Budrich UniPress Ltd.
- Schön, A., Zimmermann, K. & Johnson, N. (2012). Intrauniversitäre Ko- operation – zur gemeinsamen Entwicklung eines C-Tests durch Sprachenzentrum und Sprachlehrforschung. *Fremdsprachen und Hochschule*, 86, 61–79.
- Steurer, V. (1996). Ein neues Testverfahren beim „Bundeswettbewerb Fremdsprachen“: der C-Test. – Auch im schulischen Russischunter- richt einsetzbar? *Zielsprache Russisch*, 7 (3), 83–90.
- Wiemer, M. (2017). Forschend lernen – Selbstlernen. Selbstlernprozesse und Selbstlernfähigkeiten im Forschenden Lernen. In H. A. Mieg & J. Lehmann (Hrsg.), *Forschendes Lernen: Wie die Universität und die Fachhochschule erneuert werden kann* (47–55). Campus Verlag.
- Wildt, J. (2005). Auf dem Weg zu einer Didaktik der Lehrerbildung? *Be- träge zur Lehrerinnen- und Lehrerbildung*, 23 (2), 183–190.
- Wissenschaftsrat (2000). *Empfehlungen zur Einführung neuer Studien- strukturen und -abschlüsse (Baccalaureus/Bachelor – Magis- ter/Master) in Deutschland*. Berlin.
- Wulf, C. (2017). »From Teaching to Learning« – Merkmale und Heraus- forderungen einer studierendenzentrierten Lernkultur. In H. A. Mieg & J. Lehmann (Hrsg.), *Forschendes Lernen: Wie die Universität und die Fachhochschule erneuert werden kann* (66–75). Campus Verlag.
- Zimmermann, K. (2019). *Keine Zeit für den C-Test? Eine empirische Un- tersuchung zum Einfluss einer Geschwindigkeitskomponente auf das Konstrukt des C-Tests*. Berlin: Universitätsverlag der TU Berlin.

Anhang 1

Beispiel eines C-Test-Textes

Свадебные традиции

Большинство свадебных традиций, таких как разбрасывание риса или преломление свадебного пирога, уходят корнями в глубокое прошлое. За каж_____ обрядом ст_____ своя ист_____. Например, обы_____ надевать нев_____ на сва_____ фату, - од_____ из древн_____. Римские нев_____ носили е_____ ещё 2000 л_____ назад. Фа_____ носилась, пре_____ всего, к_____ символ скром_____ и тайны, и тол_____ муж м_____ снять е_____ после бра_____ церемонии. В_____ неко_____ восточных стр_____ фату поме_____ между мужч_____ и женщ_____ на в_____ свадебную цере_____. Делалось это для того, чтобы удостовериться, что они не увидят и не прикоснутся друг к другу до совершения брака.

Lösungen

Каждым, стоит, история, обычай, невесте, свадьбу, один, древнейших, невесты, её, лет, фата, прежде, как, скромности, только, мог, её, брачной, некоторых, помещали, мужчиной, женщиной, всю, церемонию.

Anhang 2

Zeitplan des Projekts integriert in den Seminarplan

Sitzung	DATUM	Thema	Schritte/Aufgaben Seminarprojekt
1	28.04	Testen, Prüfen, Evaluieren: Grundlagen 1	SCHRITT 1
2	05.05	Testen, Prüfen, Evaluieren: Grundlagen 2	
3	12.05	Testen allgemeiner Sprachkompetenz: C-Test	SCHRITT 2-4 2 C-Test-Texte erstellen (Abgabefrist 21.05)
4	19.05	Testen allgemeiner Sprachkompetenz: <i>Elicited Imitation Test</i>	SCHRITT 5 C-Test Texte verbessern, mit den Lösungsschlüsseeln an die studentische Hilfskraft schicken (Abgabefrist 10.06)
5	26.05	GER/Kompetenzorientiertes Testen	
6	02.06	Beispiele standardisierter Tests im Russischen: TRKI/Oral Proficiency Interview (CAL)/TELC	SCHRITT 6
7	16.06	Testen soziopragmatischer Kompetenz	SCHRITT 7
8	23.06	Vokabelkenntnisse testen	SCHRITT 7
9	30.06	Testen mehrsprachiger Kompetenzen (Fokus: HerkunftssprecherInnen/Testen vom argumentativen Schreiben)	SCHRITT 7
10	07.07	Seminarprojekt	SCHRITT 8

11	14.07	Bewertungskompetenz der RussischlehrerInnen/Selbstevaluation im Russischunterricht/Testen von Sprachmittlung (Thema nach Wahl der SeminarteilnehmerInnen)	
12	21.07	Seminarprojekt	SCHRITT 9 & 10
13	28.07	Zusammenfassung und Ausblick	

Anhang 3

Checkliste zur C-Testerstellung und Einreichung

Regeln	Check
Die endgültige Datei ist im Word-Format.	
Die Quelle des Textes ist angegeben.	
Der Text hat einen Titel.	
Das GeR-Niveau ist angegeben.	
Die Wörter des ersten sowie des letzten Satzes bleiben ungetilgt.	
Es wird jedes zweite Wort getilgt mit der Ausnahme von Eigennamen.	
Die Länge des Striches ist gleich groß, um nicht auf die Länge des Lösungswortes hinzuweisen.	
Die Wörter wurden gemäß dem Tilgungsprinzip (die halbe Buchstabenanzahl/plus eins) getilgt.	
Der C-Test enthält genau 25 Lücken.	
Die Lösungen sind in einer Wortliste präsentiert. Falls einige Lücken sich mehrmals wiederholen (siehe unten), soll der Text modifiziert werden. день, <u>рождения</u> , <u>жизни</u> , подростка, счастливый, <u>этот</u> , <u>жизнь</u> , <u>дома</u> , <u>как</u> , <u>немножко</u> , <u>не</u> , <u>уроками</u> , <u>школе</u> , <u>этот</u> , <u>не</u> , <u>можно</u> , <u>бояться</u> , <u>двойку</u> , подростком, трудно, взрослые, не	

Anhang 4

Auswahl der Texte für die Pilotierung

Text	Titel	LektorIn 1	LektorIn 2	Anastasia Drackert
1	Дружба. Мухтар и Мурка.			
2	Сегодня самый лучший день! Становление подростком.			
3	8 марта.	x		x
4	А вы откуда? Разговор на Красной Площади.	x		
5	Чувство близости. Застолье в России.			
6	На почте	x	x	x
7	Футбол – только для мальчиков?			(x)
8	Президентские состязания (Всеросийский спортивный фестиваль школьников)			(x)
9	Любимая профессия (планы на будущее)	x	x	x
10	Традиции в России (домовой)		x	
11	Пасха	x		(x)
12	Моя любимая еда		x	(x)
13	Россияне читают?	x		(x)
14	Масленица 1			
15	Лекарство от стресса	x	x	x
16	Масленица 2	x	x	(x)
17	Российские немцы - есть такая национальность?	x		x

18	Щи да каша – пища наша!	x	x
19	Где как здороваются	x	(x)
20	Розы-хамелеоны		x
21	О доступности безопасной воды	x	x
22	Медведи		

Anhang 5

Ablauf der Pilotierung

Vor der Datenerhebung:

- Sich vorstellen
- Erklären, was Sie vorhaben: Das Ziel der Datenerhebung – Pilotierung des Einstufungstests für das Seminar für Slavistik
- Einverständniserklärung lesen, besprechen und unterschreiben lassen; hervorheben, dass die Daten anonymisiert werden, die Namen der TestteilnehmerInnen allerdings für die Bestimmung der Kompetenz durch die LektorInnen gebraucht werden; die Teilnahme an der Studie kann jeder Zeit abgebrochen werden
- Anbieten, dass die Studierenden nach Wunsch ihre Punktzahl erfahren können
- Erklären, was ein C-Test ist, Ausfüllen des Fragebogens zur Person

Während der Datenerhebung:

- Für jeden Text haben die Studierenden vier Minuten Zeit⁵, plus eine Minute, um die Schwierigkeit des Textes zu bestimmen (1 einfach, 10 schwer)
- Ein Exemplar des C-Tests wird den LektorInnen gegeben.

⁵ Da die neuesten Untersuchungen (Zimmermann 2019) zeigen, dass auch weniger Lösungszeit als die üblichen fünf Minuten eine reliable Ermittlung der allgemeinen Sprachkompetenz ermöglicht, wurde entschieden, die Zeitbearbeitung jedes C-Test-Textes auf vier Minuten zu begrenzen.

Nach der Datenerhebung:

- Alle Unterlagen pro Person zusammenheften (unbedingt überprüfen, dass der Kurs angegeben wurde)
- Sich bei den Studierenden bedanken
- Alle (auch nicht ausgefüllte) Exemplare der Seminarleiterin zurückgeben

