

Maier, Uwe

Verlaufsdagnostik und Förderung im Lernbereich Sprachbetrachtung in der Sekundarstufe I mit Moodle

Empirische Sonderpädagogik 9 (2017) 2, S. 98-115



Empfohlene Zitierung/ Suggested Citation:

Maier, Uwe: Verlaufsdagnostik und Förderung im Lernbereich Sprachbetrachtung in der Sekundarstufe I mit Moodle - In: Empirische Sonderpädagogik 9 (2017) 2, S. 98-115 - URN: urn:nbn:de:0111-pedocs-149972 - <http://nbn-resolving.org/urn:nbn:de:0111-pedocs-149972>

in Kooperation mit / in cooperation with:

Zeitschrift "Empirische Sonderpädagogik"

<https://www.psychologie-aktuell.com/journale/empirische-sonderpaedagogik.html>

Nutzungsbedingungen

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

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document.

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

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

Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Mitglied der


Leibniz-Gemeinschaft

Empirische Sonderpädagogik, 2017, Nr. 2, S. 98-115
ISSN 1869-4845 (Print) · ISSN 1869-4934 (Internet)

Verlaufsdagnostik und Förderung im Lernbereich Sprachbetrachtung in der Sekundarstufe I mit Moodle

Uwe Maier

Pädagogische Hochschule Schwäbisch Gmünd

Zusammenfassung

Verlaufsdagnostik spielt zunehmend für Diagnostik und Förderung in heterogenen Klassen der Sekundarstufe eine Rolle. Die bisher entwickelten Verfahren fokussieren allerdings überwiegend Grundfertigkeiten in der Primarstufe. In dieser Studie wird deshalb der Frage nachgegangen, inwiefern eine Verlaufsdagnostik nach dem Prinzip des *mastery measurement* geeignet ist, um Lernfortschritte im Bereich Sprachbetrachtung in der Sekundarstufe I abzubilden und individuelle Übungsangebote anzustoßen. Hierfür wurden formative Tests sowie Übungen und Glossare für Teilfacetten der Sprachbetrachtung entwickelt und in einem Moodle-Kurs umgesetzt. An der dreijährigen Begleitforschung nahmen 10 Realschulen mit 34 Lehrkräften und 820 Schülerinnen und Schülern teil. Eine systematische Nutzung der Übungen und Glossare führte zu Lernzuwächsen in den einzelnen Wissensbereichen. Mit der differenzierten Analyse der Nutzerdaten konnten jedoch auch dysfunktionale Nutzungsweisen identifiziert werden. Die Befunde werden herangezogen, um den Moodle-Kurs zu optimieren.

Schlagwörter: Verlaufsdagnostik, Lernfortschrittsdiagnostik, formatives Assessment, Sprachbetrachtung, Deutschunterricht, Sekundarstufe

Student progress monitoring and mastery assessment in the domain of German grammar and spelling for secondary students

Abstract

Student progress monitoring is an important prerequisite for adaptive teaching in highly diverse secondary classrooms. To date, most of the techniques and progress monitoring measures focus primary education learning goals such as basic computation skills or reading fluency. This study aims at investigating if a progress monitoring system, based on the concept of mastery measurement, helps teachers in German secondary classrooms to adapt instruction in the field of German grammar and spelling. We therefore developed a Moodle course with different learning content modules, each of them comprising formative mastery tests for progress monitoring, glossaries and adaptive assignments. We sampled 34 secondary teachers and 820 students who worked with the Moodle course during several months. Moodle test point and assignment scores, glossary clicks and final questionnaire data were analyzed. Learning gains can be explained with a systematic use of adaptive assignments and glossaries. However, results also revealed ineffi-

cient patterns of usage, either depending on the course materials or school-side implementation problems. Evaluation data were used to improve the progress monitoring system.

Key words: Student progress monitoring, curriculum-based measurement, mastery measurement, mastery assessment, secondary education, learning management system, Moodle

Konzepte der Lernfortschrittsdiagnostik

Unter den Stichworten „Lernfortschrittsdiagnostik“ und „Verlaufsdagnostik“ wurden in den letzten Jahren verstärkt Ansätze diskutiert und Verfahren erprobt (z. B. Strathmann & Klauer, 2010; Hasselhorn, Schneider & Trautwein, 2014), die auf US-amerikanische Studien zum *curriculum-based measurement* zurückzuführen sind (z. B. Fuchs et al., 1994; Fuchs, 2004). Mit Verfahren des *curriculum-based measurement* kann die frühzeitige Diagnostik bzw. Prävention von Lernschwierigkeiten und die Überwachung von Förderinterventionen in zentralen Lernbereichen wie dem Lesen oder den Grundrechenfertigkeiten unterstützt werden. Einsatzgebiete sind vorwiegend die Primarstufe und Förderschulen. Es gibt eine Reihe von Studien, in denen die Effektivität von Lernfortschrittsdiagnostik in den Domänen Lesen und Rechnen nachgewiesen wurde (z. B. Topping & Fisher, 2003; Nunnery, Ross & McDonald, 2006; Topping, Samuels & Paul, 2007).

Das entsprechende Testkonzept wird von Fuchs und Deno (1991) als *general outcome measurement* (GOM) oder auch *long-term progress monitoring* beschrieben. Für die Lernfortschrittsdiagnostik nutzt man einen Indikator, der sich leicht testen lässt und dennoch eine zuverlässige Aussage über die Lernentwicklung in der ganzen Lerndomäne liefert. Beispielsweise wird die Leseflüssigkeit als Indikator für Fortschritte im gesamten Leselernprozess herangezogen (Walter, 2010; Souvignier, Förster & Schulte, 2014). Grundlage für die Lernfortschrittsdiagnostik ist die Generierung beliebig vieler, ungefähr gleich schwerer Paralleltestformen. Ebenso muss die Testdurchführungsdauer relativ kurz sein, um das Verfahren

tatsächlich auch im Unterrichtskontext anwenden zu können. Fuchs und Deno (1991) nennen eine Reihe von Vorteilen einer Lernfortschrittsdiagnostik nach dem Prinzip des *general outcome measurement*: Lehrkräften steht ein einfaches und beliebig oft wiederholbares Testverfahren zur Verfügung, um Lernfortschritte in zentralen Domänen über einen längeren Zeitraum zu prüfen. Diese Tests können von Experten entwickelt und standardisiert werden. Durch die langfristig angelegte Messung können auch Behaltens- und Transferleistungen erfasst werden. Ein Nachteil ist allerdings die aufwändige Prüfung der Änderungssensibilität des Indikators (z. B. Klauer & Strathmann, 2013). Es ist nicht immer klar, ob die Lernfortschrittsmessung mit einem Indikator auch alle Lernzuwächse in der Lerndomäne anzeigt.

Die in der Literatur gut dokumentierten und in Studien evaluierten Beispiele einer Verlaufsdagnostik nach dem Prinzip des *general outcome measurement* beschränken sich auf grundlegende, überwiegend prozeduralisierte Fertigkeiten wie Lesen und Rechnen in der Primarstufe. Mittlerweile spielen aber Diagnostik und individuelle Förderung auch in der Sekundarstufe I eine wichtige Rolle. Trotz unterschiedlichster Schulstrukturen gibt es in allen Bundesländern einen Trend zu Schulformen mit mehreren Bildungsgängen sowie inklusiven Konzepten. Zu Beginn der Sekundarstufe I müssen die Lehrkräfte auf ungleiche Lernvoraussetzungen reagieren. Ebenso müssen Lehrkräfte bei ihren Unterrichtsangeboten und Förderkonzepten verschiedene Bildungsabschlüsse antizipieren.

Lerndomänen in der Sekundarstufe I sind zunehmend durch deklarative Wissensbestände geprägt. Für die Beschreibung von Lernfortschritten in diesen Lerndomä-

nen muss es möglich sein, auch den festgelegten Zuwachs an Faktenwissen oder konzeptuellem Wissen beschreiben zu können. Wenn man prüfen möchte, ob eine Schülerin die Interpunktionsregeln kennt und anwenden kann, muss man direkt nach diesen Regeln fragen bzw. deren Anwendung prüfen. Es gibt keinen allgemein gültigen Indikator für Schreibkompetenz, der auch etwas über die Beherrschung von Interpunktionsregeln aussagen könnte.

Fuchs und Deno (1991) beschreiben eine Lernfortschrittsdiagnostik, bei der sämtliche Wissensfacetten mit Tests abgedeckt werden müssen, als *specific subskill mastery measurement*. Die Lerninhalte werden dabei in Teilfertigkeiten gegliedert und als einzelne Lernziele geordnet. Für die Messung wird jeder Teilfertigkeit bzw. jeder Teilfacette der Wissensdomäne ein spezifisches Set an Testitems zugeordnet. Ebenso wird bei jeder Teilfertigkeit ein *mastery criterion* für das Erreichen des Teillernziels festgelegt. Die meisten Beispiele hierzu stammen aus den 1960er und 1970er Jahren und waren Teil des Unterrichtskonzeptes *mastery learning* (Kulik, Kulik & Bangert-Drowns, 1990) und werden im Rahmen der Diskussion über *formative assessment* wieder aufgegriffen (z. B. Guskey, 2007 und 2010; Zimmermann & Dibenedetto, 2008).

Beim *mastery learning* wird der Lernstoff entlang von Begriffen oder Fertigkeiten in kleine Einheiten im Umfang von ein bis zwei Unterrichtswochen aufgeteilt. Nach jeder Einheit folgt ein formativer Kurztest, mit dem geprüft wird, was die Lernenden bereits können und wo noch Lücken sind. Da die Tests kleinschrittig angelegt sind, müssen die Lernenden den Lernstoff nahezu komplett beherrschen (*mastery*). Als Mindestpunktzahl für die Tests (*passing score, cutting score*) haben sich Werte zwischen 80% und 90% etabliert. Das Resultat des formativen Tests sind konkrete Hinweise für Lernende und Lehrkräfte, woran noch gearbeitet werden muss. Die Lehrkraft stellt für diese Phase weitere Aufgaben und Übungen zur Verfügung. Danach bearbeiten die

Lernenden einen weiteren Paralleltest, um den Erfolg der Übungsphase zu prüfen.

Learn-domäne Sprachbetrachtung

Die deutschdidaktische und sprachwissenschaftliche Literatur der letzten Jahrzehnte diskutiert unter den weitgehend synonym verwendeten Begriffen „Sprachbetrachtung“, „Sprachbewusstheit“, „metasprachliche Fähigkeiten“ oder „Sprache und Sprachgebrauch untersuchen“ eine neue Akzentuierung des traditionellen Grammatik- und Rechtschreibunterrichts im Fach Deutsch (Ossner, 2006; Bredel, 2007). Schülerinnen und Schüler sollen in die Lage versetzt werden, grammatikalische Phänomene und Regelungen in sprachlichen Handlungszusammenhängen anwenden zu können. Eichler und Nold (2007, S. 63) differenzieren hierfür zwischen Sprachreflexion/Grammatik und sprachlichem Handeln als zwei miteinander verbundenen Teilbereichen der Sprachbewusstheit. Bei der Sprachbetrachtung steht deklaratives Wissen (Grammatikwissen) im Vordergrund. Beim sprachlichen Handeln wird explizites Wissen (Grammatikwissen) oder implizites Wissen (Sprachgefühl) angewendet, um eigene oder fremde sprachliche Produktionen zu korrigieren. Sprachbewusstheit bzw. metasprachliche Fähigkeiten beziehen sich jedoch nicht nur auf die semantische oder syntaktische Ebene der Sprache. Die für den Schriftspracherwerb wichtige phonologische Bewusstheit wird beispielsweise als metasprachliche Fähigkeit auf Wortebene beschrieben. Metasprachliche Fähigkeiten spielen damit auch für die Erklärung von Teilleistungsstörungen wie der Lese-Rechtschreibschwäche eine wichtige Rolle.

Diese Vorstellung von Sprachbetrachtung als zentralem Bestandteil des Deutschunterrichts wurde in den nationalen Bildungsstandards und den darauf bezogenen Vergleichsarbeiten (VERA 3 und 8) aufgegriffen. Im Kompetenzbereich „Sprache und Sprachgebrauch untersuchen“ in den Bildungsstandards im Fach Deutsch für den

Mittleren Schulabschluss (Kultusministerkonferenz, 2004) geht es sowohl um die Reflexion als auch um die Gestaltung von mündlichen Äußerungen oder Texten in bestimmten funktionalen Verwendungszusammenhängen. Der Kompetenzbereich ist sehr komplex, weil Phänomene der deutschen Sprache auf unterschiedlichen Ebenen (Graphem-Phonem-Korrespondenz, Wort, Satz, Text) mit verschiedensten Verwendungszusammenhängen (Kontexte, Textsorten usw.) in Verbindung zu bringen sind. In den KMK-Standards wird eine Reihe von Inhalten der deutschen Grammatik im Zusammenhang mit ihrer Funktion für den Sprachgebrauch genannt (Kultusministerkonferenz, 2004, S. 19): „Satzstrukturen kennen und funktional verwenden“, „Wortarten kennen und funktional gebrauchen“, „grammatische Kategorien (z. B. Tempus, Modus, Aktiv/Passiv) und ihre Leistungen in situativen und funktionalen Zusammenhängen kennen und nutzen“ sowie „wichtige Regeln der Aussprache und der Orthographie kennen und beim Sprachhandeln berücksichtigen“.

Trotz der Bedeutung dieses Kompetenzbereichs für den Deutschunterricht gibt es keine standardisierten Testverfahren, die von Lehrkräften eingesetzt werden könnten, um das Vorwissen oder die Lernentwicklung erfassen zu können. Die Forschung hierzu fokussiert vor allem Testitems und Kompetenzmodelle zur Sprachbetrachtung für nationale Vergleichsstudien (z. B. Klieme et al. 2006; Eichler & Nold, 2007). Diese eignen sich nicht für eine gezielte Planung von Förderung.

Ziele und Fragestellungen

Ziel des Projektes war, eine digitale Lernverlaufsdagnostik nach dem Prinzip des *mastery assessment* in einem Lernbereich der Sekundarstufe I zu entwickeln und zu evaluieren. Die Lernverlaufsdagnostik sollte sich auf ein Hauptfach beziehen und

über mehrere Schuljahre hinweg anwendbar sein. Aus diesen Überlegungen heraus wurde der Lernbereich Sprachbetrachtung im Fach Deutsch ausgewählt. Im Sinne einer formativen Diagnostik sollte das Instrument Förderoptionen direkt im Anschluss an die Diagnostik anbieten (Übungen sowie Glossare mit Definitionen und Regeln für die Wiederholung). Folgende Fragestellungen standen im Zentrum der Evaluation:

1. Wie und in welchem zeitlichen Umfang setzen die Lehrkräfte die Verlaufsdagnostik in ihrem Deutschunterricht ein?
2. Welche Lernzuwächse erzielen die Schülerinnen und Schüler in den einzelnen Modulen der Verlaufsdagnostik?
3. Lassen sich Lernzuwächse mit der Bearbeitung von Übungen und der Nutzung der Glossare in verschiedenen Phasen der Verlaufsdagnostik erklären?
4. Wie werden die Feedbackinformationen der Verlaufsdagnostik von den Lehrkräften rezipiert und wie bewerten Sie diese im Vergleich zu anderen diagnostischen Informationsquellen aus ihrem Unterricht?

Stichprobe

Die Durchführung des Projektes erstreckte sich über die drei Schuljahre 2013/14, 2014/15 und 2015/16. Im Frühjahr 2013 wurde allen Realschulen in einem Schulamtsbezirk per Anschreiben eine Projektteilnahme angeboten. Mit interessierten Schulen wurden Details zunächst telefonisch oder per E-Mail geklärt. Um das Projekt näher zu erläutern und die technische Ausstattung sowie das geplante Vorgehen zu erklären, wurden schulinterne Lehrerfortbildungen abgehalten. Die endgültige Stichprobe für die Datenanalyse umfasste 34 Klassen (10 Lehrer, 24 Lehrerinnen). Im Mittel lag die Klassengröße bei 24,1 Schülerinnen und Schülern mit einem Minimalwert von 15 und einem Maximalwert von 31 Personen pro Klasse. Die teilnehmenden Klassen verteilten sich auf sieben Realschulen, eine Werkrealschule sowie zwei Schu-

len mit mehreren Schulabschlüssen. Es handelte sich überwiegend um Schulen in ländlichen Gemeinden oder Kleinstädten. Die Schülerstichprobe umfasste 820 Personen, davon 424 Schülerinnen (51,7%) und 396 Schüler (48,3%). Schwerpunktmäßig nahmen Klassen aus den Jahrgangsstufen 7 (16 Klassen) und 8 (15 Klassen) teil. Drei Klassen waren aus der Jahrgangsstufe 5/6.

Verlaufsdiagnostik für den Lernbereich Sprachbetrachtung in Moodle

Moodle ist eine nichtkommerzielle Lernplattform, die weltweit vor allem im Hochschulbereich, zunehmend aber auch an Schulen zum Einsatz kommt. Innerhalb eines Kurses können verschiedenste Aktivitäten zur Verfügung gestellt werden, die von den Kursteilnehmern online aufgerufen bzw. ausgeführt werden können. Die Funktionalität Test in Moodle eignet sich sehr gut für die flexible Programmierung einer formativen Diagnostik, vor allem durch unterschiedliche Test- und Übungsaufgabenformate und vielfältige Möglichkeiten für die Rückmeldung und Analyse von Ergebnissen.

An der Pädagogischen Hochschule Schwäbisch Gmünd konnte der Moodle-Host für das Projekt genutzt werden. Die am Projekt beteiligten Schulklassen erhielten externe Accounts. Die Lehrkräfte wurden als Tutoren angelegt und konnten die Ergebnisse ihrer Klasse über das Internet einsehen.

Die Verlaufsdiagnostik umfasst fünf Test- und Übungsmodule in einem Moodle-Kurs zu den Themen Wortarten, Kommasetzung, Satzglieder, Zeitformen sowie Groß- und Kleinschreibung. Um die inhaltliche Lernprogression abbilden zu können, wurden für jedes Modul Lernziele auf drei Niveaustufen definiert (Tabelle 1). Diese Einteilung der Lernziele orientiert sich an aktuellen Lehrwerken und dem baden-württembergischen Bildungsplan. Niveau 1 entspricht Grundschulwissen, das zu Beginn der Sekundarstufe I beherrscht werden sollte (z. B. die Wortarten Nomen, Verben, Adjektive und Artikel). Auf Niveaustufe 2 wird Wissen im Bereich Sprachbetrachtung geprüft, das im Bildungsplan für die Realschule in den Jahrgangsstufen 6 bis 8 verankert ist. Die Tests auf Niveaustufe 3 prüfen grammatikalisches Wissen, das in Schulbüchern eher selten angesprochen wird. Test- und

Tabelle 1: Übersicht Testmodule und Niveaustufen

Niveau	Wortarten	Kommasetzung	Zeitformen	Satzglieder	Groß-/Kleinschreibung
1	Verben Nomen Artikel Adjektive	Aufzählungen Satzreihen	Präsens Präteritum Perfekt Zukunft I	Subjekt Prädikat Akkusativobjekt Dativobjekt	Nomen Eigennamen Satzanfänge
2	Niveau 1 und Adverbien Pronomen Konjunktionen Präpositionen	Satzgefüge (Konjunktionalsatz) wörtliche Rede	Niveau 1 und- Plusquamperfekt Futur II	Niveau 1 und Adv. Bestimmungen des Ortes, der Zeit, der Art und Weise, des Grundes	Niveau 1 und Abstrakte Nomen Fremdwörter Nominalisierungen AnredeZeitangaben
3	Niveau 2 und Interjektion Numerale Eigennamen	Satzgefüge (Relativsatz) Infinitivgruppen Appositionen	Niveau 2 und Zeitformen bei versch. Textsorten anwenden können	Niveau 2 und Genitivobjekt Präpositionales Objekt	Niveau 2 und Nominalisierung nach Pronomen Mengenangaben Artikel

Übungsaufgaben auf Niveau 3 enthalten zudem komplexeres Wortmaterial und verlangen höhere Transferleistungen (z. B. Kenntnis der Anwendung von Zeitformen in verschiedenen Textsorten).

Die Test- und Übungsfragen bestehen aus einem Wort- und Satzmaterial, das sich an lebensweltlichen Erfahrungen vieler Schülerinnen und Schüler orientiert. Die unterschiedlichen Testaufgabenformate von Moodle werden genutzt, um spezifische Aspekte des Wissens zu prüfen (Abbildung 1). Für die Bereiche Wortarten, Zeitformen (Niveau 1 und 2) und Satzglieder wurde das Aufgabenformat Single Choice gewählt. Die Aufgaben zur Kommasetzung wurden als Lückentexte angelegt, in dem die Kommata durch ein Drop-down-Menü vervollständigt werden müssen. Das Format Lückentext wurde auch bei den Zeitformen (Niveau 3) verwendet. Zur Prüfung von Wissen im Bereich der Groß- und Kleinschreibung wurden Audiodateien in die Testitems eingebunden. Die Schülerinnen und Schüler hörten ein Wort und mussten in einem vorgegebenen Satz die Lücke ergänzen.

Im Moodle-Kurs konnten die einzelnen Lernbereiche frei gewählt werden. Die Lernenden wurden allerdings durch die Kurseinstellungen gezwungen, mit dem formativen Test auf Niveaustufe 1 zu beginnen. Beim Erreichen des Mindestwerts von 80% richtig gelöster Aufgaben (*cutting score*), wurde der nächsthöhere Test freigeschaltet. Falls nicht, konnte ein Test frühestens nach einem Tag wiederholt werden. Es standen keine Paralleltestversionen zur Verfügung. Die Schülerinnen und Schüler wiederholten denselben Test noch einmal. Die damit verbundenen Erinnerungseffekte führen natürlich zu einer Einschränkung der echten Retestreliabilität (siehe nächstes Teilkapitel). Andererseits ist es das Anliegen einer formativen Diagnostik, bereits allein durch die Durchführung von Tests Lernzuwächse anzuregen. Wenn sich eine Schülerin oder ein Schüler die Mühe macht, die Fehler bei der ersten Testdurchführung genau zu studieren, konnte sie bzw. er davon bei der zweiten Testdurchführung profitieren. Dies war didaktisch so gewollt, schränkt aber die Reliabilität der Testwiederholung ein.

The screenshot shows a Moodle PHSG test page. On the left, there is a 'TEST-NAVIGATION' grid with 40 numbered buttons (1-40). Button 1 is highlighted. To the right of the grid, there is a 'Frage 1' section. The question text is: 'Sara schaut aufgeregt, aber aufmerksam auf ihr schickes Smartphone.' Below the text, it says '(W1A_ST_1C)'. The question asks to choose one answer from a list: 'Verb', 'Adjektiv', 'Artikel', and 'Nomen'. The 'Verb' option is selected. At the bottom right of the question area, there is a 'Nächste Seite' button. At the bottom of the page, it says 'Angemeldet ist ni be (Abmelden) DU'.

Abbildung 1: Screenshot mit Beispiel für eine Testaufgabe (Wortarten, Niveau 1)

Bei Nichtbestehen eines Tests standen Übungsaufgaben für die jeweilige Niveaustufe zur Verfügung (Abbildung 2). Die Übungsaufgaben und die Tests unterschieden sich in der Art der Rückmeldung. Während bei den Tests die Rückmeldung erst nach Bearbeitung aller Aufgaben angezeigt wurde, erhielten die Lernenden unmittelbar nach jeder Übungsaufgabe eine Rückmeldung zur Korrektheit der Lösung. Danach konnte die Übungsaufgabe sofort noch einmal bearbeitet werden. Für alle Bereiche

wurden zudem Glossare angelegt, die alphabetisch sortierte Einträge (Regeln, Beispiele, Merksätze) zu den jeweiligen sprachlichen Phänomenen enthielten (Abbildung 3).

Die Lehrkräfte konnten mit der Teilnehmerübersicht das Gesamtergebnis der Klasse bzw. einzelner Schülerinnen und Schüler in den Testmodulen und für die gesamte Klasse einsehen. In einzelnen Testberichten wurden die Ergebnisse der Schülerinnen und Schüler zu einem Testmodul genauer

Die Spieler des SC Glücksburg , X verließen traurig den Platz , weil sie schon wieder verloren hatten. Obwohl _ sie sich so bemüht hatten , konnten sie das Spiel , nicht gewinnen.

(K2_A1a)

Haupt- und Nebensatz werden immer durch Komma voneinander getrennt. Die Nebensätze beginnen hier mit den Konjunktionen *obwohl* bzw. *weil*.

Abbildung 2: Übungsaufgabe Kommasetzung mit Feedback und Hinweisen

Moodle PHSG

basiswissen Deutsch (Sprache)

Wortarten

- Die Wortarten im Überblick
 - Alphabetisch
 - Test Wortarten Niveau 1
 - Übung Wortarten - Niveaustufe 1
 - Übung Wortarten - Niveaustufe 2
 - Übung Wortarten - Niveaustufe 3
- Kommasetzung
- Zeilen
- Satzglieder
- Groß- und Kleinschreibung
- Ergänzende Übungen: dass oder das
- Ergänzende Übungen: Dehnung und Schärfung

EINSTELLUNGEN

@@|A|B|C|D|E|F|G|H|I|J|K|L|M|N|O|P|Q|R|S|T|U|V|W|X|Y|Z|Alle

A

Adjektiv

Mit Adjektiven beschreibst du Lebewesen, Gegenstände, Gedanken oder Ereignisse.

Zum Beispiel:

Der Schnee war pulvrig.

Wir hatten schreckliche Angst.

Viele Adjektive lassen sich steigern: *gut – besser – am besten*; *witzig – witziger – am witzigsten*

Mit Adjektiven kannst du daher Personen und Gegenstände vergleichen: *Lena ist größer als Tim.*

Eine Reihe von Wörtern werden zu Adjektiven, wenn sie eine Adjektivendung erhalten:

Zum Beispiel: freundlich; tierisch; windig; wunderbar

Adverb

Adverbien beschreiben die genaueren Umstände eines Geschehens. Mit ihnen kannst du ausdrücken, wo (*dort, hier...*), wann (*donnerstags, jetzt...*), auf welche Art und Weise (*genauso, vergebens...*) und warum

Abbildung 3: Screenshot aus dem Glossar Wortarten

angezeigt. Richtig beantwortete Fragen wurden grün, teilweise richtig beantwortete Fragen orange und falsch beantwortete Fragen rot markiert. Durch Klicken auf eine einzelne Aufgabe öffnete sich ein Fenster mit der Aufgabenstellung, der Schülerantwort mit Markierungen zur Korrektheit und der erreichten Punktzahl.

Überprüfung der Testgüte

Formative Tests im Rahmen einer Verlaufsdagnostik sollen möglichst exakt das zu diagnostizierende Wissen erfassen und müssen deshalb vor dem Hintergrund der klassischen Testgütekriterien bewertet werden. Andererseits handelt es sich bei formativen Tests nicht um statusdiagnostische Messverfahren, von denen weitreichende Konsequenzen (z. B. Schullaufbahneempfehlungen) abhängig sind. Vielmehr sind formative Tests in den laufenden Unterricht eingebettet, sollen mit wenig Aufwand durchführbar sein und einer Schülerin oder einem Schüler den eigenen Lernfortschritt aufzeigen.

Vor allem bezüglich der Durchführungsobjektivität können die in diesem Projekt entwickelten Tests nicht mit den gleichen Maßstäben bewertet werden wie statusdiagnostische Verfahren. Beispielsweise wurden für die Testdurchführung keine fixen Zeitvorgaben gemacht. Dies würde nicht dem Prinzip des individualisierten Lernens bzw. des *mastery learning* entsprechen. Bei den Durchführungen wurde allerdings darauf geachtet, dass die Schülerinnen und Schüler die Tests eigenständig bearbeiten. In den meisten Klassen wurden die Tests entweder im Computerraum oder im Klassenzimmer mit Tablets unter Aufsicht der Lehrkraft durchgeführt. Nur bei wenigen Durchführungsterminen wurde eine verstärkte Kooperation beobachtet und konnte aufgrund der räumlichen Situation (eng stehende PC-Arbeitsplätze) auch nicht komplett unterbunden werden.

Bei Lernverlaufsdiaagnosen handelt es sich nicht um statische Messungen, deren

Zuverlässigkeit mit einer echten Retestreliabilität gemessen werden könnte. Die zugrunde liegende Schülerfähigkeit zum Messzeitpunkt t1 soll sich aufgrund des gewünschten Lerneffekts von dem Wert eines späteren Messzeitpunkts t2 unterscheiden. Allerdings ist damit zu rechnen, dass die Testwerte zu beiden Messzeitpunkten in einer gewissen Höhe miteinander korrelieren, weil bei nahe aufeinanderfolgenden Messzeitpunkten lediglich mit einem gewissen Lernzuwachs zu rechnen ist. Eine Lernfortschrittsdiagnostik nach dem Prinzip des *general outcome measurement* hat deshalb grundsätzlich den Vorteil, dass aufgrund häufiger Paralleltestdurchführungen die Reliabilität recht gut bestimmt werden kann. Strathmann und Klauer (2010) berechnen beispielsweise eine Pseudo-Retestreliabilität, indem sie die in kurzen Zeitabständen durchgeführten Paralleltests korrelieren und die Korrelationen anschließend mitteln. Im Gegensatz zu diesem Beispiel bearbeitet bei dem hier vorliegenden Konzept nur ein kleiner Teil der Schülerinnen und Schüler die Tests ein zweites oder drittes Mal. Zudem befinden sich zwischen diesen Testwiederholungen eventuell intensive Übungen, die zu einem deutlichen Lernzuwachs führen. Aus diesem Grund wird hier lediglich die interne Konsistenz als Maß für die Zuverlässigkeit der Tests angegeben.

Tabelle 2 zeigt, dass die einzelnen Tests befriedigende bis sehr gute interne Konsistenzen aufweisen. Bei den Modulen Kommasetzung und Zeitformen wurde das Moodle-Testaufgabenformat Lückentext angewendet. Hiermit lassen sich in einem inhaltlich zusammenhängenden kleinen Text beispielsweise mehrere Kommasetzungsregeln oder miteinander verbundene Zeitformen überprüfen. Der Nachteil von Moodle ist, dass lediglich Punkte pro Testaufgabe und nicht Punkte pro Lücke ausgegeben werden. Damit lässt sich der Test nicht auf dem Itemniveau auswerten, was zu einem geringeren Alpha-Wert beiträgt.

Das maximal erreichte Niveau pro Modul gibt Hinweise auf die Schwierigkeit der

Tabelle 2: Interne Konsistenzen (Cronbach's Alpha) auf Basis des jeweils ersten Testversuchs pro Modul und Niveau

	Wortarten	Kommasetzung	Zeitformen	Satzglieder	Groß- und Kleinschreibung
Niveau 1	$\alpha = 0.87$ 842 Testversuche 40 Items	$\alpha = 0.87$ 1132 Testversuche 10 Items	$\alpha = 0.76$ 490 Testversuche 16 Items	$\alpha = 0.94$ 343 Testversuche 27 Items	$\alpha = 0.92$ 366 Testversuche 42 Items
Niveau 2	$\alpha = 0.92$ 631 Testversuche 50 Items	$\alpha = 0.81$ 202 Testversuche 11 Items	$\alpha = 0.81$ 322 Testversuche 20 Items	$\alpha = 0.77$ 200 Testversuche 29 Items	$\alpha = 0.88$ 234 Testversuche 45 Items
Niveau 3	$\alpha = 0.93$ 231 Testversuche 60 Items	$\alpha = 0.80$ 130 Testversuche 10 Items	$\alpha = 0.68$ 171 Testversuche 8 Items	$\alpha = 0.88$ 192 Testversuche 29 Items	$\alpha = 0.86$ 76 Testversuche 44 Items

Tabelle 3: Anteile der maximal erreichten Niveaustufen in den einzelnen Modulen im Verhältnis zu den Modulbearbeitungen

Modul	Bearbeitungen	Maximal erreichtes Niveau in Prozent			
		0	1	2	3
Groß- und Kleinschreibung	344	19.8	49.7	17.2	13.4
Kommasetzung	601	69.2	7.8	6.0	17.0
Satzglieder	286	28.3	19.9	37.4	14.3
Wortarten	747	13.0	51.0	16.5	19.5
Zeitformen	393	22.1	36.1	23.7	18.1
Gesamt	2371	31.6	33.7	17.6	17.1

Tabelle 4: Bewertung der curricularen Validität durch die Lehrkräfte (Ratingskala von 1 = „trifft nicht zu“ bis 5 „trifft zu“)

Item	n	Mittelwert	Standard-abw.
Die Tests und Übungen decken sich inhaltlich mit dem, was auch mir im Kompetenzbereich Sprachbetrachtung (Grammatik) wichtig ist.	32	4.3	0.47
Die Tests und Übungen decken sich inhaltlich mit dem Kompetenzbereich „Sprachbetrachtung“ bei VERA 8.	25	3.7	0.75
Die drei Niveaustufen passten gut zur Leistungsverteilung in meiner Klasse.	32	3.8	1,02
Die Satz- und Wortbeispiele der Tests und Übungen entsprachen dem Leistungsniveau meiner Klasse.	32	4.0	0.72

Tests (Tabelle 3). Bei der Kommasetzung erreichten ca. zwei Drittel nicht einmal das Grundniveau 1. Allerdings gibt es einen nicht unerheblichen Anteil von 17% mit einer maximalen Niveaustufe von 3 im Kommasetzungsmodul. Der höchste Lernerfolg wurde im Modul Satzglieder erzielt. Hier erreichten über 50% der Schülerinnen und Schüler die Niveaustufen 2 oder 3. Diese Unterschiede sind bereits ein deutlicher Hinweis auf die Prioritäten im Deutschunterricht.

Die curriculare Validität einer Verlaufsdagnostik nach dem Prinzip des *mastery measurement* muss sehr hoch sein. Die einzelnen Tests müssen das von den Lehrkräften implementierte Curriculum gut abbilden. Die am Projekt beteiligten Lehrkräfte sahen dies überwiegend als gegeben an. In der Abschlussbefragung bewerteten die Lehrkräfte Aussagen zur curricularen Validität auf einer Skala von 1 „trifft nicht zu“ bis 5 „trifft zu“ (Tabelle 4). Vor allem die inhaltliche Passung der Tests und Übungen mit den Inhalten des eigenen Grammatikunterrichts wird durchgängig als sehr hoch eingeschätzt.

Befragung der Lehrkräfte

Bei Projektbeginn und nach Abschluss des Projektes an den einzelnen Schulen wurden die Lehrkräfte gebeten, einen Evaluationsbogen auszufüllen. Nachfolgend werden lediglich zwei Fragenbereiche aus der Abschluss-evaluation, zu denen Ergebnisse in diesem Aufsatz berichtet werden, beschrieben.

Auf einer fünfstufigen Skala (1 = „trifft nicht zu“ bis 5 = „trifft zu“) sollten die Lehrkräfte den eigenen Umgang mit der Verlaufsdagnostik, vor allem im Hinblick auf die Förderung von Schülerinnen und Schülern einschätzen. Beispielitem: „Bereits während der Testdurchführung habe ich leistungsschwache Schülerinnen und Schüler auf die Übungsmodule hingewiesen (oder auf weitere Hilfsmittel wie Hefteinträge, Glossare).“

Ebenso wurde der diagnostische Nutzen der Verlaufsdagnostik im Vergleich zu an-

deren diagnostischen Informationsquellen erfragt. Zu möglichen diagnostischen Informationsquellen (in Klassenarbeiten geschriebene Schülertexte, im Unterricht geschriebene Schülertexte, benotete Grammatiktests usw.) wurden Fragen zur Einschätzung (Skala von 0 = „trifft nicht zu“ bis 3 = „trifft zu“) vorgelegt, beispielsweise „... helfen mir sehr gut, die Leistungen einzelner Schülerinnen und Schüler im Bereich Sprachbetrachtung einzuschätzen“.

Aufbereitung der Ergebnis- und Interaktionsdaten aus dem Moodle-Kurs

Nach der Datenaufbereitung stand eine SPSS-Datei auf Ebene einzelner Individuen ($n = 820$ Fälle) und eine SPSS-Datei auf Ebene einzelner Modulbearbeitungen ($n = 2371$ Modulbearbeitungen) für die Datenauswertung zur Verfügung. Eine Modulbearbeitung lag vor, wenn in einem Modul (z. B. Kommasetzung) mindestens eine Testbearbeitung mit mindestens 20% korrekter Antworten vorlag. Von den 820 Schülerinnen und Schülern wurden damit im Schnitt etwas weniger als drei Module bearbeitet. Um die Reihenfolge und die zeitliche Passung der Übungen und Glossare in Bezug auf einzelne Testversuche genau erfassen zu können, wurde ein Kategoriensystem entworfen, das sich an Systeme zur Kategorisierung von Lernplattforminteraktionen aus *learning analytics*-Studien anlehnt (z. B. Agudo-Peregrina et al., 2014; You, 2016), jedoch speziell auf die hier vorliegende Verlaufsdagnostik zugeschnitten ist. Eine erste Version des Kategoriensystems wurde an Daten aus fünf Klassen erprobt und anschließend weiterentwickelt (Maier, Ramsteck & Hoffmann, 2017). Bereits in der ersten Version wurde zwischen diagnostischen Aktivitäten (Testversuchen), Übungen und Glossaren unterschieden. In der zweiten Version wurde zusätzlich die zeitliche Anordnung der Übungen bzw. Glossare vor oder nach einem Testversuch auf dem entsprechenden Niveau berücksichtigt. Dahinter verbirgt

sich die Überlegung, dass sich vor allem zeitnah nach einem Test durchgeführte Übungen auf den Lernzuwachs auswirken.

Für die Beantwortung der Forschungsfrage 3 wurden deshalb folgende abhängige und unabhängige Variablen berechnet:

- Abhängige Variable (AV): Lernzuwachs in einem Modul. Dieser wird definiert als Differenz zwischen dem Vorwissen und der maximal erreichten Niveaustufe in einem Modul. Das Vorwissen ist die Niveaustufe, die auf Anhieb ohne Scheitern erreicht wurde. Deskriptive Daten zur AV sind Gegenstand der Forschungsfrage 2 und werden in Ergebnisabschnitt 4.2 berichtet.
- Unabhängige Variablen (UV):
 - UV1: Gesamtpunktzahl der Übungen vor dem ersten Testversuch auf Niveaustufe 1
 - UV2: Gesamtpunktzahl der Übungen innerhalb der Bearbeitung einer Niveaustufe
 - UV3: Gesamtpunktzahl der Übungen nach dem letzten Testversuch auf einer Niveaustufe
 - UV4: Glossarnutzung vor der Bearbeitung einer Niveaustufe
 - UV5: Glossarnutzung innerhalb der Bearbeitung einer Niveaustufe

Zum Verständnis der berichteten Regressionskoeffizienten im Ergebnisteil sind folgende Hinweise wichtig: Pro Übung waren 100 Punkte möglich. Durch die Wiederholung von Übungen konnten einzelne Schülerinnen und Schüler mehr als 100 Übungspunkte sammeln. Mehrmaliges Aufrufen des Glossars in einer Unterrichtsstunde wurde als eine Nutzung gezählt.

Je nach Modul waren die Übungspunkte unterschiedlich verteilt. Im Modul Kommasetzung hatten die Schülerinnen und Schüler vor dem Erstversuch im Schnitt 18.8 Übungspunkte, während der Bearbeitung der Niveaustufe 1 im Schnitt 49.0 Übungspunkte und nach Abschluss der Niveaustufe 1 im Schnitt 45.5 Übungspunkte. D.h. schwerpunktmäßig wurden die Übungen

innerhalb der Niveaustufe bearbeitet. Dies lässt sich vor allem durch die Schwierigkeit des Tests auf Niveaustufe 1 erklären. Nach dem erstmaligen Scheitern wurden Übungen bearbeitet, um den Test anschließend besser bearbeiten zu können. Beim Modul Wortarten zeigte sich ein anderes Bild. Hier wurden die meisten Übungspunkte nach erfolgreichem Abschluss der Niveaustufe 1 gesammelt. Die Mehrzahl der Schülerinnen und Schüler absolvierte den Wortartentest auf Niveau 1 auf Anhieb erfolgreich. Trotzdem wurden Niveau 1-Übungen auch zu einem späteren Zeitpunkt aufgerufen und durchgeführt.

Ergebnisse

Zeitlicher Umfang und Art der Nutzung der Verlaufsdagnostik

Bereits während der Projektdurchführung zeichnete sich eine sehr unterschiedliche Nutzungsintensität ab. In der Abschlussbefragung gaben die beteiligten Lehrkräfte an, für die Arbeit mit dem Moodle-Kurs im Durchschnitt 8.6 Schulstunden zur Verfügung gestellt zu haben (SD = 4.5). Es gibt somit Klassen, bei denen die Verlaufsdagnostik nur in wenigen Unterrichtsstunden zum Einsatz kam und Klassen, die über einen längeren Zeitraum mit den Tests, Übungen und Glossaren arbeiteten. Auch die zeitliche Verteilung der Nutzung ist sehr unterschiedlich. In einigen Klassen wurde wöchentlich in einer festgelegten Zeit mit Moodle gelernt, in anderen Klassen wurde der Einsatz oft durch mehrwöchige Pausen oder Ferien unterbrochen.

Mit den Bewertungen und den Log-Daten aus Moodle konnte die Nutzung differenziert erfasst werden. In 7 der 34 Klassen wurden pro Schülerin bzw. Schüler im Schnitt 4 bis 5 Module bearbeitet. Diese Klassen nutzten alle fünf zur Verfügung stehenden Module der Verlaufsdagnostik. In 7 Klassen wurden lediglich 1 bis 2 Module pro Schülerin bzw. Schüler bearbeitet. Die

durchschnittliche Modulbearbeitung pro Klasse lag bei 2.9. Am häufigsten wurde das Wortartenmodul bearbeitet (31.5%). Danach folgten die Module Kommasetzung (25.3%), Zeitformen (16.6%), Groß- und Kleinschreibung (14.5%) sowie Satzglieder (12.1%). Ein wesentlicher Grund hierfür ist die Anordnung der Module in Moodle sowie die Vorgabe, möglichst entsprechend dieser Reihenfolge zu arbeiten.

Lernzuwächse in den einzelnen Modulen

Tabelle 5 zeigt die Lernzuwächse in den einzelnen Modulen und Niveaustufen. Zunächst wurde ermittelt, bei wie viel Prozent der Modulbearbeitungen auf einer Niveaustufe der erste Testversuch unter 80% lag und bei weiteren Testversuchen der *cutting score* überschritten werden konnte (linke Hälfte der Tabelle). Der Anteil der Schülerinnen und Schüler, die sich nach einem gescheiterten Erstversuch soweit verbessern konnten, um das Niveau erfolgreich abzuschließen, lag über alle Module hinweg bei ca. 10 bis 13 Prozent. Die höchsten Lernzuwächse konnten im Modul Kommasetzung

auf Niveau 1 und 2 verzeichnet werden. Dies lässt sich mit dem geringen Vorwissen in diesem Bereich erklären. Über 20% der Lernenden, die das Niveau 1 nicht auf Anhieb erzielten, konnten sich bei weiteren Versuchen verbessern. Die geringsten Lernzuwächse gab es beim Wortartenmodul auf Niveau 1. Dieses Niveau erzielten sehr viele Schülerinnen und Schüler bereits beim ersten Testversuch.

Die durchschnittlichen Punktzuwächse bei den Schülerinnen und Schülern mit Lernfortschritten liegen bei ca. 20 bis 25 Punkten (bei max. 100 Punkte pro Subtest) mit relativ hohen Standardabweichungen. Einigen Lernenden genügten bereits wenige Punkte, um das Niveau mit 80% (80 Punkte) erfolgreich abzuschließen. Bei anderen Lernenden konnten deutliche Zuwächse von 40 Punkten und mehr verzeichnet werden.

Erklärung der Lernzuwächse durch Glossar- und Übungsaktivitäten

Haben Schülerinnen und Schüler, die Übungen vor oder direkt nach einem Testversuch bearbeitet bzw. sich mit Glossareinträgen beschäftigt hatten, einen höheren

Tabelle 5: Erreichte Niveaustufe und Lernzuwächse nach Modulen

	Anteil erfolgreicher Niveaustufen nach erfolglosem Erstversuch (Anzahl Modulbearbeitungen)			Durchschnittliche Punktzuwächse bei Lernfortschritt (Standardabweichung)		
	Niv. 1	Niv. 2	Niv. 3	Niv. 1	Niv. 2	Niv. 3
Groß- und Kleinschreibung	6.4% (344)	15.5% (187)	4.2% (72)	19.4 (17.4)	20.5 (11.4)	22.0 (7.0)
Kommasetzung	21.3% (601)	15.7% (153)	8.2% (110)	24.8 (14.9)	24.4 (12.8)	24.4 (13.5)
Satzglieder	9.4% (286)	11.2% (169)	21.5% (107)	32.4 (15.5)	14.7 (5.6)	36.0 (13.7)
Wortarten	3.3% (747)	15.0% (494)	10.0% (201)	23.0 (12.6)	19.5 (12.9)	19.6 (13.9)
Zeitformen	9.9% (393)	7.6% (250)	10.8% (130)	24.8 (12.5)	23.7 (10.8)	23.0 (13.4)
Gesamt	10.2% (2371)	13.2% (1253)	11.1% (620)	25.0 (14.9)	20.3 (12.0)	26.5 (14.8)

Lernzuwachs zu verzeichnen? Um diese Frage beantworten zu können, wurden die Übungs- und Glossaraktivitäten im Hinblick auf ihre zeitliche Abfolge kategorisiert. Aus Platzgründen werden lediglich Analysen zum Niveau 1 berichtet.

In Tabelle 6 ist für die einzelnen Module dargestellt, wie sich die Bearbeitung der Übungen und die Glossarnutzung auf den Lernerfolg auf Niveau 1 auswirken. Vorwissen auf Niveau 1 (Punktzahl bei Versuch unter 80%), Jahrgangsstufe, Deutschnote und Geschlecht wurden kontrolliert. Es wurde geprüft, ob Niveau 1-Übungen vor (UV1), während (UV2) und nach der Bearbeitung der Niveaustufe 1 (UV3) bzw. passende Glossare vor (UV4) oder während der Bearbeitung der Niveaustufe 1 (UV5) mit dem Lernzuwachs auf Niveaustufe 1 zusammenhängen.

Das Vorwissen ist wie gehabt der stärkste Prädiktor für den Lernerfolg. Die Lernzuwächse werden in dieser Analyse zudem in drei Modulen von der Deutschnote vorhergesagt. Eine Notenstufe erklärt ca. 2 bis 3 Punkte Lernzuwachs auf Niveau 1. Das Geschlecht spielt nur bei einem Modul eine kaum bedeutsame Rolle. Die Jahrgangsstufe erklärt bei 4 von 5 Modulen einen Teil des Lernzuwachses, was ebenfalls für die curriculare Validität der Tests spricht.

Bei den Übungen sieht man sehr deutlich den Effekt der formativen Diagnostik. Wenn Übungen innerhalb des Niveau 1 bearbeitet wurden, d.h. zwischen dem ersten und letzten Testversuch, dann hat dies einen positiven Effekt auf den Lernzuwachs. Die nicht standardisierten Regressionskoeffizienten geben zudem einen Hinweis auf die Übungsintensität. Im Wortartenmodul

Tabelle 6: Zusammenhänge zwischen Lernerfolg und der zeitlichen Verteilung der Bearbeitung von Übungen und Glossaren (nicht standardisierte Regressionskoeffizienten)

	Groß- und Kleinschreibung	Kommasetzung	Satzglieder	Wortarten	Zeitformen
(Konstante)	65.96 ***	54.34 ***	45.63 ***	67.23 ***	68.39 ***
Deutschnote	n.s.	- 2.62 ***	n.s.	- 1.42 **	- 2.48 **
Geschlecht	n.s.	n.s.	n.s.	1.24 *	n.s.
Jahrgangsstufe	n.s.	7.11 **	7.13 ***	2.17 ***	2.15 *
Vorwissen N1	25.23 ***	18.55 ***	34.80 ***	22.11 ***	25.91 ***
Übungen N1 vor dem ersten Testversuch	n.s.	n.s.	n.s.	0.03 *	n.s.
Übungen innerhalb der Bearbeitung von N1	0.02 *	0.03 ***	n.s.	0.06 ***	0.04 ***
Übungen nach dem letzten Testversuch N1	n.s.	n.s.	- 0.04 *	n.s.	n.s.
Glossarnutzung vor der Bearbeitung von N1	3.49 *	n.s.	n.s.	n.s.	n.s.
Glossarnutzung innerhalb der Bearbeitung von N1	9.49 **	3.22 **	27.64 ***	3.00 ***	n.s.
Anzahl Fälle	340	597	285	743	391
Varianzaufklärung (korrigiertes R-Quadrat)	48.9%	27.8%	62.2%	55.4%	52.4%

Anmerkungen: In der Tabelle sind nicht standardisierte Regressionskoeffizienten dargestellt, n.s. = nicht signifikant, * = $p < .05$, ** = $p < .01$, *** = $p < .001$.

beispielsweise sind 100 Übungspunkte für 6 Punkte Lernzuwachs verantwortlich. Eine komplett bearbeitete Übung ergibt 100 Punkte. Somit konnten in diesem Modul Schülerinnen und Schüler mit der Bearbeitung von drei bis vier Übungen bereits einen deutlichen Lernzuwachs erzielen. In den anderen Modulen waren allerdings mehr Übungen für einen bedeutsamen Lernzuwachs notwendig. Übungen vor dem ersten Testversuch und nach Abschluss der Niveaubearbeitung wirkten sich nicht positiv aus. Nur beim Wortartenmodul gibt es einen Hinweis, dass Übungen vor dem ersten Testversuch auch positiv mit dem Lernzuwachs assoziiert sind.

Die Glossarnutzung innerhalb der Bearbeitung der Tests auf Niveau 1 wirkt sich sehr deutlich auf die Lernzuwächse aus. Auffallend sind allerdings die Unterschiede des Effekts zwischen den Modulen. Vermutlich genügte den Schülerinnen und Schülern, die beim ersten Testversuch auf Niveau 1 nicht erfolgreich waren, der Blick in das Glossar, um die Satzglieder Subjekt, Prädikat, Akkusativobjekt und Dativobjekt zu wiederholen.

Rezeption und Bewertung der Verlaufsdagnostik durch Lehrkräfte

In der Abschlussbefragung wurden die Lehrkräfte gebeten, den Nutzen der Verlaufsdagnostik mit anderen diagnostischen Informationsquellen zu vergleichen (Tabelle 7). Die höchsten Bewertungen erhalten in Klassenarbeiten geschriebene Schülertexte sowie benotete Sprach- und Grammatikaufgaben. Für die beteiligten Lehrkräfte sind somit benotete Aufgabenstellungen, die unter kontrollierten Bedingungen von den Lernenden bearbeitet werden müssen, von höchstem diagnostischem Nutzen, um Defizite im Bereich Sprachbetrachtung bei einzelnen Schülerinnen und Schülern feststellen zu können. Der im Projekt entwickelte Moodle-Kurs wird allerdings im Vergleich zu mündlichen Unterrichtsbeobachtungen und Vergleichsarbeiten in dieser Hinsicht

vergleichsweise gut bewertet. Aus Sicht der Lehrkräfte wird die Verlaufsdagnostik jedoch nicht für die differenzierte Analyse sprachlicher Fehler oder die Dokumentation der Lernentwicklung genutzt. Sie ist dagegen genauso wie Grammatikaufgaben ein Impulsgeber für konkrete Wiederholungen im Unterricht. Ebenso wird sie als nützlich für die Planung von Fördermaßnahmen betrachtet.

Zusammenfassung und Diskussion

Ein erstes Ziel der Studie war die Entwicklung einer formativen Lernverlaufsdagnostik für die Lerndomäne Sprachbetrachtung in Anlehnung an Prinzipien des *mastery measurement*. Durch eine genaue curriculare Analyse konnten für fünf Teilbereiche Tests auf jeweils drei Niveaustufen entwickelt werden. Die teststatistischen Analysen und die Rückmeldungen der Lehrkräfte zeigten, dass dieses Ziel erreicht werden konnte. Mit den Tests kann die Lernprogression im Bereich Sprachbetrachtung in der Sekundarstufe I abgebildet werden. Die Lehrkräfte bestätigten durchweg, dass die Inhalte der Module und die Testaufgaben dem Deutschunterricht der Sekundarstufe I angemessen sind. Kritische Anmerkungen bezogen sich in der Regel auf einzelne Aufgabendetails bzw. einzelne grammatikalische Phänomene.

Ein deutlicher, bereits von Fuchs und Deno (1991) angesprochener Nachteil von *mastery measurement* ist die teststatistische Optimierung der Tests. Trotz sorgfältiger curricularer Analysen war der Niveau 1-Test im Kommasetzungsmodul beispielsweise viel zu schwierig und eignete sich damit nicht für die Vorwissensdiagnostik. Bei einer Überarbeitung der Tests für den Moodle-Kurs wird deshalb ein weiteres Niveau unterhalb des Niveau 1 eingezogen, um die Lernprogression in diesem Bereich besser abbilden zu können. Ein weiteres Problem waren Länge und Wiederholung

Tabelle 7: Bewertung der Verlaufsdiagnostik im Vergleich zu anderen diagnostischen Informationsquellen

	In Klassenarbeiten geschriebene Schülertexte	Im Unterricht geschriebene Schülertexte	Benotete Grammatiktests bzw. Sprachaufgaben	Im Unterricht bearbeitete Sprach- bzw. Grammatikaufgaben	Mündliche Unterrichtsbeobachtungen	DVA 7 bzw. VERA 8 in Deutsch	Die im Projekt durchgeführten Moodle-Tests und Übungen
... helfen mir sehr gut, die Leistungen einzelner S. u. S. im Bereich Sprachbetrachtung einzuschätzen.	2,4	2,0	2,5	1,8	1,8	1,7	2,1
... werte ich im Hinblick auf sprachliche Fehler genau aus.	2,3	1,6	2,6	1,8	1,2	1,9	1,4
... nutze ich, um die Lernentwicklung der S. u. S. im Bereich Sprachbetrachtung zu dokumentieren.	1,8	1,2	2,2	1,4	1,0	1,3	1,4
... nutze ich, um Fördermaßnahmen im Bereich Sprachbetrachtung für einzelne S. u. S. zu planen.	2,0	1,6	2,2	1,5	1,1	1,5	2,0
... nutze ich, um im Unterricht noch einmal sprachliche Phänomene zu wiederholen.	2,2	2,0	2,5	2,3	1,9	1,8	2,2
... geben mir Hinweise, ob ich sprachliche Phänomene im Unterricht gut genug erklärt habe.	2,4	2,2	2,4	2,4	1,8	1,1	2,0

Anmerkungen: Rating von 0 = „trifft nicht zu“ bis 3 = „trifft zu“; S. u. S. = Schülerinnen und Schüler; Stichprobenumfang n = 34.

der Tests. Sämtliche Tests werden bei der Überarbeitung auf eine Itemzahl von 10-15 gekürzt, sodass eine maximale Bearbeitungsdauer von 5 min nicht überschritten wird. Ein mittelfristiger Entwicklungsschritt ist die Rasch-Skalierung. Hierfür muss allerdings die Itembasis deutlich erweitert werden. Ebenso müssten in allen Modulen und Niveaustufen genug Testdurchführungen vorliegen, was nur bei einer wesentlich

breiteren Nutzung des Moodle-Kurses gelingen wird.

Die Daten belegen höchst unterschiedliche Nutzungsdauern und Nutzungsweisen der Verlaufsdiagnostik in den beteiligten Klassen. Wesentliche Gründe für eine intensive Nutzung waren die Unterstützung der Schulleitung, eine geplante und mit anderen Lehrkräften koordinierte Einbindung des Moodle-Kurses in den Deutschunterricht so-

wie das Vorhandensein der entsprechenden technischen Ressourcen (W-LAN, Computerräume, Tablets).

Der Lernerfolg in den einzelnen Modulen war höchst unterschiedlich und spiegelt die Bedeutung der Lerninhalte im Deutschunterricht wider. Viele Lehrkräfte konnten bestätigen, dass die Komposition sowohl in der Primarstufe als auch in der Sekundarstufe zu wenig beachtet wird. Ein weiteres Ergebnis war, dass nur ein kleiner Teil der Lernenden Lernfortschritte innerhalb der Verlaufsdagnostik erzielte. Dies hängt einmal mit der sehr strengen Operationalisierung von Lernfortschritt zusammen. Andererseits zeigt dieser Befund, dass die Verlaufsdagnostik bisher nur in wenigen Klassen im Sinne einer formativen Förderdiagnostik eingesetzt wurde. In vielen Klassen lief der Moodle-Kurs als Differenzierungsangebot neben dem Deutschunterricht und wurde nicht von den Lehrkräften kontrolliert. Nur in wenigen Klassen reagierten die Lehrkräfte zeitnah mit Wiederholungs- und Übungseinheiten auf die Ergebnisse der formativen Tests. In diesen Fällen waren auch substanzielle Lernfortschritte zu verzeichnen. In den vorbereitenden Lehrerfortbildungen sollte deshalb die Implementation der Verlaufsdagnostik wesentlich genauer abgesprochen werden. Die Lehrkräfte sollten sich auf einen zeitlichen Umfang der Nutzung von ca. einem halben Schuljahr verständigen, eine wöchentliche Sequenzierung anstreben und durch eine abschließende Klassenarbeit den Schülerinnen und Schülern die Verbindlichkeit der Arbeit mit dem Moodle-Kurs verdeutlichen.

Mit der Kategorisierung der Moodle-Interaktionen konnten Zusammenhänge zwischen Vorwissen, Lernzuwachsen sowie Übungs- und Glossaraktivitäten nachgewiesen werden. Sowohl passende Übungen als auch die Bearbeitung passender Glossare hängen mit Lernzuwachsen zusammen. Dies spricht für die Effektivität einer Verlaufsdagnostik in Kombination mit gezielten Übungs- und Wiederholungseinheiten. Vor dem Hintergrund der sehr unterschied-

lichen Nutzungsintensität in den Klassen könnten diese Zusammenhänge bei einer besseren Einbindung in den Unterricht eventuell noch höher sein.

Der diagnostische Informationswert der Verlaufsdagnostik wurde von den Lehrkräften relativ gut bewertet. Lehrkräfte vertrauen allerdings ihren selbst entwickelten oder selbst ausgewählten diagnostischen Verfahren am meisten. Bei der Rezeption der Feedbackinformationen ist damit noch „Luft nach oben“. Vor allem in vorbereitenden und begleitenden Lehrerfortbildungen müsste in Zukunft verstärkt auf die unterschiedlichen Rückmeldeformate und Analysemöglichkeiten hingewiesen werden.

Abschließend muss auf forschungsmethodische Limitationen der Studie hingewiesen werden. Es handelt sich um ein nicht-experimentelles Design. Damit sind keine Aussagen über die Effektivität dieser Lernverlaufsdagnostik im Vergleich zu anderen Diagnose- und Förderoptionen möglich. Dies wäre Ziel weiterer Forschungsarbeiten mit standardisierten und stärker kontrollierten Implementationsbedingungen. Eine weitere Schwäche der Studie ist, dass der Lernzuwachs mit den in Moodle eingebauten formativen Tests operationalisiert wird. Durch den formativen Charakter dieser Tests ist deren Retestrelabilität eingeschränkt bzw. nur schwer kontrollierbar. In weiteren Studien wäre es deshalb wünschenswert, wenn man mit einem standardisierten Vor- und Nachtest im Paper-Pencil-Format den Lernzuwachs unabhängig von den Moodle-Tests ermitteln könnte. Nicht zuletzt wäre in weiteren Analysen ein genauer Blick in Muster individueller Lernverläufe wünschenswert. Es ist anzunehmen, dass es qualitativ unterschiedliche Strategien der Bearbeitung des Moodle-Kurses gibt. Hierfür müsste man Muster der individuellen Bearbeitung bzw. Nutzung von Übungen, Tests und Glossaren klassifizieren. Zusammen mit den Daten zum Lernzuwachs könnten auf diese Weise mehr oder weniger effektive Strategien der Bearbeitung einer formativen Lernverlaufsdagnostik identifiziert werden.

Literatur

- Agudo-Peregrina, A.F., Iglesias-Pradas, S., Conde-González, M.A. & Hernández-García, A. (2014). Can we predict success from log data in VLEs? Classification of interactions for learning analytics and their relation with performance in VLE-supported F2F and online learning, *Computers in Human Behavior*, 31, 542-550.
- Bredel, U. (2007). *Sprachbetrachtung und Grammatikunterricht*. Paderborn: Schöningh - UTB.
- Eichler, W. & Nold, G. (2007). Sprachbewusstheit. In B. Beck & E. Klieme (Hrsg.), *Sprachliche Kompetenzen - Konzepte und Messung. DESI-Studie* (S. 63-82). Weinheim: Beltz.
- Fuchs, L.S. (2004). The past, present, and future of curriculum-based measurement research. *School Psychology Review*, 33, 188-192.
- Fuchs, L.S. & Deno, S.L. (1991). Paradigmatic distinctions between instructionally relevant measurement models. *Exceptional Children*, 57, 488-500.
- Fuchs, L.S., Fuchs, D., Hamlett, C.L., Phillips, N.B. & Bentz, J. (1994). Classwide curriculum-based measurement: Helping general educators meet the challenge of student diversity. *Exceptional Children*, 60, 518-537.
- Guskey, T.R. (2007). Formative classroom assessment and Benjamin S. Bloom: theory, research, and practice. In J.H. McMillan (Ed.), *Formative classroom assessment: theory into practice* (pp. 63-78). Teachers College Columbia University New York and London: Teachers College Press.
- Guskey, T.R. (2010). Lessons of mastery learning. *Educational Leadership*, 68, 52-57.
- Hasselhorn, M., Schneider, W. & Trautwein, U. (Hrsg.) (2014). *Lernverlaufsdiagnostik. Tests und Trends N.F. Band 12*. Göttingen: Hogrefe.
- Klauer, K.J. & Strathmann, A.M. (2013). Lernverlaufsdiagnostik Mathematik: Test auf Änderungssensibilität bei rechenschwachen Grundschulern. *Psychologie in Erziehung und Unterricht*, 60, 241-252.
- Klieme, E., Eichler, W., Helmke, A., Lehmann, R., Nold, G., Rolff, H.G., Schröder, K., Thomé, G. & Willenberg, H. (2006). Unterricht und Kompetenzerwerb in Deutsch und Englisch. Zentrale Befunde der Studie Deutsch-Englisch-Schülerleistungen-International (DESI). Deutsches Institut für Internationale Pädagogische Forschung Frankfurt am Main.
- Kulik, C.C., Kulik, J.A. & Bangert-Drowns, R.L. (1990). Effectiveness of mastery learning programs: A meta-analysis. *Review of Educational Research*, 60, 265-299.
- Kultusministerkonferenz (2004). Bildungsstandards im Fach Deutsch für den Mittleren Schulabschluss. Beschluss der Kultusministerkonferenz vom 4.12.2003. München: Kluwer.
- Maier, U., Ramsteck, C. & Hoffmann, K. (2017). Formative Leistungsdiagnostik und Learning Analytics: Entwicklung, Nutzung und Optimierung eines online-basierten Kurses für die Diagnostik und Förderung von Grundwissen im Kompetenzbereich Sprachbetrachtung. *Zeitschrift für Erziehungswissenschaft*, online-first (DOI 10.1007/s11618-017-0776-9).
- Nunnery, J.A., Ross, S.M. & McDonald, A. (2006). A randomized experimental evaluation of the impact of accelerated reader/reading renaissance implementation on reading achievement in grades 3 to 6. *Journal of Education for Students Placed At Risk*, 11, 1-18.
- Ossner, J. (2006). Sprachdidaktik Deutsch. Reihe: Standardwissen Lehramt. Paderborn: Schöningh - UTB.
- Southern, E., Förster, N. & Schulte, E. (2014). Wirksamkeit formativen Assessments – Evaluation des Ansatzes der Lernverlaufsdiagnostik. In M. Hasselhorn, W. Schneider & U. Trautwein (Hrsg.), *Lernverlaufsdiagnostik (Tests und Trends N.F. Band 12)* (S. 221 - 237). Göttingen: Hogrefe.
- Strathmann, A.M. & Klauer, K.J. (2010). Lernverlaufsdiagnostik: Ein Ansatz zur länger-

- fristigen Lernfortschrittsmessung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 42, 111-122.
- Topping, K.J., & Fisher, A.M. (2003). Computerised formative assessment of reading comprehension: Field trials in the UK. *Journal of Research in Reading*, 26, 267, 279.
- Topping, K.J., Samuels, J. & Paul, T. (2007). Computerized assessment of independent reading. Effects of implementation quality on achievement gain. *School Effectiveness and School Improvement*, 18, 191-208.
- Walter, J. (2010). Lernfortschrittsdiagnostik am Beispiel der Lesekompetenz (LDL): Mess-technische Grundlagen sowie Befunde über zu erwartende Zuwachsraten während der Grundschulzeit. *Heilpädagogische Forschung*, 36, 162-176.
- You, J.W. (2016). Identifying significant indicators using LMS data to predict course achievement in online learning. *Internet and Higher Education*, 29, 23-30.
- Zimmerman, B.J. & Dibenedetto, M.K. (2008). Mastery Learning and Assessment: Implications for Students and Teachers in an Era of High-Stakes Testing. *Psychology in the Schools*, 45, 206-216.

Uwe Maier

Pädagogische Hochschule Schwäbisch
Gmünd
Oberbettringer Str. 200
73525 Schwäbisch Gmünd
uwe.maier@ph-gmuend.de

Erstmalig eingereicht: 23.02.2017

Überarbeitung eingereicht: 16.05.2017

Angenommen: 08.06.2017