

Vergleichsarbeiten im Fach Englisch – Einschätzungen von hessischen Lehrkräften

Maria Skejic,¹ Daniela Neumann² und Hogeï Mangal³

After nearly a decade of nationwide *Vergleichsarbeiten* (VERA – comparative standard-based assessment) in Germany, researchers and teachers alike are still struggling with the task of implementing educational standards and system-monitoring in schools. The VERA-test is psychometrically validated and aims at measuring competencies in an objective and reliable manner. Its purpose is to promote school development by providing feedback on students' performance in competency domains as well as a comparison with state-wide test results. The paper reviews the ongoing discussion in educational standards and the potentials and constraints of VERA with regard to foreign language learning. It then provides the views of approximately 700 English language teachers in the federal state of Hessen, who administered the test to their 8th grade students and completed an online questionnaire afterwards. While the teachers provided positive feedback on the material, the majority neither considered the test results useful in improving classroom learning nor the potential impact on school development.

1. Einleitung

Nachdem sich eine Vorgabe, ein bestimmtes Werkzeug oder eine Maßnahme zehn Jahre lang etablieren durfte, ist es Zeit, Bilanz zu ziehen. Von einer guten Bilanz wünscht man sich mindestens, dass die beiden Waagschalen ausgeglichen sind, oder besser noch, dass die Waagschale, die das Nutzen-Verhältnis beschreibt, stärker ausschlägt. Nach zehn Jahren *Vergleichsarbeiten* (VERA) wird derzeit vielfach Bilanz gezogen (vgl. Hellrung & Nachtigall 2013; Isaac 2013; Kuhn 2014; Maritzen 2014; Zimmer-Müller & Hosenfeld 2013), z.B. im Rahmen der 19. Tagung des Netzwerks "Empiriegestützte Schulentwicklung" (EMSE). Vergleichsarbeiten sind schriftliche Tests, die in der dritten und achten Jahrgangsstufe (VERA-3 bzw. VERA-8) *aller* allgemeinbildenden Schulen in Deutschland verpflichtend geschrieben werden. Dabei

-
- 1 Korrespondenzadresse: Maria Skejic, Goethe-Universität Frankfurt, Institut für England- und Amerikastudien, Didaktik, Theodor-W.-Adorno-Platz 1, 60629 Frankfurt am Main, Tel.: (069) 798 32510, E-Mail: skejic@em.uni-frankfurt.de
 - 2 Korrespondenzadresse: Daniela Neumann, Hessische Lehrkräfteakademie, III.2.2. Zentrale Lernstandserhebungen, Walter-Hallstein-Str. 5-7, 65197 Wiesbaden, Tel.: (0611) 5827 220, E-Mail: daniela.neumann@kultus.hessen.de
 - 3 Korrespondenzadresse: Hogeï Mangal, Georg-August-Zinn-Schule, Am Mühlgewann 1, 65933 Frankfurt, Tel.: (069) 212 36100, E-Mail: hogeï.mangal@gmail.com

wird untersucht, welche Kompetenzen Schülerinnen und Schüler zu einem bestimmten Zeitpunkt erreicht haben (vgl. <https://www.iqb.hu-berlin.de/vera>). Im Gegensatz zu angelsächsischen Ländern gibt es in Deutschland "keine Tradition der dauerhaften Beobachtung von Erträgen institutionalisierter Bildungsprozesse" (Baumert & Köller 1998: 12). Es stellt sich daher die Frage, wie Schulen in Deutschland mit dem Paradigmenwechsel zur Outputorientierung umgehen und inwieweit sie dabei testbasierte Ergebnisse aus den bundesweit durchgeführten Vergleichsarbeiten für schulinternes Monitoring nutzen.

Unser Bilanzfokus liegt auf der Haltung der Fremdsprachenlehrkräfte dem Instrument gegenüber und verdeutlicht, wo aus Lehrersicht noch Entwicklungs- und Unterstützungsbedarf liegt. Dabei stellt sich auch die Frage nach der Ausgeglichenheit, ob also der Nutzen an den Schulen bislang so hoch war, dass es die damit verbundenen Aktivitäten rechtfertigte bzw. sie weiterhin rechtfertigt. Die Befragungsergebnisse von ca. 700 hessischen Englischlehrkräften der 8. Jahrgangsstufe geben Aufschluss darüber, inwiefern die VERA Ergebnisse tatsächlich als hilfreich für den eigenen Unterricht empfunden und genutzt werden.

2. Bildungsstandards und Vergleichsarbeiten im fremdsprachendidaktischen und bildungspolitischen Diskurs

In den von Schulleistungsstudien erfassten Kompetenzbereichen werden solche Inhaltsbereiche fachlichen Lernens beschrieben, die einer Operationalisierung zugänglich und daher als Output überprüfbar sind. Dabei besteht innerhalb der Fachdidaktik Uneinigkeit, ob die Kompetenzorientierung von Schulleistungstests dem fremdsprachlichen Unterricht zuträglich ist oder nicht (vgl. z.B. Bredella 2009; Caspari, Kötter, Rossa, Schramm, Tesch, Vollmer & Zydatiñ 2012; De Florio-Hansen 2008; Haß 2012; Küster 2009). Küster (2009: 119) warnt:

Kompetenz- und Standardorientierung schreiben erreichbare Ergebnisse des Lernens fest. Darüber dürfen jedoch weder der Prozess noch die Inhalte des Lernens vernachlässigt werden, sonst verfehlt schulischer Fremdsprachenunterricht seine kulturelle ebenso wie seine pädagogische Dimension.

Fachdidaktiker bemängeln in diesem Zusammenhang die Schwerpunktsetzung auf funktional-kommunikative Kompetenzen, da diese relativ einfach überprüfbar sind. Damit einhergehend wird auch eine Vernachlässigung anderer relevanter Bereiche, wie etwa der interkulturellen Kompetenz und der

Rolle der Literatur, befürchtet. Besonders aus literaturdidaktischer Perspektive wird die geringe Berücksichtigung literarischer Texte in den bildungspolitischen Dokumenten des Gemeinsamen europäischen Referenzrahmens (GeR) (Council of Europe 2001), der Nationalen Bildungsstandards für die erste Fremdsprache (Englisch/Französisch) für den Mittleren Schulabschluss (KMK 2004) und der Nationalen Bildungsstandards für die erste Fremdsprache für den Hautschulabschluss (KMK 2005) kritisiert. Diese "Lerninhalte finden keinen Platz in den Skalen des GER und der Bildungsstandards" (Steiniger 2014: 17). Fremdsprachenunterricht dürfe nicht zu einem instrumentalistischen Lösen von fremdsprachlichen Problemen verkommen (vgl. Bredella 2009; Breidbach 2008; Burwitz-Melzer 2009; Hallet 2012; Küster 2009). Der drohenden theoretischen und praktischen Marginalisierung bestimmter Kompetenzbereiche soll durch neue Ansätze entgegengewirkt werden. Die Deutsche Gesellschaft für Fremdsprachenforschung (DGFF) spricht sich in ihrem Positionspapier für die Erarbeitung von Mindeststandards und Kompetenzstufenmodellen aus, die u.a. auch Mehrsprachigkeit berücksichtigen und "minimale fremdsprachliche Fähigkeiten [...] für ein gelingendes Leben" zu operationalisieren suchen (Caspari et al. 2012: 259).

Während die fachdidaktische Diskussion weiter mit Kompetenzmodellen ringt, stehen in der schulischen Praxis die konkreten Tests im Blickpunkt. Lehrerinnen und Lehrer kennen eine Vielzahl von Testverfahren (z.B. Klassenarbeiten, Material zur Leistungsermittlung vonseiten der Schulbuchverlage, diagnostische Tests, zum Teil auch empirisch validiert). Die Sinnhaftigkeit des jeweiligen Verfahrens erschließt sich aus der Nutzbarkeit gewonnener Informationen für die Leistungsbewertung, der Förderung von Schülerinnen und Schülern bzw. der Anpassung von Unterricht.

Leupold & Porsch warnen vor dem Missverständnis, dass Tests nun den guten Unterricht ablösen würden. Sie unterstreichen, dass auch die "Testmacher" erkennen, "dass die Qualität des Lehr- und Lernprozesses durch Lernaufgaben und nicht durch Testformate ausgelöst wird" (Leupold & Porsch 2011: 91). Während Lernaufgaben, beim Aufbau von Kompetenzen und Fertigkeiten helfen und viel Spielraum lassen für eine lernerorientierte, inhaltliche Ausgestaltung von Unterricht, "zielen Testaufgaben auf die Überprüfung von Kompetenzen" (Caspari, Grotjahn & Kleppin 2010: 47; vgl. Tesch 2009).

Betont wird jedoch, dass im Zuge der Kompetenzorientierung und deren Überprüfung nicht nur Ziele und Anforderungskriterien schulischen Lernens transparenter geworden seien, sondern auch "guten" Aufgaben wieder eine stärkere Gewichtung im Fremdsprachenunterricht eingeräumt worden ist (vgl. Caspari 2010; DGFF 2008). Wichtig erscheint in diesem Zusammenhang zum

einen die Entwicklung von Lern- und Testaufgaben "zur poetisch-imaginativen Funktion von Sprache, zum kritisch-argumentativen und zum kreativen Schreiben" (Tesch 2010: 30), zum anderen die Weiterentwicklung und empirische Überprüfung des *task-based language learning* (vgl. Finkbeiner, Olson & Friedrich 2013: 483; Müller-Hartmann, Schocker & Pant 2013).

Hu & Leupold (2008) weisen auf weitere positive Effekte hin, die sich aus den Testungen der kommunikativen Fertigkeiten ergeben haben. Ausgelöst durch die Bildungsstandards wurden in den vergangenen Jahren beispielsweise auch eher vernachlässigte Domänen verstärkt mit in die Lehrbücher aufgenommen und finden so vermehrt Eingang in den Unterricht. So ist aus den Ergebnissen der DESI-Studie bekannt (DESI-Konsortium 2006), dass Schülerinnen und Schüler deutliche Schwächen im Bereich des fremdsprachlichen Hörverstehens aufweisen. Hörverstehensaufgaben auf der Kompetenzstufe A (d.h. im Bereich elementarer Sprachverwendung, bei dem vor allem Vertrautes, häufig gebrauchte und einfache Sprache eine Rolle spielt) werden von gut einem Drittel der Neuntklässler "nicht mit ausreichender Sicherheit [gelöst]" (DESI-Konsortium 2006: 13). Diese Lernenden sind in allen Bildungsgängen zu unterschiedlichen Anteilen vertreten (vgl. ebd.). Da Aufgaben zum Hörverständnis in den Testheften in Englisch und Französisch eine prominente Rolle eingenommen haben, ist es im Fremdsprachenunterricht zu einer Aufwertung dieses Bereiches gekommen (vgl. Caspari et al. 2010; Maier, Ramsteck & Frühwacht 2013; Tesch 2009). So werde nun

diese Teilkompetenz in den vom IQB [Institut zur Qualitätsentwicklung im Bildungswesen] entwickelten Aufgaben systematisch trainiert und gefördert. Sie wird in Zukunft deutlich mehr Raum im Französischunterricht einnehmen als bisher (Hu & Leupold 2008: 58).

Diese Annahme bestätigt sich z.B. in einer interviewbasierten Studie: An zehn ausgewählten Gymnasien in zwei Bundesländern wurden Lehrkräfte zu ihrer Einstellung und ihrem Umgang zu den Vergleichsarbeiten befragt. Zwar schätzten sie den Nutzen von VERA für den eigenen Unterricht insgesamt als eher gering ein, gaben jedoch an, dass sie durch die Englischvergleichsarbeiten zum Nachdenken bezüglich des Hörverstehens als zentralen Aspekt der kommunikativen Kompetenz in Fremdsprachenunterricht angeregt worden seien (vgl. Maier et al. 2013: 83-86).

Im pädagogischen Diskurs um die Vergleichsarbeiten werden allerdings auch negative Wash-Back-Effekte befürchtet, welche langfristig die Schulkultur beeinflussen könnten (vgl. Breidbach 2008; Finkbeiner 2011). Die Sorge vor schlechten Ergebnissen könnte Schulen zu einem *teaching to the test* verleiten, welches sich in der einseitigen Ausrichtung des Unterrichts auf

die Testinhalte zeigen würde (vgl. Klippel 2009). Tesch hält jedoch den Befürchtungen der Skeptiker entgegen, dass aufgrund der bisherigen Erfahrungen "das Menetekel eines *teaching to the test* sich nicht bewahrheitet zu haben [scheint]" (Tesch 2009: 45). Es wird in der Unterrichtspraxis auf die fachliche und pädagogische Kompetenz der Lehrkraft ankommen, passende Lernangebote zu machen und den Unterricht nicht einseitig auszurichten. Eine Kanonisierung bestimmter, von Experten als relevant eingestufte Inhalte und Texte ist wenig erfolgversprechend. Unterstützt von Didaktikern in Universität und Fortbildung stehen besonders auch Lehrmittelverlage in der Pflicht, den Lehrenden eine große Bandbreite unterschiedlicher Materialien an die Hand zu geben, aus denen diese eine aus ihrer Sicht geeignete Auswahl treffen können (vgl. Küster 2009: 115; Leupold & Porsch 2011).

Die Deutsche Gesellschaft für Fremdsprachenforschung stellt sich in ihrem Positionspapier grundsätzlich hinter die Kompetenzorientierung und objektive, psychometrische Leistungsvergleiche: "Die Entwicklung und Bereitstellung valider diagnostischer Instrumente für Lehrkräfte und Schüler [...] ist dringend geboten" (DGFF 2008: 178). Auch andere Fachdidaktikerinnen und Fachdidaktiker heben die Qualitätsmerkmale und den diagnostischen Mehrwert für Lehrkräfte hervor. Für Nieweler (2010: 220) "liegt der entscheidende Vorteil in der Objektivität der Bewertung" (vgl. auch Harsch 2009). Grotjahn (2009a: 5) sieht "formelle und informelle Tests als unverzichtbare Instrumente einer objektiven und fairen Leistungsbeurteilung im schulischen Kontext."

Finkbeiner stellt fest, dass das Thema Leistungsbewertung im Fremdsprachenunterricht trotz der unumstrittenen Bedeutsamkeit höchstens sporadischen Stellenwert in der Lehrerbildung einzunehmen scheint (vgl. Finkbeiner 2011: 427). Auch Grotjahn weist auf einen Mangel von Testkompetenz von Lehrkräften hin, der durch die Politik in Form von systematischer Implementierung aufgefangen werden kann (vgl. Grotjahn 2009a: 5).

Innerhalb der Bundesländer sind Anstrengungen unternommen worden, um die Vergleichsarbeiten erfolgreich an Schulen zu implementieren. Es zeigt sich jedoch, dass eine breite Lehrerschaft mit dem Instrument immer noch wenig anfangen kann und es weiterer Unterstützung bedarf (vgl. Kuhn 2014). Der Erziehungswissenschaftler Böttcher (2013) kritisiert die Gesamtstrategie der KMK zum Bildungsmonitoring (KMK 2006) und warnt davor, anzunehmen, dass sich die Kenntnisse standardisierter Testverfahren und die Interpretation der Ergebnisse anhand der Kompetenzstufenmodelle "'irgendwie' von der 'Unternehmensspitze in die Klassenzimmer herunterarbeiten' werden (Böttcher 2013: 504). Positive Impulse in der Lehreraus- und -weiterbildung können zu einem profunderem Wissen im Umgang mit Tests sowie fachlicher

Expertise bei der Entwicklung von Verfahren zur Leistungsüberprüfung führen (vgl. Grotjahn 2009a, 2009b). Von einer ungesteuerten Erleuchtung im Sinne einer *deus ex machina* oder der selbstständigen und zum Selbstläufer werdenden Weitergabe von Wissen und Akzeptanz in Bezug auf VERA ist eher nicht auszugehen. Kuhn (2014: 422) schlussfolgert aus seinen Erfahrungen in der Bildungsadministration: "Der Frage, wie das Wissen in die Praxis kommt, muss künftig höhere Priorität eingeräumt werden."

Die Ergebnisse der Vergleichsarbeiten liefern diagnostisch verlässliche Informationen auf Ebene der Klasse. Die kriteriale Bezugsnorm in den Vergleichsarbeiten orientiert sich an den Bildungsstandards (Unterrichtsziele) und bietet eine Rückmeldung zu fachbezogenen Kompetenzbereichen und Kompetenzniveaus (inhaltsbezogene Lernergebnisse untergliedert in elementare, selbständige und kompetente Sprachverwendung nach dem GeR). Neben der kriterialen, erfolgt eine Rückmeldung auch in Bezug auf die soziale Bezugsnorm. Sie bezieht sich wie in Klassenarbeiten auf die Leistungen der Lerngruppe (vgl. De Florio-Hansen 2008; Grotjahn 2009b; Harsch 2009), wird jedoch um den Mittelwert des einzelnen Bundeslandes oder in manchen Bundesländern einen sogenannten "korrigierten Landesmittelwert" erweitert. Caspari et al. (2010: 48) konkretisieren:

Der Einsatz von Testaufgaben im Rahmen des Bildungsmonitoring [zielt] nicht auf eine Individualdiagnostik sondern auf einen Vergleich größerer Einheiten [...]. Auf der Basis dieser und weiterer diagnostischer Informationen können dann gegebenenfalls Maßnahmen zur konkreten Lernförderung abgeleitet werden.

Harsch (2009: 14) bezieht sich auf die Auswertungsberichte und didaktischen Begleitmaterialien, die Lehrkräften nach der Testdurchführung zur Verfügung gestellt werden und fügt hinzu:

Wenn Aufgaben einzeln charakterisiert und didaktisch kommentiert werden, wie das bei Vergleichsarbeiten Klasse 8 intendiert ist, so können sie der einzelnen Lehrkraft, die ihre Klasse damit testet, sogar diagnostisches Feedback geben im Hinblick auf die Frage, wo die Stärken und Schwächen der getesteten Klasse liegen, auch und gerade im Vergleich zu anderen Klassen.

3. Evaluation im Unterricht

Tradierte Leistungsbeurteilungen in Form von Klassenarbeiten und mündliches Abfragen nehmen viel Raum und Zeit in der Schule ein und sind aufwändig. Dabei ist der Blick oft verengt. Häufig werden die Leistungen der Lernenden lediglich mit Hilfe der Schulnoten von eins bis sechs eingestuft,

ohne differenzierte Auskunft darüber zu erteilen, was getan werden kann, um besser zu werden (vgl. Winter 2015: 11). Eine Abkehr von der bloßen Kommunikation eines "Urteils" (der Note) oder einer hauptsächlich defizitorientierten Rückmeldung hin zu einem Feedback, das die Stärken und Schwächen der Schülerinnen und Schüler kommuniziert, bleibt ein schulpraktisches Desiderat (vgl. Drücke-Noe 2014; Müller-Hartmann et al. 2013; Weskamp 2007).

VERA ermittelt die Stärken und Schwächen der Schülerinnen und Schüler in Bezug auf die abschlussbezogenen Bildungsstandards und fachlichen Kompetenzbereiche. Die Evaluation der Ergebnisse ermöglicht Erkenntnisse über den Erfolg des eigenen Unterrichts in Bezug auf die getesteten Bereiche und gibt Impulse für die Unterrichtsentwicklung an Schulen (vgl. Nachtigall, Storbeck & Landmann 2009).

VERA ersetzt aber keine anderen Verfahren wie z.B. informelle Tests mit integrativen Aufgabenformaten oder solchen Sprachtests, die eine differenzierte Individualdiagnostik ermöglichen (vgl. Grotjahn 2008). Dieser Punkt kann nicht häufig genug betont werden, da an vielen Schulen die Ablehnung der Vergleichsarbeiten mit der Annahme einhergeht, das Instrument VERA entkräfte und ersetze alle bislang in der Klasse erfolgten Leistungsfeststellungen.

Evaluationen dienen der Implementierung der kompetenzorientierten Bildungsstandards, indem sie Orientierung über die in den Fächern zu erreichenden abschlussbezogenen Ziele erlauben. Lehrerinnen und Lehrer empfinden die Durchführung und Auswertung derartiger Tests neben anderen pädagogischen Aufgaben im Schulalltag häufig als Belastung. Im Sinne eines professionellen Lehrerhandelns beginnt mit dem Erhalt der Ergebnismrückmeldung jedoch erst die Hauptaufgabe: Die Besprechung und Evaluation der Ergebnisse für die eigene Unterrichtsentwicklung, für fachliche Diskussionen in den schulischen Gremien und im Rahmen schulischer Veränderungsprozesse in den entsprechenden Steuerungs- und Schulentwicklungsgruppen. Unter inhaltlicher Leitung von Wolfgang Blum (Universität Kassel) wurde genau dieser Prozess des Besprechens, Interpretierens und Evaluierens im Projekt VELM-8 ("Verbesserung der Effektivität der Lernstandserhebungen Mathematik Klasse 8") wissenschaftlich gesteuert und begleitet. 29 hessische Lehrkräfte aus sechs Schulen wurden im Fach Mathematik unter diagnostischen, unterrichtsplanerischen und abschlussbezogenen Aspekten an das Instrument VERA-8 herangeführt. Die Evaluation des Projekts VELM-8 hat ergeben, dass sich im Rahmen des Projekts die Einstellungen und Sichtweisen der an der Stichprobe beteiligten Lehrkräfte verändert haben (vgl. Vogel, Achmetli, Krawitz & Blum 2014: 1254). Eine Änderung von Sichtweisen und Einstellungen geht im Idealfall auch mit einer Veränderung der pädagogischen und fachlichen-

didaktischen Entscheidungen in der Schule einher. So konnten durch das Projekt VELM-8 positive Veränderungen in den Einstellungen und der Sichtweise der Lehrkräfte erreicht werden. Lehrerinnen und Lehrer gaben seltener an, VERA als Belastung zu empfinden, sie sahen mehr förderdiagnostischen Nutzen in dem Instrument und konnten ihm mehr Hinweise für die zukünftige Unterrichtsgestaltung entnehmen (Vogel et al. 2014). Ein ähnliches Ergebnis wurde im Rahmen einer Befragung von Schulleiterinnen und Schulleiter an Berliner und Brandenburger Schulen erzielt. Sie gaben im Rahmen einer Befragung an, dass die Testergebnisse nicht nur für die Unterrichtsentwicklung, sondern ebenso für die Personal- und Schulentwicklung genutzt werden. Die Ergebnisse der Vergleichsarbeiten werden vor allem in den Bereichen "Lehr- und Lernprozesse initiieren" (75%), "kollegiale Unterrichtsentwicklung betreiben" (60%), "schulinterne Strukturen ausbauen" (40%) und "Personalentwicklung wie Fortbildung und Hospitationen ermöglichen" (40%) genutzt (Wurster, Richter, Schliesing & Pant 2013: 37).

4. Rezeption und Nutzung von VERA: Rückmeldung hessischer Lehrkräfte

Der Umgang mit VERA fällt in allen Bundesländern etwas unterschiedlich aus. In Hessen werden die Ergebnisse der Vergleichsarbeiten bzw. der Zentralen Lernstandserhebungen (Bezeichnung in Hessen für VERA) beispielsweise bislang nicht an die Schulaufsicht weitergeleitet. Die Schulen sind angehalten, den Test weder zu benoten noch die Ergebnisse zur Entscheidung eines Abschlusses oder der Wahl einer weiterführenden Schulform als Kriterium heranzuziehen. Vielmehr sollten die Ergebnisse aus den Lernstandserhebungen für die Schul- und Unterrichtsentwicklung genutzt werden (vgl. HKM, Erlass vom 29. Juni 2011 und HKM 2011, VOGSV §32).

Im Lernstandsportal, einem für die Schulen durch Passwort geschützten Raum, stehen den Schulen für jedes Fach ausführliche didaktische Materialien zur Weiterarbeit im Unterricht sowie verschiedene Ergebnisberichte zur Verfügung. Dies sind der Sofortbericht, die tabellarische Auswertung und der Ergebnisbericht, jeweils mit den Ergebnissen der eigenen Lerngruppe, des Weiteren ein Schulbericht, mit den Ergebnissen der Schule und seit dem Schuljahr 2014/15 ein Schülerbericht mit den Ergebnissen der einzelnen Schülerinnen und Schüler. Sofortbericht und tabellarische Auswertung enthalten keine Vergleiche und können direkt nach Eingabe der Testergebnisse der Klasse eingesehen werden. Im Schülerbericht erfolgt ein Vergleich mit

allen Schülerinnen und Schülern und im Schulbericht mit allen Schulen, die in Hessen die gleiche Testheftversion bearbeitet haben. Der Ergebnisbericht enthält als Vergleichswert den sogenannten "korrigierten Landesmittelwert" ähnlicher Lerngruppen. Dafür werden jene Hintergrundinformationen der Lerngruppe berücksichtigt, "auf die Schule und Unterricht geringen bis keinen Einfluss haben" (Korngiebel 2014: 80). Solche Einflussfaktoren sind die Schulform, das Geschlecht, die Sprache im Haushalt, der sozioökonomische Status der Eltern oder der diagnostizierte Förderbedarf. Diese Einflussfaktoren können beim Vergleich der eigenen Ergebnisse mit dem korrigierten Landesmittelwert nicht mehr als Erklärung für ein bestimmtes Abschneiden herangezogen werden. Vielmehr spielen bei der Ergebnisanalyse schulische, didaktische und methodische Begründungszusammenhänge eine entscheidende Rolle. Die Schulen erhalten den Ergebnisbericht ca. sechs Wochen nach der Testdurchführung.

4.1 Forschungsfragen

Die vorliegende Untersuchung interessiert sich für den Umgang der Lehrkräfte mit den Lernstandserhebungen (LS). Insbesondere die Einschätzungen der Lehrkräfte in Bezug auf das Instrument und die Nutzung der Ergebnisse für Schul- und Unterrichtsentwicklung, stehen im Fokus. Folgende Fragen werden im weiteren Verlauf untersucht:

1. Wie verläuft die Vorbereitung auf die Lernstandserhebungen an den Schulen?
 - a. Wie bereiten sich die Lehrkräfte selbst auf die Lernstandserhebungen vor?
 - b. Wie bereiten die Lehrkräfte ihre Schülerinnen und Schüler auf die Lernstandserhebungen vor?
2. Wie bewerten die Lehrkräfte die Materialien im Fach Englisch der 8. Jahrgangsstufe?
3. Welche Einstellungen haben die Lehrkräfte zu den Lernstandserhebungen?
4. Wie gehen die Lehrkräfte mit den Ergebnissen um?
5. Was wäre notwendig, damit die Lernstandserhebungen für die Schul- und Unterrichtsentwicklung genutzt werden können?

4.2 Beschreibung der Datengrundlage

Die Teilnahme an den Zentralen Lernstandserhebungen in Hessen war 2014 für jede Klasse der Sekundarstufe mit einem frei zu wählenden Fach verpflichtend. Insgesamt haben sich 4169 Klassen an den Lernstandserhebungen beteiligt, 2845 davon in der 8. Jahrgangsstufe. 1058 Klassen (ca. 37%) entschieden sich dafür, die Lernstandserhebungen im Fach Englisch zu schreiben. Nach der Testdurchführung hatten die Lehrkräfte die Möglichkeit, sich an einer freiwilligen schriftlichen Online-Befragung zu beteiligen. Daran nahmen im Fach Englisch 693 Lehrkräfte teil. Sie konnten mit 96 geschlossenen Fragen die Aufgabenhefte und die Durchführung bewerten, im Rahmen von sieben Fragen bestand die Möglichkeit, offene Antworten zu hinterlassen. Obwohl nur ein Teil der Lehrkräfte eine Rückmeldung zu den Lernstandserhebungen gegeben hat, kann davon ausgegangen werden, dass es sich dabei nicht nur um sehr engagierte und den Lernstandserhebungen gegenüber positiv gestimmte Kolleginnen und Kollegen handelt. Die offenen Antworten legen nahe, dass gerade auch eher kritisch eingestellte Lehrkräfte die Gelegenheit nutzten, eine Rückmeldung zum Verfahren zu geben.

Bei den verwendeten Skalen handelt es sich um Likert-Skalen, die eine Reihe von Einstellungsitems bzw. Wertitems enthalten. Antworten auf einer Likert-Skala sind formal ordinal- beziehungsweise rangskaliert, da nicht notwendigerweise angenommen werden kann, dass Testteilnehmer die verschiedenen Antwortmöglichkeiten als gleichwertig wahrnehmen. Zu jedem Item werden die Probanden nach dem Ausmaß der Übereinstimmung oder Nichtübereinstimmung gefragt bzw. werden aufgefordert, Noten von eins bis sechs zu vergeben.

4.3 Aufbereitung der Daten

Die Variablen der einzelnen Teile des Fragebogens wurden entsprechend gruppiert. Items wurden, wenn erforderlich, umcodiert, um eine positive Korrelation zu gewährleisten. Die Prüfung der Skalenreliabilität erfolgte mit dem Statistikprogramm SPSS 22.0. Berechnet wurde der Reliabilitätskoeffizient Cronbach Alpha für die einzelnen Variablengruppen (Cronbach 1951). In der sozialwissenschaftlichen Forschung gelten Skalen mit Alpha zwischen 0,7 und 1,0 als akzeptabel (vgl. Bortz & Döring 2006). Die Reliabilitäten für die berichteten Variablengruppen liegen zwischen 0.76 und 0.95.

Variablengruppen:

- | | |
|--|--------------------------|
| a. ID Schule | (Variable 1) |
| b. Schulform | (Variable 2) |
| c. Durchführung (Lehrkraft) | (Variablen 3-7) |
| d. Durchführung (Einschätzung SuS) | (Variablen 8-14) |
| e. Bewertung der Englisch-Materialien | (Variablen 15-22/23) |
| f. Eigene Vorbereitung | (Variablen 26-30) |
| g. Vorbereitung der SuS auf die LS | (Variablen 31, 32/33-40) |
| h. Einstellungen der Lehrkräfte zu den LS | (Variablen 41-49/50, 51) |
| i. Umgang mit den Ergebnissen | (Variablen 52-63) |
| j. Bewertung didaktische Materialien | (Variablen 64/65-73) |
| k. Optimierungswünsche der Lehrkräfte | (Variablen 74-80) |
| l. Konsequenzen durch Teilnahme am LS an Schule | (Variablen 81-89) |

4.4 Ergebnisse

Die folgenden Bewertungen beziehen sich nur auf die Rückmeldungen der Lehrkräfte im Fach Englisch in der 8. Jahrgangsstufe. Berichtet werden die Variablengruppen e) bis k).

4.4.1 Vorbereitung auf die Lernstandserhebungen

Im Rahmen der Befragung wurde erhoben, wie Lehrkräfte sich selbst auf die Lernstandserhebungen vorbereiten. Für die fünf Variablen dieser Gruppe besteht ein Cronbach Alpha von 0,79. Es lagen 655 Antworten vor. Nur knapp 5% der Englisch-Lehrkräfte gaben an, im Herbst des Vorjahres eine Informationsveranstaltung zu den Lernstandserhebungen besucht zu haben. Die Veranstaltung wurde von etwa 52% der Teilnehmenden als hilfreich empfunden, ca. 32% positionierten sich neutral und ca. 16% gaben an, der Besuch der Veranstaltung sei nicht hilfreich für die Durchführung und Nutzung der Lernstandserhebungen gewesen.

Insgesamt gut auf die Lernstandserhebungen vorbereitet, z.B. durch die Anschreiben vom Landesschulamt, das zur Verfügung gestellte Informationsmaterial und die Website zu den Zentralen Lernstandserhebungen, fühlten sich ca. 55% der Englischlehrerinnen und -lehrer. Ca. 15% gaben an, sie fühlten sich nicht gut vorbereitet, ca. 30% positionierten sich neutral.

Zusammenfassend kann man festhalten, dass sich zwar nur etwa die Hälfte der Englischlehrkräfte gut auf die Lernstandserhebungen vorbereitet fühlt, das Informationsangebot im Rahmen einer Informationsveranstaltung aber nur von einem sehr geringen Anteil der Lehrkräfte angenommen wurde. Hier gilt es herauszufinden, welche Art von Vorbereitung sich die Lehrkräfte wünschen, um zukünftig darauf reagieren zu können.

Für die Information der Eltern sowie der Schülerinnen und Schüler zu den Lernstandserhebungen stellt das Landesschulamt Informationsflyer zur Verfügung. Die Elternflyer wurden direkt in ausreichender Anzahl in den Testpaketen mit an die Schulen geschickt. Den Schülerflyer erhielten die Schulen als pdf-Dokument. Er musste von den Lehrkräften ausgedruckt und kopiert werden. Der Elternflyer wurde von ca. 83% der Englischlehrkräfte ausgeteilt, der Schülerflyer von ca. 80%.

Darüber hinaus wurden die Lehrkräfte gebeten, anzugeben, inwiefern sie ihre Klassen bzw. teilnehmende Schülerinnen und Schüler auf die Lernstandserhebungen vorbereitet haben (vgl. Abbildung 1). Es lagen Antworten von 654 Lehrkräften vor. Für die zehn Variablen dieser Gruppe besteht ein Cronbach Alpha von 0,76.

Knapp 80% der Lehrkräfte haben den Schülerflyer nicht nur ausgeteilt, sondern auch mit der Klasse besprochen. Ca. 70% der Lehrkräfte haben großen Wert darauf gelegt, die Schülerinnen und Schüler im Vorfeld zu motivieren. Ca. 40% der Lehrkräfte haben im Vorfeld die Aufgabenformate anhand älterer Lernstandsaufgaben besprochen. Diese drei Möglichkeiten zur Vorbereitung werden vom Landesschulamt z.B. im Rahmen der Informationsveranstaltung ausdrücklich empfohlen. Von weiteren Vorbereitungen wird eher abgeraten. Die didaktischen Materialien aus dem Vorjahr (2013) wurden im Vorfeld von ca. 30% der Lehrerinnen und Lehrer eingesetzt, um damit für die Lernstandserhebungen zu üben. Ca. 25% der Lehrkräfte haben mit älteren Lernstandsaufgaben geübt oder die Testdomänen nochmals im Unterricht wiederholt. Ca. 15% der Englischlehrkräfte gaben an, Übungsmaterial der Verlage zur Vorbereitung der Klasse eingesetzt zu haben.

Erfolgt diese Art der Vorbereitung in Maßen und eher punktuell, kann sie für die Schülerinnen und Schüler sinnvoll sein. Die Wiederholung einer Domäne, die ggf. seit längerem nicht mehr Unterrichtsgegenstand war, ist per se ja durchaus empfehlenswert. Wenn diese Art der Vorbereitung jedoch über einen längeren Zeitraum und systematisch erfolgte, wäre sie im Sinne eines *teaching to the test* kritisch zu beurteilen. Häufig gehen derartige Vorbereitungen oder Trainings mit einer Verarmung des Unterrichts einher und sind nur noch wenig kompetenzorientiert. Erfreulicherweise zeugt die Mehrheit

der Rückmeldungen davon, dass keine übertriebene Vorbereitung auf die Tests stattfindet und daher kein negativer Einfluss auf die Unterrichtsgestaltung auszugehen scheint.

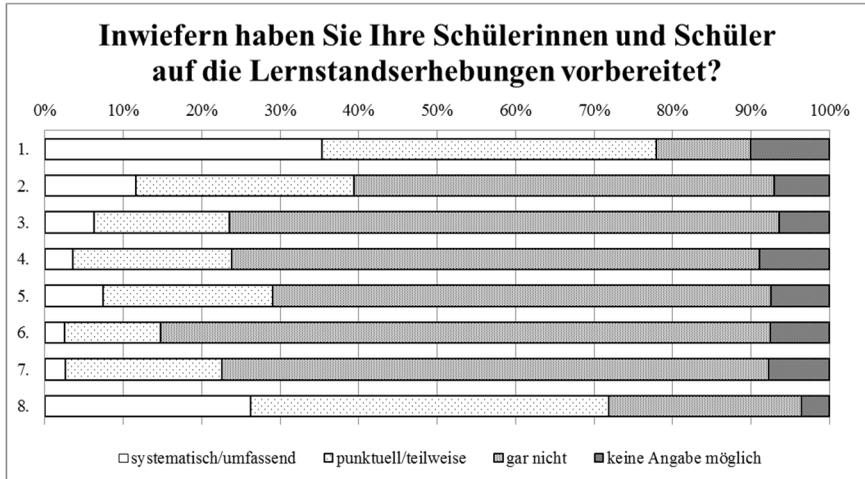


Abbildung 1: Vorbereitung auf die Lernstandserhebungen

Legende: 1. Ich habe den Informationsflyer für Schülerinnen und Schüler mit meiner Lerngruppe besprochen; 2. Ich habe im Vorfeld die Aufgabenformate anhand älterer Lernstandsaufgaben besprochen; 3. Ich habe im Vorfeld mit älteren Lernstandsaufgaben geübt; 4. Ich habe die Testdomänen nochmals im Unterricht wiederholt; 5. Ich habe mit den Didaktischen Materialien 2013 für die Lernstandserhebungen geübt; 6. Ich habe spezielles Übungsmaterial der Verlage eingesetzt; 7. Um meine Schülerinnen und Schüler während der Lernstandserhebung nicht zu demotivieren, musste ich Hilfestellungen geben; 8. Ich habe großen Wert darauf gelegt, meine Schülerinnen und Schüler im Vorfeld zu motivieren.

Hilfestellungen während der Lernstandserhebungen widersprechen dem standardisierten Verfahren und wurden von der Mehrheit der Lehrerinnen und Lehrer nicht gegeben. 22% der Lehrkräfte gaben an, Schülerinnen und Schülern während der Testdurchführung teilweise geholfen zu haben. Etwa 3% (17 Lehrkräfte) gaben an, systematische und umfassende Hilfestellungen gegeben zu haben. Die von diesen Klassen erzielten Ergebnisse dürften nicht dem tatsächlichen Fähigkeitsstand entsprechen und daher auch nur wenige aufschlussreiche Informationen für die Lehrkräfte beinhalten. Ob in diesen Klassen eine sinnvolle Auseinandersetzung mit den Ergebnisberichten und daraus eine Ableitung von Maßnahmen für den weiteren Unterricht erfolgen, ist zu bezweifeln.

4.4.2 Bewertung der Materialien im Fach Englisch der 8. Jahrgangsstufe

Die Qualität der Aufgaben im Testheft und die Begleitmaterialien zur Durchführung wurden von 687 Lehrkräften bewertet. Für neun Variablen wird ein Cronbach-Alpha von 0.82 ermittelt.

Insgesamt wurden die Materialien weitgehend positiv eingeschätzt (vgl. Abbildung 2). Das Testheft und die entsprechende Korrekturanweisung wurden von über 60% der Englischlehrkräfte mit der Schulnote 1 oder 2 bewertet. Die CD wurde etwas schlechter eingestuft. Dies könnte damit zusammenhängen, dass es sich bei den Audio-Beiträgen um authentisches Material mit leichten Störgeräuschen, z.T. leichten Akzentfärbungen und anderen Merkmalen "echter" gesprochener Sprache handelt. Diese authentischen Beiträge unterscheiden sich von didaktisierten und vereinfachten Hörtexten, die von Schulbuchverlagen angeboten werden.

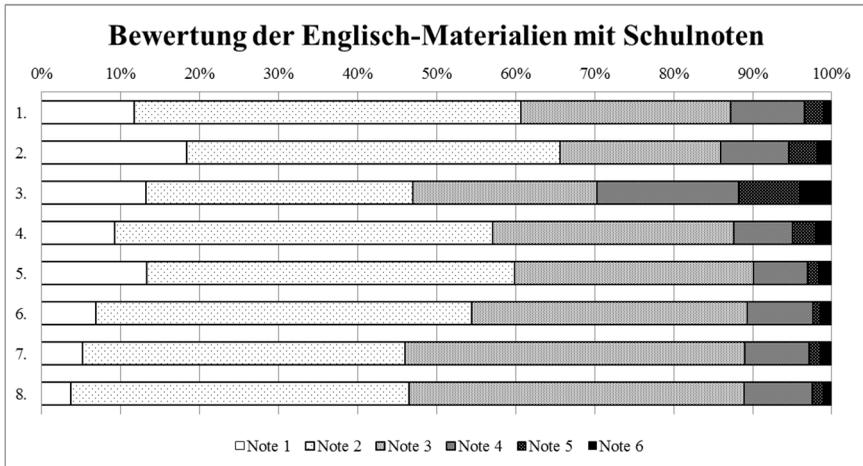


Abbildung 2: Bewertung der VERA 8 Materialien Englisch 2014

Legende: 1. Testheft; 2. Korrekturanweisung; 3. CD zum Hörverstehen; 4. Tabellarische Auswertung; 5. Sofortbericht; 6. Ergebnisbericht; 7. Schulbericht; 8. Schülerbericht; Antwortkategorien: Note 1 bis 6

Auch die Rückmeldeformate nach der Testdurchführung (Sofortbericht, tabellarische Auswertung, Ergebnisbericht, Schulbericht und Schülerbericht) wurden grundsätzlich positiv bewertet. Nur etwa 10% der Lehrkräfte bewertete die Rückmeldeformate mit der Note 4 oder schlechter. Allerdings gaben 15% bzw. 19% der Lehrkräfte an, die tabellarische Auswertung bzw.

den Sofortbericht nicht zu kennen. Diese beiden Berichte erhalten die Lehrkräfte direkt nach der Testdurchführung und Ergebniseingabe im Lernstandsportaal. 42% der Englisch-Lehrkräfte berichteten, den Ergebnisbericht nicht zu kennen und sogar 53% antworteten, den Schulbericht und auch den Schülerbericht nicht zu kennen. Ergebnis-, Schul- und Schülerbericht erscheinen zeitversetzt ca. sechs Wochen nach den Sofortberichten. Die Schulen werden über das Erscheinen informiert und aufgefordert, die Berichte aus dem Lernstandsportaal herunterzuladen bzw. die Informationen an die teilnehmenden Lehrkräfte weiterzuleiten. Für viele Lehrerinnen und Lehrer könnten die Lernstandserhebungen zu diesem Zeitpunkt bereits als abgeschlossen gelten. Dies könnte die Ursache für die vergleichsweise geringe Bekanntheit der drei Berichte sein.

4.4.3 Einstellungen der Lehrkräfte zu den Lernstandserhebungen

Die Lehrkräfte wurden gebeten, die Lernstandserhebungen im Allgemeinen zu bewerten. Sie hatten die Möglichkeit neun Aussagen zu den Lernstandserhebungen mit ++, +, 0, - oder - - einzuschätzen und so ihre Zustimmung bzw. Ablehnung zum Ausdruck zu bringen (vgl. Abbildung 3). Es lagen 653 Antworten von Lehrkräften vor. Die zugrundeliegenden neun Variablen ergeben ein Cronbach-Alpha von 0.95.

Generell zeigte sich, dass ca. 40% der Englischlehrkräfte die Lernstandserhebungen und ihren Nutzen für die schulische Arbeit eher kritisch sehen und sich in Bezug auf die Aussagen negativ positionieren. Insbesondere wurde häufig bezweifelt (Ablehnung über 50%), dass von den Lernstandserhebungen Impulse für die pädagogische Diskussion ausgingen und sie einen Beitrag zur langfristigen Qualitätsentwicklung an der Schule leisteten. Gründe für diese Ansichten könnten ein Zeit- und Ressourcenmangel sein, der es den Lehrkräften bislang noch nicht erlaubte, sich intensiv auch auf Fach- und Jahrgangsebene mit den Ergebnissen der Lernstandserhebungen auseinander zu setzen und für entsprechende Maßnahmen nicht nur für die einzelnen Schülerinnen und Schüler, die Klassen und Kurse sondern auch für das Fach, den Jahrgang und die Schule abzuleiten. Eine derart intensive Auseinandersetzung mit dem Instrument und den Ergebnissen benötigt jedoch nicht nur Zeit, sondern auch Expertise zum Vorgehen (inhaltlich, aber auch konzeptionell im Sinne von Personalführung und Anleitung). Liegt diese Expertise an einer Schule nicht vor, oder sind die strukturellen Begebenheiten ungünstig, ist es sehr

schwierig, die Lernstandserhebungen für die Weiterentwicklung von Unterricht und schulischen Prozessen zu nutzen.

Rund 30% der Lehrkräfte nahmen eine neutrale Haltung ein. Die Gründe hierfür sind unklar. Es ist nicht auszuschließen, dass keine ausreichende Beschäftigung mit den Lernstandserhebungen stattgefunden hat und noch Informationsbedarf besteht. Möglich wäre auch, dass diese Lehrkräfte zu dem Schluss gekommen sind, die Sinnhaftigkeit der Maßnahme nicht eindeutig beurteilen zu können.

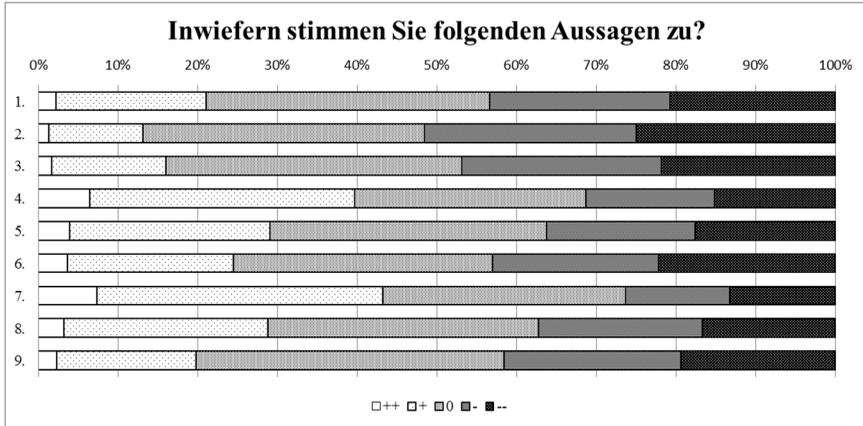


Abbildung 3: Einstellungen der Lehrkräfte zu den Lernstandserhebungen

Legende: 1. Die Schülerinnen und Schüler profitieren von der Durchführung der Lernstandserhebungen; 2. Von den Lernstandserhebungen sind wichtige Impulse für die pädagogische Diskussion in meiner Schule ausgegangen; 3. Die Lernstandserhebungen leisten einen Beitrag zur langfristigen Qualitätsentwicklung an meiner Schule; 4. Die landesweite Durchführung der Lernstandserhebungen nehme ich als eine Chance wahr, über den eigenen "Tellerrand" zu blicken; 5. Die Lernstandserhebungen nehme ich als Baustein in einem Gesamtkonzept zur Unterrichts- und Schulentwicklung wahr; 6. Ich halte die Lernstandserhebungen insgesamt für eine sinnvolle Maßnahme; 7. Die Lernstandserhebungen sind ein nützliches Instrument zur Lernstandsdiagnose; 8. Die Lernstandserhebungen sind ein nützliches Instrument für die Unterrichtsentwicklung; 9. Die Lernstandserhebungen sind ein nützliches Instrument für die Schulentwicklung.

Eine positive Einschätzung wird von ca. 30% der Befragten geteilt. Insbesondere die landesweite Durchführung der Lernstandserhebungen wird als Chance wahrgenommen, über den eigenen Tellerrand zu blicken (40% Zustimmung). Auch der Aussage, die Lernstandserhebungen seien ein nützliches Lernstandsdiagnose, konnten über 40% der Englischlehrkräfte zustimmen. Dieses an sich erfreuliche Ergebnis darf jedoch nicht darüber

hinwegtäuschen, dass die meisten Lehrkräfte offenbar an diesem Punkt stehen bleiben. Mit den Lernstandserhebungen erhalten die Lehrkräfte ein Stärken-Schwächen-Profil ihrer Lerngruppen und können durch den korrigierten Landesmittelwert erfahren, ob der Kompetenzstand der Gruppe vergleichsweise gut oder eher schwach ausgeprägt ist. Der nächste Schritt – weg von der Diagnose hin zur Förderung und durch die verstärkte Arbeit im Team auch zur Optimierung von schulischen Prozessen – wird häufig nicht gegangen.

4.4.4 Umgang mit den Ergebnissen

Bei der Auswertung der vorliegenden Fragebogendaten ist zu berücksichtigen, dass sich die Frage "Wie sind Sie mit den Ergebnissen umgegangen?" auf einen längeren Zeitraum, vom Zeitpunkt der Ergebniseingabe bis mindestens zum Schuljahresende, bezieht. Lehrkräfte, die zum ersten Mal an den Lernstandserhebungen teilnehmen, können zum Zeitpunkt der Beantwortung des Fragebogens nicht alle angegebenen Möglichkeiten des Umgangs vollzogen haben, da der Ergebnisbericht beispielsweise erst sechs Wochen später ausgegeben wird. Die dargestellten Antworten beziehen sich demnach auf Vorgehensweisen der vergangenen Jahre. Aus diesem Grund wurde die Frage auch nur den Lehrkräften vorgelegt, die bei der vorhergehenden Frage "Haben Sie persönlich im Schuljahr 2012/13 ebenfalls an den Lernstandserhebungen teilgenommen?" mit "Ja" geantwortet haben. Von den insgesamt 104 Antworten auf diese Variablen­gruppe waren 85 Lehrkräfte auch an den Lernstandserhebungen im Fach Englisch des vorhergehenden Jahres beteiligt. 29 Lehrkräfte nahmen in Deutsch bzw. Mathematik teil. Dass sich die Vorgehensweisen der Lehrkräfte nicht ausschließlich auf das Fach Englisch beziehen spielt hier nur eine sekundäre Rolle. Im Vordergrund stehen die Vertrautheit mit dem Verfahren und die Erfahrungen im Umgang mit den Berichten zu den Ergebnissen der Klasse. Die zwölf Items der Variablen­gruppe weisen ein Cronbach-Alpha von 0.88 auf.

Leistungserhebungen haben stets zweierlei Rückmeldefunktion: Einerseits erhalten die Lernenden (bzw. andere Personengruppen, wie Eltern) summative Informationen über den Kompetenzstand der Lerngruppe und andererseits dienen die Lernstandserhebungen Lehrkräften als formative Unterrichtsevaluation und darauf aufbauend als Anstoß für Qualitätsentwicklungsmaßnahmen (vgl. Maier et al 2013: 75).

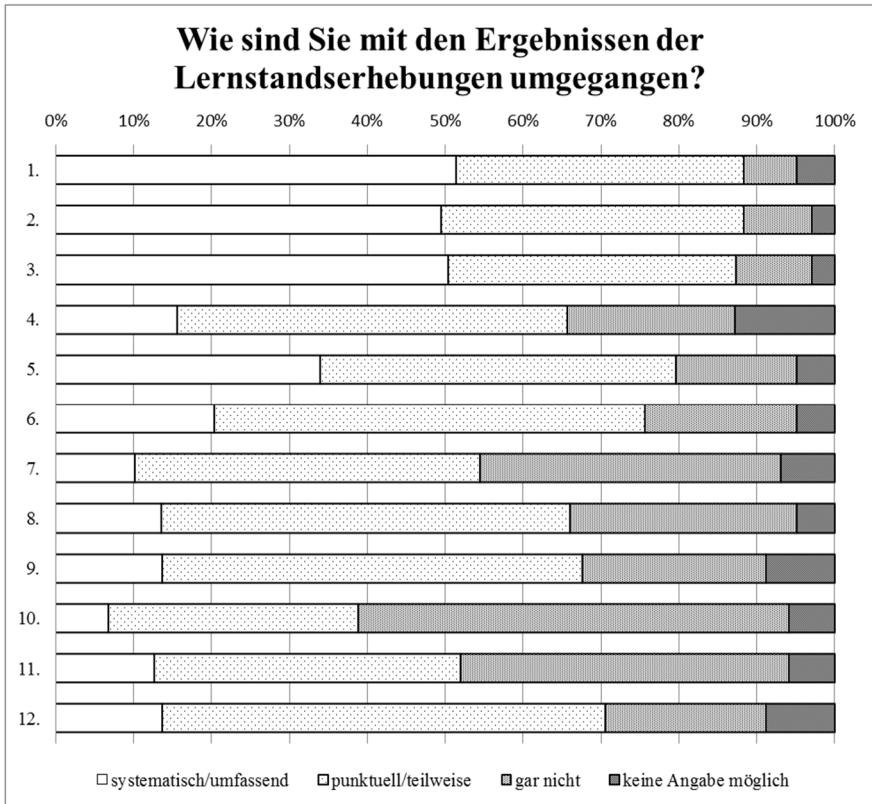


Abbildung 4: Durchsicht und Nutzung der Ergebnisberichte

Legende: 1. Durchsicht der Sofortrückmeldung; 2. Durchsicht des Ergebnisberichts; 3. Durchsicht der tabellarischen Auswertung; 4. Durchsicht des Schulberichts; 5. Besprechung der Ergebnisse mit der gesamten Klasse; 6. Besprechung der Ergebnisse mit einzelnen Schülerinnen und Schülern; 7. Besprechung der Ergebnisse mit den Eltern; 8. Besprechung der Ergebnisse im Klassen- bzw. Jahrgangsteam; 9. Besprechung der Ergebnisse in der Fachschaft; 10. Besprechung der Ergebnisse im gesamten Kollegium; 11. Besprechung der Ergebnisse mit der Schulleitung; 12. Vergleich der Ergebnisse (z.B. mit anderen Klassen, dem kor. Landesmittelwert,...)

Abbildung 4 zeigt, dass ca. 90% der Lehrerinnen und Lehrer die Ergebnisse ihrer Lerngruppen sowohl direkt nach den Lernstandserhebungen (tabellarische Auswertung und Sofortrückmeldung) zur Kenntnis nahmen, als auch später im Ergebnisbericht lasen. Die summativen Informationen zum Leistungsstand der Klasse in Bezug auf die getesteten Bereiche wurden also von der großen Mehrheit rezipiert. Der intendierte Mehrwert der Ergebnisse für formative Schul- und Unterrichtsentwicklungsprozesse fiel dagegen vergleichsweise

geringer aus. Deutlich weniger Lehrkräfte gaben an sich intensiv mit den Rückmeldungen zu beschäftigen. Am ehesten erfolgte dies noch auf Ebene der Klasse. Ca. 80% der Lehrkräfte besprachen die Klassenergebnisse mit der eigenen Lerngruppe. 35% der Lehrkräfte gaben sogar an, dies umfassend zu tun. Die Bereitschaft, die Ergebnisse mit einzelnen Schülerinnen und Schülern zu besprechen, fiel ähnlich hoch aus. Dies kann mit der Annahme zusammenhängen, die Testung sei ein nützliches Instrument zur Lernstandsdiagnose (siehe Abbildung 3). Weniger beliebt scheint die Kommunikation mit Eltern. Etwas über 50% der Lehrkräfte besprachen die Ergebnisse mit den Erziehungsberechtigten. Ca. 20 % der Lehrkräfte besprachen die Ergebnisse nicht weiter mit den Schülerinnen und Schülern bzw. ihrer Klassen. Der professionelle Austausch im Kollegium fand am häufigsten im Jahrgangsteam oder der Fachschaft statt (beides um die 70%), wenn auch in den meisten Fällen nur punktuell. Die Schulleitung wurde von etwa der Hälfte der Lehrerinnen und Lehrer in unterschiedlichem Ausmaß miteinbezogen. Ca. 30% der Lehrkräfte gaben an, keine weiteren Besprechungen der Ergebnisse in Klassen-, Jahrgangs- oder Fachteams durchzuführen. Das Vergleichen der Klassenergebnisse mit anderen Referenzklassen (z.B. dem korrigierten Landesmittelwert) wurde von der Mehrheit betrieben (ca. 70%). Allerdings zeigt sich auch hier deutlich, dass die Auseinandersetzung mit den Ergebnissen von den meisten nur eingeschränkt bzw. teilweise und punktuell erfolgte.

4.4.5 Nutzung der Lernstandserhebungen für die Schul- und Unterrichtsentwicklung

Es wird der Frage nachgegangen, was aus Sicht der Lehrkräfte notwendig wäre, um die Lernstandserhebungen für die Schul- und Unterrichtsentwicklung nutzbarer zu machen. Die sieben Variablen zu Möglichkeiten der Optimierung ergeben ein Cronbach-Alpha von 0.89. Es lagen 627 Antworten vor.

Den größten Entwicklungsbedarf sahen die Englischlehrkräfte in den Bereichen "Stärkere Verdeutlichung des Zusammenhangs von Lernstandserhebungen und Abschlussprüfungen" (ca. 62% Zustimmung) und "Unterstützung bei der Planung von Fördermaßnahmen" (ca. 60% Zustimmung) (vgl. Abbildung 5). Ca. 55% der Lehrkräfte wünschten sich mehr Informationen zur Weiterarbeit der Lernstandserhebungen im Unterricht und ca. 45% wünschten sich Hilfe bei der Interpretation. Dies verdeutlicht, dass die Lesbarkeit und Verständlichkeit der Formate weiter erhöht werden sollte, bzw. dass ein Teil der Lehrkräfte mit solchen Formaten noch nicht hinreichend vertraut ist.

Diesen Bedürfnissen versucht die Hessische Lehrkräfteakademie durch die Ausbildung von Multiplikatoren zu begegnen, die Schulen auch vor Ort zu den Lernstandserhebungen beraten können.

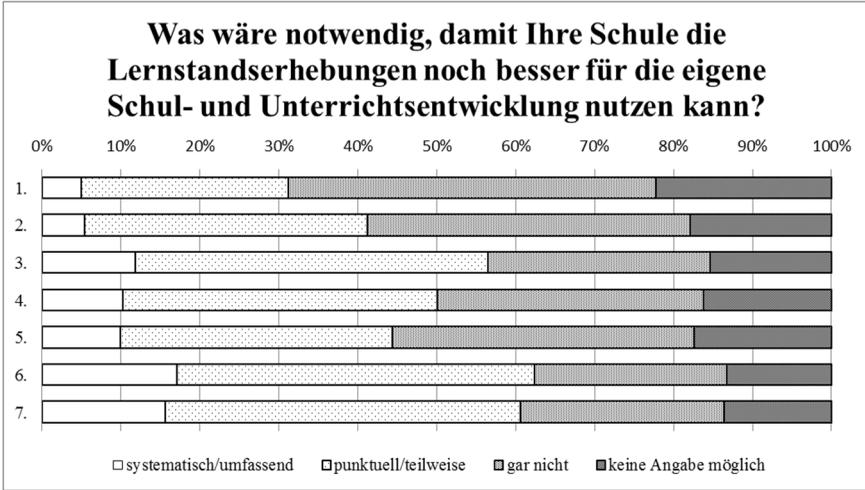


Abbildung 5: Optimierungswünsche der Lehrkräfte

Legende: 1. Austausch mit anderen Schulen; 2. Integration der Auswertung in die Schulabläufe (z.B. Analyse der Ergebnisse im Rahmen des päd. Tages); 3. Mehr Informationen zur Weiterarbeit mit den Lernstandserhebungen im Unterricht; 4. Verstärkte Kooperation in Jahrgangs-/Fachteams; 5. Hilfe bei der Interpretation der Ergebnisrückmeldungen; 6. Stärkere Verdeutlichung des Zusammenhangs von Lernstandserhebungen und Abschlussprüfungen; 7. Unterstützung bei der Planung von Fördermaßnahmen

Eher im schulstrukturellen Bereich angesiedelt sind die Wünsche nach mehr Austausch mit anderen Schulen (ca. 30%) und nach einer stärkeren Integration der Auswertung in die Schulabläufe (z.B. im Rahmen eines pädagogischen Tages) (ca. 40%). Der kollegiale Austausch in Jahrgangs- und Fachteams erscheint ca. 50% der Befragten als Optimierungsvorschlag sinnvoll. Die andere Hälfte hält eine verstärkte Kooperation nicht entscheidend für eine bessere Nutzung der Lernstandserhebungen für die eigene Schul- und Unterrichtsentwicklung. Dies bedeutet, dass dieser Teil der Lehrerschaft die stattfindende Kooperation als ausreichend empfindet oder die Testergebnisse nicht als Grund für den kollegialen Austausch angesehen werden.

Insgesamt zeigt sich, dass das Antwortverhalten als heterogen bezeichnet werden kann. Es scheint unterschiedliche Typen von Lehrerinnen und Lehrern zu geben, die mit dem Instrument unterschiedlich umgehen und daher auch unterschiedliche Bedürfnisse in Bezug auf seine Nutzung haben.

5. Schluss und Fazit

Der Aufwand im Rahmen der Durchführung und Auswertung der Lernstandserhebungen, der oft nicht auf freiwilliger Basis, sondern auf dem Verordnungswege auf sich genommen wird, mag mit verantwortlich sein für eine affektive Ablehnung des Instruments. Das eigentliche Ziel der Lernstandserhebungen, die Schul- und Unterrichtsentwicklung, scheint dabei nicht von allen Lehrkräften erkannt zu werden. Viele grundlegende Informationen zum Instrument sind noch unbekannt oder werden nicht wahrgenommen. Beispielsweise fließen die kriterialen Informationswerte der Ergebnisse häufig noch nicht in die unterrichtliche Weiterarbeit oder die Schulentwicklung ein. Vor allem Lehrerinnen und Lehrer in den sprachlichen Fächern zeigten "Schwierigkeiten beim Verständnis der Testdaten" (Kühle 2010: 76). Es zeigt sich zugleich aber auch eine Lehrerschaft, die dem Instrument positiv gegenübersteht und die Ergebnismeldungen für die Weiterentwicklung des eigenen Unterrichts nutzt. Wünschenswert wäre, wenn die Potentiale der Lernstandserhebungen in die Schulentwicklung einfließen würden. Das Potential beruht vor allem auf dem kollegialen Austausch und der gemeinsamen Diskussion der Ergebnisse.

Grundvoraussetzung dafür ist eine veränderte Wahrnehmung der Lehrerrolle, die eine Bereitschaft voraussetzt, das eigene Lehrverhalten und den Unterricht mit anderen zu reflektieren und auch zu evaluieren. Standardisierte Tests haben hierzulande keine lange Tradition, nahm Deutschland beispielsweise bis Ende der 1980er Jahre an keiner internationalen Schulleistungstudie teil (vgl. Korngiebel 2014: 9). Die Akzeptanz einer notwendigen objektiven Rückmeldung, z.B. für die Schülerdiagnostik und vereinzelt auch über die Qualität der eigenen Arbeit, erhöht sich aber zunehmend auch in der hessischen Lehrerschaft. So werden subjektive Feedbacks der Schülerinnen und Schüler, Kolleginnen und Kollegen oder die ermittelte Durchschnittsnote in den Klassenarbeiten immer häufiger auch durch standardisierte und zum Teil auch externe Verfahren ergänzt (z.B. Quop, <http://www.quop.de/>).

Kompetenzorientierung im Fremdsprachenunterricht geht einher mit einer veränderten fachlich-didaktischen und pädagogischen Einstellung, in der nicht der "Stoff" vor dem Lernprozess des Schülers stehen darf. Die VERA-Ergebnisse bieten einen guten Anlass, um die Lernprozesse zu evaluieren und eine Anschlussdiskussion nahe an relevanten fachdidaktischen Erkenntnissen führen zu können. Es ist unbestritten, dass für die erfolgreiche Implementation von Schulleistungstests zur internen Evaluation der Schulen (KMK 2012) gewisse Rahmenbedingungen herrschen müssen. Ein ausgewogenes Kosten-

Nutzen-Verhältnis kann sich nicht einstellen, wenn einerseits die Bedürfnislage seitens der Lehrkräfte groß ist, andererseits jedoch grundlegende Informationen über das Nutzungspotential der Lernstandserhebungen in Schulen und Fachschaften noch nicht angekommen sind. So wird auf Steuerungsebene die Notwendigkeit einer strukturierten und flächendeckenden Unterstützung der Lehrkräfte in Form von Fortbildungen sichtbar. Dazu gehören auch die Unterstützung der Lehrkräfte durch eine möglichst gründliche Vorbereitung im Vorfeld, Unterstützung während der Durchführungsphase und eine effiziente Begleitung bei der Datenanalyse und Interpretation, um die erkannten Stärken und Schwächen im Unterricht optimal aufzugreifen und die bestmögliche Förderung für die Schülerinnen und Schüler gewährleisten zu können. Dem wahrnehmbaren Bedürfnis der Lehrkräfte, Unterstützung bei der Auswertung der Ergebnisse zu erhalten, wird in Hessen durch den Einsatz von Multiplikatoren im Rahmen von Schulberatung und Fortbildung begegnet.

Zusammenfassend kann gesagt werden, dass mit der Kompetenzorientierung ein Prozess angestoßen wurde, der alle Akteure im Umfeld Schule, besonders auch der Lehrkräfte und der Schulleitungen, zur Weiterarbeit auffordert. De Florio-Hansen (2008: 35) erklärt: "Wir stehen, was ergebnisorientiertes Fremdsprachenlernen und dessen externe Überprüfung angeht, erst am Anfang einer langen Entwicklung." Eine Zusammenarbeit der empirischen und der didaktischen Wissenschaften an der Weiterentwicklung der Standards und der Leistungsmessung erscheint wünschenswert, gerade mit Blick auf das gemeinsame Ziel, alle Kinder und Jugendlichen durch guten Unterricht zu eigenverantwortlichen Individuen zu erziehen und sie zu erfolgreichem (Sprach-) Handeln zu befähigen. Leupold unterstreicht, wie wichtig es für renommierte Fremdsprachendidaktiker wäre, die Perpetuierung längst bekannter und oft wiederholter Einwände gegen die Bildungsstandards einzustellen, sich den aus den veränderten Strukturen erwachsenen fachdidaktischen Aufgaben und Herausforderungen zu stellen und den Weg eines konstruktiven Dialogs mit der empirischen Bildungsforschung einzuschlagen (vgl. Leupold 2010: 29).

Und so fällt die Bilanz Mut machend aus. Wenngleich nur ca. ein Drittel der Lehrkräfte angibt, vom Nutzen der Lernstandserhebungen überzeugt zu sein, geben ca. zwei Drittel der Lehrerinnen und Lehrer an, mit den Ergebnisberichten zu arbeiten (ca. 60 – 80%, je nach Bericht). Auch die Angaben dazu, was notwendig für eine bessere Nutzung der Lernstandserhebungen an der eigenen Schule wäre, lässt darauf schließen, dass die Kollegien dem Instru-

ment nicht grundsätzlich negativ gegenüber stehen, aber noch hohen Informations- und Beratungsbedarf haben. Hier anzusetzen, stellt die Aufgabe und die Herausforderung für die nächsten Jahre dar.

Eingang des revidierten Manuskripts 25.05.2015

Literaturverzeichnis

- Baumert, Jürgen & Köller, Olaf (1998), Nationale und internationale Schulleistungsstudien: Was können sie leisten, wo sind ihre Grenzen? *Pädagogik* 50, 12-18.
- Bausch, Karl-Richard; Burwitz-Melzer, Eva; Königs, Frank Gerhard & Krumm, Hans-Jürgen (Hrsg.) (2009), *Fremdsprachenunterricht im Spannungsfeld von Inhaltsorientierung und Kompetenzbestimmung*. Tübingen: Narr.
- Böttcher, Wolfgang (2013), Das Monitoring-Paradigma - Eine Kritik der deutschen Schulreform. In: Zimmer-Müller, Michael & Hosenfeld, Ingmar (Hrsg.) (2013), *Zehn Jahre Vergleichsarbeiten: Eine Zwischenbilanz aus verschiedenen Perspektiven*. *Empirische Pädagogik* 27(4), 497-509.
- Bortz, Jürgen; Döring, Nicola (2006), *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Berlin, Heidelberg: Springer.
- Bredella, Lothar (2009), Die zentrale Bedeutung von Inhalten für Bildungsziele, Lernstrategien und Kompetenzen. In: Bausch, Karl-Richard; Burwitz-Melzer, Eva; Königs, Frank Gerhard & Krumm, Hans-Jürgen (Hrsg.) (2009), 25-33.
- Breidbach, Stephan (2008), Fremdsprachliche Kompetenzen jenseits der Standardisierbarkeit. Warum mit den „Bildungsstandards“ die Probleme für eine bildungsorientierte Fremdsprachendidaktik erst anfangen. In: Lüger, Heinz-Helmut & Rössler, Andrea (Hrsg.) (2008), *Wozu Bildungsstandards? Zwischen Input- und Outputorientierung in der Fremdsprachenvermittlung*. Beiträge zur Fremdsprachenvermittlung, Sonderheft 13. Landau: *Empirische Pädagogik*, 117-133.
- Burwitz-Melzer, Eva (2009), Inhalte und Kompetenzen im Fremdsprachenunterricht. In: Bausch, Karl-Richard; Burwitz-Melzer, Eva; Königs, Frank Gerhard & Krumm, Hans-Jürgen (Hrsg.) (2009), 34-42.
- Caspari, Daniela (2010), Französischunterricht in Deutschland. Aktuelle Situation und Zukunftsperspektiven. In: Porsch, Raphaela; Tesch, Bernd & Köller, Olaf (Hrsg.) (2010), 11-24.
- Caspari, Daniela; Grotjahn, Rüdiger & Kleppin, Karin (2010), Testaufgaben und Lernaufgaben. In: Porsch, Raphaela; Tesch, Bernd & Köller, Olaf (Hrsg.) (2010), 46-68.
- Caspari, Daniela; Kötter, Markus; Rossa, Henning; Schramm, Karen; Tesch, Bernd; Vollmer, Helmut-Johannes & Zydaniß, Wolfgang (2012), Mindeststandards für Fremdsprachen am Ende der Pflichtschulzeit. Ein Positionspapier der Deutschen Gesellschaft für Fremdsprachenforschung (DGFF). *Zeitschrift für Fremdsprachenforschung* 23: 2, 243-268.
- Council of Europe (2001), *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Cronbach, Lee J. (1951), Coefficient alpha and the internal structure of tests. *Psychometrika* 16, 297-333.

- De Florio-Hansen, Inez (2008), Wer hat Angst vor Bildungsstandards? Überlegungen zur Kompetenz-, Aufgaben- und Inhaltsorientierung im Fremdsprachenunterricht. In: Lüger, Heinz-Helmut & Rössler, Andrea (Hrsg.) (2008), *Wozu Bildungsstandards? Zwischen Input- und Outputorientierung in der Fremdsprachenvermittlung*. Beiträge zur Fremdsprachenvermittlung, Sonderheft 13. Landau: *Empirische Pädagogik*, 59-86.
- DESI-Konsortium (Hrsg.) (2006), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Zentrale Befunde der Studie Deutsch-Englisch-Schülerleitungen-International (DESI)*. [Online: <http://www.dipf.de/de/pdf-dokumente/projekte-materialien/desi-zentrale-befunde>, 21.01.2015].
- Deutsche Gesellschaft für Fremdsprachenforschung (2008), Positionspapier für Vorstand und Beirat (Koordination: Hu, Adelheid, Mitarbeit: Caspari, Daniela; Grünewald, Andreas; Hu, Adelheid; Küster, Lutz; Nold, Günter; Vollmer, Helmut Johannes & Zydatis, Wolfgang). Kompetenzorientierung, Bildungsstandards und fremdsprachliches Lernen – Herausforderungen an die Fremdsprachenforschung. *Zeitschrift für Fremdsprachenforschung* 19 (2), 163-186.
- Drüke-Noe, Christina (2014), *Aufgabenkultur in Klassenarbeiten im Fach Mathematik: Empirische Untersuchungen in neunten und zehnten Klassen*. Wiesbaden: Springer Spektrum.
- Finkbeiner, Claudia (2011), Assessment and Testing and The Issue of Equity. In: Maria Eisenmann & Theresa Summer (Hrsg.) (2011), *Basic Issues in EFL- Teaching and Learning*. Heidelberg: Winter, 419-447.
- Finkbeiner, Claudia; Olson, Agnes Madeleine & Friedrich, Jennifer (2013), Foreign language learning and teaching in Germany: A review of empirical research literature from 2005 to 2010. *Language Teaching* 46 (4), 477-510.
- Grotjahn, Rüdiger (2008), Tests und Testaufgaben: Merkmale und Gütekriterien. In: Tesch, Bernd; Leupold, Eynar & Köller, Olaf (Hrsg.) (2008), *Bildungsstandards Französisch: konkret*. Berlin: Cornelsen Verlag Scriptor, 149-186.
- Grotjahn, Rüdiger (2009a), Testen im Fremdsprachenunterricht. Aspekte der Qualitätsentwicklung. *Praxis Fremdsprachenunterricht* 1, 4-8.
- Grotjahn, Rüdiger (2009b), Prüfen-Testen-Bewerten. In: Udo O. H. Jung (Hrsg.) (2009), *Praktische Handreichung für Fremdsprachenlehrer* (5. durchgesehene Aufl.). Frankfurt: Peter Lang, 221-230.
- Hallet, Wolfgang (2012), Die komplexe Kompetenzaufgabe: Fremdsprachige Diskursfähigkeit als kulturelle Teilhabe und Unterrichtspraxis. In Hallet, Wolfgang & Krämer, Ulrich (Hrsg.) (2012), *Kompetenzaufgaben im Englischunterricht. Grundlagen und Unterrichtsbeispiele*. Seelze: Klett/Kallmeyer, 8-19.
- Harsch, Claudia (2009), Externe Evaluation. Verhältnis zum fremdsprachlichen Unterricht. *Praxis Fremdsprachenunterricht* 1, 9-14.
- Haß, Frank (2012), Fertigkeitentwicklung im kompetenzorientierten Englischunterricht. In: Hallet, Wolfgang & Krämer, Ulrich (Hrsg.) (2012), *Kompetenzaufgaben im Englischunterricht. Grundlagen und Unterrichtsbeispiele*. Seelze: Klett/Kallmeyer, 20-29.
- Hellrung, Katrin & Nachtigall, Christof (2013), Zur zeitlichen Entwicklung der Rezeption von Vergleichsarbeiten. In: Zimmer-Müller, Michael & Hosenfeld, Ingmar (Hrsg.) (2013), *Zehn Jahre Vergleichsarbeiten: Eine Zwischenbilanz aus verschiedenen Perspektiven*. *Empirische Pädagogik* 27(4), 423-441.
- Hessisches Kultusministerium (2011), *Hessisches Amtsblatt*. Lernstandserhebungen in den Jahrgängen 3, 6 und 8 ab dem Schuljahr 2011/12. 8/2011, 409-410.
- Hessisches Kultusministerium (2011), *Hessisches Amtsblatt*, Verordnung zur Gestaltung des Schulverhältnisses. 9/2011, 546-581.

- Hu, Adelheid & Leupold, Eynar (2008), Kompetenzorientierung und Französischunterricht. In: Tesch, Bernd; Leupold, Eynar & Köller, Olaf (Hrsg.) (2008), *Bildungsstandards Französisch: konkret*. Berlin: Cornelsen Verlag Scriptor, 51-84.
- Isaac, Kevin (2013), Kriteriale Bezugsnormen bei Lernstandserhebungen – Bilanz und Perspektiven. In: Zimmer-Müller, Michael & Hosenfeld, Ingmar (Hrsg.) (2013), *Zehn Jahre Vergleichsarbeiten: Eine Zwischenbilanz aus verschiedenen Perspektiven. Empirische Pädagogik* 27(4), 407-422.
- Klippel, Friederike (2009), Ziele, Inhalte und Kompetenzen – Eckpunkte der Entwicklung des Fremdsprachenunterrichts. In: Bausch, Karl-Richard; Burwitz-Melzer, Eva; Königs, Frank Gerhard & Krumm, Hans-Jürgen (Hrsg.) (2009), 88-95.
- Korngiebel, Julia (2014), *Vergleichsarbeiten und ihr Potenzial für die Schul- und Unterrichtsentwicklung. Eine qualitative Untersuchung zur Nutzung der Lernstandserhebungen an hessischen Gymnasien*. Marburg: Philipps Universität Marburg.
- Kühle, Ben (2010), *Zentrale Lernstandserhebungen – Ergebnisorientierte Unterrichtsentwicklung? Schulische Strategien beim Umgang mit Ergebnissen aus den Schulrückmeldungen im Kontext der ersten Lernstandserhebungen 2004/2005 in Nordrhein-Westfalen*. Berlin: Köster.
- Kuhn, Hans-Jürgen (2014), Anspruch, Wirklichkeit und Perspektiven der Gesamtstrategie der KMK zum Bildungsmonitoring. *Die Deutsche Schule* 106 (4), 414-426.
- Küster, Lutz (2009), Gegen den Geist der Vereinheitlichung. Perspektiven einer Verbindung von Kompetenz- und Inhaltsorientierung schulischen Fremdsprachenunterrichts. In: Bausch, Karl-Richard; Burwitz-Melzer, Eva; Königs, Frank Gerhard & Krumm, Hans-Jürgen (Hrsg.) (2009), 113-121.
- Leupold, Eynar (2010), Länderübergreifende Bildungsstandards Französisch im Kontext von Bildungspolitik, Fachdidaktik und Französischunterricht. In: Porsch, Raphaela; Tesch, Bernd & Köller, Olaf (Hrsg.) (2010), 25-45.
- Leupold, Eynar & Porsch, Raphaela (2011), Von der Rezeption zur Handlungskompetenz: Bildungsstandards und ihre Konsequenzen für unterrichtliches Handeln. *Zeitschrift für Fremdsprachenforschung* 22 (1), 85-111.
- Maier, Uwe; Ramsteck, Carolin & Frühwacht, Annette (2013), Lehr-lerntheoretische Argumentationsmuster bei der Interpretation und Nutzung von Vergleichsarbeitsrückmeldungen durch Gymnasiallehrkräfte. *Die Deutsche Schule*, 12. Beiheft, 74-96.
- Maritzen, Norbert (2014), Glanz und Elend der KMK-Strategie zum Bildungsmonitoring. Versuch einer Bilanz und eines Ausblicks. *Die Deutsche Schule* 106 (4), 398-413.
- Müller-Hartmann, Andreas; Schocker, Marita & Pant, Hans Anand (2013), *Kompetenzentwicklung in der Sek. I. Lernaufgaben Englisch aus der Praxis*. Braunschweig: Diesterweg.
- Nachtigall, Christof; Storbeck, Isabell & Landmann, Meike (2009), Zur Nutzung von Vergleichsarbeiten, Lernstandserhebungen, Kompetenztests, Orientierungsarbeiten und Co. *Schulleitung und Schulentwicklung* 45, 1-17.
- Nieweler, Andreas (2010), Sprachtests: Formen und Funktion. In: Hallet, Wolfgang & Königs, Frank. G. (Hrsg.) (2010), *Handbuch Fremdsprachendidaktik*. Seelze-Velber: Klett/Kallmeyer, 219-223.
- Porsch, Raphaela; Tesch, Bernd & Köller, Olaf (Hrsg.) (2010), *Standardbasierte Testentwicklung und Leistungsmessung. Französisch in der Sekundarstufe*. Münster: Waxmann.
- Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (KMK) (2004), *Bildungsstandards im Fach Deutsch für den Mittleren Schulabschluss*. Beschluss vom 4.12.2003, München: Wolters Kluwer.

- Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (KMK) (2004), Beschlüsse der Kultusministerkonferenz. *Bildungsstandards für die erste Fremdsprache (Englisch/Französisch) für den Mittleren Schulabschluss*. Beschluss vom 04.12.2003. München: Wolters Kluwer.
- Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (KMK) (2005), Beschlüsse der Kultusministerkonferenz. *Bildungsstandards für die erste Fremdsprache (Englisch/Französisch) für den Hauptschulabschluss*. Beschluss vom 15.10.2004. München: Wolters Kluwer.
- Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (KMK) (2006), *Überblick: Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring* [Online: http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2006/2006_08_01-Gesamtstrategie-Bildungsmonitoring.pdf, 02.08.15].
- Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (KMK) (2012), *Vereinbarung zur Weiterentwicklung von VERA (Beschluss der Kultusministerkonferenz vom 08.03.2012)* [Online: http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2012/2012_03_08_Weiterentwicklung-VERA.pdf, 05.11.2014].
- Steininger, Ivo (2014), *Modellierung literarischer Kompetenz. Eine qualitative Studie im Fremdsprachenunterricht der Sekundarstufe I*. Tübingen: Narr.
- Tesch, Bernd (2009), Die Tests des IQB – Sinn und Nutzen. *Praxis Fremdsprachenunterricht* 1, 41-45.
- Tesch, Bernd (2010), *Kompetenzorientierte Lernaufgaben im Fremdsprachenunterricht. Konzeptionelle Grundlagen und eine rekonstruktive Fallstudie zur Unterrichtspraxis (Französisch)*. Frankfurt am Main: Peter Lang.
- Vogel, Sebastian; Achmetli, Kay; Krawitz, Janina & Blum, Werner (2014), Wie können die Lernstandserhebungen in Klasse 8 effektiv genutzt werden? Evaluation des Projekts VELM-8. *Beiträge zum Mathematikunterricht* (2), 1251-1254.
- Weskamp, Ralf (2007), Assessment for Learning – Leistungsbeurteilung im Klassenzimmer. In: Edelhoff, Christoph (Hrsg.) (2007), *Lernen und Leisten im Fremdsprachenunterricht. Beiträge zur Qualitätssicherung*. Braunschweig: Schrödel/Diesterweg, 49-58.
- Winter, Felix (2015), *Lerndialog statt Noten. Neue Formen der Leistungsbeurteilung*. Weinheim und Basel: Beltz.
- Wurster, Sebastian; Richter, Dirk; Schliesing, Anna & Pant, Hans Anand (2013), Nutzung unterschiedlicher Evaluationsdaten an Berliner und Brandenburger Schulen. Rezeption und Nutzung von Ergebnissen aus Schulinspektion, Vergleichsarbeiten und interner Evaluation im Vergleich. *Die Deutsche Schule* 12. Beiheft, 97-120.
- Zimmer-Müller, Michael & Hosenfeld, Ingmar (2013), Zehn Jahre Vergleichsarbeiten: eine Zwischenbilanz aus verschiedenen Perspektiven. In: Zimmer-Müller, Michael & Hosenfeld, Ingmar (Hrsg.) (2013), *Zehn Jahre Vergleichsarbeiten: Eine Zwischenbilanz aus verschiedenen Perspektiven*. *Empirische Pädagogik* 27(4), 497-509.