

# Wie verändern Bildungsstandards und zentrale Prüfungen den Fremdsprachenunterricht? Skizze eines Forschungsdesiderats zu intendierten und beobachteten Effekten der Standard- und Kompetenzorientierung

Henning Rossa<sup>1</sup>

At the beginning of the new millennium, the introduction of national educational standards, systematic assessments of learning outcomes and increasingly centralised exit exam regimes in the German federal states was prompted by the perceived need to improve the effectiveness of classroom teaching and educational outcomes in the wake of international large-scale assessment studies. The article discusses the criticism this reform has invited from scientific disciplines traditionally concerned with the contexts, conditions and processes of teaching and learning. It reviews empirical studies on the impact of central exit exams in the German context as well as international *washback* studies in the language testing research literature. In conclusion, a case is made for more explorative research in the field to support a better understanding of the ways educational standards, standardised assessment and central examinations are changing the practice of foreign language teaching and learning.

## 1. Einleitung

Die Implementation nationaler Bildungsstandards für den Hauptschulabschluss, den Mittleren Schulabschluss und die Allgemeine Hochschulreife (KMK 2003, 2004, 2012) wird wiederholt mit der politischen Forderung verknüpft, schulische Bildung in den Fremdsprachen möge sich an Kompetenzen orientieren und Lernergebnisse durch zentral gestellte Prüfungsaufgaben transparent und vergleichbar machen. Seit nunmehr einem Jahrzehnt werden die curricular spezifizierten Erwartungen an die Kompetenzentwicklung der Lernenden durch Test- und Prüfungsaufgaben operationalisiert und illustriert, die wiederum mit Blick auf Kompetenzmodelle skaliert, kalibriert und normiert werden. Diese Anhäufung messtechnischer Begriffe impliziert nur scheinbar, dass es sich bei dieser bildungspolitischen Reform um den objektiv-distanzierten Versuch handelt, Lerner-

---

1 Korrespondenzadresse: Dr. Henning Rossa, Vertr.-Prof., Fachdidaktik Englisch/Angewandte Linguistik, Institut für Anglistik und Amerikanistik, Fakultät Kulturwissenschaften, Technische Universität Dortmund, 44221 Dortmund, E-Mail: [henning.rossa@tu-dortmund.de](mailto:henning.rossa@tu-dortmund.de), Office: Emil-Figge-Straße 50, R. 3.213, Tel: +49 (0)231 755-2958, Fax: +49 (0)231 755-5450

gebnisse durch Vergleichsarbeiten, Lernstandserhebungen und zentrale Prüfungen zu erfassen. Tatsächlich ist aber eine tiefgreifende Veränderung der Unterrichtspraxis beabsichtigt, die mit der Hoffnung verknüpft ist, die Qualität – gemeint ist die "Wirksamkeit" – des Unterrichts zu steigern. In einem System, das sich selbst auf die Begriffe "Rechenschaftslegung" und "Monitoring" beruft, muss diese Absicht zwangsläufig selbst zum Gegenstand einer systematischen Überprüfung werden. Noch gibt es aber zumindest mit Blick auf den Fremdsprachenunterricht keine einschlägigen Befunde zur Frage, in welcher Weise die Implementation der Bildungsstandards und die Einführung zentraler Prüfungen auf die Unterrichtspraxis und seine Akteure wirken. Der hier vorliegende Beitrag will daher ein Forschungsdesiderat skizzieren, das Aguado (2009: 13) als Frage formuliert: "Wie verändert die Kompetenzorientierung den Fremdsprachenunterricht?"

Die Annäherung an diese Frage soll in drei Schritten gelingen. Erstens werden die intendierten Effekte von Bildungsstandards und zentralen Prüfungen in Deutschland und ihre bildungswissenschaftliche und fremdsprachendidaktische Rezeption diskutiert sowie bislang vorliegende empirische Erkenntnisse zur Wirksamkeit zentraler Prüfungen untersucht. Zweitens soll ein Überblick über forschungsmethodische Ansätze aus der internationalen (Sprachtest-)Forschung zum Konstrukt *washback* exemplarisch aufzeigen, wie die Fremdsprachenforschung ihre kritische Begleitung der Entwicklung des Fremdsprachenunterrichts in Zeiten der Standard- und Kompetenzorientierung empirisch stützen könnte. Drittens werden vorläufige inhaltliche Kategorien zur wahrgenommenen Wirkung zentraler Prüfungen auf die Alltagspraxis des Englischunterrichts vorgestellt, die den Daten einer informellen Befragung von zehn Lehrkräften entnommen wurden. Die Beobachtungen und Einschätzungen der Informanten sollen in Bezug auf die Forschungsfrage "Wie nehmen Lehrkräfte die Auswirkungen von zentralen Prüfungen auf den Englischunterricht wahr?" exemplarisch andeuten, welche Aspekte einer als verändert erlebten Alltagspraxis für zukünftige Forschungsvorhaben möglicherweise relevante Anknüpfungspunkte darstellen können.

## **2. Bildungsstandards und zentrale Prüfungen: intendierte Funktionen und Forschungsergebnisse zu ihrer Wirksamkeit**

### **2.1 Die bildungspolitische Perspektive: *Good intentions* oder der zweite Schritt vor dem ersten?**

Die Bildungsstandards werden als Mittel beschrieben, mit deren Hilfe Maßstäbe für die "Entwicklung und die Sicherung von Qualität" (KMK 2005: 5) schulischer Bildung gesetzt werden können. Diese grundlegende Doppelfunktion der Standards

und die Form ihrer Umsetzung als "externe und interne Evaluation" werden als notwendige Folgen aus den Ergebnissen internationaler Schulleistungsstudien abgeleitet. Demnach zeigen die in TIMSS, PISA und IGLU erfassten, vergleichsweise enttäuschenden Leistungen der Schülerinnen und Schüler an Schulen in Deutschland, dass die "Wirksamkeit unseres Bildungssystems regelmäßig und systematisch überprüft" (ibid.) werden müsse. Als Beleg für diese Forderung wird darauf verwiesen, "dass Staaten, in denen eine systematische Rechenschaftslegung über die Ergebnisse" (ibid.) schulischer Bildungsprozesse erfolgt, höhere Leistungen erzielen.

Die Konzeption der Bildungsstandards bezieht sich demnach, wie die Expertise "Zur Entwicklung nationaler Bildungsstandards" (Bundesministerium für Bildung und Forschung 2003), maßgeblich auf Schülerleistungen in mathematisch-naturwissenschaftlichen Fächern und im Lesen. Die Ergebnisse der DESI-Studie zum Kompetenzerwerb im Fach Englisch (DESI-Konsortium 2008) lagen zum Zeitpunkt dieser Publikation der KMK zur Konzeption und Entwicklung der Bildungsstandards noch nicht vor. Stattdessen verweisen die Bildungsstandards für die erste Fremdsprache auf den Gemeinsamen europäischen Referenzrahmen für Sprachen (Europarat 2001), der Mehrsprachigkeit als Bildungsziel und Teil kultureller Literalität definiert.

Der Bezug zum GeR äußert sich insbesondere darin, dass unter dem Schlagwort "Kompetenzorientierung", anstelle von inhaltlichen und methodischen Vorgaben, Kompetenzbeschreibungen als zentrale Orientierungspunkte für den Unterricht verankert wurden. Die Einführung der Bildungsstandards wurde so einerseits als eine Reform curricularer Vorgaben dargestellt, die zusammen mit standardkompatiblen Lehr- und Rahmenplänen helfen sollen, die Unterrichtspraxis weiter zu entwickeln. Andererseits sind die Bildungsstandards von Anfang an explizit mit der Absicht verknüpft, die Lernergebnisse in zentral organisierten Leistungsmessungen, "landesweit und länderübergreifend" zu überprüfen, zu vergleichen und für das "Bildungsmonitoring" zu nutzen (KMK 2005: 20).

Diese im Wortsinne selbstverständliche Verknüpfung verbindlicher Zielvorgaben mit systematischen, zentral durchgeführten Kompetenzmessungen hat dem bildungspolitischen Projekt der Standard- und Kompetenzorientierung auf breiter Front Kritik eingebracht. Es wird grundsätzlich bezweifelt, ob der vornehmliche Zweck der Standards, "die Qualität schulischer Bildung [...] zu sichern" (KMK 2005: 5) und zu verbessern, durch eine zentral gesteuerte, regelmäßige Überprüfung von Kompetenzen, d.h. durch testbasierten Rechenschaftsdruck auf Schülerleistungen überhaupt erfüllt werden kann (vgl. Maier 2010: 120).

Verschiedene Interessensgruppen im Kontext schulischer Bildung beklagen, dass die Funktion der Bildungsstandards für die Entwicklung von Kompetenzen hinter den Bemühungen um die Entwicklung von Testinstrumenten zur Überprüfung der Standards (Porsch, Tesch & Köller 2010; Rupp, Vock, Harsch & Köller 2008) zurückbleibt. Dieses Missverhältnis ist hauptsächlich durch die strukturelle Besonderheit bedingt, dass die Bundesländer die Implementation der nationalen Standards sehr unterschiedlich verfolgen, während die Entwicklung und Überprüfung der Standards zentral durch das Institut zur Qualitätsentwicklung im Bildungswesen verantwortet wird. So wird Schul- und Unterrichtsqualität – operationalisiert als Schülerleistung – gemessen, ohne dass vorher entsprechende Konzepte zur Schul- und Unterrichtsentwicklung ausgearbeitet und umgesetzt wurden, die hätten aufzeigen können, wie die Kompetenzentwicklung der Lernenden im Sinne der Standards konkret gefördert werden kann.

Doch obwohl die Herausgeber der Bildungsstandards betonen, "dass Qualität von Schule und Bildung immer mehr sein muss als die Erhebung von erworbenen Kompetenzen" (KMK 2010: 5), werden mittlerweile zentrale Prüfungen und andere Formen der Leistungsmessung von VERA 3/6/8 über Kompetenztests 6 und Zentrale Prüfungen 10 bis zum Zentralabitur (ab 2017) flächendeckend eingesetzt (vgl. Caspari 2014; Zimmer-Müller & Hosenfeld 2013), während eher lokale bis regionale Programme zur Unterrichtsentwicklung nur langsam vorankommen (vgl. Köller 2010: 546). Es verwundert daher kaum, dass der Ton der kritischen Analysen und Kommentare bisweilen eine ironische und abwertende Qualität annimmt, wenn die Reformvorhaben ironisch als "Welle der Kompetenzeuphorie" (Bär 2013: 98) karikiert oder als pathologische "Testeritis" (Teppe 2014: 48) gekennzeichnet werden.

## **2.2 Die pädagogisch-didaktische Perspektive auf die Standard- und Kompetenzorientierung**

Die pädagogische und (fremdsprachen-)didaktische Diskussion dieser bildungspolitischen Entwicklung ist ebenfalls von einer deutlich skeptischen Haltung geprägt (vgl. Bausch, Burwitz-Melzer, Königs & Krumm 2005, 2009; Bredella 2006; Brügelmann 2015; Christ 2006; Quetz & Vogt 2009, Quetz 2012). Diese Skepsis manifestiert sich in der Behauptung, standardorientierte curriculare Vorgaben, denen ein funktionaler Bildungsbegriff zugrunde liegt, führten zu einer Ökonomisierung schulischer Bildung und zu einer Reduktion unterrichtlicher Inhalte und Lernprozesse auf die Bewältigung standardisierter Situationen, wie sie in zentralen Tests und Prüfungen simuliert werden. Die potentiell destruktive

Wirkung von standardisierten Tests auf schulische Bildungsprozesse wurde ähnlich bereits von Madaus im Kontext der *high-stakes tests* in den USA formuliert:

[T]ests can become the ferocious master of the educational process, not the compliant servant they should be. Measurement-driven instruction invariably leads to cramming; narrows the curriculum; [...] constrains the creativity and spontaneity of teachers and students; and finally, demeans the professional judgement of teachers (Madaus 1988: 85).

So entsteht das Szenario eines standardisierten, inhaltsleeren Unterrichts, der nicht als erstrebenswerte Weiterentwicklung, sondern vielmehr als "enthumanisierender Irrweg zur Verzweckung des Menschen" (Matthes 2013: 121) erscheint (vgl. Gruschka, Lin-Klitzing, DiFuccia & Müller-Frerich 2013).

Bonnet & Breidbach (2013) argumentieren, dass der den Standards zugrunde gelegte Kompetenzbegriff in seiner kognitiven, an der Entwicklung von Testverfahren orientierten Ausprägung mit der Komplexität der Gegenstände und Ziele fremdsprachlicher Bildung nicht kompatibel sei und auf diese Weise dem "im Kompetenzdiskurs selbst formulierten Anspruch der Handlungsorientierung, des anwendungsfähigen Wissens und ganzheitlichen Könnens" (ibid.: 27) nicht gerecht werde. Die Autoren beziehen sich in ihrem Beitrag auf das literarisch-ästhetische Lernen im Fremdsprachenunterricht und bezweifeln, ob diese Dimension fremdsprachlicher Bildung "überhaupt mit den Konzepten der Standardisierung, Normierung und Testung vereinbar ist" (ibid.: 34).

Wenn sich in den Bildungsstandards nur die Deskriptoren der messbaren, funktional-kommunikativen Kompetenzen für eine Operationalisierung in Prüfungs- und Lernaufgaben eignen, würden die "persönlichkeitsbildenden und kreativ-ästhetischen Domänen des Fremdsprachenunterrichts" (Eberhardt 2013: 61) und gleichsam auch die "Individualität und Subjektivität der Lernenden" (Fäcke 2013: 43) in den standardbasierten Curricula und in der Folge auch in der Unterrichtspraxis vernachlässigt. Die Kritik an der inhaltlichen Engführung der Bildungsstandards wurde offenbar zumindest im Rahmen der Entwicklung der Standards für die Allgemeine Hochschulreife (KMK 2012) aufgegriffen, da hier mit der interkulturellen kommunikativen Kompetenz, der Text- und Medienkompetenz, der Sprachbewusstheit und der Sprachlernkompetenz (die auch schon in den Bildungsstandards von 2003 und 2004 berücksichtigt wurden) auch komplexere Dimensionen fremdsprachlicher Bildung spezifiziert und stärker in den Vordergrund gerückt werden (vgl. den Beitrag von Karin Vogt in diesem Heft).

Aus der insgesamt kritischen Rezeption der Bildungsstandards in der Fremdsprachendidaktik erwachsen folglich auch Bedenken, ob das in den Standards implizierte Konzept der Kompetenzorientierung im Zusammenspiel mit etablierten

fremdsprachendidaktischen Konzepten tatsächlich als innovativer Baustein zur Weiterentwicklung des Fremdsprachenunterrichts taugt, wie Nold (2015) vorschlägt.<sup>2</sup>

Mit Blick auf die vor mehr als vier Jahrzehnten angestoßene kommunikative Wende erkennt Quetz (2012) in der Kompetenzorientierung zumindest eine Mahnung, den kommunikativen Ansatz endlich wirksam in der Alltagspraxis des (Englisch-)Unterrichts umzusetzen. Dies scheint auch dringend geboten, wenn der Fremdsprachenunterricht aus der Perspektive der Allgemeinen Didaktik und Schulpädagogik weiterhin als Schulfach "mit aufbauender, hierarchischer Struktur" (Wiater 2013: 155) wahrgenommen wird, in dem "die persönliche Auseinandersetzung mit den Themen/Inhalten" (ibid.) nicht im Mittelpunkt stehe.

Die Fremdsprachenforschung hadert ganz offensichtlich mit der Frage, inwiefern sie bereit ist, ihre vorhandenen Konzepte zu den komplexen Inhalten, Lernprozessen und Bildungszielen des Fremdsprachenunterrichts mit Blick auf die begrifflichen Kategorien und messtheoretischen Kriterien der Standard- und Kompetenzorientierung anzupassen, oder ob sie dies ablehnt und sich stattdessen dafür einsetzt, unterrichtliche "Freiräume für interkulturelle, reflexive, ethische und ästhetische Aspekte sprachlichen Lernens" (Hu 2008: 11) zu schützen (vgl. Burwitz-Melzer 2005; Caspari, Kötter, Rossa, Schramm, Tesch, Vollmer & Zydati 2012; Hu, Caspari, Grnewald, Kster, Nold, Vollmer & Zydati 2008).

Unabhngig von der Positionierung zu dieser Frage kann die Realitt der in der Schulpraxis curricular eingeforderten Kompetenzorientierung meines Erachtens jedoch nicht einfach mit einem Verweis auf die Inkompatibilitt der bestimmenden Begriffe mit dem etablierten Fachdiskurs in der Fremdsprachenforschung abgetan werden.

Angesichts des Handlungsdrucks, der insbesondere fr die Lehrkrfte aus der Forderung nach Rechenschaftslegung ber das fremdsprachliche Lernen erwchst, ist es dringend notwendig, die Entwicklung fremdsprachendidaktischer Konzeptionen zur Frderung des Kompetenzerwerbs in einer interkulturellen und kommunikativen Unterrichtspraxis voranzubringen. Diese Vernderungsprozesse sollten jedoch nicht nur durch vereinzelte Unterrichtsrezepte und eine eher impressionistische Kritik an standardorientierten Aufgaben (vgl. De Florio-Hansen 2015) kommentiert, sondern auch durch Theoriebildung reflektiert und forschend begleitet werden.<sup>3</sup>

Lehrkrfte mssen die didaktischen Problemstellungen eines kompetenzorientierten Fremdsprachenunterrichts – z.B. die Auswahl konkreter Inhalte und Ziele fr Unterrichtsvorhaben, die Diagnose von Lernvoraussetzungen und die Frderung der Kompetenzentwicklung der Lernenden unter Bercksichtigung ihrer

---

2 Zum Potential von Lernstandserhebungen fr die Unterrichtsentwicklung siehe auch Bial (2014).

3 Siehe z.B. das interdisziplinre Projekt zur Frage "How does change happen?", FaBiT (Doff, Bikner-Ahsbahs, Grnewald, Komoss, Lehmann-Wermser, Peters & Rovir 2014).

individuellen Spracherwerbsgeschichten – professionell bewältigen; d.h. unter Rückgriff auf theoretisches Wissen, didaktische Strategien und eine kritische Reflexion des eigenen Handelns.

Empirisch überprüfte und fachdidaktisch spezifizierte Modelle zur Entwicklung fremdsprachlicher Kompetenzen, die als kohärenter theoretischer Referenzrahmen für die oben beschriebenen didaktischen Problemstellungen genutzt werden könnten, liegen noch nicht vor. Weder der GeR noch die Bildungsstandards können dies leisten (vgl. Burwitz-Melzer & Quetz 2006). So kann die Entwicklung kompetenzorientierter Lerngelegenheiten (vgl. Hallet 2013) eher als ein exploratives Projekt beschrieben werden, das Lehrkräfte in ihren jeweiligen lokalen Kontexten kooperativ, etwa im Sinne der Aktionsforschung (vgl. Burns 2010), verfolgen müssten.

Mit den Publikationen zur Konkretisierung der Bildungsstandards, die aus der Zusammenarbeit von Bildungsforschung, Fremdsprachendidaktik, Lehreraus- und -fortbildung und Lehrkräften aus der Praxis hervorgegangen sind, liegen nun Lernaufgaben, didaktische Kommentierungen und zusätzliche Lehr-/Lernmaterialien vor. Diese Materialien zeigen exemplarisch auf, wie im Englisch- bzw. Französischunterricht die in den Bildungsstandards formulierten Kompetenzen durch eine Verknüpfung von Themen, Inhalten und Aufgaben sinnvoll gefördert werden können (vgl. Tesch et al. 2008, 2016).

### **2.3 Die empirische Perspektive der Bildungsforschung: Forschungsstand zur Wirkung von zentralen Prüfungen**

Die bisherigen Ausführungen lassen erkennen, dass sich die Kritik an der Standard- und Kompetenzorientierung insbesondere an der Wahrnehmung entzündet, dass diese bildungspolitischen Vorgaben im Kern den Kriterien von *large-scale* Tests und zentralen Prüfungen verpflichtet seien und somit eine inhaltliche Reduktion von Bildungszielen auf messbare Kompetenzen in Kauf nehmen.

Die Ergebnisse der Testverfahren sollen neben dem Anspruch, Lernergebnisse transparent zu machen und langfristig eine bessere Vergleichbarkeit von Bildungsabschlüssen zu erreichen, vor Allem zu einer systematischen, datengestützten Verbesserung der Qualität des Bildungssystems beitragen. Angesichts dieser ehrgeizigen Ziele stellt sich die Frage, in welcher Weise und unter welchen Bedingungen sich zentrale Tests und Prüfungen auf das Bildungssystem, die Schulentwicklung und die Unterrichtspraxis auswirken (vgl. Meyerhöfer, Linklitzing, DiFuccia & Müller-Frerich 2013). In einem Forschungsüberblick stellen Klein et al. allerdings fest, dass zentrale Prüfungen in Deutschland implementiert

wurden, "ohne dass empirisch fundierte Ergebnisse zu deren Wirksamkeit – im Hinblick auf intendierte und nicht-intendierte Wirkungen – vorlagen" (Klein, Krüger, Kühn & van Ackeren 2014: 8).

Mit Blick auf die in den letzten zehn Jahren eingeführten *low-stakes tests* (z.B. Vergleichsarbeiten, Lernstandserhebungen, Kompetenztests) und die in den Bundesländern sehr unterschiedlich etablierten zentralen Abschlussprüfungen (zentrale Prüfungen für den Mittleren Schulabschluss und das Zentralabitur) wurden einige Forschungsprojekte – mehrheitlich zum Zentralabitur – durchgeführt. Es liegen aber noch keine empirischen Untersuchungen vor, die sich aus einer fachdidaktischen Perspektive explizit auf den Fremdsprachenunterricht beziehen.

Empirische Studien der Bildungsforschung prüfen auf der Makroebene der Bildungspolitik und -administration, inwiefern Effekte der verschiedenen Steuerung intentionen (z.B. Zertifizierung, Vergleichbarkeit, Schulentwicklung) im Kontext zentraler Prüfungen nachgewiesen werden können. Auf der Mesoebene der Schule steht die Frage im Vordergrund, inwiefern die Implementation zentraler Prüfungen zu einer größeren Kooperation zwischen Lehrkräften führt, und welche (Leitungs-)Kompetenzen und Unterstützungsmaßnahmen hierfür erforderlich sind. In der Erforschung der Mikroebene des Unterrichts liegen Studien zur Rezeption zentraler Prüfungen bzw. ihrer Ergebnisse und Rückmeldungen bei Lehrkräften sowie bei Schülerinnen und Schülern vor (vgl. Klein, Krüger, Kühn & van Ackeren 2014: 23f.). Ein zweiter Schwerpunkt liegt naturgemäß auf der Untersuchung der Frage, ob zentrale Prüfungen zu besseren Schülerleistungen führen, denn "[t]heoretically, teachers as well as students should react to central exit exams by increasing effort" (Jürges, Schneider, Senkbeil & Carstensen 2012: 64).

#### *Führen zentrale Prüfungen zu besseren Schülerleistungen?*

In einem Vergleich von Schülerleistungen, die in der Ergänzungsstudie PISA-I Plus zwischen 2003 und 2004 zu mathematischem Wissen (*curricular knowledge*) und mathematischer Grundbildung (*literacy*) erhoben wurden, stellen Jürges et al. (2012) fest, dass die Leistungen in Bundesländern mit bereits etablierten zentralen Abschlussprüfungen (*central exit exams*, CEEs) nur im Bereich "mathematisches Wissen" und nur in "nicht-akademischen" Bildungsgängen höher sind als in Ländern ohne zentrale Abschlussprüfungen. Außerdem zeigen sich negative Effekte in Bezug auf die Einstellungen der Schülerinnen und Schüler zum Fach Mathematik: "There appears to be a downside to CEEs because students in CEE-states do like mathematics less. They find it less easy and more boring than those in non-CEE states" (ibid.: 64).

Maag Merki (2012) kann in einer Analyse der Schülerleistungen in den Fächern Englisch und Mathematik in den ersten drei Jahren nach Einführung des



Zentralabiturs in den Ländern Bremen und Hessen keinen generellen Effekt feststellen; allein im Englisch Leistungskurs lässt sich in beiden Bundesländern im Verlaufe der drei aufeinander folgenden Abiturjahrgänge eine positive Veränderung der Leistungen belegen.

*Werden Leistungsbewertungen im Zentralabitur vergleichbarer?*

Holmeier (2013) findet in ihrer Analyse der Leistungsbewertung im Zentralabitur Englisch und Mathematik für Bremen und Hessen Unterschiede zwischen den Fächern und den Kursarten. Eine bessere Vergleichbarkeit der Leistungsbewertung im Zentralabitur kann demnach nicht als genereller Effekt des Zentralabiturs bestätigt werden.

In einem Vergleich zwischen zwei Ländern mit zentralen (Baden Württemberg) bzw. dezentralen Abiturprüfungen (Rheinland-Pfalz) und einem Land (Nordrhein-Westfalen) im Übergang können van Ackeren et al. (2012) in den Fächern Mathematik, Deutsch und Biologie nur geringe Unterschiede in der Bewertung der Leistungen finden. So zeigt sich z.B. in Grund- und Leistungskursen in der zentralen Prüfung eine geringere individuelle und soziale Bezugsnormorientierung bei der Benotung (vgl. van Ackeren, Block, Klein & Kühn 2012). Neumann et al. (2011) können für die Fächer Mathematik und Englisch in Hamburg und Baden-Württemberg ebenfalls nachweisen, dass die Bewertung im Zentralabitur nur gering von schulspezifischen sozialen Bezugsmaßstäben beeinflusst wird (vgl. Neumann, Trautwein & Nagy 2011).

In ihrer Analyse zentraler Abituraufgaben im Fach Englisch stellt Lorenz (2013) fest, dass die Abiturleistungen des Abiturjahrgangs 2009 in Nordrhein-Westfalen geschlechts- und schulformspezifische Unterschiede erkennen lassen, die auch auf der Ebene der Teilaufgaben zu unterschiedlich ausgeprägten Leistungen führen. Die Skalierung der Abituraufgaben zeigt, dass verschiedene Teilaufgaben, aus denen die Prüflinge wählen konnten, teilweise deutlich unterschiedliche Schwierigkeitswerte aufweisen. In der Aufgabenentwicklung waren die Teilaufgaben aber demselben Anforderungsniveau zugeordnet worden. Das Ziel einer besseren Vergleichbarkeit der erfassten Abiturleistungen durch zentral gestellte Aufgaben wird so noch nicht erreicht.

Insgesamt stehen diese uneinheitlichen Befunde im Widerspruch zu dem von Schwerdt & Wößmann (2015) in den Metaphern der Wirtschaftswissenschaft formulierten Ergebnis ihrer Studie zum Beitrag zentraler Prüfungen zur differenzierten Auswahl von Absolventinnen und Absolventen für den Arbeitsmarkt: "Grades obtained in central exams are a less noisy signal for effective human capital" (Schwerdt & Wößmann 2015: 18).

*Tragen zentrale Lernstandserhebungen und Prüfungen zur Unterrichtsentwicklung bei?*

Kühn (2010) untersucht zentral gestellte Abituraufgaben in den Naturwissenschaften und kommt nach einer international vergleichenden Nachfolgestudie (2011) zu dem Schluss, dass sich für die Annahme, zentral gestellte Aufgaben seien ein wirksames Instrument zur Steuerung der Unterrichtsentwicklung, keine Belege finden lassen. Der Erklärungsansatz der Autorin stützt sich auf eine Annahme, die aus Sicht der Unterrichtsforschung sicher den Erwartungen entspricht: "[T]he innovation potential of individual teachers seems to be higher than that of a central exam commission" (Kühn 2011: 194).

Otto & Kühn (2014) analysieren zentrale Prüfungsaufgaben im Fach Deutsch und zeigen, dass diese durchaus fachdidaktische Innovationen hinsichtlich der Aufgabenformate und der Auswahl der zugrunde gelegten Texte aufgreifen. Dieser Befund gilt allerdings stärker für die Aufgaben zum mittleren Schulabschluss, und insgesamt dominieren weiterhin traditionelle Formate und Textsorten, von denen keine nachhaltige Steuerung der Unterrichtsentwicklung zu erwarten ist.

Eine Studie zur Rezeption der Vergleichsarbeiten in der Sekundarstufe 1 durch Lehrkräfte und Schulleitungen (Nachtigall & Hellrung 2013) findet in der Auswertung von Nutzerbefragungen zu Kompetenztests in Thüringen ebenfalls Anhaltspunkte, die die langfristige Wirksamkeit von zentralen Tests in Bezug auf die Unterrichtsentwicklung zumindest in Frage stellen. Die Daten aus den Jahren 2003 bis 2012 zeigen einen rückläufigen Trend in Bezug auf die Intensität, mit der Lehrkräfte und Schulleitung sich mit den Ergebnissen und Rückmeldungen der Kompetenztests auseinandersetzen, und mit Blick auf die Frage, wie sie die Nützlichkeit der Rückmeldungen einschätzen.

Richter et al. (2014) stellen dagegen im Kontext VERA 3, Vergleichsarbeiten in der Grundschule, einen schwachen positiven Zusammenhang zwischen den Überzeugungen von Lehrkräften in Bezug auf VERA als Instrument der Unterrichtsentwicklung und der erfassten Lesekompetenz der Schülerinnen und Schüler fest (vgl. Richter, Böhme, Becker, Pant & Stanat 2014).

*Wie verändert sich die Unterrichtspraxis nach Einführung des Zentralabiturs aus der Schülerperspektive?*

Holmeier & Maag Merki (2012) erfassten im Rahmen der bereits oben angeführten Längsschnittstudie (Maag Merki 2012) zur Einführung des Zentralabiturs in Bremen im Vergleich zu dem bereits etablierten System in Hessen Veränderungen in den Wahrnehmungen der Schülerinnen und Schüler zur Unterrichtspraxis. Nur in Englisch-Leistungskursen in Bremen nehmen die Schülerinnen und Schüler mit der Einführung zentraler Abiturprüfungen eine etwas stärkere Kompetenz- und

Autonomieunterstützung ( $d = 0,14$  bzw.  $d = 0,20$ )<sup>4</sup>, eine größere Motivierungsfähigkeit ihrer Lehrpersonen ( $d = 0,34$ ) und ein stärkere kognitive Aktivierung im Unterricht ( $d = 0,39$ ) wahr, die über den Erhebungszeitraum von drei Jahren (2007 bis 2009) hinweg andauern. In Hessen zeigt sich im selben Zeitraum nur ein schwacher positiver Effekt ( $d = 0,13$ ) in Bezug auf die Wahrnehmung der Motivierungsfähigkeit der Lehrpersonen (vgl. Holmeier & Maag Merki 2012: 157f.).

*Wie verändern sich Selbstwirksamkeitsüberzeugungen und die Arbeitszufriedenheit der Lehrkräfte im Zusammenhang mit der Einführung des Zentralabiturs?*

Mit Blick auf einen wichtigen Indikator für die Wirkung des Zentralabiturs auf der Schulebene, die Selbstwirksamkeitsüberzeugungen der Lehrkräfte, zeigt sich in derselben Studie (Jäger 2012) ein schwacher negativer Zusammenhang zwischen der Wahrnehmung der individuellen Selbstwirksamkeit und der wachsenden Erfahrung mit zentralen Prüfungen (standardisierter Koeffizient  $-0,08 - p < .001$ ).<sup>5</sup> Analog ist eine sinkende Arbeitszufriedenheit (bzw. ein leichter Anstieg auf der Skala "Unzufriedenheit") bei zunehmender Erfahrung mit dem Zentralabitur zu beobachten (standardisierter Koeffizient  $0,03 - p < 0.1$ ). Die Autorin vermutet, dass "das Instrument Zentralabitur in die Autonomie der Lehrpersonen eingreift, was als Zuschreibung von Inkompetenz in diesem Bereich erlebt werden könnte" (Jäger 2012: 90). Zudem könne die Umstellung auf das neue Prüfungssystem als Belastung erlebt werden und die Handlungsspielräume der Lehrpersonen insbesondere bei der Gestaltung des Unterrichts reduzieren (vgl. *ibid.*).

*Ist das Zentralabitur im Sinne der Standard- und Kompetenzorientierung wirksam?*

Die überschaubaren empirischen Ergebnisse zum Zentralabitur in Deutschland sind so uneinheitlich wie das Zentralabitur selbst, dessen Grad an Standardisierung in seinen föderal ausdifferenzierten Ausprägungen bisher stark variiert. Mit den Bildungsstandards für die Allgemeine Hochschulreife (KMK 2012) und dem Beschluss über einen gemeinsamen Aufgabenpool für die zentralen Abiturprüfungen ab 2017 (KMK 2015) sind weitere Bemühungen zu erkennen, die Standardisierung des Abiturs in Richtung einer tatsächlich zentralen Abschlussprüfung zu forcieren. Dennoch wird zunächst offen bleiben, wie wirksam das Zentralabitur in Deutschland sein kann:

[D]oubts remain as to the general capacity of the statewide Abitur to bring about the changes intended by the education authorities – especially in view of the diverse forms in which exit exams are currently used across the German states (van Ackeren et al. 2012: 16).

---

4 Cohens  $d$ : Effektstärke für Mittelwertunterschiede zwischen zwei Gruppen. Nach Cohen bedeuten Werte  $d > 0,2$  einen kleinen,  $d > 0,5$  einen mittleren und  $d > 0,8$  einen starken Effekt (vgl. Cohen 2013).

5 Regressionskoeffizienten aus einer Mehrebenenregressionsanalyse werden nach Hox (2002) standardisiert, damit die Werte vergleichbar werden.

Aber auch für die traditionell etablierten zentralen Abschlussprüfungen in anderen Staaten lassen sich keine stabilen, systematischen Effekte im Sinne der beabsichtigten Qualitätssicherung und -entwicklung nachweisen. Zu unterschiedlich sind offenbar sowohl die nationalen Bildungskontexte, in denen jeweils ein spezifisches, komplexes Netzwerk interagierender Bedingungen und Faktoren die Qualität von Bildungsprozessen bestimmt, als auch die vielfältigen Verfahren, wie zentrale Prüfungen implementiert und ausgewertet werden (vgl. Bol, Witschge, Van de Werfhorst & Dronkers 2014). Klein et al. (2014: 26f.) kommen daher zu diesem Schluss: "Die empirischen Befunde deuten [...] insgesamt eher auf eine – gemessen an den Ansprüchen – eingeschränkte Steuerungswirksamkeit zentraler Abschlussprüfungen hin".

### **3. Ausblick: zur Erforschung des Fremdsprachenunterrichts in Zeiten des Monitoring-Paradigmas**

Ich habe bereits an anderer Stelle dafür plädiert, dass die Fremdsprachenforschung sich ihrer Verantwortung für die kritische Begleitung der kompetenzorientierten Messung von Schülerleistungen bewusst werden und dieses Forschungsfeld nicht mit einem bloßen Hinweis auf grundsätzliche Bedenken der Bildungsforschung und der pädagogischen Psychologie überlassen möge (Rossa 2007, 2012). Stattdessen sollte die Fremdsprachenforschung ihre methodisch vielfältigen Bemühungen um die empirische Erforschung des Fremdsprachenunterrichts als Ressourcen begreifen, die eine mehrperspektivische Betrachtung der Auswirkungen des sogenannten "Monitoring-Paradigmas" (Böttcher 2015) auf das Lehren und Lernen von Fremdsprachen ermöglichen. So könnten Beiträge zu der von Bonnet & Breidbach (2013: 34) formulierten Aufgabe, "die Implikationen der Rede von Kompetenzorientierung auszubuchstabieren und ihre Folgen abzuschätzen", auch mit empirisch gewonnen Erkenntnissen gestützt werden.

#### **3.1 Forschungsmethodische Ansätze in der Sprachtestforschung zum Konzept *washback***

In der Sprachtestforschung zählen die Auswirkungen von Tests und der Interpretation ihrer Ergebnisse in Kontexten des Lehrens und Lernens spätestens seit Veröffentlichung der grundlegenden Beiträge von Messick (1989; 1996) zu den zentralen Gegenständen theoretisch und empirisch orientierter Arbeiten. Unter dem Begriff "consequential aspect of construct validity" fordert Messick (1996: 241), dass beabsichtigte und unbeabsichtigte Konsequenzen der Interpretation

und Verwendung von Testergebnissen zu untersuchen seien, um Belege für die Validität eines Tests zusammenzutragen. Diese Forderung bestimmt auch weiterhin theoretische Modelle, die sich mit dem Gütekriterium der Validität, z.B. von testbasierten bildungspolitischen Vorhaben, auseinandersetzen (vgl. Chalhoub-Deville 2015).

Im Rahmen aller denkbaren (z.B. gesellschaftlichen, politischen usw.) Wirkungen von Tests (*impact*) definieren Alderson & Wall (1993) *washback* als die Einflüsse, die *high-stakes tests* – d.h. Prüfungen, von deren Ergebnissen für die Testteilnehmenden bedeutsame Entscheidungen abhängen (z.B. Zugang zu einem Bildungsangebot) – auf die Praxis in Bildungskontexten ausüben. In einem Überblick zur *washback*-Forschung im Kontext des Englischunterrichts identifiziert Spratt (2005) fünf Aspekte, die in Studien zu den Auswirkungen des Einsatzes von Sprachtests untersucht werden: Unterrichtsmaterialien, Unterrichtsmethoden, Einstellungen und Emotionen der Akteure und die Prozesse bzw. Produkte des Lehrens und Lernens (vgl. auch Cheng, Sun & Ma 2015).

Studien der Sprachtestforschung zum Konzept *washback* unterscheiden sich durch diese differenzierte inhaltliche Ausrichtung erkennbar von den in Kapitel 2.3 referierten empirischen Beiträgen der Bildungsforschung zur Wirksamkeit zentraler Prüfungen, die schwerpunktmäßig an den Einstellungen der Lehrkräfte und den erfassten Kompetenzen interessiert sind. In der forschungsmethodischen Umsetzung äußert sich dieser Unterschied darin, dass in einer Reihe von *washback*-Studien die üblichen quantitativen Daten aus Fragebögen, Kompetenztests und anderen Leistungsmessungen und -bewertungen durch explorativ-interpretative Methoden der Datenerhebung und -analyse komplementiert werden (siehe Tabelle 1).

Diese forschungsmethodische Öffnung ergibt sich nach Bailey sachlogisch aus den im Diskurs der Sprachtestforschung unstrittigen Elementen des theoretischen Konzepts *washback*:

If the core of any definition of washback has to do with the effects of tests on learning and teaching, then it is necessary to document those effects – both by asking about and by watching teaching and learning (Bailey 1999: 36).

Wenn das Forschungsziel also darin besteht, die Wirkungen von Tests auf Bildungsprozesse nicht nur zu beschreiben und ihre relativen Beiträge zu bzw. korrelativen Zusammenhänge mit den gemessenen Veränderungen zu schätzen, sondern diese Veränderungen in ihren verschiedenen Ausprägungen zu verstehen, dann ist der Einsatz ethnographisch orientierter Methoden unumgänglich, wie auch Wall & Alderson in ihrer viel zitierten Studie berichten:

Above all, we would not have known that the exam had virtually no impact on methodology if we had not observed classes. However, we would not have been able to understand why

the exam had no impact on how teachers taught without discussions with teachers after having observed their classes. [...] Without observations the researchers are unlikely to know all the questions they should be asking and may not understand (or be sufficiently critical of) the answers they are given (Wall & Alderson 1993: 65).

Tabelle 1: Studien zum Konzept *washback*: Gegenstände und Forschungsmethoden

Quelle	Gegenstände	Methoden
Andrews, Fullilove & Wong 2002	Use of English Test in der Sekundarstufe, mündliche Sprachproduktion	Videographie von Übungstests Bewertung mündlicher Prüfungen Diskursanalyse
Read & Hayes 2003	IELTS, Hochschulzugang in Neuseeland	Fragebogen (Lehrkräfte und Lernende) Unterrichtsbeobachtung (COLT, UCLES) Schriftliche <i>pre-</i> und <i>post-tests</i>
Saif 2006	<i>needs-based assessment</i> mündliche Sprachproduktion Ausbildung von <i>Teaching Assistants</i>	Interviews Unterrichtsbeobachtungen Tests der mündlichen Sprachproduktion
Latimer 2009	Cambridge PET und FCE in der Primar- bzw. Sekundarstufe	Unterrichtsbeobachtungen im Sinne der <i>grounded theory</i> Fragebogen (Lernende) Interviews (Lehrkräfte, Schulleitung)
Munoz & Alvarez 2010	<i>formative evaluation</i> mündliche Sprachproduktion	Fragebogen (Lehrkräfte und Lernende) Unterrichtsbeobachtungen, Analyse der von den Lehrkräften genannten Lernziele und der Gelegenheiten für <i>self-assessment</i> Externe Bewertung der mündlichen Sprachproduktion der Lernenden
Wall & Horak 2011	TOEFL iBT	Analyse von Unterrichtsmaterialien Interviews mit Lehrkräften Schriftliche Befragung via E-Mail Unterrichtsbeobachtung
Booth 2012	TOEIC, Überzeugungen der Lernenden	Fokusgruppeninterviews Halb-strukturierte Lerntagebücher Interviews mit einzelnen Lernenden

Diese Einsicht mag überraschen, da der Sprachtestforschung angesichts ihres wesentlichen Untersuchungsgegenstands typischerweise eine positivistische erkenntnistheoretische Position zugeschrieben wird. Tatsächlich lässt sich in den einschlägigen Publikationen erst mit Beginn der 1990er Jahre eine stärkere Öffnung hin zu qualitativen Forschungsansätzen beobachten (vgl. Lazaraton & Taylor 2007), die durch den Diskurs zur Validität als zentralem Gütekriterium (Messick 1989) und zur Frage "Does washback exist?" (Alderson & Wall 1993) ganz offensichtlich gefördert wurde. Angesichts der sehr uneinheitlichen Ergebnisse bisheriger Studien zur Wirksamkeit zentraler Prüfungen in Deutschland mag es lohnenswert sein, das Phänomen *washback* mit einem vielfältigeren Methodenrepertoire zu erforschen und sich dabei methodisch an den oben aufgeführten Studien zu orientieren.

### **3.2 Skizze einer Forschungsfrage und beispielhafte inhaltliche Kategorien: Wie nehmen Lehrkräfte die Auswirkungen von zentralen Prüfungen auf den Englischunterricht wahr?**

Im Folgenden sollen vorläufige inhaltliche Kategorien zur wahrgenommenen Wirkung zentraler Prüfungen auf die Alltagspraxis des Englischunterrichts vorgestellt werden, die einer informellen Befragung von zehn Lehrkräften entnommen wurden. Die Beobachtungen und Einschätzungen der Informanten sollen mit Blick auf die Forschungsfrage "Wie nehmen Lehrkräfte die Auswirkungen von zentralen Prüfungen auf den Englischunterricht wahr?" exemplarisch aufzeigen, welche Anknüpfungspunkte sich im Rahmen zukünftiger Forschungsvorhaben zu *washback*-Effekten im Kontext der Standard- und Kompetenzorientierung in Deutschland möglicherweise als relevant erweisen könnten.

Die Informanten (N = 10) wurden im Sinne des *purposeful sampling* (Patton 2002: 220 ff.) gezielt ausgewählt. Alle Informanten haben längere (10-25 Jahre) unterrichtspraktische Erfahrungen im Fach Englisch an den Schulformen Gymnasium, Gesamtschule und Realschule in Nordrhein-Westfalen gesammelt; und zwar vor und nach der Einführung von zentralen Prüfungen und Lernstandserhebungen.<sup>6</sup> Vier Informanten verfügen zusätzlich über Erfahrungen als Fachleiterin bzw. Fachleiter in der zweiten Phase der Lehrerbildung. Diese Position ist in der Regel mit der Aufgabe verbunden, regelmäßig Unterricht zu beobachten und die Unterrichtsqualität einzuschätzen. In dieser schriftlichen Online-Befragung haben sich die Informanten per E-Mail<sup>7</sup> zu drei Fragen geäußert:

1. Beschreiben Sie bitte vor dem Hintergrund Ihrer auf die Schule bezogenen Erfahrungen, Beobachtungen und Erkenntnisse, inwiefern sich die Alltagspraxis des Englischunterrichts seit Einführung der zentralen Prüfungsaufgaben verändert hat.
2. Welche positiven Impulse für die Unterrichtsentwicklung im Fach Englisch können von zentralen Prüfungen ausgehen?
3. Welche negativen Wirkungen auf die Unterrichtsentwicklung im Fach Englisch sind durch zentrale Prüfungen zu befürchten?

Es war beabsichtigt, durch die Formulierung der Fragen erstens Einschätzungen zu Veränderungsprozessen in der Alltagspraxis des Englischunterrichts seit der

---

6 Zentrale Prüfungen: ab 2006 "Zentrale Prüfungen 10" für den Mittleren Schulabschluss bzw. den Hauptschulabschluss nach Klasse 10, ab 2007 Zentralabitur. Lernstandserhebungen: ab 2004 "Lernstand 9", ab 2006 ersetzt durch "Lernstand 8".

7 Diese Methode wurde gewählt, weil sie die Kriterien der Freiwilligkeit und Flexibilität der Teilnahme an der Befragung am ehesten gewährleistet (vgl. Porter 2008).

Einführung zentraler Prüfungen zu gewinnen und zweitens positive bzw. drittens negative Bewertungen dieser Veränderungen zu elizitieren. Diese Absicht führt in der Konsequenz bei den zuspitzenden Fragen Nr. 2 und Nr. 3 zu einem gewissen suggestiven Charakter, der üblicherweise in schriftlichen Befragungen vermieden werden sollte. Im Rahmen dieser informellen Vorstudie mit einer sehr kleinen Stichprobe schien diese Zuspitzung aber nötig zu sein, damit eine möglichst große Bandbreite verschiedener Einschätzungen erfasst werden kann.

Die folgenden Kategorien einer ersten qualitativen Analyse fassen mehrere inhaltlich ähnlich formulierte Wahrnehmungen und Beobachtungen zusammen. Die Daten erheben nicht den Anspruch, auf andere Kontexte generalisierbar zu sein und sind zudem durch ihre Vermengung von zwei unterschiedlichen Prüfungssystemen und drei Schulformen nur eingeschränkt auswertbar. Zukünftige Studien könnten die hier dargelegten Kategorien dennoch aufgreifen und zur Orientierung bei der Entwicklung von Hypothesen, theoretischen Propositionen und Forschungsfragen nutzen.

Die prototypischen Äußerungen der Informanten wurden für die exemplarische Darstellung der inhaltlichen Kategorien in diesem Beitrag sprachlich leicht geglättet und ergänzt, um die Lesbarkeit und Verständlichkeit zu erhöhen. Die Ergänzungen stehen in Klammern, die Äußerungen der Informanten sind durch doppelte Anführungszeichen markiert.

In den Äußerungen, die sich auf die erste Frage beziehen, sind implizit eher neutral bewertete Beobachtungen zu erkennen, die einen generellen Effekt des Zentralabiturs bzw. der Zentralen Prüfungen 10 auf den Unterricht ausmachen.

*Kategorie 1: Der Unterricht orientiert sich an inhaltlichen Vorgaben und zielt auf die Entwicklung von Kompetenzen.*

"Die Alltagspraxis des Englischunterrichts richtet sich thematisch natürlich nach den Vorgaben für das Zentralabitur".

(Im Vergleich beobachte ich) "gelenkteres Lernen" (und ein) "Training bestimmter Kompetenzen".

(Der) "konsequente Kompetenzbezug (des Unterrichts) wird durch die zentralen Prüfungen viel stärker eingefordert".

Ein Informant berichtet abweichend mit Blick auf die Sekundarstufe I an einer Realschule, dass eine "systematische Vorbereitung auf die Abschlussprüfung [...] im Sinne eines kompetenzorientierten Unterrichts, der auf kumulatives Lernen setzt" nicht zu beobachten sei.

Die folgenden, tendenziell eher negativ bewerteten Beobachtungen spiegeln einige der Sorgen wider, die in den in Kapitel 2.2 referierten, didaktisch bzw. pädagogisch ausgerichteten Publikationen dargelegt wurden: eine Reduktion der Inhalte, *teaching to the test*, eine Weiterentwicklung des Unterrichts wird durch zentrale Prüfungen nicht gefördert.



Kategorie 2: *Der Unterricht wird inhaltlich beschnitten*

"Die Zentralabitur-Vorgaben führen einerseits natürlich zu einem gewissen *streamlining* der Kurse."

"Englisch (wird) als Werkzeug (verstanden), um sich in der Berufswelt (bzw.) auf globalisierten Märkten durchzusetzen. Künstlerisches, Literarisches, Kreatives tritt in der Tendenz in den Hintergrund".

"Die Auswahl an Romanen oder (thematischen) Schwerpunkten, wie sie früher war, ist eingeschränkter".

Kategorie 3: *Der Unterricht orientiert sich an den Prüfungsformaten.*

(Die) "Kolleginnen und Kollegen orientieren sich an dem, was abgefragt wird".

(Es ist das Phänomen) "*teaching to the test*" (zu beobachten).

Kategorie 4: *Die Lehrkräfte nutzen im Unterricht vermehrt Unterrichtsmaterial aus Lehrwerken.*

"Die Tendenz geht jetzt dahin, wieder Lehrbuchunterricht zu machen und somit fehlt auch den Schülerinnen und Schülern der Reiz an der Oberstufe".

"Der geringe Anteil der sprachlichen Richtigkeit an der Gesamtbewertung der Arbeiten (führt) nicht (wie beabsichtigt) zu einer Abkehr vom lehrwerksorientierten, auf grammatischer Progression basierenden Unterricht".

Kategorie 5: *Lehrkräfte werden durch die Abiturvorgaben stärker belastet und in ihrer Freiheit beschränkt.*

"Lehrkräfte haben Angst, irgendetwas in der Vorbereitung auf die Abiturprüfung vergessen zu haben".

"Neue Klausurformate" (Integrative Aufgaben, Überprüfung von Sprachmittlung und Hör- bzw. Hörsehverstehen<sup>8</sup> bedeuten) "weitere Anforderungen für die Kollegen, die freie Unterrichtsplanung wird (so noch) stärker eingeschränkt".

Einige Informanten äußern auch Einschätzungen, die dieser Wahrnehmung widersprechen:

"Die Abiturvorgaben erleichtern die Unterrichtsvorbereitung".

(Die Abiturvorgaben) "schaffen gewisse Freiräume, die man thematisch/fachmethodisch ausnutzt".

Die folgenden, eher positiv wahrgenommenen Veränderungen scheinen mit den intendierten Effekten zentraler Prüfungen kongruent oder zumindest kompatibel zu sein: mehr Kooperation unter den Lehrkräften, verbesserte Unterrichtsqualität, bessere Lernbedingungen für die Lernenden und vergleichbarere Prüfungsleistungen.

Die verbalen Daten deuten an, dass zukünftige, stärker explorativ-interpretativ ausgerichtete Studien helfen können, diese Veränderungsprozesse und ihre Wahrnehmung durch die Akteure genauer zu verstehen. Dieser Bereich berührt schließ-

---

8 Vgl. Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen 2015.

lich den Kern der intendierten Wirkungen der Standard- und Kompetenzorientierung und zugleich einen bedeutsamen Aspekt der professionellen Kompetenzen und Überzeugungen der Lehrkräfte.

*Kategorie 6: Lehrkräfte arbeiten stärker zusammen; dies steigert die Unterrichtsqualität.*

"Ich stelle eine verstärkte Zusammenarbeit im Kollegium fest; ein größeres Interesse daran, was der Andere macht, paralleles Unterrichten. Da war früher mehr Einzelkämpfertum."

"Die Unterrichtsqualität steigt. Dies liegt daran, dass Unterrichtsvorhaben gemeinsam besprochen werden und diese somit ausgereifter sind, als wenn dies einzelne Lehrkräfte tun".

*Kategorie 7: Die Vorgaben zu zentralen Prüfungen verbessern die Unterrichtsqualität.*

"Die Vorgabe, Sprachmittlung und Hör- bzw. Hörsehverstehen mit abprüfen zu müssen, sorgt dafür, dass der Unterricht nicht mehr schwerpunktmäßig auf Schreiben, Leseverstehen (und das) Unterrichtsgespräch ausgerichtet ist, sondern eben auch auf andere Kompetenzen, die im 'echten Leben' relevant sind."

"Kolleginnen und Kollegen, die zu Zeiten der Inputorientierung bereits im Dienst waren, bekommen durch die Zentralen Prüfungen Impulse für eine kompetenzorientierte Unterrichtsentwicklung. Das ist meines Erachtens sehr positiv und schränkt ein 'inputorientiertes Dahinwurschteln' ein."

*Kategorie 8: Die Lernenden profitieren*

"Der Blick für die Heterogenität der Lerngruppen, der Blick für (einzelne) Schülerinnen und Schüler wird geschärft, (denn) alle Schülerinnen und Schüler sollen erfolgreich durch die Prüfungen (kommen)".

"Für die Schüler wird dadurch Vieles transparenter, vergleichbarer und somit weniger lehrerabhängig. Überhaupt scheinen mir die Schüler die Nutznießer zu sein."

Insgesamt zeigt diese Zusammenschau einen Ausschnitt der Wahrnehmungen und Beobachtungen zur Veränderung des Englischunterrichts im Kontext der Einführung zentraler Prüfungen, der von einer erstaunlich großen Homogenität und gleichzeitig hohen Differenziertheit der Äußerungen gekennzeichnet ist. Nur zu wenigen Aspekten, z.B. zur Frage, ob eine zielgerichtete Kompetenzorientierung in der Alltagspraxis des Englischunterrichts zu erkennen ist, lassen sich in den Äußerungen gegensätzliche Deutungen und Wahrnehmungen finden. Alle Informanten benennen sowohl positive als auch negative Effekte; in der relativen Verteilung der Menge ihrer Beobachtungen unterscheiden sich die Informanten allerdings deutlich. Einzelne Informanten benennen eine ganze Liste von negativ bewerteten Aspekten und können in der Einführung zentraler Prüfungen nur hypothetisch positive Wirkungen erkennen, die aber in der Realität nicht wahrnehmbar seien. Die Beobachtungen anderer Informanten wiederum sind entgegengesetzt verteilt, während die meisten Informanten eine tendenziell positive bis ausgewogene Einschätzung der wahrgenommenen Veränderungen erkennen lassen.

## 4. Fazit

Im Forschungsfeld der Standard- und Kompetenzorientierung operieren unterschiedlich motivierte, in der Sache und in ihren Deutungen der relevanten Begriffe in mancher Hinsicht unversöhnliche Akteure in einem hochkomplexen Kontext. Dies hat zur Folge, dass der Versuch einer Bestandsaufnahme, wie am hier vorliegenden Beitrag abzulesen ist, eine ganze Reihe von ungeklärten Fragen und Widersprüchen anerkennen muss: Die empirische Basis zur Frage der Wirksamkeit von Bildungsstandards und zentralen Prüfungen zur Steuerung der Qualitätsentwicklung im Bildungssystem ist dünn. Die Befunde empirischer Studien sind uneinheitlich. Die wissenschaftliche Rezeption dieser bildungspolitischen Reform, die als notwendige Innovation und als bedeutsamer Impuls zur Unterrichtsentwicklung legitimiert wird, ist in der Pädagogik und in der Fremdsprachendidaktik von Skepsis und Ablehnung geprägt.

Eines aber ist unverkennbar: Die implementierten Standards, Lernstandserhebungen und zentralen Prüfungen haben die Alltagspraxis in Schulen und im Fremdsprachenunterricht bereits verändert und stellen die Lehrkräfte vor didaktische und pädagogische Probleme. In diesem undurchsichtigen Feld ist es dringend geboten, die Veränderungen und Problemstellungen in der Praxis schulischer Bildungsprozesse durch einen explorativen, stärker qualitativen Forschungsansatz genauer zu erfassen und besser zu verstehen. Erst dann wird es möglich sein, dass eine fremdsprachendidaktisch informierte, datengeleitete (*bottom-up*) Unterrichtsforschung einen positiven *washback*-Effekt auf das bislang im Kern konzeptgeleitete (*top-down*) Bildungsmonitoring ausüben kann.

Eingang des revidierten Manuskripts 17.12.2015

## Literaturverzeichnis

- Aguado, Karin (2009), Am Anfang ist das Ziel... Für eine lern(er)zielorientierte Bestimmung von Unterrichtsinhalten. In: Bausch, Karl-Richard; Burwitz-Melzer, Eva; Königs, Frank G. & Krumm, Hans-Jürgen (Hrsg.) (2009), *Fremdsprachenunterricht im Spannungsfeld von Inhaltsorientierung und Kompetenzbestimmung. Arbeitspapiere der 29. Frühjahrskonferenz zur Erforschung des Fremdsprachenunterrichts*. Tübingen: Narr, 9-15.
- Alderson, J. Charles & Wall, Dianne (1993), Does Washback Exist? *Applied Linguistics* 14: 2, 115-129.
- Andrews, Stephen; Fullilove, John & Wong, Yama (2002), Targeting washback – a case-study. *System* 30: 2, 207-223.
- Bailey, Kathleen M. (1996), Working for washback. A review of the washback concept in language testing. *Language Testing* 13: 3, 257-279.
- Bailey, Kathleen M. (1999), *Washback in Language Testing*. Princeton, New Jersey: ETS.

- Bär, Marcus (2013), Standardentwicklung vs. Individualisierung: Zur Rolle von (komplexen) Lernaufgaben im kompetenzorientierten Fremdsprachenunterricht. In: Grünewald, Andreas; Plikat, Jochen & Wieland, Katharina (Hrsg.) (2013), 98-109.
- Bausch, Karl-Richard; Burwitz-Melzer, Eva; Königs, Frank G. & Krumm, Hans-Jürgen (Hrsg.) (2005), *Bildungsstandards für den Fremdsprachenunterricht auf dem Prüfstand. Arbeitspapiere der 25. Frühjahrskonferenz zur Erforschung des Fremdsprachenunterrichts*. Tübingen: Narr.
- Bausch, Karl-Richard; Burwitz-Melzer, Eva; Königs, Frank G. & Krumm, Hans-Jürgen (Hrsg.) (2009), *Fremdsprachenunterricht im Spannungsfeld von Inhaltsorientierung und Kompetenzbestimmung. Arbeitspapiere der 29. Frühjahrskonferenz zur Erforschung des Fremdsprachenunterrichts*. Tübingen: Narr.
- Bial, Jessica (2014), Lernstandserhebungen 8: Fluch und Segen eines Diagnoseinstruments. *Praxis Fremdsprachenunterricht* 11: 4, 7-11.
- Bol, Thijs; Witschge, Jacqueline; Van de Werfhorst, Herman G. & Dronkers, Jaap (2014), Curricular Tracking and Central Examinations. Counterbalancing the Impact of Social Background on Student Achievement in 36 Countries. *Social Forces* 92: 4, 1545-1572.
- Bonnet, Andreas & Breidbach, Stephan (2013), Blut ist im Schuh. Wie gut kleidet der Kompetenzbegriff die literarisch-ästhetische Bildung beim Tanz auf dem Hofball der Standardisierung? In: Grünewald, Andreas; Plikat, Jochen & Wieland, Katharina (Hrsg.) (2013), 20-35.
- Booth, Dawn K. (2012), *Exploring the Washback of the TOEIC in South Korea. A sociocultural perspective on student test activity*. Ph.D. dissertation, The University of Auckland, Auckland, NZ [Online: <https://researchspace.auckland.ac.nz/bitstream/handle/2292/19379/who1e.pdf?sequence=2>, 30.09.2015].
- Böttcher, Wolfgang (2015), Unterricht in Zeiten des Monitoring-Paradigmas. Eine curriculumsoziologische Analyse. In: Rolff, Hans-Günter (Hrsg.) (2015), 182-194.
- Bredella, Lothar (2006), Bildungsstandards und ihre Umsetzung. In: Timm, Johannes-Peter & Vollmer, Helmut J. (Hrsg.) (2006), 105-122.
- Brügelmann, Hans (2015), *Vermessene Schulen – standardisierte Schüler. Zu Risiken und Nebenwirkungen von PISA*, Hattie, VerA & Co. Weinheim: Beltz.
- Bundesministerium für Bildung und Forschung (Hrsg.) (2003), *Zur Entwicklung nationaler Bildungsstandards. Eine Expertise* [Online: [http://www.bmbf.de/pub/zur\\_entwicklung\\_nationaler\\_bildungsstandards.pdf](http://www.bmbf.de/pub/zur_entwicklung_nationaler_bildungsstandards.pdf), 13.01.2016].
- Burns, Anne (2010), *Doing action research in English language teaching. A guide for practitioners*. New York: Routledge.
- Burwitz-Melzer, Eva (2005), Kompetenzen für den Literaturunterricht heute. Ein Beitrag zur standardorientierten Didaktik des Fremdsprachenunterrichts. *Fremdsprachen Lehren und Lernen* 34, 94-111.
- Burwitz-Melzer, Eva & Quetz, Jürgen (2006), Trügerische Sicherheit: Referenzniveaus als Passepartout für den Fremdspracheunterricht? In: Timm, Johannes-Peter & Vollmer, Helmut J. (Hrsg.) (2006), 355-372.
- Caspari, Daniela (2014), Von der Hattie-Studie zur Diagnose. Oder: Tests als Lernchance für Lernende und Lehrende. *Praxis Fremdsprachenunterricht* 11: 4, 5-6.
- Caspari, Daniela; Kötter, Markus; Rossa, Henning; Schramm, Karen; Tesch, Bernd; Vollmer, Helmut J. & Zydaitis, Wolfgang (2012), Mindeststandards für Fremdsprachen am Ende der Pflichtschulzeit. Ein Positionspapier Deutschen Gesellschaft für Fremdsprachenforschung. *Zeitschrift für Fremdsprachenforschung* 23: 2, 243-268.
- Chalhoub-Deville, Micheline (2015), Validity theory. Reform policies, accountability testing, and consequences. *Language Testing*, first published on August 27, 2015 as doi:10.1177/0265532215593312.

- Cheng, Liying; Sun, Youyi & Ma, Jia (2015), Review of washback research literature within Kane's argument-based validation framework. *Language Teaching* 48: 4, 436-470.
- Christ, Herbert (2006), Lernaufgaben als Steuerungselemente? In: Bausch, Karl-Richard; Burwitz-Melzer, Eva; Königs, Frank G. & Krumm, Hans-J. (Hrsg.) (2006), *Aufgabenorientierung als Aufgabe. Arbeitspapiere der 26. Frühjahrskonferenz zur Erforschung des Fremdsprachenunterrichts*. Tübingen: Narr, 43-51.
- Cohen, Jacob (2013), *Statistical Power Analysis for the Behavioral Sciences* (2. Aufl.). Hoboken: Taylor and Francis.
- De Florio-Hansen, Inez (2015), *Standards, Kompetenzen und fremdsprachliche Bildung. Beispiele für den Englisch- und Französischunterricht*. Tübingen: Narr Francke Attempto.
- DESI-Konsortium (Hrsg.) (2008), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie*. Weinheim: Beltz.
- Doff, Sabine; Bikner-Ahsbahs, Angelika; Grünewald, Andreas; Komoss, Regine; Lehmann-Wermser; Peters, Maria & Roviró, Bárbara (2014), Change and continuity in subject-specific educational contexts: Research report of an interdisciplinary project group at the University of Bremen. *Zeitschrift für Fremdsprachenforschung* 25: 1, 73-88.
- Eberhardt, Jan-Oliver (2013), *Interkulturelle Kompetenzen im Fremdsprachenunterricht. Auf dem Weg zu einem Kompetenzmodell für die Bildungsstandards*. Trier: WVT Wissenschaftlicher Verlag Trier.
- Europarat (2011), *Gemeinsamer europäischer Referenzrahmen für Sprachen. Lernen, Lehren, Beurteilen*. Berlin: Langenscheidt.
- Fäcke, Christiane (2013), „Weiche Kompetenzen“ als Prüfstein zur Integration von Bildung und Standards. In: Grünewald, Andreas; Plikat, Jochen & Wieland, Katharina (Hrsg.) (2013), 36-46.
- Grünewald, Andreas; Plikat, Jochen & Wieland, Katharina (Hrsg.) (2013), *Bildung – Kompetenz – Literalität. Fremdsprachenunterricht zwischen Standardisierung und Bildungsanspruch* (1. Aufl.). Seelze: Klett/Kallmeyer.
- Gruschka, Andreas; Lin-Klitzing, Susanne; DiFuccia, David & Müller-Frerich, Gerhard (2013), Was wird aus der Kompetenzorientierung? In: Lin-Klitzing, Susanne; DiFuccia, David & Müller-Frerich, Gerhard (Hrsg.) (2013), 167-179.
- Hallet, Wolfgang (2013), Die komplexe Kompetenzaufgabe. *Der fremdsprachliche Unterricht Englisch* 47: 124, 2-8.
- Holmeier, Monika (2013), *Leistungsbeurteilung im Zentralabitur*. Wiesbaden: Springer.
- Holmeier, Monika & Maag Merki, Katharina (2012), Unterstützung im Unterricht im Kontext der Einführung zentraler Abiturprüfungen. In: Maag Merki, Katharina (Hrsg.) (2012), 155-178.
- Hox, Joop (2002), *Multilevel analysis. Techniques and applications*. Mahwah, NJ: Erlbaum.
- Hu, Adelheid (2008), Interkulturelle Kompetenz. Ansätze einer Dimensionierung und Evaluation einer Schlüsselkompetenz fremdsprachlichen Lernens. In: Frederking, Volker (Hrsg.) (2008), *Schwer messbare Kompetenzen. Herausforderungen für die empirische Fachdidaktik*. Baltmannsweiler: Schneider, 11-35.
- Hu, Adelheid; Caspari, Daniela; Grünewald, Andreas; Küster, Lutz; Nold, Günter; Vollmer, Helmut J. & Zydariß, Wolfgang (2008), Kompetenzorientierung, Bildungsstandards und fremdsprachliches Lernen. Herausforderungen an die Fremdsprachenforschung. Positionspapier von Vorstand und Beirat der DGFF. *Zeitschrift für Fremdsprachenforschung* 19: 2, 163-186.
- Jäger, Daniela J. (2012), Schulklima, Selbstwirksamkeit und Arbeitszufriedenheit aus Sicht der Lehrpersonen und Schüler/-innen in Hessen und Bremen. In: Maag Merki, Katharina (Hrsg.) (2012), 65-94.

- Jürges, Hendrik; Schneider, Kerstin; Senkbeil, Martin & Carstensen, Claus H. (2012), Assessment drives learning. The effect of central exit exams on curricular knowledge and mathematical literacy. *Economics of Education Review* 31: 1, 56-65.
- Klein, E. Dominique; Krüger, Mirko; Kühn, Svenja M. & van Ackeren, Isabell (2014), Wirkungen zentraler Abschlussprüfungen im Mehrebenensystem Schule. Eine Zwischenbilanz internationaler und nationaler Befunde und Forschungsdesiderata. *Zeitschrift für Erziehungswissenschaft* 17: 1, 7-33.
- KMK = Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2003), *Bildungsstandards für die erste Fremdsprache (Englisch/Französisch) für den Mittleren Schulabschluss. Beschluss vom 4.12.2003* [Online: [http://www.kmk.org/fileadmin/veroeffentlichungen\\_beschluesse/2003/2003\\_12\\_04-BS-erste-Fremdsprache.pdf](http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2003/2003_12_04-BS-erste-Fremdsprache.pdf), 11.09.2015].
- KMK = Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2004), *Bildungsstandards für die erste Fremdsprache (Englisch/Französisch) für den Hauptschulabschluss. Beschluss vom 15.10.2004* [Online: [http://www.kmk.org/fileadmin/veroeffentlichungen\\_beschluesse/2004/2004\\_10\\_15-Bildungsstandards-erstefS-Haupt.pdf](http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_10_15-Bildungsstandards-erstefS-Haupt.pdf), 11.09.2015].
- KMK = Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2005), *Bildungsstandards der Kultusministerkonferenz. Erläuterungen zur Konzeption und Entwicklung* [Online: [http://www.kmk.org/fileadmin/veroeffentlichungen\\_beschluesse/2004/2004\\_12\\_16-Bildungsstandards-Konzeption-Entwicklung.pdf](http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Bildungsstandards-Konzeption-Entwicklung.pdf), 13.01.2016].
- KMK = Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (Hrsg.) (2010), *Konzeption der Kultusministerkonferenz zur Nutzung der Bildungsstandards für die Unterrichtsentwicklung*. Köln: Link.
- KMK = Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2012), *Bildungsstandards für die fortgeführte Fremdsprache (Englisch/Französisch) für die Allgemeine Hochschulreife. Beschluss der Kultusministerkonferenz vom 18.10.2012* [Online: [http://www.kmk.org/fileadmin/veroeffentlichungen\\_beschluesse/2012/2012\\_10\\_18-Bildungsstandards-Fortgef-FS-Abi.pdf](http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2012/2012_10_18-Bildungsstandards-Fortgef-FS-Abi.pdf), 13.01.2016].
- KMK = Kultusministerkonferenz (2015), KMK ermöglicht erstmals Blick auf den Abituraufgabenpool. Pressemitteilung 12.06.2015 [Online: <http://www.kmk.org/presse-und-aktuelles/meldung/kmk-ermoeeglicht-erstmal-blick-auf-den-abituraufgabenpool.html>, 26.09.2015].
- Köller, Olaf (2010), Bildungsstandards. In: Tippelt, Rudolf & Schmidt, Bernhard (Hrsg.) (2010), *Handbuch Bildungsforschung* (1. Aufl.). Wiesbaden: VS Verlag für Sozialwissenschaften, 529-548.
- Kühn, Svenja M. (2010), *Steuerung und Innovation durch Abschlussprüfungen?* Wiesbaden: VS Verlag für Sozialwissenschaften.
- Kühn, Svenja M. (2011), Exploring the use of statewide exit exams to spread innovation. The example of Context in science tasks from an international comparative perspective. *Studies in Educational Evaluation* 37: 4, 189-195.
- Latimer, David G. (2009), *Washback effects of the Cambridge Preliminary English Test at an Argentinean bilingual school*. Ph.D. dissertation, University of Georgia, Athens, GA [Online: [https://getd.libs.uga.edu/pdfs/latimer\\_david\\_g\\_200908\\_phd.pdf](https://getd.libs.uga.edu/pdfs/latimer_david_g_200908_phd.pdf), 30.09.2015].
- Lazaraton, Anne & Taylor, Lynda (2007), Qualitative Research Methods in Language Test Development and Validation. In: Fox, Janna D.; Wesche, Mari; Bayliss, Doreen; Cheng, Liying; Turner, Carolyn E. & Doe, Christine (Hrsg.) (2007), *Language testing reconsidered*. Ottawa: University of Ottawa Press, 113-130.
- Lin-Klitzing, Susanne; DiFuccia, David & Müller-Frerich, Gerhard (Hrsg.) (2013), *Zur Vermessung von Schule. Empirische Bildungsforschung und Schulpraxis*. Bad Heilbrunn: Klinkhardt.

- Lorenz, Ramona (2013), *Das Zentralabitur im Kontext der Bildungsgerechtigkeit. Schwierigkeit und Fairness der Abituraufgaben im Fach Englisch in NRW*. Münster: Waxmann.
- Maag Merki, Katharina (2012), Die Leistungen der Gymnasialistinnen und Gymnasialisten in Mathematik und Englisch. In: Maag Merki, Katharina (Hrsg.) (2012), 263-292.
- Maag Merki, Katharina (Hrsg.) (2012), *Zentralabitur. Die längsschnittliche Analyse der Wirkungen der Einführung zentraler Abiturprüfungen in Deutschland*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Madaus, George F. (1988), The Influence of Testing on the Curriculum. In: Tanner, Laurel N. (Hrsg.) (1988), *Critical Issues in curriculum*. Chicago, Ill.: University of Chicago Press, 83-121.
- Maier, Uwe (2010), Effekte testbasierter Rechenschaftslegung auf Schule und Unterricht. Ist die internationale Befundlage auf Vergleichsarbeiten im deutschsprachigen Raum übertragbar? *Zeitschrift für Pädagogik* 56: 1, 112-128.
- Matthes, Eva (2013), Kompetenzorientierter Unterricht. *Bildung und Erziehung* 66: 2, 121-125.
- Messick, Samuel (1989), Validity. In: Linn, Robert L. (Hrsg.) (1989), *Educational Measurement* (3. Aufl.). Washington, D.C.: American Council on Education, 13-103.
- Messick, Samuel (1996), Validity and washback in language testing. *Language Testing* 13: 3, 241-256.
- Meyerhöfer, Wolfram; Lin-Klitzing, Susanne; DiFuccia, David & Müller-Frerich, Gerhard (2013), Unterrichten – standardisiertes Testen – erziehen. Effekte der Standardisierung. In: Lin-Klitzing, Susanne; DiFuccia, David & Müller-Frerich, Gerhard (Hrsg.) (2013), 181-205.
- Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen (2015), *Konstruktionshinweise: Neue Aufgabenformate in den modernen Fremdsprachen Englisch, Französisch, Spanisch, Italienisch, Russisch, Niederländisch, Türkisch, Portugiesisch, Neugriechisch. Aufgabenart 1.1: Schreiben mit Leseverstehen (integriert) und einer weiteren Teilkompetenz (Sprachmittlung/Hör-/Hörsehverstehen) in isolierter Überprüfung* [Online: [https://www.standardsicherung.schulministerium.nrw.de/abitur-gost/getfile.php?file=3714\\_14.09.2015](https://www.standardsicherung.schulministerium.nrw.de/abitur-gost/getfile.php?file=3714_14.09.2015)].
- Munoz, Anna P. & Alvarez, Marta E. (2010), Washback of an oral assessment system in the EFL classroom. *Language Testing* 27: 1, 33-49.
- Nachtigall, Christof & Hellrung, Katrin (2013), Zur zeitlichen Entwicklung der Rezeption von Vergleichsarbeiten. *Empirische Pädagogik* 27: 4, 423-441.
- Neumann, Marko; Trautwein, Ulrich & Nagy, Gabriel (2011), Do central examinations lead to greater grading comparability? A study of frame-of-reference effects on the University entrance qualification in Germany. *Studies in Educational Evaluation* 37: 4, 206-217.
- Nold, Günter (2015), Englisch: Fremdsprachendidaktische Konzepte der Unterrichtsentwicklung. In: Rolff, Hans-Günter (Hrsg.) (2015), 413-424.
- Otto, Stephan & Kühn, Svenja M. (2014), Zwischen Tradition und Innovation. Eine Analyse zentraler Prüfungsaufgaben im Fach Deutsch im nationalen Vergleich. *Schulpädagogik heute* 5: 9, 1-16.
- Patton, Michael Quinn (2002), *Qualitative research and evaluation methods* (3. Aufl.). Thousand Oaks: SAGE.
- Porsch, Raphaela; Tesch, Bernd & Köller, Olaf (Hrsg.) (2010), *Standardbasierte Testentwicklung und Leistungsmessung. Französisch in der Sekundarstufe I*. Münster: Waxmann.
- Porter, Stephen R. (2008), Email Survey. In: Lavrakas, Paul (Hrsg.) (2008), *Encyclopedia of Survey Research Methods*. Thousand Oaks: SAGE, 231-233.
- Quetz, Jürgen & Vogt, Karin (2009), Nationale Bildungsstandards für die erste Fremdsprache: Sprachenpolitik auf unsicherer Basis. *Zeitschrift für Fremdsprachenforschung* 20: 1, 63-89.

- Quetz, Jürgen (2012), Bildungsstandards für die erste Fremdsprache Englisch. Ihre Umsetzung in einem kompetenzorientierten Unterricht. In: Bauer, Karl-Oswald (Hrsg.) (2012), *Effektive Bildung. Zur Wirksamkeit und Effizienz pädagogischer Prozesse*. Münster: Waxmann, 183-197.
- Read, John & Hayes, Belinda (2003), The impact of IELTS on preparation for academic study in New Zealand. *IELTS International English Language Testing System Research Reports* 4, 153-206.
- Richter, Dirk; Böhme, Katrin; Becker, Michael; Pant, Hans Anand & Stanat, Petra (2014), Überzeugungen von Lehrkräften zu den Funktionen von Vergleichsarbeiten. Zusammenhänge zu Veränderungen im Unterricht und den Kompetenzen von Schülerinnen und Schülern. *Zeitschrift für Pädagogik* 60: 2, 225-244.
- Rolff, Hans-Günter (Hrsg.) (2015), *Handbuch Unterrichtsentwicklung* (1. Aufl.). Weinheim: Beltz.
- Rossa, Henning (2007), Die Erfassung des fremdsprachlichen Hörverstehens: Zur Validierung eines Testkonstrukts. In: Doff, Sabine & Schmidt, Torben (Hrsg.) (2007), *Fremdsprachenforschung heute. Interdisziplinäre Impulse, Methoden und Perspektiven*. Frankfurt am Main: Peter Lang, 141-153.
- Rossa, Henning (2012), *Mentale Prozesse beim Hörverstehen in der Fremdsprache. Eine Studie zur Validität der Messung sprachlicher Kompetenzen*. Frankfurt am Main: Peter Lang.
- Rupp, André A.; Vock, Miriam; Harsch, Claudia & Köller, Olaf (2008), *Developing standards-based assessment tasks for English as a first foreign language: Context, processes, and outcomes in Germany*. Münster: Waxmann.
- Saif, Shahrzad (2006), Aiming for positive washback. A case study of international teaching assistants. *Language Testing* 23: 1, 1-34.
- Schwerdt, Guido & Wößmann, Ludger (2015), *The information value of central school exams* [Online: [http://www.cesifo-group.de/DocDL/cesifo1\\_wp5404.pdf](http://www.cesifo-group.de/DocDL/cesifo1_wp5404.pdf), 13.01.2016].
- Spratt, Mary (2005), Washback and the classroom: the implications for teaching and learning of studies of washback from exams. *Language Teaching Research* 9: 1, 5-29.
- Tepe, Marlis (2014), Zu Viele Tests und Leistungsvergleiche? *Pädagogik* 66: 9, 48.
- Tesch, Bernd; Hammerstein, Xenia von; Stanat, Petra & Rossa, Henning (Hrsg.) (erscheint 2016), *Bildungsstandards aktuell: Englisch/Französisch in der Sekundarstufe II*. Braunschweig: Diesterweg.
- Tesch, Bernd; Leupold, Eynar & Köller, Olaf (Hrsg.) (2008), *Bildungsstandards Französisch: konkret. Sekundarstufe I: Grundlagen, Aufgabenbeispiele und Unterrichts Anregungen* (1. Aufl.). Berlin: Cornelsen Scriptor.
- Timm, Johannes-Peter & Vollmer, Helmut J. (Hrsg.) (2006), *Fremdsprachenlernen und Fremdsprachenforschung. Kompetenzen, Standards, Lernformen, Evaluation: Festschrift für Helmut Johannes Vollmer*. Tübingen: Narr.
- van Ackeren, Isabell; Block, Rainer; Klein, Esther Dominique & Kühn, Svenja Mareike (2012), The Impact of State-Wide Exit Exams in Germany. A Descriptive Case Study. *Education Policy Analysis Archives* 20, 1-28.
- Wall, Dianne & Alderson, J. Charles (1993), Examining washback. The Sri Lankan Impact Study. *Language Testing* 10: 1, 41-69.
- Wall, Dianne & Horak, Tanja (2011), *The Impact of Changes in the TOEFL Exam on Teaching in a Sample of Countries in Europe: Phase 3, The Role of the Coursebook – Phase 4: Describing Change*. Princeton, New Jersey: ETS.
- Wiater, Werner (2013), Kompetenzorientierung des Unterrichts. Alter Wein in neuen Schläuchen? Anfragen seitens der Allgemeinen Didaktik. *Bildung und Erziehung* 66: 2, 145-161.
- Zimmer-Müller, Michael & Hosenfeld, Ingmar (2013), Zehn Jahre Vergleichsarbeiten: Eine Zwischenbilanz aus verschiedenen Perspektiven. Themenheft *Empirische Pädagogik* 27: 4.