

Die Auswirkungen von standardisierten Testverfahren auf den Unterricht und das Lernen im Fach Englisch auf der Sekundarstufe I

Claudia Mewald¹

Zusammenfassung

Kompetenzorientierung, die Einführung der E8 Bildungsstandards und standardisierter Überprüfungen, ein neuer Lehrplan für Lebende Fremdsprachen, die Implementierung der Neuen Mittelschule und die damit verbundene neue pädagogische Relevanz der Differenzierung von Unterricht und Beurteilung sowie die Chancen und Herausforderungen des kooperativen Unterrichtens haben Auswirkungen auf den Englischunterricht der Sekundarstufe I. Entsprechend verändert sind die Anforderungen an die Lehrenden, welche diese Neuerungen in der Praxis umsetzen sollen.

Diese Washbackstudie beschreibt ausgehend von einer Ist-Stand Erhebung aus dem Jahr 2009 die Auswirkung der Implementierung der E8 Bildungsstandards, des neuen Lehrplans sowie der standardisierten E8 BIST Überprüfungen auf den Englischunterricht der Sekundarstufe I.

The impact of standardised testing on teaching and learning in English at secondary school level

This contribution explores the impact of several innovations of teaching and learning in English as a foreign language at secondary school level. It discusses the washback of competence orientation, the implementation of E8 standards and standardised tests, a new curriculum for foreign languages, and the New Middle School including the need for differentiation of teaching and assessment. Moreover, it explores the changing requirements for teachers, who are putting these innovations into practice.

Taking the results from an online survey in 2009 as a starting point, the washback of E8 testing on teaching and learning is fleshed out.

Schlüsselwörter:

E8 Bildungsstandards
Standardisierte Überprüfung
Schularbeiten & Leistungsbeurteilung

Keywords:

E8 standards
Standardised testing
Classroom testing & assessment

1 Einleitung

Im Jahr 2009 wurden in Österreich Bildungsstandards für Fremdsprachen in der Sekundarstufe I (E8 BIST) erlassen. Ziel dieser Standardisierung war es, durch Kompetenzorientierung Veränderungen und Innovationen im Fremdsprachenunterricht herbeizuführen. Dies inkludierte das Vorhaben, die Lernergebnisse der Schüler/innen vergleichbar zu machen und verbindliche Ziele für die achte Schulstufe zu formulieren (BIFIE, 2012).

Nach einem fünfjährigen Zeitraum der Entwicklung und Pilotierung wurde eine Reihe von Kompetenzbeschreibungen beschlossen, welche auf Deskriptoren aus dem Gemeinsamen Europäischen

¹ Pädagogische Hochschule Niederösterreich, Mühlgasse 67, 2500 Baden.

Korrespondierende Autorin. E-Mail: claudia.mewald@ph-noe.ac.at

Referenzrahmen (GERS; Europarat, 2001) beruhen. Diese Kann-Beschreibungen wurden auf die lokalen und curricularen Bedürfnisse zugeschnitten, um der österreichischen Bildungswelt zu entsprechen. Zusätzlich listet der aktuelle Lehrplan für zehn- bis vierzehnjährige Lernende in der fünften bis achten Schulstufe ebenfalls Ziele auf, welche auf GERS-Deskriptoren von A1 bis B1 basieren. Der erwartete Prozess der Innovation und Veränderung im Lehren, Lernen und Beurteilen von Fremdsprachen, welcher durch die Einführung der Standards erwartet wurde, sollte neben besseren fremdsprachlichen Kompetenzen auch die Aus-, Fort- und Weiterbildung der Lehrkräfte betreffen und den „Wunsch nach einem kontinuierlichen Verbesserungsprozess von Unterricht und Schule“ (BIFIE, 2011, S. 8) verwirklichen.

Ziel dieses Top-Down-Prozesses organisierter pädagogischer Veränderung für geregelten Output war es, Lernziele zu definieren, welche die Kompetenzen, die Lernende nach acht Jahren Fremdsprachenunterricht erreicht haben sollten, festlegen. Zudem war geplant, durch standardisierte Testungen herauszufinden, in welchem Ausmaß die definierten Ziele erreicht worden waren. Daher entwickelten das Language Testing Centre der Alpen-Adria-Universität Klagenfurt und das BIFIE Salzburg einen Low-Stakes Test, der alle vier Jahre zum Zwecke der Systemüberwachung durchgeführt werden sollte. Die Ergebnisse sollten Schulen und Lehrkräfte mit Informationen zur Auswirkung ihrer Arbeit versorgen und Erkenntnisse für die gezielte Planung von Curricula in der Erstausbildung und Fortbildung von Lehrkräften liefern, jedoch keinen Einfluss auf die Noten der Testteilnehmer/innen haben. Eine Baselinestudie ging 2009 dem ersten standardisierten Test (E8-Test), welcher im April 2013 vom BIFIE Salzburg durchgeführt wurde, voraus. In der ersten Standardsüberprüfung wurden beinahe 76.728 Schüler/innen in 1.410 Schulen in den Bereichen Hörverständnis, Leseverständnis und Schreiben getestet; 2.744 Schüler/innen wurden im Sprechen überprüft (BIFIE, 2014, S. 13 und 68).

Ursprünglich wurden die E8-Standardtests entwickelt, um zusätzlich zu dem übergeordneten Ziel der Systemüberwachung auch diagnostisches Feedback zu liefern. Daher waren die Testspezifikationen der durchgeführten E8-Baselinestudie darauf ausgerichtet, dass die Ergebnisse beide Zwecke erfüllen könnten (Gassner, Mewald und Sigott, 2007; Mewald, Gassner und Sigott, 2007; Sigott, Gassner et al., 2007; Gassner, Mewald und Sigott, 2008; Mewald, Gassner und Sigott, 2009). Das Vorhaben, die Leistungen von Schülern und Schülerinnen am Ende des achten Jahres zu testen, erforderte die Entwicklung von Testkonstrukten, welche es ermöglichten, die erwartete kommunikative Kompetenz der Schüler/innen den curricularen Erwartungen entsprechend überprüfen und diagnostisches Feedback geben zu können. In den produktiven Kompetenzen wurde dies durch die Beschreibung der Konstrukträume (Construct Spaces), welche die Themen, Textarten und kommunikativen Funktionen definieren, erreicht. Des Weiteren wurden die Kriterien für die Bewertung der Leistungen der Testteilnehmer/innen in Form von Bewertungs- und Beurteilungsskalen dargestellt (Gassner, Mewald, Brock, Lackenbauer, & Siller, 2011; Mewald, Gassner, Lackenbauer, Brock, & Siller, 2013). In den rezeptiven Fertigkeiten definierten die Konstrukträume die Textarten und die Zielgruppen der Lese- und Hörtexte sowie die zugrundeliegenden Strategien, welche Hör- oder Lesekompetenzen entwickeln sollten (Gassner, Mewald, & Sigott, 2007; Mewald, Gassner, & Sigott, 2007). Zusammen mit der Veröffentlichung der verwendeten Testformate wurde angenommen, dass die Überprüfung der Bildungsstandards für die Fremdsprache Englisch auf der achten Schulstufe positive Auswirkungen (Washback) zeigen und somit zur erhofften Innovation und Veränderung des Lehrens und Lernens von Fremdsprachen führen würde.

Anders als in der Baselinestudie wurde das Konzept der Diagnose in der ersten E8-Überprüfung im Jahr 2013 nicht verfolgt. Die Testteilnehmer/innen erhielten zwar Rückmeldung über inhaltliche Aspekte ihrer Performanzen, auf diagnostisches Feedback wurde für diese Zielgruppe jedoch verzichtet. Lehrer/innen erhielten detailliertere Informationen zum Abschneiden ihrer Gruppe(n), wodurch sie gegebenenfalls ihr zukünftiges Unterrichten adaptieren konnten. Schulen wurden genaue Auskünfte mit dem Fokus auf Qualitätssicherung gegeben, und Institutionen für Lehrerbildung erhielten Impulse zur Gestaltung von Programmen, um die erkennbaren Lücken als Antwort auf die Beschreibung von schulischen Leistungen schließen zu können (BIFIE, 2012).

Im Sinne der Ziele der E8-Überprüfung vergleicht dieser Beitrag Schularbeiten im Fach Englisch in zwei Phasen (2004-2011; 2012-2014) mit einer früheren Studie, die 1992 durchgeführt wurde (Gerngross und Puchta, 1992). Zudem werden Daten aus Beobachtungen, Interviews und Dokumentenrecherche verwendet, um den Stand der Beurteilung in Hauptschulen, Neuen Mittelschulen und Allgemeinbildenden Höheren Schulen zu beschreiben.

2 Theoretische Überlegungen zur Beschreibung des Washbacks

Theoretische Überlegungen zum Washback der E8 Bildungsstandards auf das Lehren und Lernen von Englisch als Fremdsprache erfordern eine Auseinandersetzung mit den Annahmen und Überzeugungen der Lehrenden und Lernenden darüber, wie Fremdsprachen erworben werden, sowie eine Beschreibung der bestehenden Unterrichtspraxis. Im Kontext der österreichischen Bildungslandschaft, welche laut Lehrplan seit 1999/2000 (BMBWF, 2000) eine kommunikative Orientierung in Bereich der Fremdsprachenerziehung verfolgt, muss sich eine Auseinandersetzung mit der gängigen Praxis deshalb mit dem Kommunikativen Ansatz des Fremdsprachenerwerbs beschäftigen. Zusätzlich ist im Kontext der Kompetenzorientierung der Begriff der Kompetenz, welcher im Fremdsprachenunterricht als „Kommunikative Kompetenz“ (Hymes, 1972) schon viel länger grundlegend diskutiert wird als im Rest des Fächerkanons, klarzustellen. Des Weiteren muss eine Diskussion darüber geführt werden, wie die Faktoren der kommunikativen Kompetenz und der Standardisierung durch Low-Stakes-Tests beeinflusst werden können.

Die Annahme, dass eine landesweite Überprüfung der vorhandenen Kompetenzen Auswirkungen darauf haben würde, wie Lehrende, Lernende und Eltern das Lehren und Lernen von Fremdsprachen wahrnehmen und dass Lehrer/innen in Hinblick auf die Überprüfung testorientiert unterrichten würden, wenn sie mit dem Inhalt und Format vertraut wären, ist in der Literatur verankert (Alderson & Wall, 1993; Bachman & Palmer, 2010; Brown & Abeywickrama, 2010; Cheng, 2005; Cheng, Watanabe, & Curtis, 2004; Green, 2007; Wall, 2005; Watanabe, 1997). Daher konnten pädagogische Veränderungen bei der Vermittlung von Fremdsprachen sowie die damit verbundenen Auswirkungen auf den Fremdsprachenerwerb erwartet werden (van der Branden, 2011).

2.1 Kompetenzorientierung

Der Begriff der Kompetenz sowie Modelle des kompetenzorientierten Unterrichts fanden in den 1970er Jahren ihren Weg in das Bildungswesen. Gleichzeitig und sogar vor Pädagogen der Allgemeinbildung wie zum Beispiel Klafki (1993; 1963), Habermas (1979; 1987) oder Roth (1971) entwickelten Sprachforscher/innen linguistische Modelle kommunikativer Kompetenz (Canale & Swain, 1980; Canale, 1983; Hymes, 1972; Swain & Lapkin, 1995).

In der jüngsten Vergangenheit ist Kompetenz zu einem Begriff geworden, der sehr häufig in Verbindung mit Standards und der Standardisierung des Testens sprachlicher Leistungen verwendet wird (Beck & Klieme, 2007; Rickheit & Strohner, 2010; Shohamy E., 1996). Allerdings verwenden Forscher/innen die Begriffe Kompetenz und Leistung auf unterschiedliche Weise. Aufgrund dessen scheint eine gründliche Definition dieser Begriffe notwendig (Brown, 1996; Celce-Murcia & Dörnyei, 1995), wenn die aktuelle Vorstellung von kommunikativer Kompetenz und von Standards, die sie beschreiben sollen, festgelegt werden soll. Dabei müssen frühe Modelle linguistischer Kompetenz, die auf eher eingeschränkte Konzepte von Kompetenz und Leistung basiert sind, durch Aspekte kommunikativer und pragmatischer Kompetenz (Lyons, 1996) ergänzt werden. Das Werk „On communicative competence“ (1972) von Dell Hymes wird allgemein als Fundament für alles, was zu diesem Thema folgte, gesehen, wie zum Beispiel die neueren Modelle von Canale & Swain (1980), Canale (1983) oder Celce-Murcia & Dörnyei (1995).

Während kommunikative Kompetenz allgemein als das Wissen und die Fähigkeit, eine Sprache erfolgreich zum Kommunizieren zu verwenden, verstanden wird, richten kompetenzorientierte Ansätze zusätzliche Aufmerksamkeit auf die funktionellen, interaktionalen und sozialen Aspekte einer Sprache in realen Kontexten (Richards & Rodgers, 2014). Zusätzlich zu diesen Aspekten legt Kompetenzorientierung einen besonderen Schwerpunkt auf Beurteilung. Kriterien und Indikatoren in Form von Kann-Beschreibungen (Can Do-Statements) sollen dabei helfen, die Leistungen der Lernenden und die Lernergebnisse zu beschreiben. Ferner bilden sie die Basis für die Beurteilung von Sprachkompetenzen hinsichtlich der wesentlichen Fähigkeiten des Hörverständnisses, des Lesens, des Sprechens und des Schreibens. Neben den Sprachkenntnissen werden das Wissen, die Einstellungen und die Verhaltensweisen, die für die erfolgreiche Durchführung von konkreten Aufgabenstellungen oder Tätigkeiten notwendig sind, als grundlegende Bestandteile kommunikativer Kompetenz gesehen, welche immer als sprachliche Leistung zum Vorschein kommt (Richards & Rodgers, 2014; Weinert, 2001). In diesem Zusammenhang ist Leistung die „Anwendung der Kompetenz oder des Wissens über die Regeln einer Sprache auf tatsächliche Kommunikation“ (Davies et al., 1999, S. 143). Diese sprachliche Konzeptualisierung kommunikativer Kompetenz ist in der allgemeineren Definition der Kompetenzorientierung,

auf der die österreichische Standardsbewegung beruht, enthalten. Weinert (2001) beschreibt Kompetenz als kognitive Fähigkeit kombiniert mit motivationaler, volitionaler und sozialer Bereitschaft, Probleme in verschiedenen Situationen erfolgreich und verantwortlich zu lösen. Eine linguistische Definition davon, was beurteilt werden kann, unterscheidet sich von Weiners Konzept in der Tatsache, dass eine funktionelle und interaktionale Perspektive der Natur der Sprache, welche von kompetenzorientierter Fremdsprachenausbildung eingenommen wird, Kommunikation als mit dem sozialen Kontext, in dem sie verwendet wird, verbunden sieht. Sprache wird als Medium der Interaktion und Kommunikation zwischen Menschen gesehen, die bestimmte Ziele und Absichten erreichen wollen, besonders in Situationen, in denen sie gewisse Rollen mit Sprachkenntnissen erfüllen müssen, die für den entsprechenden Kontext vorausgesehen oder bestimmt werden müssen (Richards & Rodgers, 2014). Somit basiert jegliche Bewertung der kommunikativen Kompetenz auf sprachlichen Leistungen, welche als Performanzen beobachtbar werden. Die Bewertung stellt folglich eine Annahme der zugrundeliegenden Kompetenz auf der Basis der Performanz dar.

2.2 E8-Standards, Innovation und Veränderung

Die zweite Annahme, welche die Idee unterstützt, dass sich Innovation und Veränderung ohne Weiteres mit der Umsetzung und Standardisierung entwickeln würden, stützte sich auf den Umstand, dass der GERS (Europarat, 2010) und die dazugehörigen Deskriptoren bereits einige Jahre verfügbar gewesen waren, als die E8-Standards eingeführt wurden. Das führte zu dem Schluss, dass Lehrer/innen für die Umsetzung der Standards bereit wären, da sie bereits mit einem ähnlichen System vertraut waren. Die Annahme, dass sich Washback beim Überprüfen unweigerlich einstellen würde und dass Lehrer/innen bereitwillig in Richtung Testung unterrichten würden, verließ sich auf die Vorstellung, dass das Lehren und Testen von Englisch als Fremdsprache (EaF) in Österreich bereits aneinander sowie an kommunikative und kriterienorientierte Lehrgrundsätze und Bewertungsverfahren angeglichen waren.

Allerdings berücksichtigten diese Annahmen nicht, wie die meisten Lehrer/innen tatsächlich den Englischunterricht planten und umsetzten. Diese Vermutungen berücksichtigten auch nicht, wie und was die Schüler/innen wirklich lernten. Zu einem gewissen Grad schien die österreichische Situation zu Beginn der offiziellen Umsetzung und Standardisierung im Jahre 2009 das widerzuspiegeln, was aus anderen Kontexten bekannt war: Das Reformieren garantiert nicht automatisch Verbesserung (Stigler & Hiebert, 1999, S. xi). Zudem war zu Beginn der Standardbewegung nicht klar, wie Lehrer/innen auf den neuen Ansatz der kompetenzorientierten Fremdsprachenausbildung reagieren würden. Würden sie ausreichend mit den Methoden vertraut sein, wären sie bereit, die Verfahren anzuwenden, die für ihr eigenes Unterrichten empfohlen wurden?

In der Tat schienen diese ursprünglichen Annahmen die Umstände zu vernachlässigen, dass Lehrer/innen, die sich neue Überzeugungen aneignen müssen, die ihren eigenen Überzeugungen und Kenntnissen widersprechen, ablehnend reagieren würden (Borg, 2006; Richards & Rodgers, 2014). Zusammen mit den Veränderungen durch vorgeschriebene Kompetenzorientierung und der zusätzlichen Herausforderung von zunehmend diversen Klassen in Bezug auf ethnische, sprachliche und kulturelle Hintergründe (Organization for Economic Co-operation and Development, 1994) mussten Lehrer/innen der Sekundarschulbildung auch mit dem Ende der Leistungsgruppen und der Einführung heterogener Klassen in Neuen Mittelschulen zurechtkommen. Dies erforderte eine weitaus breitere Dimension von Innovation, was bedeutete, dass Lehrer/innen plötzlich mit mindestens drei erheblichen Veränderungsbewegungen konfrontiert waren. Vielen Lehrkräften schien es daher nur schwer möglich, alle Neuerungen in die Tat umzusetzen.

Folglich wurde die Annahme, dass die Überprüfung der Bildungsstandards automatisch die Einstellungen, Verhaltensweisen und die Motivation zur Kompetenzorientierung im Lehren, Lernen und Beurteilen beeinflussen würde, zunehmend in Frage gestellt. Eine zusätzliche Herausforderung zur beabsichtigten Veränderung war die Tatsache, dass Washback am einfachsten geschieht, wenn der Test für die Zukunft der Schüler/innen sehr wichtig ist, und wenn die Erfolgsquoten als Maßstab für den Erfolg von Lehrer/innen verwendet werden (Buck, 2001), und wenn er einen wesentlichen Einfluss auf das Leben der Testteilnehmer/innen sowie wichtige Folgen hat (Shohamy E. , 1992; Shohamy, Donitsa-Schmidt, & Ferman, 1996). Low-Stakes Tests ohne unmittelbare Auswirkung auf die Beurteilung der Testteilnehmer/innen oder Tests ohne die oben genannten wichtigen Folgen, welche die E8-Überprüfungen nicht ausweisen können, werden vermutlich nur geringen Washback aufweisen und daher kaum dazu geeignet sein, das Verhalten von Lehrern/Lehrerinnen und Lernenden zu verändern. Diese Tatsache hatte zur Folge, dass die Erwartungen an den Washback der E8-Überprüfung nicht hoch sein konnten. Dennoch war die Vorstellung, dass standardisierte

Testverfahren Auswirkungen auf das Lehren, Lernen und Beurteilen haben, das Fundament für jede Investition und Leistung, die im Kontext der E8 durchgeführt wurden.

2.3 E8-Überprüfung und Washback

Der theoretische Hintergrund dieser Untersuchung bezieht sich auf wesentliche Definitionen (Shohamy E., 1993), Hypothesen (Alderson & Wall, 1993) und Dimensionen (Watanabe, 1997; Watanabe, 2004) des Washbacks. Somit versucht die Studie die Beschreibung des Washbacks in Bezug auf seine *allgemeinen* oder *spezifischen Wirkungen* auf das Lehren, Lernen und Beurteilen (*von Beurteilung gesteuertes Lehren*) sowie die *Intensität* dieser Wirkungen zu umspannen. Ferner betrachtet sie die *Anpassung des Lehrplans* und die *systembezogene Validität* dahingehend, dass sie die Verbindung zwischen dem Beurteilen und dem Lehrplan sowie der Integration der E8 in das Bildungssystem beurteilt. Hierbei werden *Intensität, Dauer, Absicht* und *Wert* der vorgenommenen Maßnahmen betrachtet.

Alderson & Wall (1993) lieferten eine ausführliche Reihe an Hypothesen für die Beobachtung und Beurteilung von Washback, welche in zahlreichen Studien bereits ausgiebig erprobt wurden. Insbesondere in der Erstellung von Untersuchungen, die darauf abzielen, die Auswirkungen des Testens auf das Lehren und Lernen zu ermitteln, waren diese bedeutsam. Allerdings erforschten Untersuchungen zum Washback häufig die Ansichten und Überzeugungen von Lehrern/Lehrerinnen ohne einen direkten Blick in ihre Unterrichtspraxis zu werfen. Dies führte daher oft zu einem unscharfen Bild, welches Abweichungen zwischen den Aussagen von Lehrern/Lehrerinnen über ihren Unterricht und ihre wirkliche Unterrichtsweise aufwies (Wall, 2005). Aus diesem Grund schien es wichtig, die aktuelle Studie so zu gestalten, dass die Ansichten und Überzeugungen der Lehrenden und Lernenden sowie ihre Vorgehensweisen berücksichtigt werden.

3 Design und Ablauf der Untersuchung

Diese „Mixed-Methods“ Studie wurde von 2009 bis 2014 in drei Phasen und auf drei Stufen umgesetzt. In der Phase 1 (2009) wurde der Status quo der Unterrichtspraxis im Englischunterricht durch eine österreichweite Online Befragung erhoben. In den Phasen 2 und 3 (2010-2014) folgten weitere Befragungen, Beobachtungen und Interviews. Die Stufe 1 umfasste die *organisierenden Einrichtungen* und Erneuerer (z.B. das Unterrichtsministerium, Schulbehörden etc.). Die *ausführenden Einrichtungen* und Erneuerer, wie zum Beispiel Universitäten, Pädagogische Hochschulen, Lehrbuchautoren und -autorinnen, Herausgeber/innen etc. bildeten das Sample der Stufe 2. Die Stufe 3 umfasste die *umsetzenden Einrichtungen* und Erneuerer wie Schulleiter/innen, Lehrer/innen, Schüler/innen und Eltern. Während in der Phase 1 alle Lehrkräfte, die 2009 EaF unterrichteten, angesprochen wurden, waren in den Stufen 2 und 3 die Teilnehmer/innen der Studie interessierte und freiwillige Lehrkräfte sowie Experten und Expertinnen. Bei der Auswahl der Lehrkräfte wurden absichtlich solche Personen angesprochen, von denen eine Implementierung der Bildungsstandards erwartet wurde, weil sie Aus- und/oder Fortbildungen zu dem Thema besucht hatten. Zusätzlich zu den an den Bildungsstandards interessierten und informierten Lehrkräften wurden in allen Schulen auch mindestens ebenso viele Lehrkräfte beobachtet und befragt, die keinen direkten Bezug zu den Bildungsstandards durch Aus- oder Fortbildungen hatten.

Das Forschungsinteresse der verschiedenen Stufen wird in Abbildung 1 dargestellt.



Abbildung 1: Forschungsinteresse der drei Forschungsstufen.

Die Forschungsteilnehmer/innen an den organisierenden Einrichtungen berichteten über die beabsichtigten Ziele und die erwarteten Bedürfnisse der Lehrenden. Zudem wurden sie gebeten, Auskunft über aktuelle und zukünftige Ziele für Lehrerausbildung und Entwicklung zu geben. Neben Innovationen aus diesem Bereich wurden Dokumente gemäß den Forschungszielen analysiert. Teilnehmer/innen aus ausführenden Einrichtungen wurden gebeten, über die unterstützenden Maßnahmen zu berichten, die bereits umgesetzt worden waren oder gerade umgesetzt wurden sowie über zukünftige Ziele. Die Teilnehmer/innen aus umsetzenden Einrichtungen beschrieben die von ihnen wahrgenommene Wirkung der Standardisierung auf das Lehren, Lernen und die Beurteilung. Beobachtungen mit dem Fokus auf das Lehren und Lernen wurden österreichweit an Hauptschulen, Neuen Mittelschulen und Allgemein Bildenden Höheren Schulen durchgeführt. In jeder Schule wurden Lehrkräfte an ein bis zwei aufeinanderfolgenden Tagen in ein bis zwei Unterrichtssequenzen beobachtet.

Das theoretische Konstrukt für die Datenerhebung und -analyse stützte sich auf zehn der fünfzehn Hypothesen zum Washback von Anderson und Wall (1993) und auf Untersuchungen von Cheng (2005), Wall (2005), Watanabe (2004), Shohamy (1993) und Green (2007). Die Forschungsergebnisse von Wall zur Diskrepanz zwischen den Aussagen von Lehrern/Lehrerinnen über ihren Unterricht und ihr eigentliches Verhalten im Unterricht (Wall, 2005, S. 14) ließen einen Schwerpunkt auf Beobachtung in der Überprüfung von Untersuchungs- und Interviewdaten durch Triangulation wichtig erscheinen. Folgende Forschungsfragen wurden aus den zehn Washbackhypothesen nach Alderson und Wall (1993) abgeleitet:

	Hypothesis, Number
Have the E8 Standards and tests influenced teaching/ created washback?	1
Have the E8 Standards and tests influenced learning?	2
What do teachers teach?	3
How do teachers teach?	4
What do learners learn?	5
How do learners learn?	6
What is the rate and sequence of teaching?	7
What is the rate and sequence of learning?	8
What is the degree and depth of teaching?	9
What is the degree and depth of learning?	10
What is the attitude to methods/approaches/activities?	11
... and how have E8 Standards and tests influenced 3-11	

Tabelle 1: Forschungsfragen basierend auf den Washbackhypothesen nach Alderson und Wall (1993).

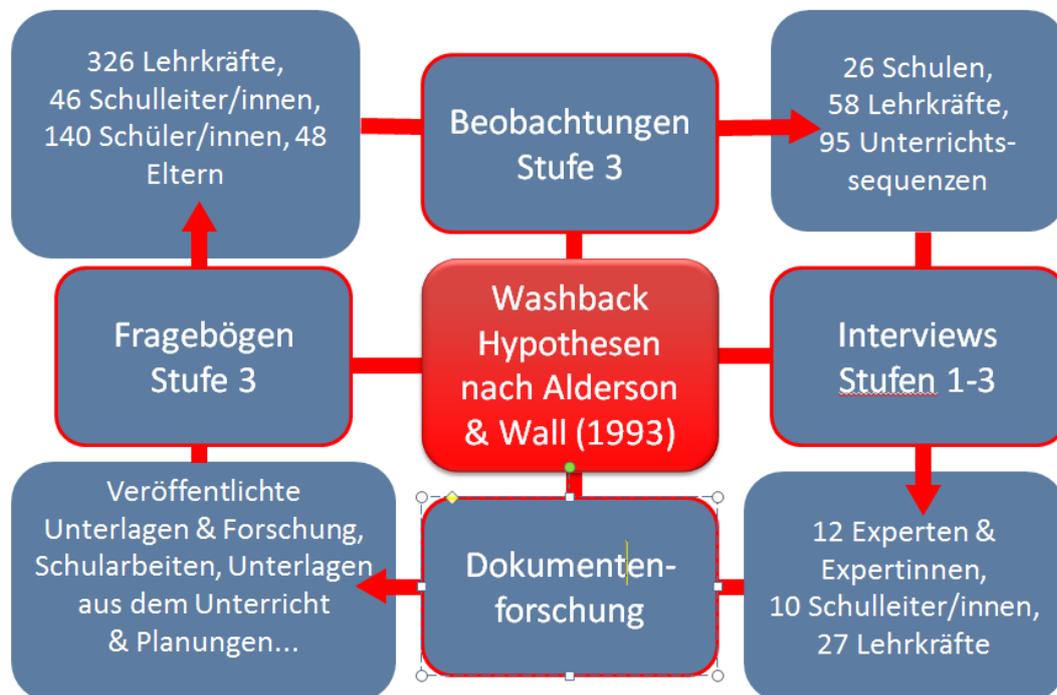


Abbildung 2: Triangulation der Datenerhebung und -analyse.

Die Datenerhebung erfolgte zyklisch und führte mehrmals zurück ins Feld. Dies war besonders deshalb wichtig, weil die Umsetzung der Neuen Mittelschulen mit heterogenen Leistungsgruppen beinahe gleichzeitig zur Einführung der standardisierten Überprüfung der Bildungsstandards stattfand und die Lehrer/innen, die in diesen Schulen Beurteilungen durchführten, kontinuierlich neue Anforderungen umzusetzen hatten (Schlichterle, Weißkopf-Prantner, & Westfahl-Greitner, 2013).

Die Datenanalyse folgte einem integrativen Design, welches konvergierende Triangulation auf der Basis des oben beschriebenen theoretischen Konstrukts zum Washback anwendete. Das bedeutet, dass quantitative Ergebnisse aus den Online-Untersuchungen sowie Daten aus den Fragebögen mit den Daten aus qualitativen Quellen (Beobachtungen, Interviews und Dokumentenanalyse), welche anhand der „Grounded Theory“ Methode (Strauss & Corbin, 1990) analysiert wurden, verglichen und kontrastiert wurden (Ziegler & Kang, 2016).

4 Ergebnisse

Als die Ergebnisse der E8-Überprüfung im Januar 2014 veröffentlicht wurden, waren die meisten Testteilnehmer/innen nicht mehr an derselben Schule, da nach dem achten Jahr der Übergang zur Sekundarstufe II oder zur Berufsausbildung stattfindet. Während die eigentlichen Testteilnehmer/innen auf ihre Ergebnisse mit einem Code, den sie beim Test erhalten hatten, online zugreifen konnten, wurden die Schulberichte Eltern vorgelegt, deren Kinder nicht getestet worden waren und nicht getestet werden würden, da die nächsten Tests für 2019 vorgesehen waren. Wie 2013 werden die Ergebnisse keine Auswirkungen auf die Beurteilung der Schüler/innen haben. Daher wird die E8-Überprüfung in Bezug auf ihre Wirkung auf das Lernen der Schüler/innen als Low Stakes-Test gesehen. Dennoch wurde spezifischer Washback als Reaktion auf das Lehren und Lernen im österreichischen Englischunterricht erwartet, da die Schulen und Lehrer/innen auf Verlangen kontrolliertes Feedback auf Basis der Testergebnisse der Schüler/innen erhielten. Durch den Status des Tests als Low Stakes-Test wurde außerdem erwartet, dass der Einfluss der Standardisierung auf das Lehren und Lernen gering wäre, doch es wurde als Folge ihres gesetzlichen Erlasses dennoch eine nachhaltige und beabsichtigte Auswirkung vorausgesagt.

4.1 Ein Einblick in den Englischunterricht in der Sekundarstufe I

Die Analyse der Unterrichtszeit, die den vier Fertigkeiten im EaF Unterricht gewidmet wurde, sollte Information über die Entwicklung des Lehrens zwischen 2009 und 2014 liefern. Es wurde erwartet, dass gemäß den neuen Lehrplanvorgaben allen Fertigkeiten die gleiche Aufmerksamkeit im Unterricht sowie in der Beurteilung zukommen würde.

Die Analyse der Unterrichtszeit basiert auf zwei Untersuchungen (Online Befragung und Befragungen zwischen 2010 und 2014) und auf 95 Unterrichtsbeobachtungen. Die markantesten Unterschiede im Laufe der fünf Jahre wurden in den Kompetenzbereichen Sprechen und Schreiben festgestellt. Im Vergleich zu der ersten Online-Untersuchung gaben Lehrer/innen in den Fragebögen an, dass sie die Unterrichtszeit für die Entwicklung der Schreibkompetenz von 18% auf 10% reduziert hatten, während die Unterrichtszeit für die Sprechfertigkeiten von etwa 25% auf 30% anstieg. Die Beobachtungen bestätigten dies allerdings nicht. Zudem gaben Lehrer/innen in Interviews nach der Beobachtung an, dass sie vor allem Änderungen am Lehren des Sprechens und Schreibens vorgenommen hatten und weniger an der Beurteilung. Sie bezogen sich auf Änderungen am Umfang und an der Intensität des Unterrichts (Hypothese 9, Alderson und Wall, 1993), jedoch nicht auf die Bewertung des Sprechens oder Schreibens. Die Unterrichtszeit, welche der Lese- und Hörkompetenz gewidmet wurde, stieg ebenfalls an. Die Zeit, die der Grammatik gewidmet wurde, blieb hingegen beinahe unverändert. Beobachtungen lieferten ein ausgeglichenes Bild von der Zeit, welche für die rezeptiven Fähigkeiten aufgewendet wurde, obwohl nur eine der 95 beobachteten Unterrichtseinheiten Hinweise für das explizite Unterrichten von Hörkompetenz lieferte. In allen anderen Einheiten war kein Fokus auf das Unterrichten von Lese- oder Hörkompetenz zu erkennen. Vielmehr wurde einer auf das „Tun“ beobachtet, wie es von Graham und Santos als „Doing“ beschrieben wurde (2015): Dabei wird zwar gelesen und zugehört, aber ohne diese Kompetenzen bewusst oder explizit zu fördern.

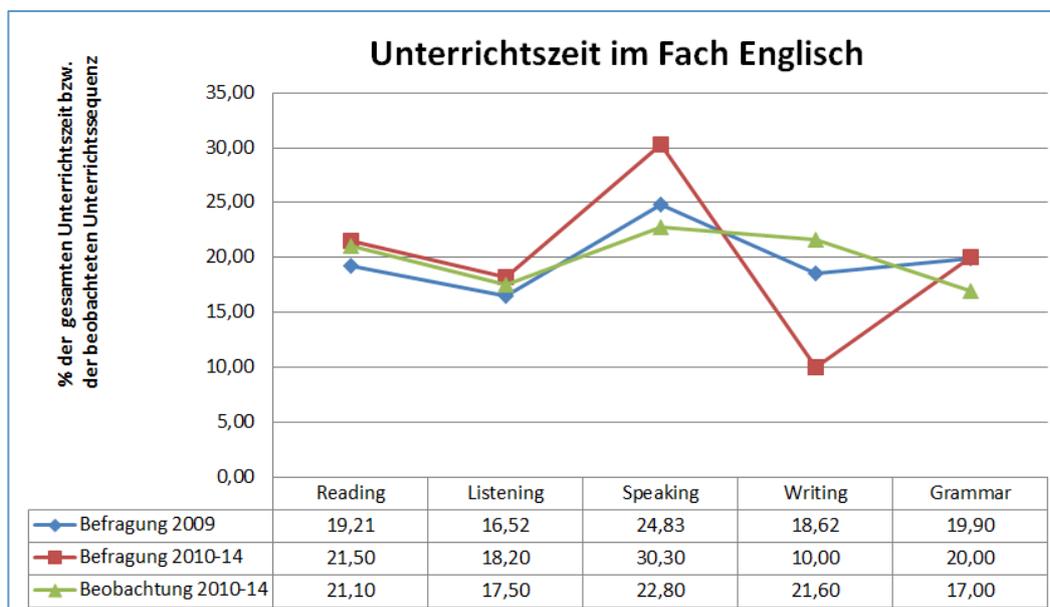


Tabelle 2: Unterrichtszeit.

In Übereinstimmung mit der Unterrichtszeit scheint das Sprechen den Untersuchungen entsprechend reichlich Washback in der Kategorie des Überprüfens erzeugt zu haben. Dies konnte jedoch nicht durch Beobachtungen, Interviews oder Dokumentenforschung bestätigt werden. Lehrer/innen sprachen mit Leidenschaft über die Herausforderungen, mit denen sie in der Beurteilung gesprochener Leistungen konfrontiert waren und beschrieben ihre Schwierigkeiten damit. Sie konnten jedoch nicht erklären, wie sie Redefertigkeiten tatsächlich überprüfen oder wie diese zu den Noten der Schüler/innen beitragen.

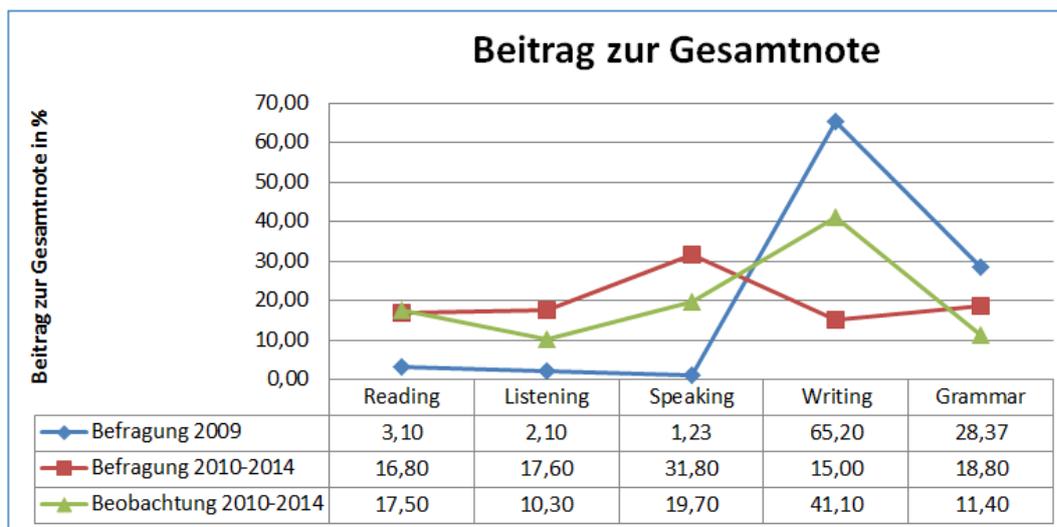


Tabelle 3: Beitrag der Teilbereiche zur Gesamtnote in %.

Im Vergleich zu der Untersuchung von 2009, wo Schreiben und Grammatik 93% der Gesamtnote ausmachten und wo Hör- und Lesekompetenz kaum zur Note beitrugen, kann gesagt werden, dass positiver Washback in Richtung eines homogeneren Beitrags aller vier Fähigkeiten zur Gesamtnote erreicht wurde (siehe Tabelle 3).

4.2 Schularbeiten

Zusätzlich zur kontinuierlichen Beurteilung, die den Interviewdaten entsprechend immer noch auf impressionistischen Notizen zu den Lernergebnissen der Schüler/innen basiert war, sollte eine vergleichende Untersuchung zum Aufbau von Schularbeiten Licht auf den Washback in dieser Dimension werfen.

Gerngross & Puchta hatten 1992 Schularbeiten untersucht, um mehr über den Übergang von der Sekundarstufe I zur Sekundarstufe II zu erfahren (Gerngross & Puchta, 1992). Ihre Ergebnisse zeigten einen deutlichen Schwerpunkt auf dem Schreiben, also Textproduktion, dem Befüllen von Lückentexten oder Übersetzung oder der Beantwortung von Fragen durch memorierte Inhalte, was sich insgesamt auf 69% belief. Schularbeiten, die vor der E8-Überprüfung im Zeitraum zwischen 2004 und 2011 gesammelt wurden, zeigen, dass das Überprüfen von Wortschatz und Grammatik im Vergleich zur Stichprobe aus 1992 von 38% auf 33,7% gesunken war. Ein Rückgang war auch bei den Schularbeiten, die zwischen 2012 und 2014 (siehe Tabelle 4) gesammelt wurden, zu verzeichnen. Allerdings wurden 31,6% der Gesamtpunktzahl der Prüfungen dennoch der Grammatik zugeschrieben, welche hauptsächlich in Form von Lückentexten überprüft wurde. Die späteren Schularbeiten zeigen die stetige Zunahme der Überprüfung der Lese- und Hörkompetenzen. Allerdings waren Schularbeiten aus dem Jahr 2014 dennoch weit entfernt vom Ziel einer gleichmäßigen Verteilung von Lese-, Hör- und Schreibkompetenz, wie es vom Lehrplan vorgegeben ist. Der schriftliche Teil, bestehend aus Grammatik, Textproduktion und der schriftlichen Wiedergabe von memorierten Inhalten, ging zwar auf 67% zurück, jedoch nicht auf das vorgeschlagene Drittel.

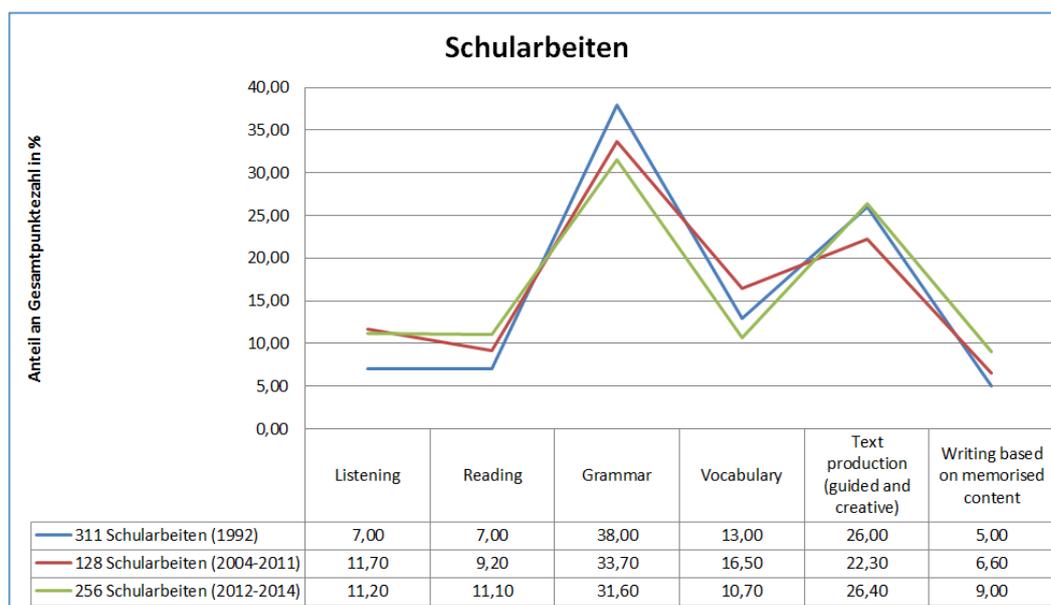


Tabelle 4: Schularbeiten, die zwischen 1992 und 2014 gesammelt wurden.

Was in der Analyse der Schularbeiten auffiel, war die verstärkte Überprüfung von scheinbar auswendig gelernten Inhalten. In Interviews gaben Lehrer/innen an, dass diese Inhalte inkludiert worden waren, um schwächeren Schülern/Schülerinnen eine Chance zu geben, positive Beurteilungen zu erzielen. Dies führte zu einer genaueren und getrennten Betrachtung der Schularbeiten aus Hauptschulen und Allgemeinbildenden Höheren Schulen. Schularbeiten aus Allgemeinbildenden Höheren Schulen (siehe Tabelle 5) zeigten, dass Items, die mit auswendig gelernten Informationen beantwortet werden mussten, in den früheren Tests auf 0,3% zurückgegangen und dann in den späteren Jahren wieder auf 2,4% angestiegen waren. Das Prüfen von Wortschatz wurde auf 6,5% reduziert, während das Prüfen von Grammatik zunahm (33,1%). Das Überprüfen der Hörkompetenz war auf 14,6% angestiegen und jenes in der Lesekompetenz war leicht zurückgegangen (10,5%). Das Schreiben von Texten stieg auf beinahe 33% an, verglichen mit dem kleinen Durchschnittswert aus allen Schulen von 1992. Schularbeiten der Hauptschulen und Neuen Mittelschulen unterschieden sich nicht allzu viel von jenen an Allgemeinbildenden Höheren Schulen. Allerdings machten auswendig gelernte Inhalte beinahe 10% aus und Hörkompetenz war mit nur 10,3%, also um rund 4% weniger, vorhanden. Grammatik und Wortschatz waren anteilmäßig beinahe gleichwertig vorhanden, doch die Textproduktion lag nur bei 27,6% der gesamten Punktezahl der Schularbeiten (siehe Tabelle 6).

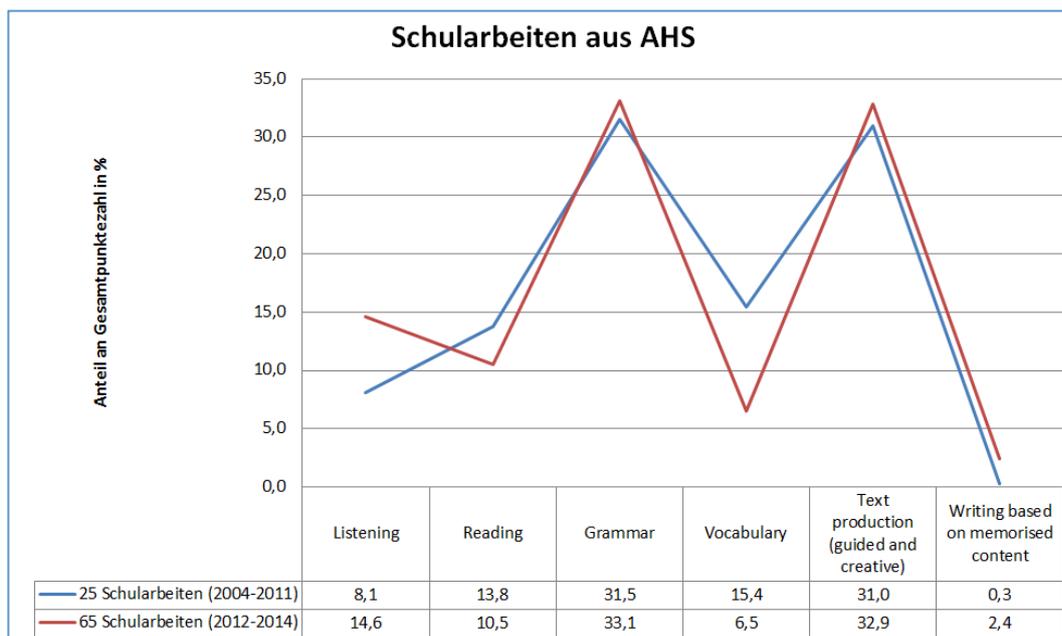


Tabelle 5: Schularbeiten in Allgemeinbildenden Höheren Schulen (AHS).

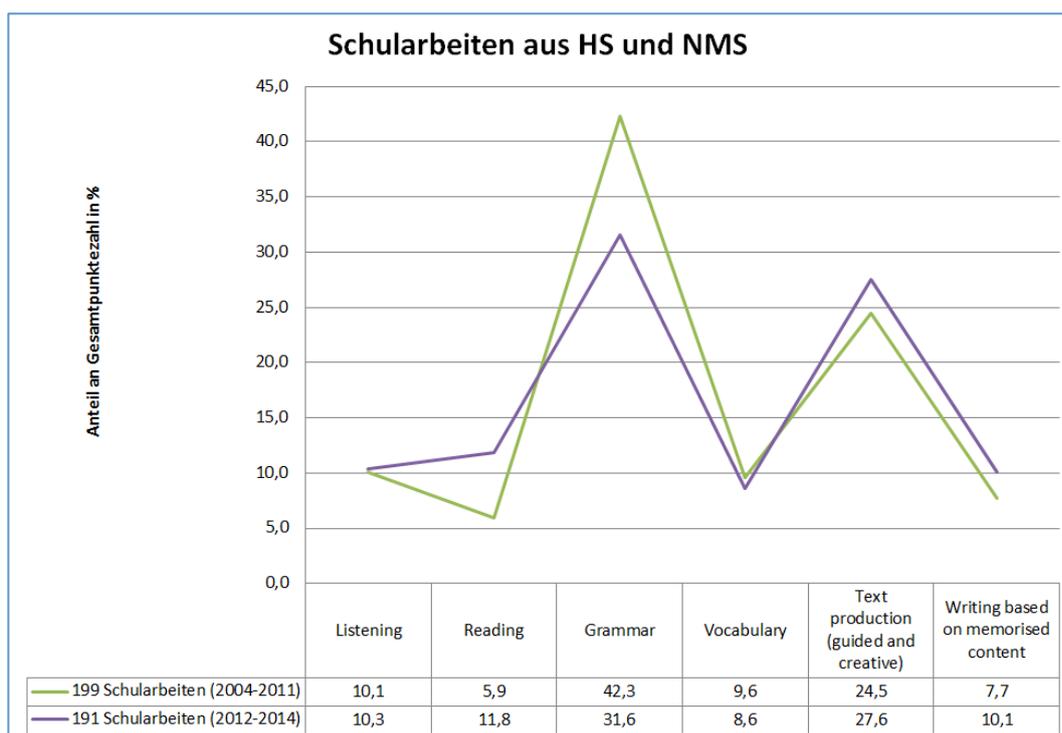


Tabelle 6: Schularbeiten in Hauptschulen (HS) und Neuen Mittelschulen (NMS)

Wenn man berücksichtigt, dass die Unterrichtszeit für Schreibkompetenz auch zurückgegangen war, ist es interessant, die Ergebnisse im nationalen Bericht zur E8-Überprüfung zu betrachten.

Tabelle 7 zeigt die Ergebnisse in einem Maßstab, der bereits in der Erstuntersuchung von 2009 verwendet worden war. Während das Durchschnittsergebnis 2009 bei 500 Punkten lag, waren es im Jahr 2013/14 519 Punkte.

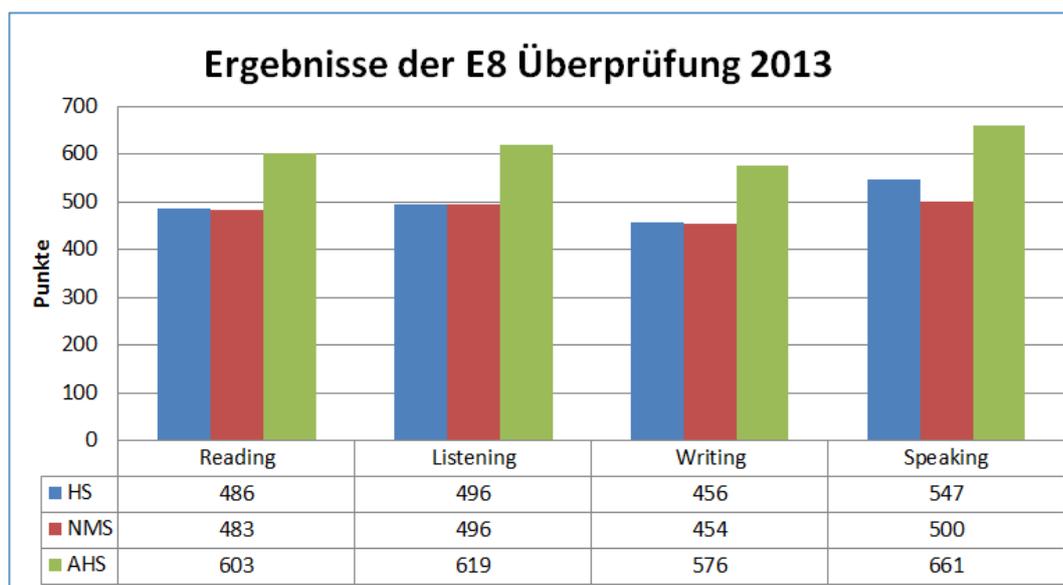


Tabelle 7: Ergebnisse der E8-Überprüfung von 2013 (adaptiert nach BIFIE, 2014, S. 34).

Die Testteilnehmer/innen aus Hauptschulen und Neuen Mittelschulen erzielten in allen Fertigkeiten niedrigere Ergebnisse als jene aus Allgemeinbildenden Höheren Schulen. Der Unterschied in der Schreibkompetenz war markant: In GERS-Niveaus umgewandelt erreichten mehr als die Hälfte der Schüler/innen in Hauptschulen oder Neuen Mittelschulen nur das Niveau A1. Das Ziel, B1, wurde von 61% der Testteilnehmer/innen in Allgemeinbildenden Höheren Schulen erreicht.

Die besten Ergebnisse wurden in der Redekompetenz erzielt, wo der Durchschnitt für die Schüler/innen Allgemeinbildender Höherer Schulen 661 Punkte war und wo die Schüler/innen Neuer Mittelschulen mit 519 Punkten nur leicht unter dem Gesamtdurchschnitt lagen. Das zweitbeste Ergebnis wurde in der Hörkompetenz erreicht, wo die Ergebnisse 496, 496 und 519 Punkte betragen. Der Prozentsatz der Testteilnehmer/innen, die in dieser Fertigkeit nicht das A2-Niveau erreichten, war mit 4,5% der niedrigste in den vier Fertigkeiten. Keine davon stammten aus Allgemeinbildenden Höheren Schulen. Im Gegenteil, 86% der Schüler/innen in Allgemeinbildenden Höheren Schulen erreichte das B1-Niveau in der Überprüfung der Hörkompetenz (BIFIE, 2014).

4.3 Lehren, Lernen und Beurteilen

Bezüglich des Lehrens und Lernens deuten Daten aus den Fragebögen darauf hin, dass Lehrer/innen ihre Lehrmethoden nicht wirklich verändert hatten. In der späteren Umfrage und den Interviews nach den Beobachtungen gaben nur 25% aller Lehrer/innen an, dass sie verändert hatten, was sie unterrichteten, und der Prozentsatz der Lehrer/innen, die angaben, dass sie ihre Methoden und Strategien nicht verändert hatten, war mit 20% relativ hoch. Auffällig war, dass 70% der Lehrer/innen angaben, dass sie ihre Methoden zur Überprüfung von Lernleistungen wegen der standardisierten Testverfahren nicht verändert hatten.

Die letzten beiden Kategorien in Tabelle 8 stammen aus Fragebögen, die von Schülern/Schülerinnen ausgefüllt wurden. Nur 16,7% gaben an, dass sie ihr Lernverhalten wegen der E8-Überprüfung verändert hatten und niemand dachte, dass die Überprüfung zu mehr Motivation beim Lernen führte. In Interviews bestätigten Lehrer/innen und Schulleiter/innen dieses Ergebnis.

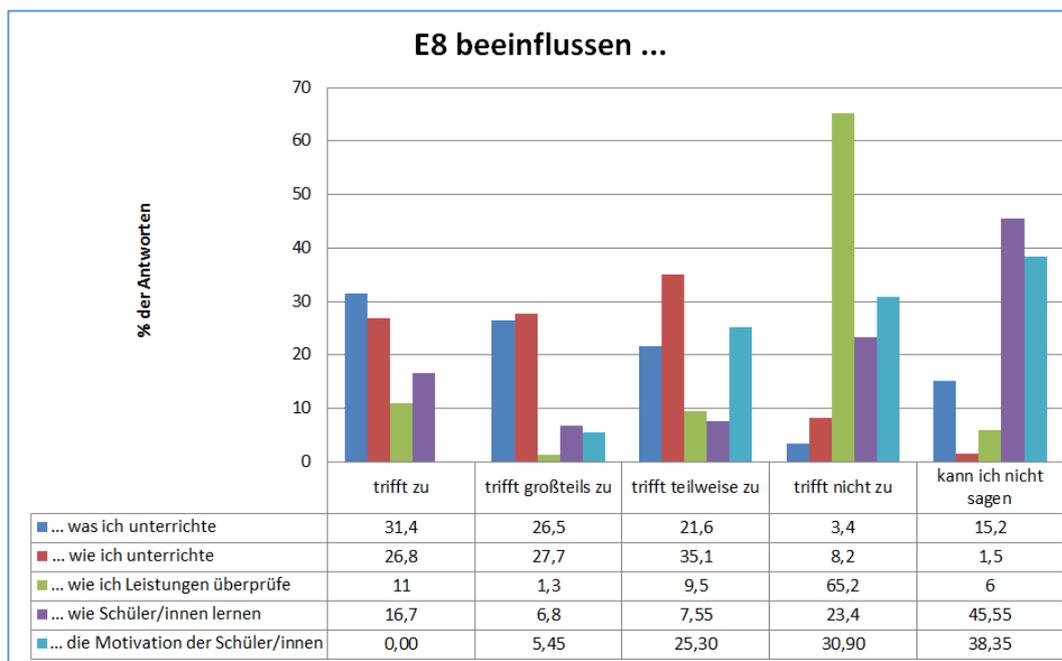


Tabelle 8: Fragebögen von Lehrern/Lehrerinnen und Schülern/Schülerinnen: E8 beeinflussen ...

Daten aus Interviews nach den Unterrichtsbeobachtungen, welche transkribiert und mit Hilfe der MAXQDA-Software analysiert wurden, zeigten verschiedene Kategorien und Trends wie eine „positive Einstellung“ zu den Bildungsstandards und deren Überprüfung. Diese positive Haltung wurde vor allem von durch das BIFIE ausgebildeten Lehrpersonen (Writing Raters, Speaking Assessors & Interlocutors) ausgedrückt. Ausgebildete Beurteiler/innen (Raters, Assessors) benutzen die Skalen für die Bewertung schriftlicher Leistungen regelmäßig und unterrichten und prüfen gesprochene Leistung mit Hilfe des standardisierten Testformats. Allerdings gaben sie auch an, dass sie dafür ihre eigenen Unterrichtsmaterialien erstellen müssten, da Lehrbücher keine geeigneten Materialien enthielten, die mit ihren persönlichen Vorstellungen zum Lehren und Beurteilen kompatibel wären. Lehrer/innen, die kein E8-Training besucht und nicht an den Pilotierungen teilgenommen hatten, verwendeten hauptsächlich die veröffentlichten Testmappen der Schulbücher und berichteten, dass sie selten eigene Ideen in die Designs von Schularbeiten einbringen würden. Sowohl Lehrer/innen als auch Schulleiter/innen gaben an, dass Schüler/innen kaum über die E8-Standards Bescheid wüssten. Dies wird als der Grund dafür vermutet, dass Lehrer/innen wie auch Schulleiter/innen auch nicht an eine bewusste Veränderung des Lernens ihrer Schüler/innen glaubten. Dies wurde in den Fragebögen der Schüler/innen bestätigt.

5 Schlussfolgerungen

Zusammenfassend kann festgestellt werden, dass standardisierte Testverfahren und ein neu angepasster Lehrplan mit Sicherheit Washback auf das Lehren und Lernen von Englisch als Fremdsprache verursacht haben. Allerdings sind die Auswirkungen nicht gleichmäßig auf die vier Fertigkeiten sowie auf Lehren, Lernen, Beurteilen und Motivation aufgeteilt. Positive Auswirkungen wurden vor allem in der Redefähigkeit am deutlichsten sichtbar. Im Fertigkeitsbereich des Sprechens ist der Washback besonders dort spezifisch, stark und positiv, wo ausgebildete Gesprächspartner/innen (Interlocutors) und mündliche Prüfer/innen (Assessoren/Assessorinnen) beobachtet und befragt wurden. Außerdem zeigte sich positiver Washback ganz besonders in Schulen, in denen die mündliche Überprüfung durchgeführt wurde. Die starke Ausprägung des Washbacks an diesen Schulen mag einerseits an der gezielten Vorbereitung der Schüler/innen liegen (Schulen wurden über die Durchführung der E8 im Bereich des Sprechens informiert, um administrative Vorkehrungen zu treffen), andererseits könnte auch die persönlichere Testadministration durch Assessors & Interlocutors für die positive Einstellung verantwortlich sein. Was die anderen Fertigkeiten betrifft, so ist der Washback zwar spezifisch (Hörverständnis wird in Schularbeiten überprüft, Leseverständnis wird verstärkt überprüft), aber

schwach (es gibt noch keine gleichmäßige Verteilung der Fertigkeiten in Schularbeiten, und Lehrkräfte unterrichten Lese- und Hörstrategien nicht explizit). Bezüglich der schriftlichen Kompetenzen haben Lehrbücher bereits das standardisierte Format der Schreibenanlässe (Writing Prompts) übernommen. Allerdings enthalten sie nicht genügend Lehr- oder Lernstrategien, die bewährte Schreibstrategien entwickeln und Lückentexte zugunsten von Schreiben auf Textebene reduzieren würden. Des Weiteren vermitteln nur Lehrer/innen, die Klassen in der Sekundarstufe II betreuen, explizit, wie Lese- oder Hörstrategien effektiv den Kompetenzerwerb fördern. In manchen Klassen der Sekundarstufe I ist vielmehr das laute Vorlesen immer noch eine übliche Vorgehensweise und Lehrer/innen lehren nicht wie man effektiv liest oder zuhört. Sie beschränken sich vielmehr auf Aktivitäten, die Lesen und Hören inkludieren und erachten das bloße „Tun“ als ausreichend für die effektive Förderung von Lese- und Hörkompetenzen (Graham und Santos, 2015).

Der beabsichtigte Washback auf Schularbeiten war bei ausgebildeten Beurteilern/Beurteilerinnen (Raters, Assessors) ebenso wie bei ausgebildeten Lerndesignern und Lerndesignerinnen stark. Dennoch stellten die Lehrpersonen nicht die wesentliche Verbindung zwischen dem Testkonstrukt und den Konsequenzen der Schularbeiten auf Klassenebene her. Daraus ist zu schließen, dass Lehrer/innen immer noch viel Unterstützung bei der professionellen Leistungsbeurteilung (Green, 2014) brauchen. Es scheint, dass die bloße Implementierung von Standards und standardisierten Testverfahren nicht die erwünschte Wirkung hat. Es erfordert ausführliche und zielgerichtete Fortbildung, sollen die Bildungsstandards sowie die Testungen konkrete und dauerhafte Veränderungen hervorrufen.

Die Motivation der Schüler/innen durch standardisierte Testverfahren war positiv, allerdings nur bei jenen Schülern/Schülerinnen, die im Sprechen überprüft wurden und die währenddessen positives Feedback durch erfolgreiche Kommunikation oder Interlocutor-Feedback erhielten. Beim Zusammenfassen des Feedbacks aus den Fragebögen der Schüler/innen zeigte sich außerdem, dass die Wirkung der E8-Überprüfung auf die Motivation zum damaligen Zeitpunkt als die schwächste der vier Dimensionen gesehen werden musste.

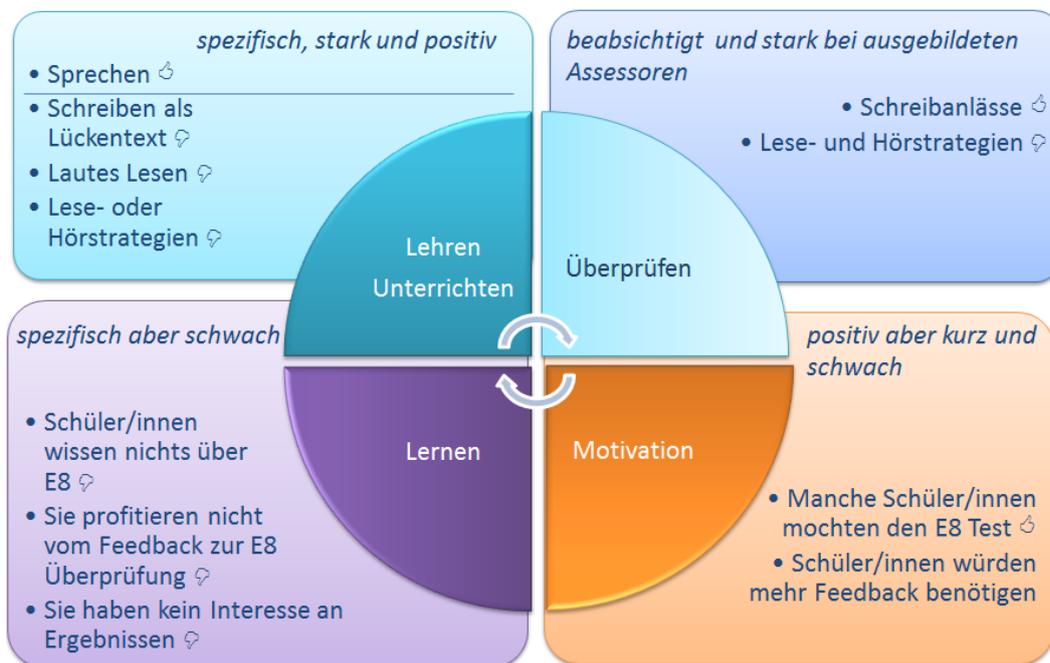


Abbildung 3: E8-Washback in vier Dimensionen: Lehren/Unterrichten, Überprüfen, Lernen und Motivation.

Eine kurze Zusammenfassung der Ergebnisse deutet daher darauf hin, dass

- Standards und standardisierte Testverfahren das Ausmaß und die Qualität des Unterrichts von Sprechfertigkeiten gesteigert haben, doch die Sprechfertigkeit dennoch nicht wirklich überprüft wird. Sprechfertigkeiten tragen kaum oder gar nicht zur Gesamtnote bei.
- Lehrer/innen, die keine Klassen der Sekundarstufe II unterrichten oder nicht über Lese- oder Hörstrategien Bescheid wissen, diese nicht unterrichten; Schüler/innen haben in der Folge Schwierigkeiten mit Aufgaben, welche die Verwendung von Strategien erfordern, besonders mit dem Inferenzieren.

- die meisten Lehrer/innen nicht über Textsorten Bescheid wissen; sie wissen ebenfalls nicht, wie man die erforderlichen Strategien unterrichtet, um verschiedene Textsorten erfolgreich zu produzieren.
- Lehrbücher jene Strategien oder sprachliche Mittel, mit denen Aufgaben strategisch bearbeitet werden können, kaum vermitteln.
- Schüler/innen und Eltern über das späte Feedback zu den standardisierten Überprüfungen frustriert sind; es sollte überlegt werden, rascheres Feedback zu liefern. Anderenfalls werden Schüler/innen und Eltern in der Zukunft keinen Grund sehen, wegen der Standards oder standardisierten Testverfahren ihr Verhalten zu verändern.
- sich das Prüfen in Schulen nicht verändert hat. Es wird immer noch ein besonderer Schwerpunkt auf schriftliche Leistungen gelegt, besonders in Form von Lückentexten; Textproduktion wird oft durch Schreibenlässe initiiert, die zwar mit dem Testformat übereinstimmen, aber kaum komplexes Denken erfordern.
- Veränderungen in der Einstellung zum Lehren und Lernen von Englisch als Fremdsprache in den beobachteten Unterrichtseinheiten kaum erkennbar sind.

Die potentielle Auswirkung auf Kompetenzorientierung, die durch Standards und standardisierte Testverfahren erzielt werden soll, könnte verloren gehen, wenn theoretische Testkonstrukte in der Zukunft nicht bekannter und bewusster gemacht werden. Dies kann durch expliziteren und strategischeren Input für Lehrer/innen sowie durch rascheres Feedback für Schüler/innen, bevor sie die Schule verlassen oder die Sommerferien beginnen, erreicht werden. Die informelle Kompetenzmessung der Lernenden auf standardisierte Art und Weise, wie sie vom BIFIE umgesetzt wurde (<https://www.bifie.at/node/3099>), ist ein guter Schritt, um stärkeren und spezifischeren Washback auf Klassenebene zu erreichen. Allerdings würden mehr Informationen für Lehrer/innen über die zugrundeliegenden Konstrukte die Investition effektiver machen. Zurzeit sind die Materialien ein Service, das mit wenig Auswirkung auf das Lehren und Lernen verwendet wird.

Insgesamt weisen die Ergebnisse dieser Untersuchung zum Washback der E8-Überprüfung auf das Lehren, Lernen und Beurteilen darauf hin, dass sich das Unterrichten und Prüfen im Englischunterricht nicht maßgeblich verändert haben. Das Prüfen konzentriert sich immer noch zu stark auf schriftliche Leistungen, besonders in Form von Lückentexten, obwohl das Überprüfen von Lese- und Hörkompetenzen zunimmt. Obwohl Textproduktion oft auf Schreibenlässen basiert, welche jenen der E8 Überprüfung entsprechen, erfordern schriftliche Aufgaben dieser Art oft nicht jenes komplexere Denken, das in der Neuen Mittelschule erwartet wird. Dasselbe gilt für Lese- und Höraufgaben in veröffentlichten Testmaterialien. Zudem wurden von den meisten befragten Lehrkräften keine Veränderungen in der Einstellung zum Prüfen und Beurteilen bekundet.

Daraus lässt sich schlussfolgern, dass die Aus- und Fortbildung von Lehrkräften aktiver und expliziter zur Kompetenzentwicklung der Lehrer/innen in Hinblick auf die strategische Vermittlung von Strategien sowie auf die formative wie die summative Beurteilung von Lernleistungen beitragen sollte. Wenn die E8-Überprüfung das beabsichtigte Ziel pädagogischer Veränderung und Innovation erreichen soll, müssen ihre Grundlagen besser bekannt und konkreter vermittelt werden. Dies kann durch die Veröffentlichung der Testkonstrukte und die Stärkung des Bewusstseins von Lehrern/Lehrerinnen für die Grundlagen des E8-Testverfahrens erreicht werden.

Literatur

- Alderson, C. J., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14, pp. 115-129.
- Bachman, L., & Palmer, A. (2010). *Language Assessment in Practice*. Oxford: OUP.
- Beck, B., & Klieme, E. (Eds.). (2007). *Sprachliche Kompetenzen. Konzepte und Messung. DESI Studie*. Weinheim & Basel: Beltz.
- BIFIE. (2011). *Bildungsstandards in Österreich. Überprüfung und Rückmeldung. Informationen für Lehrer/innen und Schulleiter/innen*. Salzburg: BIFIE.
- BIFIE (Ed.). (2012). *Handbuch zur Standardüberprüfung in Englisch, 8. Schulstufe*. Retrieved August 20, 2014, from BIFIE: https://www.bifie.at/system/files/dl/bist-ue_Handbuch-Standardueberpruefung-E8_20130204.pdf.

- BMBWF. (2000). *Lehrplan: LEBENDE FREMDSPRACHE (Erste, Zweite)*. Retrieved from Bundesministerium für Bildung, Wissenschaft und Forschung:
https://bildung.bmbwf.gv.at/schulen/unterricht/lp/ahs8_782.pdf?61ebzr.
- Borg, S. (2006). *Teacher Cognition and Language Education*. London: Continuum.
- Brown, D. H., & Abeywickrama, P. (2010). *Language Assessment. Principles and Classroom Practices*. New York: Pearson.
- Brown, G. (1996). Introduction. In G. Brown, K. Malmkjaer, & J. Williams (Eds.), *Performance and Competence in Second Language Acquisition* (pp. 1-8). Cambridge: Cambridge University Press.
- Buck, G. (2001). *Assessing Listening*. Cambridge: Cambridge University Press.
- Canale, M. (1983). From communicative competence to communicative language pedagogy. In C. Richards, & R. Schmidt (Eds.), *Language and communication* (pp. 1-27). New York: Longman.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1, pp. 1-47.
- Celce-Murcia, M., & Dörnyei, Z. (1995). Communicative Competence: A Pedagogically Motivated Model with Content Specifications. *Issues in Applied Linguistics*, 6(2), 5-35.
- Cheng, L. (2005). *Changing Language Teaching through Language Testing*. Cambridge: CUP.
- Cheng, L., Watanabe, Y., & Curtis, A. (Eds.). (2004). *Washback in Language Testing: Research Contexts and Methods*. Mahwah: Laurence Erlbaum & Associates.
- Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T., & McNamara, T. (1999). *Dictionary of language testing* (Studies in Language Testing 7 ed.). Cambridge: Cambridge University Press.
- Europarat. (2001). *Gemeinsamer Europäischer Referenzrahmen für Sprachen: lernen, lehren, beurteilen*. (Goethe-Institut, Hrsg., & J. Quetz, Übers.) Berlin: Langenscheidt.
- Gassner, O., Mewald, C., & Sigott, G. (2007). *Technical Report Testing Reading*. Universität Klagenfurt: Language Testing Centre.
- Gassner, O., Mewald, C., Brock, R., Lackenbauer, F., & Siller, K. (2011). *Testing Writing for the E8 Standards. Technical Report*. (BIFIE, Ed.) Salzburg.
- Graham, S., & Santos, D. (2015). *Strategies for Second Language Listening. Current Scenarios and Improved Pedagogy*. Basingstoke: Palgrave Macmillan.
- Green, A. (2007). *IELTS Washback in Context*. Cambridge: Cambridge University Press.
- Habermas, J. (1979). *Communication and the Evolution of Society*. Toronto: Beacon Press.
- Habermas, J. (1987). *The Theory of Communicative Action, Volume 2: Lifeworld and System: A Critique of Functionalist Reason*. Boston: Beacon Press.
- Hymes, D. H. (1972). On communicative competence. In J. Pride, & J. Holmes (Eds.), *Sociolinguistics: selected readings*. (pp. 269–293). Harmondsworth: Penguin.
- Klafki, W. (1963). *Studien zur Bildungstheorie und Didaktik*. Weinheim: Beltz Verlag.
- Klafki, W. (1993). *Neue Studien zur Bildungstheorie und Didaktik: Zeitgemäße Allgemeinbildung und kritisch-konstruktiver Didaktik*. Weinheim: Beltz Verlag.
- Lantolf, J. (2001). *Sociocultural theory and second language acquisition*. Oxford: Oxford University Press.
- Larsen-Freeman, D. (1997). Chaos/Complexity Science on Second Language Acquisition. *Applied Linguistics*, 18(2), pp. 141-165.
- Lyons, J. (1996). On competence and performance and related notions. In G. Brown, K. Malmkjaer, & J. Williams (Eds.), *Performance and Competence in Second Language Acquisition* (pp. 11-32). Cambridge: Cambridge University Press.
- Mewald, C., Gassner, O., & Sigott, G. (2007). *Technical Report: Testing Listening*. Universität Klagenfurt: Language Testing Centre.
- Mewald, C., Gassner, O., Lackenbauer, F., Brock, R., & Siller, K. (2013). *Testing Speaking for the E8 Standards*. Salzburg: BIFIE Salzburg. Retrieved 03 10, 2014, from BIFIE:
https://www.bifie.at/system/files/dl/TR_Speaking_130805.pdf.
- Organization for Economic Co-operation and Development. (1994). *School. A matter of choice*. Paris: Centre for Educational Research and Innovation, OECD.

- Richards, J. C., & Rodgers, T. S. (2014). *Approaches and Methods in Language Teaching*. Cambridge: Cambridge University Press.
- Rickheit, G., & Strohner, H. (Eds.). (2010). *Handbook of Communication Competence* (Vol. 1). Berlin: De Gruyter.
- Roth, H. (1971). *Pädagogische Anthropologie. Band 2. Entwicklung und Erziehung. Grundlagen einer Entwicklungspädagogik*. Hannover: Schroedel .
- Schlichterle, B., Weißkopf-Prantner, V., & Westfahl-Greitner, T. (2013). Kriterienorientierte Leistungsfeststellung mit der 4.0-Skala. Zentrum für lernende Schulen. NMS Entwicklungsbegleitung.
- Shohamy, E. (1992). Beyond proficiency testing: A diagnostic feedback testing model for assessing foreign language learning. *Modern Language Journal*, 76(4), 513-521.
- Shohamy, E. (1993). *The power of tests: The impact of language tests on teaching and learning. NFLC Occasional Paper*. Washington: National Foreign Language Center.
- Shohamy, E. (1996). Competence and performance in language testing. In G. Brown, K. Malmkjaer, & J. Williams (Eds.), *Perforamce & Competence in Second Language Acquisition* (pp. 138-151). Cambridge: Cambridge University Press.
- Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996). Test impact revisited: Washback effect over time. *Language Testing*, 13(3), 298-317.
- Stigler, J. W., & Hiebert, J. (1999). *The Teaching Gap. Best Ideas from the World's Teachers for Improving Education in the Classroom*. New York: The Free Press.
- Strauss, A., & Corbin, J. (1990). *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. Newbury Park: Sage Publications.
- Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics*, 371-391.
- Wall, D. (2005). *The impact of high-stake testing on teaching and learning. A case study using insights from testing and innovation theory*. Cambridge: Cambridge University Press.
- Watanabe, Y. (1997). *Washback effects of the Japanese entrance examination: Classroom-based research*. Lancaster University: Unpublished doctoral dissertation.
- Watanabe, Y. (2004). Methodology in Washback Studies. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in Language Testing* (pp. 19-36). Mahwah, NJ: Lawrence Erlbaum Associates.
- Weinert, F. E. (2001). *Leistungsmessungen in Schulen*. Weinheim: Beltz Verlag.
- Ziegler, N., & Kang, L. (2016). Mixed methods designs. In A. J. Moeller, J. W. Creswell, & N. Saville (Eds.), *Second Language Assessment and Mixed Methods Research* (pp. 51-83). Cambridge: Cambridge University Press.