



universität  
wien

# Diplomarbeit

Titel der Diplomarbeit

## Der Backwascheffekt der GCSE-Prüfung

Eine empirische Untersuchung zur Wahrnehmung des Backwascheffekts durch  
Deutschlehrkräfte in England vor und nach der GCSE-Reform 2009

Verfasserin

Elisabeth Denscher

angestrebter akademischer Grad

Magistra der Philosophie (Mag.phil.)

Wien, im Mai 2011

Studienkennzahl lt. Studienblatt:

A190 344 333

Studienrichtung lt. Studienblatt:

UF Deutsch

Betreuer:

em. Univ.-Prof. Dr. Hans-Jürgen Krumm



## **Ich danke...**

meinen ehemaligen Kolleginnen und Kollegen an der John Warner School, insbesondere Vanessa Lamb, die für Fragen jederzeit bereit stand,

meinem Diplomarbeitsbetreuer Univ.-Prof. Dr. Hans-Jürgen Krumm, der mir kurz vor seiner Emeritierung noch die Betreuung dieser Arbeit zugesagt hat,

Univ.-Prof. Dr. Camilla Badstübner-Kizik für die wissenschaftliche Begleitung,

Mag.<sup>a</sup> Silvia Flotzinger-Aigner, M.A., von KulturKontakt Austria, die den Kontakt zu den derzeit tätigen SprachassistentInnen hergestellt hat,

allen, die an der Verbreitung meiner Fragebögen mitgewirkt haben, sowie

allen Lehrerinnen und Lehrern, die an meiner Umfrage teilgenommen haben und nicht zuletzt

meiner Familie, die mich während meines Studiums immer unterstützt hat.



# Inhalt

<b>1</b>	<b>Einleitung .....</b>	<b>1</b>
<b>2</b>	<b>Theoretischer Hintergrund .....</b>	<b>5</b>
2.1	TESTGÜTEKRITERIEN .....	5
2.1.1	Hauptgütekriterien .....	5
2.1.1.1	Objektivität.....	5
2.1.1.2	Reliabilität.....	6
2.1.1.3	Validität.....	6
2.1.2	Nebengütekriterien.....	8
2.1.3	Gütekriterien für kommunikatives Testen und Prüfen .....	8
2.1.3.1	Authentizität.....	8
2.1.3.2	Interaktivität.....	10
2.1.3.3	Integrativität.....	10
2.1.3.4	Unvorhersehbarkeit .....	11
2.1.4	Zusammenfassung: Balance der Merkmale .....	12
2.2	DER KOMMUNIKATIVE ANSATZ.....	14
2.2.1	Geschichte und Modelle .....	14
2.2.2	Grammatik im kommunikativen Unterricht .....	15
2.2.3	Kritik am kommunikativen Ansatz für den Fremdsprachenunterricht in England .....	16
2.3	DER ‚BACKWASHEFFEKT‘ .....	17
2.3.1	Entwicklung und Strömungen der Backwashforschung.....	17
2.3.2	Watanabes Backwashdimensionen .....	18
2.3.3	Schifkos Aspekte des Backwasheffekts.....	19
2.3.4	Bedingungen für positiven Backwash.....	20
2.3.5	Negativer Backwash.....	21
2.3.6	Zusammenfassung.....	24
2.4	FREMSPRACHENUNTERRICHT UND PRÜFUNGSSYSTEME IN ENGLAND.....	24
2.4.1	Die Situation von Fremdsprachen im Vereinigten Königreich .....	24
2.4.2	Organisation und Umfang des Fremdsprachenunterrichts in Schulen: ein Fallbeispiel.....	25
2.4.3	Prüfungssystem und Qualifikationen .....	26
2.4.3.1	Das GCSE.....	27
2.4.3.2	Die A-Levels .....	28
2.4.3.3	Exam Boards.....	29
2.4.4	Zusammenfassung.....	30

<b>2.5</b>	<b>DAS DEUTSCH-GCSE BIS 2008/2010 .....</b>	<b>31</b>
2.5.1	Kurzdarstellung .....	31
2.5.2	Das GCSE im Kontext der Hauptgütekriterien .....	32
2.5.2.1	<i>Spezialfall: Coursework</i> .....	32
2.5.3	Das GCSE als kommunikative Prüfung .....	33
2.5.4	Das GCSE im Kontext des kommunikativen Prüfens: Authentizität und Interaktivität.....	35
2.5.4.1	<i>Rollenspiele</i> .....	35
2.5.4.2	<i>Konversationen</i> .....	37
2.5.4.3	<i>Fazit</i> .....	40
2.5.5	Der ‚Backwash‘-Effekt des GCSE bis 2010.....	40
2.5.5.1	<i>Der Backwasheffekt nach Schifko (2001)</i> .....	40
2.5.5.2	<i>Der Backwasheffekt nach Watanabe (2004)</i> .....	43
2.5.5.3	<i>Das GCSE im Kontext der Bedingungen für positiven Backwash</i> .....	44
2.5.6	Zusammenfassung.....	46
<b>2.6</b>	<b>DAS GCSE SEIT 2010.....</b>	<b>48</b>
2.6.1	Controlled Assessment .....	48
2.6.2	Weitere Veränderungen .....	50
2.6.3	Einschätzung und Bewertung der Veränderungen.....	51
<b>3</b>	<b>Empirische Untersuchung.....</b>	<b>53</b>
3.1	METHODISCHE ÜBERLEGUNGEN .....	53
3.1.1	Triangulation.....	55
3.1.2	Durchführung der Untersuchung.....	55
3.1.3	Der Fragebogen.....	56
3.1.4	Datenauswertung .....	57
3.1.5	Kategorienbildung.....	58
3.1.6	Hypothesen.....	59
3.2	ERGEBNISSE.....	60
3.2.1	Der allgemeine Backwasheffekt.....	60
3.2.1.1	<i>Quantitativ ausgerichtete Fragen</i> .....	60
3.2.1.2	<i>Verbale Daten und Interpretation</i> .....	62
3.2.1.3	<i>Fazit</i> .....	64
3.2.2	Intensität und Ausrichtung.....	65
3.2.2.1	<i>Quantitativ ausgerichtete Fragen</i> .....	65
3.2.2.2	<i>Verbale Daten und Interpretation</i> .....	66
3.2.2.3	<i>Fazit</i> .....	68
3.2.3	Lernstrategie ‚Auswendiglernen‘ .....	68

3.2.3.1	<i>Quantitative Daten</i> .....	68
3.2.3.2	<i>Verbale Daten und Interpretation</i> .....	70
3.2.3.3	<i>Fazit</i> .....	74
<b>3.2.4</b>	<b>Validität</b> .....	<b>74</b>
3.2.4.1	<i>Quantitative Daten</i> .....	74
3.2.4.2	<i>Verbale Daten und Interpretation</i> .....	76
3.2.4.3	<i>Fazit</i> .....	78
<b>3.2.5</b>	<b>Freies und kreatives Unterrichten</b> .....	<b>78</b>
3.2.5.1	<i>Quantitative Daten</i> .....	79
3.2.5.2	<i>Verbale Daten und Interpretation</i> .....	80
3.2.5.3	<i>Fazit</i> .....	83
<b>3.3</b>	<b>ABSCHLIEßENDE DISKUSSION</b> .....	<b>85</b>
<b>4</b>	<b>Zusammenfassung</b> .....	<b>89</b>
<b>5</b>	<b>Glossar</b> .....	<b>90</b>
<b>6</b>	<b>Bibliographie</b> .....	<b>92</b>
6.1	<b>SEKUNDÄRLITERATUR</b> .....	<b>92</b>
6.2	<b>INTERNETQUELLEN</b> .....	<b>96</b>
6.3	<b>ONLINEZEITUNGEN</b> .....	<b>98</b>
<b>7</b>	<b>Anhang</b> .....	<b>99</b>
7.1	<b>BILDUNG DES KATEGORIENSYSTEMS</b> .....	<b>99</b>
7.1.1	<b>Einschätzung und Bewertung der Veränderungen</b> .....	<b>99</b>
7.1.2	<b>Hypothesen</b> .....	<b>99</b>
7.1.3	<b>Deduktive Kategorien</b> .....	<b>99</b>
7.1.4	<b>Induktive Kategorien</b> .....	<b>100</b>
7.1.5	<b>Das Kategoriensystem</b> .....	<b>101</b>
7.2	<b>FRAGEBOGEN</b> .....	<b>102</b>
7.3	<b>ERGEBNISSE – QUANTITATIVE AUSWERTUNG</b> .....	<b>113</b>
7.3.1	<b>In der Arbeit verwendete Tabellen</b> .....	<b>113</b>
7.3.2	<b>Weitere Ergebnisse</b> .....	<b>117</b>
7.4	<b>ERGEBNISSE – VERBALE DATEN</b> .....	<b>120</b>
7.5	<b>E-MAIL-KONVERSATION MIT FRAU LAMB, HEAD OF MFL AN DER JOHN WARNER SCHOOL</b> 135	
7.6	<b>ABSTRACTS</b> .....	<b>137</b>

## **Tabellen- und Abbildungsverzeichnis**

### ***A. Abbildungen***

Abbildung 1: Schematische Darstellung der Beziehungen zwischen Testgütekriterien, kommunikativen Prüfungs- und Kompetenzmodellen und Backwashkriterien .....	2
Abbildung 2: Bachmans Modell der 'communicative language ability'.....	14
Abbildung 3: Rollenspielkarte von Edexcel .....	36
Abbildung 4: Prüfungsangaben für Controlled Assessment (Speaking).....	49
Abbildung 5: Prüfungsangaben für Controlled Assessment (Writing).....	49



## **B. Tabellen**

Tabelle 1: Prüfungsmodi des alten GCSE.....	31
Tabelle 2: Beurteilungsraster von Edexcel.....	34
Tabelle 3: Beurteilungsraster für den Konversationsteil der mündlichen Prüfung von AQA .....	38
Tabelle 4: Prüfungsmodi des neuen GCSE .....	48
Tabelle 5: Antworten auf die Frage: "Would teaching be different if there were no nation-wide exams at all?" .....	61
Tabelle 6: Antworten auf die Frage: "Would teaching be different if there were no league tables?" .....	62
Tabelle 7: Antworten auf die Frage: "Would this difference be rather positive or negative?" .....	62
Tabelle 8: Antworten auf die Frage: "How strongly did the old GCSE affect your teaching?" .....	65
Tabelle 9: Antworten auf die Frage: "How strongly does the new GCSE affect your teaching?" .....	65
Tabelle 10: Antworten auf die Frage: "Was the impact of the old GCSE on your teaching rather positive or negative?" .....	66
Tabelle 11: Antworten auf die Frage: "Is the impact of the new GCSE on your teaching rather positive or negative?" .....	66
Tabelle 12: Antworten auf die Frage: "Which strategies are your students using / did the students use to prepare for the assessment?" .....	68
Tabelle 13: Antworten auf die Frage: "Do you think that Controlled Assessment says more or less about students' actual abilities than the previous test types?" .....	74
Tabelle 14: Antworten auf die Frage: "Do / will the results reflect your students' actual communicative competence?" .....	75
Tabelle 15: Antworten auf die Frage: "How will the introduction of controlled assessment affect the students' overall German proficiency?" .....	75

Tabelle 16: Antworten auf die Frage: "How much room do you now have for personalised and creative teaching (in comparison to the previous GCSE)?" .....	79
Tabelle 17: Antworten auf die Frage: "How do you think controlled assessment will affect the time available for content not treated in the exam?" .....	79
Tabelle 18: Antworten auf die Frage: "How time-consuming was the preparation for the GCSE previously?" .....	80
Tabelle 19: Antworten auf die Frage: „How time-consuming is the preparation for the GCSE now?" .....	80

# 1 Einleitung

Ziel dieser Arbeit ist es, den Backwasheffekt der GCSE-Prüfung vor und nach der Reform von 2009 zu untersuchen. Mit dem Ausdruck ‚Backwasheffekt‘ werden die Rückwirkungen einer Prüfung auf den ihr vorangehenden Unterricht bezeichnet.

Das GCSE ist eine Prüfung des englischen Schulsystems, welche das Ende der Schulpflicht nach der 11. Schulstufe markiert und deren Resultate bedeutende Auswirkungen auf das weitere Leben der KandidatInnen haben. Die Prüfung wird schon seit Jahren immer wieder kritisiert. Zentrale Kritikpunkte sind dabei der Druck zum *Teaching to the Test*<sup>1</sup> und die stetige Vereinfachung der Prüfungsanforderungen, die in den Fremdsprachen ein stures Auswendiglernen vorformulierter Sätze als Lernstrategie zur Folge haben. Am Institut für Germanistik der Universität Wien wurden bereits Diplomarbeiten veröffentlicht, die dieses Thema am Rande berührten<sup>2</sup>. Als Sprachassistentin lernte ich die Auswirkungen dieser Prüfung auf den Deutschunterricht von 2008 bis 2009 acht Monate lang selbst kennen, was letztendlich nicht nur zu Reflexion und Kritik führte, sondern mich auch zu diesem Diplomarbeitsthema inspirierte.

Der Verbesserungsbedarf wurde auch von den verantwortlichen Institutionen erkannt: Das GCSE befand sich schon während meiner Tätigkeit als Sprachassistentin im Reformprozess. Seit Herbst 2009 wird nun mit dem neuen GCSE-System gearbeitet, doch aufgrund der zweijährigen Kurse wurde noch nicht damit geprüft. Das Thema ist demnach höchst aktuell, denn die ersten Abschlüsse wird es erst im Sommer 2011 geben. Auch wissenschaftliche Analysen dieser Umstellung gibt es bisher noch nicht. Da der Unterricht aber trotzdem schon von der Reform betroffen ist, müsste bereits eine Veränderung des Backwasheffekts erkennbar sein. Das motiviert mich zu folgender Herangehensweise: zu einer Analyse der Prüfung in Bezug auf Kriterien für positiven und negativen Backwash einerseits, und zu einer empirischen Untersuchung, die die Einstellungen und Meinungen von Lehrkräften zu Tage bringen soll andererseits. Die Arbeit ist folgendermaßen aufgebaut:

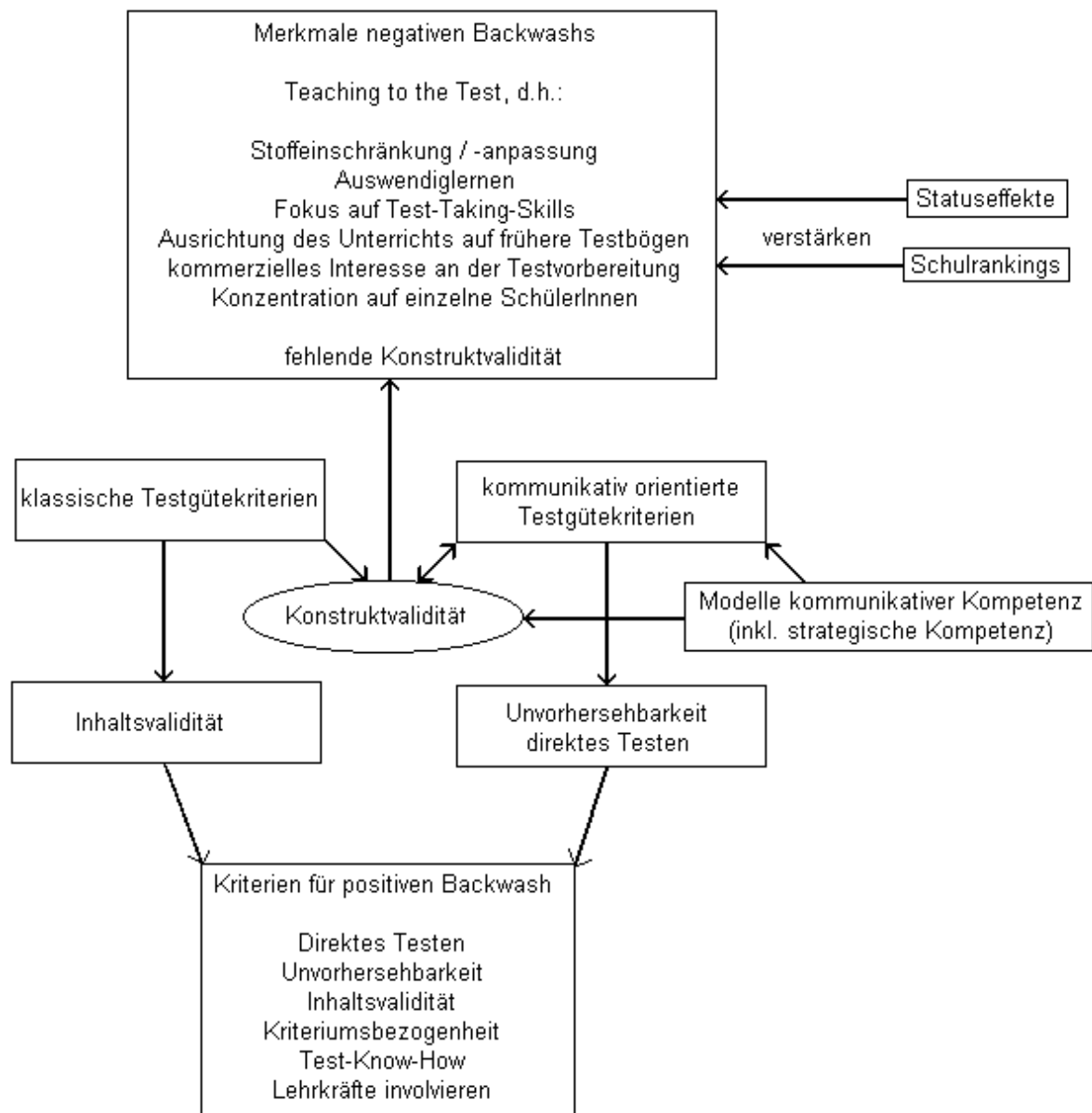
---

<sup>1</sup> Wenn es keine etablierten Übersetzungen gibt, verwende ich in dieser Arbeit die englischsprachigen Bezeichnungen (z.B. *Teaching to the Test*, *Controlled Assessment*), wobei ich Nomen zur besseren Lesbarkeit in Anlehnung an das Deutsche großschreibe.

<sup>2</sup> Vgl. z.B. Obermaier (1999); Koglbauer (2003).

### Theoretischer Hintergrund.

Zunächst werden die Testgütekriterien und das Konstrukt ‚Kommunikative Kompetenz‘ sowie die Praxis kommunikativen Unterrichts kurz erläutert. Das darauf folgende Kapitel erklärt die Dimensionen und Aspekte des Backwasheffekts nach Watanabe (2004) und Schifko (2001) und enthält mit den Bedingungen für positiven und negativen Backwash die Basis für die Analyse der GCSE-Prüfung. Der Zusammenhang von Gütekriterien, kommunikativem Prüfen, Modellen kommunikativer Kompetenz und dem Backwasheffekt ist komplex und linear schwer vermittelbar. Folgende schematische Darstellung soll daher eine Orientierungshilfe bieten:



**Abbildung 1: Schematische Darstellung der Beziehungen zwischen Testgütekriterien, kommunikativen Prüfungs- und Kompetenzmodellen und Backwashkriterien**

Abbildung 1 zeigt, wie manche der Testgütekriterien (Inhaltsvalidität, Unvorhersehbarkeit, direktes Testen, Konstruktvalidität) einen wichtigen Beitrag zu positivem Backwash leisten. Eine besondere Rolle spielt das klassische Testgütekriterium Konstruktvalidität: Wenn es sich vordergründig um einen kommunikativ ausgerichteten Test handelt, zeigt eine Evaluierung der Konstruktvalidität, ob der Test den kommunikativ orientierten Gütekriterien und Modellen kommunikativer Kompetenz entspricht. Wenn dies nicht der Fall ist, ist negativer Backwash die Folge. Prinzipiell kann negativer Backwash durch Druck von außen – sei es durch den hohen Status einer Prüfung oder durch Schulrankings – verstärkt werden. Auf die genauen Zusammenhänge wird in den entsprechenden Kapiteln noch hingewiesen.

### Darstellung des Fremdsprachenunterrichts und Prüfungssystems in England

Beginnend mit einer Darstellung der Situation von Fremdsprachen im Vereinigten Königreich allgemein werden hier der schulische Fremdsprachenunterricht und das Prüfungs- und Qualifikationssystem erklärt. Dieses Wissen ist notwendig, um die Wirkungsweisen und Kritikpunkte des alten GCSE nachvollziehen zu können, und um zu verstehen, wie es zur aktuellen Reform kommen konnte.

### Darstellung und Analyse der GCSE-Prüfungen

Die alte GCSE-Prüfung wird einer Analyse im Hinblick auf die Kriterien für positiven bzw. negativen Backwash unterzogen. Eine wichtige Rolle spielen dabei jene Punkte, die der meisten Kritik ausgesetzt waren und dadurch für die Reform ausschlaggebend waren: die Rollenspiele bei der mündlichen Prüfung und die *Coursework*. Es folgt eine Darstellung der reformierten Prüfung, die sich hauptsächlich durch die Einführung von *Controlled Assessment* auszeichnet. Der Abschnitt ‚Einschätzung und Bewertung der Veränderungen des GCSE‘ ist ein Produkt der oben genannten Analyse und von anschließenden Gesprächen mit DeutschlehrerInnen<sup>3</sup> der Schule, an der ich selbst tätig war. Im Sinne des Prinzips der Triangulation soll der Gegenstand aus unterschiedlichen Perspektiven betrachtet bzw. mit unterschiedlichen Methoden untersucht werden. Meine theoriegeleitete

---

<sup>3</sup> Obwohl ich in dieser Arbeit anstelle des generischen Maskulinums durchgängig Strategien zur Geschlechtsspezifikation wie z.B. die Binnen-I-Schreibung (z.B. LehrerInnen) und zur Geschlechtsneutralisierung (z.B. Lehrkräfte) verwende, muss ich kritisch anmerken, dass ich nicht alle Vorschläge der feministischen Sprachwissenschaft umsetze und auch queere Schreibweisen (z.B. Lehrer\*innen) nicht zur Anwendung kommen.

Analyse soll daher um eine empirische Untersuchung erweitert werden. Diese soll zeigen, ob Lehrkräfte meine Einschätzungen teilen.

### Empirische Untersuchung

Eine qualitativ-quantitative Untersuchung soll einen tieferen Einblick in das Thema bieten. Es handelt sich hierbei um ein Feld, das aufgrund seiner Aktualität noch nicht wissenschaftlich erforscht wurde. Diese Untersuchung stellt daher keinen Anspruch auf Generalisierbarkeit, sondern soll lediglich dazu dienen, meine Analyse um die Perspektive von Lehrkräften zu ergänzen. Meine Hypothesen werden demnach nicht nach streng wissenschaftlichen Kriterien verifiziert oder falsifiziert, sondern erweitert, einem Vergleich unterzogen und vorläufig bestätigt oder widerlegt.

Im Zentrum des Interesses meiner empirischen Untersuchung stehen die Fragen: Wie hat sich der Backwasheffekt durch die Reform verändert? Wird der Effekt im Vergleich zur alten Prüfung heute als positiver oder negativer wahrgenommen? Welche konkreten Auswirkungen hat die reformierte Prüfung auf den Unterricht und wie wird das kritisierte *Teaching to the Test* davon beeinflusst?

## 2 Theoretischer Hintergrund

### 2.1 Testgütekriterien

Testgütekriterien sind insofern für diese Arbeit wichtig, als sich die Kriterien für positiven Backwash mit manchen Testgütekriterien überschneiden. Von besonderem Interesse sind die kommunikativ ausgerichteten Testgütekriterien, da es sich bei den Sprachprüfungen des GCSE offiziell um kommunikative Prüfungen<sup>4</sup> handelt. Sollte es sich herausstellen, dass dieser Anspruch nicht erfüllt wird, so ist das GCSE nicht konstruktvalide (siehe Abschnitt 2.1.1.3), was zu einem negativen Backwasheffekt führt.

#### 2.1.1 Hauptgütekriterien

In der klassischen Testtheorie werden zumeist drei Hauptgütekriterien genannt: Objektivität, Reliabilität und Validität. In der englischsprachigen Literatur werden dabei die Reliabilität und die Validität häufig gesondert hervorgehoben, während die Objektivität vernachlässigt wird. Dies ist damit zu erklären, dass manche AutorInnen die Objektivität in den anderen beiden Hauptkriterien integriert sehen.<sup>5</sup>

##### 2.1.1.1 Objektivität

Nach Sacher (2004) bezeichnet die Objektivität einer Messung „den Grad, in welchem ihre Ergebnisse unabhängig von der Person des Messenden sind.“<sup>6</sup> Sie lässt sich weiter in die Durchführungs-, Auswertungs- und Interpretationsobjektivität unterteilen. Um die Durchführungsobjektivität zu gewährleisten muss beispielsweise reglementiert sein, welche Hilfsmittel erlaubt sind und wie lange die Bearbeitung dauern darf. Eine höhere Auswertungsobjektivität erreicht man, wenn vor der Prüfung festgelegt wird, welche Antworten als richtig oder falsch gelten und wie welche Zweifelsfälle gehandhabt werden. Für die Interpretationsobjektivität ist eine gemeinsame Bewertungsgrundlage notwendig, nach der gleiche Leistungen zu gleichen Schlussfolgerungen (z.B. Noten) führen.<sup>7</sup>

---

<sup>4</sup> Obwohl mir die unterschiedlichen Definitionen für ‚Test‘ und ‚Prüfung‘ (Vgl. z.B. Krause / Sändig 2002: 172, 174) bewusst sind, verwende ich die Begriffe in dieser Arbeit synonym.

<sup>5</sup> Vgl. Bachman (1990), Hughes (2003) einerseits und Sacher (2004), Jürgens (2005) andererseits; zur Klarstellung: Krause / Sändig (2002), S. 84.

<sup>6</sup> Sacher (2004), S. 35.

<sup>7</sup> Vgl. Grotjahn (2000), S. 309-310; Jürgens (2005), S. 74-75; Sacher (2004), S. 35-36.

### 2.1.1.2 Reliabilität

„Die Reliabilität, bzw. Zuverlässigkeit einer Messung bezeichnet die Genauigkeit und Sicherheit“<sup>8</sup> mit der eine bestimmte Eigenschaft gemessen wird. Eine große Zahl an Items, die diese Eigenschaft messen, ermöglicht eine höhere Reliabilität. Die Reliabilität kann mit unterschiedlichen Verfahren gemessen werden, die als Wiederholungsmethode, Paralleltestmethode und Halbierungsmethode bezeichnet werden. Die Retest- bzw. Wiederholungstestreliabilität wird gemessen, indem derselbe Test zu einem späteren Zeitpunkt noch einmal durchgeführt wird, was allerdings durch zwischenzeitlichen Unterricht zu fehlerhaften Resultaten führen kann. Für die Bestimmung der Paralleltestreliabilität benötigt man zwei strukturäquivalente Tests, die gleichzeitig in derselben LernerInnengruppe durchgeführt werden. Bei der Halbierungsmethode werden schließlich zwei Hälften eines Tests getrennt ausgewertet, wobei sich diese Hälften in ihrer Struktur gleichen müssen. Die Reliabilität wird mit einem Reliabilitätskoeffizienten in Form einer Ziffer zwischen 0 (nicht zuverlässig) bis 1 (absolut zuverlässig) angegeben. Generell gilt, dass kein Test absolut reliabel ist; so soll etwa der Koeffizient guter Vokabel-, Struktur- und Leseverstehenstests zwischen 0.90 und 0.99 liegen, während Testungen mündlicher Sprachproduktion meist eine Reliabilität zwischen 0.70 und 0.79 aufweisen. Wie der Koeffizient berechnet werden kann, sowie weitere Details zur Reliabilität, können Bachman (1990) entnommen werden.<sup>9</sup>

### 2.1.1.3 Validität

Sacher (2004) beschreibt Validität folgendermaßen: „Validität bzw. Gültigkeit einer Messung ist dann gegeben, wenn Gewähr leistet ist, dass tatsächlich das gemessen wird, was man vorgibt zu messen“.<sup>10</sup> Grundsätzlich wird zwischen Inhaltsvalidität, kriterienbezogener Validität, Augenscheinvalidität und Konstruktvalidität unterschieden, wobei unterschiedliche Kategorisierungen und Termini in Verwendung sind.<sup>11</sup> Für diese Arbeit sind ausschließlich die Inhaltsvalidität und die Konstruktvalidität von Bedeutung, wobei die Konstruktvalidität nach Hughes (2003) ein übergeordnetes Kriterium darstellt und meines Erachtens tatsächlich eine zentrale Rolle spielen kann (siehe Abbildung 1, Einleitung).

---

<sup>8</sup> Sacher (2004), S. 36.

<sup>9</sup> Vgl. Bachman (1990), S. 160-235; Grotjahn (2000), S. 310-312; Hughes (2003), S. 38-40.

<sup>10</sup> Sacher (2004), S. 37.

<sup>11</sup> Vgl. z.B. externe und interne Validität bei Alderson / Clapham / Wall (1995), S. 171; Testfairness bei Sacher (2004), S. 38-41. Meine Unterteilung basiert auf Grotjahn (2000), S. 312-313.



Für die Inhaltsvalidität ist die Relevanz (vgl. ‚content relevance‘) und Repräsentativität (vgl. ‚content coverage‘) des Inhalts in Bezug auf die Testspezifikationen, den Lehrplan oder generell den Unterrichtsstoff ausschlaggebend.<sup>12</sup> Sie bezieht sich damit auf die Frage, inwiefern sich der Prüfungsstoff auf das gesamte Spektrum des relevanten Unterrichtsstoffes bezieht und nicht etwa nur auf Teilbereiche oder Nebensächliches. Unter ‚Inhalt‘ werden hierbei nicht nur Themen, sondern ebenso Fertigkeiten und Strategien verstanden (z.B. unterschiedliche Lesestrategien). Wie später noch gezeigt wird, ist die Inhaltsvalidität auch ein wichtiges Kriterium für positiven Backwash.<sup>13</sup>

Konstruktvalidität kann auf unterschiedliche Arten festgestellt werden. Eine Möglichkeit besteht darin, zu überprüfen, inwiefern der Test tatsächlich auf der zugrundeliegenden Theorie (i.e. auf dem jeweiligen theoretischen Konstrukt) basiert. Nach Hughes (2003) ist die Konstruktvalidität ein übergeordnetes Kriterium, das sich aus den anderen zusammensetzt. Konstruktvalidität ist für diese Arbeit wichtig, da die Frage, ob es sich beim GCSE tatsächlich um eine kommunikative Prüfung handelt, um eine Frage nach der Konstruktvalidität handelt. In vielen Fällen werden als Konstrukt Modelle kommunikativer Kompetenz, wie beispielsweise von Bachman (1990) oder Bachman / Palmer (1996), herangezogen, die in Abschnitt 2.2.1 erklärt werden. Normalerweise wird das Konstrukt in den Testspezifikationen angegeben.<sup>14</sup>

Diese drei Hauptgütekriterien sind voneinander abhängig: die Objektivität beeinflusst die Reliabilität und diese beeinflusst wiederum die Validität in der Art und Weise, dass ein Test nie reliabler als objektiv und nie valider als reliabel sein kann. Ein Test mit hoher Validität erfüllt demnach immer auch die anderen beiden Hauptgütekriterien, doch ein objektiver und reliabler Test muss noch lange nicht valide sein. Interessant ist das Verhältnis von Reliabilität und Validität, denn ein Kriterium kann immer nur zu Lasten des anderen maximiert werden. Ein besonders valider Test ist also nicht besonders reliabel, sowie auch ein besonders reliabler Test nicht besonders valide ist. Dieses Spannungsverhältnis zeichnet sich auch im

---

<sup>12</sup> Für die Begriffe ‚content relevance‘ und ‚content coverage‘ cf. Bachman (1990), S. 244-247.

<sup>13</sup> Vgl. Alderson / Clapham / Wall (1995), S. 173-176.

<sup>14</sup> Vgl. Alderson / Clapham / Wall (1995), S. 183-186; Grotjahn (2000), S. 312-315; Hughes (2003), S. 26-35; Jürgens (2005), S. 77-81; Sacher (2004), S. 37-41.

Verhältnis der kommunikativen Gütekriterien zu den Kriterien Objektivität und Reliabilität aus.<sup>15</sup>

### **2.1.2 Nebengütekriterien**

Abseits der Hauptgütekriterien gibt es noch eine Reihe von Nebengütekriterien, deren Einschätzungen aber noch weiter auseinander gehen. Diese wären etwa Praktikabilität, Ökonomie, Nützlichkeit und Zumutbarkeit, Akzeptanz, Transparenz, Normierung, Replikabilität und weiters mit dem Schwerpunkt auf kommunikativen Prüfungen die Authentizität, Interaktivität und Lernerzentriertheit. Es würde jedoch den Rahmen dieser Arbeit sprengen, all diese genauer zu erläutern.<sup>16</sup>

### **2.1.3 Gütekriterien für kommunikatives Testen und Prüfen**

Für das kommunikative Testen und Prüfen sind spezielle Gütekriterien notwendig, mit denen sich besonders Weir (1990), Glaboniat (1998), Kühhaas (2000), Krause / Sändig (2002) und Prantl (2008) beschäftigt haben. Die Bedingungen variieren zwar etwas zwischen den verschiedenen AutorInnen, doch gibt es auch Merkmale, über die weitgehend Einigkeit herrscht. Diese wären: Interaktivität, Authentizität, Integrativität und direktes Testen. Es wird ebenso von mehreren AutorInnen darauf aufmerksam gemacht, dass der Backwasheffekt zu berücksichtigen ist. Kühhaas nennt weiters das Kriterium Unvorhersehbarkeit, dem im Rahmen des GCSE eine besondere Rolle zukommt.<sup>17</sup>

Glaboniat nennt noch die Lernerzentriertheit, Prantl das kriteriumsorientierte Messen und Kühhaas die Orientierung an realen kommunikativen Bedürfnissen und den kommunikativen Zweck der Aufgabe. Auch eine qualitative Bewertung der Leistungen wird vorgeschlagen. Auf diese Kriterien werde ich allerdings nicht näher eingehen, da sie für die vorliegende Arbeit unerheblich sind.<sup>18</sup>

#### **2.1.3.1 Authentizität**

Authentizität meint Realitätsnähe bzw. –bezogenheit und ist unumstritten ein wichtiges Merkmal kommunikativ orientierter Tests, wobei hundertprozentige

---

<sup>15</sup> Vgl. Alderson / Clapham / Wall (1995), S. 186-188; Bolton (1985), S. 63, 82-85.

<sup>16</sup> Vgl. Bolton (1997), S. 19-22; Glaboniat (1998), S. 48-52; Grotjahn (2000), S. 317-318; Krause / Sändig (2002), S. 82-91; Preußler (1997), S. 14; Rauscher (2001), S. 24-27; Sacher (2004), S. 42.

<sup>17</sup> Vgl. Glaboniat (1998), S. 50-51; Krause / Sändig (2002), S. 82-92; Kühhaas (2000), S. 31-41; Prantl (2008), S. 11-24; Weir (1990), S. 10-14.

<sup>18</sup> Ebenda.

Authentizität in Prüfungssituationen nie erreicht werden kann. Unterschieden wird hier zwischen Authentizität von Texten und Authentizität von Aufgaben, wobei sich letztere auf die Realitätsnähe der Situation bezieht. Nach Krause / Sändig (2002) handelt es sich um eine authentische Kommunikationssituation, „wenn der Kommunikationsanlass echt (nicht simuliert bzw. gestellt) ist, ein reales Kommunikationsziel existiert und das Kommunikationsmedium adäquat ist“.<sup>19</sup> Die größte Schwierigkeit kommunikativer Tests besteht im Erreichen realer Kommunikationsanlässe und –ziele. Wie Spolsky (1985) formuliert: „Any language test is by its very nature inauthentic“.<sup>20</sup> Damit drückt er aus, dass es sich bei Sprachprüfungen immer um künstliche Sprechanlässe handelt, bei denen es ausschließlich darum geht, sprachliche Fertigkeiten unter Beweis zu stellen, während der eigentliche Informationsaustausch in Wahrheit keine Rolle spielt. Dies ist per definitionem nicht authentisch.

Trotzdem kann versucht werden, auf künstliche Weise möglichst reale Sprechanlässe zu schaffen, wie es etwa beim Aufgabentypus ‚Information Gap‘ der Fall ist. Dieser Typus gewährleistet, dass es ein reales Kommunikationsziel gibt, weil die beiden interagierenden Personen einen unterschiedlichen Wissensstand haben, den sie ausgleichen möchten. Auch Interviews könnten hier eingereicht werden. Solche didaktisch aufbereiteten Kommunikationsanlässe werden als ‚quasiauthentisch‘ bezeichnet.<sup>21</sup>

Durch das Üben und Wiederholen solcher Gesprächsaufgaben wird ein reales Kommunikationsziel wieder aufgehoben, da die Antworten schon bekannt sind. Das ‚Information Gap‘, also die Informationslücke, ist schließlich bereits geschlossen. Exzessives *Teaching to the Test*, bei dem Gespräche und Interviews zum gleichen Thema immer wieder wiederholt werden, steht damit im Kontrast zur Authentizität.

Leichter zu erreichen ist Authentizität in Texten, also in Lese- oder Hörverstehensaufgaben. Hier ist der Entstehungszusammenhang der Texte relevant: Texte, die nicht gezielt für den Fremdsprachenunterricht erstellt wurden, sondern aus dem Alltag von MuttersprachlerInnen stammen, gelten als authentisch. Auch hier gibt es verschiedene Zwischenstufen, wenn etwa authentische Texte (Zeitungsartikel, U-

---

<sup>19</sup> Krause / Sändig (2002), S. 85.

<sup>20</sup> Spolsky (1985), S. 39.

<sup>21</sup> Vgl. Grotjahn (2000), S. 318-319; Krause / Sändig (2002), S. 85-89.

Bahn-Durchsagen, u.ä.) für den Unterricht oder die Prüfung vereinfacht oder auf sonstige Weise bearbeitet werden. Hier spricht man dann von Semiauthentizität.<sup>22</sup>

Im Zusammenhang mit authentischen Texten nennt Bolton auch ganzheitlichere Testaufgaben, die etwa das Globalverstehen testen, und nicht nur gezielt einzelne Wörter und grammatische Strukturen, die bereits als bekannt vorausgesetzt werden.<sup>23</sup>

### **2.1.3.2 Interaktivität**

Ein weiteres Gütekriterium ist die Interaktivität. Die traditionelle Unterteilung in produktive und rezeptive Fähigkeiten entspricht nicht der realen Sprachverwendung, die meist interaktiv ist. Im weitesten Sinne sind alle Äußerungen interaktiv, da selbst in Textproduktionsaufgaben der Rezipient beachtet werden muss, um einen situations- und partneradäquaten Text zu erstellen.<sup>24</sup> Eine engere Definition ist hier allerdings hilfreicher: Das angemessene Reagieren auf die Äußerungen des Gegenübers und somit der ständige Wechsel zwischen Produktion und Rezeption ist dann die Grundlage für das Gelingen von Kommunikation. Auch die Verwendung von sogenannten Kompensations- oder Umwegstrategien, die Teil der strategischen Kompetenz<sup>25</sup> sind, fällt in den Bereich ‚Interaktion‘. Es gibt offensichtlich unterschiedlich starke Ausprägungen von Interaktivität, beispielsweise im Falle eines simulierten Interviews im Vergleich zu einem realen Interview, dessen Verlauf vom Prüfling mitbestimmt werden kann. Kühhaas (2000) macht hier darauf aufmerksam, dass „Interaktion bedeutet [...], dass der Kandidat auch selbst Fragen stellen, seine eigene Meinung und seine eigenen Vorstellungen einbringen kann“<sup>26</sup>. Demnach ist das Kriterium Interaktivität in simulierten Interviews und Gesprächen, deren Verlauf nicht mitbestimmt werden kann, nicht erfüllt.

### **2.1.3.3 Integrativität**

Häufig wird unter diesem Kriterium lediglich verstanden, dass sich VertreterInnen integrativer Ansätze gegen das isolierte Testen verschiedener Teilkompetenzen, u.a. etwa auch gegen grammatische Subtests aussprechen. Doch Integrativität geht noch weiter:

---

<sup>22</sup> Vgl. Glaboniat (1998), S. 70-72; Krause / Sändig (2002), S. 85-89.

<sup>23</sup> Vgl. Bolton (1997), S. 19-22.

<sup>24</sup> Vgl. Glaboniat (1998), S. 51-52.

<sup>25</sup> siehe Abschnitt 2.2.1.

<sup>26</sup> Kühhaas (2000), S. 33.

Integrativität bezieht sich auf die Forderung, den Lerner als Subjekt mit all seinen Kompetenzen und Eigenschaften zu mobilisieren und in das Testgeschehen einzubeziehen, insbesondere natürlich mit seiner kommunikativen Kompetenz.<sup>27</sup>

Hierbei sollen Bildungsprofil, kulturelle Prägung, Allgemeinwissen, Problemlösungsstrategien und noch vieles mehr berücksichtigt werden. Es könnte daher behauptet werden, dass die Integrativität mit der Validität unvereinbar ist, da für die Validität wichtig ist, „dass tatsächlich das gemessen wird, was man vorgibt zu messen“<sup>28</sup>. Wer vorgibt, Sprachkenntnisse zu testen aber bei der Prüfung eigentlich Allgemeinwissen verlangt, erstellt theoretisch keinen validen Test. Dabei kritisiert Sacher (2004) unter anderem, dass die Selektionsfunktion der Schule genau dadurch noch erhöht wird, dass im Sprachunterricht häufig Kompetenzen geprüft werden, die außerhalb der Schule erworben werden müssen.<sup>29</sup> Krause / Sändig (2002) sprechen in der Tat von einer „partielle[n] Unvereinbarkeit der einzelnen Testqualitäten“<sup>30</sup>. Welchen Merkmalen in einem speziellen Test der Vorrang gegeben wird, soll jedenfalls von Ziel und Zweck der Überprüfung abhängen.<sup>31</sup>

#### **2.1.3.4 Unvorhersehbarkeit**

Besonders wichtig erscheint das Kriterium der Unvorhersehbarkeit. Da „beinahe jede authentische Sprachproduktion extrem variabel, komplex, abhängig von äußeren Faktoren, ungeplant und unvorhersehbar ist“<sup>32</sup>, sollten auch TestkandidatInnen die Freiheit haben, ihre Aussagen nach Belieben zu formulieren. Dies steht im Konflikt mit Testaufgaben, die gleichzeitig bestimmte grammatikalische Strukturen verlangen. Unvorhersehbarkeit kann aber auch noch etwas anderes meinen: die Unvorhersehbarkeit, d.h. Unberechenbarkeit, der Aufgaben. Bei letzterem handelt es sich allerdings nicht um ein Gütekriterium für kommunikatives Prüfen, sondern um eine Bedingung für positiven Backwash (siehe Abschnitt 2.3.4).

---

<sup>27</sup> Krause / Sändig (2002), S. 89.

<sup>28</sup> Sacher (2004), S. 37.

<sup>29</sup> Vgl. Sacher (2004), S. 38.

<sup>30</sup> Krause / Sändig (2002), S. 91.

<sup>31</sup> Vgl. Krause / Sändig (2002), S. 89-92

<sup>32</sup> Kühhaas (2000), S. 37.

#### 2.1.4 Zusammenfassung: Balance der Merkmale

Eine Balance zwischen den verschiedenen Eigenschaften zu finden ist nicht nur in der Praxis, sondern auch in der Theorie nicht einfach. Krause / Sändig (2002) sind beispielsweise der Überzeugung, dass es „den idealen Test“<sup>33</sup> nicht gibt und je nach Verwendungszweck immer ein Merkmal zum Nachteil eines anderen dominieren wird. Speziell wird hier das Verhältnis der kommunikativ orientierten Qualitätsmerkmale zu den übrigen Gütekriterien hervorgehoben.<sup>34</sup>

Insgesamt kann man zusammenfassen, dass sich jene Gütekriterien, die sich auf die Kommunikationsfähigkeit beziehen, eher von der Überprüfung einzelner Teilkompetenzen entfernen und stattdessen stärker einen ganzheitlichen Ansatz verfolgen: sei es durch das vermehrte Prüfen von Globalverstehen, durch den Vorzug von Interaktivität gegenüber rein produktiven und rezeptiven Aufgaben oder durch den integrativen Zugang im Allgemeinen. Im Gegensatz dazu stehen die Hauptgütekriterien, bei denen Messbarkeit im Vordergrund steht, was häufig zu einer verstärkten Aufteilung in überprüfbare Teilbereiche führt. Problematisch ist das Verhältnis der beiden kommunikativen Kriterien Integrität und Unvorhersehbarkeit: Einerseits sollen in integrativen Prüfungen grammatische Subtests vermieden werden, doch andererseits kann auch nicht verlangt werden, dass in einem offenen Dialog bestimmte grammatische Strukturen angewendet werden. Dieses Spannungsfeld von Grammatik in kommunikativen Tests wird uns noch bei unserer Analyse des GCSE begegnen.

Ebenso wie sich im Laufe der Geschichte auch die Bevorzugung verschiedener Gütekriterien verändert hat, haben sich auch die Testmethoden geändert: Von den psychometrisch-strukturalistisch beeinflussten ‚discrete point‘-Tests, die hauptsächlich präzise, reliabel und wissenschaftlich sein sollten, über pragmatisch-integrative Tests bis hin zu kommunikativen Tests. Heute stehen diese Testmethoden selbstverständlich parallel zur Auswahl, je nachdem welcher theoretischen Schule man angehört.<sup>35</sup>

In ‚Die Gütebestimmung kommunikativer Tests‘ schreibt Bolton (1985), dass in kommunikativen Tests Reliabilität und Objektivität nicht mehr das Hauptanliegen der

---

<sup>33</sup> Krause / Sändig (2002), S. 91. Hervorhebung im Text durch Krause / Sändig.

<sup>34</sup> Vgl. Grotjahn (2000), S. 320-322; Krause / Sändig (2002), S. 91-92; Rauscher (2001), S. 24-27.

<sup>35</sup> Vgl. Brown (1996), S. 23-26; McNamara (1990), S. 13-17.

TestkonstrukteurInnen sein sollten, und zwar insbesondere, wenn die interaktiven Fähigkeiten überprüft werden sollen. Die auch von Krause / Sändig (2002) thematisierte Unvereinbarkeit von Objektivität und Reliabilität mit kommunikativen Kriterien kann also nur durch eine klare Bekennung zu letzteren gelöst werden.<sup>36</sup>

Zumindest beim alten GCSE handelt es sich um einen hochgradig standardisierten Test, der allerdings auf kommunikativen Prinzipien aufbauen soll. Das überarbeitete GCSE lässt den Schulen hingegen wieder mehr Freiraum. Interessant ist nun zu untersuchen, wie dieses Spannungsfeld zwischen Kommunikativität und Objektivität gemeistert wird.

---

<sup>36</sup> Vgl. Bolton (1985), S. 63, 82-85.

## 2.2 Der kommunikative Ansatz

### 2.2.1 Geschichte und Modelle

Der Begriff der ‚Kommunikativen Kompetenz‘ geht auf wissenschaftliche Auseinandersetzungen zwischen Noam Chomsky und Dell Hymes in den Sechziger- und Siebzigerjahren des zwanzigsten Jahrhunderts zurück. Damals entwickelte Chomsky das Konzept der ‚linguistischen Kompetenz‘, in dem von einem idealen Sprecher-Hörer ausgegangen wird. Hymes reagierte darauf, indem er Chomskys Konzept um die pragmatische Ebene erweiterte und so den Begriff ‚Kommunikative Kompetenz‘ schuf.<sup>37</sup> Das Konzept wurde schließlich immer weiter entwickelt: In den 80er-Jahren übte das Modell von Canale / Swain (1980) großen Einfluss aus, das die Kommunikative Kompetenz in drei Bereiche unterteilt: grammatische bzw. linguistische Kompetenz, soziolinguistische Kompetenz sowie strategische Kompetenz. Weiters wird zwischen Kompetenz und Performanz unterschieden. Das Modell wurde von Canale um die Diskurskompetenz erweitert und schließlich durch das Bachman-Modell von 1990 ersetzt. Das Modell von Bachman (1990) gliedert die Teilkompetenzen neu. Wichtig ist, dass Bachman die strategische Kompetenz ins Zentrum der kommunikativen Sprachfähigkeit (‚communicative language ability‘) rückt, wie folgendes Diagramm veranschaulicht:<sup>38</sup>

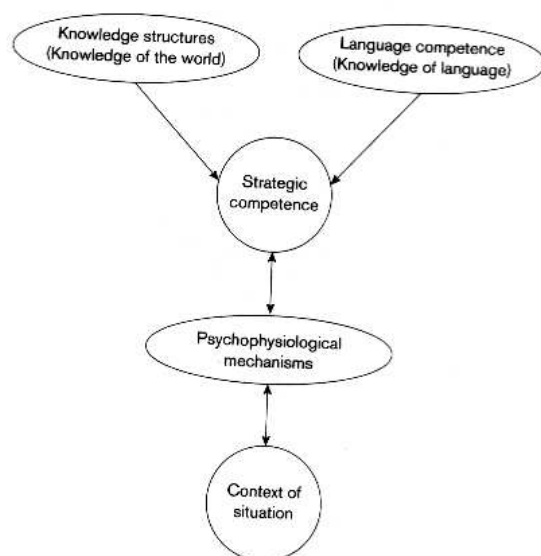


Abbildung 2: Bachmans Modell der 'communicative language ability'<sup>39</sup>

<sup>37</sup> Vgl. Werlen (2007), S. 16-19.

<sup>38</sup> Vgl. Bachman (1990), S. 81-100; Byram (2004), S. 125-126; Mitchell (1994); Skehan (1991), S. 8.

<sup>39</sup> Bachman (1990), S. 85.



Bachmans Modell gilt als weit verbreitet und diente bereits vielen kommunikativen Tests als Konstrukt (siehe Abschnitt 2.1.1.3). Die strategische Kompetenz beinhaltet sowohl Kompensations- und Umwegstrategien für Situationen, in denen die Kommunikation sonst nicht gelingen könnte, als auch Strategien zur Verbesserung und Erleichterung der Kommunikation allgemein, was etwa auch nonverbale Kommunikation mit einschließt.<sup>40</sup> In den erstgenannten Bereich fällt beispielsweise die Bitte um Wiederholung oder Neuformulierung, die gerade für Sprachlernende sehr wichtig ist. Es scheint daher verständlich, dass der Einbezug der strategischen Kompetenz für das kommunikative Testen und Prüfen notwendig ist; insbesondere für das Kriterium Interaktivität.<sup>41</sup>

### 2.2.2 Grammatik im kommunikativen Unterricht

Entgegen der häufig vorherrschenden Ansicht, dass formale Aspekte keine Rolle spielen, ist, wie oben beschrieben, die grammatische bzw. linguistische Kompetenz vertreten. Dazu Hedge: „An important point for the teacher to note is that linguistic competence is an integral part of communicative competence“.<sup>42</sup> Ausschlaggebend für kommunikative Ansätze ist nämlich nicht das Fehlen von Grammatik, von schriftlichen Texten oder Instruktionen durch Lehrende, sondern eine Erweiterung um vorher vernachlässigte Bereiche. Wie auch in Byram (2004) bemerkt wird:

CLT [Communicative Language Teaching, Anmerkung von mir] is not exclusively concerned with face-to-face oral communication [...] CLT does not require small group or pair work [...] CLT does not exclude a focus on metalinguistic awareness or knowledge of rules of syntax, discourse and social appropriateness.<sup>43</sup>

Dass durch eine Erweiterung um neue Bereiche (z.B. mündliche Kommunikationsfähigkeit) andere Bereiche in den Hintergrund rücken müssen, ist angesichts gleichbleibender Unterrichtsdauer selbstverständlich. In der Praxis wurde dies allerdings teilweise stark übertrieben, was dem kommunikativen Ansatz ab den 1990er-Jahren viel Kritik einbrachte und etwa auch zu Hedges Richtigstellung dieser scheinbaren „misconception about communicative language teaching“<sup>44</sup> führte.

---

<sup>40</sup> Vgl. Bachman (1990), S. 98-100.

<sup>41</sup> Vgl. Bachman (1990), S. 98-100.

<sup>42</sup> Hedge (2000), S. 47.

<sup>43</sup> Byram (2004), S. 128.

<sup>44</sup> Hedge (2000), S. 47.

### 2.2.3 Kritik am kommunikativen Ansatz für den Fremdsprachenunterricht in England

Auch in England erfuhr der kommunikative Fremdsprachenunterricht seit den 80er-Jahren einen enormen Aufschwung und ist bis heute offiziell der dominierende Ansatz.<sup>45</sup> Es gibt allerdings auch Kritikpunkte, die nicht außer Acht zu lassen sind. Wie Grenfell bereits 1994 schrieb: „Our National Syllabus – GCSE – turns out to be a transactional wolf in a communicative sheep’s clothing“<sup>46</sup>. Dieses treffende Zitat beschreibt, wie auch hier der Grammatikunterricht zugunsten eines reinen ‚Transaktionsunterricht‘ vernachlässigt wird, der allerdings nicht auf den kommunikativen Grundsätzen der Interaktivität und Authentizität beruht. Nach Grenfell (1994) ist der Unterricht vielerorts in den Drillübungen der direkten Methode stehengeblieben, ohne tatsächliche Konversation zu forcieren.<sup>47</sup>

Doch dies ist nicht der einzige Kritikpunkt. Es stellt sich die Frage, ob die Reduktion von Sprachenlernen auf das Ermöglichen von Kommunikation überhaupt für ein Land geeignet ist, in dem die heutige Lingua franca gesprochen wird. Seit etwa 20 Jahren lernen SchülerInnen im Deutschunterricht nun, mit welchen Phrasen sie im Hotel einchecken, im Restaurant Essen bestellen oder in der Touristeninformation nach Sehenswürdigkeiten fragen können. All das können sie in deutschsprachigen Ländern jedoch ebenso gut – wenn nicht besser – auch auf Englisch erledigen. Der Motivation ist das nicht gerade dienlich. Ein stärker linguistisch ausgerichteter Ansatz wäre in einem Land wie England eventuell sinnvoller.<sup>48</sup>

---

<sup>45</sup> Vgl. Mitchell (1994).

<sup>46</sup> Grenfell (1994), S. 54.

<sup>47</sup> Vgl. Grenfell (1994), S. 54-57.

<sup>48</sup> Vgl. Williams (2000).

## 2.3 Der ‚Backwasheffekt‘

Unter dem ‚Backwash-‘, ‚Washback-‘, oder auch ‚Rückkopplungseffekt‘<sup>49</sup> versteht man die Auswirkungen von Tests auf das Bildungssystem und dessen Institutionen, sowie auf die darin involvierten Personen, bzw. einfacher formuliert, auf das Lehren und Lernen.<sup>50</sup>

Grundsätzlich werden bei der Diskussion des Backwasheffekts zwei Fragen unterschieden: erstens, wen betreffen diese Auswirkungen, und zweitens, auf welche Art und Weise können verschiedene Bereiche betroffen sein?<sup>51</sup>

In Bezug auf die erste Frage unterscheiden Bachman / Palmer (1996) zwischen Auswirkungen auf Individuen und Auswirkungen auf das Bildungssystem und die Gesellschaft. Watanabe (2004) unterscheidet in Anlehnung an Bailey (1996) hingegen zwischen dem Einfluss auf die Lernenden und Auswirkungen auf das Programm, wobei letzteres die Lehrkräfte mit einschließt. Die Backwashforschung hat sich bisher auf die Auswirkungen auf das Programm konzentriert. Obwohl in dieser Arbeit versucht wird, auch LernerInnenstrategien zu erfassen (z.B. das Auswendiglernen), muss betont werden, dass Lernende selbst hier leider nicht zu Wort kommen, da mir der entsprechende Zugang zum Feld fehlt.<sup>52</sup>

### 2.3.1 Entwicklung und Strömungen der Backwashforschung

Im Jahr 1993 veröffentlichten Alderson und Wall ihren Artikel „Does Washback exist?“ und legten damit den Grundstein für diese neue Forschungsrichtung.

Eingehender beschäftigen sich damit heute Cheng, Curtis und Watanabe; Überblicksdarstellungen und einzelne Studien liegen weiters von Bailey, Hughes, Messick, Rea-Dickins / Scott sowie Shohamy / Donitsa-Schmidt / Ferman vor.<sup>53</sup> Im deutschsprachigen Raum hat Schiffko das Thema für das Handbuch ‚Deutsch als Fremdsprache‘ bearbeitet.<sup>54</sup> Insgesamt ist das Thema in Österreich jedoch ein

---

<sup>49</sup> Im englischsprachigen Raum ist der Terminus ‚Washback‘ geläufiger, im deutschsprachigen Raum hingegen ‚Backwash‘.

<sup>50</sup> Vgl. z.B. Bachman / Palmer (1996), S. 30-31, Cheng / Curtis (2004), S. 3; Hughes (2003), S. 53; Krause / Sändig (2002), S. 90.

<sup>51</sup> Vgl. z.B. Cheng / Curtis (2004); Cheng (2005); Watanabe (2004).

<sup>52</sup> Vgl. Bachman / Palmer (1996), S. 30-36; Watanabe (2004), S. 21.

<sup>53</sup> Vgl. z.B. Bailey (1996); Cheng / Curtis (2004); Cheng (2005); Hughes (2003); Messick (1996); Shohamy / Donitsa-Schmidt / Ferman (1996); Rea-Dickins / Scott (2007) und Watanabe (2004).

<sup>54</sup> Vgl. Schiffko (2001).

Randgebiet, zu dem kaum wissenschaftliche Literatur verfügbar ist. Das kann unter Umständen auf die (noch) fehlenden standardisierten Prüfungen zurückzuführen sein, deren zwar oft positiv intendierter, aber trotzdem häufig negativer Backwash in anderen Ländern mittlerweile ein populärer Forschungsgegenstand ist.

Grundsätzlich werden in der Backwashforschung zwei Hauptströmungen unterschieden: jene, die untersucht, ob der intendierte Backwash einer bestimmten Prüfung tatsächlich positiv ist, und jene Strömung, die sich hauptsächlich mit nicht intendiertem, negativem Backwash beschäftigt. Diese Arbeit steht in der Tradition der letzteren, da ich auf die Intensität und die Ausprägung von nicht intendiertem Backwash fokussiere.<sup>55</sup>

Die beiden Strömungen unterscheiden sich aber nicht nur hinsichtlich ihres Untersuchungsgegenstandes, sondern auch in ihrer Einstellung zum Backwash. Während eine Richtung der Ansicht ist, dass Tests den Unterricht tatsächlich beeinflussen sollen (vgl. measurement driven instruction) und dementsprechend gut ausgearbeitet sein müssen, spricht die andere Richtung eher von unerwünschten Nebenwirkungen, die unumgänglich sind. So tritt Spolsky beispielsweise dafür ein, nur die nicht intendierten – und daher oft eher negativen – Effekte als Backwash zu bezeichnen.<sup>56</sup>

### **2.3.2 Watanabes Backwashdimensionen**

Betreffend die Frage, auf welche Art und Weise verschiedene Bereiche betroffen sein können, nennt Watanabe in ‚Washback in language testing: research content and methods‘ fünf Dimensionen, in denen sich der Backwasheffekt unterscheiden kann:<sup>57</sup>

#### 1. Spezifität: allgemein – spezifisch

Der spezifische Backwasheffekt bezieht sich auf bestimmte Teile eines konkreten Tests, während der allgemeine Backwasheffekt durch mehrere unterschiedliche Testtypen ausgelöst werden kann.

#### 2. Intensität: stark – schwach

---

<sup>55</sup> Vgl. Cheng / Curtis (2004), S. 7-11.

<sup>56</sup> Vgl. Cheng / Curtis (2004), S. 3-14, Cheng (2005), S. 29-30, Spolsky (1994), S. 55 zitiert nach Cheng / Curtis (2004), S. 8.

<sup>57</sup> Vgl. Bachman / Palmer (1996), S. 30-36; Watanabe (2004), S. 20-21.

Während ein starker Backwascheffekt nahezu den gesamten Unterricht beeinflusst, hat ein schwacher Effekt nur Auswirkungen auf manche Lehrkräfte, auf manche SchülerInnen oder nur auf einzelne Teile des Unterrichts.

### 3. Länge: kurz – lang

Der Backwascheffekt dauert lange an, wenn er nicht nur auf die Zeit der speziellen Prüfungsvorbereitung beschränkt ist, sondern auch noch nach der Prüfung andauert.

### 4. Intentionalität: intendiert – nicht intendiert

Wie bereits oben angesprochen existieren bezüglich dieser Dimension unterschiedliche Strömungen.

### 5. Ausrichtung („value“): positiv – negativ

Der Effekt kann, je nach Qualität der Prüfung, positiv oder negativ sein, wobei meist beides der Fall ist. Prinzipiell gilt, dass qualitativ minderwertige Tests auf lange Sicht keine positiven Auswirkungen haben können, sich aber gut ausgearbeitete Tests trotzdem negativ auswirken können. Für letzteres sind unter anderem die testexternen Faktoren verantwortlich, wie etwa Prüfungsangst, die Vermittlungskompetenz der Lehrkraft oder ein Mangel an Ressourcen. Auch der Status einer Prüfung kann sich negativ auswirken. Weiters gilt, dass positiver Backwash meist intendiert und negativer Backwash nicht intendiert ist.<sup>58</sup>

Wichtig ist, dass diese Dimensionen miteinander in Zusammenhang stehen. So wäre etwa die Aussage ‚*High-Stakes-Testing* hat häufig einen starken Backwascheffekt‘ eine allgemeine Aussage, die sich auf die Intensität bezieht. Die Aussage ‚Die Rollenspiele der GCSE-Prüfung haben einen negativen Backwascheffekt‘ wäre hingegen eine spezifische, die sich auf die Ausrichtung bezieht.

### **2.3.3 Schifkos Aspekte des Backwascheffekts**

Schifko (2001) unterscheidet ebenfalls fünf verschiedene Aspekte, die er allerdings anders klassifiziert. Da wäre zunächst der statusbezogene Aspekt zu nennen, der in

---

<sup>58</sup> Vgl. Hughes (2003), S. 53; Krause (2002), S. 90; Schifko (2003), S. 14-15, Madaus / Russel / Higgins (2009), S. 139-164, Watanabe (2004), S. 21.

Zusammenhang mit Watanabes Intensität steht: Demnach sei der Effekt stärker, je wichtiger eine Prüfung ist. Weiters gibt es den curricularen Aspekt, welcher die Beeinflussung des Lehrplans beschreibt und mit Watanabes Auswirkungen auf das Programm zusammenhängt, sowie den didaktisch-methodischen, den lernstrategischen und den motivationspsychologischen Aspekt, wobei die letzten drei individuenbezogen sind. Mit Hilfe des statusbezogenen Aspekts lassen sich gut die Eigenschaften von *High-Stakes-Tests* erklären, weswegen ich dieses Kriterium für sinnvoll halte. Individuenbezogene Auswirkungen werden – obwohl durchaus relevant – von Watanabe nicht angesprochen, so dass ich diesen Aspekt Schifko (2001) entnehme. Abgesehen von diesen notwendigen Ergänzungen halte ich Watanabes Gliederung für sehr gelungen.<sup>59</sup>

### **2.3.4 Bedingungen für positiven Backwash**

Um zu untersuchen, ob ein Test tatsächlich jene positiven Effekte aufweist, die beabsichtigt waren, sind Analyse Kriterien notwendig. Sowohl Schifko (2001), Hughes (2003) als auch zumindest am Rande Bachman / Palmer (1996) listen solche Kriterien auf, die ich hier als Bedingungen für positiven Backwash bezeichne. Auffällig ist dabei der Zusammenhang mit den Testgütekriterien, der schon in der Einleitung skizziert wurde.

Erstens soll ein Test jene Fähigkeiten und Fertigkeiten prüfen, die man fördern möchte. So ist es beispielsweise sinnvoll, eine mündliche Prüfung durchzuführen, wenn auch die mündliche Kommunikationsfähigkeit gefördert werden soll. Das klingt zwar selbstverständlich, doch oft steht das Nebengütekriterium ‚Praktikabilität‘ im Vordergrund, so dass eher getestet wird, was leicht zu überprüfen ist. Diese Forderung ist mit der Forderung nach ‚direktem Testen‘ vergleichbar: wenn es etwa das Ziel ist, dass SchülerInnen wissenschaftliche Texte lesen können, dann soll auch die Testaufgabe darin bestehen und nicht etwa in bloßen Vokabelübungen zu diesem Thema.<sup>60</sup>

Zweitens soll die Auswahl der Aufgaben in Bezug auf Themen wie Sprachhandlungen repräsentativ und unvorhersehbar sein, was mit der Inhaltsvalidität – also der Repräsentativität und Relevanz des Inhalts in Bezug auf

---

<sup>59</sup> Vgl. Schifko (2001), S. 831.

<sup>60</sup> Vgl. Hughes (2003), S. 53-54; Schifko (2001), S. 832-833.

den zu messenden Bereich - in Zusammenhang steht. Doch nicht nur die Repräsentativität, sondern auch die Unvorhersehbarkeit muss gewährleistet sein. Fragenkataloge stellen demnach keine sinnvolle Grundlage für Sprachprüfungen dar, denn die Unvorhersehbarkeit ist nicht mehr gegeben, was z.B. Auswendiglernen ermöglicht.<sup>61</sup>

Ein weiterer Punkt ist die Kriteriumsorientiertheit. Im Gegensatz zu normorientierten Tests, die lediglich soziale Ränge innerhalb der Stichprobe ermitteln, beziehen sich kriterienbezogene Tests auf vorher festgelegte Standards. Die erreichte Note ist damit unabhängig vom Klassen-, Schul- oder nationalen Durchschnitt, sondern drückt aus, welcher Level erreicht wurde. Laut Hughes (2003) ist es sinnvoll, dass unterschiedlich schwierige Tests zur Auswahl stehen, wobei in jedem Fall der Großteil der Aufgaben richtig gelöst werden muss, um die Prüfung zu bestehen.<sup>62</sup>

Was Schiffko als ‚Testfairness‘ bezeichnet, beschreibt Hughes als Vertrautheit und Transparenz der Anforderungen. Dazu zählen aussagekräftige Punkte- und Prozentwerte, Informationen zu Hilfsmitteln, Ablauf und Bewertung sowie das Absolvieren eines Modelltests.<sup>63</sup>

Schließlich fordern beide mehr Unterstützung und Test-Know-How für Lehrkräfte.<sup>64</sup> Bachman / Palmer (1996) schlagen vor, die zu testenden Personen in die Erstellung von Prüfungen zu involvieren und sie zu ihrer Wahrnehmung des Tests zu befragen. Diese Forderung möchte ich u.a. mit meiner empirischen Untersuchung erfüllen, die sich mit den Meinungen von Lehrkräften zur Reform des GCSE beschäftigt, wenn ich auch keinerlei Möglichkeit habe, mittels dieser Erkenntnisse die Prüfung zu revidieren.

### 2.3.5 Negativer Backwash

Negativer Backwash wird hauptsächlich im Zusammenhang mit *High-Stakes-Testing* thematisiert. Unter *High-Stakes-Tests* werden in der englischsprachigen Literatur Prüfungen verstanden, deren Resultate zumindest in der Wahrnehmung der KandidatInnen große Auswirkungen auf das weitere Leben haben.<sup>65</sup> Meist dienen

---

<sup>61</sup> Vgl. Hughes (2003), S. 54; Schiffko (2001), S. 832.

<sup>62</sup> Vgl. Hughes (2003), S. 55; Schiffko (2001), S. 833.

<sup>63</sup> Vgl. Hughes (2003), S. 55-56; Schiffko (2001), S. 831-832.

<sup>64</sup> Vgl. Bachman / Palmer (1996), S. 32; Hughes (2003), S. 55-56; Schiffko (2001), S. 833-834.

<sup>65</sup> Vgl. Cheng (2005), S. 43.

diese Prüfungen als Grundlage für wichtige Entscheidungen, wie z.B. die Studienberechtigung. Diese haben – in Übereinstimmung mit dem oben beschriebenen statusbezogenen Aspekt – einen starken Backwasheffekt. Ein starker Backwasheffekt ist für sich genommen noch nichts Negatives, da die Ausrichtung entscheidend ist. Problematisch ist allerdings, dass bei *High-Stakes-Tests* vielfach ein negativer Backwash zu erkennen ist.<sup>66</sup> Nach Madaus / Russel / Higgins (2009) können u.a. die folgenden negativen Auswirkungen auftreten: die Einschränkung des Unterrichtsstoffes auf den Prüfungsstoff, die Anpassung des Lehrplans – auch von viel niedrigeren Schulstufen – an den Test, Auswendiglernen, Fokus auf *Test-Taking-Skills*, kommerzielle Verbreitung von Materialien für die Prüfungsvorbereitung, eine Ausrichtung des Unterrichts auf das Üben mit früheren Testbögen, sowie die Konzentration auf einzelne SchülerInnen, die den Test gerade noch bestehen könnten, während andere vernachlässigt werden.<sup>67</sup> Häufig werden einige dieser Auswirkungen schlicht als *Teaching to the Test* zusammengefasst. Zum Fokus auf *Test-Taking-Skills* nennen Madaus / Russel / Higgins Forschungsergebnisse: 85% der LehrerInnen in US-Bundesstaaten mit *High-Stakes-Tests* behandelten im Unterricht *Test-Taking-Skills*, aber nur 67% in *Low-Stakes*-Staaten. Dies ist insofern problematisch, als zwar die Testergebnisse besser zu werden scheinen, aber eigentlich die Validität der Ergebnisse darunter leidet.<sup>68</sup> Krause / Sändig (2002) thematisieren die Ausrichtung des Unterrichts auf High Stakes Tests im Zusammenhang mit dem motivationspsychologischen Aspekt: „[e]ine negative Haltung zu Tests [] auch dadurch aufgebaut werden [kann], dass – wie in manchen Ländern üblich – der gesamte Unterricht strikt auf zentrale Tests und Prüfungen ausgerichtet ist.“<sup>69</sup> Wie lange diese Ausrichtung auf den Test andauern kann, beschreiben Bachman / Palmer (1996):

In some countries, where high-stakes nation-wide public examinations are used for selection and placement into higher levels of the school system or into universities, teaching may be focused on the syllabus of the test for up to several years before the actual test, and the techniques needed in the test will be practised in class.<sup>70</sup>

---

<sup>66</sup> Vgl. Cheng (2005), S. 43-45; Madaus / Russel / Higgins (2009), S. 139-164; Wall (2005), S. 36-41.

<sup>67</sup> Vgl. Madaus / Russel / Higgins (2009), S. 139-164.

<sup>68</sup> Vgl. Madaus / Russel / Higgins (2009), S. 144, 155.

<sup>69</sup> Vgl. Krause / Sändig (2002), S. 90.

<sup>70</sup> Bachman / Palmer (1996), S. 31.



Das Zitat zeigt, dass die Phase des *Teaching to the Test* sogar mehrere Jahre andauern kann, wobei der Lehrstoff an den Prüfungsstoff angepasst, bzw. auf ihn reduziert wird, um stattdessen *Test-Taking-Skills* zu üben.

Ein besonders starker Druck zum *Teaching to the Test* entsteht dabei, wenn die Resultate auch dazu verwendet werden, Schulen und Lehrkräfte zur Rechenschaft zu ziehen, etwa durch Schulrankings oder notengebundene Bezahlung. Schon 1911 schrieb ein Kritiker über das ‚Payment by Results‘-System, das seit 1862 in britischen Grundschulen herrschte: „The children [...] were drilled in the contents of those books until they knew them almost by heart. [...] To get him through the yearly examination by hook or by crook was the one concern of the teacher.“<sup>71</sup> Für eine aktuellere Kritik an Noten als Basis für Schulbewertungen siehe z.B. Davis (1998). Schulrankings und ähnliche Bewertungsverfahren führen jedenfalls dazu, dass es für Lehrkräfte kaum noch eine Möglichkeit gibt, frei zu unterrichten, was etwa auch die intrinsische Motivation der Lehrkräfte schwächen kann. Die oben genannten negativen Effekte treten durch Schulrankings flächendeckender und verstärkt auf. Die empirische Untersuchung soll u.a. feststellen, wie die teilnehmenden DeutschlehrerInnen den Einfluss der League Tables – also der nationalen Schulrankings – einschätzen.<sup>72</sup>

*Teaching to the Test* wird demnach von mehreren AutorInnen negativ dargestellt. Interessant ist dies insofern, als vorher Vertrautheit („Testfairness“) als ein Kriterium für positiven Backwash ausgemacht wurde. Das Kriterium der Vertrautheit wird also damit wieder etwas relativiert, dass eine zu intensive Vertrautheit ebenfalls negative Auswirkungen hat. Auch das Kriterium der Unvorhersehbarkeit könnte davon negativ betroffen sein: was jahrelang intensiv als Prüfungsvorbereitung geübt wird, ist nicht mehr unvorhersehbar.

Doch nicht nur *Teaching to the Test* wird kritisiert. Mehrere AutorInnen erwähnen, dass fehlende Konstruktvalidität zu negativem Backwash führt, wobei Anderson / Wall betonen, dass daraus nicht der umgekehrte Schluss gezogen werden dürfe.

---

<sup>71</sup> Holmes (1911), S. 107-108

<sup>72</sup> Vgl. Cheng / Curtis (2004), S. 9; Wall (2005), S. 40.

Folglich hat ein Test, der den eigenen Anspruch kommunikativ zu sein, nicht erfüllt, ebenfalls negativen Backwash.<sup>73</sup>

### **2.3.6 Zusammenfassung**

In diesem Kapitel wurde ein kurzer Überblick über die Grundzüge der Backwashforschung geboten. Es wurde gezeigt, welche Bereiche von diesen Auswirkungen betroffen sein können und entlang welcher Dimensionen sich der Effekt unterscheiden kann. Für die weiteren Teile dieser Arbeit sind besonders die Bedingungen für positiven Backwash und die Merkmale von negativem Backwash wichtig, wobei negativer Backwash meist mit *Teaching to the Test* und *High-Stakes-Testing* zusammenhängt.

## **2.4 Fremdsprachenunterricht und Prüfungssysteme in England**

### **2.4.1 Die Situation von Fremdsprachen im Vereinigten Königreich**

Im Vergleich zum EU-Durchschnitt sind die Fremdsprachenkenntnisse im Vereinigten Königreich von Großbritannien und Nordirland (UK) sehr gering. Während laut einer Eurobarometer-Studie der Europäischen Kommission 50% der EuropäerInnen angeben, in mindestens einer Fremdsprache eine Konversation führen zu können, sind es im Vereinigten Königreich nur 30% - das Königreich liegt damit vor Ungarn an vorletzter Stelle der EU25.<sup>74</sup>

Das Ergebnis scheint verständlich, wenn man bedenkt, dass im Vereinigten Königreich eine Lingua franca gesprochen wird, deren Kenntnis für die Kommunikation mit weiten Teilen der Welt ausreicht.<sup>75</sup> Selbst vorhandene Deutschkenntnisse können kaum angewendet werden, da die Englischkenntnisse in deutschsprachigen Ländern oft höher sind, so dass hier mit britischen Urlaubern meist auf Englisch kommuniziert wird. Wenn Fremdsprachen für die Kommunikation nicht notwendig sind, trägt das nicht unbedingt zur Motivation bei.<sup>76</sup> Hier wäre auch die Kritik am (ausschließlich) kommunikativen Ansatz für den Fremdsprachenunterricht in einem englischsprachigen Land zu nennen (vgl. Abschnitt 2.2.3).

---

<sup>73</sup> Vgl. Anderson / Wall (1993), S. 116; Cheng / Curtis (2004), S. 9; Cheng (2005), S. 29; Purpura (2008), S. 304-305.

<sup>74</sup> Vgl. European Commission (2005), S. 3.

<sup>75</sup> Vgl. Wilson (2008), S. 111.

<sup>76</sup> Vgl. Wilson (2008), S. 111, 114.

Nicht außer Acht zu lassen ist allerdings auch die Stellung des Fremdsprachenunterrichts im Schulsystem. Insgesamt gilt Deutsch als sehr anspruchsvolles Fach und wird oft nur leistungsstarken SchülerInnen empfohlen.<sup>77</sup> Erst seit der Einführung eines Lehrplans für Fremdsprachen als Teil der ‚Languages for all‘-Direktive im Jahr 1992 ist der Fremdsprachenunterricht im Sekundarbereich verpflichtend, was einige für den Anstieg an Deutschlernenden in den 90er-Jahren verantwortlich machen. Macaro (2008) argumentiert allerdings mit Hilfe von Statistiken, dass nicht der verpflichtende Unterricht, sondern die Einführung des GCSE Ende der 80-er Jahre zu einem Anstieg an Deutschlernenden in der Key Stage 4 führte.<sup>78</sup> Seit Ende der 90-er Jahre fällt jedenfalls die Zahl der GCSE-Abschlüsse in Fremdsprachen wieder konstant. Um dem zu begegnen wurde zwar die ‚National Language Strategy‘<sup>79</sup> beschlossen, doch an dem Trend hat sich in den letzten Jahren leider nichts geändert: Im Gegenteil, seit 2004 Fremdsprachen in der ‚Key Stage 4‘ (10. und 11. Schulstufe) wieder optional wurden, verstärkte sich der Rückgang noch mehr.<sup>80</sup> Das bedeutet, dass heute nur mehr von der 7. bis zur 9. Schulstufe eine Sprache gelernt werden muss. Zumindest, dass seit dem Beschluss der ‚National Language Strategy‘ immer mehr Grundschulen Fremdsprachenunterricht anbieten, ist positiv zu bewerten.<sup>81</sup>

#### **2.4.2 Organisation und Umfang des Fremdsprachenunterrichts in Schulen: ein Fallbeispiel**

Aufgrund der hohen Schulautonomie ist es schwer, allgemein gültige Aussagen über den Deutschunterricht in England zu treffen. Es folgt daher ein Fallbeispiel, um einen Einblick in das System zu bekommen: der Fremdsprachenunterricht an der John Warner School, Hoddesdon. An dieser Schule war ich im Schuljahr 2008/2009 für acht Monate als Sprachassistentin tätig.<sup>82</sup>

Bis zum Schuljahr 2008/2009 wählten SchülerInnen der 7. Schulstufe an der John Warner School<sup>83</sup> zwei Fremdsprachen, von denen sie eine nach nur einem Jahr

---

<sup>77</sup> Vgl. Reershemius (2010), S. 1675.

<sup>78</sup> Vgl. Macaro (2008).

<sup>79</sup> Vgl. DfES (2002); Pachler (2002); Ofsted (2008).

<sup>80</sup> Vgl. Macaro (2008), S. 101-107.

<sup>81</sup> Vgl. European Commission 2008, Reershemius (2010), S. 1674-1680.

<sup>82</sup> Die Informationen können zwar keiner offiziellen Quelle entnommen werden, wurden aber von Vanessa Lamb (Head of MFL Department) in einem Gespräch bestätigt.

<sup>83</sup> Allgemeine Informationen zur John Warner School unter:

The John Warner School. 2010. <http://www.johnwarner.org.uk> (18. 3. 2010).

wieder abwählen konnten. Die zweite Fremdsprache besuchten sie mindestens bis zur 9. Schulstufe. Seit dem Schuljahr 2009/2010 wird in der 7. Schulstufe nur mehr Französisch angeboten. Insgesamt stehen die Sprachen Französisch, Italienisch und Deutsch zur Auswahl, wobei Französisch am weitesten verbreitet ist – das deckt sich auch mit den nationalen Statistiken.<sup>84</sup> In jedem Schuljahr bis zum Ende der Schulpflicht gibt es zwei Wochenstunden pro Fremdsprache. Nur wer sich nachher noch auf eine Fremdsprache spezialisieren möchte und seine A-Level-Prüfung darin ablegen will, hat mehr als zwei Wochenstunden in diesem Fach, denn erst ab dann gibt es fünf Wochenstunden.

An der John Warner School gibt es sechs verschiedene Leistungsgruppen („sets“), davon drei obere („top sets“) und drei untere („bottom sets“). In der 9. Schulstufe werden dafür zwei unterschiedliche Schulbücher verwendet.

In der 8. und 9. Schulstufe gibt es somit jeweils sechs Deutschklassen, in denen meistens 30 bis 35 SchülerInnen sitzen. In der 10. und 11. Schulstufe sind meistens etwa zwölf SchülerInnen in einer Deutschklasse und die Anzahl der Klassen ist bereits reduziert, später gibt es gar nur mehr zwei bis zwölf SchülerInnen in einem Jahrgang. Das deckt sich auch mit einer Untersuchung, nach der 2007 nur 2,2% eines Jahrgangs Deutsch für die A-Levels wählten.<sup>85</sup> Die meisten DeutschschülerInnen haben das Fach heute demnach nur mehr zwei Jahre lang zwei Stunden pro Woche.

### **2.4.3 Prüfungssystem und Qualifikationen**

Ich beziehe mich hier explizit auf England, da in Schottland, Wales und Nordirland zumindest in Teilbereichen anders geprüft wird, was eine Verallgemeinerung erschwert, wenn nicht gar unmöglich macht. Insbesondere seit 1999 verfolgen Schottland und Wales eine jeweils eigene Bildungspolitik mit zunehmend unterschiedlichen Lehrplänen, eigenen Schulabschlusszeugnissen und Qualifikationen. Wenn also die Rede von nationalen Tests ist, beziehen sich diese nicht auf das ganze Vereinigte Königreich, sondern nur auf England.<sup>86</sup>

---

<sup>84</sup> Vgl. Reershemius (2010), S. 1674.

<sup>85</sup> Vgl. Reershemius (2010), S. 1676

<sup>86</sup> Vgl. Hodgson / Spours (2008), S. 13-15.

Während der empirischen Untersuchung meldeten sich zwei Lehrkräfte aus Nordirland, die der Meinung waren, das nordirische Prüfungssystem wäre dem englischen in Bezug auf die Inhalte meiner Untersuchung ähnlich genug.

Die ‚Qualifications and Curriculum Development Agency‘ (QCDA) ist für die Entwicklung der nationalen Tests verantwortlich. Grundsätzlich unterschieden wird dabei zwischen den National Curriculum Tests in Mathematik, Englisch und Naturwissenschaften und Tests, mit denen gleichzeitig Qualifikationen erworben werden. Zu letzteren zählen u.a. das General Certificate of Secondary Education (GCSE), die A-Levels und das Diplom („Diploma“). Die National Curriculum Tests, die im Volksmund oft als ‚Sats‘ bezeichnet werden, sind zwar für Diskussionen des englischen Bildungs- und Prüfungssystems nicht unerheblich, doch für den Deutschunterricht irrelevant.<sup>87</sup>

Jene Tests, die als Qualifikationen dienen, sollen hier von Interesse sein. Deren Geschichte kann bei Wegner (1999), Koglbauer (2003), Hansen / Vignoles (2005) sowie bei Hodgson / Spours (2008) mit teilweise unterschiedlichen Schwerpunktsetzungen nachgelesen werden, weshalb ich hier lediglich die neueren Entwicklungen darstellen möchte. Im Jahr 2008 wurde vom Bildungsministerium (damals ‚Department for children, schools and families‘, kurz: DCSF) eine Strategie zur Erarbeitung neuer Qualifikationen für Vierzehn- bis Neunzehnjährige („strategy for 14-19 qualifications“) vorgestellt. Diese Strategie beinhaltet die Veränderung, Vereinfachung und leichtere Kombinierbarkeit verschiedener Qualifikationssysteme. Ziel ist, dass es bis 2013 vier Routen gibt, die zu unterschiedlichen Qualifikationen führen, und zwar die GCSE- / A-Level-Route, das Diplom, die Lehre und ‚Foundation Learning‘. Problematisch ist dabei der sogenannte ‚goldene Standard‘ der A-Levels, an dem alle anderen Qualifikationen gemessen werden. Dieser erschwerte es schon bisher den zahlreichen berufsbildenden Qualifikationen Anerkennung zu erlangen, weswegen GCSEs und A-Levels im Sekundarbereich dominieren: Praktisch alle SchülerInnen legen zumindest eine GCSE-Prüfung ab; 40% erlangen mindestens zwei A-Levels. Es ist fraglich, ob die erneute Überarbeitung die öffentliche Meinung ändern kann.<sup>88</sup>

#### **2.4.3.1 Das GCSE**

Das GCSE wird normalerweise am Ende der Schulpflicht mit etwa 15 oder 16 Jahren absolviert, kann aber auch im späteren Leben zu Zusatzqualifikationen führen. Es stehen über 50 Fächer in unterschiedlichen Kurslängen zur Verfügung, die je nach

---

<sup>87</sup> Vgl. DCSF (2010), QCDA (2009).

<sup>88</sup> Vgl. DCSF (2010); Hodgson / Spours (2008), S. 7-8.

Dauer unterschiedlich viel – zwischen  $\frac{1}{2}$  und 2 GCSEs - wert sind. Weiters gibt es zwei verschiedene Niveaustufen – ‚foundation tier‘ und ‚higher tier‘ – die zu unterschiedlichen Noten führen können. Wer die höhere Niveaustufe absolviert, kann ein A\*, A, B, C oder D erreichen, wer hingegen die leichtere Prüfung wählt, erhält lediglich ein C, D, E, F, oder G. Wer die minimale Punktegrenze für das entsprechende Niveau nicht erreicht, wird mit ‚U‘ (‚unclassified‘) negativ beurteilt. Die Anzahl der Fächer ist frei wählbar: der Großteil der SchülerInnen wählt mindestens 5 Fächer, eine Minderheit über 10 Fächer, wobei Einzelfälle sogar bis zu 16 Kurse belegen. Um für den weiterführenden Schulbesuch zugelassen zu werden sind üblicherweise fünf Noten von A\* bis C erforderlich. Insgesamt gibt es damit bei der Fächerauswahl, Kursdauer und dem zu erzielenden Niveau eine große Flexibilität.<sup>89</sup>

Da sich der konkrete Aufbau der Deutschprüfung mit der Überarbeitung des GCSE stark verändert hat und diese Entwicklungen für meine Erhebung sehr wichtig sind, wird dem Prüfungsaufbau und der Gegenüberstellung der beiden Konzepte ein eigenes Kapitel gewidmet.

#### **2.4.3.2 Die A-Levels**

Nach dem Ende der Schulpflicht besteht die Option, die Schulausbildung fortzusetzen und mit den A-Level-Prüfungen die Hochschulreife zu erwerben. Auch hier werden zwei unterschiedliche Kurslängen angeboten: ein einjähriger Kurs, der zur AS-Level-Prüfung führt, und darauf aufbauend ein weiterer einjähriger Vorbereitungskurs für die A2-Prüfung. Die meisten SchülerInnen, die sich für den Weg zur Hochschulreife entscheiden, absolvieren vier bis fünf AS-Prüfungen und wählen dann in Folge aus diesen Fächern zwei bis drei Kurse für den A2-Level. Je nachdem, wie viele Prüfungen sie in welchen Fächern mit welchen Noten absolvieren, gestalten sich auch die Möglichkeiten eine bestimmte Universität zu besuchen. Für Cambridge sind beispielsweise mindestens vier ‚A‘-Noten auf AS-Level und ein ‚A\*‘ sowie zwei ‚A‘-Noten auf A2-Level in bestimmten Fächerkombinationen notwendig.<sup>90</sup> Dementsprechend hoch ist der Druck, viele gute Noten zu sammeln.<sup>91</sup>

---

<sup>89</sup> Vgl. Hodgson / Spours (2008), S. 40; Obermaier (1999), S. 22-25; Ofqual. GCSEs. The official student guide to the system.

<sup>90</sup> Vgl. University of Cambridge (2009).

<sup>91</sup> Vgl. Koglbauer (2003), S. 29-33; Obermaier (1999), S. 22-25.

### 2.4.3.3 Exam Boards

Exam Boards sind für die Erstellung, Verteilung und Korrektur der nationalen Tests verantwortlich. Derzeit arbeitet die QCDA mit fünf Exam Boards zusammen, wovon eines für Wales und eines für Nordirland zu ständig ist.<sup>92</sup> In England stehen demnach drei Prüfungsanbieter zur Verfügung:

- AQA (Assessment and Qualifications Alliance)
- Edexcel
- OCR (Oxford Cambridge and RSA Examinations)

Für das Verständnis erschwerend ist die Tatsache, dass verschiedene Exam Boards unterschiedliche Prüfungen anbieten oder auch andere Bezeichnungen verwenden. Die A-Level-Prüfung von Edexcel ist beispielsweise unter dem Namen GCE bekannt, außerdem gibt es von diesem Anbieter noch das IGCSE und verschiedene BTEC-Prüfungen. Ich versuche daher, mich möglichst auf die Gemeinsamkeiten zu konzentrieren, um so allgemeine Aussagen treffen zu können.<sup>93</sup>

Jede Schule wählt einen, bzw. für unterschiedliche Fächer auch mehrere Prüfungsanbieter. Diese sind offiziell gleichwertig und erstellen auch relativ ähnlich aufgebaute Tests, wobei sich meist die Gewichtung der verschiedenen Fertigkeiten unterscheidet. Die Vergleichbarkeit wird dabei regelmäßig von der Institution Ofqual überprüft.<sup>94</sup> Zu beachten ist, dass es sich bei den Exam Boards um private Unternehmen handelt, die untereinander in Konkurrenz stehen. Dementsprechend wichtig sind Marketing und Wettbewerb. Da den Kunden hauptsächlich gute Noten für ihre SchülerInnen wichtig sind, bieten Exam Boards auch entsprechende Prüfungsvorbereitungsmaterialien sowie – teilweise kostenpflichtige – Seminare an, in denen erklärt wird, wie die Prüfung am effizientesten zu bestehen ist. Auch diese Praxis verstärkt das *Teaching to the Test* und ist daher fragwürdig. Insgesamt spielen Exam Boards im Schulwesen jedenfalls eine tragende Rolle, da sie für die Umsetzung der Reform des Ministeriums verantwortlich sind.<sup>95</sup>

---

<sup>92</sup> Vgl. QCDA (2010b).

<sup>93</sup> Vgl. Edexcel (2011).

<sup>94</sup> Vgl. Ofqual (2010b).

<sup>95</sup> Mehr zur Rolle der Exam Boards siehe Lloyd (1999).

#### **2.4.4 Zusammenfassung**

Grundsätzlich haben Fremdsprachen in einem Land, in dem die heutige Lingua franca gesprochen wird, einen eher schweren Stand. Obwohl seit Anfang der 90er-Jahre immer wieder mit verschiedenen Maßnahmen (z.B. ‚languages for all‘-Direktive) versucht wurde den Fremdsprachenunterricht zu fördern, sind die Zahlen der GCSE- und A-Level-AbsolventInnen in Deutsch rückläufig. Das könnte auf zwei Gründe zurückzuführen sein: erstens werden Fremdsprachen erst relativ spät und nur mit geringer Wochenstundenanzahl unterrichtet, was zur Folge hat, dass die Kompetenz der SchülerInnen keinen sehr hohen Level erreicht. Das ist der Motivation nicht besonders dienlich. Zweitens spielen Qualifikationen eine tragende Rolle im Schulsystem. Deutsch gilt als besonders schwieriges Fach, in dem für eine gute Note viel Einsatz erforderlich ist. Dadurch, dass bei den Qualifikationsprüfungen eine freie Fächerwahl besteht und die spätere Schul- und Berufslaufbahn von GCSE-Noten abhängig ist, wählen nur relativ wenige SchülerInnen Deutsch.



## 2.5 Das Deutsch-GCSE bis 2008/2010

### 2.5.1 Kurzdarstellung

Tabelle 1: Prüfungsmodi des alten GCSE<sup>96</sup>

Fertigkeit	Gewichtung	Prüfungsmodi
Hörverstehen	25%	Endprüfung, externe Bewertung
Sprechen	25%	Endprüfung, externe oder interne Bewertung – je nach Verfügbarkeit der PrüferInnen
Leseverstehen	25%	Endprüfung, externe Bewertung
Schreiben	25%	Endprüfung, externe Bewertung  Option zu <i>Coursework</i> : interne Bewertung

Bis 2008/2010<sup>97</sup> bestand das GCSE fast ausschließlich aus einer umfassenden Endprüfung: die einzige Ausnahme stellte die optionale *Coursework* im Bereich Schreiben dar. Für alle Prüfungsteile standen je nach Prüfungsanbieter vier bis fünf Themenbereiche zur Verfügung, die zum Termin der Abschlussprüfung alle beherrscht werden mussten. Die Endprüfung setzte sich aus vier Komponenten zusammen: Hörverstehen (‘Listening and responding’), Leseverstehen (‘Reading and responding’), Schreiben (‘Writing’) und Sprechen (‘Speaking’), wobei jeweils unterschiedliche Niveaustufen (‘tiers’) angeboten wurden. Alle vier Prüfungsteile enthielten zumindest teilweise englischsprachige Instruktionen und Stimuli oder verlangten eine Antwort auf Englisch. Die Prüfungsbögen für das Hör- und das Leseverstehen beinhalteten dabei zahlreiche kurze Aufgaben in unterschiedlichen Aufgabenformaten. Die mündliche Prüfung setzte sich aus zwei Rollenspielen und zwei Konversationen zusammen, die Textproduktion je nach Niveau aus entweder vier kürzeren oder zwei längeren Aufgaben.<sup>98</sup> Die Lese- und Hörverstehensaufgaben waren meines Erachtens stets weniger Kritik ausgesetzt als die produktiven Prüfungsteile und haben sich bei der Überarbeitung kaum geändert. Die meiste Kritik galt den Rollenspielen und der *Coursework*, die im Zuge der Reform

<sup>96</sup> Adaptiert von Edexcel (2000), S. 2.

<sup>97</sup> Letzter Kursbeginn: Herbst 2008, letzte Prüfung: Juni 2010.

<sup>98</sup> Vgl. Edexcel (2000), S. 1-16.

gestrichen wurden – diesen gilt daher besondere Aufmerksamkeit. In den folgenden Abschnitten werde ich das GCSE bzw. dessen kritisierten Teile zunächst hinsichtlich der Hauptgütekriterien, sowie anschließend in Bezug auf die Kriterien für kommunikatives Prüfen und den Backwasheffekt analysieren. Damit soll folgendes geklärt werden: Welche Gütekriterien werden erfüllt – die klassischen Hauptgütekriterien oder die kommunikativ ausgerichteten? Ist das GCSE konstruktvalide, d.h., ist die Prüfung tatsächlich kommunikativ? Werden die Kriterien für positiven Backwash erfüllt?

## 2.5.2 Das GCSE im Kontext der Hauptgütekriterien

Objektivität spielt beim GCSE eine dominierende Rolle. Um die Durchführungsobjektivität zu gewährleisten ist die Bearbeitungszeit der standardisierten Testbögen genau geregelt; außerdem werden mündliche Prüfungen mit Tonband aufgezeichnet und die Vorbereitungszeit wird gestoppt und überwacht. PrüferInnenseminare sollen der Auswertungsobjektivität dienlich sein; ebenso wie Beurteilungsraster („Mark Schemes“), die auflisten, welche Antworten als richtig und welche als falsch gelten. Grundsätzlich wird extern ausgewertet, so dass die SchülerInnen den AuswerterInnen nicht bekannt sind. So soll sichergestellt werden, dass „Ergebnisse unabhängig von der Person des Messenden sind“<sup>99</sup> – also die Prüfung objektiv ist.<sup>100</sup>

### 2.5.2.1 Spezialfall: Coursework

Nur ein Prüfungsteil erfüllt das Kriterium der Durchführungsobjektivität nicht, denn dieser kann – optional – von der zentralen Abschlussprüfung ausgenommen werden: die *Coursework*, welche die schriftliche Prüfung ersetzt. Dabei handelt es sich um mehrere kurze Texte, die im Verlauf der 11. Schulstufe teilweise zuhause und teilweise im Unterricht geschrieben werden. Von diesen sollen schließlich drei Texte eingereicht werden, von denen – je nach Exam board verschieden – etwa ein Drittel unter kontrollierten Bedingungen verfasst worden ist. Wie viel Hilfe erlaubt ist, ist streng geregelt und jede Hilfestellung seitens der Lehrkraft muss dokumentiert werden. Die Schule muss bestätigen, dass die Arbeit tatsächlich ohne fremde menschliche Hilfe geschrieben wurde und auch keine Passagen aus anderen Texten

---

<sup>99</sup> Sacher (2004), S. 35.

<sup>100</sup> Vgl. Ofqual. GCSEs. The official student guide to the system; vgl. auch Edexcel (2000)

abgeschrieben wurden, was nicht immer leicht ist.<sup>101</sup> Internetseiten bieten mittlerweile *Coursework* – sowohl getippt als auch handgeschrieben – gegen Gebühren zum Download an.<sup>102</sup> Ob Eltern oder Freunde einen Beitrag liefern ist ebenfalls oft nicht feststellbar. Es wurden auch immer wieder Fälle bekannt, in denen selbst LehrerInnen beim Schummeln halfen, da der hohe Notendruck nicht nur die SchülerInnen betrifft, sondern durch die League Tables auch die Schule. Das setzte schließlich noch mehr Lehrkräfte unter Druck, die Grenzen voll auszuschöpfen oder gar zu übertreten.<sup>103</sup> Aus Adams (2002) geht hervor, dass *Coursework* hauptsächlich von schwächeren SchülerInnen in Anspruch genommen wird, was ebenfalls darauf hinweist, dass diese leichter ist als die Endprüfung.<sup>104</sup> Mit der Reform des GCSE wurde diese Prüfungsform schließlich abgeschafft und durch ‚*Controlled Assessment*‘ ersetzt. Ursprünglich wurde die *Coursework* nach der Einführung des GCSE auf Wunsch von LehrerInnen eingeführt, da die schriftliche Prüfung zu schwer erschien. *Coursework* stellte damit ein Schlupfloch für schlechtere Schulen dar und wurde angeblich von einigen LehrerInnen ausgenützt.

Um die Reliabilität von bundesweiten Tests zu untersuchen wurde von der Organisation Ofqual das ‚Reliability Programme‘ ins Leben gerufen, wobei bisher hauptsächlich Studien zur Wahrnehmung der Reliabilität von diversen Prüfungen in der Öffentlichkeit durchgeführt wurden.<sup>105</sup>

Objektivität und Reliabilität sind die Grundlage für einen validen Test. Zwei Bereiche der Validität – die Inhaltsvalidität und die Konstruktvalidität – werden unten genauer untersucht. An dieser Stelle soll nur kurz daran erinnert werden, dass Inhaltsvalidität mit dem Kriterium für positiven Backwash ‚repräsentative und unvorhersehbare Auswahl‘ zusammenhängt (vgl. Abschnitte 2.1.1.3 und 2.3.4). Für die Konstruktvalidität ist wichtig, dass die Ergebnisse gültige Indikatoren für ein theoretisches Konstrukt sind. (vgl. Abschnitt 2.1.1.3).

### **2.5.3 Das GCSE als kommunikative Prüfung**

Bevor die Überprüfung der Konstruktvalidität sinnvoll sein kann, muss noch gezeigt werden, dass das GCSE nach Eigendefinition eine kommunikative Prüfung sein soll.

---

<sup>101</sup> Vgl. z.B. Edexcel (2001).

<sup>102</sup> Vgl. Mansell (2010), S. 7.

<sup>103</sup> Vgl. Lloyd (1999), S. 108-109.

<sup>104</sup> Vgl. Adams (2002), S. 3.

<sup>105</sup> Vgl. Ofqual (2010a).

Üblicherweise dienen dazu die Testspezifikationen, doch auch Beurteilungsraster können aussagekräftig sein.

Die Richtlinien der QCA für die Erstellung der Testspezifikationen durch die Exam Boards sind eher vage gehalten:

GCSE specifications in modern foreign languages must enable learners to:

- develop understanding of the language in a variety of contexts
- develop knowledge of the language and language learning skills
- develop the ability to communicate effectively in the language
- develop awareness and understanding of countries and communities where the language is spoken.<sup>106</sup>

Aus dieser Richtlinie geht zwar schon hervor, dass Kommunikation eine Rolle spielt, doch die Beurteilungsraster zeigen dies noch deutlicher:

**Tabelle 2: Beurteilungsraster von Edexcel<sup>107</sup>**

Paper/task	Communication and Content	Knowledge and Application of Language	Mark Allocation	Total for paper
1F – Foundation	50	-	50	50
1H – Higher	50	-	50	50
2F – Foundation				
Role play A	10	-	10	50
Role Play B	10	-	10	
Conversation Topic 1				
Conversation Topic 2	10	20*	30	
2H – Higher				
Role play B	10	-	10	50
Role Play C	10	-	10	
Conversation Topic 1				
Conversation Topic 2	10	20*	30	
3F – Foundation	50	-	50	50
3H – Higher	50	-	50	50
4F – Foundation				
Task 1	10	-	10	50
Task 2	5	5	10	
Task 3	5	5	10	
Task 4	10	10**	20	
4H – Higher				
Task 1	10	10**	20	50
Task 2	15	15**	30	
Coursework	10	10**	20	20 x 3

\* Includes 10 marks for accuracy

\*\* Includes 5 marks for accuracy

<sup>106</sup> QCA (2007), S. 3.

<sup>107</sup> Edexcel (2000), S. 42.

Aus der Tabelle folgt, dass die Bereiche 1 (Hörverstehen) und 3 (Leseverstehen), ebenso wie auch die Rollenspiele aus Bereich 2 (Sprechen) und der erste Task der leichteren Niveaustufe in Bereich 4 ausschließlich hinsichtlich der Kommunikationsleistung der KandidatInnen bewertet werden. Damit liegt eindeutig ein Schwerpunkt auf kommunikativen Fähigkeiten. Das GCSE ist also dann konstruktvalid, wenn ein gutes Testergebnis auch bedeutet, dass der / die SchülerIn erfolgreich in der Fremdsprache kommunizieren kann. Inwiefern das GCSE tatsächlich kommunikative Kompetenz prüft, wird im folgenden Abschnitt erläutert.

## **2.5.4 Das GCSE im Kontext des kommunikativen Prüfens: Authentizität und Interaktivität**

### **2.5.4.1 Rollenspiele**

Die mündliche Prüfung des GCSE bis 2010 bestand unter anderem aus zwei Rollenspielen, deren einziges Ziel es war, die Kommunikationsfähigkeit der SchülerInnen zu prüfen. Das Rollenspiel soll an dieser Stelle im Hinblick auf die zwei wichtigsten Kriterien von kommunikativen Prüfungen analysiert werden: hinsichtlich Authentizität und Interaktivität.

Zunächst einmal muss festgehalten werden, dass es für das Rollenspiel strenge Anweisungen gab und der Ablauf genau geregelt war. So wurde mittels Bildimpulsen, englischen Anweisungen und einem vorgedruckten Text auf der Rollenkarte des Prüfers / der Prüferin genau vorgegeben, was wann gesagt werden sollte. Die PrüferInnen waren sogar explizit dazu angehalten, sich an den vorgedruckten Text zu halten, da die Antworten sonst ungültig wurden:

Rephrasing of a question renders the candidate's response invalid no matter how good it is. For example, rephrasing *Wann möchten Sie bei uns essen?* to *Um wie viel Uhr möchten Sie bei uns essen?* may give a candidate an unfair advantage. Similarly altering *Wo kann man hier gut essen?* in C8 to *Was für Restaurants gibt es hier in der Nähe?* was not permissible [sic!].<sup>108</sup>

Theoretisch bieten Rollenspiele die Möglichkeit zur freien Interaktion, was in dieser Prüfung allerdings nicht der Fall war. Das Zitat zeigt deutlich, wie Objektivität das dominierende Kriterium war, wodurch die Interaktivität nicht mehr gewährleistet war. Die meisten Rollenkarten waren so angelegt, dass alle Antworten zum vorgedruckten Text der Lehrkraft passen. Hier ist ein Beispiel für ein solches Rollenspiel:

---

<sup>108</sup> Edexcel (2009a), S. 11.

## C3

You have visited a German school. Now your German exchange partner is visiting you. The Examiner will play the part of the German student.

**Wochenprogramm für den Englandbesuch**

Montag:	Unterricht
Dienstag:	Ausflug nach London (mit Austauschpartnern)
Mittwoch:	Unterricht
Donnerstag:	Ausflug nach Bath (nur deutsche Gruppe)
Freitag:	Unterricht
Wochenende:	bei der Partnerfamilie

1. **Pläne für die Woche in der Schule**
2. !
3. **Beschreibung deiner Schulkleidung**
4. **Meinungen über die Schule?**
5. !

## C3

*Introduction to be read by the Examiner.*

**Du bist in deiner Schule. Ich bin dein deutscher Austauschpartner/ deine deutsche Austauschpartnerin.**

**Exam:** Was machen wir diese Woche in der Schule?

**Cand:** 1.

**Exam:** Was sind deine Lieblingsfächer und warum?

**Cand:** 2.

**Exam:** Ich mag Mathe. Was trägst du normalerweise in deiner Schule?

**Cand:** 3.

**Exam:** Interessant!

**Cand:** 4.

**Exam:** Ich gehe sehr gern in die Schule. Was hat dir an der Schule in Deutschland gefallen?

**Cand:** 5.

**Exam:** Ich freue mich auf diese Woche.

**Abbildung 3: Rollenspielkarte von Edexcel<sup>109</sup>**

Dieses Rollenspiel erinnert eher an einen Fragenkatalog, dessen Fragen der Reihe nach vorgelesen werden. Es stellt bloß einen Rahmen für das Abwechseln von Hören und Sprechen dar, doch ohne Möglichkeit, den Verlauf des Gesprächs mitzubestimmen, denn der/die PrüferIn kann sich schon aufgrund der Prüfungsbestimmungen gar nicht auf die Aussagen des Schülers oder der Schülerin beziehen. Dementsprechend durften auch keine Umwegstrategien wie etwa das Nachfragen angewendet werden (siehe ‚strategische Kompetenz‘ in Abschnitt 2.2.1). Dies entspricht nicht Kühhaas‘ Definition von Interaktivität (siehe Abschnitt 2.1.3.2).

Wir haben bereits in einem früheren Abschnitt (2.1.3.1) definiert, dass Authentizität gegeben ist „wenn der Kommunikationsanlass echt (nicht simuliert bzw. gestellt) ist, ein reales Kommunikationsziel existiert und das Kommunikationsmedium adäquat ist“<sup>110</sup>. Da es sich hier um einen simulierten und gestellten Kommunikationsanlass ohne reales Kommunikationsziel handelt, ist das das Rollenspiel nicht als authentisch zu werten. Abgesehen davon fanden sich teilweise Verkaufsgespräche, die im realen

<sup>109</sup> Edexcel (2009a), S. 36.

<sup>110</sup> Krause / Sändig (2002), S. 85.

Leben nicht so stattfinden würden (z.B.: ‚Ich möchte Seife, bitte‘ – ‚Ja, welche Farbe?‘ – ‚Grün, bitte‘<sup>111</sup>) und daher auch nicht authentisch sind.

Obwohl das Rollenspiel offensichtlich nur zur Überprüfung der Kommunikationsfähigkeit hätte dienen sollen („All role-plays will be marked for communication only.“<sup>112</sup>), wurden also die zwei wichtigsten Kriterien einer kommunikativen Prüfung nicht erfüllt.

#### **2.5.4.2 Konversationen**

Zusätzlich gab es bei der mündlichen Prüfung noch zwei Konversationen zu je ca. 4 Minuten. Für diese gab es von den jeweiligen Prüfungsanbietern Fragenkataloge mit etwa zehn bis zwanzig Basisfragen pro Themenbereich für das niedrigere Prüfungsniveau und wenigen Zusatzfragen für das höhere Niveau, wobei die Zusatzfragen meist Begründungen verlangten.

Beispiele für solche Basisfragen zum Thema ‚Medien, Unterhaltung und Jugendkultur‘ sind etwa: ‚Welchen Sport treibst du am liebsten?‘ und ‚Was trägst du am Wochenende?‘. Erweiterungsfragen zu diesem Thema sind: ‚Ist die Mode für junge Leute wichtig? Warum/warum nicht?‘ und ‚Findest du, dass junge Leute heutzutage zu viel fernsehen?‘.<sup>113</sup>

In den Konversationen wurde zusätzlich zur Kommunikationsfähigkeit auch die Komplexität der Sprache („Knowledge and Application of Language“) und die Korrektheit („Accuracy“) überprüft. Dabei bekamen SchülerInnen mehr Punkte, wenn sie eine bestimmte Anzahl an unterschiedlichen Zeiten verwenden konnten. Ein Beispiel für ein solches Bewertungsraster des Exam Boards AQA ist hier ersichtlich:

---

<sup>111</sup> Edexcel (2008), S. 19.

<sup>112</sup> Edexcel (2000), S. 10.

<sup>113</sup> Vgl. Edexcel (2008), S.47.

Tabelle 3: Beurteilungsraster für den Konversationsteil der mündlichen Prüfung von AQA<sup>114</sup>

Communication Mark	Range & Complexity Mark	Pronunciation & Accuracy Mark
0	0	0
1	1 or 2 marks	1 or 2
2	2 or 3 marks 3 marks <b>must</b> include <b>two</b> of Past/Present/Future	Max 4
3	2, 3, 4 marks 3 marks <b>must</b> include <b>two</b> of Past/Present/Future  4 marks <b>must</b> include Past <b>and</b> Present <b>and</b> Future <b>And</b> opinion	Max 5
4	4, 5, 6 marks 4 marks or more <b>must</b> include Past <b>and</b> Present <b>and</b> Future <b>and</b> opinion Revert to 3 if 2 time frames only Revert to 1 if 1 time frame only	Max 6

Das Bewertungskriterium ‚Komplexität‘ ist damit offensichtlich sehr genau definiert, was dazu führt, dass SchülerInnen hier sichere Punkte bekommen, wenn sie die entsprechende Anzahl an Zeiten verwenden und eine Meinung äußern. Wie bereits in Abschnitt 2.1.3.4 behandelt, ist es bezüglich des Kriteriums Unvorhersehbarkeit fraglich, ob eine solche Verbindung von – theoretisch – spontaner Konversation und der Forderung, bestimmte grammatikalische Strukturen zu produzieren, sinnvoll ist. MuttersprachlerInnen machen sich jedenfalls in den seltensten Fällen darüber Gedanken, ob sie in ihren Äußerungen zum Thema ‚Familie‘ auch mindestens drei Zeiten verwenden.<sup>115</sup>

Auch die Prüfungsregulierungsbehörde Ofqual (2008) hat hier ein Problem erkannt:

The need for candidates to show use of past, present and future tenses, together with expression of opinions with justification, led to very formulaic and sometimes highly artificial questions.<sup>116</sup>

<sup>114</sup> AQA (2008b), S. 5.

<sup>115</sup> Vgl. Tenberg / Jones 1993: 26-27.

<sup>116</sup> Ofqual (2008), S. 18.



Das Zitat weist mit „very formulaic“ (‚sehr formelhaft‘) und „highly artificial“ (‚stark gekünstelt‘) auf auswendig gelernte Phrasen hin. Ein Ratgeber für SchülerInnen empfiehlt gezielt Auswendiglernen als Lernstrategie: „The more model answers you learn by heart, the better you will do in your exam.“<sup>117</sup> Diese Modellantworten werden schließlich so dargestellt, dass einzelne unterstrichene Wörter selbst ausgetauscht werden können. Kommentiert werden sie beispielsweise mit: „Some good references to the past, well worth learning by heart and a good reference to the present introduced by ‘aber’“<sup>118</sup>. Hier kann also der Zusammenhang zwischen dem Zwang, grammatische Kompetenz zu zeigen, und formelhaften Antworten hergestellt werden. An dieser Stelle sei noch einmal auf das kommunikativ orientierte Gütekriterium Unvorhersehbarkeit erinnert (siehe Abschnitt 2.1.3.4).

Als Prüfungsvorbereitung wurde den SchülerInnen meist schon von Anfang an nahegelegt ‚gute‘ Antworten zu geben, anstatt ihre Meinung zu kommunizieren. Selbst manche LehrerInnen empfahlen ihren SchülerInnen, bei den Prüfungen einfache Wörter wie ‚interessant‘ zu wählen, als zu lange zu überlegen um die Wahrheit zu sagen, denn nur so konnten ihre SchülerInnen bei den zentralen Prüfungen gute Ergebnisse erzielen. Ein Ratgeber für LehrerInnen empfiehlt beispielsweise: „When teaching students to express opinions, select adjectives which are not too difficult to pronounce and which students should have little difficulty spelling correctly“<sup>119</sup>. Das zeigt nicht nur, dass der Unterrichtsstoff oftmals an die Prüfungsinhalte angepasst wird, sondern auch einen Fokus auf *Test-Taking-Skills* und ein kommerzielles Interesse an der Testvorbereitung. Dies sind Merkmale von *Teaching to the Test* und damit Anzeichen von negativem Backwash.

In den vergangenen Jahren wurden oft SprachassistentInnen eingesetzt, um den SchülerInnen entsprechend komplexe Antworten auf diese Fragen bereitzustellen und diese anschließend zu üben. Viele haben dabei die Erfahrung gemacht, dass die SchülerInnen ausschließlich den Fragenkatalog auswendig lernen und zu den gleichen Themen nicht spontan kommunizieren könnten, wenn die Fragen anders formuliert wären. Obermaier (1999) hat sich eingehender mit der Rolle der SprachassistentInnen und den Prüfungsvorbereitungen für das GCSE beschäftigt,

---

<sup>117</sup> Buckby / Pugh (1997), S. 56.

<sup>118</sup> Buckby / Pugh (1997), S. 61.

<sup>119</sup> Buckby / Corney (1999), S. 12.

wobei meine Erfahrung als Sprachassistentin im Schuljahr 2008/09 auch zehn Jahre später noch sehr ähnlich war.

Diese Beispiele zeigen, dass der Konversationsteil der mündlichen Prüfung in der Praxis oft so gehandhabt wurde, dass die Kriterien Authentizität und Interaktivität nicht erfüllt wurden.

### **2.5.4.3 Fazit**

Das alte GCSE weist besonders bei der Überprüfung der produktiven Fertigkeiten einige Mängel auf. Obwohl die Prüfung eigentlich als kommunikativ gilt, werden bei der mündlichen Prüfung die zwei Hauptkriterien von kommunikativen Tests - Authentizität und Interaktivität – aufgrund der hohen Anforderungen an die Objektivität nicht erfüllt. Damit ist diese Prüfung nicht konstruktvalid, was einen negativen Backwash zur Folge hat. Im Spannungsfeld zwischen kommunikativen und klassischen Testgütekriterien fiel die Entscheidung offensichtlich auf letztere. Die *Coursework* gilt dagegen nicht als objektiv genug, da kein ausreichender Schutz vor Plagiaten gegeben ist.

## **2.5.5 Der ‚Backwash‘-Effekt des GCSE bis 2010<sup>120</sup>**

### **2.5.5.1 Der Backwasheffekt nach Schifko (2001)**

#### 2.5.5.1.1 Statusbezogener Aspekt

Für den statusbezogenen Aspekt ist relevant, wie wichtig eine Prüfung ist. Das GCSE nimmt hier einen ganz besonderen Stellenwert in der Bildungspolitik und am Arbeitsmarkt ein. Gute Noten sind einerseits für die weitere Schulbildung nötig, weil mindestens 5 Noten von A\* bis C nötig sind, um von einer Schule für einen A-Level-Kurs angenommen zu werden, aber auch besonders für die zukünftigen ArbeitgeberInnen von Interesse, da für Menschen mit schlechteren Abschlusszeugnissen nicht alle Berufe offen stehen. Damit ist das GCSE auch viel bedeutsamer als es etwa die National Curriculum Tests sind. In der Literatur werden immer wieder die erheblichen Auswirkungen des GCSE auf die Zukunft der KandidatInnen thematisiert. Um als *High-Stakes-Test* zu gelten, ist es allerdings auch notwendig, dass die Prüfung als wichtig wahrgenommen wird. Dazu schreibt z.B.

---

<sup>120</sup> Bereiche, die sich nicht auf die Deutschprüfung beziehen, sondern z.B. auf den Stellenwert in der Gesellschaft oder institutionelle Zuständigkeiten, haben sich mit der Überarbeitung kaum bis gar nicht verändert, was in den entsprechenden Abschnitten auch gekennzeichnet wird.

Denscombe (2000), dass britischen Jugendlichen immer mehr gute GCSE-Noten abverlangt werden, was eine häufige Ursache für Stress ist.<sup>121</sup> Auch Putwain (2009) und Rogers / Hallam (2010) haben in ihren Studien festgestellt, dass sich SchülerInnen vom Druck der Prüfung stark belastet fühlen.<sup>122</sup>

Doch betroffen sind nicht nur die individuellen Schülerinnen und Schüler, deren Karrierechancen und weitere Bildungsmöglichkeiten davon abhängen, sondern auch Schulen und Lehrkräfte: Auf Basis des GCSE werden nämlich die League Tables<sup>123</sup> erstellt: ein Schulranking, das Eltern zur Auswahl der Schule für ihre Kinder zu Rate ziehen. Schlechte GCSE-Resultate können sogar zur Schließung von Schulen führen<sup>124</sup>, was zusätzlichen Druck auf Lehrkräfte ausübt. Für die individuelle Lehrkraft ist es dadurch beinahe unmöglich, dem *Teaching to the Test* zu widerstehen. Der statusbezogene Aspekt hat sich mit der Überarbeitung des GCSE nicht verändert.

#### 2.5.5.1.2 Curricularer Aspekt

Der curriculare Aspekt bezieht sich auf die Auswirkungen von Prüfungen auf den Lehrplan. Diese Auswirkungen fallen in England besonders drastisch aus, da der Lehrplan für die zwei Jahre vor dem GCSE in manchen Fächern ganz abgeschafft wurde. Derzeit gibt es für die Key Stage 4 nur mehr in den Pflichtfächern einen Lehrplan.<sup>125</sup> Das macht in allen anderen Fächern und somit auch Deutsch de facto die GCSE-Kriterien zum Lehrplan, zumindest insofern unter einem Lehrplan die Lehr- und Lerninhalte verstanden werden. Die PrüfungsentwicklerInnen haben damit – obwohl demokratisch nicht legitimiert – einen weitreichenden Einfluss auf den Deutschunterricht. Hodgson / Spours (2008) betonen, dass die Rolle des Lehrplans im Vergleich zu den Qualifikationen eine viel geringere ist.<sup>126</sup>

Doch auch der Lehrplan für Pflichtfächer hängt stark mit der Entwicklung des GCSE zusammen. Die ‚Qualifications and Curriculum Development Agency‘, kurz QCDA<sup>127</sup>, ist sowohl für die nationalen Prüfungen (z.B. A-Levels, GCSE, The Diploma, ...) als

---

<sup>121</sup> Vgl. Denscombe (2000), S. 359-360.

<sup>122</sup> Vgl. Putwain (2009), S. 407-409; Rogers / Hallam (2010), S. 153-163.

<sup>123</sup> Vgl. DCSF (2010).

<sup>124</sup> Vgl. Paton, Graeme [12. 1. 2011]. GCSE league tables: more than 200 schools 'face closure'. <http://www.telegraph.co.uk/education/educationnews/8254817/GCSE-league-tables-more-than-200-schools-face-closure.html> (21. 3. 2011). (The Telegraph, Onlineausgabe).

<sup>125</sup> Vgl. QCDA (2011).

<sup>126</sup> Vgl. Hodgson / Spours (2008), S. 7-8.

<sup>127</sup> Bis 1. April 2010: QCA (Qualifications and Curriculum Agency)

<http://www.dcsf.gov.uk/14-19/index.cfm?go=site.home&sid=3&pid=445&ctype=None&ptype=Contents>

auch für die Entwicklung des Lehrplans zuständig. Obwohl der Lehrplan Schulen nur einen Rahmen bieten soll, in dem sie selbst schulweite Richtlinien für ‚Assessment‘ entwickeln können, wird immer wieder auf nationale Standards Bezug genommen. Einerseits ist dabei die Rede von Individualisierung, Personalisierung und mehr Flexibilität, doch andererseits wird klar auf die neu entwickelten Qualifikationsprüfungen der QCDA verwiesen:

To enable all learners to demonstrate their achievements, schools will be able to offer a range of new and modernised qualifications. These have been developed as an integral part of the curriculum, rather than as schemes of assessment that are separate from teaching and learning.<sup>128</sup>

Damit bezieht sich die Organisation unter anderem auf die Überarbeitung des GCSE. Die Tatsache, dass dieselbe Organisation für Lehrplan und nationale Prüfungen zuständig ist, zeigt, dass durchaus eine Wechselwirkung besteht. Weiters arbeitet die QCDA eng mit der Organisation Ofqual (Office of Qualifications and Examinations Regulation) zusammen, die für die Erhaltung von Standards zuständig ist und das Prüfungssystem kontrolliert und reguliert. Ofqual ist etwa auch für die Veröffentlichung der Prüfungskriterien zuständig, die von QCDA ausgearbeitet werden.<sup>129</sup> Diese Kriterien bieten schließlich den Rahmen, in dem die Exam Boards ihre Tests erstellen können. An dieser Stelle bleibt nur festzuhalten, dass eine starke Verschränkung – sowohl in organisatorischer, als auch in inhaltlicher Hinsicht, zwischen Curriculum und dem Prüfungssystem besteht. Diese Verschränkung gab es – wenn auch institutionell anders verankert – schon früher, doch sie gilt auch für das neue GCSE.<sup>130</sup>

#### 2.5.5.1.3 Individuenbezogene Aspekte

Die Aufsichtsbehörde Ofqual gab eine Studie zur Vorhersehbarkeit der Aufgaben in GCSE Prüfungen in Auftrag, die 2008 unter anderem zu folgendem Ergebnis führte:

Both learners and teachers recognised the highly predictable and formulaic nature of the examinations and understood the reasons behind it. Both groups also appreciated these aspects in terms of making preparation for and performance in the examination easier to manage, while finding the process narrow and rather frustrating.<sup>131</sup>

---

<sup>128</sup> QCDA (2007), S. 8.

<sup>129</sup> Vgl. QCDA (2010).

<sup>130</sup> Vgl. QCDA (2010); Ofqual (2010).

<sup>131</sup> Ofqual (2008), S. 19.

Dieses Zitat beschreibt den Backwasheffekt in Bezug auf didaktisch-methodische, lernstrategische und motivationspsychologische Aspekte. Obwohl sowohl Lehrkräfte als auch SchülerInnen über die Vorhersehbarkeit der Fragen froh waren und sich dementsprechend auf die Prüfung vorbereiteten, war die motivationspsychologische Wirkung eine negative. Das alte GCSE befand sich damit im Spannungsfeld zwischen positiver Vertrautheit mit den Anforderungen einerseits und negativem *Teaching to the Test* andererseits. Ob sich das geändert hat, wird noch zu klären sein.

Die didaktisch-methodischen und lerntechnischen Strategien für die schriftliche Prüfung werden in einem Ratgeber für LehrerInnen anschaulich dargestellt. In diesem Buch soll die Frage geklärt werden, ob es sich um Schummeln handelt, wenn sich SchülerInnen folgendermaßen auf ihre *Coursework* vorbereiten: "The student has learnt a good piece of their own writing off by heart and reproduces it under controlled conditions." Die Antwort der AutorInnen lautet: „This is what many of us have been doing for years to prepare students for writing in the terminal examinations!“<sup>132</sup> Dass für die schriftliche Prüfung auswendig gelernt wird, ist offensichtlich kein Einzelfall, sondern allgemein bekannt. Die Vorbereitung auf die Rollenspiele und Konversationen wurde bereits in Abschnitt 2.5.3 ausführlich dargestellt: die Lernstrategie bestand ebenfalls zu einem großen Teil aus Auswendiglernen.

#### **2.5.5.2 Der Backwasheffekt nach Watanabe (2004)**

Watanabe unterscheidet fünf verschiedene Backwashdimensionen, von denen hier drei von Interesse sind: Spezifität, Intensität und Ausrichtung. Da es sich um eine Untersuchung in der Tradition der Strömung handelt, die nicht intendierten Backwash untersucht, erübrigt sich die Dimension der Intentionalität. Mit der Dimension Länge meint Watanabe nicht die Dauer der Auswirkungen vor der Prüfung, sondern die Frage, ob der Effekt auch nach der Prüfung noch anhält. Dies kann ich leider nicht untersuchen, weil viele SchülerInnen nach dem GCSE die Schule verlassen und somit nichts über ihren weiteren Werdegang bekannt ist.

Aufgrund des starken statusbezogenen Effekts ist auch die Intensität sehr hoch. In Abschnitt 2.3.5 wurde ferner bereits erklärt, dass *High-Stakes-Tests* zumeist einen

---

<sup>132</sup> Adams (2002), S. 20-21.

negativen Backwascheffekt haben. Nun stellt sich die Frage nach der Spezifität: gibt es einzelne Prüfungsteile, deren Backwascheffekt in einer bestimmten Richtung abweicht? Die rezeptiven Prüfungsteile wurden bisher kaum kritisiert. Aus meiner Erfahrung als Sprachassistentin weiß ich, dass diese auch im Unterricht weniger speziell vorbereitet werden. Da sie aus vielerlei verschiedenen, kurzen Aufgaben bestehen, ist auch das Kriterium ‚repräsentative und unvorhersehbare Auswahl‘ erfüllt, was *Teaching to the Test* generell erschwert. Der Effekt ist hier also positiver. Das könnte auf zwei verschiedene Arten ausgedrückt werden: die Intensität ist schwächer (= der Effekt ist weniger stark negativ) oder die Ausrichtung ist eine andere (= die Ausrichtung ist positiv). Da die starke Intensität im Status der Prüfung begründet liegt, ist es unwahrscheinlich, dass hier Abweichungen bestehen. Ich tendiere daher zur zweiten Deutungsart.

### **2.5.5.3 Das GCSE im Kontext der Bedingungen für positiven Backwash**

Die von Ofqual (2008) durchgeführte Studie ermittelte die Vorhersehbarkeit der Prüfungsfragen in ausgewählten Fächern und steht damit eng im Zusammenhang mit dem Kriterium für positiven Backwash ‚Repräsentative und unvorhersehbare Auswahl‘, und somit auch mit dem Gütekriterium ‚Inhaltsvalidität‘.

#### **2.5.5.3.1 Repräsentative und unvorhersehbare Auswahl**

For speaking, there was evidence of a high level of predictability in the speaking assessment in terms of the topic coverage in the various role-plays, as there was in the writing papers. There was variation in the extent of the predictability across the awarding bodies, with the narrative role-play from one seen as very rigid and unchanging.<sup>133</sup>

Die Studie stellt einen hohen Level an Vorhersehbarkeit fest, wobei die Rollenspiele eines bestimmten Exam Boards als sehr starr und gleichbleibend bezeichnet werden. Dieses Kriterium für positiven Backwash wurde bei der alten GCSE Prüfung somit nicht erfüllt. Weiters wird in einer Studie der Organisation Ofsted kritisiert, dass nur wenige SchülerInnen fähig sind eigene Sätze zu formulieren um damit eine vorher nicht einstudierte Situation zu meistern:

Overall, there was insufficient emphasis on helping students to use the language spontaneously for real situations. Consequently, too few students could speak creatively, or beyond the topic they were studying, by making up

---

<sup>133</sup> Ofqual (2008), S. 18.

their own sentences in an unrehearsed situation. Several students said that being able to say what they wanted to say would improve their enjoyment.<sup>134</sup>

Rollenspiele auf mittlerem und höherem Niveau enthielten zwar ein unvorhersehbares Element („unpredictable element“), das durch ein Rufzeichen an der entsprechenden Stelle auf der Rollenkarte markiert war, doch auch dieses war – nach einigem Üben mit Tests aus Vorjahren – eigentlich relativ berechenbar: „Although this question is referred to as the unpredictable element in the role plays, it is, in fact, often quite predictable!“<sup>135</sup>.

#### 2.5.5.3.2 Fertigkeiten testen, die gefördert werden sollen

Diese Bedingung für positiven Backwash bezieht sich auf das häufige Versäumnis, der mündlichen Fertigkeit bei der Prüfung genug Raum zu verschaffen. So wurde etwa häufig betont, dass die mündliche Kommunikation mehr zu fördern sei, während sie gleichzeitig bei Prüfungen kaum von Nutzen war. Dementsprechend kamen mündliche Produktionsaufgaben auch im Unterricht häufig zu kurz.

Dies ist beim GCSE nicht der Fall: alle Fertigkeiten werden gleichermaßen getestet. Hörverstehen, mündliches Kommunizieren, Leseverstehen und schriftliches Kommunizieren nehmen in den fachspezifischen Kriterien für die GCSE-Prüfung („subject criteria“) jeweils 20 - 30 Prozent ein. Wenn es also das Ziel ist, alle Fertigkeiten gleichermaßen zu fördern, ist diese Bedingung erfüllt.<sup>136</sup>

#### 2.5.5.3.3 Kriteriumsbezogenheit

Das GCSE ist offiziell eine kriterienbezogene Prüfung, was unter anderem an den publizierten Prüfungskriterien erkennbar ist. Nach Hughes (2003) sind etwa auch die zur Wahl stehenden Niveaustufen ein Indikator dafür.<sup>137</sup> Es soll also ermittelt werden, welches Lernpensum bewältigt wurde, statt dass soziale Ränge ermittelt werden. Interessant ist dabei allerdings die immerwährende Diskussion um die Erhaltung von Standards: Seit vielen Jahren werden die Noten immer besser, so dass diskutiert wird, ob der Standard tatsächlich steigt, oder ob die Prüfung leichter wird. Ofqual ist unter anderem dafür zuständig, die Kriterien zu erleichtern oder zu verschärfen, so dass die Prüfung in der Öffentlichkeit als gleichbleibend schwer wahrgenommen

---

<sup>134</sup> Ofsted (2008), S. 12.

<sup>135</sup> Buckby / Corney (1999), S. 20.

<sup>136</sup> Vgl. QCA (2007), S. 5.

<sup>137</sup> Vgl. Hughes (2003), S. 53-55.

wird. Demnach ist gilt zwar innerhalb eines Jahrgangs die Kriteriumsbezogenheit, über die Jahre hinweg spielt die soziale Norm allerdings doch eine große Rolle.<sup>138</sup>

Folgendes Zitat illustriert diese Tatsache:

Once all the marking is complete, an 'awarding meeting' takes place. At the meeting the exam board decides on the number of marks needed for each grade (called setting the grade boundaries) and ensures that the standard is the same as in previous years.<sup>139</sup>

Die für eine Note erforderliche Punktzahl wird demnach erst festgelegt, wenn alle Prüfungsbögen ausgewertet wurden. Die Beurteilungsraster der Exam Boards regeln nur, wie viele Punkte für eine Aufgabe vergeben werden, was noch nichts über die Note aussagt. Die Exam Boards liefern zwar auch ‚grade descriptions‘, doch diese sind von der Punktzahl unabhängige Kompetenzbeschreibungen.

In einem Jahr, in dem die SchülerInnen besonders gut sind, werden also die Noten nach unten gesetzt, während sie in einem Jahr, in dem die SchülerInnen vergleichsweise schlecht abschneiden, nach oben gesetzt werden können. Mit der gleichen Punktzahl können also SchülerInnen in einem Jahr eine andere Note bekommen als im darauf folgenden Jahr. Rein kriterienbezogen ist die Prüfung damit nicht.

Andererseits muss auch festgehalten werden, dass die Prüfungskriterien den SchülerInnen durchaus bekannt sind und eine gezielte Vorbereitung auf das GCSE ermöglichen. Auch wenn Noten (und damit auch die Unterscheidung zwischen negativ und positiv) im Vorhinein nicht feststehen, ist es für SchülerInnen möglich, an der erreichten Punktzahl das bewältigte Lernpensum abzulesen. Für einen positiven Backwash könnte das genügen. Die Debatte um die Standards ist allerdings in den Medien derart wichtig, dass ich sie hier nicht außer Acht lassen wollte.

### **2.5.6 Zusammenfassung**

Meiner Analyse lagen drei verschiedene Quellen bzw. Kategorisierungen zugrunde: erstens, Schifkos Unterteilung in statusbezogene, curriculare und individuenbezogene Aspekte; zweitens, Watanabes Backwashdimensionen und drittens, die Kriterien für positiven Backwash nach Hughes (2003) und Schifko (2001).

---

<sup>138</sup> Da die Debatte um die Vergleichbarkeit und Erhaltung von Standards schon Jahre andauert und auch heute noch aktuell ist, ist keine Veränderung durch das neue GCSE feststellbar.

<sup>139</sup> Ofqual. GCSEs. The official student guide to the system. S. 13.



In Bezug auf Schifkos Kategorisierung hat sich gezeigt, dass der statusbezogene Backwasheffekt sehr stark ist und der Lehrplan stark von der Prüfung beeinflusst wird. Beides kann sich durch die Reform nicht geändert haben und wird daher nicht untersucht. Weiters stellte ich fest, dass die Lehr- und Lernstrategien stark in Richtung Auswendiglernen tendieren und auch die Motivation zu wünschen übrig lässt. Dies kann sich geändert haben und soll in meiner Untersuchung erfasst werden.

Von Watanabes Backwashdimensionen sind lediglich Spezifität, Intensität und Ausrichtung relevant, die anderen beiden sind für diese Arbeit hinfällig. Es wurde eine hohe Intensität und eine negative Ausrichtung, zumindest der produktiven Prüfungsteile, festgestellt. In Bezug auf die Spezifität werde ich zwischen einem allgemeinen und einem spezifischen Backwash unterscheiden und insbesondere die kritisierten Teile untersuchen.

Im Hinblick auf die Kriterien für positiven Backwash bot sich ein differenziertes Bild. Von ihnen werden in meiner Fragebogenstudie die meisten erfasst. Diese sind: die repräsentative und unvorhersehbare Auswahl, das Testen von Fertigkeiten, die gefördert werden sollen, die Vertrautheit und Transparenz der Anforderungen, sowie die Unterstützung der Lehrkräfte. Die letzten beiden stehen allerdings nicht im Zentrum des Interesses, so dass diesbezügliche Ergebnisse den Tabellen im Anhang entnommen werden müssen.

In mehreren Bereichen wurden Merkmale negativen Backwashes sichtbar: die mündliche Prüfung des alten GCSE war nicht konstruktvalid. Die starre Struktur der Rollenspiele ermöglichte keine freie Kommunikation und die Konversationsprüfung ermöglichte durch eine Einschränkung und Anpassung des Unterrichtsstoffs an die Testinhalte Auswendiglernen. Das gezielte Üben von Rollenspielen und Konversationsthemen weist auf eine Konzentration auf *Test-Taking-Skills* hin. Auch Hinweise auf ein kommerzielles Interesse an Vorbereitungsmaterialien und –seminaren wurden gezeigt. Der Fokus der empirischen Untersuchung wird hauptsächlich auf den Bereichen Stoffeinschränkung, Auswendiglernen und *Teaching to the Test* liegen.

## 2.6 Das GCSE seit 2010

Tabelle 4: Prüfungsmodi des neuen GCSE<sup>140</sup>

Fertigkeit	Gewichtung „short course“	Gewichtung „full course“	Prüfungsmodi
Hörverstehen	40%	20%	Endprüfung, externe Bewertung
Sprechen	60%	30%	<i>Controlled Assessment</i> : interne Bewertung
Leseverstehen	40%	20%	Endprüfung, externe Bewertung
Schreiben	60%	30%	<i>Controlled Assessment</i> : interne Durchführung und externe Bewertung

### 2.6.1 Controlled Assessment

Die größte Veränderung betrifft die Umstellung von Endprüfungen zu *Controlled Assessment*. In Deutsch und einigen anderen Fächern sind nun 60% *Controlled Assessment* erforderlich, so dass nur mehr 40% in einem Abschlusstest abgeprüft werden dürfen. In Fremdsprachen werden mit diesem neuen Testmodus die produktiven Fertigkeiten erfasst. Folglich sind dadurch jetzt Sprechen und Schreiben etwas stärker gewichtet als die rezeptiven Fertigkeiten. Unterteilungen in unterschiedliche Niveaustufen gibt es nur mehr bei der Endprüfung, aber nicht in diesem neuen Prüfungsformat. Insgesamt müssen mindestens zwei Sprech- und zwei Schreibaufgaben im Rahmen des *Controlled Assessment* bewältigt werden, wobei diese zu jedem Zeitpunkt im Schuljahr durchgeführt werden können. Vor jeder Aufgabe findet eine beliebig lange Vorbereitungsphase statt, in der das Prüfungsthema erarbeitet wird (Phase 1). Phase 2 beginnt, sobald die Prüfungsaufgabe ausgeteilt wird. Ein Beispiel für eine mündliche Prüfung des Anbieters AQA sieht folgendermaßen aus:

---

<sup>140</sup> Tabelleninhalte vgl. Edexcel (2010), S. 4-7.

**Exemplar Task B ij – Leisure**

**Task: Holidays**

You are going to have a conversation with your teacher about holidays.

Your teacher will ask you the following;

- what is your favourite type of holiday and why?
- do you think holidays are important? Why/why not?
- do you think people have too many holidays? Why/why not?
- describe your last holiday.
- what do you like doing on holiday and why?
- !

! Remember, at this point, you will have to respond to something you have not yet prepared.

The dialogue will last between 4 and 6 minutes.

Notes for Teachers

! The unpredictable task could be:

- Where will you go on your next holiday and why?

**Abbildung 4: Prüfungsangaben für Controlled Assessment (Speaking)<sup>141</sup>**

Die Anweisungen für die schriftliche Prüfung sehen fast genauso aus. Wenn schriftlich und mündlich das gleiche Thema gewählt wird, muss sich lediglich mindestens ein Punkt unterscheiden:<sup>142</sup>

**Exemplar Task 3 – Leisure**

**Task: Holidays**

You decide to enter a competition for the chance of winning a two week holiday to Europe. You have to write an account of your holiday experiences and preferences.

You could include:

- a description of a recent holiday you have been on – where you stayed, how you got there, when you went etc;
- what you did whilst you were there;
- your opinion of the holiday;
- what type of holidays you like best and why;
- your plans for your holiday next year;
- why you think you should win the competition.

Remember, in order to score the highest marks you must answer the task fully, developing your response where it is appropriate to do so.

**Abbildung 5: Prüfungsangaben für Controlled Assessment (Writing)<sup>143</sup>**

Ab diesem Zeitpunkt dürfen sich die SchülerInnen mit Wörterbüchern, Notizen, manchmal auch mithilfe des Internets und allerlei anderen Quellen gezielt auf die Prüfung vorbereiten, doch um Plagiate zu verhindern müssen sie dabei unter

<sup>141</sup> AQA (2008a), S. 69.

<sup>142</sup> Vgl. AQA (2010c), S. 9.

<sup>143</sup> AQA (2008a), S. 72.

Aufsicht sein. Phase 2 darf höchstens sechs Stunden innerhalb von zwei Wochen dauern. In dieser Zeit dürfen auch Notizen mit Stichwörtern gemacht werden, die bei der Prüfung verwendet werden dürfen, wobei je nach Prüfungsanbieter eine unterschiedliche Wortanzahl (ca. 40) erlaubt ist und auch sonst unterschiedliche Regelungen bezüglich Zeichnungen und konjugierten Verben gelten. Ein Prüfungsanbieter erwähnt explizit, dass die SchülerInnen in dieser Phase auch eine ausformulierte Antwort für ihre mündliche Prüfung erstellen dürfen: „Students can produce a draft version [...] A draft is a fully scripted version of the student’s response to the task“<sup>144</sup>. Diese darf jedoch nicht in den Prüfungsraum mitgenommen werden. Alleine schon die explizite Erlaubnis legt durchaus die Frage nahe, ob dieser Prüfungsmodus das Auswendiglernen fördert. Auch die lange Vorbereitungszeit spricht jedenfalls dafür: Selbst wenn AQA eine Vorbereitungszeit von nicht mehr als zwei Stunden empfiehlt, ist das für eine nur vier bis sechs Minuten dauernde Prüfung immer noch sehr lange. Edexcel und OCR nennen gar nur die Obergrenze von sechs Stunden. Diese Zeiten gelten übrigens auch für die Vorbereitung der Textproduktion, die je nach Niveau etwa 100 bis 300 Wörter enthalten soll.<sup>145</sup>

## 2.6.2 Weitere Veränderungen

Abgesehen von der Einführung des *Controlled Assessment* bietet die GCSE-Prüfung heute durch die Möglichkeit zur Spezialisierung auf Themen oder Fertigkeiten viel mehr Flexibilität. Hier folgt ein kurzer Überblick über die weiteren Veränderungen.

Es bestand zwar auch schon früher die Möglichkeit, einen kürzeren Kurs (‘short course’) zu wählen, doch der unterschied sich nicht wesentlich von einem vollständigen Kurs (‘full course’). Jetzt muss bei einem Kurzkurs zwischen mündlichem und schriftlichem Deutsch gewählt werden.

Mündlich und schriftlich ist nun die Spezialisierung auf ein einzelnes Thema möglich. Für die Hör- und Leseverstehenstests müssen allerdings trotzdem vier Themen vorbereitet werden.

Insgesamt dürfte die Objektivität bei der mündlichen Prüfung nicht mehr so stark im Vordergrund stehen wie früher, so dass die Prüfung flexibler werden konnte. Die Prüfungsaufgaben dürfen nun von den Lehrkräften abgeändert oder auch selbst

---

<sup>144</sup> AQA (2010b), S.6.

<sup>145</sup> Vgl. AQA (2010a); Edexcel (2009b), OCR (2009).

erstellt werden. Die Rollenspiele wurden abgeschafft und durch offene Interaktionen, Diskussionen mit visuellem Stimulus oder Präsentationen mit anschließender Diskussion ersetzt.<sup>146</sup>

### 2.6.3 Einschätzung und Bewertung der Veränderungen

Die Einschätzung und Bewertung der Veränderungen basiert nicht nur auf theoriegeleiteten eigenen Analysen, sondern auch auf Gesprächen mit mehreren Lehrkräften, die meine Überlegungen größtenteils teilten.

Zu den Abschaffungen:

- Die Abschaffung der *Coursework* als Alternative zur schriftlichen Endprüfung war eine positive Entscheidung, da diese oft auf unfaire Art und Weise ausgenutzt wurde und so manchen Schulen einen Vorteil verschaffte, d.h. nicht (durchführungs-)objektiv war.
- Die Abschaffung der Role Plays war eine positive Entscheidung, da diese weder authentisch noch interaktiv waren, und somit den eigenen Anspruch, eine kommunikative Prüfungsform zu sein, nicht erfüllten, d.h. nicht konstruktvalide waren.

Zur Einführung des *Controlled Assessment*:

- Das wohl bedeutendste Element des *Controlled Assessment* ist die lange Vorbereitungszeit, in der die Prüfungsaufgabe schon bekannt ist. Das reduziert die erforderliche Spontaneität, die Unvorhersehbarkeit und die Authentizität der Prüfungssituation. Gleichzeitig wird das Auswendiglernen begünstigt, da nur so die Vorbereitungszeit voll ausgeschöpft werden kann.
- SchülerInnen können sich bei mündlichen und schriftlichen Prüfungen nun auf ein einzelnes Thema spezialisieren. Der geringere Stoffumfang bedeutet, dass der Test weniger inhaltsvalide ist. Die Repräsentativität und Unvorhersehbarkeit des Inhalts nimmt ab, was Auswendiglernen ebenfalls tendenziell erleichtert. Eine geringere Validität bedeutet, dass *Controlled Assessment* weniger über die tatsächlichen Fähigkeiten der SchülerInnen

---

<sup>146</sup> Vgl. Edexcel (2009b), S. 5-6, 10-13.

aussagt, als die bisherigen Prüfungsformen der mündlichen und schriftlichen Endprüfung.

- Die Einführung von *Controlled Assessment* verringert die Dauer des ‚*Teaching to the Test*‘ von etwa zwei Jahren auf nur mehr vier Mal wenige Wochen. Es kann beispielsweise Themen geben, zu denen SchülerInnen gar nicht geprüft werden. Dadurch kann es in Zukunft mehr Zeit für freies und kreatives Unterrichten geben. Dieser zusätzliche Freiraum für kreatives Unterrichten ist positiv zu bewerten. Er kann positive Auswirkungen auf die Motivation der LehrerInnen und SchülerInnen haben.
- Die Beurteilungskriterien wurden vager. Es ist unklar, welchen Stellenwert die Grammatik nun einnimmt. Theoretisch sollte sie weniger wichtig sein, doch in der Praxis liegt es im Ermessen der AuswerterInnen. In Bezug auf die Transparenz und die Vertrautheit mit den Anforderungen ist das jedenfalls nicht positiv zu werten. Laut Mark Scheme werden in Zukunft authentisch, spontan und sicher klingende KandidatInnen besser bewertet. Ob diese KandidatInnen allerdings tatsächlich spontaner interagieren können, ist aufgrund der langen Prüfungsvorbereitungszeit unklar.

Insgesamt könnte man also zusammenfassen, dass sich die Prüfung aufgrund ihrer geringen Aussagekraft zwar negativ entwickelt hat, aber trotzdem positive Auswirkungen auf den Unterricht haben könnte, da die Prüfung nicht mehr so stark im Vordergrund steht.

### 3 Empirische Untersuchung

#### 3.1 Methodische Überlegungen

Am Beginn jeder empirischen Untersuchung stellt sich die Frage nach der Forschungsmethode. Häufig wird die Entscheidung für qualitative oder quantitative Methoden abhängig vom aktuellen Forschungsstand, der im jeweiligen Forschungsbereich üblichen Methodik und dem Ziel der Untersuchung getroffen. Erstere wirken dabei meist hypothesengenerierend, während letztere herangezogen werden, um bestehende Hypothesen zu prüfen. Im vorliegenden Fall ist aufgrund der hohen Aktualität noch keine wissenschaftliche Literatur verfügbar, doch mittels einer Analyse des neuen GCSE, basierend auf einer Backwashtheorie und anhand von Gesprächen mit Lehrkräften ist es möglich, im Vorfeld der Untersuchung erste Hypothesen zu entwickeln. Es wird also weder „wissenschaftliches Neuland betreten“<sup>147</sup>, noch ist die „Überprüfung einer gut begründeten Hypothese zulässig“<sup>148</sup>. Grundsätzlich wären also beide Zugänge möglich.<sup>149</sup>

In „Working for washback: a review of the washback concept in language testing“ beschreibt Bailey (1996) die für diesen Forschungsbereich übliche Methodik:

In a broad sense, most research involves observing or asking [...] and current research on language testing washback is no exception. The most complete designs for researching washback include both observing classes and asking the participants about their views and experiences (either through interviews or written questionnaires).<sup>150</sup>

Aus dem Zitat geht hervor, dass sich Washback-, bzw. Backwashstudien hauptsächlich qualitativer Forschungsmethodologie bedienen, wobei Triangulation (siehe Abschnitt 3.1.1) anzustreben ist. Ob dies in der vorliegenden Studie möglich ist, hängt allerdings vom Zugang zum Feld ab.

Eine weitere wichtige Entscheidungsgrundlage sind daher der Zugang zum Feld sowie die Art der Stichprobe und die Stichprobengröße.<sup>151</sup> Ich habe leider nur eingeschränkte Möglichkeiten, Deutschunterricht in England zu observieren. Die

---

<sup>147</sup> Reicher (2005), S. 91.

<sup>148</sup> Reicher (2005), S. 91.

<sup>149</sup> Vgl. Grotjahn (2006); Reicher (2005).

<sup>150</sup> Bailey (1996), S. 273.

<sup>151</sup> Vgl. Reicher (2005), S. 97-102.

Schule, zu der ich den besten Kontakt habe, möchte ich bewusst nicht für diese Studie heranziehen, da persönliche Verbundenheit meine Wahrnehmung und somit auch das Ergebnis verzerren könnte. Es erscheint daher sinnvoll, eher in die Breite als in die Tiefe zu gehen. Ich habe mich für das Instrument Fragebogen entschieden, da ich als ehemalige Sprachassistentin die Möglichkeit habe, über KulturKontakt Austria die derzeitigen SprachassistentInnen zu kontaktieren, um diese zu ermutigen, den Link zu meinem Onlinefragebogen an ihren jeweiligen Schulen zu verbreiten. Ein besseres Netzwerk von Deutschlehrkräften existiert nach Angaben einer ehemaligen Kollegin nicht. Im Schuljahr 2010/2011 sind 68 SprachassistentInnen in England tätig. Davon ausgehend, dass jedeR AssistentIn zwei bis drei DeutschlehrerInnen mit GCSE-Erfahrung erreichen kann, nehmen maximal 136 bis 204 LehrerInnen an meiner Umfrage teil. Für eine rein quantitative Untersuchung ist diese Zahl zu gering, da eine niedrige Rücklaufquote alleine schon durch meinen nur indirekten Zugang zum Feld entstehen kann. Andererseits ist die Stichprobe klar begrenzt und relativ repräsentativ für die Grundgesamtheit aller DeutschlehrerInnen in England. Die Analyse der neuen GCSE-Prüfung und eine Backwashtheorie lieferten bereits erste Hypothesen. Im Oktober 2010 besuchte ich meine ehemalige Schule, um drei Lehrkräfte sowie den Vizerektor zu ihren Ansichten zur Reform des GCSEs zu befragen. Meine eigenen Hypothesen wurden dabei größtenteils bestätigt und durch neue Impulse erweitert. Das Ergebnis des Gesprächs schlägt sich im Kapitel ‚Einschätzung und Bewertung der Veränderungen‘ nieder. Obwohl es sich um die wissenschaftliche Erstbearbeitung dieses Themas handelt, können daher bereits konkrete Fragen gestellt werden, für die Multiple-Choice-Antworten oftmals ausreichen. Qualitative und quantitative Methoden schließen sich heute nicht mehr aus, sondern können sinnvoll kombiniert werden, um die Schwächen des jeweils anderen Zugangs teilweise zu kompensieren. Ich habe mich daher für einen qualitativ-quantitativen Fragebogen entschieden, der den LehrerInnen die Möglichkeit bietet, einzelne Phänomene genauer zu beschreiben (exploratives Design), aber gleichzeitig quantitativ auswertbar ist, um bestehende Hypothesen zu überprüfen (explanatives Design). Auch wenn zusätzliche Unterrichtsbeobachtungen erstrebenswert wären, sind sie nicht zwingend notwendig, da ich auf die Wahrnehmung des Backwasheffekts fokussiere:



Some research on participants' *perceptions* of washback, whether negative or positive, has not included the observational component for collecting baseline or follow-up data (after the introduction of a new test).<sup>152</sup>

### 3.1.1 Triangulation

Unter Triangulation wird ein „mehrperspektivisches Vorgehen“<sup>153</sup> verstanden, bei dem unterschiedliche Methoden, Daten oder Theorien sinnvoll kombiniert werden, um ein ausgewogeneres Bild zu erhalten.<sup>154</sup> Mein qualitativ-quantitativer Fragebogen soll einer solchen Methodentriangulation entsprechen. Auch meine Einschätzung im Vorfeld der Untersuchung basierte auf dem Konzept der Triangulation. Dabei wurden Daten aus unterschiedlichen Quellen (Bildungsministerium, Exam Boards, Backwashtheorie, Aussagen von Lehrkräften) miteinander kombiniert. Bei der Analyse des alten GCSE kamen ferner noch wissenschaftliche Untersuchungen von unabhängigen ForscherInnen einerseits und von Organisationen wie Ofsted und Ofqual andererseits sowie diverse GCSE-Guides für SchülerInnen und LehrerInnen hinzu.

### 3.1.2 Durchführung der Untersuchung

Ich hatte weder direkten Kontakt zu einem Netzwerk von DeutschlehrerInnen in England, noch zu den derzeit tätigen SprachassistentInnen. Die gesamte Kommunikation lief über die Koordinatorin von KulturKontakt Austria, die meine E-Mails an SprachassistentInnen weiterleitete, die ihrerseits ihre KollegInnen an den Schulen über meine Studie informieren sollten. Aufgrund meines nur indirekten Zugangs zum Feld war es besonders wichtig, die Rücklaufquote so weit wie möglich zu erhöhen. Daher hatte ich für die SprachassistentInnen Unterrichtsmaterialien ins Internet gestellt, die als kleines Dankeschön dazu ermuntern sollten, den Link zur Online-Umfrage tatsächlich weiterzuleiten.

Anfang Dezember fand ein Pretest statt, bei dem einzelne mir bekannte Lehrerinnen kritische Anmerkungen zum Fragebogen machen konnten und die Dauer der Teilnahme ermitteln sollten. Da keine Verbesserungen notwendig waren, konnte ich diese Daten später auch für die Auswertung benutzen. Die erste offizielle Aussendung erfolgte Mitte Dezember 2010, doch da bis Anfang Jänner nur sechs

---

<sup>152</sup> Bailey (1996), S. 274, Hervorhebung im Original.

<sup>153</sup> Reicher (2005), S. 91.

<sup>154</sup> Grotjahn (2006), S. 259-260; Reicher (2005), S. 92.

Fragebögen eintrafen, ließ ich die Einladung im Jänner 2011 erneut ausschicken. Als nach wenigen Tagen noch immer keine weiteren Fragebögen vorlagen, entschied ich mich dazu, private Kontakte zu ehemaligen KollegInnen an mehreren Schulen zu nutzen und meinen Aufruf in ein weithin bekanntes LehrerInnenforum<sup>155</sup> zu stellen. Darunter könnte zwar insofern die Repräsentativität leiden, als ich die Stichprobe nicht mehr kontrollieren kann, doch immerhin trafen auf diese Weise innerhalb von einer Woche weitere fünfundzwanzig Fragebögen ein. Insgesamt haben bis Anfang Februar fünfunddreißig Lehrkräfte an meiner Umfrage teilgenommen. Davon haben elf über Kontakte zur SprachassistentInnen, vier über Kontakte zu anderen Lehrkräften und neunzehn über Foren bzw. Websites im Internet von der Umfrage erfahren, wobei letztere fast ausschließlich auf <http://community.tes.co.uk/forums/28.aspx> davon gelesen haben. Eine Person gibt an, von einem ehemaligen Kollegen bzw. einer ehemaligen Kollegin davon erfahren zu haben. Zehn weitere Personen haben die Umfrage nach einigen Fragen abgebrochen. Da diese Stichprobengröße für viele Untersuchungen (z.B. psychologische Tests) zu gering wäre, muss ich anmerken, dass es sich durch die Art der Stichprobe und das teils offene Design quasi um schriftliche ExpertInneninterviews handelt und die Daten somit aussagekräftig sind. Trotzdem werde ich mich angesichts der wenigen Rückmeldungen besonders auf die qualitative Datenanalyse konzentrieren. Die quantitativen Daten sollen lediglich dazu dienen, Trends widerzuspiegeln und die qualitativen Daten zu ergänzen bzw. zu veranschaulichen und werden nur als aussagekräftig behandelt, wenn die Ergebnisse stark in eine Richtung tendieren.

### 3.1.3 Der Fragebogen

Der Fragebogen besteht aus quantitativ auswertbaren Multiple-Choice-Fragen einerseits und qualitativ auswertbaren offenen Fragen andererseits. So wie auch die Analyse der GCSE-Prüfung bis 2010 stützt sich diese Untersuchung hauptsächlich auf drei Quellen: Watanabes (2004) Backwashdimensionen Intensität, Ausrichtung und Spezifität, Schifkos (2001) individuenbezogene Aspekte und die Kriterien für positiven Backwash nach Hughes (2003) und Schifko (2001).

---

<sup>155</sup> TES Connect. Forums / Subjects / Modern Foreign Languages.  
<http://community.tes.co.uk/forums/28.aspx> (21. 3. 2011)

Weiters soll erhoben werden, wie groß der Einfluss der League Tables ist. Diese können zu einem besonders intensiven (allgemeinen) Backwash führen. Obwohl sich die Prüfung unter Umständen in manchen Bereichen verbessert hat, kann der Druck zum *Teaching to the Test* so weiter bestehen und eventuell negative Auswirkungen haben.

### 3.1.4 Datenauswertung

Die Datenauswertung erfolgt mittels qualitativer Inhaltsanalyse. Im Zentrum dieser Auswertungsmethode steht die Entwicklung eines Kategoriensystems. Diese Kategorienbildung kann entweder induktiv (Bottom-up-Strategie) oder deduktiv (Top-down-Strategie) erfolgen. Der induktive Zugang soll größtmögliche Offenheit bieten und ist daher besonders für explorative Studiendesigns geeignet (vgl. auch *Grounded Theory*). Die Kategorien werden dabei direkt aus dem Forschungsmaterial heraus gewonnen. Bei den systematischeren deduktiven Zugängen wird zwischen einer Kategoriengewinnung aus dem Forschungsinteresse heraus (z.B. aus Hypothesen) einerseits und aus abgesicherten Theorien andererseits unterschieden.<sup>156</sup>

In Bezugnahme auf Flick (1991) weist Reinhoffer (2005) darauf hin, dass sich die beiden Zugänge nicht ausschließen sondern „in einem Spannungsverhältnis zwischen Theorie und Empirie“<sup>157</sup> befinden und dadurch in der Praxis häufig, wenn auch mit unterschiedlich stark ausgeprägten Schwerpunktsetzungen, kombiniert werden.

Auch in dieser Arbeit sollen sich die beiden Zugänge ergänzen, wobei Regelgeleitetheit und Systematik eine zentrale Rolle einnehmen müssen. Aus diesem Grund werde ich zunächst zwei getrennte Kategoriensysteme erstellen, um anschließend Überschneidungen und Kombinationsmöglichkeiten aufzuzeigen.

Die ersten fünf Fragen waren für die quantitative Auswertung gedacht und werden nicht ausgewertet. Damit hätten Korrelationen zwischen Erfahrungen mit anderen Prüfungssystemen (vor der Einführung des GCSE) und der Einstellung gegenüber *Controlled Assessment* festgestellt werden sollen. Weiters hätte ich ausgewertet, ob

---

<sup>156</sup> Vgl. Reinhoffer (2005), S. 124-128, Mayring (2003), S. 74-76.

<sup>157</sup> Reinhoffer (2005), S. 128.

die Daten zum Schulstandort mit der Einstellung zur *Coursework* und mit der Wahrnehmung des Drucks der League Tables in Verbindung stehen.

### **3.1.5 Kategorienbildung**

Für die deduktive Kategorienbildung habe ich die Einschätzung und Bewertung der Veränderungen aus Abschnitt 2.6.3 und die in Abschnitt 3.1.6 genannten Hypothesen auf Schlagworte reduziert um daraus vorläufige Kategorien abzuleiten.

Für die induktive Kategoriengewinnung habe ich mich an der Technik der zusammenfassenden Inhaltsanalyse nach Mayring (2003) orientiert. Dafür werden die verbalen Daten aus der Untersuchung zunächst paraphrasiert und schließlich in mehreren Schritten reduziert bis das gewünschte Abstraktionsniveau erreicht ist.<sup>158</sup>

Anschließend habe ich die Kategoriensysteme auf Basis der Überschneidungen und der besonders häufigen Kategorien zusammengeführt. Im Anhang befindet sich eine kurze Übersicht, welche Kategorien im Zuge der Kombination und Reduktion zusammengeführt wurden und welche davon aus der induktiven bzw. deduktiven Kategoriengewinnung entnommen sind. Die Restkategorie enthält all jene Bereiche, die nur unregelmäßig thematisiert werden und für die Analyse nicht zentral sind.

---

<sup>158</sup> Vgl. Mayring (2003), S. 59-77.

### 3.1.6 Hypothesen

Die Hypothesen leiten sich aus der Darstellung des Fremdsprachenunterrichts (vgl. 2.4), aus meiner Analyse der alten GCSE-Prüfung (vgl. 2.5.4 und 2.5.5) und aus der Bewertung und Einschätzung der Reform (vgl. 2.6.3) ab.

1. Es gibt einen allgemeinen Backwasheffekt, der durch den Notendruck der League Tables entsteht bzw. verstärkt wird.
2. Der Backwasheffekt wird mit der Reform des GCSE stärker und/oder negativer.<sup>159</sup>
3. Die Lernstrategie ‚auswendig lernen‘ wird bei den *Controlled Assessment* Tasks dominieren.
4. *Controlled Assessment* reduziert die Validität.
5. Während der Phasen des *Controlled Assessment* kommt es zu einem verstärkten *Teaching to the Test*
6. Es kann in Zukunft mehr Zeit für freies und kreatives Unterrichten geben. Dies ist auf folgendes zurückzuführen:
  - a. einerseits darauf, dass nun nicht mehr alle Themen, die unterrichtet werden, geprüft werden und daher in diesen Bereichen mehr Freiraum entsteht;
  - b. andererseits auf die Möglichkeit, Prüfungsaufgaben selbst zu erstellen und so der jeweiligen LernerInnengruppe anzupassen. Individuelle Schwerpunktsetzungen müssten dadurch möglich werden.

---

<sup>159</sup> Klar abgrenzen lassen sich Ausrichtung und Intensität hier nicht, da eine gleichbleibende Intensität bei stärkerer negativer Ausrichtung in der subjektiven Wahrnehmung der Lehrkräfte zum gleichen Ergebnis führt wie ein intensiverer, aber von der Ausrichtung her unverändert negativer Backwasheffekt.

## 3.2 Ergebnisse

Die Gliederung der Ergebnisse erfolgt analog zur Darstellung der Hypothesen. Ich wähle bewusst nicht dieselbe Gliederung wie im Kapitel zur alten GCSE-Prüfung oder eine Gliederung nach Kategorien, da es zu Überschneidungen kommen würde, die Unübersichtlichkeit zur Folge hätten. In jedem Abschnitt soll zunächst ein Überblick geboten werden, der auf den Antworten zu den quantitativ ausgerichteten Fragen basiert. Da es sich um eine relativ kleine Studie mit nur 35 TeilnehmerInnen<sup>160</sup> handelt, sind die quantitativen Daten nicht generalisierbar, sondern bloß dazu geeignet, die eher komplex zu veranschaulichen verbalen Daten systematisch darzustellen und so zu untermauern. Anschließend sollen Antworten aus den offenen Fragen zu einem tieferen Verständnis führen.

### 3.2.1 Der allgemeine Backwasheffekt

Hypothese 1:

Es gibt einen allgemeinen Backwasheffekt, der durch den Notendruck der League Tables entsteht bzw. verstärkt wird.

#### 3.2.1.1 Quantitativ ausgerichtete Fragen

Washback on a general dimension [...] addresses the question „would teaching/learning become different if there were no exams“.<sup>161</sup>

Watanabe (2004) empfiehlt in „Methodology in washback studies“ zunächst einmal die Frage zu stellen, ob Backwash überhaupt existiert. Damit knüpft er an den grundlegenden Artikel „Does Washback exist?“ von Alderson und Wall an, die damit 1993 erstmals diesen Effekt ins Zentrum des Interesses rückten.<sup>162</sup> Ich habe mich dieser Empfehlung angeschlossen und die folgende Frage gestellt: „Would teaching be different if there were no nation-wide exams at all?“ Die Antworten dazu sind in Tabelle 5 ersichtlich.

---

<sup>160</sup> Ich habe in der Studie nicht nach dem Geschlecht der TeilnehmerInnen gefragt, da es für die Analyse unerheblich ist. Dies hat Konsequenzen für die sprachliche Referenz auf die einzelnen Teilnehmerinnen und Teilnehmer.

<sup>161</sup> Watanabe (2004), S. 28.

<sup>162</sup> Vgl. Watanabe (2004), S. 24 und Alderson / Wall (1993)

**Tabelle 5: Antworten auf die Frage: "Would teaching be different if there were no nation-wide exams at all?"**

Antwortmöglichkeit	Anzahl
1 = no	1
2 = yes, a bit different	5
3 = yes, quite different	11
4 = yes, completely different	18
nicht beantwortet	0

Die Antworten sind eindeutig: etwa die Hälfte der Befragten gibt an, dass das Unterrichten vollkommen anders wäre, wenn es keine landesweiten Tests gäbe. Fast ein weiteres Drittel meint, dass das Unterrichten immerhin noch ‚ziemlich‘<sup>163</sup> anders wäre. Insgesamt nennen damit knapp über 80% der TeilnehmerInnen einen mittelstarken bis starken Unterschied, der durch das Wegfallen von nationalen Tests entstehen würde. Die Frage, ob ein (allgemeiner) Backwascheffekt existiert, kann also durchaus bejaht werden.

Testexterne Faktoren wie der Einfluss von League Tables als Verstärkung des statusbezogenen Effekts können ebenfalls zu einem allgemeinen Backwascheffekt beitragen. Als nächstes soll daher geklärt werden, ob Druck durch die League Tables wahrgenommen wird, und wenn ja, wie dieser den Unterricht beeinflusst.

Die erste Frage zu diesem Themenkomplex lautete: „Do you feel pressure to perform well in league tables?“<sup>164</sup>. Bei nur einer Enthaltung wählten 32 Lehrkräfte die Antwort ‚ja‘ und nur zwei die Antwort ‚nein‘. Fast alle Befragten nehmen somit den Druck der League Tables wahr.

Die nächste Frage lautete: „How strongly does it affect teaching?“. Bei wieder einer Enthaltung antwortet nur jeweils eine Person, dass dieser Druck keinen bzw. einen schwachen Einfluss auf den Unterricht hat. Zehn TeilnehmerInnen sprechen von einem starken und achtzehn gar von einem sehr starken Einfluss.

Auf die Frage „Is the impact rather positive or negative?“ antwortet niemand, dass der Effekt sehr positiv sei. Fünf Personen geben an, dass der Effekt eher positiv sei,

---

<sup>163</sup> Bei Übersetzungen von Antwortalternativen können Probleme mit der Intervallskalierung auftreten, was jedoch für die Ermittlung von generellen Trends nur wenig Ausschlag geben sollte. Engl. ‚quite‘ hat zwar mehrere Bedeutungen, soll aber hier in Abgrenzung zur Angabe ‚completely‘ als ‚ziemlich‘ verstanden und übersetzt werden.

<sup>164</sup> Alle Tabellen sind im Anhang ersichtlich.

im Gegensatz zu dreiundzwanzig Befragten, die einen eher negativen, und weiteren sechs, die einen sehr negativen Effekt feststellen.

Jeweils siebzehn Lehrkräfte sind der Ansicht, dass sich der Druck der League Tables durch die Reform verstärkt habe, bzw. gleich geblieben sei. Niemand vertritt die Meinung, der Effekt sei schwächer geworden.

Analog zur Frage, ob der Unterricht anders wäre, wenn es keine landesweiten Tests gäbe, habe ich als nächstes die Frage gestellt, ob das Unterrichten ohne League Tables anders wäre. Aus den Tabellen 6 und 7 geht hervor, dass etwa 65% der TeilnehmerInnen einen zumindest mittelstarken Unterschied ausmachen würden und dass etwa 60% diesen Unterschied als positiv bezeichnen, während die verbleibenden 40% weder von einem positiven noch von einem negativen Unterschied sprechen.

**Tabelle 6: Antworten auf die Frage: "Would teaching be different if there were no league tables?"**

Antwortmöglichkeit	Anzahl
1 = no	4
2 = yes, a bit different	8
3 = yes, rather different	13
4 = yes, completely different	10
Nicht beantwortet	0

**Tabelle 7: Antworten auf die Frage: "Would this difference be rather positive or negative?"**

Antwortmöglichkeit	Anzahl
1 = rather negative	0
2 = just different: neither positive nor negative	13
3 = rather positive	21
Nicht beantwortet	1

### 3.2.1.2 Verbale Daten und Interpretation

Das Thema 'League Tables' wird in den offenen Antworten vier Mal konkret angesprochen. Interessanterweise wird in dreien dieser Beiträge geäußert, dass der Druck der League Tables das Schummeln fördert. *Coursework* und *Controlled Assessment* sind dafür anfälliger als die Abschlusstests, weil erstere direkt von den verantwortlichen LehrerInnen erstellt und ausgewertet werden:

There is also (because of results and league tables) huge pressure on staff to inflate, even if ever so slightly, the marks for the speaking assessment which is not submitted to the board - what is to stop me from adding a few marks onto



every candidate? This non-submitted speaking assessment does not even have to be recorded... (F9T18)

Teachers feel pressure and I believe many may inflate grades due to pressure of school, parents, students, dept to gain high grades/perform well in league tables. (F9T5)

My school actually did coursework but I feel that the final written exam gives a better picture of the student overall. Coursework can leave teachers open to pressure from senior leaders to "assist" with the work because of league tables. (F1T24)

Auch die vierte Lehrkraft thematisiert League Tables in einem negativen Zusammenhang:

League tables just cannot help. They only inform parents how well a student is prepared for an exam and not a reflection on the quality of teaching and learning. Remove this pressure and you will see schools experimenting and taking risks and doing exciting things that will increase uptake of university courses and a higher quality of student. (F9T13)

Ähnlich, wenn auch nicht konkret auf Schulrankings bezogen, schreibt auch diese Lehrkraft über den Druck, der von Prüfungen ausgeht:

I feel that we are no [sic<sup>165</sup>] dominated by examinations at all levels, and this leads to massive dishonesty, at all levels. (F9T20)

Von der gleichen Person stammt auch der folgende Kommentar. Es handelt sich dabei um die Begründung, warum er/sie beim alten GCSE *Coursework* präferiert hat:

We are driven by results. Coursework is under the teacher's control; the final written exam is not. (F1T20)

Es stellt sich die Frage, warum eine Prüfung, die ‚unter der Kontrolle‘ der Lehrkraft stattfindet, zu besseren Noten führen soll als der Abschlusstest. In Verbindung mit F9T20 und auch im Kontext der Antworten von anderen TeilnehmerInnen lässt sich daraus ein versteckter Hinweis auf mögliches Schummeln herauslesen. Im Rahmen der Antwort auf die Frage, ob die TeilnehmerInnen bei der alten GCSE-Prüfung die *Coursework* oder den Abschlusstest präferiert haben, wurde das Thema ‚Schummeln‘ mehrmals angesprochen, und zwar durchwegs von TeilnehmerInnen, die selbst den Abschlusstest vorgezogen haben. Die BefürworterInnen der *Coursework* nannten zwar mehrheitlich bessere Noten als Grund für ihre Präferenz, allerdings ohne Schummeln auch nur ein einziges Mal zu erwähnen. Im Folgenden

---

<sup>165</sup> Einige Zitate weisen im Original Flüchtigkeitsfehler auf, die ich hier unverändert wiedergebe. Aus Gründen der besseren Lesbarkeit werde ich nicht jedes Mal mit [sic!] darauf hinweisen.

beschreiben TeilnehmerInnen, welchem Druck sie ausgesetzt gewesen wären, wenn sie bei der alten GCSE-Prüfung *Coursework* gewählt hätten:

Did not want the ethical dilemma of 'how much help to give' which I saw in other subjects (F1T33)

Harder for teachers to 'cheat'. The exam forced you to teach the language, not just give them loads of help sheets to copy from. They really had to understand the language. Because loads of schools gave students too much help with coursework, the standard of work increased so much, that it was really hard for students to do well if they were not given similar help sheets. The exam actually demanded less complex language from them and less accuracy. (F1T30)

*Coursework* wurde für die Verleitung zum Schummeln kritisiert und daher von vielen Lehrkräften abgelehnt. Das Problem waren jedoch nicht, zumindest nicht ausschließlich, die schummelnden SchülerInnen, sondern auch die Verlockung für das Lehrpersonal. Wie bereits weiter oben gezeigt fühlen sich LehrerInnen durch die League Tables stark unter Druck gesetzt, ihren SchülerInnen zu guten Noten zu verhelfen. Um diesem Zwiespalt zu entfliehen wählten einige Lehrkräfte den schriftlichen Abschlusstest anstelle von *Coursework*. *Controlled Assessment* soll zwar bezüglich der Möglichkeit zu schummeln theoretisch eine Verbesserung von *Coursework* darstellen, doch das betrifft lediglich die Genauigkeit der Richtlinien und die Überwachung von SchülerInnen. Solange es den Druck der League Tables gibt, wird auch die Verlockung für Lehrkräfte und damit ein persönlicher Zwiespalt bestehen bleiben. Da mit der Reform die Abschlussprüfung abgeschafft wurde, gibt es nun keine Möglichkeit mehr, diesem Zwiespalt zu entkommen.

### 3.2.1.3 Fazit

Die quantitativen Daten deuten auf einen starken allgemeinen Backwascheffekt hin, der durch den Notendruck der League Tables entweder entsteht oder verstärkt wird. Aus den verbalen Daten zum Thema League Tables geht hervor, dass dieser Druck Lehrkräfte zum Schummeln anregt, was angesichts der Tatsache, dass LehrerInnen durch solche verbotenen Aktionen ihre Karriere aufs Spiel setzen, von einem relativ starken Druck zeugt.

### 3.2.2 Intensität und Ausrichtung

Hypothese 2:

Der Backwasheffekt wird mit der Reform des GCSE stärker und/oder negativer. Dies zeigt sich durch ein verstärktes *Teaching to the Test* in den *Controlled Assessment* Phasen.

#### 3.2.2.1 Quantitativ ausgerichtete Fragen

**Tabelle 8: Antworten auf die Frage: "How strongly did the old GCSE affect your teaching?"**

Antwortmöglichkeit	Anzahl
1 = no impact	4
2 = weak impact	1
3 = mid impact	19
4 = strong impact	10
nicht beantwortet	1

**Tabelle 9: Antworten auf die Frage: "How strongly does the new GCSE affect your teaching?"**

Antwortmöglichkeit	Anzahl
1 = no impact	0
2 = weak impact	4
3 = mid impact	12
4 = strong impact	18
nicht beantwortet	1

Etwa die Hälfte der Befragten stellt seit der Reform starke Auswirkungen des GCSE auf den Unterricht fest. Immerhin ein Drittel gibt eine mittelstarke Intensität an und lediglich vier von fünfunddreißig Personen sprechen von einem geringen Einfluss. Die Angaben für die das alte GCSE sehen noch etwas anders aus: Nur etwas weniger als 30% der Befragten bezeichnen die Auswirkungen der früheren Prüfung als stark, aber knapp mehr als die Hälfte nahm die Intensität bereits als mittelstark wahr. Vier Lehrkräfte geben an, dass das alte GCSE keinerlei Auswirkungen auf den Unterricht gehabt hätte, was von der neuen Prüfung niemand mehr behauptet. Tendenziell hat sich die Wahrnehmung des Backwasheffekts in dieser Stichprobe damit von einer mittleren hin zu einer hohen Intensität verschoben.

**Tabelle 10: Antworten auf die Frage: "Was the impact of the old GCSE on your teaching rather positive or negative?"**

Antwortmöglichkeit	Anzahl
1 = very negative	0
2 = rather negative	11
3 = rather positive	19
4 = very positive	2
nicht beantwortet	3

**Tabelle 11: Antworten auf die Frage: "Is the impact of the new GCSE on your teaching rather positive or negative?"**

Antwortmöglichkeit	Anzahl
1 = very negative	5
2 = rather negative	16
3 = rather positive	12
4 = very positive	0
nicht beantwortet	2

Während 21 Personen angeben, dass die Auswirkungen der alten GCSE-Prüfung eher bis sehr positiv waren und nur elf Lehrkräfte der Ansicht sind, dass der Einfluss eher negativ war, hat sich dieses Bild nun nahezu ins Gegenteil verkehrt: Aus 21 positiven wurden 21 negative Angaben. Immerhin fünf von 35 Lehrkräften betrachten die Auswirkungen heute sogar als sehr negativ, während gleichzeitig niemand mehr von sehr positiven Auswirkungen spricht. Insgesamt wird die Ausrichtung des Effekts somit heute als negativer wahrgenommen.

### 3.2.2.2 Verbale Daten und Interpretation

Um zu erfassen, wie sich der Backwascheffekt durch die Reform verändert hat, habe ich die Frage gestellt: „If the impact of the GCSE on your teaching changed in any way, please explain this briefly: what exactly changed and why?“.

Die hohe Intensität des Backwascheffekts wird durch die Antwort von TeilnehmerIn 11 deutlich:

I no longer teach question words as the questions in the exam are in English. This means that I will have to teach them to 6th form when they start the A-Level course. There is also an effect on how we assess at Key Stage 3. (F8T11)

TeilnehmerIn 21 beschreibt, wie das Rollenverständnis von Lehrkräften in Verbindung mit dem Notendruck zu einem besonders starken Backwascheffekt führt:

It really is difficult to say. Our job as teachers is to prepare the pupils for the exams and facilitate them achieving their best. So, of course, every change in the exam syllabus has a knock on effect on our teaching. (F8T26)

LehrerInnen fühlen sich zwar nicht unmittelbar durch die Reform für die Testresultate verantwortlich, doch wie im folgenden Zitat gezeigt wird, hat sich der Druck verstärkt, indem sich die Möglichkeiten der Einflussnahme verändert haben: „I have to think about which topics will get the students the best results“ (F8T21). Auch

Hausübungen werden nun gezielt als Testvorbereitung eingesetzt: „become more focussed on how I set hwk as a preparation for the controlled assessment“ (F8T29)

Im folgenden Kommentar wird ein regelrechter Zwang zum *Teaching to the Test* geschildert, der auf die Einführung von *Controlled Assessment* zurückgeht. Eine negative Haltung gegenüber dieser Entwicklung ist dabei deutlich erkennbar:

My teaching has become much more focussed to the exam. I dislike teaching to the test but now that we are preparing for a CA its all I can do! (F8T23)

Ein verstärktes *Teaching to the Test* kritisiert auch TeilnehmerIn 19 vehement:

I now spoon feed pupils to pass silly tasks. This has been the worst change to GCSE German in the history of the exam. Paper work for teachers has increased. Teachers now have to mark and award a grade for the Speaking - this will be inconsistent accross the country and is not fair. (F8T19)

Gleichzeitig zeigt dieser Kommentar die größere Verantwortung der Lehrkräfte, die durch das selbstständige Beurteilen entsteht. Im Kontext des Notendrucks der League Tables ist durchaus verständlich, wieso dies als unfair bezeichnet wird (siehe auch Abschnitt 3.2.1.2).

Die Hypothese zum allgemeinen Backwascheffekt in Abschnitt 3.2.1 hat die Themen Druck, bzw. Notendruck zwar ebenfalls behandelt, allerdings im Kontext der League Tables und des GCSE allgemein. In diesem Abschnitt wird von Veränderungen in konkreten Bereichen, die von der Reform betroffen sind, berichtet. Zwischen beiden Problemen besteht ein Zusammenhang: Der allgemeine Druck hat eine starke Orientierung an Tests und deren Ergebnissen zur Folge, die sich in diversen von der Reform betroffenen Bereichen (z.B. freie Themenwahl, Verantwortung für die Aufgabenerstellung und -bewertung) widerspiegelt.

Auch die Frage: “Was the teaching of GCSE classes different two or more years ago? Please explain, what was different and how it changed.” steht damit in Zusammenhang. Es wird oftmals (vgl. F7T11, F7T15, F7T16, F7T23 im Anhang) auf die durch *Controlled Assessment* entstandene Zeitnot hingewiesen, wie z.B. hier von TeilnehmerIn 32:

there was time to work systematically through the grammar and vocabulary, whereas now we have to cover short topics and prepare for rote learning of a piece. (F7T32)

Ebenso wie in den Kommentaren F8T21, F8T23 und F8T26 weiter oben wird auch in diesem Beitrag das Gefühl des Zwangs geäußert. Es ist nicht der Wunsch dieser LehrerInnen, auswendig lernen zu fördern und *Teaching to the Test* zu betreiben,

aber sie sind der Auffassung, dass sie keine andere Möglichkeit haben, wie die Aussagen „we have to“(F7T32) und „its all I can do“(F8T23) illustrieren.

### 3.2.2.3 Fazit

Die Analyse der quantitativen Daten zeigt, dass LehrerInnen den Backwasheffekt heute als stärker und negativer empfinden als vor der Reform des GCSE. Auf die Frage, wie sich der Einfluss („Impact“) verändert habe, thematisieren mehrere Lehrkräfte einen starken Druck zum *Teaching to the Test* und einen starken Notendruck. Durch die Reform haben LehrerInnen mehr Einfluss auf die Resultate, da sie die Prüfungsaufgaben selbst erstellen und durch eine geringere Themenvielfalt, explizit auf *Controlled Assessment* zugeschnittene Hausübungen und das gezielte Einüben und Auswendiglernen von Texten die Noten positiv beeinflusst werden können. Auch wenn es den Druck bereits vorher gab (siehe Abschnitt 3.2.1), hat er sich durch die Einflussmöglichkeiten in der Wahrnehmung verstärkt.

### 3.2.3 Lernstrategie ‚Auswendiglernen‘

Hypothese 3: Die Lernstrategie ‚auswendig lernen‘ wird bei den *Controlled Assessment Tasks* dominieren.

#### 3.2.3.1 Quantitative Daten

Tabelle 12: Antworten auf die Frage: "Which strategies are your students using / did the students use to prepare for the assessment?"<sup>166</sup>

	writing a list of bullet points	practising with their peers	writing a full draft (written exam)	writing a script (oral exam)	studying vocabulary	studying grammar
1 = none of them	1	3	0	1	2	5
2 = some	13	6	5	5	13	14
3 = many	5	12	3	3	9	7
4 = (nearly) all of them	14	14	26	24	11	9
nicht beantwortet	2	0	1	2	0	0

<sup>166</sup> Diese Tabelle setzt sich aus den Antworten auf die zwei getrennten Fragen für jene, die Controlled Assessment bereits absolviert haben und für jene, die zum Zeitpunkt der Durchführung das erste Mal damit gearbeitet haben zusammen und ist dementsprechend anders formuliert. Ein Vergleich (siehe Tabellen 26 und 27 im Anhang) hat keine signifikanten Unterschiede zwischen den Gruppen gezeigt.

	looking for similar texts on the internet	writing notes / some paragraphs	recording themselves	rote learning / memorising	practising to become fluent	copying sentences from various sources with slight changes	trying to formulate in their own way
1 = none of them	23	3	6	1	2	2	1
2 = some	8	10	18	5	12	5	22
3 = many	1	10	8	6	9	20	8
4 = (nearly) all of them	1	12	1	23	7	8	4
nicht beantwortet	2	0	2	0	5	0	0

Die Tabelle zeigt, wie häufig welche Strategien für die Vorbereitung des *Controlled Assessment* verwendet wurden. Besonders interessant sind die Angaben für Antwortmöglichkeit vier: 26 Lehrkräfte (74%) geben an, dass fast alle SchülerInnen in ihren Klassen eine Draftversion für die schriftliche Prüfung schreiben bzw. geschrieben haben. Ähnliche Ergebnisse gibt es mit 24 Angaben (69%) für das Script bei der mündlichen Prüfung. Immerhin zwei Drittel der TeilnehmerInnen schildern, dass nahezu die ganze Klasse für *Controlled Assessment* Prüfungen auswendig lernt.

Mit 23, 24 und 26 Nennungen und damit von jeweils mindestens 66% der TeilnehmerInnen genannt, sind diese drei Strategien ganz klar die dominierenden. Erst mit einigem Abstand folgen das Schreiben einer Liste mit Stichworten und das Üben mit KollegInnen (jeweils 14 Nennungen, 40%), sowie das Schreiben von Notizen bzw. einigen Absätzen (12 Nennungen) und das Lernen von Vokabeln (11 Nennungen).

Ein Blick auf Antwortmöglichkeit eins soll die besonders selten verwendeten Strategien zeigen. Nur wenige SchülerInnen suchen ähnliche Texte im Internet: 66% der Befragten halten fest, dass keinE einziger SchülerIn diese Strategie nutzt. Eine Lehrkraft nimmt darauf Bezug, indem sie im Feld für weitere Strategien schreibt: „i have told them to avoid the internet as it is full of rubbish“ (F6T19). Fünf bzw. sechs Personen geben an, dass niemand in ihren Klassen als Vorbereitung Grammatik lernt bzw. sich selbst aufnimmt. Sonst wird diese Antwortmöglichkeit kaum gewählt.

Relativ wenige SchülerInnen versuchen, die Antworten auf ihre Prüfungsaufgaben selbst zu formulieren. Etwa zwei Drittel der Lehrkräfte wählten hier Kategorie eins oder zwei, im Gegensatz zu nur einem Drittel, das Kategorie drei oder vier wählte.

Insgesamt zeigt die Auswertung der quantitativen Daten hier ein sehr klares Bild: Texte vollständig auszuformulieren und (diese<sup>167</sup>) auswendig zu lernen ist die gängige Praxis.

### 3.2.3.2 Verbale Daten und Interpretation

Bisher hat die Auswertung nur gezeigt, dass auswendig gelernt wird, aber nicht, wie und warum es dazu kommt.

Um mehr über die Hintergründe zu erfahren habe ich die offene Frage gestellt: „How does the introduction of controlled assessment affect rote learning as exam preparation?“ Die Frage wurde von 29 TeilnehmerInnen ausgefüllt, wovon einmal nur auf eine frühere Antwort verwiesen wird. Es zeigen sich zwei Haupttendenzen:

- 1) Es wird mehr auswendig gelernt als früher.
- 2) Es wurde immer schon auswendig gelernt. Durch die Reform hat sich kaum etwas verändert.

Manche Kommentare thematisieren zwar die Dominanz dieser Strategie, aber fallen nicht in eine der beiden Kategorien, weil sie nur die heutige Situation darstellen. Sieben Antworten sind nicht kategorisierbar und sind lediglich im Anhang ersichtlich. In ihnen wird u.a. angesprochen, dass statt ganzen Texten flexiblere Strukturen gelernt werden sollten, dass es bei korrekter Durchführung nicht zum Auswendiglernen kommen sollte und dass vor allem schlechtere SchülerInnen darauf angewiesen sind. Einzelne betonen, dass SchülerInnen, die auswendig lernen, die besten Noten bekommen. Da es sich hierbei um eine Frage nach der Validität handelt, werden diese Antworten in einem späteren Abschnitt ausführlicher behandelt.

#### Antworten zu: 1) Es wird mehr auswendig gelernt als früher:

1. It encourages rote learning and I have seen this in practice. (F5T3)
2. Pupils are being encouraged to learn whole texts off by heart. Hugely negative. [...] At the moment, grades depend too much on how much the teachers cheat and on how much the pupils learn by rote. (F5T7)
3. It encourages rote learning. We now train our pupils in Years 7-9 to rote-learn before assessments. (F5T18)
4. You can now pass a GCSE by rote learning your prepared piece. (F5T19)

---

<sup>167</sup> Diese Verbindung kann allein durch die Auswertung der quantitativen Daten selbstverständlich nicht hergestellt werden. Erst die Analyse der qualitativen Daten zeigt diese Verbindung.



5. It makes it unavoidable. Students with the best memories will do best. Those who don't have a good memory are already saying, "I can't do that." (F5T20)
6. I will happen even more - an impossible task for some students (F5T21)
7. they do it more than before. (F5T22)
8. They do a lot more of this, previously they had to have a very good understanding of the grammar and structures so they could adapt this to the written exam tasks. Now they just learn write something for a topic and learn it off by heart (and then immediately forget it). (F5T28)
9. Increases the chances of this being the key factor in the exam. In stage 1 of the controlled assessment, teachers are allowed to get students to practise language they will use in the CA, so they will make sure they have somewhere, already written an answer to the CA, had it marked, before Stage 2. This means students will have a perfect answer to learn by heart. They'd be silly not to do this, as they are more likely to score the marks this way. (F5T30)
10. It will be used more. (F5T31)
11. it means that rote learning is becoming the only way of learning for many. (F5T32)

Nur Zitat 9 beschreibt den Prüfungsprozess, woraus sich Gründe für das verstärkte Auswendiglernen ableiten lassen. Demnach wird auswendig gelernt, weil sich erstens die Möglichkeit dazu bietet und zweitens so leichter bessere Noten erreicht werden könnten. Der erste Grund wird auch in Zitat 4 mit „You can now pass [...]“ angesprochen. Auffällig ist auch Zitat 3: sie bezieht sich auf die Stärke und Dauer des Backwasheffekts<sup>168</sup>.

Antworten zu: 2) Es wurde immer schon auswendig gelernt. Durch die Reform hat sich kaum etwas verändert.

- 1) Some will still do it, it depends on the type of learner they are. (F5T12)
- 2) I don't think that it does affect it much because even with the controlled assessments, they get to know what the task is and can still learn a presentation off-by-heart and likely questions. (F5T16)
- 3) I believe that in many schools rote learning of topics will still take place. (F5T24)
- 4) The same pupils that needed to rote learn for the previous exam still need to learn off by heart for this exam; but this time what they are learning is more than likely incorrect German because a teacher has not checked it. (F5T26)
- 5) It doesn't - those who can, will continue to do so, whilst those who can't, will continue to score less well (F5T27)

---

<sup>168</sup> Auch wenn Watanabe unter Dauer bzw. Länge das Anhalten des Effekts *nach* einer Prüfung meint, bezeichne ich den hier beschriebenen Effekt als Dauer.

- 6) they still do it but have less to do at any one time.  
Rote learning is not necessarily a bad thing - it does have its place (F5T33)

Es wird zwar weniger explizit darauf eingegangen, doch auch aus diesen Antworten lässt sich herauslesen, dass auswendig gelernt wird, weil die Möglichkeit dazu besteht, wie z.B. Zitat 2 („they [...] can still learn a presentation off-by –heart [...]“) und Zitat 5 („those who can, will continue to do so“) zeigen.

Antworten, die sich ausschließlich auf die heutige Situation beziehen.

- 1) It is not a far move from learning by hear as student panic and try to learn huge chunks of info prior to their exam (F5T5)
- 2) They just learn by rote and regurgitate their prepared pieces (F5T15)
- 3) It's all a lot of students do....and I think many teachers don't teach the nuts and bolts, so students have no choice but to rote learn, and they hate it. (F5T23)
- 4) They learn it all off by heart - simple. Out of my group, I don't think there was one that hadn't written it all out in advance and learnt it, or added anything spontaneously in the assessment session. (F5T29)

Hier werden weitere Gründe für das Auswendiglernen genannt: schlechte LehrerInnen, die den SchülerInnen keine andere Wahl lassen, und Prüfungsangst, wobei letzteres mit dem Notendruck zusammenhängen könnte.

Nach der Auswertung der qualitativen Daten zu dieser Frage gibt es zwar sehr eindeutige Hinweise auf die Häufigkeit und Dominanz dieser Strategie, aber noch immer nur wenige Begründungen dafür. Bisher scheint Notendruck der einzige Antrieb zu sein, die Möglichkeit auswendig zu lernen auszunützen. Es gibt allerdings noch mehr und teilweise auch aufschlussreichere Kommentare, die in Antworten auf andere Fragen zu finden sind. Die aussagekräftigsten möchte ich hier zitieren.

I feel they can do these assessments simply learning by rote as they focus on one topic area and they have up to 6 lessons to prepare the task in advance (F4T16)

Diese Lehrkraft nennt den Themenfokus und die lange Vorbereitungszeit als Gründe für das Ermöglichen dieser Strategie, was meiner Vermutung in Abschnitt 2.6.3 entspricht. Im nächsten Zitat wird die Überlegung angestellt, dass der Themenfokus zu besseren Noten führen könnte. Auf die Strategie Auswendiglernen wird dabei zwar nicht Bezug genommen, doch angesichts der vielen Kommentare weiter oben, nach denen wegen besseren Noten auswendig gelernt wird, könnte eine Verbindung aufgestellt werden:

Although the controlled assessment speakings take a lot more time, they are much more pupil friendly and I am hoping that they will achieve better marks

as a result of only having to concentrate on one topic at a time. It is more accessible for all abilities, whereas the old style was a horrible experience for lower ability pupils. (F3T12)

Im folgenden Beitrag werden die fehlende Spontaneität und die daraus resultierende fehlende Authentizität beklagt. Aufgrund der Möglichkeit, für GCSE-Prüfungen auswendig zu lernen, wird ein negativer Langzeitbackwascheffekt, der die A-Levels betrifft, befürchtet:

The speaking now has no spontaneity - my pupils learn 6 mini-speeches (one for each bullet point). It is not authentic to talk only on one topic. Similarly with the writing - pupils with good memories will succeed - it must be a nightmare for those who cannot memorise extended chunks of language. I am concerned about the longterm impact in terms of moving onto A Level where you simply cannot memorise your "speech" beforehand. (F3T18)

Sehr ähnlich sieht es die Person, von der dieser Kommentar stammt:

Controlled speaking is worse than controlled writing in my opinion - there is nothing spontaneous or unexpected at all about it. they only learn 6 brief statements and that is it - old style orals, they at least had to learn about all topic areas, even if they were only asked a few questions on a couple of areas. possibility of same topic - shocking. did actually consider doing it myself to let students get a good grade (just change one bullet point and you have a new task) but decided against it in the end. (F3T29)

Auch hier wird die fehlende Spontaneität beklagt, ebenso wie die zu starke Vorhersehbarkeit. Die Verbindung von Themenfokus und besseren Noten wird erneut thematisiert. Interessanterweise wird zwar vehement abgelehnt, dass für mündliche und schriftliche Prüfungen das gleiche Thema gewählt werden darf, aber es wurde trotzdem in Betracht gezogen. Das dürfte auf einen sehr starken Notendruck hinweisen.

Schließlich wird in diesem Beitrag angesprochen, dass auswendig lernen zu besseren Noten führt:

Controlled Assessment is better than coursework as it is more strictly controlled, but to do well, it still encourages students to learn by heart lots of prose that they have previously had marked, rather than to make the language up on the spot. (F3T30)

Das Thema ‚auswendig lernen für bessere Noten‘ wird im Abschnitt zur Validität noch ausführlicher behandelt. An dieser Stelle sei festgehalten, dass es ein Antriebsgrund für das Verwenden dieser Strategie ist.

### 3.2.3.3 Fazit

Notendruck dürfte tatsächlich der größte Antrieb für das Auswendiglernen sein. Ermöglicht wird es durch die lange Vorbereitungszeit, den Themenfokus und die fehlende Spontaneität. Obwohl sich viele LehrerInnen kritisch äußern und diese Strategie zwar theoretisch nicht für sinnvoll halten, können sie in der Praxis oft nicht widerstehen. Das vorherrschende Rollenverständnis dürfte folgendes sein: “Our job as teachers is to prepare the pupils for the exams and facilitate them achieving their best.” (F8T26). LehrerInnen sind demnach also hauptsächlich dafür verantwortlich, ihren SchülerInnen zu guten Noten zu verhelfen, auch wenn dafür Lehr- und Lernstrategien erforderlich sind, die nicht unbedingt dazu beitragen, eine Sprache selbstständig zu beherrschen. Es zeigt sich ein gewisser Zwiespalt zwischen dieser notenorientierten Rolle einerseits und dem Wunsch, SchülerInnen eine Sprache beizubringen andererseits. Dieser Zwiespalt, der meines Erachtens mit fehlender Validität in Zusammenhang steht, wird durch einen starken allgemeinen Backwash in Form von Notendruck verstärkt.

### 3.2.4 Validität

Hypothese 4: *Controlled Assessment* reduziert die Validität.

#### 3.2.4.1 Quantitative Daten

**Tabelle 13: Antworten auf die Frage: “Do you think that Controlled Assessment says more or less about students’ actual abilities than the previous test types?”**

Antwortmöglichkeit	Anzahl
1 = rather more	5
2 = rather less	20
3 = the same	4
4 = don't know	6
nicht beantwortet	0

Die Frage nach den ‘actual abilities’, also nach den wahren Fähigkeiten und Fertigkeiten, ist eine Frage nach der Validität. Tendenziell soll sich die Validität im Vergleich zur früheren GCSE-Prüfung verschlechtern haben.

**Tabelle 14: Antworten auf die Frage: "Do / will the results reflect your students' actual communicative competence?"**

Antwortmöglichkeit	Anzahl
1 = they (will) correspond very much	2
2 = they (will) correspond well enough	12
3 = they won't/ don't really correspond	15
4 = I can't judge	3
5 = I haven't got any results yet	3
nicht beantwortet	

Auch die Frage, ob die Ergebnisse die wahre kommunikative Kompetenz widerspiegeln, ist eine Frage nach der Validität. Kaum jemand hält *Controlled Assessment* für sehr valide. Während 12 TeilnehmerInnen die Prüfung als ausreichend valide bezeichnen, sind 15 andere Lehrkräfte der Ansicht, dass die Ergebnisse nicht wirklich die kommunikative Kompetenz der KandidatInnen zeigen. Von jenen, die dazu eine Meinung haben, sind das knapp mehr als 50%. Angesichts der geringen Stichprobe ist das Ergebnis allerdings nicht besonders aussagekräftig.

Weiters stellt sich die Frage, in welcher Weise *Controlled Assessment* die Deutschkompetenz der SchülerInnen beeinflusst. Wenn ein Test negative Auswirkungen auf die Deutschkompetenz der SchülerInnen hat, testet er vermutlich etwas anderes und ist somit nicht sehr valide. Aus diesem Grund steht auch die folgende Frage indirekt mit der Validität in Zusammenhang.

**Tabelle 15: Antworten auf die Frage: "How will the introduction of controlled assessment affect the students' overall German proficiency?"**

Antwortmöglichkeit	Anzahl
1 = very negatively	7
2 = a bit negatively	19
3 = no changes	6
4 = a bit positively	0
5 = very positively	1
nicht beantwortet	2

Aus Tabelle 15 geht hervor, dass die Einführung von *Controlled Assessment* die Deutschkompetenz der SchülerInnen eher negativ beeinflusst, was nicht nur von einem negativen Backwascheffekt zeugt, sondern auch die Frage aufwirft, was die neue Prüfungsform tatsächlich fördert und fordert. Dies wird im folgenden Abschnitt behandelt.

### 3.2.4.2 Verbale Daten und Interpretation

Zu dem Thema, ob mit *Controlled Assessment* die wahren Fähig- und Fertigkeiten der SchülerInnen getestet werden, wird u.a. wie folgt geantwortet:

1. I don't think we're testing proficiency in German. I think we're testing how well they can prepare for an exam. (F4T20)
2. I feel that instead of learning the language and how it works in many situations, the pupils are simply being trained to pass an exam. (F3T32)
3. it shows how good students' memory is, rather than their ability in german (F4T32)
4. In CA, students can simply copy out corrected work from their books and reproduce it- which does not show any of their true ability. The people who get the best marks are the ones who can learn it the best. (F4T34)
5. The new CA is a test of memory and doesn't say very much at all about a students ability - especially considering the amount of teacher input allowed in early stages. (F4T23)
6. The L and R skills are also now worth 20% each, not 25%, so the exam rewards rote learners in the controlled assessments more than competent linguists who can work out meaning (F2T11)

Nach den Aussagen dieser TeilnehmerInnen testet *Controlled Assessment* nicht die wahren Fähig- und Fertigkeiten der KandidatInnen. Stattdessen soll entweder die Fähigkeit, sich auf Prüfungen vorzubereiten (Zitate 1 und 2) oder die Fähigkeit, auswendig zu lernen (Zitate 3, 4, 5 und 6) überprüft werden. TeilnehmerIn 11 weist in diesem Zusammenhang auf die ungleiche Gewichtung der Fertigkeiten hin, die das Auswendiglernen fördern soll. Niemand verteidigt *Controlled Assessment* diesbezüglich (siehe Antworten im Anhang).

Auch in Abschnitt 3.2.3.2 wurde bereits am Rande erwähnt, dass Auswendiglernen als Prüfungsvorbereitungsstrategie belohnt wird. Theorie und Praxis scheinen nicht überein zu stimmen. Edexcel (2009) beschreibt in einem Leitfaden, wie *Controlled Assessment* theoretisch auszusehen hat:

It is not in the spirit of the new GCSE controlled assessments for speaking for teachers to simply provide their students with a list of pre-set questions to be rote-learned and regurgitated. All tasks – open interaction, picture-based discussion and presentation should aim for natural interaction.<sup>169</sup>

---

<sup>169</sup> Edexcel (2009b), S. 14.

Die Praxis beschreibt TeilnehmerIn 15 jedoch mit den Worten: „They just learn by rote and regurgitate their prepared pieces” (F5T15), was nicht nur dem Gegenteil der im Leitfaden beschriebenen Wunschvorstellung entspricht, sondern erstaunlicherweise auch mit dem gleichen Wortlaut ausgedrückt wird. Dieser Widerspruch zwischen Theorie und Praxis dürfte darin begründet liegen, dass die Prüfung nach Ansicht der Lehrkräfte Auswendiglernen tatsächlich in der Art begünstigt, dass andere Strategien schlechtere Noten zur Folge haben:

Memorised chunks seem to be better rewarded than spontaneous, natural communication with a few more errors. (F3T11)

It makes it unavoidable. Students with the best memories will do best. Those who don't have a good memory are already saying, "I can't do that." (F5T20)

It doesn't - those who can, will continue to do so, whilst those who can't, will continue to score less well (F5T27)

Offensichtlich ist keinerlei Sprachkompetenz nötig um sich auf diese Prüfung vorzubereiten:

The students only learn the text by heart which was almost written by the teacher. Besides this text, the pupils know virtually nothing of German. (F4T6)

Besonders problematisch erscheint, dass selbst die Exam Boards, die in ihren offiziellen Begleitschreiben betonen, dass es keine auswendig gelernten Fragenkataloge geben sollen, bei Fortbildungsseminaren etwas ganz anderes durchklingen lassen:

They still learn things off by heart, for speaking assessments especially, even the training courses run by the exam board tell you to get them to learn it "it's a memory test" said the AQA examiner to the room full of teachers. The only thing now is, because you can't mark the drafts, they learn incorrect German off by heart. (F3T22)

Scheinbar tragen die Exam Boards eine gewisse Mitschuld an der fehlenden Validität. Diese widersprüchliche Haltung müsste jedenfalls noch genauer untersucht werden.

Bezüglich der Aussagekraft der Ergebnisse wird zwar eine Verbesserung im Vergleich zur *Coursework* genannt, doch gleichzeitig soll *Controlled Assessment* eine Verschlechterung gegenüber dem Abschlusstest darstellen. Dabei wird insbesondere die fehlende Spontaneität beklagt, die richtigerweise mit Validität in Relation gebracht wird:

Depends. Definitely shows more than coursework, but a lot less than written exam.

The speaking exams were also much harder, so CAs say less about a

students' ability. There was more speaking on the spot in the speaking exams. (F4T30)

Obwohl nach dieser Aussage und auch nach Tabelle 13 (s.o.) *Controlled Assessment* weniger über die tatsächlichen Fähigkeiten und Fertigkeiten aussagt als die frühere GCSE-Prüfung, bedeutet das nicht, dass die alte GCSE-Prüfung tatsächlich valide war. Der folgende Kommentar zeigt, dass auch bisher SchülerInnen nicht unbedingt zur Kommunikation fähig sein mussten um gute Noten zu erlangen:

I encounter many pupils, sometimes even those doing A level languages, who have gained high grades in GCSE but who cannot put together a simple sentence, and who cannot answer simple greetings and enquiries as to their health. Many are almost unaware of the difference between singular and plural forms of the second person. Very few can use personal pronouns. (F9T7)

### 3.2.4.3 Fazit

Dass die alte GCSE-Prüfung das Kriterium Validität nicht erfüllte, habe ich bereits in Abschnitt 2.5.4 festgestellt. Die neue GCSE-Prüfung wird als noch weniger aussagekräftig und in Bezug auf die kommunikative Kompetenz als nicht besonders valide angesehen. Die quantitative Auswertung zeigt, dass etwa die Hälfte der KandidatInnen keine bzw. kaum eine Korrespondenz zwischen den Prüfungsergebnissen und der fremdsprachlichen Kommunikationsfähigkeit ihrer SchülerInnen erkennen kann, während nicht ganz die andere Hälfte die Validität noch für akzeptabel hält. Nur drei Lehrkräfte erkennen eine große Übereinstimmung zwischen Testergebnissen und kommunikativer Kompetenz. Die qualitative Auswertung zeigt, dass stattdessen die Fähigkeit sich auf eine Prüfung vorzubereiten und die Fähigkeit auswendig zu lernen überprüft wird. Das Kriterium für positiven Backwash ‚Fähigkeiten testen, die gefördert werden sollen‘ ist somit nicht erfüllt.

In Bezug auf die Hypothese formuliert lässt sich zusammenfassen, dass *Controlled Assessment* tatsächlich die Validität reduziert, doch dass die Ansichten, ob der Validitätslevel noch akzeptabel ist, auseinander gehen.

### 3.2.5 Freies und kreatives Unterrichten

Hypothese 6:

Es kann in Zukunft mehr Zeit für freies und kreatives Unterrichten geben.



### 3.2.5.1 Quantitative Daten

Ich habe diese Hypothese auf zweierlei zurückgeführt: einerseits darauf, dass nun nicht mehr alle Themen, die unterrichtet werden, geprüft werden und daher in diesen Bereichen mehr Freiraum entsteht, und andererseits auf die Möglichkeit, Prüfungsaufgaben selbst zu erstellen und so der jeweiligen LernerInnengruppe anzupassen. Dadurch müssten individuelle Schwerpunktsetzungen möglich werden. Die Ergebnisse zeigen allerdings eher das Gegenteil.

**Tabelle 16: Antworten auf die Frage: “How much room do you now have for personalised and creative teaching (in comparison to the previous GCSE)?”**

Antwortmöglichkeit	Anzahl
1 = much less	15
2 = a bit less	7
3 = about the same	10
4 = a bit more	2
5 = much more	1
nicht beantwortet	0

Zweiundzwanzig Personen sind der Ansicht, dass sie nun weniger Raum für personalisiertes und kreatives Unterrichten haben. Dem gegenüber stehen drei Lehrkräfte, die der Meinung sind, dass es mehr Zeit dafür geben wird. Die zehn verbleibenden TeilnehmerInnen nennen keine große Veränderung. Eine weitere Frage zum Zeitmanagement lieferte noch eindeutigere Ergebnisse:

**Tabelle 17: Antworten auf die Frage: “How do you think controlled assessment will affect the time available for content not treated in the exam?”**

Antwortmöglichkeit	Anzahl
1 = there will be considerably less time	23
2 = there will be a bit less time	6
3 = it will remain the same	4
4 = there will be a bit more time	1
5 = there will be considerably more time	0
nicht beantwortet	1

Tabelle 17 zeigt die Antworten auf die Frage: ‘How do you think controlled assessment will affect the time available for content not treated in the exam?’ Dreiundzwanzig von 35 TeilnehmerInnen gaben an, dass ab jetzt ‚wesentlich weniger Zeit‘ (‚considerably less time‘), und weitere sechs, dass nun ‚ein bisschen weniger Zeit‘ (‚a bit less time‘) für Inhalte zur Verfügung stehen werde, die bei der Prüfung nicht geprüft werden. Im Gegensatz dazu ist nur einE TeilnehmerIn der Meinung, dass es mehr Zeit dafür geben wird.

Auch die Daten zur Zeitintensität der Prüfungsvorbereitung vor (Tabelle 18) und nach (Tabelle 19) der Reform sind aufschlussreich. Heute finden 24 von 35 TeilnehmerInnen die Prüfungsvorbereitung sehr zeitintensiv. Weitere 10 Personen finden sie immer noch eher zeitintensiv. Für die Zeit vor der Reform sehen die Angaben noch etwas gemäßiger aus: 13 Lehrkräfte sind der Ansicht, dass die Vorbereitungen damals nur ein bisschen Zeit gekostet haben.

**Tabelle 18: Antworten auf die Frage: "How time-consuming was the preparation for the GCSE previously?"**

Antwortmöglichkeit	Anzahl
1 = not time-consuming	1
2 = a bit time-consuming	13
3 = rather time-consuming	15
4 = very time-consuming	5
nicht beantwortet	1

**Tabelle 19: Antworten auf die Frage: „How time-consuming is the preparation for the GCSE now?“**

Antwortmöglichkeit	Anzahl
1 = not time-consuming	0
2 = a bit time-consuming	1
3 = rather time-consuming	10
4 = very time-consuming	24
nicht beantwortet	0

### 3.2.5.2 Verbale Daten und Interpretation

Die qualitativen Daten zeigen, dass für kreativen Unterricht zu wenig Zeit zur Verfügung steht. Es wird oftmals kritisiert, dass *Controlled Assessment* so viel Unterrichtszeit kostet, dass manche Themen ausgelassen werden müssen oder nur mehr sehr oberflächlich behandelt werden können. Doch nicht nur die Unterrichtszeit, sondern auch die benötigte Administrationszeit wird als problematisch gesehen.

TeilnehmerIn 11 weist darauf hin, dass sowohl die Vorbereitung, die Durchführung als auch die Nachbereitung von *Controlled Assessment* sehr zeitintensiv ist. Das führt zu einer Vernachlässigung der rezeptiven Fertigkeiten, die erst am Ende geprüft werden:

Controlled assessments require 2 weeks' worth of lessons for preparation, sitting the assessment(s) and then marking and giving feedback. If we do three sets of assessment throughout the course, we lose a half-term of teaching. As all the assessments have to be in by May of Year 11, there is a very dry spell between May and June where we panic and try to teach them

comprehension skills. The course therefore has to be carefully planned to include comprehension skills from earlier on (F7T11)

Oben wurde angesprochen, dass für die rezeptiven Fertigkeiten nur mehr wenig Zeit bleibt. Im folgenden wird beschrieben, wie Themen, die nicht mittels *Controlled Assessment* überprüft werden, aufgrund von Zeitmangel nur mehr oberflächlich und nur mehr für die Rezeption unterrichtet werden:

Focus on one topic in depth whilst other topics not needed for CA are taught with much less detail (because there is less time) and really just for comprehension. (F8T28)

Auch hier wird die Zeitintensität thematisiert, und zwar mit einem Fokus auf der Durchführungs-, Nachbereitungs- und Administrationszeit:

The marking is a headache too - I have 56 candidates (2 classes) - just to physically LISTEN to their 2 Speaking assessments (2 X 6 minutes per candidate) will take me over 11 hours of solid listening. The speaking is difficult to arrange and now we have to do two lots, we are struggling with timetabling to get these done. The other difficulty is marking these, as it is very time consuming and access to the sound files can be limited. (F9T18)

Aus den nächsten Zitaten geht hervor, dass besonders die Prüfungsvorbereitung – also hier nicht die Durchführung – sehr zeitintensiv ist.

The new GCSE has potential to be very narrow. You spend a disproportionate amount of time preparing for the test rather than teaching a wide range of topics (F3T15)

I am unable to cover as many topic areas in depth as well as focus on grammar in more depth as it is very time consuming preparing for two speaking and two written assessments (F8T16)

Im folgenden Zitat wird der Druck zum *Teaching to the Test* thematisiert, der nicht viel Zeit für eine ausgewogene Behandlung aller Themen führt. Im Gegensatz dazu wird das alte GCSE erwähnt, bei dem kreatives Unterrichten möglich gewesen sein soll:

I now feel under pressure to over-teach certain topics and skim others. Each topic is taught with the bullet points in mind - everything is geared towards that.

In the old GCSE I taught a range of topics, often creatively, and then only in the final term or so did we start looking at actual exam papers and exam technique. (F8T18)

Es gibt zwar auch positive Stimmen zum *Controlled Assessment* (vgl. "Although the controlled assessment speakings take a lot more time, they are much more pupil

friendly [...]” (F3T12) in Abschnitt 3.2.3.2), aber nicht im Hinblick auf das Zeitmanagement.

Weiter oben rechneten bereits zwei TeilnehmerInnen (F7T11 und F9T18) vor, wie viel Zeit *Controlled Assessment* in Anspruch nimmt. Es gibt mehrere Lehrkräfte, die die Zeitintensität mit Hilfe dieser ‚harten Fakten‘ illustrieren.

we had more time with pupils. Controlled assessment stops us from teaching, in the time they are revising for the task and when we have to conduct speaking tasks. This adds up to nearly 8 weeks of teaching time in a year. (F7T22)

Controlled assessment (5 lessons preparation time) takes 3 months out of pupils' learning time over the course of the GCSE. (F5T17)

Die Ergebnisse dieser Rechnungen variieren zwischen acht Wochen im Jahr und drei Monaten bzw. einem halben Trimester (üblicherweise sechs bis sieben Wochen) während des zweijährigen Kurses. Wie viel Zeit für *Controlled Assessment* benötigt wird ist damit zwar individuell verschieden, doch subjektiv offensichtlich zu viel.

Meine Hypothese, dass es durch die Einführung von *Controlled Assessment* mehr Zeit für kreatives Unterrichten geben könnte, ist somit widerlegt. Interessant ist jedoch, dass diese Überlegung ursprünglich im Gespräch mit einer Lehrerin der John Warner School entstanden ist. Da sie meinen Fragebogen bereits in der Pretestphase ausgefüllt hat, waren ihre Antworten für mich leicht zu ermitteln: Auch sie befürchtete beim Ausfüllen im Dezember eine zu hohe Zeitintensität. Dieser Sinneswandel von Oktober bis Dezember beschäftigte mich, so dass ich mich entschloss, per E-Mail nachzufragen. Meine Frage lautete:

I am writing to you regarding some of the findings of my survey. When we discussed the impact of controlled assessment in October, you mentioned that there might be more room for creative teaching than before. At that time, you taught the topic “environment” to a year 11 class and told me that there won’t be any assessments relating to it, so that you can approach the topic in a more interesting way than in the previous years. In my survey, however, only three out of 35 participants said that there will be more room for “personalised and creative teaching”, so I had a look at your questionnaire, and you too answered that there will be “much less” room for that than before. Now I wonder why that is: Did you change your opinion between October and December – possibly because you started with CA or discussed this with other teachers? – or did you interpret the question in a different way?

Hier die Antwort der betreffenden Lehrerin:

I think my opinion changes about the the new GCSE as we go through it. Overall, I and other teachers have feared that there is less room for 'presonalised and creative' teaching given that we have to give over 2 weeks

each time to the preparation of the written and spoken CA. If you therefore wish to have the pupils produce a formal written piece for each topic they do, then you are looking at 8 to 10 weeks of teaching time lost to that, and then a further 4 weeks for the preparation for their speaking tasks. That is practically a whole term.

In theory though, there seems to be more room for creativity as the general topic areas have been simplified, on the face of it, in terms of what pupils need to be able to produce either written or spoken, and the guidelines are very loose - so yes, in theory there is room to be creative, if teachers find sufficient time and are willing to plan their own schemes of work more so than they did before.

There are no suggested questions and answers to use for preparing for the speaking CA, so again there is the room to personalize. It is the usual problem though of things in theory looking inviting, but in reality being very challenging because of time constraints, and it is also worth mentioning that teachers could get it all horribly wrong if given too little guidance. [...] <sup>170</sup>

Wie auch schon in der Fragebogenstudie mehrmals erwähnt, ist die Vorbereitung auf das *Controlled Assessment* zu zeitintensiv um frei unterrichten zu können. Gleichzeitig wird allerdings angesprochen, dass zumindest theoretisch mehr Raum für kreativen Unterricht wäre. Das Gespräch im Oktober, auf dem meine Hypothese basierte, fand vor dem ersten *Controlled Assessment* statt und konnte die Auswirkungen somit nur in der Theorie behandeln. Als diese Lehrerin im Dezember an meiner Studie teilnahm, konnte sie bereits Praxiserfahrung vorweisen und schätzte die Reform negativer ein. Der Widerspruch zwischen meiner Hypothese und den Ergebnissen ist demnach höchstwahrscheinlich ein Widerspruch zwischen Theorie und Praxis.

### 3.2.5.3 Fazit

Die Hypothese, dass durch die Reform des GCSE mehr Zeit für kreatives Unterrichten zur Verfügung stehen würde, ist leider widerlegt. Es handelte sich hierbei um die einzige positive Backwash-Hypothese. Sowohl die quantitativen als auch die qualitativen Daten lassen eine hohe Zeitintensität des reformierten GCSE erkennen. Der Zusammenhang zwischen kreativem Unterricht und der Zeitintensität des GCSE wird schließlich im E-Mail einer Lehrerin aufgestellt. Demnach gibt es zwar theoretisch Raum für kreativen Unterricht, der allerdings in der Praxis nicht ausgefüllt werden kann.

---

<sup>170</sup> Das ganze E-Mail ist im Anhang ersichtlich.

Der reformbegleitende Kontakt zu dieser Lehrerin hat sich als nützlich erwiesen, da dadurch sowohl die Erwartungen vorher als auch die gewonnenen Erfahrungen erfasst werden konnten. Nur dadurch konnte der Widerspruch zwischen Theorie und Praxis aufgedeckt werden.

### 3.3 Abschließende Diskussion

Ich habe dieses Thema ursprünglich aufgrund meiner negativen Erfahrungen mit der alten GCSE-Prüfung gewählt. Aus der theoretischen Auseinandersetzung mit Testgütekriterien, Bedingungen und Eigenschaften von positivem wie negativem Backwash und den Prüfungsbedingungen habe ich gefolgert, dass die hohe Standardisierung für die fehlende Kommunikativität und somit für den negativen Backwash verantwortlich ist. In Abschnitt 2.1.4 wurde das Spannungsfeld zwischen den verschiedenen Testgütekriterien beschrieben. Insbesondere die kommunikativ ausgerichteten Gütekriterien Authentizität und Interaktivität sind unvereinbar mit den klassischen Testgütekriterien, bei denen Messbarkeit im Vordergrund steht. In Abschnitt 2.5.4 habe ich analysiert, dass dieses Spannungsfeld bei der alten mündlichen GCSE-Prüfung zugunsten der Objektivität gelöst wurde, so dass die Prüfung die hohen Anforderungen kommunikativer Tests nicht erfüllte.

Heute steht die Objektivität bei den mündlichen GCSE-Prüfungen nicht mehr im Vordergrund, was zur Folge hat, dass die Prüfung theoretisch kommunikativer gestaltet sein könnte. Doch schon bei der ersten Beschäftigung mit der überarbeiteten Prüfung (vgl. Abschnitt 2.5) hat sich gezeigt, dass *Controlled Assessment* noch weniger valide ist, noch weniger echte Kommunikation fördert, Auswendiglernen erleichtert und damit insgesamt einen negativeren Backwash als die frühere Prüfung hat. Die Ergebnisse der empirischen Untersuchung bestätigen diese Einschätzung. Tabelle 14 zeigt beispielsweise, dass die Ergebnisse des *Controlled Assessment* nicht die kommunikative Kompetenz der KandidatInnen widerspiegeln. Es wurden im Zuge der Untersuchung mehrfach Aussagen getätigt, wonach natürliche Interaktion schlechter bewertet wird als auswendig gelernte Phrasen (F3T11, F5T20, F5T27 in Abschnitt 3.2.4.2). Kommentare in Abschnitt 3.2.3.2, wonach die Prüfung keinerlei Spontaneität erfordert (F3T18, F3T29) und selbst ganze Klassen die mündliche Prüfung absolvieren, ohne auch nur ein einziges Wort spontan hinzuzufügen (F5T29) oder auch die Tabelle in Abschnitt 3.2.3.1, die zeigt, dass freies Formulieren eine vergleichsweise seltene Strategie ist, weisen auf eine fehlende Authentizität und Interaktivität hin. Die kommunikativ orientierten Gütekriterien sind damit wieder nicht erfüllt.

Trotzdem sah ich die zumindest in der Theorie bestehende Möglichkeit, die Prüfung offener und kreativer zu gestalten, persönliche Schwerpunkte zu setzen und so die

Motivation zu steigern. Diese – einzige! – positive Erwartung hat sich nicht erfüllt. Obwohl ich also insgesamt eine negative Entwicklung vorhergesagt habe, ist das Ausmaß an negativen Reaktionen doch überraschend. In Anbetracht dieser negativen Einstellung zur Reform drängt sich die Frage auf: Könnte es sein, dass die alte Prüfung doch nicht so schlecht war, wie ich aufgrund von theoretischen Überlegungen und Gesprächen mit einzelnen Lehrkräften angenommen habe?

Die Kommentare zeigen oft eine relativ positive Einstellung zu den Rollenspielen und zur mündlichen und schriftlichen Abschlussprüfung allgemein. Weiters zeigt Tabelle 10 (Antworten auf die Frage: „Was the impact of the old GCSE on your teaching rather positive or negative?“), dass der alten GCSE-Prüfung ein positiver Backwash zugeschrieben wird. Wie ist das angesichts meiner Analysen zu deuten?

Ich interpretiere, dass diese positive Bewertung mit der Enttäuschung über die Reform zusammenhängt. Aus mehreren Kommentaren geht schließlich deutlich hervor, dass für die frühere mündliche Prüfung ebenso auswendig gelernt wurde und dass auch schon früher SchülerInnen mit herausragenden GCSE-Noten nicht unbedingt fähig dazu gewesen sein mussten, einen einfachen Satz frei zu formulieren. Das Zitat „Role plays made sure that pupils learned some useful transactional phrases, even learning these off by heart is better than nothing.“ (F3T22) zeigt, dass früher zwar ebenso auswendig gelernt wurde wie heute, aber dass die Inhalte sinnvoller waren. Auch die Studie<sup>171</sup> zur Vorhersehbarkeit der GCSE-Fragen darf nicht außer Acht gelassen werden. Dass die Prüfung früher schon diverse Probleme hatte, war nicht nur meine eigene Einschätzung, sondern auch die der evaluierenden Organisationen Ofqual und Ofsted<sup>172</sup>. Wenn das GCSE damals gut funktioniert hätte, wäre es vermutlich nicht mit großem Aufwand reformiert worden. Meine Interpretation lautet also, dass die Einstellung zum alten GCSE heute relativ positiv ist – mit der starken Betonung auf ‚relativ‘. Es wäre sehr interessant zu wissen, wie groß die Zufriedenheit vor der Reform war – doch für eine diesbezügliche Studie ist es heute leider zu spät.

Es stellt sich nun die Frage, warum sich die Auswirkungen der Prüfung und damit die Einstellung der teilnehmenden Lehrkräfte ihr gegenüber derart verschlechtert haben. Wenn meine ursprüngliche Hypothese besagte, dass der negative Backwash eine

---

<sup>171</sup> Vgl. Ofqual (2008).

<sup>172</sup> Vgl. Ofsted (2008)



Folge von der Unvereinbarkeit von Standardisierung einerseits und kommunikativen Prüfungsbedingungen andererseits war – wieso ist dieses Spannungsverhältnis heute nicht anders gelöst? *Controlled Assessment* bietet schließlich theoretisch die Möglichkeit dazu.

Dazu finde ich nur eine Antwort: Der starke Notendruck der League Tables ist dafür verantwortlich.

Schulrankings haben eigentlich nur dann Sinn – falls sie überhaupt einen Sinn haben – wenn die Ergebnisse vergleichbar sind, d.h. bei standardisierten Prüfungen. Die Abkehr von standardisierten Tests bei gleichzeitiger Beibehaltung der League Tables schafft einen Zwiespalt, der LehrerInnen unter Druck setzt. Lehrkräfte müssen dann zwar noch immer notenorientiert unterrichten, sind aber gleichzeitig dafür verantwortlich, Prüfungsaufgaben zu erstellen, die ihren SchülerInnen einen Vorteil verschaffen. Dieser Druck hat in England verstärktes *Teaching to the Test* zur Folge. Je genauer LehrerInnen wissen, was ihre SchülerInnen können müssen, um die Prüfung zu bestehen, desto leichter kann dafür auswendig gelernt werden. Außerdem wird in meiner Studie angesprochen, dass dadurch LehrerInnen zum Schummeln verleitet werden. Eine Verbindung von nicht-standardisierten Prüfungen und League Tables ist also nicht sinnvoll.

Nun gibt es zwei Möglichkeiten: zurück zu den standardisierten Tests, oder hin zu einer Abschaffung der League Tables. Nachdem die frühere Situation – so positiv sie vielleicht im Vergleich zu heute erscheinen mag – ebenfalls stark kritisiert wurde, scheint ersteres nicht der richtige Weg zu sein. Die League Tables werden hingegen – ich erinnere an die Tabellen 6 und 7 – von vielen LehrerInnen negativ wahrgenommen, woraus ich schließe, dass deren Abschaffung positiv wäre. Die Diskussion über eine Abschaffung der League Tables ist nicht neu. Ofsted kritisiert das starke *Teaching to the Test* als Folge der League Tables, und die LehrerInnengewerkschaft plädiert ebenso für deren Abschaffung wie auch eine Studie von AQA<sup>173</sup>

---

<sup>173</sup> Lipsett, Anthea [14. 1. 2008]: Unions make case for scrapping league tables. <http://www.guardian.co.uk/education/2008/jan/14/schools.uk1?INTCMP=SRCH> (18. 3. 2011). (The Guardian, Onlineausgabe)  
 Paton, Graeme [21. 7. 2008]. Ofsted: Schools 'teaching to the test'. <http://www.telegraph.co.uk/education/2440091/Ofsted-Schools-teaching-to-the-test.html> (21. 3. 2011). (The Telegraph, Onlineausgabe)

Doch selbst nach einer Abschaffung der League Tables wird die positive Veränderung vielleicht nicht sofort eintreten. Jahrzehntelange Erfahrungen mit standardisierten Tests und Schulrankings haben das Rollenbild von LehrerInnen beeinflusst. Das Bewusstsein, nicht für die Noten der SchülerInnen und die Rangliste der Schule, sondern stattdessen für das Interesse an Sprachen und für das Fördern von Sprachkompetenz verantwortlich zu sein, müsste vielleicht erst geschaffen werden. Angesichts der vielen LehrerInnen, die die fehlende Validität und die Sinnlosigkeit und den Frust der Vorbereitungen von *Controlled Assessment* thematisieren, bin ich diesbezüglich allerdings zuversichtlich. Wichtig ist, dass sich diese LehrerInnen nicht von Stimmen aus der Bevölkerung entmutigen lassen, die sich nach einer Abschaffung der League Tables über fallende Standards beklagen. In Wales wurden die League Tables 2001 abgeschafft, woraufhin die Noten bei GCSE-Prüfungen schlechter wurden. Was meines Erachtens ein positives Zeichen für die Abkehr von notenorientiertem Unterricht ist, wird in Großbritannien von manchen als Zeichen negativer Schulentwicklung interpretiert. So näherte sich im Oktober 2010 eine Studie<sup>174</sup> der Universität Bristol dieser Problematik mit folgender Logik:

“[i]f results improved faster in England, league tables must be a good thing. If they improved faster in Wales they must be bad. The researchers found that results in England had improved faster - so league tables must be a good thing.”<sup>175</sup>

Dieser Logik gilt es entgegenzuhalten – schlechtere GCSE-Noten bedeuten nicht schlechteren Unterricht, sondern in erster Linie einen weniger stark notenorientierten Unterricht, was ich positiv bewerte.

Als weiteres Forschungsvorhaben könnte es interessant sein, eine Backwashstudie mit walisischen Lehrkräften durchzuführen um zu ermitteln, welche Auswirkungen deren GCSE auf den Deutschunterricht hat und was dies für die Kompetenz der SchülerInnen bedeutet. Im Gegensatz zur Studie der Universität Bristol stehen dann nicht Noten im Vordergrund, sondern das, was wirklich zählt – das Lernen einer Sprache.

---

Paton, Graeme [16. 9. 2009]: Scrap school league tables, say experts. <http://www.telegraph.co.uk/education/6193787/Scrap-school-league-tables-say-experts.html> (18. 3. 2011). (The Telegraph, Onlineausgabe)

<sup>174</sup> Vgl. Burgess, Wilson, Worth (2010).

<sup>175</sup> Mansell, Warwick [12. 11. 2010]: The untruth about Wales and league tables. <http://www.tes.co.uk/article.aspx?storycode=6063124> (18. 3. 2011). (TES Connect)

## 4 Zusammenfassung

*“I dislike teaching to the test but now that we are preparing for a CA its all I can do!”*  
(F8T23)

Eine Analyse der alten GCSE Prüfung hat zunächst den Schluss nahegelegt, dass das Spannungsverhältnis zwischen Objektivität und den kommunikativ orientierten Testgütekriterien und die aus dieser Unvereinbarkeit resultierende fehlende Konstruktvalidität für den negativen Backwasheffekt verantwortlich sind. Daher stellte sich die Frage, inwiefern die Reform des GCSE, die eine Abkehr von zentralen Tests beinhaltete, den Backwasheffekt verändern würde.

Schon die theoretische Darstellung der überarbeiteten Prüfung ließ einen negativen Backwash befürchten, und zwar insbesondere aufgrund der langen Vorbereitungszeit und dem möglichen Themenfokus. Beides erleichtert die Strategie Auswendiglernen und fördert das *Teaching to the Test*. Gleichzeitig stellte ich allerdings die Hypothese auf, dass die Möglichkeit, die Unterrichtsthemen selbst zu wählen und anschließend selbst Prüfungsaufgaben zu erstellen, mehr Raum für individualisierten, kreativen Unterricht bieten würde. Die Abkehr von standardisierten Tests könnte dann die Motivation steigern und kommunikativen Unterricht fördern.

Die Ergebnisse dieser Untersuchung zeigen eine große Unzufriedenheit mit der Reform des GCSE. Verstärktes *Teaching to the Test*, sehr starker Zeit- und Notendruck, sowie noch mehr Auswendiglernen als schon bisher werden beklagt. Im Gegensatz zu meinen Erwartungen gibt es heute weniger Raum für freies und kreatives Unterrichten. Insgesamt deuten die Daten auf einen starken allgemeinen Backwasheffekt hin, der mit dem Notendruck der League Tables in Zusammenhang steht. Aufgrund dieses starken Drucks können Lehrkräfte nicht aus dem System des *Teaching to the Test* ausbrechen, obwohl viele es als negativ empfinden. Dies lässt den Schluss zu, dass eine Abschaffung der League Tables positive Auswirkungen auf den Unterricht haben sollte.

*League tables just cannot help. They only inform parents how well a student is prepared for an exam and not a reflection on the quality of teaching and learning. Remove this pressure and you will see schools experimenting and taking risks and doing exciting things... (F9T13)*

## 5 Glossar

6 <sup>th</sup> Form	12. – 13. Schulstufe Bezeichnung für die zwei Jahre zwischen dem Vollenden der Schulpflicht (GCSE) und dem Erreichen der Hochschulreife (A-Levels)
A-Levels	‚Advanced Level General Certificate of Education‘ Höchster Schulabschluss und Studienberechtigung Besteht aus den Advanced Subsidiary (AS)-Levels am Ende der 12. Schulstufe und der A2-Prüfung am Ende der 13. Schulstufe
AQA	Assessment and Qualifications Alliance Ein Exam Board
DCSF, DfES, DfE	Das Unterrichts-/Bildungsministerium Bis Juni 2007: DfES (Department for Education and Skills) Bis Mai 2010: DCSF (Department for Children, Schools and Families) Seit Mai 2010: DfE (Department for Education)
Edexcel	Ein Exam Board
Exam Board	Eine Organisation, die für die Erstellung, Verbreitung und Bewertung von nationalen Prüfungen verantwortlich ist (z.B. Edexcel, AQA, OCR)
GCSE	‚General Certificate of Secondary Education‘ Prüfung, die am Ende der Schulpflicht nach der 11. Schulstufe abgelegt wird und vor kurzem reformiert wurde
Key Stage (KS)	Die Schullaufbahn ist in ‚Key Stages‘ unterteilt: Key Stage 1: 1.-2. Schulstufe Key Stage 2: 3.-6. Schulstufe Key Stage 3: 7.-9. Schulstufe Key Stage 4: 10.-11. Schulstufe
League Tables	Schulranglisten Je nach Key Stage werden hier unterschiedliche Informationen aufgelistet (z.B. ‚National Curriculum Test‘-Ergebnisse in der KS2, Prozentsatz der GCSE-KandidatInnen mit fünf oder mehr A*-C in der KS4 und Durchschnittspunktwerte der A-Level-KandidatInnen in der 6th Form. <sup>176</sup>

<sup>176</sup> DCSF. Publication of 2010 test and examination results in the school and college achievement and attainment tables 2010. <http://www.education.gov.uk/performance/tables/Statement-of-Intent-Final-2010-April.pdf>

OCR	Oxford Cambridge and RSA Examinations Ein Exam Board
Ofqual	Office of Qualifications and Examinations Regulation Für die Regulierung von nationalen Tests verantwortlich
Ofsted	Office for Standards in Education, Children's Services and Skills Für die Schulinspektion und Publikation von schulspezifischen Berichten verantwortlich
QCA, QCDA	Qualifications and Curriculum (Development) Agency Seit April 2010: QCDA Verantwortlich für die Entwicklung des Lehrplans und der Prüfungskriterien für alle landesweiten Tests. Die Institution wird im Zuge der aktuellen Bildungsreform geschlossen. Die Website der QDCA wird noch bis Herbst 2011 erreichbar sein.

Die Beschreibungen sind, sofern dies möglich war, inhaltlich dem Glossar<sup>177</sup> auf der Homepage des Department for Education (DfE) entnommen. In allen anderen Fällen stammen die Informationen von den offiziellen Websites der jeweiligen Organisationen und Institutionen<sup>178</sup>.

---

<sup>177</sup> <http://www.education.gov.uk/help/atozandglossary/>

<sup>178</sup> Siehe Bibliographie

## 6 Bibliographie

### 6.1 Sekundärliteratur

Adams, Julie: On course for coursework. Second edition. London: CiLT 2002 (Pathfinder 35).

Alderson, J. Charles / Clapham, Caroline / Wall, Dianne: Language test construction and evaluation. Cambridge: Cambridge University Press 1995. (Cambridge Language Teaching Library)

Alderson, J. Charles / Wall, Dianne: Does washback exist? In: Applied Linguistics 14/2 (1993), S. 115-129.

Bachman, Lyle F. / Palmer, Adrian S.: Language testing in practice: designing and developing useful language tests. Oxford: Oxford University Press 1996.

Bachman, Lyle F.: Fundamental considerations in language testing. Oxford: Oxford University Press 1990.

Bailey, Kathleen M: Working for washback: a review of the washback concept in language testing. In: Language Testing 13/3 (1996), S. 257-279.

Bolton, Sybille: Die Gütebestimmung kommunikativer Tests. Tübingen: Narr 1985.

Bolton, Sybille: Probleme der Bewertung bei den produktiven Fertigkeiten in den Deutschprüfungen des Goethe-Instituts. In: Kühn, Ingrid (Hg.): Sprachtests in der Diskussion. Frankfurt am Main: Peter Lang 2005, S. 53-59.

Bolton, Sybille: Tests im Wandel theoretischer Prämissen. In: Gardenghi, Monica / O'Connell, Mary (Hg.): Prüfen, Testen, Bewerten im modernen Fremdsprachenunterricht. Frankfurt am Main: Peter Lang 1997, S. 17-23.

Brown, James Dean: Testing in language programs. New Jersey: Prentice Hall 1996.

Buckby, Michael / Corney, Kate: Getting the best results at GCSE. London: CiLT 1999. (Pathfinder 39. A CILT series for language teachers)

Buckby, Michael / Pugh, John: Revise for German GCSE. Listening and speaking. Oxford: Heinemann 1997.

Byram, Michael (Hg.): Routledge encyclopedia of language teaching and learning. London and New York: Routledge 2004.

Canale, Michael / Swain, Merrill: Theoretical bases of communicative approaches to second language teaching and testing. In: Applied Linguistics 1/1 (1980), S. 1-47.

Cheng, Liying (Hg.): Changing language teaching through language testing. A washback study. Cambridge: Cambridge University Press / UCLES 2005. (Studies in Language Testing)

Cheng, Liying / Curtis, Andy: Washback or backwash. A review of the impact of testing on teaching and learning. In: Cheng, Liying / Watanabe, Yoshinori / Curtis, Andy (Hg.): Washback in language testing. Research contexts and methods. New Jersey: Lawrence Erlbaum Associates 2004, S. 3-17.

Davis, Andrew: The limits of educational assessment. Oxford: Blackwell 1998.

Denscombe, Martyn: Social conditions for stress: young people's experience of doing GCSEs. In: British Educational Research Journal 26/3 (2000), S. 359 – 374.

Fuhs, Burkhard: Qualitative Methoden in der Erziehungswissenschaft. Darmstadt: Wissenschaftliche Buchgesellschaft 2007.

Glaboniat, Manuela: Kommunikatives Testen im Bereich Deutsch als Fremdsprache. Eine Untersuchung am Beispiel des Österreichischen Sprachdiploms Deutsch. Innsbruck, Wien: Studien-Verlag 1998. (Theorie und Praxis – österreichische Beiträge zu Deutsch als Fremdsprache, Bd. 2)

Grenfell, Michael: Communication: sense and nonsense. In: Swarbrick, Ann (Hg.): Teaching modern languages. London: Routledge 1994.

Grotjahn, Rüdiger: Testtheorie. Grundzüge und Anwendungen in der Praxis. In: Wolff, Armin / Tanzer, Harald (Hg.): Sprache - Kultur - Politik. Regensburg: Fachverband Deutsch als Fremdsprache 2000. (Materialien Deutsch als Fremdsprache 53), S. 304-341.

Hansen, Kirstine / Vignoles, Anna: The United Kingdom education system in a comparative context. In: Hansen, Kirstine / Vignoles, Anna (Hg.): What's the good of education? The economics of education in the UK. Princeton, New Jersey: Princeton University Press 2005, S. 13-36

Hedge, Tricia: Teaching and learning in the language classroom. Oxford: Oxford University Press 2000. (Oxford Handbooks for Language Teachers)

Hodgson, Ann / Spours, Ken: Education and training 14-19. Curriculum, qualifications and organization. London: SAGE Publications 2008.

Holmes, Edmond. What is and what might be. London: Constable and Co Ltd. 1911. <http://www.educationengland.org.uk/documents/holmes> (29. 3. 2011)

Hughes, Arthur: Testing for language teachers. Second edition. Cambridge: Cambridge University Press 2003. (Cambridge Language Teaching Library)

Jürgens, Eiko: Leistung und Beurteilung in der Schule. Sechste, aktualisierte und stark erweiterte Auflage. Sankt Augustin: Academia Verlag 2005.

Koglbauer, René Thomas: Deutschmatura in England und Ungarn am Beispiel der Fertigkeit Sprechen: eine Gegenüberstellung. Diplomarbeit. Universität Wien, 2003.

Krause, Wolf-Dieter / Sändig, Uta: Testen und Bewerten kommunikativer Leistungen im Unterricht Deutsch als Fremdsprache. Frankfurt am Main u.a.: Peter Lang 2002. (Sprache. System und Tätigkeit Bd. 43)

Kühhaas, Ulrike: Wie viel Grammatik verträgt ein kommunikativer Test? Diplomarbeit. Universität Wien, 2000.

Lloyd, Sarah: An introductory guide to education finance in England. London: CIPFA 2008.

Macaro, Ernesto: The decline in language learning in England: getting the facts right and getting real. In: Language Learning Journal 36/1 (2008), S. 101-108.

Madaus, George / Russel, Michael / Higgins, Jennifer: The paradoxes of high-stakes testing. How they affect students, their parents, teachers, principals, schools, and society. Charlotte, NC: Information Age Publishing 2009.

Mansell, Warwick: GCSEs move to controlled assessment. In: Education Journal 121 (2010), S. 7.

Mayring, Philipp: Qualitative Inhaltsanalyse. Grundlagen und Techniken. 8. Auflage. Weinheim und Basel: Beltz 2003.

- McNamara, Tim: Language testing. Oxford: Oxford University Press 1990. (Oxford Introductions to Language Study)
- Messick, Samuel: Validity and washback in language testing. In: Language Testing 13/3 (1996), S. 241-256.
- Mitchell, Rosamond: The communicative approach to language teaching: an introduction. In: Swarbrick, Ann (Hg.): Teaching modern languages. London: Routledge 1994, S. 33-42.
- Mummendey, Hans Dieter / Grau, Ina: Die Fragebogen-Methode. Fünfte, überarbeitete und erweiterte Auflage. Göttingen u.a.: Hogrefe 2008.
- Obermaier, Alexandra: Österreichische FremdsprachenassistentInnen in England und Wales. Praxis, Ziele und Perspektiven eines Austauschprogramms – eine Evaluation. Diplomarbeit. Universität Wien, 1999.
- Pachler, Norbert: Foreign language learning in England in the 21st century. Language Learning Journal 25 (2002), S. 4-7  
(<http://www.itmfl.org.uk/modules/policy/4d/paper4d3.pdf>)
- Prantl, Christoph: Zur Rolle der Authentizität in kommunikativen Tests. Eine Untersuchung am Beispiel des Österreichischen Sprachdiploms. Dissertation. Universität Wien, 2008.
- Preußler, Walburga: Grundbegriffe der klassischen Testtheorie In: Gardenghi, Monica / O'Connell, Mary (Hg.): Prüfen, Testen, Bewerten im modernen Fremdsprachenunterricht. Frankfurt am Main: Peter Lang 1997, S. 11-16.
- Purpura, James: The impact of large-scale and classroom-based language assessments on the individual. In: Taylor, Linda / Weir, Cyril (Hg.): Language testing matters: investigating the wider social and educational impact of assessment. Proceedings of the ALTE Cambridge Conference April 2008. Cambridge: Cambridge University Press 2009. (Studies in language testing 31), S. 310-325.
- Putwain, David William: Assessment and examination stress in Key Stage 4. In: British Educational Research Journal 35/3 (2009), S. 391-411.
- Rauscher, Gabriele: Qualitätssicherung bei Sprachtests im Bereich Deutsch als Fremdsprache. Diplomarbeit. Universität Wien, 2001.
- Rea-Dickins, Pauline / Scott, Catriona: Washback from language tests on teaching, learning and policy: evidence from diverse settings. In: Assessment in Education 1/14 (2007), S. 1-7.
- Reershemius, Gertrud: In: Krumm, Hans-Jürgen / Fandrych, Christian / Hufeisen, Britta / Riemer, Claudia (Hg.): Deutsch als Fremd- und Zweitsprache. Ein internationales Handbuch. Zweiter Halbband. Berlin / New York: De Gruyter 2010, S. 1674-1680.
- Reicher, Hannelore: Die Planung eines Forschungsprojektes – Überlegungen zur Methodenauswahl. In: Stigler, Hubert / Reicher, Hannelore (Hg.): Praxisbuch Empirische Sozialforschung in den Erziehungs- und Bildungswissenschaften. Innsbruck: StudienVerlag Innsbruck 2005, S. 85 – 104.
- Reinhoffer, Bernd. Lehrkräfte geben Auskunft über ihren Unterricht. Ein systematisierender Vorschlag zur deduktiven und induktiven Kategorienbildung in der Unterrichtsforschung. In: Mayring, Philip / Gläser-Zikuda, Michaela (Hg.): Die Praxis der Qualitativen Inhaltsanalyse. Weinheim und Basel: Beltz 2005.



- Rogers, Lynne / Hallam, Susan: Changes in perception of studying for the GCSE among Year 10 and Year 11 pupils. In: *Educational Studies* 36/2 (2010), S. 153-163.
- Sacher, Werner: Leistungen entwickeln, überprüfen und beurteilen. Vierte, überarbeitete und erweiterte Auflage. Bad Heilbrunn: Verlag Julius Klinkhardt 2004.
- Schifko, Manfred: Prüfungen, Zertifikate, Abschlüsse als Planungskategorien für den Unterricht. In: Helbig, Gerhard / Götze, Lutz / Henrici, Gert / Krumm, Hans-Jürgen (Hg.): *Deutsch als Fremdsprache. Ein internationales Handbuch. Zweiter Halbband.* Berlin / New York: De Gruyter 2001, S. 827-834.
- Shohamy, Elana / Donitsa-Schmidt, Smadar / Ferman, Irit: Test impact revisited: washback effect over time. In: *Language Testing* 13/3 (1996), S. 298-317.
- Skehan, Peter: Progress in language testing - the 1990s. In: Alderson, J. Charles / North, Brian (Hg.): *Language testing in the 1990s: the communicative Legacy.* London / Basingstoke: Macmillan 1991, S. 3-21.
- Spolsky, Bernard: The limits of authenticity in language testing. In: *Language Testing* 2/1 (1985), S. 31-40.
- Tenberg, Reinhard / Jones, Roger: *German studies in the United Kingdom. A survey of German in schools and universities.* Cambridge: European Business Associates 1993.
- Wall, Dianne: *The impact of high-stakes examinations on classroom teaching. A case study using insights from testing and innovation theory.* Cambridge: Cambridge University Press 2005. (Studies in language testing 22)
- Watanabe, Yoshinori: Methodology in Washback Studies. In: Cheng, Liying / Watanabe, Yoshinori / Curtis, Andy. (Hg.): *Washback in language testing. Research contexts and methods.* New Jersey: Lawrence Erlbaum Associates 2004, S. 19-36.
- Wegner, Anke: *100 Jahre Deutsch als Fremdsprache in Frankreich und England . Eine vergleichende Studie von Methoden, Inhalten und Zielen.* München: Iudicium 1999.
- Weir, Cyril J.: *Communicative Language Testing.* New York: Prentice Hall 1990.
- Werlen, Erika: Kommunikative Kompetenz und Mehrsprachigkeit. Zwei Seiten einer Medaille. In: Werlen, Erika (Hg.): *Kommunikative Kompetenz und Mehrsprachigkeit . Diskussionsgrundlagen und unterrichtspraktische Aspekte.* Baltmannsweiler: Schneider-Verlag Hohengehren 2007, S. 9-28.
- Williams, Kevin: *Why teach foreign languages in school? A philosophical response to curriculum policy.* London: Philosophy of Education Society of Great Britain 2000. (Impact 5)
- Wilson, Rosemary: Foreign languages: a guide for the inhibited. In: *Language Learning Journal* 36/1 (2008), S. 111-115.

## 6.2 Internetquellen

AQA. Controlled Assessment Guidance. GCSE Modern Foreign Languages. 2010a. <http://store.aqa.org.uk/qual/newgcse/pdf/AQA-4645-4655-4665-W-TRB-CAG-2011.PDF> (3. 3. 2011)

AQA. GCSE Chinese (Mandarin) / French / German / Italian / Spanish / Urdu Controlled Assessment Handbook. Speaking and Writing. 2010b. <http://store.aqa.org.uk/qual/newgcse/pdf/AQA-MFL-W-TRB-CAHA10.PDF>. (11. 10. 2010.)

AQA. GCSE Specification German. Full Course – for exams June 2010 onwards and certification June 2011 onwards. Short Course – for exams June 2010 onwards and certification June 2010 onwards. 2010c. <http://store.aqa.org.uk/qual/newgcse/pdf/AQA-4665-W-SP.PDF> (2. 3. 2011)

AQA. New GCSE Chinese (Mandarin) / French / German / Italian / Spanish / Urdu. Frequently Asked Questions Controlled Assessment Speaking & Writing. 2010d. <http://store.aqa.org.uk/qual/newgcse/pdf/AQA-MFL-W-TRB-FAQCASAUT10.PDF> (3. 3. 2011).

DCSF. 14-19 Qualification Strategy. 2010. <http://www.dcsf.gov.uk/14-19/index.cfm?go=site.home&sid=3&pid=529&ctype=None&ptype=Contents>. (30. 9. 2010)

DCSF. Publication of 2010 test and examination results in the school and college achievement and attainment tables 2010b. <http://www.education.gov.uk/performance/tables/Statement-of-Intent-Final-2010-April.pdf> (21. 3. 2011)

DfES. Languages for All: Languages for Life. A Strategy for England. 2002. <http://www.education.gov.uk/publications/eOrderingDownload/DfESLanguagesStrategy.pdf>

Edexcel. Specification. Edexcel GCSE in German (1231). 2000. [http://www.edexcel.com/migrationdocuments/Current%20GCSE/17352\\_uk\\_qualifications\\_gcse\\_mfl\\_gcse\\_1231\\_17352.PDF](http://www.edexcel.com/migrationdocuments/Current%20GCSE/17352_uk_qualifications_gcse_mfl_gcse_1231_17352.PDF) (3. 3. 2011)

Edexcel. GCSE German. Coursework guide. 2001. [http://www.edexcel.com/migrationdocuments/Current%20GCSE/139230\\_uk\\_qualifications\\_gcse\\_mfl\\_gcse\\_1231\\_17358.PDF](http://www.edexcel.com/migrationdocuments/Current%20GCSE/139230_uk_qualifications_gcse_mfl_gcse_1231_17358.PDF)

Edexcel. GCSE German. Paper 2F – Foundation Oral and Paper 2H – Higher Oral. 2009a. [http://www.edexcel.com/migrationdocuments/QP%20Current%20GCSE/June%202009/n33666ra\\_German.pdf](http://www.edexcel.com/migrationdocuments/QP%20Current%20GCSE/June%202009/n33666ra_German.pdf) (3. 3. 2011)

Edexcel. GCSE Modern Foreign Languages Controlled Assessment. Teacher Support Book – Speaking. 2009b. <http://www.edexcel.com/migrationdocuments/CPS/GCSE%20from%202009/GCSE%20MFL%20Controlled%20Assessment%20Teacher%20Support%20Book%20%20-%20Speaking.pdf> (3. 3. 2011)

Edexcel. Deutsch. Specification. 2010. <http://www.edexcel.com/migrationdocuments/GCSE%20New%20GCSE/GCSE-German-spec-Issue-3-UG025113.pdf>. (4. 10. 2010)

- Edexcel. Qualifications Home. 2011.  
<http://www.edexcel.com/quals/Pages/default.aspx>
- European Commission. Europeans and Languages. 2005.  
[http://ec.europa.eu/public\\_opinion/archives/ebs/ebs\\_237.en.pdf](http://ec.europa.eu/public_opinion/archives/ebs/ebs_237.en.pdf) (25.2.2010)
- European Commission. Key Data on Teaching Languages at School in Europe. 2008  
[http://eacea.ec.europa.eu/about/eurydice/documents/KDL2008\\_EN.pdf](http://eacea.ec.europa.eu/about/eurydice/documents/KDL2008_EN.pdf) (25.2.2010)
- OCR. GCSE. French/ German / Spanish. 2009.  
[http://www.ocr.org.uk/download/kd/ocr\\_9962\\_kd\\_spec.pdf](http://www.ocr.org.uk/download/kd/ocr_9962_kd_spec.pdf) (3. 3. 2011)
- Ofqual. GCSEs. The official student guide to the system.  
[http://www.ofqual.gov.uk/files/GCSE\\_Guide.pdf](http://www.ofqual.gov.uk/files/GCSE_Guide.pdf). (28. 8. 2010)
- Ofqual. Predictability studies report. A study of GCSE and GCE level examinations. 2008. [http://www.ofqual.gov.uk/files/Predictability\\_Studies\\_final\\_report.pdf](http://www.ofqual.gov.uk/files/Predictability_Studies_final_report.pdf) (15.3.2010)
- Ofqual. Who we are. 2010b. <http://www.ofqual.gov.uk/who-we-are>. (28. 9. 2010)
- Ofqual. What is Ofqual's Reliability Programme?. 2010a.  
<http://www.ofqual.gov.uk/research-and-statistics/research-reports/92-articles/20-reliability> (3.3. 2011)
- Ofsted. The changing landscape of languages. 2008.  
[http://www.ofsted.gov.uk/content/download/5958/53674/file/The%20changing%20landscape%20of%20languages%20\(PDF%20format\).pdf](http://www.ofsted.gov.uk/content/download/5958/53674/file/The%20changing%20landscape%20of%20languages%20(PDF%20format).pdf)
- QCA. GCSE subject criteria for MFL. 2007. [http://www.ofqual.gov.uk/files/qca-07-3468\\_gcsecriteriamfl.pdf](http://www.ofqual.gov.uk/files/qca-07-3468_gcsecriteriamfl.pdf). (28. 9. 2010)
- QCA. Modern foreign languages. Programme of study for key stage 3 and attainment targets. 2007. [http://curriculum.qcda.gov.uk/uploads/QCA-07-3340-p\\_MFL\\_KS3\\_tcm8-405.pdf](http://curriculum.qcda.gov.uk/uploads/QCA-07-3340-p_MFL_KS3_tcm8-405.pdf). (30. 9. 2010)
- QCDA. The new secondary curriculum. What has changed and why? 2007.  
[http://curriculum.qcda.gov.uk/uploads/overview\\_doc\\_tcm8-1839.pdf](http://curriculum.qcda.gov.uk/uploads/overview_doc_tcm8-1839.pdf). (15. 9. 2010)
- QCDA. Introducing the Qualifications and Curriculum Development Agency. 2009.  
[http://orderline.qcda.gov.uk/gempdf/1445900645/qcda\\_introduction\\_leaflet.pdf](http://orderline.qcda.gov.uk/gempdf/1445900645/qcda_introduction_leaflet.pdf). (30. 9. 2010)
- QCDA. About GCSEs. 2010. <http://www.qcda.gov.uk/qualifications/gcses/68.aspx>. (28. 9. 2010)
- QDCA. GCSE awarding organisations. 2010b.  
<http://www.qcda.gov.uk/qualifications/gcses/572.aspx>
- QDCA. National curriculum. Subjects. Key Stage 3 & 4. 2011.  
<http://curriculum.qcda.gov.uk/key-stages-3-and-4/subjects/index.aspx> ( 3. 3. 2011)
- TES Connect. Forums / Subjects / Modern Foreign Languages.  
<http://community.tes.co.uk/forums/28.aspx> (21. 3. 2011)
- The John Warner School. 2010. <http://www.johnwarner.org.uk>. (15. 3. 2011)
- University of Cambridge. Undergraduate Admissions. AS/A levels. 2009.  
<http://www.cam.ac.uk/admissions/undergraduate/apply/asalevels.html> (7. 3. 2010)

### 6.3 Onlinezeitungen

Lipsett, Anthea [14. 1. 2008]: Unions make case for scrapping league tables. <http://www.guardian.co.uk/education/2008/jan/14/schools.uk1?INTCMP=SRCH> (18. 3. 2011). (The Guardian, Onlineausgabe)

Mansell, Warwick [12. 11. 2010]: The untruth about Wales and league tables. <http://www.tes.co.uk/article.aspx?storycode=6063124> (18. 3. 2011). (TES Connect)

Mansell, Warwick [21.4.2009]: Inside out. Examiners offering guidance in how to do well in exams. <http://www.guardian.co.uk/education/2009/apr/21/gcses-alevels-sats> (18. 3. 2011). (The Guardian, Onlineausgabe)

Paton, Graeme [12. 1. 2011]. GCSE league tables: more than 200 schools 'face closure'. <http://www.telegraph.co.uk/education/educationnews/8254817/GCSE-league-tables-more-than-200-schools-face-closure.html> (21. 3. 2011). (The Telegraph, Onlineausgabe)

Paton, Graeme [16. 9. 2009]: Scrap school league tables, say experts. <http://www.telegraph.co.uk/education/6193787/Scrap-school-league-tables-say-experts.html> (18. 3. 2011). (The Telegraph, Onlineausgabe)

Paton, Graeme [21. 7. 2008]. Ofsted: Schools 'teaching to the test'. <http://www.telegraph.co.uk/education/2440091/Ofsted-Schools-teaching-to-the-test.html> (21. 3. 2011). (The Telegraph, Onlineausgabe)

Simon Burgess, Simon; Wilson, Deborah; Worth, Jack: A natural experiment in school accountability: the impact of school performance information on pupil progress and sorting. Bristol: Centre for Market and Public Organisation 2010. <http://www.bristol.ac.uk/cmpo/publications/papers/2010/wp246.pdf>

## 7 Anhang

### 7.1 Bildung des Kategoriensystems

#### 7.1.1 Einschätzung und Bewertung der Veränderungen

Coursework (CW) = unfair, schummeln

Role Plays (RP) = nicht spontan / interaktiv / authentisch

Controlled Assessment (CA) = lange Vorbereitungszeit, nicht spontan, nicht unvorhersehbar, nicht authentisch, begünstigt auswendiglernen

Themenfokus = begünstigt auswendig lernen, zeigt tatsächliche Fähigkeiten nicht (nicht valide), nicht repräsentativ, nicht unvorhersehbar

CA = kürzere TTTT-Phasen, weniger zeitintensiv, Themen die nicht geprüft werden, mehr Zeit für kreativen Unterricht, mehr Motivation

Grammatik vs. Spontaneität in der CA-Bewertung: Wenig Transparenz / Vertrautheit mit Anforderungen

CA: weniger spontane Sprachverwendung

#### 7.1.2 Hypothesen

1) League Tables = (Noten-)druck, allgemeiner Backwash

2) Neues GCSE / CA = stärkerer / negativerer Backwash

3) CA = begünstigt auswendig lernen

4) CA = weniger valide

5) CA = intensiveres TTTT

6) CA = mehr Zeit für kreativen Unterricht: Themen die nicht geprüft werden, Prüfungsaufgaben selbst erstellen

Unterstrichen = Überschneidung von deduktiven und induktiven Kategorien

#### 7.1.3 Deduktive Kategorien

CW / CA / RW

Fairness / schummeln

Spontaneität

Interaktivität

Authentizität (Rest)

Vorhersehbarkeit (Rest)

Repräsentativität (Themenfokus)

begünstigt

Auswendiglernen

Lange Vorbereitungszeit (Rest)

Validität: zeigt tatsächliche Fähigkeiten (nicht)

TTTTZeitintensität

Themen die nicht geprüft werden (Rest)

Freier / kreativer Unterricht (Rest)

Motivation vs. Langeweile (Rest)

Transparenz / Vertrautheit mit Anforderungen (Rest)

League Tables (Rest)

(Noten-)druck

Backwash: Spezifität, Stärke, Ausrichtung (Rest)

**7.1.4 Induktive Kategorien**FT / CW / CA / RP

gute / schlechte SchülerInnen

Theoretisch / praktisch

Wird mehr gefördert als / wird belohnt / ist wichtiger als / begünstigt

Vorhandene / nicht vorhandene / mehr / weniger Zeit zum Lernen

Mehr / weniger / wichtig / unwichtig / Grammatik/Sprachreflexion

Zeitintensität (Administration, Unterrichtszeit)

Zeigt tatsächliche Fähigkeiten (nicht)

Fairness / Schummeln

Auswendig lernen

bessere/ schlechtere Noten

mehr / weniger / viel / wenig Druck / Stress (für SchülerInnen / durch League Tables)

vorhandene / fehlende Themenvielfalt, Themenfokus

Auswirkungen auf KS3/6th form

Spontane Sprachverwendung / Echte Konversation / „transactional“ / „functional“

TTTT = intensive Testvorbereitung, Vorbereitungszeit, vorgeschriebene Texte

Risiken eingehen (Rest)

Starke Veränderung (Rest)

Überprüft Rezeption (Rest)

Grammatik/Sprachreflexion zum Angeben (Rest)

Integrativität (Rest)

Balance der Skills erhalten (Rest)

Alles Englisch (Rest)

Keine Korrektur (Rest)

Regt zum Lernen an (für die Prüfung / für das Leben) (Rest)

Keine Sprachkenntnisse (Rest, Teil von Grammatik/Sprachreflexion, Spontaneität)

Authentizität (Rest)

### 7.1.5 Das Kategoriensystem

Coursework	CW
Controlled Assessment	CA
Role Plays	RP
Final Test	FT
gute / schlechte SchülerInnen	+S, -S
theoretisch / praktisch	theoretisch, praktisch
wird als +/- eingeschätzt	= pos, = neg
gut, mehr, stark, vorhanden	+
schlecht, weniger, schwach, nicht vorhanden	-
begünstigt, führt zu	→
Fairness	fair
Schummeln	schummeln
Spontaneität / echte Konversation	Spontaneität
Auswendiglernen	auswendig
Zeigt tatsächliche Fähigkeiten	ZTF
Teaching to the Test	TTTT
Zeitintensität	Zeitintensität
Zeit zum Lernen bzw. Lehren	ZZL
Grammatik/Sprachreflexion	GR
bessere / schlechtere Noten	Noten
Druck / Stress	Druck
Themenvielfalt	Themenvielfalt
Auswirkungen auf KS3/6th form	KS3, 6THF
Restkategorie	R
Keine Angabe	kA

## 7.2 Fragebogen

### 1. Which subjects do you teach?

<input type="checkbox"/>	German	<input type="checkbox"/>	ICT
<input type="checkbox"/>	other MFL	<input type="checkbox"/>	Math
<input type="checkbox"/>	Art & Design	<input type="checkbox"/>	Music
<input type="checkbox"/>	Citizenship	<input type="checkbox"/>	Personal, social, health and economic wellbeing
<input type="checkbox"/>	Design & Technology	<input type="checkbox"/>	Physical education
<input type="checkbox"/>	English	<input type="checkbox"/>	Religious education
<input type="checkbox"/>	Geography	<input type="checkbox"/>	Science
<input type="checkbox"/>	History	<input type="checkbox"/>	Other: <input type="text"/>

### 2. What type(s) of exam did you take as a student?

<input type="checkbox"/>	CSE
<input type="checkbox"/>	GCE O-Levels
<input type="checkbox"/>	GCSE
<input type="checkbox"/>	A-Levels
<input type="checkbox"/>	Other: <input type="text"/>

### 3. In what year did you start teaching (approximately)?

### 4. When did you start teaching GCSE (or CSE / O-Level) courses? Please give the year approximately.

### 5. In what kind of area is your school?

<input type="checkbox"/>	urban
<input type="checkbox"/>	rural
<input type="checkbox"/>	rather low income
<input type="checkbox"/>	rather high income
<input type="checkbox"/>	mid income



**6. Have you already done controlled assessment?**

<input type="checkbox"/>	no
<input type="checkbox"/>	I am currently doing it for the first time
<input type="checkbox"/>	yes

**7. In the old GCSE, did you prefer coursework or the final written exam?**

<input type="checkbox"/>	coursework
<input type="checkbox"/>	final written exam
<input type="checkbox"/>	no preference for either

**8. Why? Please give at least one reason.**

--

**9. How do you see... ?**

	very negatively	rather negatively	rather positively	very positively	no opinion
the abolishment of coursework	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
the abolishment of the final written exam	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
the abolishment of role plays	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
the introduction of controlled assessment (writing)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
the introduction of controlled assessment (speaking)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
the possibility to focus on one specific topic for writing and speaking	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
the change in short courses (only written or only spoken language)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**10. Please explain your attitude towards the changes listed above briefly.**

--

**11. How do you think the possibility to focus on one specific topic for writing and speaking will affect...**

	negatively	positively	don't know
the students' overall German proficiency	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
the breadth of topics German teachers will cover	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
the students' motivation	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**12. There were hardly any changes to the listening and reading part of the GCSE. How do you see that? Would you have wanted any more changes?**

<input type="checkbox"/>	I would not have wanted any more changes
<input type="checkbox"/>	I would have wanted the following change(s):

**13. How does the introduction of controlled assessment affect rote learning as exam preparation? Please explain briefly.**

--

**14. How do you think controlled assessment will affect the time available for content not treated in the exam?**

<input type="checkbox"/>	there will be considerably less time
<input type="checkbox"/>	there will be a bit less time
<input type="checkbox"/>	it will remain the same
<input type="checkbox"/>	there will be a bit more time
<input type="checkbox"/>	there will be considerably more time

**15. How will the introduction of controlled assessment affect the students' overall German proficiency?**

<input type="checkbox"/>	very negatively
<input type="checkbox"/>	a bit negatively
<input type="checkbox"/>	no changes
<input type="checkbox"/>	a bit positively
<input type="checkbox"/>	very positively

**16. Do you think that controlled assessment says more or less about students' actual abilities than the previous test types?**

<input type="checkbox"/>	rather more
<input type="checkbox"/>	rather less
<input type="checkbox"/>	the same
<input type="checkbox"/>	don't know

**17. Please explain the choice of your answer above shortly:**

--

**18 (Version A<sup>179</sup>). You are currently doing controlled assessment. Which skills are you testing?**

<input type="checkbox"/>	Writing
<input type="checkbox"/>	Speaking
<input type="checkbox"/>	Both

**18 (Version B). You have already done controlled assessment. Which skills did you test?**

<input type="checkbox"/>	writing
<input type="checkbox"/>	speaking
<input type="checkbox"/>	both

---

<sup>179</sup> Frage 6 ist eine Filterfrage. Je nachdem, ob LehrerInnen bereits Controlled Assessment durchgeführt haben, erhielten sie andere Fragen. Version A: führt derzeit CA durch, Version B: hat bereits CA durchgeführt

**19 (Version B). How often have you done controlled assessment?**

--

**19 (Version A). Which strategies are your students using to prepare for the assessment?**

	none of them	some	many	(nearly) all of them
they are writing a list of bullet points	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
they practise with their peers	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
they write a full draft (written exam)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
they write a script (oral exam)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
they study vocabulary	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
they study grammar	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
they look for similar texts on the internet	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
they write notes / some paragraphs	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
they record themselves	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
they rote learn / memorise	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
they practise to become fluent	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
they copy sentences from various sources with slight changes	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
they try to formulate in their own way	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**20 (Version A). Do you think that the results will reflect your students' actual communicative language competence?**

<input type="checkbox"/>	they will probably correspond very much
<input type="checkbox"/>	they will probably correspond well enough
<input type="checkbox"/>	they probably won't really correspond
<input type="checkbox"/>	I can't judge

**20 (Version B). How much do the results reflect your students' actual communicative language competence?**

<input type="checkbox"/>	they correspond very much
<input type="checkbox"/>	they correspond well enough
<input type="checkbox"/>	they don't really correspond
<input type="checkbox"/>	I can't judge
<input type="checkbox"/>	I haven't got any results yet

**21 (Version B). Which strategies did the students use to prepare for the assessment?**

	none of them	some	many	(nearly) all of them
they wrote a list of bullet points	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
they practised with their peers	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
they wrote a full draft (written exam)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
they wrote a script (oral exam)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
they studied vocabulary	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
they studied grammar	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
they looked for similar texts on the internet	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
they wrote notes / some paragraphs	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
they recorded themselves	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
they rote learned / memorised	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
they practised to become fluent	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
they copied sentences from various sources with slight changes	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
they tried to formulate in their own way	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**21 (A) / 22 (B). Is there any additional important strategy which I did not list?**

<input type="checkbox"/>	no
<input type="checkbox"/>	yes, the following: <input type="text"/>

**22 (A) / 23 (B). How strongly did the old GCSE affect your teaching?**

<input type="checkbox"/>	no impact
<input type="checkbox"/>	weak impact
<input type="checkbox"/>	mid impact
<input type="checkbox"/>	strong impact

**23 (A) / 24(B). Was the impact of the old GCSE on your teaching rather positive or negative?**

<input type="checkbox"/>	very negative
<input type="checkbox"/>	rather negative
<input type="checkbox"/>	rather positive
<input type="checkbox"/>	very positive

**24 (A) / 25 (B). How strongly does the new GCSE affect your teaching?**

<input type="checkbox"/>	no impact
<input type="checkbox"/>	weak impact
<input type="checkbox"/>	mid impact
<input type="checkbox"/>	strong impact

**25 (A) / 26 (B). Is the impact of the new GCSE on your teaching rather positive or negative?**

<input type="checkbox"/>	very negative
<input type="checkbox"/>	rather negative
<input type="checkbox"/>	rather positive
<input type="checkbox"/>	very positive

**26 (A) / 27 (B). If the impact of the GCSE on your teaching changed in any way, please explain this briefly: what exactly changed and why?**

--

**27 (A) / 28 (B). Do you feel that you and your students know enough about the changes to the GCSE?**

<input type="checkbox"/>	yes
<input type="checkbox"/>	no

**28 (A) / 29 (B). Would teaching be different if there were no nation-wide exams at all?**

<input type="checkbox"/>	no
<input type="checkbox"/>	yes, a bit different
<input type="checkbox"/>	yes, quite different
<input type="checkbox"/>	yes, completely different

**29 (A) / 30 (B). How time-consuming was the preparation for the GCSE previously?**

<input type="checkbox"/>	not time-consuming
<input type="checkbox"/>	a bit time-consuming
<input type="checkbox"/>	rather time-consuming
<input type="checkbox"/>	very time-consuming


**30 (A) / 31 (B). How time-consuming is the preparation for the GCSE now? This refers to both controlled assessment and the final exams.**

<input type="checkbox"/>	not time-consuming
<input type="checkbox"/>	a bit time-consuming
<input type="checkbox"/>	rather time-consuming
<input type="checkbox"/>	very time-consuming

**31 (A) / 32 (B). How much room do you now have for personalised and creative teaching (in comparison to the previous GCSE)?**

<input type="checkbox"/>	much less
<input type="checkbox"/>	a bit less
<input type="checkbox"/>	about the same
<input type="checkbox"/>	a bit more
<input type="checkbox"/>	much more

**32 (A) / 33 (B). Is it now more or less motivating to prepare for the GCSE?**

	much less	a bit less	the same	a bit more	much more
					
for you	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
for your students	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**33 (A) / 34 (B). Was the teaching of GCSE classes different two or more years ago?**

<input type="checkbox"/>	it hardly changed
<input type="checkbox"/>	some things changed
<input type="checkbox"/>	it changed a lot

**34 (A) / 35 (B). Please explain, what was different and how it changed.**

--

**35 (A) / 36 (B). Do you feel pressure to perform well in league tables?**

<input type="checkbox"/>	yes
<input type="checkbox"/>	no

**36 (A) / 37 (B). If so, how strongly does it affect teaching?**

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
no impact	weak impact	mid impact	strong impact	very strong impact

**37 (A) / 38 (B). Is the impact of league tables rather positive or negative?**

<input type="checkbox"/>	very negative
<input type="checkbox"/>	rather negative
<input type="checkbox"/>	rather positive
<input type="checkbox"/>	very positive

**38 (A) / 39 (B). (How) did the amount of pressure change through the revision of the GCSE?**

<input type="checkbox"/>	the pressure became stronger
<input type="checkbox"/>	the pressure became weaker
<input type="checkbox"/>	the pressure did not change

**39 (A) / 40 (B). Would teaching be different if there were no league tables?**

<input type="checkbox"/>	No
<input type="checkbox"/>	yes, a bit different
<input type="checkbox"/>	yes, rather different
<input type="checkbox"/>	yes, completely different



**40 (A) / 41 (B). Would this difference be positive or negative?**

<input type="checkbox"/>	rather negative
<input type="checkbox"/>	just different: neither positive nor negative
<input type="checkbox"/>	rather positive

**41 (A) / 42 (B). If you had the choice, how would you design the German GCSE?**

	no	rather no	rather yes	yes	no answer
controlled assessment	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Coursework	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
terminal examinations	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
external examination	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
internal examination	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
personalised tasks	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
standardised tasks	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
focus on specific topics or skills	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
breadth of topics and/or skills	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
possibility to prepare a list of questions and answers	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
unpredictable exam questions	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**You have just answered questions about**

- your attitude towards the changes to the GCSE
- what you expect from controlled assessment
- the strength of impact of the GCSE
- how positive or negative that impact is
- the influence of league tables

**42 (A) / 43 (B). I very much appreciate any additional comments, as these will help to approach the issue of backwash from different perspectives. Whether you want to explain one aspect in further detail, give examples from your experience in school or simply want to describe your personal views: please feel free to use this text box.**

**44. One last question: where did you hear of this survey?**

<input type="checkbox"/>	foreign language assistant
<input type="checkbox"/>	other teachers
<input type="checkbox"/>	forum / website (please specify which one) <input type="text"/>
<input type="checkbox"/>	other (please specify): <input type="text"/>

## 7.3 Ergebnisse – quantitative Auswertung

### 7.3.1 In der Arbeit verwendete Tabellen

Antworten auf die Frage: "Would teaching be different if there were no nation-wide exams at all?" (= Tabelle 5)

Antwortmöglichkeit	Anzahl
1 = no	1
2 = yes, a bit different	5
3 = yes, quite different	11
4 = yes, completely different	18
nicht beantwortet	0

Antworten auf die Frage: "Would teaching be different if there were no league tables?" (= Tabelle 6)

Antwortmöglichkeit	Anzahl
1 = no	4
2 = yes, a bit different	8
3 = yes, rather different	13
4 = yes, completely different	10
nicht beantwortet	0

Antworten auf die Frage: "Would this difference be rather positive or negative?" (= Tabelle 7)

Antwortmöglichkeit	Anzahl
1 = rather negative	0
2 = just different: neither positive nor negative	13
3 = rather positive	21
nicht beantwortet	1

Antworten auf die Frage: "How strongly did the old GCSE affect your teaching?" (= Tabelle 8)

Antwortmöglichkeit	Anzahl
1 = no impact	4
2 = weak impact	1
3 = mid impact	19
4 = strong impact	10
nicht beantwortet	1

Antworten auf die Frage: "How strongly does the new GCSE affect your teaching?" (= Tabelle 9)

Antwortmöglichkeit	Anzahl
1 = no impact	0
2 = weak impact	4
3 = mid impact	12
4 = strong impact	18
nicht beantwortet	1

**Antworten auf die Frage: "Was the impact of the old GCSE on your teaching rather positive or negative?" (= Tabelle 10)**

Antwortmöglichkeit	Anzahl
1 = very negative	0
2 = rather negative	11
3 = rather positive	19
4 = very positive	2
nicht beantwortet	3

**Antworten auf die Frage: "Is the impact of the new GCSE on your teaching rather positive or negative?" (= Tabelle 11)**

Antwortmöglichkeit	Anzahl
1 = very negative	15
2 = rather negative	16
3 = rather positive	12
4 = very positive	0
nicht beantwortet	2

**Antworten auf die Frage: "Which strategies are your students using / did the students use to prepare for the assessment?" (= Tabelle 12)**

	writing a list of bullet points	practising with their peers	writing a full draft (written exam)	writing a script (oral exam)	studying vocabulary	studying grammar		
1 = none of them	1	3	0	1	2	5		
2 = some	13	6	5	5	13	14		
3 = many	5	12	3	3	9	7		
4 = (nearly) all of them	14	14	26	24	11	9		
nicht beantwortet	2	0	1	2	0	0		
	looking for similar texts on the internet	writing notes / some paragraphs	recording themselves	rote learning / memorising	practising to become fluent	copying sentences from various sources with slight changes	trying to formulate in their own way	
1 = none of them	23	3	6	1	2	2	1	
2 = some	8	10	18	5	12	5	22	
3 = many	1	10	8	6	9	20	8	
4 = (nearly) all of them	1	12	1	23	7	8	4	
nicht beantwortet	2	0	2	0	5	0	0	

Antworten auf die Frage: "Do you think that Controlled Assessment says more or less about students' actual abilities than the previous test types?" (= Tabelle 13)

Antwortmöglichkeit	Anzahl
1 = rather more	5
2 = rather less	20
3 = the same	4
4 = don't know	6
nicht beantwortet	0

Antworten auf die Frage: "Do / will the results reflect your students' actual communicative competence?" (= Tabelle 14)

Antwortmöglichkeit	Anzahl
1 = they (will) correspond very much	2
2 = they (will) correspond well enough	12
3 = they won't/ don't really correspond	15
4 = I can't judge	3
5 = I haven't got any results yet	3
nicht beantwortet	

Antworten auf die Frage: "How will the introduction of controlled assessment affect the students' overall German proficiency?" (= Tabelle 15)

Antwortmöglichkeit	Anzahl
1 = very negatively	7
2 = a bit negatively	19
3 = no changes	6
4 = a bit positively	0
5 = very positively	1
nicht beantwortet	2

Antworten auf die Frage: "How much room do you now have for personalised and creative teaching (in comparison to the previous GCSE)?" (= Tabelle 16)

Antwortmöglichkeit	Anzahl
1 = much less	15
2 = a bit less	7
3 = about the same	10
4 = a bit more	2
5 = much more	1
nicht beantwortet	0

Antworten auf die Frage: "How do you think controlled assessment will affect the time available for content not treated in the exam?" (= Tabelle 17)

Antwortmöglichkeit	Anzahl
1 = there will be considerably less time	23
2 = there will be a bit less time	6
3 = it will remain the same	4
4 = there will be a bit more time	1
5 = there will be considerably more time	0
nicht beantwortet	1

**Antworten auf die Frage: "How time-consuming was the preparation for the GCSE previously?"  
(= Tabelle 18)**

<b>Antwortmöglichkeit</b>	<b>Anzahl</b>
1 = not time-consuming	1
2 = a bit time-consuming	13
3 = rather time-consuming	15
4 = very time-consuming	5
nicht beantwortet	1

**Antworten auf die Frage: „How time-consuming was the preparation for the GCSE previously?“  
(= Tabelle 19)**

<b>Antwortmöglichkeit</b>	<b>Anzahl</b>
1 = not time-consuming	0
2 = a bit time-consuming	1
3 = rather time-consuming	10
4 = very time-consuming	24
nicht beantwortet	0

### 7.3.2 Weitere Ergebnisse

Antworten auf die Frage: „Have you already done controlled assessment?“ (= Tabelle 20)

Antwortmöglichkeit	Anzahl
No	0
I am currently doing it for the first time	15
Yes	20

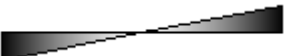
Antworten auf die Frage: „In the old GCSE, did you prefer coursework or the final written exam?“ (= Tabelle 21)

Antwortmöglichkeit	Anzahl
Coursework	11
Final written exam	20
no preference for either	4

Antworten auf die Frage: “How do you see...?“ (= Tabelle 22)

	the abolishment of coursework	the abolishment of the final written exam	the abolishment of role plays	the introduction of controlled assessment (writing)	the introduction of controlled assessment (speaking)	the possibility to focus on one specific topic for writing and speaking	the change in short courses (only written or only spoken language)
nichts ausgewählt	2	0	1	0	0	1	0
don't know	7	7	4	2	2	1	23
very negatively	0	12	4	8	8	6	4
rather negatively	5	13	16	15	13	11	2
rather positively	11	2	5	10	8	13	3
very positively	10	1	5	0	4	3	3

Antworten auf die Frage: “How do you think the possibility to focus on one specific topic for writing and speaking will affect...?“ (= Tabelle 23)

	Negatively				positively	don't know	nichts ausgewählt
the students' overall German proficiency	15	7	8	1	1	2	1
the breadth of topics German teachers will cover	11	10	10	2	0	1	1
the students' motivation	4	4	14	8	2	2	1

**Antworten auf die Frage: “Do the results reflect your students’ actual communicative competence?” (= Tabelle 24)**

Antwortmöglichkeit	Anzahl
1 = they correspond very much	2
2 = they correspond well enough	6
3 = they don't really correspond	7
4 = I can't judge	2
5 = I haven't got any results yet	3
nicht beantwortet	0

**Antworten auf die Frage: Will the results reflect your students’ actual communicative competence?” (= Tabelle 25)**

Antwortmöglichkeit	Anzahl
1 = they will correspond very much	0
2 = they will correspond well enough	6
3 = they won't really correspond	8
4 = I can't judge	1
nicht beantwortet	0

**Antworten auf die Frage: “Which strategies did your students use to prepare for the assessment?” (= Tabelle 26)**

	they wrote a list of bullet points	they practised with their peers	they wrote a full draft (written exam)	they wrote a script (oral exam)	they studied vocabulary	they studied grammar
1 = none of them	1	1	0	1	0	1
2 = some	8	2	3	3	8	11
3 = many	2	8	1	2	7	4
4 = (nearly) all of them	8	9	16	13	5	4
nicht beantwortet	1	0	0	1	0	0

	they looked for similar texts on the internet	they wrote notes / some paragraphs	they recorded themselves	they rote learned / memorised	they practised to become fluent	they copied sentences from various sources with slight changes	they tried to formulate in their own way
1 = none of them	13	1	2	1	0	2	0
2 = some	5	7	11	3	8	3	13
3 = many	0	7	5	5	7	9	4
4 = (nearly) all of them	0	5	0	11	2	6	3
nicht beantwortet	2	0	2	0	3	0	0



Antworten auf die Frage: "Which strategies are your students using to prepare for the assessment?" (= Tabelle 27)

	they are writing a list of bullet points	they practise with their peers	they write a full draft (written exam)	they write a script (oral exam)	they study vocabulary	they study grammar
1 = none of them	0	2	0	0	2	4
2 = some	5	4	2	2	5	3
3 = many	3	4	2	1	2	3
4 = (nearly) all of them	6	5	10	11	6	5
nicht beantwortet	1	0	1	1	0	0

	they look for similar texts on the internet	they write notes / some paragraphs	they record themselves	they rote learn / memorise	they practise to become fluent	they copy sentences from various sources with slight changes	they try to formulate in their own way
1 = none of them	10	2	4	0	2	0	1
2 = some	3	3	7	2	4	2	9
3 = many	1	3	3	1	2	11	4
4 = (nearly) all of them	1	7	1	12	5	2	1
nicht beantwortet	0	0	0	0	2	0	0

Antworten auf die Frage: "Is it now more or less motivating to prepare for the GCSE?" (= Tabelle 28)

	for you	for your students
1 = much less	14	7
2 = a bit less	6	7
3 = the same	13	9
4 = a bit more	2	10
5 = much more	0	2
nicht beantwortet	0	0

Antworten auf die Frage: "Do you feel that you and your students know enough about the changes?" (= Tabelle 29)

Antwortmöglichkeit	Anzahl
1 = yes	25
2 = no	8
nicht beantwortet	2

## 7.4 Ergebnisse – Verbale Daten

Code	7. In the old GCSE, did you prefer coursework or the final written exam? (1 = Coursework, 2 = Final written exam, 3 = no preference)	8. Why? Please give at least one reason.	
F1T1	2	Pupils learn more thoroughly over the course of 2 years to use language structures which they can then apply to any topic.	FT = +ZZL, +GR
F1T2	3	I have only been teaching for a short time.	kA
F1T3	2	More academic	FT = +Akademisch (R)
F1T4	2	Coursework is an administrative nightmare	CW = +zeitintensiv
F1T5	3	i have not much experience of the old GCSE but feel the controlled assessment to be very stressful to organise and the rule around ensuring that it is in fact 'controlled' are not black or white leading much cause for confusion. A final written exam would take more skill on behalf of the students but would be a truer and fairer representation of their capabilities.	CA = +zeitintensiv, Transparenz (R) FT = +ZTF, +fair
F1T6	3	have only started this year, don't know how the old GCSE worked	kA
F1T7	2	Final exam is a fairer test. It avoids cheating. Coursework was wide open to abuse AND took too much teacher time.	FT = +fair, -schummeln, CW = +schummeln, +zeitintensiv
F1T8	2	students not ready till end and it takes too much time	FT = +ZZL, CW = +zeitintensiv
F1T9	2	less admin and own students' work rather than rely on teacher	CW = +zeitintensiv, +schummeln
F1T10	1	Opportunity to revise - less reliance on memory	FT = +auswendig
F1T11	2	Gave pupils more time to build up language ability. Less stress during course, more time to teach.	CW = +zeitintensiv
F1T12	1	Achieved better results. Less pressure	CW = +Noten, -Druck
F1T13	2	A proper and fair test of ability to manipulate the German language	FT = +ZTF, +fair, +GR
F1T14	1	because I knew my students would get higher marks	CW = +Noten
F1T15	2	Better reflection of students' ability	FT = +ZTF
F1T16	1	We could prepare students reasonably well for this and grades were better	CW = +Noten (+ZZL/TTTT)
F1T17	3	coursework takes a lot of preparation time/marking for teacher, but achieves better results for pupil. Exam is out of our control - if they haven't learnt it, it's their fault, plus there's no marking for us. But get a better idea of predicted grade outcome from coursework. Coursework suits well-organised girls better, but allows lower ability pupils to get a higher mark through redrafting.	CW = +zeitintensiv, +Noten, - S → +Noten
F1T18	2	Teaching a high-ability cohort - it suited them - more flexibility of topics throughout the 2 years	FT = pos für +S, +Themenvielfalt

F1T19	2	Easy to teach and easy to get a C grade - just had to teach pupils to put in 3 tenses (past, present and future)	FT = +Noten, +GR
F1T20	1	We are driven by results. Coursework is under the teacher's control; the final written exam is not.	CW = +Noten; Insgesamt: +Druck
F1T21	1	because students got better marks and had time to think about how language works	CW = +Noten, +GR
F1T22	2	Fair way to test. Easier to administrate / organise. Pupils learned to write and speak better. Teach in all boys school - boys overall seem to prefer exams to coursework. BUT the flexibility of being able to choose was also good - you could please everyone.	FT = +fair, -zeitintensiv
F1T23	1	It was easier for students to gain higher marks and gave them access to more materials while writing their answer.	CW = +Noten
F1T24	2	My school actually did coursework but I feel that the final written exam gives a better picture of the student overall. Coursework can leave teachers open to pressure from senior leaders to "assist" with the work because of league tables.	FT = +ZTF, CW = +Druck → +schummeln
F1T25	2	We do not approve of cheating and since coursework could be written at home, this was cheating	CW = +schummeln
F1T26	2	The pupils were not ready to take the second foreign language before then.	FT = +ZZL
F1T27	1	pupils did better	CW = +Noten
F1T28	2	This is my answer for German! I felt it ensured students knew the grammar and structures well and over a wide variety of topics rather than finding out for a piece of coursework and then immediately forgetting. However I also teach Italian and as these students go from 0 - GCSE in one year I always found coursework was a much easier option for them as they were able to be really good at a few topics and just have the others ready for oral / comprehension. It means there's a lot more to do with them in the sixth form though!	FT = +GR, +Themenvielfalt; CW = +Noten, +Druck in 6THF
F1T29	1	never taught for a written exam	kA
F1T30	2	Harder for teachers to 'cheat'. The exam forced you to teach the language, not just give them loads of help sheets to copy from. They really had to understand the language. Because loads of schools gave students too much help with coursework, the standard of work increased so much, that it was really hard for students to do well if they were not given similar help sheets. The exam actually demanded less complex language from them and less accuracy.	CW = +Druck → +schummeln; FT = leichter (R)
F1T31	1	Students got better grades	CW = +Noten
F1T32	2	I never did the coursework option so it's difficult to compare	kA
F1T33	2	Felt it prepared better for A level Did not want the ethical dilemma of 'how much help to give' which I saw in other subjects	FT = pos für 6THF, CW = +Druck → +schummeln
F1T34	2	True reflection of student's competence/skill- fairly challenging but accessible for all ability ranges	FT = +ZTF, pos für +/-S
F1T35	1	We only start German in Y9 and coursework gave less academic students a chance to build up to their best pieces. It also reduced the stress of the final exam.	CW = pos für -S, -Druck
Z1	1	I don't like the written exams	FT = neg
Z2	1	Less pressure on the children, however it	CW = -Druck
Z3	1	Pupils score higher	+Noten
Z4	1	students tend to get better results	+Noten

<b>Code</b>	<b>12. There were hardly any changes to the listening and reading part of the GCSE. How do you see that? Would you have wanted any more changes?</b>	
F2T7	I hope the questions will become less about logic and more about testing linguistic ability. At the moment, too many of the questions are rather obscure and hang too much on often fussy interpretations of one or two words.	?
F2T10	The balance between the four AT skills to have been kept.	Wunsch: Balance der Skills erhalten
F2T11	All questions and answers in English - this was a big change and will seriously affect pupils if they move onto A-Level where questions and answers are in German. The L and R skills are also now worth 20% each, not 25%, so the exam rewards rote learners in the controlled assessments more than competent linguists who can work out meaning and use linguistic/grammatical knowledge for comprehension.	Alles Englisch = neg für 6THF CA → +auswendig
F2T18	I am not happy with the comprehension questions being totally in English, to be answered in English - as I said above, this makes the leap to AS even greater - what was wrong with some assessment of comprehension in English and some in German?	Alles Englisch = neg für 6THF
F2T19	There have actually been A LOT of changes. The rubric and answers are now in ENGLISH - this is completely different from the old GCSE (except CCEA exam board, which has maintained some academic rigour).	Alles Englisch = starke Veränderung
F2T22	i am happy that questions are now in english. th only positive change to the whole thing. i would not like any more changes to this	Alles Englisch = pos
F2T24	This is quite difficult as the topics in these two skills should also be more "realistic". However, this would make it difficult for a scheme of work and specification.	Realistisch (R)
F2T25	You are wrong: the changes are quite significant and far-reaching	Starke Veränderung
F2T26	It is hard to cover the breadth of topics and vocabulary needed when we are concentrating on such small areas for the speaking and the writing.	-Themenvielfalt = neg
F2T33	the questions are now all in English which i think is good as it actually checks understanding more rigorously	Alles Englisch = pos, überprüft Verständnis (R)
F2T35	I wanted no changes to the weighting of the exams. 25% for each was good. I do however think that a bit more overlap between Higher and Foundation tiers would be a good thing.	Wunsch: Balance der Skills erhalten, mehr Überschneidung der Tiers (R)

<b>Code</b>	<b>9. How do you see...?</b>  the abolishment of coursework, the abolishment of the final written exam, the abolishment of role plays the introduction of controlled assessment (writing), the introduction of controlled assessment (speaking), the possibility to focus on one specific topic for writing and speaking, the change in short courses (only written or only spoken language)  <b>10. Please explain your attitude towards the changes listed above briefly.</b>	
F3T1	Coursework allowed for cheating to take place which is very negative and the controlled assessemnt can take up too much valuable teaching time for the administraion, and also different teachers will set about	CW = +schummeln, CA = +zeitintensiv, -Durchführungs-

	preparing for it differently despite the guidelines given. The speaking assessment has the potential to be liberating but I feel pupils should be obliged to do more than topic area to really test them. Pupils after all need to be prepared for A levels.	objektivität (R) Wunsch: +Themenvielfalt → pos für 6THF
F3T3	All the changes that I view negatively seem to me to be simplification of the subject and so-called "dumbing down".	Dumbing down (R)
F3T4	To much reliance on memory and less on spontaneous language production skills	CA = +auswendig, -Spontaneität
F3T5	I feel that the students know some topics really well but due to time constraints - some topics have to be lefty out.	CA = -Themenvielfalt, +zeitintensiv
F3T6	The pupils can rewrite and redraft their controlled assessment as often as possible, there is no limit and hence little pressure on them to learn it properly for the first time. Basically, the teacher corrects the texts, sometimes almost writes the texts for the pupils, who then have to learn it. When the assessed text is not good enough, they have another chance. This procedure is not very effective and does not contribute to a high level of German.	CA = -Druck, ?schummeln, +TTTT, +auswendig
F3T7	I am not sure that the controlled assessment (writing) realistically shows pupils' ability to construct the language.	CA = -ZTF (~GR)
F3T8	The old exam needed simplifying to work but the new one is too student centred and it does not work.	CA = zu schülerInnen- orientiert (R), neg
F3T9	not sure about what results we will get and not liking the admin side of it but i think gives a better chance to my students who want to do well	CA = +zeitintensiv, +Noten
F3T11	The old speaking exam wasn't ideal either, and relied on rote learning as the controlled assessments do, but the controlled assessments aren't an improvement in any way. Memorised chunks seem to be better rewarded than spontaneous, natural communication with a few more errors. The best pupils will still do very well - my top-set pupils love sticking in real show-off grammar phrases! - but lower-ability pupils now have to start assessments much earlier in the course and this can knock their confidence. It's a particular problem for schools who teach a language ab initio in year 10 and don't have time to put the groundwork in before the first round of assessment.	früher (mündlich) / CA = +auswendig GR zum Angeben (R) CA = neg für -S
F3T12	Although the controlled assessment speakings take a lot more time, they are much more pupil friendly and I am hoping that they will achieve better marks as a result of only having to concentrate on one topic at a time. It is more accessible for all abilities, whereas the old style was a horrible experience for lower ability pupils.	CA = +zeitintensiv, schülerInnen- orientiert (R) hoffentlich: -Themenvielfalt → +Noten
F3T13	We are dumbing down the challenge faced by students. We always bemoan as a country the dropping in standards. Until now languages stood up very well as an academic challenge and not simply a test of ability to learn. Now i don't believe i can say the same and the test favours those people who have a good memory and access to external support.	CA → Dumbing down (R), +auswendig, ?schummeln
F3T14	hated the pressure of the old format oral exam, I prefer the Controlled Assessment in speaking. I find the preparation required for the written assessments excessively time-consuming	Früher (mündlich) = neg, CA(speaking) = pos, CA(writing) = +zeitintensiv
F3T15	The new GCSE has potential to be very narrow. You spend a disproportionate amount of time preparing for the test rather than teaching a wide range of topics	CA = -Themenvielfalt, +zeitintensiv, +TTTT
F3T16	The writing and speaking assessments take up a lot of time in the year and I find that other topic areas suffer due to lack of time. The role-plays were good as they covered a variety of different topic areas and had an element of unpredictability. The new assessment seems to be rather restrictive and based on one topic area. At least with the former oral exam, there was time to cover more topic areas, which also helped	CA = +zeitintensiv, -Themenvielfalt, RP = +Themenvielfalt, +Spontaneität

	with the reading and listening exams.	
F3T17	Coursework better for lower ability sets (old-style CSE) and final exam for higher/grammar sets (old-style O Level exam)	CW = pos für -S, FT = pos für +S
F3T18	The speaking now has no spontaneity - my pupils learn 6 mini-speeches (one for each bullet point). It is not authentic to talk only on one topic. Similarly with the writing - pupils with good memories will succeed - it must be a nightmare for those who cannot memorise extended chunks of language. I am concerned about the longterm impact in terms of moving onto A Level where you simply cannot memorise your "speech" beforehand.	CA = -Spontaneität, +auswendig, -Themenvielfalt = -Authentizität (R), auswendig → +Noten, neg für 6THF
F3T19	Communicative language teaching requires pupils to Listen, Read, Speak and Write. With controlled assessment we spend weeks at a time teaching one skill. For example, in preparing for a writing task, my pupils come in for around six weeks and do no reading, listening or speaking. They copy from the board so that they have a range of phrases to put in to their writing controlled assessment.	CA = -Integrativität (R), +TTTT
F3T20	One problem with coursework is the time that it takes. Controlled assessment has a deadline. Students must keep to that deadline. Otherwise, what should take place in two weeks can take six months.	CW = +zeitintensiv
F3T21	I feel that students can still just copy phrases from their books without thinking about how language works - great for weaker student's results but not so good in making independent language thinkers	CA = +TTTT, -S → +Noten
F3T22	Many schools cheated in coursework, so that had to change. I am so upset by the abolishment of final exam. It was better for pupils, less stressful for teachers, and above all the ONLY FAIR assessment. Role plays made sure that pupils learned some useful transactional phrases, even learning these off by heart is better than nothing. Controlled assessment rules and regulations are too confusing. Even the pupils think we have gone mad when we tell them what they can and can't do. They still learn things off by heart, for speaking assessments especially, even the training courses run by the exam board tell you to get them to learn it "it's a memory test" said the AQA examiner to the room full of teachers. The only thing now is, because you can't mark the drafts, they learn incorrect German off by heart. Simple solution - Writing exam and Terminal Speaking exam marked by teacher. Russian GCSE still has a writing exam, why aren't we allowed one? It's all to save exam board money, getting us to mark speaking, and also to stop complaints of inconsistent marking of speaking exams. in short it's stupid and i hate it!	CW = +schummeln FT = -Druck, +fair RP = pos, +Interaktivität (R) trotz +auswendig; CA = keine Korrektur → lernen Fehler auswendig (R)
F3T23	I like the concept of controlled assessment - it DOES reduce ability to plagiarise etc BUT it has become an exercise in rote learning for a lot of students and this is counter-productive in so many ways!	CA = -schummeln > CW, praktisch = +auswendig
F3T24	Coursework - as stated in previous answer. The idea behind controlled assessment is good - more choice of topics, real conversations instead of pre-learned questions and answers. However, I am not sure if this is what will really happen in all schools.	CA = theoretisch +Themenvielfalt, +Spontaneität, -auswendig; praktisch: ?
F3T25	Coursework was not coursework at all but rather homework and did not reflect students' real level of attainment. CAs allow students huge flexibility and ownership of the task. Given that it happens in school, this is a more reliable	CW = -ZTF, CA = -schummeln > CW
F3T26	I have not yet seen the whole of this new exam go through and can therefore not really comment on it.	kA
F3T27	re short courses - that pupils can focus on either written or spoke) at a time, possibly taking the other part afterwards	Kommentar zu eigenen Angaben
F3T28	Pretty much the same as above. Role plays were useful as we know longer focus much on the functional language that is what they really need when they are abroad, but they were quite boring and quite hit and miss so your student may be just unlucky!	RP = +Spontaneität, +authentisch, +langweilig (R)

F3T29	for me, the controlled assessment in writing is no different from coursework really, except they have to learn 3 pieces, and in theory actually get less help now from teachers when it comes to preparing a draft etc. Used to like roleplays - they were actually transactional and students had to ask a question as well (at higher). Controlled speaking is worse than controlled writing in my opinion - there is nothing spontaneous or unexpected at all about it. they only learn 6 brief statements and that is it - old style orals, they at least had to learn about all topic areas, even if they were only asked a few questions on a couple of areas. possibility of same topic - shocking. did actually consider doing it myself to let students get a good grade (just change one bullet point and you have a new task) but decided against it in the end.	CW/ CA = kein Unterschied theoretisch: -schummeln RP = +Spontaneität, pos, CA (speaking) neg > CA (writing) CA = -Spontaneität, -Themenvielfalt → +Noten
F3T30	Controlled Assessment is better than coursework as it is more strictly controlled, but to do well, it still encourages students to learn by heart lots of prose that they have previously had marked, rather than to make the language up on the spot.	CA → +auswendig , -Spontaneität, -schummeln > CW
F3T31	Many of the changes removed the parts our students scored worse on. The negative changes to me is short course, although it's not something out school offers. It's impossible to separate the skills in such a way, you still need to teach the whole content topic wise and i just don't see how it can work.	-Integrativität = neg (R)
F3T32	I feel that instead of learning the language and how it works in many situations, the pupils are simply being trained to pass an exam.	CA → TTTT
F3T33	role play gave opportunity for useful functional language controlled assessment makes pupils see that every piece of written work done is valuable as it could contain the language they need for their CA I think that the CA speaking is even less of a real conversation than the old oral exam students actually have to learn less language for active use - helps to get a good gcse grade, but less use in the real world	RP = +Spontaneität CA= regt zum Lernen an (R) CA = -Spontaneität, +Noten, -authentisch (R)
F3T34	Abolishment of role plays- these are the conversations most likely to be had in the 'real' world, give students a chance to see if they would be able to understand life in Germany Introduction of CA/focusing on one topic area- study in greater depth and more complex language structures, often easier for weaker candidates but can push the best ones	RP = +Spontaneität, +authentisch CA & -Themenvielfalt = +GR, +ZZL, pos für +/-S
F3T35	The speaking is difficult to arrange and now we have to do two lots, we are struggling with timetabling to get these done. The other difficulty is marking these, as it is very time consuming and access to the sound files can be limited. I think for the more able the abolishment of the final exam is a pity. The pressure to get CA writing done, means that there is less teaching time before a student writes and therefore they cannot as well. The possibility of focussing on one topic will result in students moving on to AS and A2 having a limited vocabulary. As we did CW in school the CA writing is not so different, however, if students are off sick getting those controlled, monitored sessions to write once they return is difficult.	CA = +zeitintensiv, +Druck, -ZZL -Themenvielfalt → neg für 6THF FT = pos für +S

	<p><b>14. How do you think controlled assessment will affect the <u>time available for content not treated in the exam</u>?</b></p> <p><b>15. How will the introduction of controlled assessment affect the <u>students' overall German proficiency</u>?</b></p> <p><b>16. Do you think that controlled assessment says more or less about students' actual abilities than the previous test types?</b></p> <p><b>17. Please explain the choice of your answer above shortly:</b></p>	
F4T1	At least the pupils will have learned the script and written it from memory unlike the old style of 'uncontrolled' coursework, but in the end the final written exam is best.	CA = +auswendig, -schummeln > CW
F4T3	They are more likely to get help from the teacher.	CA = +schummeln
F4T4	Difficult to monitor if they received outside help	CA = +schummeln
F4T5	They know the topic in advance and can prepare for it. A truer, yet more difficult exam (as in Ireland), the students have to write on a topic and only see it on the day of the exam when they open their exam paper.	CA = -Unvorhersehbarkeit (R), +TTTT, -ZTF, -Spontaneität
F4T6	The students only learn the text by heart which was almost written by the teacher. Besides this text, the pupils know virtually nothing of German.	CA = +TTTT, +auswendig, keine Sprachkompetenz (R)
F4T7	Coursework, and prepared spoken presentations, were equally ineffective. Controlled assessment is no worse. And definitely no better.	CW / CA = neg
F4T8	Students work to task and there is no time to broaden their horizons -little time to learn about the culture of the country. No time to develop their artistic sides.	CA = +TTTT, -Kreativität (R)
F4T11	Narrower topic focus means that they can't speak about as many topics confidently. Abolishment of role-play makes it hard to work on spontaneity. However, the preparation time allows strong pupils to really work on including impressive grammar structures and this is a plus.	CA = -Themenvielfalt, -Spontaneität > RP +GR zum Angeben (R)
F4T12	As they will be able to focus and work with their strengths and weaknesses rather than just being weak at all of it. It is not such a nerve racking experience for them	CA = -Druck
F4T13	If done fairly and properly and without external help then students still require a grasp of the grammar - the only problem is that the grammar is potentially limited to a topic and students do not see the bigger picture. This is very much the role of the teacher to ensure it does not impact disastrously for A Level	CA = falls -schummeln → =GR → neg für 6THF?
F4T15	To get the results you prepare them specifically for the task and they don't take risks as you may not correct their work	CA = TTTT → +Noten, keine Korrektur → keine Risiken (R)
F4T16	I feel they can do these assessments simply learning by rote as they focus on one topic area and they have up to 6 lessons to prepare the task in advance	-Themenvielfalt & +Vorbereitungszeit → +auswendig
F4T17	less input/redrafting from teacher.	Keine Korrektur (R)
F4T18	They can memorise whole chunks straight out of the textbook without actually understanding what they are saying/writing. Also, there is nothing to stop a pupil getting a native speaker to write their "script" before they memorise it.	CA = +auswendig, keine Sprachkompetenz (R), +schummeln,
F4T20	I don't think we're testing proficiency in German. I think we're testing how well they can prepare for an exam.	CA = -ZTF, +TTTT
F4T23	I think it is less because with old style coursework two of the pieces were put together independently by the student. The new CA is a test of memory and doesn't say very much at all about a student's ability - especially considering the amount of teacher input allowed in early	CA = -ZTF, +auswendig



	stages.	
F4T24	It is the same as the controlled piece of coursework for the written exam except with a bit less information for students in advance.	CAwriting ~ CW
F4T25	If carried out properly, it allows the students to perform to their best ability. The tasks need to be set appropriately - which is often not the case - and the teachers need to conduct them properly - we should not be using preset banks of questions for the oral for example, but I fear this is happening	Falls –schummeln & +Unvorhersehbarkeit (R) = pos
F4T26	This is the first year of the course.	
F4T28	As above	
F4T29	can learn great big chunks of grammatically accurate German, without understanding anything about manipulating the language for their own purposes.	CA = +auswendig, -Sprachkompetenz (R)
F4T30	Depends. Definitely shows more than coursework, but a lot less than written exam. The speaking exams were also much harder, so CAs say less about a student's ability. There was more speaking on the spot in the speaking exams.	CA = +ZTF > CW, -ZTF < FT, RP CA = -Spontaneität
F4T31	It's a fact, it will rely heavily on rote learning, but I've found I can focus more on grammar and students creating their own language because with them all preparing and doing the same topics there is more desire for them to create their own individual response requiring more manipulation of the language in the preparation stage. Also for the speaking we used to have to prepare 6 topics plus role play, and three courseworks, now just four topics covers writing and speaking so there is more time to focus on grammar when teaching those topics and to study other topics more passively from a vocab point of view.	CA = +auswendig, -Themenvielfalt → +ZZL, +GR
F4T32	it shows how good students' memory is, rather than their ability in German	CA = -ZTF, +auswendig
F4T33	because they can be directly prepared for the questions they will face. There is no 'thinking on their feet' as in the old exam and as you have to do in real life	CA = +TTTT, -Spontaneität, -Unvorhersehbarkeit (R)
F4T34	In CA, students can simply copy out corrected work from their books and reproduce it - which does not show any of their true ability. The people who get the best marks are the ones who can learn it the best.	CA = +TTTT, -ZTF, +auswendig → +Noten
F4T35	As students have a time limit for those whose memories are not so good on lexical items, but whose grammar is good they have limited opportunity to look forgotten or new more interesting words up and to use them appropriately. Students also have less time to check work under the new system.	CA = -Vorbereitungszeit neg für +GR / -auswendigS

Code	13. How does the introduction of controlled assessment affect rote learning as exam preparation? Please explain briefly.	
F5T1	It is on the one hand positive to make children learn, but it is better for them to learn adaptable, flexible structures to use spontaneously, rather than a script. The pupils have found learning 200 words off by heart a huge challenge and it is not fair if they can use structures but cannot memorise scripts.	CA = +auswendig, -Spontaneität → -fair
F5T3	It encourages rote learning and I have seen this in practice. This is detrimental to ultimate fluency and ability to use the language in situ.	CA = +auswendig, -Spontaneität
F5T5	It is not a far move from learning by heart as student panic and try to learn huge chunks of info prior to their exam	CA = +auswendig
F5T7	Pupils are being encouraged to learn whole texts off by heart. Hugely negative. A fairer exam (for writing) would be for them to be given texts to translate (one from English into TL, the other from TL to English). That's the only way we can have a level playing field. At the moment, grades	CA = +auswendig; +schummeln & +auswendig →

	depend too much on how much the teachers cheat and on how much the pupils learn by rote.	+Noten Wunsch: Übersetzung (R)
F5T8	Badly. We cannot help students prepare for the tasks other than direct them to resources or each other and they do not have the maturity to manage.	TTTT erschwert
F5T11	See above answers.	kA
F5T12	Some will still do it, it depends on the type of learner they are.	CA = =auswendig
F5T13	I can envisage a situation where students do their controlled work but supplement it with help from a language speaker outside the classroom and then bring it straight into a controlled examination. The same applies for speaking as for writing	CA = ?schummeln
F5T15	They just learn by rote and regurgitate their prepared pieces	CA = +auswendig
F5T16	I don't think that it does affect it much because even with the controlled assessments, they get to know what the task is and can still learn a presentation off-by-heart and likely questions.	RP / CA = +auswendig, -Unvorhersehbarkeit (R)
F5T17	today's pupils are incapable of thinking for themselves and do not have the discipline/motivation to learn in class as pupils did twenty years ago. Controlled assessment (5 lessons preparation time) takes 3 months out of pupils' learning time over the course of the GCSE.	CA = +zeitintensiv
F5T18	It encourages rote learning. We now train our pupils in Years 7-9 to rote-learn before assessments.	CA = +auswendig, auch in KS3
F5T19	You can now pass a GCSE by rote learning your prepared piece. A lot of teachers are abandoning grammar teaching.	CA = +auswendig, -GR
F5T20	It makes it unavoidable. Students with the best memories will do best. Those who don't have a good memory are already saying, "I can't do that."	CA = +auswendig
F5T21	I will happen even more - an impossible task for some students	CA = +auswendig
F5T22	they do it more than before.	
F5T23	It's all a lot of students do....and I think many teachers don't teach the nuts and bolts, so students have no choice but to rote learn, and they hate it.	CA = +auswendig, -Motivation (R)
F5T24	I believe that in many schools rote learning of topics will still take place. Students in my own school have asked me how to learn / revise. Staff should focus on teaching skills and grammar and how to apply this but in reality teachers may consider that this cannot happen with current year 10 and 11. I personally have adapted my teaching in years 8-9 to ensure that students are more equipped to deal with this. This should have a positive effect on AS and A level.	CA = =auswendig Wunsch: +GR +GR in KS3 → pos für 6THF
F5T25	If carried out properly, there will be no rote learning for CAs. The great problem is that many teachers have failed to grasp the main point of CA.	CA = theoretisch: -auswendig
F5T26	The same pupils that needed to rote learn for the previous exam still need to learn off by heart for this exam; but this time what they are learning is more than likely incorrect German because a teacher has not checked it.	früher / CA = =auswendig CA = keine Korrektur → Fehler auswendig lernen (R)
F5T27	It doesn't - those who can, will continue to do so, whilst those who can't, will continue to score less well	früher / CA = =auswendig; +auswendig → +Noten, -

		auswendig → -Noten
F5T28	They do a lot more of this, previously they had to have a very good understanding of the grammar and structures so they could adapt this to the written exam tasks. Now they just learn write something for a topic and learn it off by heart (and then immediately forget it).	CA = +auswendig, FT = +GR
F5T29	They learn it all off by heart - simple. Out of my group, I don't think there was one that hadn't written it all out in advance and learnt it, or added anything spontaneously in the assessment session.	CA = +auswendig, -Spontaneität
F5T30	Increases the chances of this being the key factor in the exam. In stage 1 of the controlled assessment, teachers are allowed to get students to practise language they will use in the CA, so they will make sure they have somewhere, already written an answer to the CA, had it marked, before Stage 2. This means students will have a perfect answer to learn by heart. They'd be silly not to do this, as they are more likely to score the marks this way.	CA = +TTTT, +auswendig → +Noten
F5T31	It will be used more.	CA = +auswendig
F5T32	it means that rote learning is becoming the only way of learning for many.	CA = +auswendig
F5T33	they still do it but have less to do at any one time. Rote learning is not necessarily a bad thing - it does have its place	CA = =auswendig, weniger Stoff (R)
F5T34	CA builds on language already learnt that is useful in exams	
F5T35	As their teacher and particularly with less able linguist I have to encourage them to rote learn some structures. Last year this resulted in too many students writing very similar things, which the exam board did not like.	CA = -S → +auswendig → -Noten

Code	20. Is there any additional important strategy which I did not list?	
F6T11	They revised using homework/notes which the teacher had specifically prepared as controlled assessment preparation. (This is a big problem, but we all do it!)	CA = +TTTT, ?schummeln
F6T13	I would be amazed if students somewhere in the country didn't have their work written for them by a native speaker and then rote learned it. I do not know for certain but the possibility must exist.	CA = ?schummeln
F6T17	AfL. Peer marking of draft texts	
F6T18	In preparation (Stage One), I set homeworks which relate directly to the bullet points which the pupils will be given at a later date - I correct their work and then they copy out a corrected version....which they then memorise. They also use text-to-speech websites to check pronunciation.	CA = +TTTT, +auswendig
F6T19	i have told them to avoid the internet as it is full of rubbish	
F6T20	the teacher's ability to correct drafts	CA = ?schummeln/TTTT
F6T31	using teacher made recordings to practice	Aufnahmen (R)
F6T34	Revision cards	Kärtchen (R)
F6T35	For the oral they practised with teachers or the Fremdsprachassistent.	LehrerIn / FLA (R)

<b>Code</b>	<b>32. Was the teaching of GCSE classes different two or more years ago?</b> <b>33. Please explain, what was different and how it changed.</b>	
F7T1	As stated before there is a lot more reading and less listening going on. The pupils are having to deal with a lot of texts to draw language from which is difficult for weaker pupils who do not enjoy the written word so much.	-Balance der Skills (R), neg für -S
F7T4	a lot of time wasted on explaining the administration to pupils	+zeitintensiv
F7T5	I don't know as I did not teach it	kA
F7T6	unfortunately I cannot give any comments on all the questions which refer to the old GCSE	kA
F7T7	no change	Früher = heute
F7T8	Students do not like assessed tasks in so many subjects are immune to them now. The only positive about the new GCSE is that we can conduct the orals in a more relaxed way and that the student constructs the oral himself. Otherwise very dull.	Insgesamt mehr Tests (R) = neg CA = -Druck
F7T11	Controlled assessments require 2 weeks' worth of lessons for preparation, sitting the assessment(s) and then marking and giving feedback. If we do three sets of assessment throughout the course, we lose a half-term of teaching. As all the assessments have to be in by May of Year 11, there is a very dry spell between May and June where we panic and try to teach them comprehension skills. The course therefore has to be carefully planned to include comprehension skills from earlier on.	CA = +zeitintensiv
F7T13	The challenge of both written and speaking final examinations focused the mind and made students more alert and driven. The motivation for examination seems to have dropped and i find myself thinking more and more that there is a greater possibility for students and staff to cheat the system (which i abhor). It will increase the gap still further between comprehensive and grammar/private students who are not simply learning to a task.	CA = -Motivation, +schummeln, Unterschied +/-S (R) +S = -TTTT
F7T15	You could explore coursework areas in depth and prepare thoroughly for oral	CW = +ZZL > CA
F7T16	More topics covered, more ind epth grammar practice, not as pushed for time as I am now	früher = +Themenvielfalt, +GR, -zeitintensiv > heute
F7T17	tried to rejiggle everything we were already doing towards the new changes.	+admin (R)
F7T18	new textbook, new topics (e.g. technology), new assessments spread throughout the two year course - different subjects competing for slots in the busy school year calendar so that controlled assessments do not clash.	CA = +zeitintensiv
F7T19	It is now demotivating and the absence of role-plays is a shame. They never study the train station, the cafe etc. Now when they arrive in Germany they won't be able to book train tickets, go to a restaurant etc.	CA = -Motivation, RP = authentisch (R)
F7T20	Controlled assessment is now worth 60%, and it dominates. Listening and reading, at 40%, are less important.	Balance der Skills (R)
F7T22	we had more time with pupils. COnrolled assessment stops os from teaching, in the time they are revising for the task and when we have to conduct speaking tasks. This adds up to nearly 8 weeks of teaching time in a year.	Früher: +ZZL, CA = +zeitintensiv
F7T23	As before, teaching to the test a lot more now. Not enough time to go deeply into anything.	CA = +TTTT, -ZZL
F7T24	There was less pressure in my current school but this changed following an Ofsted inspection.	Früher = -Druck (Ofstedinspektion)
F7T25	We still have to cover the content and the grammar, we still have to teach students transferable skills: in short we have to teach them a language	Früher = heute
F7T26	To be quite honest I am finding this questionnaire difficult to answer as	kA

	we are still in the early days of the new course.	
F7T28	As I said before	kA
F7T30	Two years ago, students did GCSE over 2 years and they had a written exam, so grammar had to be learnt (albeit only 1st person verbs in many cases). Now students only get one year to complete the course in my school - it is an absolute nightmare to fit it in.	Früher = +GR, CA = +zeitintensiv
F7T31	The pressure towards the final oral exam and trying to prepare that as well as preparing for reading and listening has changed. Now I envisage there being much more time to prepare specifically for listening and reading towards the end.	?Druck, +ZZL/TTTT?
F7T32	There was time to work systematically through the grammar and vocabulary, whereas now we have to cover short topics and prepare for rote learning of a piece.	Früher = +GR, CA = +TTTT, +auswendig
F7T33	no role play practice no practice at interpreting questions less practice at finding strategies to work around limited knowledge	CA = -Spontaneität?
F7T34	Students nowadays feel that if they mess up one CA, they still have other chances, so are motivated all the time. 3 years ago, all they were concerned about was not messing up their final exam. I find it easier to teach students now	CA = +Motivation
F7T35	We could finish CW, perhaps do 3, 4 or even 5 pieces and then choose the best to enter. Now do the the highly controlled environment this is no longer possible. Then we would have chance to concentrate on the speaking exam. Role plays were far easier for less academic student and they always gained marks on them.	CA neg < CW CA = -ZZL/-TTTT für FTspeaking RP = -S → +Noten

Code	25. If the impact of the GCSE on your teaching changed in any way, please explain this briefly: what exactly changed and why?	
F8T1	The pupils have a lot of reading texts with exercises to lift useful phrases from the text and to then apply to other similar questions - this is a bit repetitive and formulaic teaching.	CA = -Motivation, +TTTT, -Themenvielfalt?
F8T4	when the speaking and writing are finished it leaves little variety to do in the classroom – that is boring for the pupils	
F8T5	I can't really comment for 23 - 27 as I did not teach old GCSE.	kA
F8T6	at the moment there is only teaching towards the exam	CA = +TTTT
F8T7	no change	Früher ~ CA
F8T8	As above, teach to targets. Stress to have to mark own orals and prepare documentation for board - feel that they are employing us to do their work. No spontaneity or joy and students just as bored with "new" topics.	CA = +TTTT, +zeitintensiv, -Spontaneität, -Motivation
F8T9	modules get assessed so students know what they are studying for	CA = regt zum Lernen an? (R)
F8T10	No other subject requires the amount of memorization that MFL still require - pupils are simply not used to it, and many are fazed by the challenge.	CA = +auswendig
F8T11	I no longer teach question words as the questions in the exam are in English. This means that I will have to teach them to 6th form when they start the A-Level course. There is also an effect on how we assess at Key Stage 3.	CA → KS3 & 6THF
F8T13	Changing topic areas, less time spent on developing oral fluency.	CA = -Spontaneität, ?Themenvielfalt
F8T15	I don't enjoy teaching to an exam that had very little to do with Germany or things the students would find useful	CA = +TTTT, -authentisch
F8T16	I am unable to cover as many topic areas in depth as well as focus on	CA =

	grammar in more depth as it is very time consuming preparing for two speaking and two written assessments	-Themenvielfalt, -GR, +zeitintensiv
F8T17	different challenges. had to relearn everything - again!	CA = starke Veränderung (R)
F8T18	I now feel under pressure to over-teach certain topics and skim others. Each topic is taught with the bullet points in mind - everything is geared towards that. In the old GCSE I taught a range of topics, often creatively, and then only in the final term or so did we start looking at actual exam papers and exam technique.	CA = +Druck → -Themenvielfalt, +TTTT, -Kreativität, +zeitintensiv
F8T19	I now spoon feed pupils to pass silly tasks. This has been the worst change to GCSE German in the history of the exam. Paper work for teachers has increased. Teachers now have to mark and award a grade for the Speaking - this will be inconsistent across the country and is not fair.	CA = +TTTT, +zeitintensiv, -fair
F8T20	I feel that I am preparing for the exam, nothing else.	CA = +TTTT
F8T21	I have to think about which topics will get the students the best results	CA = -Themenvielfalt → +Noten
F8T22	makes it dull, boring, exam focussed and produces weak linguists	CA = -Motivation, +TTTT, -GR
F8T23	My teaching has become much more focussed on the exam. I dislike teaching to the test but now that we are preparing for a CA it's all I can do!	CA = +Druck → +TTTT
F8T24	More grammar and skills focused, less content.	Früher? = +GR, -Themenvielfalt?
F8T25	we no longer have to practise the very boring and mostly futile role plays	RP = -Motivation
F8T26	It really is difficult to say. Our job as teachers is to prepare the pupils for the exams and facilitate them achieving their best. So, of course, every change in the exam syllabus has a knock on effect on our teaching.	Job (R) = +TTTT → +Noten, starker Backwascheffekt (R)
F8T28	Focus on one topic in depth whilst other topics not needed for CA are taught with much less detail (because there is less time) and really just for comprehension. I still teach the grammar in detail because I see this as being extremely important.	CA = +zeitintensiv → -Themenvielfalt, =GR
F8T29	become more focussed on how I set h/w as a preparation for the controlled assessment	CA = +TTTT
F8T30	Have to focus on fewer topics - so less breadth. Have to do the CAs sooner - not enough time for students to actually understand the grammar and acquire the vocab so it has to be really focused on the test.	CA = -Themenvielfalt, -ZL, +TTTT
F8T31	I seem to be able to teach more grammar, although my experience of the old GCSE was always with mixed ability classes, my current class is a higher ability class so I can't be sure that the changes in my teaching are only affected by the changes in assessments.	CA = +GR bei +S, früher: -S
F8T32	we have to have covered the topics so much earlier in the year that I feel my students have not had enough time to produce their best performance.	CA = -ZL
F8T33	it is narrowed as I know the questions they will face	CA = -Unvorhersehbarkeit, -Themenvielfalt
F8T35	Less vocab covered, possibly pushing students to use more complicated structures, before they are really ready to. We do the writing in Y10 and have only started German in Y9. Listening and reading activities have been left to one side a bit more as we have to focus on the 30% exams first. We received our results for the writing and as the grades awarded were not so good the students have been less motivated. We also found the grades not comparable to the old CW ones.	CA = +GR, -ZL, -Integrativität, -Noten, -Motivation

Code	Open question	
F9T1	Whilst complete predictability can be negative, an exam cannot be entirely unpredictable. Pupils must have the chance to prepare for something and therefore need to know what to prepare for. The new speaking test should combine some pre-prepared answers / presentation which then lead into more 'unpredictable' though not unguessable areas.	Wunsch = ~Unvorhersehbarkeit
F9T4	it is rather early to conduct this survey. Our thoughts are not yet clear. But the questions are sensible	kA
F9T5	I feel strongly that having an external examiner is better.  Teachers feel pressure and I believe many may inflate grades due to pressure of school, parents, students, dept to gain high grades/perform well in league tables.  External examiner makes it more serious and exam-like for students and I believe it will help them to focus more.  In Ireland we do not have different exam boards so everyone in the whole country sits the exact same exam on the exact same day. I believe this to be a fairer judgement of ability and when students go on to college/university they will have all come from doing the same exams.	Beobachtung: +Druck → +schummeln → +Noten  Wunsch: Extern, keine Exam Boards
F9T6	I have only started teaching in England this year and I am doing at the moment my GTP training year. My experiences so far show me that German is less and less popular in England as pupils find it really hard. It obviously also depends which school one is teaching in, but this general trend can be noticed.	Beobachtung: Deutsch = neg, schwierig
F9T7	I encounter many pupils, sometimes even those doing A level languages, who have gained high grades in GCSE but who cannot put together a simple sentence, and who cannot answer simple greetings and enquiries as to their health. Many are almost unaware of the difference between singular and plural forms of the second person. Very few can use personal pronouns. Teaching strategies seem to be chaotic. There seems to be a view that pupils will learn grammar and structure for themselves just by being exposed to the target language in "topics". This emphatically does not happen.	Beobachtung: Oft keine Sprachkompetenz -GR in 6THF
F9T11	Noone has yet looked in detail at the effect the revised GCSE will have on Key Stage 3 and Key Stage 5.  Teachers tend to be very negative about change! In 5 years, answers to such a survey might be rather different.	Beobachtung: Veränderungen immer negativ, welcher Effekt auf KS3 & 6THF?
F9T13	The emphasis should be simply on grammar as the vocab is transitive. A student could develop grammar and apply it to where it is required according to personal and cultural interests. Final examinations or honest internal assessment would make this challenging and relevant and would create A Level standard students. League tables just cannot help. They only inform parents how well a student is prepared for an exam and not a reflection on the quality of teaching and learning. Remove this pressure and you will see schools experimenting and taking risks and doing exciting things that will increase uptake of university courses and a higher quality of student. Coursework simply does not test a student's real cognitive ability in any subject as the potential for bending the rules is too great.	Wunsch: +GR, FT oder intern & -schummeln → pos für 6THF  Beobachtung: League tables = neg, -ZTF, +TTTT, +Druck, -Kreativität CW → +schummeln
F9T16	I have a rather negative attitude about the new GCSE - I feel that nothing is really very clear and that there are too many assessments to prepare for. This is very time consuming and affects how much can actually be achieved in the classroom. I also feel that that	Beobachtung: -Transparenz, +zeitintensiv, -Motivation,

	students' enjoyment is affected as we constantly seem to be preparing for an assessment and this increases the pressure for the teachers and the students who have many other subjects to study for at the same time.	+TTTT, +Druck
F9T17	Today's GCSE does not prepare pupils for the rigours of A Level. A/A* grade pupils cannot speak enough of the language/know enough grammar to be able discuss in the language topics they cannot often verbalise in their own mother tongue! Especially now there is the pressure to take AS in nine months.	Beobachtung: CA = -GR → neg für 6THF
F9T18	I feel that my teaching has become dull because each half-term of lessons points towards a controlled assessment and every lesson counts towards that. It is a nightmare trying to juggle controlled assessments for Year 10 and Year 11 - if nothing else, it is extremely time-consuming. The marking is a headache too - I have 56 candidates (2 classes) - just to physically LISTEN to their 2 Speaking assessments (2 X 6 minutes per candidate) will take me over 11 hours of solid listening. There is also (because of results and league tables) huge pressure on staff to inflate, even if ever so slightly, the marks for the speaking assessment which is not submitted to the board - what is to stop me from adding a few marks onto every candidate? This non-submitted speaking assessment does not even have to be recorded....so there is no moderating, even within the school. I am also concerned about total lack of parity between exam boards. For EDEXCEL listening and reading are only based on certain topics, whereas for AQA listening and reading are based on the entire syllabus. Why the difference? We needed a shake up, but this is hell.	Beobachtung: +zeitintensiv +Druck durch League Tables & +Noten → +schummeln, Exam boards ungleich schwer
F9T20	I feel that we are no dominated by examinations at all levels, and this leads to massive dishonesty, at all levels. Two levels dominate: one is that students' higher grades than 15 years ago suggest they are better learners: they are not, it is the standards that have changed. The second is the need by senior managers to cling to this fiction.  I don't care if standards fall or rise, or change to meet different educational needs and changing times. I just wish we were honest about it.  None of the changes in any subject has increased proficiency, for one very simple reason: the distribution of intelligence from year to year goes up or down. Everything else is a lie.	Beobachtung: +Druck durch +Tests → +schummeln Wunsch: -schummeln, ?Standards
F9T22	Ich bin sehr dankbar, dass Sie mir die Gelegenheit gegeben haben, meine Meinung zum Thema niederzuschreiben. Die "Exam Boards" hier in England haben uns keine solche Gelegenheit gegeben - die haben vielleicht Angst! Es wuerde uns Lehrern in England sehr helfen, wenn Sie diese Umfrage an AQA schicken koennten, mit dem Vorschlag, dass sie uns um unsere Meinungen beten. Danke	Beobachtung: -Transparenz (R) Wunsch: + Transparenz (R)
F9T35	The changes have been implemented too quickly and without sufficient back-up and guidance. Our experience has been very negative so far. Our team of 7 teachers used to feel confident about marking CW and invariably knew what made for a C,B,A or A* grade. After our results last year it was obvious that this was no longer the case. It has made our students doubt us and created a very negative atmosphere in the languages area.  This has also come on top of the fairly recent changes to the A level and has required a lot of time and effort to implement. At the moment we are not sure that our students will benefit as much from the new GCSE.	Beobachtung: -Transparenz (R), neg, +zeitintensiv



### **7.5 E-Mail-Konversation mit Frau Lamb, Head of MFL an der John Warner School**

Von: Denscher, Elisabeth  
gesendet am: 22. Februar 2011, 14:08  
An: Lamb, Vanessa

Dear Vanessa,

I am writing to you regarding some of the findings of my survey. When we discussed the impact of controlled assessment in October, you mentioned that there might be more room for creative teaching than before. At that time, you taught the topic “environment” to a year 11 class and told me that there won’t be any assessments relating to it, so that you can approach the topic in a more interesting way than in the previous years. In my survey, however, only three out of 35 participants said that there will be more room for “personalised and creative teaching”, so I had a look at your questionnaire, and you too answered that there will be “much less” room for that than before. Now I wonder why that is: Did you change your opinion between October and December – possibly because you started with CA or discussed this with other teachers? – or did you interpret the question in a different way?

Best wishes,

Elisabeth

---

Von: Lamb, Vanessa  
gesendet am: 24 Februar 2011, 14:00  
An: Denscher, Elisabeth

Hi Elisabeth,

Thanks for your email, good to hear from you. I think my opinion changes about the the new GCSE as we go through it.

Overall, I and other teachers have feared that there is less room for 'personalised and creative' teaching given that we have to give over 2 weeks each time to the preparation of the written and spoken CA. If you therefore wish to have the pupils

produce a formal written piece for each topic they do, then you are looking at 8 to 10 weeks of teaching time lost to that, and then a further 4 weeks for the preparation for their speaking tasks. That is practically a whole term.

In theory though, there seems to be more room for creativity as the general topic areas have been simplified, on the face of it, in terms of what pupils need to be able to produce either written or spoken, and the guidelines are very loose - so yes, in theory there is room to be creative, if teachers find sufficient time and are willing to plan their own schemes of work more so than they did before.

There are no suggested questions and answers to use for preparing for the speaking CA, so again there is the room to personalize. It is the usual problem though of things in theory looking inviting, but in reality being very challenging because of time constraints, and it is also worth mentioning that teachers could get it all horribly wrong if given too little guidance.

I hope that helps you. At the moment, I think all teachers are contradicting themselves as they blindly fumble their way through the new exam course.

It isn't hugely different in many ways, we are still teaching the same level of language and using roughly the same topics, just updated a bit, but it is the getting used to time scales and new assessment procedures that has most people perplexed and anxious, and the sheer lack of a 'safety net' that we had before with suggested question banks and role plays to base our preparation on.

I still would prefer a final written exam!

Take care

Good luck!

Vanessa

## **7.6 Abstracts**

Die vorliegende Arbeit untersucht den Backwasheffekt des Deutsch-GCSE an englischen Schulen vor und nach der Reform von 2009. Zunächst wird eine Verbindung zwischen Testtheorie und den Grundlagen der Backwashforschung hergestellt, indem gezeigt wird, wie Kommunikative Kompetenz, Konstruktvalidität und positiver Backwash in Zusammenhang stehen. Darauf basiert eine Analyse des GCSE vor der Reform, die auf einen negativen Backwash hinweist.

Im Anschluss an eine Darstellung der reformierten Prüfung folgt eine empirische Untersuchung zur Wahrnehmung des Backwasheffekts durch Deutschlehrkräfte in England. Die Ergebnisse deuten auf einen starken und negativen allgemeinen Backwasheffekt hin, der mit dem Notendruck der League Tables zusammenhängt. Viele der befragten Lehrkräfte beklagen, dass ihr Unterricht nun nur mehr der Prüfungsvorbereitung dient und SchülerInnen auf Auswendiglernen angewiesen sind, statt dass die Sprache für kommunikative Zwecke gelehrt und gelernt werden kann. Aufgrund dieser negativen Ergebnisse komme ich zu dem Schluss, dass die Abschaffung der Schulrankings eine sinnvolle Maßnahme wäre.

---

This thesis investigates the washback effect of the German GCSE within the English education system with a special focus on the changes triggered by the reform of 2009. First, I will draw a connection between two related yet distinct fields of research: test theory and washback research – thereby, the interrelation of communicative competence, construct validity and positive washback will be revealed. The following analysis of the previous GCSE indicates negative backwash.

After a description of the new GCSE, the second part consists in an empirical study which explores teachers' perceptions of washback. The findings suggest a strong and negative general washback that is related to pressure from league tables. To be more specific, many teachers feel that they are now solely training their students to pass an exam using rote learning strategies, instead of teaching the language for communicative purposes. Finally, I will argue that school rankings shall be abolished because of their harmful impact.



# Curriculum Vitae

## Angaben zur Person

Vor- und Zuname Elisabeth Denscher

Geburtsdatum 26. 7. 1986

## Ausbildung und studienbezogene Berufserfahrung

2004 Matura am BRG 4, Waltergasse 7, 1040 Wien

2004 - 2011 Studium an der Universität Wien, Lehramt Englisch und Deutsch

2007 - 2008 Freie Mitarbeiterin der Wiener Kinderfreunde

Sommer 2007 Lernbetreuerin für Deutsch der Kinderfreunde Niederösterreich

2008 Übersetzungstätigkeiten für Baumgartner, Andreas et al. (Hg.):  
Wer widerstand? Biografien von WiderstandskämpferInnen aus  
ganz Europa im KZ Mauthausen und Beiträge zum  
internationalen Symposium 2008. Wien: Edition Mauthausen  
2008.

2008-2009 Sprachassistentin an der John Warner School, Hoddesdon,  
Herts, UK

2009 - 2011 Freie Mitarbeiterin der Wiener Kinderfreunde