

FREMSPRACHENDIDAKTIK **INHALTS- UND LERNERORIENTIERT**

Herausgegeben von Gabriele Blell und Rifa Kupetz
Mitbegründet von Karlheinz Hellwig

24

SYLKE BAKKER

FRAGEN DES ASSESSMENTS **AUS SICHT VON** **ENGLISCHLEHRKRÄFTEN**

EMPIRISCHE ANNÄHERUNGEN
DURCH QUALITATIVE INHALTSANALYSEN



PETER LANG
EDITION

Fragen des Assessments aus Sicht von Englischlehrkräften

FREMSPRACHENDIDAKTIK

INHALTS- UND LERNERORIENTIERT

Herausgegeben von Gabriele Blell und Rita Kupetz
Mitbegründet von Karlheinz Hellwig

Band 24



PETER LANG
EDITION

SYLKE BAKKER

**FRAGEN DES ASSESSMENTS
AUS SICHT VON
ENGLISCHLEHRKRÄFTEN**

**EMPIRISCHE ANNÄHERUNGEN
DURCH QUALITATIVE INHALTSANALYSEN**



**PETER LANG
EDITION**

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Zugl.: Oldenburg, Univ., Diss., 2012

D 715

ISSN 1430-8150

ISBN 978-3-631-64070-8 (Print)

E-ISBN 978-3-653-02546-0 (E-Book)

DOI 10.3726/978-3-653-02546-0

© Peter Lang GmbH

Internationaler Verlag der Wissenschaften

Frankfurt am Main 2013

Alle Rechte vorbehalten.

Peter Lang Edition ist ein Imprint der Peter Lang GmbH

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlages unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

www.peterlang.de

Inhaltsverzeichnis

Abkürzungsverzeichnis.....	10
Abbildungsverzeichnis	11
Tabellenverzeichnis	12
Vorwort.....	15
1. Einleitung.....	17
1.1 Lernen, Lehren und Bewerten vor neuen Horizonten	17
1.2 Aufbau der Dissertationsschrift.....	21
2. Theoretische Fundierung der Studie	25
2.1 Einordnung der Studie in die empirische Unterrichtsforschung.....	25
2.1.1 Einflüsse und Entwicklungslinien.....	25
2.1.2 Methodologische Verortung des Forschungsprojekts	29
2.1.3 Einordnung in die Sprachlehr- und Lernforschung.....	34
2.1.4 Die Studie im Schnittpunkt qualitativer und quantitativer Forschung.....	39
2.2 Diagnostik	44
2.2.1 Thematische Annäherung und Überblick.....	44
2.2.2 Kompetenzorientierung und Kompetenzdiagnostik	48
2.2.3 GER und Bildungsstandards als Bezugssysteme	53
2.2.4 Diagnostik in der Aus- und Fortbildung von Fremd- sprachenlehrkräften	57
2.3 Das „Handwerk“ des Beurteilens, Testens und Evaluierens	59
2.3.1 Verknüpfung von Theorie und Praxis	59
2.3.2 <i>Assessment</i> in formativer und summativer Ausprägung.....	64
2.4 <i>Assessment</i> aus der Perspektive der Bildungsgangforschung.....	69
2.4.1 Der Einfluss subjektiver Theorien geringer Reichweite	69
2.4.2 Entwicklung und Entwicklungsaufgaben von Fremd- sprachenlehrern und -lernern.....	73
2.4.3 Rekonstruktion der Erfahrungen mit Leistungsbewertung.....	75

2.5	Der Fremdsprachenlehrer als Experte	77
2.5.1	Epistemologische Veränderungen	77
2.5.2	Expertentum und Expertise	79
2.5.3	Grenzen der Professionalisierung	85
2.5.4	Die Wissensbasis von Fremdsprachenlehrkräften	86
2.5.5	Wissen über Sprache	88
2.5.6	Wissen über Spracherwerb	90
2.5.7	„Languaculture“	91
2.5.8	Exkurs: Professionalität, Identität und Diskurs im Zeitalter der „Superkomplexität“	93
2.6	Zusammenfassung der Ergebnisse	97
3.	Untersuchungsdesign	101
3.1	Methodologische Grundlagen	101
3.1.1	Allgemeine Überlegungen	101
3.1.2	Ebenen der Abstraktion im Forschungsprozess	104
3.1.3	Forschungsfragen	106
3.1.4	Konstruktion des Interviewleitfadens	108
3.1.4.1	Grundlegende Prinzipien	108
3.1.4.2	Der Leitfaden	109
3.2	Experteninterviews	113
3.2.1	Qualitative Interviews im Überblick	113
3.2.2	Das Experteninterview als Forschungsinstrument	117
3.2.3	Organisation und Repräsentation von Expertenwissen	118
3.2.4	Experten suchen und finden: der Zugang zum Feld	121
3.2.5	Experteninterviews zwischen Information und Narration	125
3.2.6	Überlegungen zur Auswertung der Experteninterviews	127
3.2.6.1	Qualitativ-interpretative und deduktiv-nomologische Verfahren im Vergleich	127
3.2.6.2	Entwicklung einer passenden Auswertungsstrategie ...	130
3.2.7	Kommunikative Validierung	133
3.3	Vorbereitung der Schritte ins Feld	135
3.3.1	Forschungsverlauf	135
3.3.2	Mögliche Vorschaltung eines Fragebogens	137
3.3.3	Erhebungsbogen berufsbiographischer Daten	138
3.3.4	Der Umgang mit den Daten	139
3.3.4.1	Datensicherheit und Datensicherung	139
3.3.4.2	Transkription	140
3.3.5	Die Interviewsituation	141
3.3.5.1	Der Rahmen	141

3.3.5.2	Der Einfluss subjektiver Faktoren	144
3.3.6	Konsequenzen aus der Pilotstudie.....	148
3.4	Zusammenfassung der Ergebnisse	152
4	Auswertung der Interviewstudie	155
4.1	Erfahrungen im Feld	155
4.1.1	Die Gruppe der Befragten	155
4.1.2	Altersstruktur	156
4.1.3	Fächerkombination.....	157
4.1.4	Besondere Funktionen und Expertenstatus	158
4.1.5	Nationalität und Mehrsprachigkeit.....	161
4.2	Verlauf und Grenzen der Erhebung.....	162
4.2.1	Der Verlauf im Überblick	162
4.2.2	Grenzen	163
4.3	Anpassung der Auswertungsstrategie	164
4.4	<i>Assessments</i> im eigenen Spracherwerbsprozess	166
4.4.1	Erfahrungen mit der Fremdsprache in Kindheit und Schulzeit.....	166
4.4.2	Emotional prägender Erstkontakt.....	166
4.4.3	Sprachenlernen in der Institution Schule.....	173
4.4.3.1	Die Rolle der Lehrenden.....	174
4.4.3.2	Positive Erfahrungen mit Leistungsbewertung im schulischen Fremdsprachenunterricht.....	176
4.4.3.3	Ungünstige Erfahrungen mit Leistungsbewertung im schulischen Fremdsprachenunterricht	183
4.4.4	Vorstellungen über und Erfahrungen mit <i>Assessment</i> vor dem Hintergrund von Auslandserfahrungen	189
4.4.4.1	Auslandsaufenthalte im Überblick.....	189
4.4.4.2	Reisen oder Daheimbleiben	190
4.4.4.3	Schlüsselerlebnisse in Bezug auf <i>Assessment</i> im Ausland.....	195
4.4.4.3.1	Institutionelle Lernzusammenhänge	196
4.4.4.3.2	Informelle Lernzusammenhänge	199
4.4.4.3.3	Schlüsselerlebnisse und zugrunde liegenden Mechanismen	203
4.4.5	Vorstellungen über Maßstäbe und Prozesse gerechter Leistungsbeurteilung	204
4.5	Leistungsbeurteilung als Bestandteil erfolgreichen Fremdsprachenlernens.....	210
4.5.1	Vorstellungen eines idealen Fremdsprachenunterrichts.....	211

4.5.2	Umgang mit Fragen des <i>Assessments</i> unter idealen Bedingungen	212
4.5.3	Vergleich von Ideal und Realität.....	221
4.5.4	Maßstäbe, Orientierungspunkte und Prozesse.....	227
4.5.4.1	Bewertungspraxis im „traditionellen“ Fremdsprachenunterricht.....	228
4.5.4.2	Der GER in Theorie und Praxis.....	231
4.5.5	Möglichkeiten und Grenzen kollegialer Kooperation	237
4.5.5.1	Ergebnisse im Überblick.....	237
4.5.5.2	Konflikthafte Kooperation.....	240
4.5.5.3	Gelingende Kooperation.....	245
4.5.5.4	Gelingensbedingungen für Kooperation	246
4.6	Leistungsbeurteilung, Curriculumentwicklung und Professionalisierungsprozesse	252
4.6.1	Vom intendierten zum erreichten Curriculum.....	256
4.6.2	Einstellungen zum Kerncurriculum im Überblick	264
4.6.3	Kenntnisse und Einstellungen zu den Materialien für kompetenzorientierten Unterricht	268
4.6.3.1	Rahmenbedingungen	268
4.6.3.2	Verbreitung, Verwendung und Nutzen der Materialien.....	269
4.6.4	Kerncurriculum und Fachgruppenarbeit	272
4.6.4.1	Perspektiven und Phasen des Implementierungsprozesses.....	272
4.6.4.2	Perspektiven für die Umsetzung curricularer Vorgaben im Bereich der Leistungsbewertung.....	279
4.6.5	Die VERA-Lernstandserhebungen als Möglichkeit der Unterrichtsentwicklung	282
4.7	Perspektiven für die Fort- und Weiterbildung im Bereich <i>Assessment</i>	290
4.7.1	Aufbau diagnostischer Kompetenz– erste Überlegungen zu einem Curriculum	293
4.7.2	Fortbildungswünsche im Bereich <i>Assessment</i>	299
4.7.3	Erwartungen an ein Training im Bereich der Diagnostik.....	303
4.7.4	Faktoren für mehr Berufszufriedenheit im Bereich <i>Assessment</i>	308
4.8	Einzelfallanalysen und Typenbildung	310
4.8.1	Vorüberlegungen.....	310
4.8.2	Einzelfallanalysen	311
4.8.2.1	Einzelfallanalyse S17	312

4.8.2.1.1	Entwicklung und Praxis eines Konzepts zum Umgang mit Fragen des Assess- ments.....	312
4.8.2.1.2	Umgang mit Assessment – Alltag und Perspektiven.....	319
4.8.2.1.3	Zusammenfassung und Typenbildung	325
4.8.2.2	Einzelfallanalyse S10	326
4.8.2.2.1	Entwicklung und Praxis eines Konzepts zu Fragen des Assessments.....	326
4.8.2.2.2	Umgang mit Assessment – Alltag und Perspektiven.....	329
4.8.2.2.3	Zusammenfassung und Typenbildung	335
4.8.3	Darstellung, Kontrastierung und Zuordnung der Typen	336
4.9	Zusammenfassung der Ergebnisse	344
5	Ergebnisse und Schlussfolgerungen.....	351
5.1	Kritischer Rückblick auf die Untersuchungsstrategie	351
5.2	Beantwortung der Forschungsfrage und Bestimmung des Geltungsbereichs	355
5.3	Prägungen, Praktiken und Perspektiven in Fragen des <i>Assessments</i> ...	358
5.3.1	Erste Ergebnisse im Überblick.....	358
5.3.2	Biographische Prägungen durch <i>Assessment</i> im eigenen Spracherwerbsprozess	363
5.3.3	Die <i>Assessment</i> -Praxis der Befragten: Theorie und Praxis, Ideal und Realität	370
5.3.4	Curriculum, Kooperation und Professionalisierung	379
5.3.5	Reformbereitschaft und Vision eines „neuen“ Fremdsprachenunterrichts	383
5.4	Möglichkeiten und Grenzen der Theoriebildung.....	389
5.5	Forschungsperspektiven	395
5.6	Implikationen für Lehrerbildung und Fachdidaktik	397
6	Literaturverzeichnis	403
Anhang 1:	Erhebungsbogen berufsbiographischer Daten.....	431
Anhang 2:	Interviewleitfaden der Hauptstudie	432
Anhang 3:	Beispiel einer Kodierung.....	436
Anhang 4:	Einstellungen der Befragten zum Kerncurriculum.....	441
Anhang 5:	Beispiel einer synoptischen Darstellung (Themenblock 1 für die Befragten S1-S3)	443

Abkürzungsverzeichnis

DESI	Deutsch Englisch Schülerleistungen International
GER	Gemeinsamer Europäischer Referenzrahmen für Sprachen
IQB	Institut zur Qualitätsentwicklung im Bildungswesen
KAL	Knowledge about Language
KC	Kerncurriculum [in Tabellen und Fußnoten]
KMK	Kultusministerkonferenz
MARKUS	Projekt Mathematik-Gesamterhebung Rheinland-Pfalz: Kompetenzen, Unterrichtsmerkmale, Schulkontext
OECD	Organisation für wirtschaftliche Zusammenarbeit und Entwicklung
OLAW	Kooperation Oldenburg-Leer-Aurich-Wilhelmshaven
PISA	Programme for International Student Assessment
PLG	Professionelle Lerngemeinschaft
QTS	Qualified Teacher Status
SALVE	Systematische Analyse des Lernverhaltens und des Verständnisses in Mathematik: Entwicklungstrends und Fördermöglichkeiten
SLA	Second Language Acquisition
SLTE	Second Language Teacher Education
VERA 8	Vergleichsarbeiten Jahrgangsstufe 8

Abbildungsverzeichnis

Abbildung 1:	Kumulativer Kompetenzaufbau	52
Abbildung 2:	Beispiel eines CAN-DO-Statements aus dem GER nach Quetz	54
Abbildung 3:	Kompetenzbereiche der Bildungsstandards	56
Abbildung 4:	„Coordinating assessment and evaluation“	66
Abbildung 5:	Übersicht der Transkriptionsregeln nach Kuckartz.....	140
Abbildung 6:	Skala zur Direktheit des Interviewverhaltens nach Cohen, Manion und Morrison.....	142
Abbildung 7:	Übersicht der Aufgaben der Fachkonferenz	160
Abbildung 8:	Arbeitsschritte im Auswertungsprozess	165
Abbildung 9:	Idealvorstellungen zum Thema „Diagnostik“	218
Abbildung 10:	Ziele, Nachweise und Anhaltspunkte in den Bereichen Lehrerprofessionalität und Lehrerverkooperation gemäß Orientierungsrahmen	247
Abbildung 11:	Schulqualität gemäß Orientierungsrahmen	254
Abbildung 12:	Vom intendierten zum praktizierten, zum gelernten Curriculum	257
Abbildung 13:	Verbesserungsvorschläge im Bereich der Curriculumentwicklung und -implementierung	267
Abbildung 14:	Kriterien einer wissenschaftlichen Diagnostik.....	294
Abbildung 15:	Bausteine einer diagnostischen Wissensbasis nach Helmke.....	295
Abbildung 16:	Diagnosezyklus nach Helmke	298
Abbildung 17:	„Curiously separate“	313
Abbildung 18:	„The end-of-unit test model“	313
Abbildung 19:	Lernprinzipien und <i>Assessment</i> nach Jones und Wiliam ...	318
Abbildung 20:	„Assessment within teaching and learning“	319
Abbildung 21:	„Feedforward“	339

Tabellenverzeichnis

Tabelle 1:	Abstraktionsebenen im Forschungsprozess.....	106
Tabelle 2:	Gegenüberstellung der Themenblöcke Mayer (2002) und der vorliegenden Studie.....	110
Tabelle 3:	Verschiedene Interviewtypen nach Helfferich	115
Tabelle 4:	Übersicht zu Tätigkeitsbereichen des Experten-Samples.....	123
Tabelle 5:	Sechs Schritte der Auswertung nach Meuser und Nagel.....	127
Tabelle 6:	Auswertung von Experteninterviews nach Gläser und Laudel	129
Tabelle 7:	Modell empirisch begründeter Stufenbildung nach Kelle und Kluge.....	133
Tabelle 8:	Phasen und Schritte im Forschungsprozess.....	135
Tabelle 9:	Altersstruktur der Befragten	157
Tabelle 10:	Übersicht der Fächerkombinationen der Befragten.....	157
Tabelle 11:	Übersicht der Zusatzfunktionen der Befragten.....	159
Tabelle 12:	Der Einfluss der Fremdsprachenlehrkräfte aus der eigenen Schulzeit.....	175
Tabelle 13:	Konstruktion eines Falles zu klaren Vorstellungen über gerechte Leistungsbewertung	205
Tabelle 14:	Umgang mit Leistungsbewertung: Ideal und Realität	221
Tabelle 15:	Hindernisse auf dem Weg zum Ideal.....	226
Tabelle 16:	Übersicht zu Bewertungsmaßstäben	227
Tabelle 17:	Merkmale kollegialer Zusammenarbeit.....	238
Tabelle 18:	Bewertungskriterien für eine sprachliche Leistung im Bereich Mediation	242
Tabelle 19:	Standards für die Lehrerbildung, Beschluss der Kultusministerkonferenz vom 16.12.2004	248
Tabelle 20:	Hemmende Faktoren bei der Umsetzung kompetenzorientierten Unterrichts nach Schratz und Westfall-Greiter, übertragen auf den Fremdsprachenunterricht.....	259
Tabelle 21:	Merkmale kompetenzorientierten Fremdsprachenunterrichts nach Tesch im Vergleich zu curricularen Vorgaben in Niedersachsen	262

Tabelle 22:	Bekanntheitsgrad und Verwendung der Materialien für kompetenzorientierten Unterricht Englisch.....	270
Tabelle 23:	Umgang mit dem Kerncurriculum innerhalb der Fachgruppe....	274
Tabelle 24:	Verbesserungsvorschläge zur Umsetzung von Reformen an die Adresse des Kultusministeriums	279
Tabelle 25:	Rezeption der VERA 8 Lernstandserhebungen im Fach Englisch durch die Befragten.....	284
Tabelle 26:	Fortbildungswünsche der Befragten nach Themen geordnet	299
Tabelle 27:	Kontrastierung zweier Typiken	336

Vorwort

Das Interesse für das mit dieser Dissertationsschrift vorliegende Forschungsprojekt zum Umgang mit Fragen des *Assessments* entstand vor allem aus meiner Auseinandersetzung mit eben dieser Thematik in Schule, Universität und Studienseminar. Dem Engagement sowie der Diskurs- und Reflexionsfähigkeit meiner 20 Interviewpartnerinnen und -partnern verdanke ich die empirische Basis meiner Forschung. Allen Lehrkräften möchte ich ausdrücklich für die Bereitschaft danken, ihre Erfahrungen und Erkenntnisse mit mir zu teilen.

Danken möchte ich ganz besonders meinem Doktorvater, Herrn Prof. Dr. Wolfgang Gehring, der über vier Jahre den Entstehungsprozess der Dissertation sehr verlässlich, kompetent und hilfreich betreut hat. Frau Prof. Dr. Rita Kupetz bin ich für die konstruktive und effiziente Begleitung gerade in der Abschlussphase meines Projekts sehr verbunden. Vielfältige Anregungen für die thematische Ausrichtung meiner Arbeit erhielt ich vom Leiter des Studienseminars Leer, Herrn Prof. Dr. Johann Sjuts. Ihm möchte ich zudem für sein Vorbild und die Ermutigung danken, die Dissertation berufsbegleitend in Angriff zu nehmen.

Herrn Dr. Peter Vollmers, Frau Simone Krächter und Frau Frauke Thees danke ich herzlich für ihre äußerst kompetente, verlässliche und geduldige Arbeit am Manuskript. Ihre freundlichen und freundschaftlichen Kommentare habe ich stets als konstruktiv und motivierend empfunden.

Schließlich möchte ich mich auch für die Unterstützung in meinem privaten Umfeld bedanken. Dieser Dank gilt meiner Mutter, Helga Erdmann, der gesamten Familie sowie zahlreichen Freundinnen und Freunden. Auf meine eigenen Schulzeit zurückblickend, möchte ich die Förderung durch meinen damaligen Lehrer, Herrn Carl-Clemens Andresen, besonders hervorheben und ihm nach all den Jahren noch einmal meinen Dank aussprechen.

Widmen möchte ich diese Arbeit meinem Mann, Jan Janssen Bakker, der mein Projekt in allen Phasen liebevoll, fröhlich und konstruktiv unterstützt hat. Sein Interesse, seine Zuversicht und seine Gelassenheit bildeten die Grundlage meines berufsbegleitenden Studiums und trugen wesentlich zu unserer geteilten Freude am Entstehungsprozess der Dissertation bei.

Dezember 2012

Sylke Bakker

1. Einleitung

1.1 Lernen, Lehren und Bewerten vor neuen Horizonten

Spätestens seit Veröffentlichung der Ergebnisse der ersten PISA-Studie (*Programme for International Student Assessment* [PISA]) im Jahre 2000 hat die Sicherung und Entwicklung von Unterrichtsqualität an deutschen Schulen enorm an Bedeutung gewonnen. Die Frage, wie das deutsche Schulsystem leistungsfähiger werden könnte, hat zu Überlegungen auf der Ebene des Systems, aber auch auf der Mikroebene der Lehrkollegien und der Fachgruppen und nicht zuletzt auf Seiten der Lehrkräfte, Eltern und Schüler geführt. Der Bildungsforscher Fend spricht gar von „kleinen Revolutionen“, die sich „post-PISA“ im Nachdenken über die Ursachen der Leistungsfähigkeit von Bildungssystemen vollzogen hätten (Fend 2008: 13).

Bedingt durch die föderale Bildungslandschaft in Deutschland, obliegt es den einzelnen Bundesländern, Schulqualität zu sichern und dabei auch wissenschaftliche Erkenntnisse, etwa zur Kompetenzorientierung, einzubeziehen. In Niedersachsen findet die Neuausrichtung des Unterrichts ihren Niederschlag in Dokumenten wie dem Orientierungsrahmen Schulqualität (Niedersächsisches Kultusministerium 2006e) und in der Umstellung der „alten“ Rahmenrichtlinien auf neue, an Kompetenzen und Output orientierte Kerncurricula in allen Fächern. Eines der ersten vorliegenden und gültigen Kerncurricula in einem Hauptfach ist das für die Sekundarstufe I im Fach Englisch (Niedersächsisches Kultusministerium 2006a), welches für die Sekundarstufe II im Jahre 2010 fortgeschrieben wurde. Länderübergreifend einigte man sich auf kultusministerieller Ebene auf Bildungsstandards für die Fächer Deutsch, Mathematik und die erste Fremdsprache (Englisch oder Französisch) und ist dabei, bundesweit einheitliche Anforderungen für das Abitur (EPA) zu erstellen. Neben der Annäherung hinsichtlich der Kompetenzen, welche von Schülerinnen und Schülern erworben werden sollen, liegen auch Standards für die Lehrerbildung vor, auf die sich die Kultusministerkonferenz einigte (Kultusministerkonferenz, Beschluss vom 16.12.2004).

Einen internationalen Rahmen für das Thema der Schulqualität und Leistungsfähigkeit von Bildungssystemen bieten neben Schulleistungstests auch

Standards zur Lehrerbildung, etwa aus dem angelsächsischen Raum oder auch der Schweiz (vgl. Helmke 2009: 143-151). Auf Schulebene oder für einzelne Kolleginnen und Kollegen finden sich Handreichungen zur Evaluation und Selbstevaluation von Unterricht mit dem Ziel einer reflektierten Praxis (vgl. Farrell 2007a). Die in angelsächsischen Ländern recht verbreitete Erforschung des eigenen Unterrichts, *Action Research*, ermöglicht eine Diagnose und Evaluation, welche dann eine gezielte Qualitätsverbesserung bewirken soll (vgl. Burns 2009; Altrichter & Feindt 2008).¹

Damit ein gezielter Blick auf Lehren und Lernen an deutschen Schulen überhaupt möglich wird und Evaluationszirkel in Bewegung gesetzt werden können, sind laut Fend (2008) unterschiedliche Betrachtungsebenen von Belang: die Makroebene des Bildungssystems als Ganzes, die Mesoebene der Schulen in ihrem lokalen Umfeld und die Mikroebene der Lehrkräfte und ihres professionellen Handelns (Fend 2008: 17). Die Komplexität dieser Ebenen und Bedingungsgefüge erschweren Veränderungen im System, weshalb die Erkenntnis von Verbesserungsbedarf im Hinblick auf Schülerleistungen, in extremer Weise beim „PISA-Schock“, zunächst auf Unterricht und Lernleistungen der Schülerschaft scheinbar kaum Auswirkungen hatte.

Die Tatsache, dass Ergebnisse von PISA „schockierten“, weist auf die vergleichsweise geringe Bedeutung von Diagnostik in den Jahren vor 2000 hin (Helmke 2009: 269). Die mittlerweile in den KMK-Standards verankerte professionelle Kompetenz des Bewertens und Beurteilens tauchte zwar bundesweit in den jeweiligen Richtlinien auf und war bereits seit den siebziger Jahren Thema in der pädagogischen Fachliteratur; der Effekt auf die erste und zweite Phase der Lehrerausbildung sowie die Lehrerfortbildung erschien jedoch eher gering. Kenntnisse und Fähigkeiten im Bereich der Testerstellung, des Anfertigens und der Interpretation von Bewertungsskalen, der Nutzung von Computertechnologie zur Berechnung von Zensuren sowie der Beherrschung verschiedener Feedback-Techniken sind beispielsweise in Großbritannien verpflichtend (vgl. Fautley & Savage 2008: 1) und werden unter den Standards *Assessing, Monitoring and Giving Feedback* zusammengefasst.² Das in Niedersachsen verbindliche Kerncurriculum für die Sekundarstufe I beschränkt sich hingegen in seinen Ausführungen zur „Leistungsfeststellung und Leistungsbewertung“, beispielsweise im

1 Die Frage, „what actually works in schools to improve learning“, beantwortet der neuseeländische Forscher John Hattie mit seiner Synthese von über 800 Meta-Analysen zum Thema Schülerleistungen (Hattie 2009). Er liefert so auch deutschen Forschern und interessierten Praktikern Anhaltspunkte zur Evaluation von Lernen in schulischen Kontexten.

2 Eine detaillierte Darstellung der QTS-Standards findet sich auf dieser Seite des Bildungsministeriums http://www.tda.gov.uk/trainee-teacher/qts-standards.aspx?sc_lang=en-GB [Letzter Zugriff am 10.12.2012].

Fach Englisch, auf Ausführungen von lediglich zwei Seiten (Niedersächsisches Kultusministerium 2006a: 26 f.).³

Das Thema „Leistungsfeststellung und Leistungsbewertung“ – oder in angelsächsischer Terminologie *Assessment* – nimmt meiner Einschätzung nach eine Schlüsselstellung im Bereich der Unterrichtsentwicklung und der Lehreraus- und -weiterbildung ein. Wenn auf Lehrerseite die Fähigkeit besteht, Schülerleistungen präzise einzuschätzen, d.h. eine hohe diagnostische Kompetenz vorliegt, kann adaptiver, auf die konkreten Bedürfnisse der jeweiligen Lernenden und Lerngruppen abgestimmter Unterricht geplant und durchgeführt werden. Die Entscheidungen zur Unterrichtsplanung können auf der Basis einer präzisen Diagnose als „informed decisions“ bezeichnet werden (vgl. Hattie 2012: 26). Neben den im eigenen Klassenraum erhobenen Daten spielen ebenso Befunde von Schulleistungstests, etwa DESI für den Englischunterricht (vgl. Beck & Klieme 2007), eine Rolle. Die Ergebnisse summativer Leistungsmessungen, beispielsweise die VERA-Tests in den Fächern Mathematik, Englisch und Deutsch, bieten neben ihrem Informationsgehalt über den Leistungsstand der Lernenden auch die Möglichkeit einer formativen Nutzung, welche auf Verbesserung des Unterrichts und des Lernens der Schüler zielt.

Zusammenfassend soll noch einmal die Schlüsselstellung, die *Assessment* im Bereich der Unterrichtsentwicklung einnimmt, hervorgehoben werden. Durch eine möglichst präzise Bestandsaufnahme des Ist-Zustands auf der Makro-, Meso- und Mikroebene im Bildungssystem kann Steuerung von Evaluations- und Adaptionsprozessen wesentlich erleichtert werden, da die Ausgangsbasis des Qualitätszirkels transparenter wird als bei bloßen Eindrücken zu Schülerleistungen und somit Leistungsmöglichkeiten des Schulsystems. Wie Helmke in seinen Ausführungen zu Evaluationskultur und Innovationsklima herausstellt, hängt der Qualitätsfortschritt auf der Mikroebene von der an der jeweiligen Schule dem Thema „Evaluation“ entgegengebrachten Wertschätzung ab:

In manchen Schulen sind Leistungsmessung, Schülerfeedback zum Unterricht und andere Formen der Selbstevaluation eine Selbstverständlichkeit, in anderen Schulen finden sich Vorbehalte, Ängste, Abwehr und Tabus. (Helmke 2009: 318)

Die vorliegende Studie nimmt einen spezifischen Bereich in den Blick, in welchem *Assessment* auf nationaler und internationaler Ebene besonders relevant und interessant erscheint: das Feld des Fremdsprachenlernens und -lehrens. Angesichts der

3 In den untersuchten Kerncurricula für das Fach Englisch in Hauptschulen, Realschulen, Integrierten Gesamtschulen sowie Gymnasien finden sich in Wortlaut und Umfang fast identische Texte zu diesem Thema. Im Fach Französisch entsprechen die Bestimmungen sinngemäß denjenigen im Fach Englisch, der Wortlaut ist jedoch jeweils ein anderer. Eine genauere Analyse am Beispiel des Kerncurriculums Englisch erfolgt unter Punkt 2.3.1.

Bestrebungen nach vergleichbaren und praktikablen Maßstäben der Beurteilung von Sprachenlernen in Europa durch den Gemeinsamen Europäischen Referenzrahmen für Sprachen (GER) (Europarat 2001) sowie der Erfahrungen mit Vergleichsarbeiten wie VERA und auch der Erkenntnisse aus der umfangreich und komplex angelegten DESI-Studie (Beck & Klieme 2007) werden auf Bundes- und Länderebene Bemühungen deutlich, sich über Fragen der Leistungsbewertung im Fremdsprachenunterricht zu verständigen.

Durch die Umstellung von inputorientierten hin zu kompetenzorientierten Unterrichtskonzepten stellt sich die Frage nach der Bewertung von fremdsprachlichen Leistungen für die in der fremdsprachlichen Ausbildung Lehrenden und Lernenden neu. Traditionelle Vorstellungen von sprachlicher Richtigkeit werden angesichts des GER hinterfragt. Auch lassen sich vielfach die aus einem handlungs-, produkt- und prozessorientierten Fremdsprachenunterricht hervorgehenden Schülerleistungen gar nicht mit „klassischen“ Bewertungsmustern der sprachlichen Richtigkeit erfassen. Folglich scheint es angebracht, für die vorliegende Studie eine weite Begriffsdefinition von Leistungsbeurteilung zu wählen:

Das Wort „Assessment“ wird zunehmend in deutschsprachigen Ländern verwendet. „Assessment“ ist im englischen Sprachraum der Überbegriff für Leistungsbeurteilung in der Pädagogik. Das Wort stammt vom lateinischen „assessere“ ab und bedeutet, nicht nur ein Urteil über eine Leistung zu liefern, sondern auch Begleitung und Wahrnehmung neben der Richterrolle. (Schratz & Westfall-Greiter 2010: 98)

Die vorliegende Erhebung beschäftigt sich mit der Rezeption von fachdidaktischen und curricularen Neuerungen im Bereich des Lehrens und Lernens von Fremdsprachen und nimmt dabei aufgrund der Bedeutsamkeit persönlicher schulischer Eindrücke (Wahl 2006: 17) die Erfahrungen mit *Assessment* im eigenen Spracherwerbsprozess zum Ausgangspunkt. In 20 Experteninterviews wird der institutionelle Rahmen von Leistungsbewertung mit persönlichen Erfahrungen der Befragten sowohl in der Schüler- als auch in der Lehrerrolle in Beziehung gesetzt, um auszuloten, welche Vorstellungen der Praxis des Bewertens zugrunde liegen. Überdies soll erforscht werden, welche Anpassungen möglicherweise nötig sind, um die berufliche Zufriedenheit mit der Rolle der Beurteilenden (*Rater*) zu steigern sowie Lehr- und Lernprozesse im Sinne des *Assessment for Learning* (Black & Wiliam 1998) insgesamt zu verbessern.⁴ Um ein möglichst hohes Diskurs- und Reflexionsniveau zu erreichen, werden in der Hauptstudie Lehrkräfte befragt, die in Lehrerausbildung und Fortbildung tätig sind oder eine Fachgruppe leiten und deshalb mit neueren fachdidaktischen Konzepten und den curricularen Rahmenbedingungen vertraut sind. Die Studie ist schulformüber-

4 Bis auf einen Teilnehmer der Pilotstudie unterrichten alle Befragten das Fach Englisch, häufig in Kombination mit einer weiteren Fremdsprache (vgl. 4.1.3).

greifend und wird im Sinne der Vergleichbarkeit ausschließlich in Niedersachsen durchgeführt.

1.2 Aufbau der Dissertationsschrift

Nach einer thematischen Einführung zum Thema *Assessment* und einer ersten Skizzierung der geplanten Interviewstudie erfolgt im **zweiten Kapitel** eine Einordnung in die empirische Unterrichtsforschung. Unterschiedliche Entwicklungslinien innerhalb der qualitativen Sozialforschung werden nachvollzogen, um schließlich eine angemessene Positionierung des Forschungsprojekts vornehmen zu können. Einen weiteren Schritt bildet die Engführung auf Lehr- und Lernforschung im Bereich des Fremdsprachenunterrichts sowie eine Auseinandersetzung mit qualitativen und quantitativen Forschungsansätzen, deren Relevanz für die Studie überprüft wird.

Die theoretische Fundierung zum Thema *Assessment* erfolgt auf mehreren Ebenen. Zunächst werden im Überblick zentrale Erkenntnisse zum Themenkomplex „Diagnostik im Fremdsprachenunterricht“ dargestellt. Dabei kommen sowohl theoretische Aspekte im Bereich des Testens und Evaluierens von Leistungen in der Fremdsprache zum Tragen wie auch die praktische Anwendung im Lande Niedersachsen auf der Basis der gültigen Kerncurricula. Einen weiteren Erkenntnishorizont bildet die Lehrerwissensforschung. Die Frage, wie Lehrkräfte Leistungen in der Fremdsprache beurteilen, wird auf der Folie der Wissens- und Erfahrungsbasis von Fremdsprachenlehrerinnen und -lehrern⁵ beleuchtet. Besonderes Augenmerk gilt dem Expertenstatus von Fremdsprachenlehrkräften. Es soll Aufschluss darüber erhalten werden, was unter „diagnostischer Expertise“ (Helmke 2009: 268) im Bereich des Fremdsprachenunterrichts zu verstehen ist, um für die Erhebung selbst einen Maßstab zu gewinnen.

Das **dritte Kapitel** spezifiziert das Untersuchungsdesign der Studie. Ausgehend von einer Positionierung innerhalb der qualitativen Bildungsforschung werden im Folgenden die Ebenen der Abstraktion im Forschungsprozess entwickelt. In einem anschließenden Prozess der Operationalisierung werden schließlich die Forschungsfragen formuliert und in einem weiteren Schritt in Fragen des Leitfadens überführt. Es schließen sich Überlegungen zur Wahl einer passenden Interviewform an. Das Experteninterview bildet den Ausgangspunkt für konkrete Planungen der Erhebung. Auf dem Hintergrund relevanter Literatur wird erörtert, wie das Experteninterview als Forschungsinstrument in der vorliegenden Studie

5 Im Text der vorliegenden Arbeit wird aus stilistischen Gründen vereinzelt nur ein Geschlecht verwendet, das jeweils synonym für beide Formen gelten soll.

produktiv zur Anwendung kommen kann. Thematisiert wird auch, wie ein Zugang zum Feld erlangt werden soll. Das Verhältnis von Narration und Information in den geplanten Interviews ist ein Punkt, der die Anpassung des zuvor erstellten Leitfadens an die besonderen Anforderungen des Experteninterviews vorbereitet. Allgemeine Überlegungen zur Auswertungsmethodik werden anhand eines Vergleichs qualitativ-interpretativer und deduktiv-nomologischer Verfahren spezifiziert und schließlich zu einer passenden Auswertungsstrategie für die eigene Studie verarbeitet.

Eine weitere Konkretisierung vollzieht sich auf der Ebene der Vorbereitung der Schritte ins Feld. Nach einem Überblick zum Forschungsverlauf folgen Überlegungen zur Nutzung unterschiedlicher Forschungsinstrumente. Der Einsatz eines der Interviewstudie vorgeschalteten Fragebogens wird ebenso erörtert wie die Nutzung und Nützlichkeit eines Erhebungsbogens berufsbiographischer Daten. Der geplante Umgang mit Daten wird auf verschiedenen Ebenen und zu verschiedenen Zeitpunkten der Studie dargestellt. Einen Schwerpunkt der Vorbereitung der Erhebung bildet die konkrete Interviewsituation. Für diese muss zunächst ein allgemeiner Rahmen geschaffen werden, der wissenschaftlichen Anforderungen an eine Erhebung genügt. Eine Reflexion der Subjektivität der Forscherin schließt sich an. Es wird kritisch diskutiert, welche Rolle sie und ihre berufliche Biographie in der Interviewsituation spielen.

Das dritte Kapitel schließt mit der Darstellung und Reflexion der Pilotstudie ab. Dabei werden begründet Schlüsse aus den Erfahrungen im Feld und der ersten Analyse der erhobenen Daten gezogen. Dementsprechend erfolgt eine Anpassung der Planungen für die Hauptstudie in Bezug auf die Stichprobe, das Interview- sowie das Auswertungsverfahren.

Das **vierte Kapitel** gibt zunächst einen Überblick zum tatsächlichen Forschungsverlauf. Dabei spielen sowohl der Zugang zum Feld als auch die Zusammensetzung des Samples eine Rolle. Nach einer Übersicht zur geplanten Vorgehensweise und dem Aufzeigen der Grenzen der Erhebung erfolgt eine begründete Anpassung der Untersuchungsstrategie. Die Darstellung und Auswertung der erhobenen Daten beginnt mit einem Blick auf die biographische Prägung im Bereich der Leistungsbewertung. Dabei kommen unter anderem emotional als bedeutsam empfundene Erfahrungen mit Sprachenlernen und Bewertungen in Kindheit und Jugend zur Sprache. Es wird versucht, das Bedingungsgefüge zwischen unterschiedlichen Arten des *Assessments* und der Motivation ebenso zu erkunden wie die Rolle, die Lehrende in diesem Zusammenhang spielten. Neben dem gesteuerten Fremdsprachenlernen in der Institution Schule wird auch die Rolle von Auslandsaufenthalten und des zumeist ungesteuerten Spracherwerbs dort analysiert. Dabei soll auch Klarheit darüber erlangt werden, ob *Assessment*-Methoden im zielsprachlichen Ausland eine andere Qualität haben,

als dies in Deutschland der Fall ist. Mögliche Schlüsselerlebnisse im Bereich der Leistungsbewertung dienen dazu, die Mechanismen von Bewertungsvorgängen zu erschließen. Auf der Basis der Erfahrungen mit unterschiedlichen Formen des *Assessments* im In- und Ausland sollen Maßstäbe der Befragten zu angemessener, gerechter Leistungsbeurteilung zusammengetragen und analysiert werden.

Der nächste Schritt der Datenauswertung besteht in der Rekonstruktion der Vorstellungen der Befragten von der Durchführung von *Assessments* unter idealen Bedingungen. Erst in einem weiteren Abschnitt geht es um Möglichkeiten und Grenzen kompetenzorientierten Lehrens, Lernens und Bewertens im Vergleich zu traditionelleren Ansätzen. Dabei kommt der GER als Werkzeug ebenso zur Sprache wie die erlebte Praxis kollegialer Kooperation und erste Überlegungen zu Gelingensbedingungen in diesem Bereich. Die Ebene der gültigen curricularen und schulgesetzlichen Vorgaben sowie die vom Kultusministerium veröffentlichten Handreichungen werden im Kapitel zu Curriculumsentwicklung und Professionalisierungsprozessen im Zusammenhang mit Leistungsbewertung erkundet. Dabei wird der Weg vom intendierten zum erreichten Curriculum nachgezeichnet, indem Rahmenbedingungen analysiert und verschiedene Faktoren im Implementierungsprozess zueinander in Beziehung gesetzt und evaluiert werden. Es kommt auch eine neue Aufgabenkultur zur Sprache, die sich z.B. in standardisierten Vergleichsarbeiten wie VERA 8 manifestiert.

Das sich anschließende Unterkapitel eruiert, ausgehend von ersten Überlegungen zu einem Curriculum zur Förderung diagnostischer Expertise bei Fremdsprachenlehrkräften, die Wahrnehmung von Fortbildungsbedarf in diesem Bereich unter den Befragten. Darüber hinaus werden weitere Fortbildungswünsche und Faktoren zur Förderung der Berufszufriedenheit im Zusammenhang mit *Assessment* thematisiert.

Die bisher auf das gesamte Sample bezogenen Ergebnisse der Erhebung werden in einem weiteren Schritt auf zwei zu kontrastierende Fälle fokussiert, anhand derer eine Typenbildung vorgenommen werden soll. Dabei wird aus dem Datenkorpus jeweils das von den beiden Interviewpartnerinnen und -partnern entwickelte Konzept zu Fragen des *Assessments* herausgearbeitet und auf dem Hintergrund der konkreten Praxis im Schulalltag analysiert. Dabei spielen wiederum die Rezeption curricularer Vorgaben und deren Umsetzung in der Fachgruppe eine zentrale Rolle. Anschließend werden die von den Befragten erarbeiteten Zukunftsperspektiven in Bezug auf Leistungsbewertung analysiert. Abschließend erfolgen eine Typenbildung und eine Kontrastierung sowie der Versuch einer Zuordnung der übrigen Interviewpartnerinnen und -partner zu einer der beiden Typiken.

Das **fünfte Kapitel** nimmt bilanzierend die gewählte Untersuchungsstrategie kritisch in den Blick und versucht dann, Aussagen zur Beantwortung der For-

schungsfrage zusammenzutragen. Dabei werden die Ergebnisse der Studie auf der Folie aktueller Forschungsbefunde im Bereich der Fachdidaktik und der Lehrerbildungsforschung analysiert, um schließlich zu einer Theoriebildung zu gelangen. Abschließend werden Forschungsperspektiven und Implikationen für die unterschiedlichen Phasen der Ausbildung von Fremdsprachenlehrkräften aufgezeigt.

2. Theoretische Fundierung der Studie

2.1 Einordnung der Studie in die empirische Unterrichtsforschung

2.1.1 Einflüsse und Entwicklungslinien

Bevor konkrete Schritte ins Feld unternommen werden, erscheint eine Einordnung der Studie in theoretische Zusammenhänge angebracht.

Das qualitative Forschungsparadigma im Bereich der Erziehungswissenschaft (*education*) ist begrifflich sehr weit gefasst und kann viele unterschiedliche Zugänge repräsentieren, die ihrerseits durch die Person der Forscherin und ihrer Weltsicht innerhalb einer pluralistischen Gesellschaft geprägt sind (Punch 2009: 115; Kelle & Kluge 2010: 10). Keith F. Punch unterstreicht in seinen Ausführungen zur Forschungsmethodik die politische Dimension von Bildung und Sozialforschung. Forschungsstile sind folglich nicht „neutral“ und „objektiv“, was besonders angesichts postmoderner Ansätze, die Modelle von Ursache und Wirkung ablehnen, deutlich wird (Punch 2009: 83).

Diese Auffassung findet sich auch bei Pierre Bourdieu, dessen Vorstellung der *Bias* im Sinne einer gesellschaftlich geteilten und unbewussten Voreingenommenheit das wissenschaftliche Vorgehen *per se* einer kritischen Analyse unterzieht und eine bewusste Distanzierung von einer eingeschränkten, vorurteilsbehafteten Sichtweise fordert (Prenzel et al. 2010: 22 f.). Als Schutz vor unnötigen Verzerrungen in der eigenen Forscherperspektive bietet sich eine Offenlegung des gesamten Forschungsprozesses an, um einerseits für Transparenz zu sorgen und andererseits die Begrenztheit der eigenen Erkenntnismöglichkeiten aufzuzeigen. Im Zuge der historischen Entwicklung der Wissenschaft wurden und werden neue Untersuchungsinstrumente entwickelt, um Komplexität in sozialen Handlungsfeldern angemessen abzubilden und sich somit gegen die Annahme eines „naiven Empirismus“ zu positionieren (Prenzel et al. 2010: 22 f.).

Auch Hedda Bennewitz konstatiert eine breite Vielfalt von qualitativen Ansätzen, „die sich in ihrem Gegenstandsverständnis und ihrem methodischen Vorgehen deutlich unterscheiden“ (Bennewitz 2010: 45; vgl. auch Kelle & Kluge

2010: 11). Sie verweist auf die Bedeutungszuschreibung und Interpretation sowie Interaktion der in einem sozialen Feld Handelnden, welche soziale Wirklichkeit erst konstituieren (Bennewitz 2010: 45). Sozialwissenschaftliches Forschen bedeutet somit im Sinne von Alfred Schütz eine Konstruktion zweiter Ordnung, indem eine Rekonstruktion der Konstruktion sozial Handelnder erfolgt (Bennewitz 2010: 45). Diese Pluralität begründet sich in dem zunehmenden Einfluss der Sozialwissenschaften und deren Forschungsmethodik seit den sechziger und siebziger Jahren des 20. Jahrhunderts. Im Zuge einer „Alltagswende“ rückte „Verstehen“ von Handelnden in einem sozialen Feld zunehmend in den Mittelpunkt und knüpfte so an „Traditionen des Symbolischen Interaktionismus, der Phänomenologie, der Ethnomethodologie und der Ethnologie [und auch an] hermeneutisch[e] Traditionen“ an (Prenzel et al. 2010: 20).

Die vorliegende Studie ist vor dem Hintergrund verschiedener wissenssoziologischer Konzepte zu betrachten, deren Relevanz es jeweils auszuloten gilt. Im Bereich der pragmatischen Wissenssoziologie sowie des symbolischen Interaktionismus und sozialen Behaviorismus sind besonders die Arbeiten des Amerikaners George Herbert Mead (1863-1931) zu nennen, welche die normative Funktion der Gruppe als Resultat der Instanz des „generalisierten Anderen“ beschreiben und somit von der „Verankerung kollektiver Repräsentation als soziale[m] Prozess mit sozial-ordnungsstiftender Wirkung“ ausgehen (Maasen 2009: 28). Die Kategorie der Lern- und Berufsgruppe spielt in der konzipierten Studie ebenso wie die Frage nach Wissensrepräsentationen und deren Ordnung sowie deren Formation in sozialen Systemen und Institutionen durchgängig eine Rolle.

Einen weiteren relevanten Bezug bietet Meads Vorstellung, dass die Anwendung des optisch-physikalisch geprägten Begriffs der Perspektivität „ein unerwartetes Geschenk der Physik an die Philosophie“ sei (Mead zitiert nach Prenzel et al. 2010: 26). Prenzel und Mitarbeiter nennen im Hinblick auf die Perspektivität einige zentrale Aspekte, mit denen eine Ausschärfung der qualitativen Forschung möglich werde: „Größendimension, Standort, Blickrichtung, motivationaler Grundzug, Horizont und Dynamik“ (Prenzel et al. 2010:25). Forschende können mit diesem Instrumentarium kontinuierlich ihre Sichtweise auf das Forschungsfeld und den Forschungsgegenstand überprüfen und so beispielsweise den Wechsel zwischen Mikro- und Makroperspektive gewinnbringend für den Erkenntnisprozess nutzen, um somit Komplexität vielfältiger, aber auch fokussierter zu erfassen.

Karl Mannheim (1893-1947) nutzt laut Maasen zwar nicht das „Potential der Sozialphänomenologie für [seine] Wissenssoziologie“ (Maasen 2009: 32), er liefert jedoch mit seinem Postulat der „Seinsadäquatheit des Denkens“ (Maasen 2009: 26) insofern einen Anknüpfungspunkt, als er das Konzept der „freischwebenden Intelligenz“ entwickelt, d.h. einer Personengruppe, welche über eine

mehrperspektivische Wahrnehmung gesellschaftlicher Gruppen verfügt und somit eine Vermittler- und Synthetisierungsfunktion einnimmt (Maasen 2009: 25).

Des Weiteren basiert die „Dokumentarische Methode“ auf Mannheims Unterscheidung zwischen kommunikativem und konjunktivem Wissen. Hinsichtlich des Zugriffs auf Konstruktionen sozialer Wirklichkeit vertritt Mannheim die Vorstellung, „dass im Alltag bereits Methoden gegeben sind, über die wir sozusagen intuitiv verfügen und die es zu rekonstruieren gilt“ (Bohnsack 2008: 23). Gemäß den Vorstellungen Mannheims sind sozialwissenschaftliches Denken und Handeln stets bereits vorstrukturiert, sodass es dem Forscher obliegt, die zugrunde liegenden Entwürfe nachzuzeichnen. Bohnsack nennt diesen Vorgang „verstehende[n] Nachvollzug“, welcher möglich werde, da Wissen und Handeln „typengeleitet, wissensgeleitet, entwurfsorientiert“ seien (Bohnsack 2008: 23).

Sowohl Mannheim als auch Schütz (1899-1959) sind im Gegensatz zu Karl Popper der Auffassung, dass sich naturwissenschaftliche und sozialwissenschaftliche Forschung deutlich unterscheiden. Die Begründung liege in der besonderen Qualität des Gegenstandsbereichs und dem Verhältnis von Gegenstand und Theorie (Bohnsack 2008: 22). Schütz geht es um die Herstellung von „Sinn in der sozialen Welt“ (Maasen 2009: 32). Er nimmt an, dass durch „Mechanismen der Typisierung und Idealisierung“ Wissen innerhalb einer Gesellschaft organisiert, verbreitet und eingebunden wird und somit „Lebenswelten“ bestimmt. Die resultierenden Konstrukte seien jeweils „typisierend und idealisierend“ und spiegeln demnach nicht nur das Wissen „jedermanns“, sondern vielmehr deren Grundlage wider (Maasen 2009: 32). Schütz' phänomenologische Wissenssoziologie kann insofern kritisiert werden, als die angestrebte intersubjektive Grundlage der Genese von Wissen nicht dahingehend genutzt werden kann, über eine „individualistisch[e] Perspektive“ hinauszukommen (Maasen 2009: 33). Den Anknüpfungspunkt für die vorliegende Studie liefert Schütz nicht nur durch seine Überlegungen zu der Entstehung und Beschaffenheit von Wissen, sondern auch durch die Unterscheidung von „Experten“, dem „Mann auf der Straße“ und dem „gut informierten Bürger“. Laut Schütz wird erst durch die institutionelle Zuschreibung von Kompetenz zur „Konstruktion von Wirklichkeit“ der Expertenstatus konstituiert (Meuser & Nagel 2010: 461).

Pierre Bourdieu (1930-2002) entwickelte in Abgrenzung zu Phänomenologie und Objektivismus eine „praxeologische Sichtweise, der es um die gruppenspezifischen Erzeugungsbedingungen von Praxisformen und Repräsentationen geht“ (Maasen 2009: 46). Die von Bourdieu entwickelten Begriffe „Habitus“ und „Doxa“ sind für mein Forschungsprojekt relevant, da es bei der Erforschung von Lehrer-kognitionen im Praxisfeld selbst und in der Reflexion derselben um mentale Strukturen geht, die „wie eine *Handlungs-, Wahrnehmungs- und Denkmatrix* funktionier[en] und dank der analogen Übertragung von Schemata [...] es ermög-

lich[en], unendlich differenzierte Aufgaben zu erfüllen“ (Bourdieu zitiert nach Maasen 2009: 47).

„Doxa“ bezieht sich auf implizites Wissen, welches Bourdieu als „praktisches Wissen“ bezeichnet. Ein phänomenologischer Ansatz wird sowohl „Habitus“ als auch „Doxa“ insofern nicht gerecht, als beide im Sinne des generativen Prinzips auf jeweils unterschiedliche Schemata bzw. Objekte angewendet werden können (Polythetie und Polysemie), weshalb Wissen in mehrfacher Hinsicht als dynamisch betrachtet werden kann, da es sowohl geordnet als auch sich selbst ordnend auftritt und gemäß den spezifischen Interessen der Wissensträger eine jeweils spezifische Form annimmt (Maasen 2009: 49). Die in dieser Studie thematisierten Antinomien des Lehrerhandelns lassen sich mit Bourdieus Vorstellung der „Doxa“ in Beziehung setzen, da auch er von Dissonanzen spricht, welche die ansonsten impliziten Wissensbestände erst in Frage stellen und somit bewusst machen (vgl. Maasen 2009: 49).

Nikola Mayer bezieht sich in ihren Überlegungen zur Interviewsituation ebenfalls auf Bourdieu, welcher in seiner Monographie „Das Elend der Welt“ von 1997 die Vorstellung des Interviews als „geistige[r] Übung“ entwickelt und laut Mayer auf methodologischer Ebene die Subjektivität von Forscherin und Gesprächssituation betont, woraus sich ein „ganzheitliches Verständnis“ ergibt, welches Mayer als Orientierung im Bereich der Forschungsethik und der Interviewtechnik dient (Mayer 2002: 114 f.).

Obschon meine Studie zu Vorstellungen über Leistungsbewertung im Fremdsprachenunterricht weniger „ganzheitlich“ orientiert ist, da statt der von Mayer verwendeten narrativen Interviews die Form der Experteninterviews gewählt wird, spielen auch hier die Subjektivität der Interviewsituation, aber auch die subjektiv geprägte Sicht und Erfahrung und Praxis der Leistungsbewertung sowie die „wissenschaftliche Intimität“ eine Rolle (vgl. Mayer 2002: 116). Eine klare Abgrenzung muss jedoch hinsichtlich Bourdieus „therapeutischen Anspruchs der Interviewsituation“ vorgenommen werden, da die in meinem Forschungsprojekt verwendete Form des Experteninterviews sich darauf beschränkt, einen bestimmten Weltausschnitt aus dem Blickwinkel einer ausgewählten Gruppe von Wissensträgern zu erforschen, und es nicht darum geht, die Ebene einer „Selbstanalyse“ in der Interviewsituation herzustellen (Mayer 2002: 121).

Auf Basis des symbolischen Interaktionismus Meads lieferte die von Glaser und Strauss Ende der sechziger Jahre publizierte *Grounded Theory* ein forschungspraktisches Gerüst, welches in Abgrenzung zu hypothesenprüfenden Verfahren und Methodologien, wie etwa dem Kritischen Rationalismus, die Alltagsintuition für Theoriebildung in den Vordergrund stellte. Während ethnomethodologische Zugänge bewusst zwischen der rekonstruktiven Haltung des Sozialforschers einerseits und den lediglich auf „Alltag“ gerichteten Einstellungen

andererseits unterschieden, galt für Glaser und Strauss die „Alltagsrationalität“ als konstitutiv für kreatives und innovatives Forschen (Bohnsack 2008: 27). Zentral für den Forschungsansatz der *Grounded Theory* sind Gegenstandsangemessenheit und Gegenstandsbezogenheit im Sinne einer möglichst objektiven und genauen Erfassung der Problemstellung und des Forschungsgegenstands sowie eine sichere Fundierung im Datenmaterial, aus welchem Hypothesen „emergieren“ und nicht im Vorfeld bereits quasi feststehen sollen (vgl. Bennewitz 2010: 47).

Kelle und Kluge sehen eine Stärke qualitativer Verfahren wie der *Grounded Theory* grundsätzlich in der „hypotheseengenerierende[n], heuristische[n], explorative[n] Funktion“ (Kelle & Kluge 2010: 11), kritisieren jedoch das ihrer Meinung nach vorliegende „induktivistische Selbstmissverständnis der qualitativen Methodenlehre“, als Forscher möglichst unvoreingenommen und unbeeinflusst von Wissensbeständen an das Datenmaterial herangehen zu können (Kelle & Kluge 2010: 13). Die Autoren halten jegliche Art von Wahrnehmung für stets auf irgendeine Weise vorgeprägt und sehen überdies ein Problem in der Vorstellung, man könne als Forschende theoretische Konzepte quasi automatisch aus dem Datenmaterial „emergieren“ lassen (Kelle & Kluge 2010: 19).

2.1.2 Methodologische Verortung des Forschungsprojekts

Bei der Konzeption der vorliegenden Interviewstudie ist zunächst auf allgemeiner Ebene das Erkenntnisinteresse zu bestimmen, aus welchem sich im konkreten methodischen Zuschnitt Forschungsfragen ableiten und operationalisieren lassen. In einem zweiten Schritt gilt es, über allgemeine methodologische Überlegungen hinaus die besondere Perspektive der Fremdsprachendidaktik mit einzubeziehen und die konkrete Forschungsmethodik entsprechend zu konzipieren.

Im Sinne Annedore Prengels und ihrer Mitarbeiter zeichnet sich das zu planende Forschungsprojekt durch „die Suche nach dem Wissen, den Handlungsweisen und des [*sic*] Sinndeutungsmustern lokaler Akteure im Erziehungs- und Bildungswesen und den unbewussten Motiven und unbewussten regelhaften Strukturen“ aus (Prengel et al. 2010: 28). Bezogen auf die vier Erkenntnismodi wissenschaftlicher Forschung (Prengel et al. 2010: 34) handelt es sich vorwiegend um ein qualitativ geprägtes Forschungsprojekt, wobei im Sinne der Triangulation vereinzelt auch quantitative Verfahrensweisen zum Einsatz kommen. Obschon die Interviewstudie im Zentrum steht, werden auch Daten zum Lebenslauf und zur beruflichen Tätigkeit gesammelt, welche mit Hilfe von quantitativen Methoden ausgewertet werden. Gleiches gilt für quantifizierende Verfahren bei der Kodierung und Auswertung des komplett transkribierten Interviewtextes.

Da es zu erkunden gilt, wie Fremdsprachenlehrerinnen und -lehrer über *Assessment* denken, welche Erfahrungen sie in konkreten Berufszusammenhängen gemacht haben, wie sie diese bewerten, und nicht zuletzt wie sie Zukunftsperspektiven für den produktiven Umgang mit Leistungsbewertung im Fremdsprachenunterricht sehen, bietet es sich an, den Schwerpunkt der Studie qualitativ auszurichten:

Qualitative Forschung sucht die Annäherung an komplexe Einzelfälle, sie interessiert sich für das Zusammenspiel von Makro-, Meso- und Mikroebene. Sie widmet sich den subjektiven und biographisch entstandenen Wirklichkeiten der Adressaten und Akteure pädagogischer Arbeit, den sozialen Interaktionen, den Konstruktionen von Sinn und Bedeutung im Leben von Einzelnen, den Gruppenprozessen und kollektiven Mustern. Es geht darum Denken, Fühlen, Handeln oder Nicht-Handeln von Menschen zu verstehen. (Prenzel et al. 2010: 34)

Neben der Sammlung einiger berufsbiographischer Daten ist zum Verständnis der institutionellen Rahmenbedingungen und Vorgaben, welche die gesetzliche Basis für das professionelle Handeln der befragten Fremdsprachenlehrkräfte bilden, eine Analyse der curricularen Vorgaben und der schulgesetzlichen Bestimmungen nötig. Zu den Datensätzen der Interviews und dem berufsbiographischen Erhebungsbogen müssen folglich auch Auszüge aus Kerncurricula sowie aus anderen relevanten kultusministeriellen Dokumenten hinzugezogen werden, um auch die bereits erwähnte Makro- und die Mesoebene in den Blick nehmen zu können.

„Verstehen als Basishandlung sozialwissenschaftlicher Forschung“ (Gläser & Laudel 2009: 32) spielt als Mittel der Erkundung des speziellen Weltausschnitts rund um Kognitionen und Praxis des *Assessments* eine zentrale Rolle. Daher bilden die Daten der Interviews auch den umfangreichsten und bedeutendsten Teil des Datenkorpus'. Das Prinzip des Verstehens ist im Verbund mit den Prinzipien der Offenheit sowie des theorie- und regelgeleiteten Vorgehens aufzufassen (Gläser & Laudel 2009: 30 ff.). Dabei geht es im Sinne Heinz-Hermann Krügers und Ulrike Deppes zwar auch um „Lerngeschichten“ (Krüger & Deppe 2010: 62), da der Ausgangspunkt der Interviews eine Bestandsaufnahme der persönlichen Erfahrungen mit Bewertungen des eigenen Sprachlernprozesses ist und zudem im zweiten Themenblock möglicherweise Lerngeschichten der eigenen Schülerinnen und Schüler zur Sprache kommen. Der Zugang umfasst jedoch lediglich narrative Elemente und ist grundsätzlich eher auf die möglichst präzise Erkundung eines bestimmten Realitätsbereichs durch gezielte Informationsentnahme ausgerichtet.

Im Vordergrund steht also nicht die Person selbst, sondern vielmehr deren Funktion. Der Begriff der „Ganzheitlichkeit“, so wie ihn einerseits Nikola Mayer (2002) und andererseits Krüger und Deppe (2010: 63) prägen, ist folglich nur bedingt anwendbar. Andererseits erscheint es für das Verständnis der Genese

von Lehrerkognitionen und möglicherweise daraus resultierenden Handlungs- und Praxismustern ergiebig, die Sprachlerngeschichte der Lehrenden selbst in den Blick zu nehmen, um möglicherweise Verbindungslinien zur gegenwärtigen Bewertungspraxis ziehen zu können.

Da die Umsetzung von *Assessment* im schulischen Rahmen nicht in einem individuell geprägten Vakuum stattfindet, sind verschiedene potenziell prägende Faktoren in die Untersuchung einzubeziehen. Hier ist gemäß Marotzki und Tiefel nicht nur die subjekttheoretische, sondern auch die gemeinschaftstheoretische Perspektive einzubeziehen, um „Erkenntnisse über Lern- und Bildungsprozesse in einer dynamischen Moderne“ (Marotzki & Tiefel 2010: 73) zu gewinnen und dabei auch Prozesse der Systemsteuerung und der Qualitätssicherung innerhalb eines Bildungssystems zu betrachten (vgl. Fend 2008).

Winfried Marotzki und Sandra Tiefel beziehen sich auf Arnd-Michael Nohl, welcher die „handlungstheoretische Grundlegung qualitativer Bildungsforschung [...] in der pragmatischen Tradition“ verortet und für seine Forschung die dokumentarische Methode favorisiert. Nohl sieht Bildung „auf der Schnittstelle zwischen Individualität und Kollektivität“. Im Sinne der Vorstellung des konjunktiven Wissens Mannheims bilden „subjektive Sinnzuschreibungen“ einerseits und „milieuorientierte Handlungen“ andererseits die Determinanten des zu analysierenden Weltausschnitts (Marotzki & Tiefel 2010: 75). Meine Studie zu Fremdsprachenlehrkräften lässt sich ebenfalls auf der von Nohl genannten „Schnittstelle zwischen Individualität und Kollektivität“ einordnen, da es sowohl um individuelle und kollektive Bildungserfahrungen in der Position der Lernenden als auch um die Rekonstruktion der professionellen Situation von Amtsträgern im Staatsdienst geht.

In Bezug auf den Umgang mit Leistungsbewertung spielt in beiden Positionen *Assessment* eine Rolle, da zwar bis zur staatlich autorisierten Übernahme des Lehramtes kontinuierlich wertende Rückmeldungen über den fachlichen Leistungsstand der Lernenden gegeben werden, sich mit Beginn der Tätigkeit als Fremdsprachenlehrkraft das Verhältnis jedoch radikal umkehrt: Während man zuvor lediglich auf die eigene Person bezogen jeweils eine Beurteilung erfährt, müssen nun hunderte Lernende bewertet und die dazugehörigen Bewertungsgrundlagen transparent gemacht werden. Das Verhältnis von Individualität und Kollektivität erfährt folglich insofern eine Neuakzentuierung, als die individuelle Praxis der Leistungsbewertung einer Lehrkraft sowohl durch das Kollektiv der Schülerschaft und der Elternschaft, die Abnehmer, als auch durch das Kollegium und indirekt das Kultusministerium eingeschätzt und potenziell in Frage gestellt wird.

Laut Marotzki und Tiefel kommt es in Berufsbiographien zu „Brüchen und Krisen, Orientierungsverlusten wie Neukalibrierungen“ (Marotzki & Tiefel 2010: 76). Ein wichtiger Teil meiner Studie besteht in der Rekonstruktion des

Umgangs von Fremdsprachenlehrkräften mit der Kritik, die sie in Bezug auf die Lehrerfunktion des Beurteilens von Leistungen erfahren. Dies geschieht vor dem Hintergrund der Ideale, die sie bezüglich einer „gerechten“ Bewertung im Fremdsprachenunterricht hatten bzw. haben.

Marotzki und Tiefel betrachten „Bildung als Reflexivwerden von Lernprozessen“ und wenden Elemente der Narrationsanalyse nach Franz Schütze an, um die Schritte der „Veränderung von Selbst- und Weltverhältnissen“ zu rekonstruieren (Marotzki & Tiefel 2010: 74). Die von Marotzki postulierte „Reflexivitätsthese“ (Marotzki & Tiefel 2010: 74) hat eine erhebliche praktische Relevanz für die tägliche Evaluationspraxis von Fremdsprachenunterricht. Fred Genesee, Jack C. Richards und John A. Upshur skizzieren den Evaluationsprozess wie folgt:

Second language evaluation is primarily about decision making. Indeed, the overall purpose of second language evaluation is to make sound choices that will improve second language teaching and enhance second language learning. Such decisions are based on informed judgement. They require the careful collection of relevant information and a thoughtful interpretation of that information. Three essential components of evaluation are information, interpretation and decision making. It is important to distinguish among these three components because they are distinct. Information about teaching and learning is rarely meaningful by itself. It becomes meaningful when it is interpreted. Meaningful interpretations are needed in order to decide what actions to take or what changes to make to instruction. (Genesee et al. 1996: 4)

Die im Zitat beschriebenen Entscheidungs-, Evaluations- und Adaptionprozesse werden innerhalb der Lehrerkognitionsforschung untersucht. Während Forschungsergebnisse in den späten siebziger Jahren erstmals die Komplexität von Unterrichtstätigkeit aufzeigen und somit allzu lineare und vordergründig rationale Modelle als unzureichend ausweisen (Borg 2006: 9), ergeben Befunde in den achtziger Jahren, dass es keinen eindeutig nachweisbaren Zusammenhang zwischen Denken und Handeln im Sinne eines quasi automatisch ablaufenden Handlungsschemas bei Lehrkräften gibt und sich stattdessen bestimmte Routinen und Skripte hartnäckig als handlungsleitend erweisen, da anscheinend eine Prämisse des Lehrerhandelns und der zugrunde liegenden Entscheidungen darin besteht, den Unterrichtsfluss nicht zu unterbrechen (Borg 2006: 11).

In seiner Untersuchung zu Kognitionen von Sprachenlehrkräften stellt Borg heraus, dass trotz der genannten Dominanz der vor allem in der eigenen Schulzeit „gelernten“ Skripte und deren ungünstiger Wirkung auf neu zu erwerbende Denk- und Handlungsmuster (Borg 2009: 164) eine signifikante Beziehung zwischen Lehrerkognitionen und Handeln besteht. Er weist jedoch darauf hin, dass es keine lineare Veränderung von Lehrerkognitionen gebe und auch bei Referendarinnen und Referendaren zu beobachtende Handlungsveränderungen nicht unbedingt Zeichen einer veränderten Denkweise sein müssten (Borg 2009: 164;

166). Des Weiteren stellt er heraus, dass eine mangelnde Stimmigkeit von Überzeugungen (*teacher beliefs*) einerseits und Handeln andererseits nicht als unglaubwürdig angesehen werden sollte (Borg 2009: 167). In diesem Zusammenhang ist auch John S.C. Farrells Befund zur „Halbwertszeit“ von Elementen der Lehrerbildung zu nennen: Bereits ein Jahr nach Abschluss des Referendariats dominierten Überzeugungen aus der Zeit vor der Ausbildung und neu erworbene Handlungsmuster seien anscheinend verschwunden (Farrell 2009: 183).

Für meine Studie ergeben sich in Bezug auf die skizzierten Forschungsergebnisse aus der Kognitionsforschung folgende methodologische Überlegungen: Es erscheint grundsätzlich zielführend, sich mit Lehrerkognitionen zum Thema *Assessment* im Rahmen rekonstruktiver Sozialforschung zu beschäftigen. Dabei sind jedoch die Perspektiven *pre-service* und *in-service* der Lehrkräfte zu unterscheiden, um die professionelle Entwicklung und die Genese der Vorstellungen zu Leistungsbewertung im Fremdsprachenunterricht möglichst komplex erfassen zu können und bei den Befunden deutlich über die Ebene der subjektiven Theorien geringerer Reichweite (Wahl 2006: 24) hinauszugehen. Eine Voraussetzung für die Teilnahme an der geplanten Interviewstudie ist daher ein recht hohes Maß an Reflexionsfähigkeit im Sinne eines Konzepts des *Reflective Language Teaching* (2007).

Die diskursive Entwicklung von Identität als Fremdsprachenlehrkraft bedarf der Navigation verschiedener Diskurse (Hedgcock 2009: 146) und spielt laut Jennifer Miller eine zentrale Rolle für die professionelle Identitätsentwicklung (Miller 2009: 172). Gurmit Singh und Jack C. Richards nennen die Teilnahmefähigkeit am Diskurs über Fremdsprachenlehren und -lernen gar „a new kind of literacy“ (Singh & Richards 2009: 204). Paula Golombek charakterisiert die möglicherweise schmerzhafteste Bestandsaufnahme der eigenen Sprachlerngeschichte als förderlich für die Reflexion und gezielte Introspektion als Fremdsprachenlehrkraft (Golombek 2009: 157). Miller sieht diesen Weg der Identitätsfindung als „process of continual emerging and becoming“ (Miller 2009: 173) und weist verschiedene Diskursthemen als „lens of identity negotiation“ aus (Miller 2009: 178). Zusammenfassend ist deshalb festzustellen, dass die Interviewgruppe nicht nur diskursfähig, sondern auch bereit sein muss, sich mit der eigenen Spracherwerbgeschichte auseinanderzusetzen, selbst wenn dies bedeutet, dass die eigene Expertise, im angelsächsischen Bereich auch als *Personal Practical Knowledge* (PPP) geläufig, einer Neubewertung unterzogen wird.⁶

6 Golombek nennt diesen Prozess „to frame and reframe teachers' PPK“ (Golombek 2009: 158).

2.1.3 Einordnung in die Sprachlehr- und Lernforschung

Übertragen auf den Bereich der Fremdsprachendidaktik hat „Lehren, Lernen und Bewerten vor neuen Horizonten“⁷ vielfältige Bedeutungen und Implikationen. Wie erwähnt hat die Diskussion um Leistungsbewertung im Fremdsprachenunterricht gerade in der ersten Dekade des neuen Jahrtausends durch die Publikation des GERund durch die Entwicklung und Durchführung wichtiger Studien wie DESI sowie aktuellen Ländervergleichsstudien zu fremdsprachlicher Kompetenz⁸ wesentliche neue Impulse in Richtung Standardisierung und auch *Outcome*-Orientierung erhalten. Der Sprachunterricht muss sich daran messen lassen, welche Ergebnisse er generiert.

Auch Vergleichsarbeiten im Rahmen der länderübergreifend in den Hauptfächern durchgeführten Lernstandserhebung VERA in den Schuljahrgängen 6 und 8 tragen zur Auseinandersetzung mit Fragen der Lern- und Leistungskultur bei. Die von der Universität Koblenz-Landau entwickelten Testunterlagen stellen möglicherweise für Lehrkräfte eine Erstbegegnung mit professionell erstellten und durchgeführten Sprachtests dar. Die Vorbereitung auf VERA sowie die Auswertung und Rückmeldung an das Kultusministerium haben sicherlich einen Effekt auf die Testkultur innerhalb der Fachgruppe. Ob von einem *Washback*-Effekt auf Lehren und Prüfen im Sinne der angelsächsischen Prägung dieses Begriffs gesprochen werden kann, ist meines Wissens noch nicht systematisch untersucht worden.⁹ Ähnliche Auswirkungen sind jedoch denkbar.

Da mittlerweile nicht nur für die Aufnahme eines Studiums Sprachkenntnisse gemäß der Stufen des GER nachgewiesen werden müssen – Bewerbungsvoraussetzung etwa für den gehobenen Polizeidienst in Niedersachsen ist das Level B2¹⁰ –, halten Sprachzertifikate zunehmend Einzug in deutsche Klassenzimmer, sodass Bewertungen der Lehrkräfte nicht mehr in einem Vakuum stattfinden.

Ein Beispiel für die Unzuverlässigkeit von Bewertungen deutscher Englischlehrkräfte ist die Studie von Reinhard David aus dem Jahr 1991. Die untersuchten *Rater* von schriftlichen Abiturprüfungen unterschieden sich stark in ihren Urteilen und erwiesen sich deshalb in Bezug auf *Test-Fairness* und *Inter-Rater-Reliability* als deutlich unzuverlässig (David 1991; Arras 2007). Diesem Umstand wird an vielen Universitäten Rechnung getragen, indem bestimmte Punktzahlen und Niveaus in professionellen Sprachtests wie *TOEFL* und *Cambridge* verlangt

7 Diese Formulierung erfolgt in Anlehnung an den Titel der Zeitschrift SEMINAR vom März 2009 „Beraten, Beurteilen, Prüfen in neuen Horizonten“.

8 Die Ergebnisse der Ländervergleichsstudie wurden am 23.06.2010 veröffentlicht.

9 Die Eindrücke zu VERA 8 sind Bestandteil meines Interviewleitfadens.

10 http://www.polizei.niedersachsen.de/dst/pdol/pd_oldenburg/beruf/start.html [Letzter Zugriff am 10.12.2012].

werden und die im Abiturzeugnis attestierten Lehrerurteile ihre *Gatekeeper*-Funktion zu einem Sprachenstudium einbüßen. Ulrike Arras (2007: 18) verweist allerdings auf die Singularität von Studien wie der von David und somit auf die Grenzen der Generalisierbarkeit der Befunde.

Die Auseinandersetzung mit dem GER und seiner Funktionalität in der Praxis des Fremdsprachenunterrichts findet sowohl im wissenschaftlichen Diskurs als auch in den Fachgruppen der jeweiligen modernen Fremdsprachen statt. Zu nennen wären beispielsweise die Arbeitspapiere der 22. Frühjahrskonferenz zur Erforschung des Fremdsprachenunterrichts im Jahre 2003, die im Sammelband zum Thema „Der Gemeinsame Europäische Referenzrahmen für Sprachen in der Diskussion“ veröffentlicht wurden. Die tatsächliche Anbindung der Forschung an die unterrichtliche Realität an deutschen Schulen ist in kleinem Rahmen durch vereinzelte Studien gewährleistet, welche aber wegen ihres Umfangs nicht immer als repräsentativ gelten können. Die Einschätzungen von Praktikern bleiben folglich oft ungehört, da sich die Kooperation von Universitäten und Schulen selten systematisch und in größerem Rahmen mit Fragen der Sprachdidaktik (Gehring 2010: 13) beschäftigt.

Der GER kann bei der Arbeit mit unterschiedlichen Altersgruppen und Schulformen für die Fremdsprachenlehrkraft sicherlich eine gewisse Orientierung geben. Eine konkrete Leistungsmessung unter Nutzung der GER-Skalen als kriterialer Norm ist jedoch im täglichen Unterrichtsgeschehen nur begrenzt möglich, da der eigene, auf die Lerngruppe und das Curriculum abgestimmte Unterricht speziell angepasste Raster erfordert, die für eine aussagekräftige Messung erst erstellt werden müssten.¹¹ Zydatið beklagt in seiner Analyse der Anwendbarkeit der GER-Deskriptoren „die Vagheit der Skalierung, die weitgehend über graduelle Adjektive, Adverbien und Mengenangaben verläuft“ (Zydatið 2005a: 7).

Auch Glenn Fulcher und Fred Davidson kritisieren die mangelnde Messgenauigkeit der GER-Skalen, da ihrer Meinung nach eine zu enge oder zu breite Definition des jeweiligen Kompetenzlevels vorliegt und so eine Zuordnung allenfalls vage vorzunehmen sei (Fulcher & Davidson 2007: 100). Außerdem weisen sie auf die fehlende Passung von externem *Large-Scale Testing* einerseits und *Classroom Assessment* andererseits hin (Fulcher & Davidson 2007: 35). Gemäß den Kriterien des GER handelt es sich bei Klassenarbeiten deutschen Stils um „punktueller, kriteriumsorientierte, analytische, summative, (*idealiter*) objektive Sprachstandsbeurteilung (*achievement test*) in Form der Fremdbeurteilung (durch die Lehrkraft)“ (Thaler 2008: 6).

11 Eine umfassende Analyse der Möglichkeiten und Grenzen des GER findet sich bei Bausch, Christ & Königs (2003).

Das in Deutschland noch wenig bekannte und beforschte Feld des *Language Testing* wird Praktikern in einschlägigen Fachdidaktiken, wie beispielsweise der Wolfgang Gehrigs (Gehrig 2010), in theoretischer und praktischer Weise näher gebracht und auch von kultusministerieller Seite aus liegen Anregungen zu einer veränderten Testpraxis vor. Die Ende 2008 erschienenen „Materialien für kompetenzorientierten Unterricht in der Sekundarstufe I“ (Niedersächsisches Kultusministerium 2008) sehen neben der *Output/Outcome*-Orientierung und der Handlungsorientierung den *CAN-DO-Approach* als dritte Säule einer veränderten Unterrichts- und Bewertungspraxis im Englischunterricht vor. Die Autoren unterstreichen die Notwendigkeit, den Erwerb von sprachlichen Kompetenzen mit Hilfe von geeigneten Testformaten zu überprüfen, und heben in diesem Zusammenhang die Bedeutung des GER, der Bildungsstandards und des Kerncurriculums hervor (Niedersächsisches Kultusministerium 2008: 4-7).

Vor dem Hintergrund der bisher skizzierten Situation des Sprachenlernens, -lehrens und -prüfens liegt mein Forschungsinteresse darin, genauere Informationen und auch Reflexionen über die Praxis des *Assessments* bei Fremdsprachenlehrkräften zu erhalten. Ziel der präsentierten qualitativen Interviewstudie ist es, induktiv und theoriegenerierend Kausalmechanismen, die sozialem Handeln zugrunde liegen, zu beschreiben und deren Geltungsbereich zu bestimmen (vgl. Gläser & Laudel 2009: 25 f.). Die Frage nach den der Bewertung von Schülerleistungen zugrunde liegenden Kausalmechanismen ist den durchzuführenden Interviews übergeordnet. Es geht weiterhin darum, welchen Grad der Komplexität bei Prozessen des *Assessments* die Befragten tatsächlich wahrnehmen und artikulieren können, sowie in welcher Art von Diskurs sie zu diesem Themenfeld stehen. Das theoretische Hintergrundwissen der Befragten und dessen Verarbeitung sind dabei relevant, denn es geht in den Experteninterviews nicht nur um die subjektiv-spezifische Konstruktion, sondern um verarbeitetes Fachwissen. Da es in den verschiedenen Bundesländern deutliche Unterschiede in der Rezeption und curricularen Umsetzung etwa des GER oder der Bildungsstandards gibt, werde ich mich in meiner Studie in der Auswahl meines Samples und der Analyse der Curricula sowie anderer schulrechtlicher Bestimmungen zu Leistungsmessung und Leistungsbewertung auf Niedersachsen beschränken.

Ich stimme mit Matthias Trautmann dahingehend überein, dass Sprachlehrforschung zu wenig auf die (deutsche) Schule bezogen ist, und zudem ein genaueres Verständnis über den Umgang von Lehrenden und Lernenden mit Fremdsprachen eine stärkere Einbeziehung erziehungswissenschaftlicher und schulpädagogischer Befunde notwendig macht (Trautmann 2007b: 196). Dennoch ist bei einem Dissertationsprojekt wie dem vorliegenden eine Fokussierung auf eine begrenzte Zahl von Perspektiven nötig, um der Forschung eine klare Richtung zu geben. Dies gilt sowohl für den Umfang der Datenmenge als auch

für die Auswertung. Katharina Kunze und Bernhard Stelmaszyk sprechen in diesem Zusammenhang von der Herstellung eines „systematischen und verdichteten Zusammenhang[s] zwischen Datenerhebung und Datenmengen“ und favorisieren eine möglichst extensive Interpretation überschaubarer Datenmengen (Kunze & Stelmaszyk 2008: 833). Die Autoren beklagen hinsichtlich der Lehrerbiographieforschung eine Beliebigkeit und mangelnde Fokussierung gerade aufgrund der Vielfältigkeit des Feldes:

Insgesamt hat sich ein beeindruckend umfangreiches Feld an Arbeiten zu vielen Facetten von Lehrerbiographien aufgetan. Es erweckt jedoch eher den Eindruck eines „bunten“ Patchworks als den einer systematisch ausdifferenzierten Forschungsrichtung. (Kunze & Stelmaszyk 2008: 832)

Die Arbeiten von Terhart (2000; 2001) geben Aufschlüsse über Berufskultur und Berufsbiographie von Lehrkräften. Die Fachlehrerperspektive ist bisher jedoch wenig erforscht. In Bezug auf die Biographieforschung bei Englischlehrkräften ist die Arbeit „Wie werden Englischlehrer professionell?“ von Dirks (2000) zu nennen, die sich allerdings mit den speziellen Bedingungsgefügen ostdeutscher Lehrersozialisation beschäftigt.

Aufbauend auf den Erkenntnissen Nikola Mayers (2002) zum Thema „Ganzheitlichkeit und Sprache“ sowie besonders Daniela Casparis zum beruflichen Selbstverständnis von Fremdsprachenlehrern (2003) lässt sich meine Studie jedoch recht gut in die Lehrerforschung in der Tradition der Biographieforschung einordnen, da das Thema „Leistungsbewertung“ im Fremdsprachenunterricht im doppelten Sinne einen Kristallisationspunkt der eigenen Biographie als Fremdsprachenlernende und -lehrer bietet. Zunächst geht es um die eigenen Erlebnisse mit Bewertung des Spracherwerbsprozesses in Ausbildungs- bzw. Erwerbssituationen und erst in einem zweiten Schritt um die Erfahrungen mit Bewertung und Begleitung der Spracherwerbsprozesse bzw. des Sprachstandes von Schülerinnen und Schülern in der Rolle als Lehrkraft, kurz um diagnostische Expertise.

Forschungen zur Lehrerexpertise haben zum Ziel, Lehrerkognitionen offenzulegen. Gemäß Bromme und Haag liegt derzeit „kein empirisch und psychologisch abgesichertes Modell der Entwicklung von Expertise, aus dem ein normatives Modell der Lehrerausbildung abgeleitet werden könnte“, vor. Die bloße Berufserfahrung führe nicht automatisch zu mehr Expertise und es sei ohnehin fraglich, inwieweit der Erwerb von „domänenspezifische[m] Fachwissen“ zur Ausübung des Lehrerberufs überhaupt notwendig sei (Bromme & Haag 2008: 811).

Der handlungsleitende Einfluss der Erfahrungen aus der eigenen Schulzeit (Bromme & Haag 2008: 811) wird in meiner Studie insofern aufgegriffen, als im ersten Themenblock das Verhalten der eigenen Lehrerinnen und Lehrer thematisiert und auch die Möglichkeit zu kritischer Reflexion gegeben wird. Da das Thema „Leistungsbewertung“ einerseits sehr konkret und andererseits auch sehr

emotional besetzt ist, erscheint eine „Verwässerung“ der Daten durch die Äußerung von Allgemeinplätzen eher unwahrscheinlich.

Die Experten- und Novizenforschung bietet einen Anknüpfungspunkt, um Aufschlüsse über die Wissensrepräsentation bei Lehrkräften zu erhalten. Bromme und Haag verweisen jedoch darauf, dass kein „qualitativer sprunghafter Übergang gemäß einem Alles-oder-Nichts-Prinzip in der Entwicklung der Expertise“ anzunehmen ist (Bromme & Haag 2008: 806). Die Autoren warnen davor, Fachwissen und Unterrichtswissen gleichzusetzen, und weisen in diesem Zusammenhang auf die Existenz von Mediationsprozessen hin (Bromme & Haag 2008: 809). Die vorliegende Studie greift die Erkenntnisse der Experten- und Novizenforschung insofern auf, als bei der Auswahl der Interviewpartnerinnen und -partner der Hauptstudie darauf geachtet wurde, dass sie auf einem ähnlichen Wissensstand sind. Es ist allerdings nicht durchgängig davon auszugehen, dass die von Bromme und Haag postulierte „kategoriale Wahrnehmung bei Experten“ (Bromme & Haag 2008: 807) auch tatsächlich ausgeprägt ist.

Der Titel „Voices from the Language Classroom“, den Kathleen M. Bailey und David Nunan (1996) für ihren Band mit Aufsätzen zu hauptsächlich qualitativ orientierter Lehrerforschung wählen, wäre auch für meine Studie ein passender Arbeitstitel, da es hier ebenfalls darum geht, Stimmen aus fremdsprachlichen Klassenzimmern zu Wort kommen zu lassen. Zwar sprechen „nur“ Lehrkräfte, dies geschieht jedoch im Rückblick und Rückgriff auf eigene Erfahrungen mit Fremdsprachenlernen und den damaligen Lehrenden. Bromme und Haag fordern angesichts der starken Wirkung des als Schüler rezipierten Lehrerverhaltens, welches sogar durch eigene Berufstätigkeit erst recht reaktiviert werden kann, eine gezielte Reflexion der eigenen Schulzeit (Bromme & Haag 2008: 811).

Angesichts von „biographisch aufgeschichteten Deutungsbeständen“, in denen Lehrerhandeln wurzelt, sehen auch Combe und Kolbe Reflexivität als Schlüsselkompetenz an (Combe & Kolbe 2008: 859). „Stimmen aus dem Fremdsprachenklassenzimmer“ sind zur Beantwortung der Forschungsfrage, wie sich Vorstellungen zur Leistungsbewertung im Fremdsprachenunterricht entwickeln, nur dann aussagekräftig, wenn *Reflective Practitioner* zu Wort kommen und auch bereit sind, mögliche Widersprüche im Gespräch aufzudecken und auszuhalten. Dabei können symbolisch-interaktionistische Konzepte der widersprüchlichen Situation an Schulen und der Erfahrung unaufhebbarer Antinomien Ausdruck verleihen (Combe & Kolbe 2008: 857). Das Aushandeln von Bedeutung stellen Combe und Kolbe wie folgt dar:

[...] dass jeder Einzelne dauernd vor der Aufgabe steht, die unklaren und widersprüchlichen Ansprüche der anderen nicht nur auszubalancieren, sondern auch aktiv über verbale bzw. non-verbale Gesten zu interpretieren und zu beeinflussen. (Combe & Kolbe 2008: 857)

Für diese Studie hat die oben skizzierte Fähigkeit, mit der Vielfältigkeit und möglicherweise auch Krisenhaftigkeit der Anforderungen an den Lehrerberuf produktiv umzugehen, eine zentrale Bedeutung, denn gerade im Bereich der Lehrerfunktion der Leistungsbewertung besteht ein hoher Erklärungs- und möglicherweise Rechtfertigungsdruck auf Seiten der Lehrkräfte. Eine bloße Äußerung der eigenen Befindlichkeit und mögliche Rechtfertigungen der eigenen Praxis helfen bei der Erkundung dieses Bereichs wenig weiter, da sie lediglich Stimmungen, nicht aber Funktionsweisen des zu erforschenden Weltausschnitts beleuchten. Diese Annahme hat Konsequenzen auf die Auswahl des Samples der Interviewpartner, da Experten gewählt werden sollen, deren Aussagen einen möglichst vielfältigen Einblick in das Forschungsfeld gewähren.

2.1.4 Die Studie im Schnittpunkt qualitativer und quantitativer Forschung

Ein Einblick in die bei Fremdsprachenlehrkräften während der Bewertung von Schülerleistungen ablaufenden mentalen Prozesse ist grundsätzlich schwer zu erlangen, da die Einschätzungen selten artikuliert und expliziert werden. Die von Terhart (2000: 39) problematisierte, schulgesetzlich verankerte Verpflichtung, Leistungen von Schülerinnen und Schülern zu bewerten und dies in Form von Ziffernnoten den Lernenden und deren Eltern mitzuteilen, findet an allgemeinbildenden Schulen auf einer Skala von 1 bis 6 statt. Obschon beispielsweise der GER laut KMK-Beschluss in Deutschland bundesweit als Bezugsmaßstab gilt und sich dies auch auf die ländereigenen Curricula auswirkt, kann nicht fraglos davon ausgegangen werden, dass dieses Dokument oder auch die KMK-Standards für die Lehrerbildung sich handlungsleitend auf die Bewertungspraxis von Fremdsprachenlehrkräften auswirken. Auch ist es angesichts von Defiziten im Bereich der Diagnostik in Lehrerausbildung und -fortbildung (Hesse & Latzko 2009) fraglich, ob die Prozesse, die einem für schulische Zusammenhänge möglichst objektiven, validen und reliablen Testinstrumentarium und Testprozedere zugrunde liegen, überhaupt bekannt sind. Selbst wenn die theoretischen Grundlagen als deklaratives Wissen geläufig wären, bliebe immer noch offen, ob von diagnostischer Expertise im Sinne Helmkes in der prozeduralen Wissensausprägung bei Fremdsprachenlehrkräften gesprochen werden kann.

Um verlässliche Aussagen darüber zu erhalten, wie akkurat Fremdsprachenlehrerinnen und -lehrer Leistungen in der Fremdsprache messen, wären recht groß angelegte quantitative Studien denkbar, in denen für verschiedene Kompetenzbereiche Einschätzungen und Prognosen von Fremdsprachenlehrkräften zu Leistungsstand, Förderbedarf und Entwicklung der sprachlichen Fähigkeiten von

Schülerinnen und Schülern abgegeben und in einem anschließenden qualitativ orientierten Forschungsabschnitt zudem Maßstäbe der Leistungsbewertung reflektiert würden. Diese Befunde könnten dann wie bei der PISA-Lesestudie mit den tatsächlichen, mit wissenschaftlichen Tests erhobenen Ergebnissen des *Assessments* verglichen werden. So könnten auch Aussagen zur *Inter-Rater-Reliability* gemacht werden, um, besonders mit Blick auf zunehmend stattfindende Sprechprüfungen, Perspektiven für die schulische Praxis aufzuzeigen. Lehrkräfte könnten in einem Interviewteil die ihren Bewertungen zugrunde liegenden Konzepte zu sprachlicher Korrektheit und Maßstäbe von Leistungsbewertung angeleitet reflektieren bzw. bei mündlichen Prüfungen oder Bewertung kommunikativer Leistungen im Unterricht Videosequenzen analysieren (vgl. Aguado 2009; Trautmann & Sacher 2010), sodass eine Triangulation quantitativ und qualitativ erhobener Daten möglich würde.

Mit Fremdsprachlehrerinnen und -lehrern über Fragen des *Assessments* ins Gespräch kommen, impliziert jedoch eine Perspektive, welche aus dem Unterrichtsalltag heraus „spricht“, d.h. Praktiker ohne vorherige Lenkung ihre eigenen Erfahrungen mit und Vorstellungen über Sinn und Funktion sowie Praxis von Leistungsbewertung im Fremdsprachenunterricht darlegen bzw. im Interview selbst erst rekonstruieren lässt. Die Komplexität von Leistungsbewertung im alltäglichen Fremdsprachenunterricht weist sicherlich über das Abschneiden von Lehrkräften in der besonderen Situation von Schulleistungsstudien hinaus und hat viele Facetten, die mit der eigenen Biographie und Sprachlerngeschichte, eigenen Erfahrungen mit Prüfungen und Auslandsaufenthalten sowie Vorstellungen über Sprache und Angemessenheit sprachlicher Äußerungen und auch Lehrerprofessionalität zu tun haben. Christiane Kallenbach definiert die Beziehungen von vergangenen Ereignissen mit Blick auf Gegenwart und Zukunft folgendermaßen:

Im Gegensatz zum tatsächlichen Erleben ist ein Ereignis in der Erinnerung als Ganzes und vor allem in seinen Konsequenzen zu überblicken. Damit gewinnen vergangene Erlebnisse, selbst wenn sie nicht mehr in ihrem genauen Hergang erinnert werden und in der Hinsicht auch nicht mehr bedeutsam sind, Relevanz für die Gegenwart und Zukunft – gewissermaßen als *Essenz einer gelebten Erfahrung* [Hervorhebung – S. B.]. (Kallenbach 2007: 219)

Die Studie von Hanna Cronjäger (2007) zur Erfassung von Lernemotionen im Französischunterricht bietet hier einen wichtigen Anknüpfungspunkt, da auf den von der Forscherin genutzten Skalen neben dem recht breit erforschten Phänomen der Angst auch positive Emotionen wie Stolz erfasst werden, welche auch für die Einschätzung von emotionalen Reaktionen auf Leistungsbewertungen relevant sind. Das Interesse der vorliegenden Studie liegt unter anderem in der reziproken Wirkung von selbst erlebten Lernemotionen auf die Art und Weise, wie später selbst mit Leistungsbewertung im eigenen Unterricht umgegangen

wird. Es ist zu überprüfen, ob auch im Bereich der Leistungsbewertung der Faktor *Apprenticeship of Observation* (Bardhun & Johnson 2009: 60) handlungsleitend wirkt oder ob durch Erfahrungen und Reflektionen über die eigene Schulzeit hinaus neue Maßstäbe entwickelt und angewendet werden.

Andreas Müller-Hartmann und Marita Schocker-von Ditfurth betonen, dass qualitative Forschung nicht linear verlaufe, und verweisen auf die zentrale Bedeutung von Gegenstandsangemessenheit, Ganzheitlichkeit und Kontextgebundenheit. Im Mittelpunkt der Forschung stehe das Bemühen, die Bedeutungszuschreibungen derer zu erfassen, welche in dem jeweiligen Feld handeln. Die gewählten Analyseinstrumente müssten in einem stimmigen Verhältnis zu den komplexen Bedingungen der zu erforschenden „sozialen Lebens- und Handlungszusammenhänge“ stehen (Müller-Hartmann & Schocker-von Ditfurth 2001: 3). Die Autoren plädieren für eine offene Herangehensweise und betonen die Möglichkeit der stimmigen Ergänzung quantitativer und qualitativer Forschungsmethoden. Weiterhin nennen sie sechs Merkmale qualitativer Forschung:

- didaktische Orientierung, d.h. Entwicklung von Vermittlungskonzepten,
- Teilnehmerorientierung,
- erweiterter Forschungsbegriff, der flexibel, inklusiv und gegenstands begründet ist,
- Prozessorientierung, Reflexivität und Transparenz,
- Forscher als Teil des Erkenntnisprozesses,
- Ermittlung geteilter wie auch divergierender Sinnstrukturen (Müller-Hartmann & Schocker-von Ditfurth 2001: 4 ff.).

In der praktischen Umsetzung impliziert qualitative Forschung im Bereich des Fremdsprachenlehrens und -lernens in Deutschland oft eine Orientierung an dem von Scheele und Groeben Mitte der siebziger Jahre konzipierten Forschungsprogramm Subjektive Theorien (FST) (Scheele & Groeben 1988). Die subjektive Sicht eines Individuums soll mit Hilfe dieses Verfahrens rekonstruiert werden, indem komplexe individuelle Kognitionssysteme erhoben werden (Caspari 2003: 72). Ein weiterer Schritt im Forschungsprozess besteht in der Handlungsvalidierung, d.h. in der Überprüfung, ob die rekonstruierte subjektive Theorie sich tatsächlich als handlungsleitend erweist (Caspari 2003: 72). Groeben und Scheeles Konzept bewegt sich zwischen den beiden Polen der engen und weiten Begriffsexplikation hinsichtlich der Festigkeit und Verbindlichkeit der jeweiligen Theorie. Der Schritt der Aggregierung besteht in der Zusammenfassung individueller subjektiver Theorien zu übergreifenden Theoriestructuren, was in gewisser Weise eine nomothetische Sichtweise voraussetzt (Caspari 2003: 76).

Beispiele in der Fremdsprachenforschung für die konkrete Anwendung des FST sind die Studien von Kallenbach (1996), Mayer (2002) und Caspari (2003)

sowie Schart (2003). Letzterer verändert den Ansatz jedoch an einer zentralen Stelle, indem er bewusst auf die kommunikative Validierung mittels der von Groeben und Scheele entwickelten Struktur-lege-Technik (1988) verzichtet (vgl. 3.2.8 zur kommunikativen Validierung). Caspari setzt eine ihrer Ansicht nach nur teilweise bewusste, allerdings recht stabile subjektive Großtheorie zum Thema „Beruf als Bestandteil eines Gesamtselbstverständnisses von Fremdsprachenlehrern“ voraus (Caspari 2003: 15; 36). Für ihre Zwecke kommt das FST nur in seiner weiten Explikation in Frage, um eine unnötige Einengung der Perspektive zu vermeiden. Ohnehin ist ein zentraler Orientierungspunkt die Gegenstandsan-gemessenheit (Caspari 2003: 72; 77), welche ggf. auch eine Abwandlung des Forschungsparadigmas zur Folge hat, etwa bei Schart (2003) oder Wieser (2008). Das Ziel von Casparis Studie besteht in einer möglichst detaillierten, interpretativen Beschreibung des beruflichen Selbstverständnisses der von ihr interviewten Fremdsprachenlehrkräfte. Das dabei vorausgesetzte Menschenbild ist das von rationalen, reflexiven und kommunikativen Gesprächspartnern (Caspari 2003: 88).

Wie unter anderen Müller-Hartmann und Schocker-von Ditfurth betonen, ist qualitative Forschung oft explorativ und muss deshalb Flexibilität in der Vorgehensweise ermöglichen. Adelheid Hu geht sogar einen Schritt weiter, indem sie auf postmoderne oder poststrukturalistische Annahmen zur Forschungsmethodik verweist, welche die Vorstellung veränderter Validitätskriterien im Sinne „multipler Wahrheiten“ vertritt (Hu 2001: 30 f.). Hu ist sich der Subjektivität ihrer Forschung zum Thema „Lebensweltliche Mehrsprachigkeit und schulischer Fremdsprachenunterricht. Die Perspektive von Lehrenden und Lernenden“ bewusst, führt aber an, dass die *per se* subjektiv-interpretativ ausfallende Darstellung wissenschaftlichen Ansprüchen genüge, da die Subjektivität transparent gemacht werde und ihr Forschungszugriff es ermögliche, die „Stimmen“ der Gesprächspartner möglichst authentisch „sprechen zu lassen“ (Hu 2001: 30).

Die wachsende Beliebtheit qualitativ-interpretativer Forschungsparadigmen, welche sich gegenüber analytisch-nomologischen Ansätzen vermehrt durchsetzen konnten, ging mit dem steigenden Interesse an der „Binnensicht“ aus kognitionspsychologischer Perspektive einher. Die nunmehr „reflexive[n] und intentionale[n] Aktanten“ emanzipierten sich in der Folge aus der Rolle des „Daten liefernden Subjektes“ (Hu 2001: 13). Dieser Trend hat sich ein Jahrzehnt später bestätigt, zumal die Hirnforschung komplexere Strukturen nachweisen konnte, als dies bei Entstehung des FST Ende der siebziger Jahre möglich war, und zudem computergestützte Auswertungsprogramme von Daten eine schnellere Analyse quantitativ und qualitativ erhobener Daten erlauben und auch eine Triangulation von Daten leichter möglich ist.

Ebenso wie bei dem FST ist auch bei dem Begriff der Ethnographie ein enger und weiter Gebrauch möglich. Hu beklagt zu Recht, „dass eine deutliche Konfusion

in Bezug auf ethnographische Ansätze herrsche“ (Hu 2001: 15). Sie verortet Sprachenlernen in den sozialen und kulturellen Zusammenhängen der Gruppe und geht dabei über die bloße Reduktion auf mentale Prozesse hinaus (Hu 2001: 16). Kontextgebundenheit, ein interpretativer Zugang sowie Ganzheitlichkeit weisen Hus Studie als dem ethnographischen Paradigma zugehörig aus (Hu 2001: 17). Unter kritischer Reflexion der in der Gesellschaft und im wissenschaftlichen Diskurs vorhandenen Konzepte entscheidet sich Hu in Abgrenzung von der analytischen Linguistik für den ihrer Ansicht nach komplexeren anthropologischen Sprachbegriff (Hu 2001: 22). Diese Entscheidung passt zu den epistemologischen Vorannahmen der Selbstreferentialität von Sprache, z.B. der Unterscheidung von wissenschaftlichen und fiktionalen Texten (Hu 2001: 25).

Auch Michael Schart sieht „keine Notwendigkeit, den Erzählungen der Lehrenden eine ontologische Qualität zuzuschreiben“ (Schart 2003: 24) und geht im Hinblick auf seine Person als Forscher noch einen Schritt weiter, indem er seine eigenen persönlichen Empfindungen im Sinne der Authentizität des Forschungsprozesses gezielt einer „aperspektivischen Objektivität“ entgegensetzt (Schart 2001: 43 f.). Zwar stimme ich mit Schart dahingehend überein, dass die Forscherperspektive besonders bei kleineren Untersuchungen verstärkt Berücksichtigung finden muss. Im Gegensatz zur Arbeit in Forscherteams kann bei Forschungsprojekten im Alleingang kaum eine retrospektive Reflexion der Erhebungsmethoden stattfinden, beispielsweise der Fragetechnik bei Interviews. Deshalb ist in der Tat eine umfassende Reflexion der eigenen Rolle sinnvoll. Meines Erachtens sollte die von Schart postulierte „reflektierte Subjektivität“ (Schart 2003: 42 f.) dennoch ein möglichst hohes Maß an Vergleichbarkeit der Erhebungs- und Auswertungssituation implizieren, da ansonsten die Daten kaum generalisierbar sind.

Natürlich ist das forschungsmethodische Prinzip der Offenheit (Gläser & Laudel 2009: 31) in Bezug auf unerwartete Informationen und Theorien für meine Studie sehr wichtig. Ein weiteres für mich leitendes Prinzip besteht in der Umsetzung von „Verstehen als Basishandlung“ (Gläser & Laudel 2009: 31). In Bezug auf die externe Validität teile ich zwar Scharts pointierte Aussage, dass qualitative Forschung „keine Glaubenssicherheit“ schaffe, meine aber, dass eine möglichst transparente Darstellung zu einem gegenstandsangemessenen Zugriff auf den Forschungsgegenstand führen kann und sollte, sodass qualitativ erhobene Ergebnisse dem Kriterium der Transferabilität in etwas höherem Maße entsprechen könnten, als Schart dies prognostiziert.

2.2 Diagnostik

2.2.1 Thematische Annäherung und Überblick

Die Vorstellung, mit Fremdsprachlehrerinnen und -lehrern über Leistungsbewertung und somit über Fragen des *Assessments* im weitesten Sinne ins Gespräch zu kommen, findet im Schnittpunkt verschiedener Forschungsrichtungen statt.

Angesichts des thematischen Schwerpunkts der Fragen zur Leistungsbewertung spielt das Forschungsfeld zu den Bereichen Diagnostik, *Language Testing* und *Assessment* naturgemäß eine zentrale Rolle. Im Bereich der deutschen Forschung sind zunächst die grundlegenden Werke von Karl-Heinz Ingenkamp und Urban Lissmann (2008) sowie Franz E. Weinert (2001) zu nennen. Andreas Helmke weist in der Neubearbeitung seines Buches zur Unterrichtsqualität dem Thema „Diagnostik“ als einer „Basiskompetenz für erfolgreichen Unterricht“ (Helmke 2009: 121) ebenfalls eine zentrale Rolle zu. Er wählt den Begriff der diagnostischen Expertise, da dieser methodisches, prozedurales und konzeptuelles Wissen umfasse und somit über den Begriff der diagnostischen Kompetenz im Sinne der Urteilsgenauigkeit hinausweise (Helmke 2009: 122). Um adaptiven Unterricht zu planen und durchzuführen, sei ein Abgleich zwischen Lernvoraussetzungen und zu stellenden Anforderungen notwendig, v.a. wenn es darum gehe, individuelle Förderung umzusetzen (Helmke 2009: 124).

Angesichts internationaler Schulleistungsvergleichsstudien verweisen die Autorinnen Ingrid Hesse und Brigitte Latzko auf die Relevanz diagnostischer Kompetenz bei deutschen Lehrkräften. Am Rande der ersten PISA-Studie erwiesen sich beispielsweise Deutschlehrerinnen und -lehrer als nur mangelhaft in der Lage, schwache Leser als solche zu identifizieren und somit geeignete Fördermaßnahmen zu initiieren. Hesse und Latzko halten allerdings eine Übertragung dieses Befundes auf die gesamte deutsche Lehrerschaft und deren umfangreiche diagnostische Tätigkeiten für übertrieben (Hesse & Latzko 2009: 14). Dennoch sprechen sie von einer „Renaissance pädagogischer Diagnostik“, deren Begriff bereits 1969 von Ingenkamp geprägt wurde (Hesse & Latzko 2009: 15).

Eiko Jürgens und Werner Sacher begründen ihrerseits die Notwendigkeit pädagogischer Diagnostik in einem gesamtgesellschaftlichen Zusammenhang, indem sie den schulgesetzlich verankerten Erziehungs- und Bildungsauftrag auf dem Hintergrund einer sich „rasant“ verändernden Gesellschaft betrachten und zu dem Schluss kommen, dass gerade der Bereich der Leistungsbeurteilung eine bewusst gestaltete Schülerpartizipation erfordere (Jürgens & Sacher 2008: 11; 16 f.). Eine zentrale Bedeutung weisen die Autoren in diesem Zusammenhang metakognitiven Kompetenzen zu (Jürgens & Sacher 2008: 33). Auch beklagen sie den für eine positive Leistungsentwicklung von Schülerinnen und Schülern

kontraproduktiven „subjektneutrale[n] Leistungsbetrieb an Schulen“ und beziehen sich auf Alexejew Nikolajew Leontjew (1977), der unterstreicht, dass Lernanforderungen nicht automatisch in Lernhandlungen münden (Jürgens & Sacher 2008: 33). Die Autoren warnen vor einem Missbrauch von Leistungsbewertung als Disziplinierungsmittel (Jürgens & Sacher 2008: 36) und schlagen stattdessen vor, Leistungs- und Unterrichtskultur in ein harmonisches Verhältnis zu überführen und dabei Lernen und Leisten sinnvoll miteinander zu verschränken und innerhalb einer veränderten Feedbackkultur so aufeinander zu beziehen, dass in klar unterschiedenen Lern- und Leistungssituationen Orientierung für das Weiterlernen gegeben wird (Jürgens & Sacher 2008: 47-50).

Die in der Folge von PISA 2004 veröffentlichten KMK-Standards für die Lehrerbildung nennen im Kompetenzbereich Beurteilen verschiedene, die theoretische und praktische Ausbildung betreffende Anforderungen. Damit Lehrerinnen und Lehrer „ihre Beurteilungsaufgabe gerecht und verantwortungsvoll“ ausüben, wird in den Kompetenzen 7 und 8 herausgestellt, dass angemessene Förderung und Beratung von Schülern und Eltern auf der Basis genauer Analyse der Lernvoraussetzungen erfolgen solle und dass eine Transparenz der Bewertungsmaßstäbe Grundlage der Leistungsmessung sei (Kultusministerkonferenz 2004: 11). Die Aufschlüsselung der beiden Kompetenzen stimmt in weiten Teilen mit denen der von Helmke zitierten *American Foundation of Teachers* überein (Helmke 2009: 125), wobei zu beachten ist, dass die amerikanischen Standards etwas umfangreicher und detaillierter ausfallen und zudem die im angelsächsischen Raum genutzten professionell erstellten Testverfahren mitberücksichtigen. Durch die mittlerweile verpflichtende Durchführung von Schulleistungstests wie VERA in den Fächern Mathematik, Deutsch und Englisch ist jedoch die wissenschaftlich abgesicherte Testpraxis mittlerweile auch an deutschen Schulen Bestandteil der Testkultur und kann so Grundlage für fachlichen Diskurs werden.

Die durch die Kultusministerkonferenz vorgelegten Standards, die unterschiedliche Kompetenzen ausweisen und sowohl Orientierung für die theoretische wie praktische Ausbildung geben, müssen laut dem von Hesse und Latzko zitierten Modell Osers eine theoretische Fundierung haben, auf Grundlagenforschung beruhen, an Kriterien ausgerichtet, evaluierbar und schließlich Bestandteil der schulischen Praxis sein. Oser schlägt auf diesen Annahmen aufbauend eine Ausbildung von Lehramtsanwärtern im Studium vor, welche auf einer Verzahnung zwischen psychologischem, pädagogischem und fachdidaktischem Wissen einerseits und Arbeit mit Forschungsberichten sowie Videomaterial der Praxis von Experten und Novizen andererseits basiert (Hesse & Latzko 2009: 21). Die von der Kultusministerkonferenz lediglich als Desiderat formulierten Standards wurden von Fritz Oser und Jürgen Oelkers 2001 in der Schweizer Lehrerbildung einer umfangreichen empirischen Prüfung unterzogen. Als Ergebnis