Rezeption und Nutzung von Diagnose- und Vergleichs-arbeiten an Schulen

Eine Interviewstudie mit baden-württembergischen Lehrkräften an Haupt-, Realschulen und Gymnasien

Von der Pädagogischen Hochschule Schwäbisch Gmünd
zur Erlangung des Grades
eines Doktors der Philosophie (Dr. phil.)
angenommene Dissertation

von

Sibylle Jäger

Gerstetten

Erstgutachterin:	Frau Prof. Dr. Marita Kampshoff Pädagogische Hochschule Schwäbisch Gmünd
Zweitgutachter:	Herr Prof. Dr. Axel Gehrmann Technische Universität Dresden
Fach:	Erziehungswissenschaft
Abgabetermin der Dissertation:	10. Oktober 2011

Inhaltsverzeichnis

1 Einführung 1

Theoretischer Teil: 9

2 Qualität, Steuerungsmodelle und Maßnahmen der Qualitätssicherung im deutschen Bildungswesen 9

2.1 Die Ausgangssituation im Rückblick 9

2.2 Der Qualitätsbegriff im Bildungswesen 12

2.3 Modelle zur Steuerung des Bildungswesens 18

2.3.1 Das Input-Prozess-Output-Modell 18
2.3.2 Educational Governance 21
2.3.3 Kombiniertes Steuerungsmodell 22
2.3.4 Grenzen dieser Steuerungsmodelle 24
2.3.5 Zwischenfazit 29

2.4 Maßnahmen zur Qualitätssicherung im deutschen Bildungswesen 30

2.4.1 Überblick 30
2.4.2 Evaluation als Verfahren der Qualitätsüberprüfung 31
2.4.3 Einrichtung von Qualitätsagenturen 36
2.4.4 Bildungsberichterstattung 37
2.4.5 Teilnahme an internationalen und nationalen Schulleistungsstudien 39
2.4.6 Einführung von Bildungsstandards 41
2.4.7 Überlegungen zu einer gelingenden Evaluation im Bildungssystem 43

3 Vergleichsarbeiten als Instrument der Qualitätssicherung im deutschen Bildungssystem 44

3.1 Zur Entstehungsgeschichte von Vergleichsarbeiten 44

3.1.1 Konzeption, Zielsetzung und Funktionen von Vergleichsarbeiten 46
3.1.2 Kritik an Vergleichsarbeiten 51

3.2 Vergleichsarbeiten in der Praxis 57

3.2.1 Länderübergreifende Projekte in der Grundschule – das Projekt VERA - ein Kooperationsprojekt der Bundesländer 57
3.2.2 Länderspezifische Projekte in der Sekundarstufe – am Beispiel: Vergleichsarbeiten in Baden-Württemberg 58

3.3 Implementierung von Vergleichsarbeiten als Instrument der Qualitätsentwicklung und -sicherung 65

3.3.1 Allgemeine Überlegungen und Begriffsbestimmungen 65
3.3.2 Gelingensbedingungen für die Implementierung von Vergleichsarbeiten im Mehrebenensystem 66
3.3.3 Schwierigkeiten und Probleme der Implementierung 74

3.4 Ergebnisse aus der Rezeptionsforschung zur Implementierung von Vergleichsarbeiten 77
3.4.1 Ausgewählte Rezeptionsstudien zum Schulmonitoring 78
3.4.2 Befunde zu Rezeptionstypen 91
3.4.3 Modelle zur Rezeption von Rückmeldedaten 97

Empirischer Teil: 103

Eine qualitative Interviewstudie zur Rezeption und Nutzung von Vergleichsarbeiten an baden-württembergischen Haupt-, Realschulen und Gymnasien. 103

4 Fragestellung und Forschungsdesign 104

4.1 Fragestellung 105
4.1.1 Individuelle Bedingungen der Lehrpersonen 106
4.1.2 Rezeption der Vergleichsarbeiten 106
4.1.3 Reflexion der Ergebnisse aus Vergleichsarbeiten 106
4.1.4 Mögliche Aktionen, Maßnahmen 107
4.1.5 Schulische und Kontextbedingungen 107
4.1.6 Verdichtung der Kategorien, Typenbildung 107

4.2 Forschungsdesign - Material und Methoden 108
4.2.1 Ziehung und Umfang der Stichprobe – Beschreibung der Stichprobe 109

4.3 Erhebungsmethoden 114
4.3.1 Das Leitfaden-Interview zur Erforschung subjektiver Theorien 115
4.3.2 Durchführung der Interviews - Datenerhebung und Transkription 121

4.4 Datenauswertung 125
4.4.1 Auswertung der Interviews mit der Methode der Qualitativen Inhaltsanalyse 125
4.4.2 Computergestützte Inhaltsanalyse und Darstellung der Ergebnisse 135

4.5 Typenbildung 143

4.6 Inhaltsanalytische Gütekriterien 145

5 Ergebnisse 147

5.1 Individuelle Bedingungen – Funktion der VA 148

5.2 Ergebnisse zur Rezeption 152
5.2.1 Rezeption - Hauptkategorie I: Durchführung der Vergleichsarbeit 152
5.2.2 Rezeption – Hauptkategorie II: Verständlichkeit der Rückmeldung 158

5.3 Ergebnisse zur Reflexion 163
5.3.1 Reflexion – Hauptkategorie I: Auswertung der Vergleichsarbeit 163
5.3.2 Reflexion – Hauptkategorie II: Aufgaben und Zusammenstellung der Vergleichsarbeit 171
5.3.3 Reflexion - Hauptkategorie III: Niveau der Vergleichsarbeit 177
5.3.4 Reflexion - Hauptkategorie IV: Zeitpunkt der Vergleichsarbeit 182
5.3.5 Reflexion – Hauptkategorie V: Förderdiagnostischer Nutzen der Vergleichsarbeit 187

5.4 Ergebnisse zur Aktion 196
5.4.1 Auswirkungen der Vergleichsarbeiten auf die Unterrichtspraxis 196

5.5 Ergebnisse zur Kommunikation im schulischen Kontext 202

5.6 Gesamtübersicht aller Kategorien 207

5.7 Typenbildung und typologische Analyse – zur Beschreibung spezifischer Rezeptions- und Nutzungstypen von Vergleichsarbeiten 210

5.7.1 Der begeisterte, überzeugte und befürwortende Typ 217
5.7.2 Der positiv mitspielende Typ 226
5.7.3 Der desinteressierte, gelassene Typ 235
5.7.4 Der negativ kritische Typ 241
5.7.5 Der ablehnende, belastete Typ 251
5.7.6 Zuordnung der einzelnen Interviews zu den fünf Rezeptions- und Nutzungstypen 257

6 Zusammenfassung der Ergebnisse, Diskussion und Ausblick 258

6.1 Qualitätsbegriff und Steuerungsmodelle 258

6.2 Individuelle Bedingungen am Bsp. der Funktion von Vergleichsarbeiten 262

6.3 Rezeption von Vergleichsarbeiten 267

6.4 Reflexion von Vergleichsarbeiten 273

6.5 Aktionen und abgeleitete Maßnahmen aus Vergleichsarbeiten 284

6.6 Innerschulische Kommunikation und Unterstützung 285

6.7 Fünf unterschiedliche Rezeptions- und Nutzungstypen 288

6.8 Ausblick 290

7 Literaturverzeichnis 291

8 Abbildungs- und Tabellenverzeichnis 305

1 Einführung

> *„Die Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring dient zum einen der systematischen Beschaffung von Informationen über das Bildungssystem. Zum anderen verknüpft sie diese Erkenntnisse mit Maßnahmen zur Unterrichts- und Qualitätsentwicklung, die der pädagogischen Arbeit an jeder Schule zugutekommen. Ich bin zuversichtlich, dass wir international wieder an die Spitze kommen, wenn wir diesen Weg konsequent gehen.“* (Ute Erdsiek-Rave, Präsidentin der KMK im Jahre 2006)

„Dieser Weg“, den die damalige Präsidentin der Ständigen Konferenz der Kultusminister und Kultussenatoren der Länder in der Bundesrepublik (KMK) nennt, bezieht sich auf die Umsetzung zahlreicher Maßnahmen der KMK zur Qualitätssicherung und Qualitätsentwicklung des Bildungssystems, die als Reaktion auf die moderaten Ergebnisse deutscher Schülerinnen und Schüler bei internationalen Schulleistungsvergleichen sukzessive im Bildungssystem implementiert wurden. Angestrebtes Ziel aller Maßnahmen dieser Gesamtstrategie ist die Verbesserung, Entwicklung und Sicherung der Qualität auf allen Ebenen des deutschen Bildungssystems (vgl. KMK 2007a, KMK 2007b). So umfasst sie einerseits Maßnahmen, die Erkenntnisse auf der Systemebene liefern sollen im Sinne eines Systemmonitorings. Hierzu zählen z.B. die kontinuierliche Teilnahme an internationalen und nationalen Schulleistungsstudien und die Bildungsberichterstattung von Bund und Ländern. Andererseits beinhaltet sie Instrumente, die für die Arbeit der Einzelschule und für die darin handelnden Akteure konzipiert sind. Dazu gehören neben Bildungsstandards auch Vergleichsarbeiten, mit deren Hilfe die Leistungsfähigkeit der einzelnen Schule landesweit und länderübergreifend überprüft werden soll (vgl. KMK 2006, KMK 2010b).

Die vorliegende Arbeit beschäftigt sich mit der Implementierung von Vergleichsarbeiten als Instrument der Qualitätssicherung auf der Ebene der Einzelschule und des Unterrichts in Baden-Württemberg. Adressaten sind demnach Lehrerinnen und Lehrer des allgemein bildenden Schulsystems, die verpflichtet sind, in bestimmten Jahrgängen und Klassenstufen vom Bundesland vorgegebene, standardi-

sierte Vergleichsarbeiten durchzuführen. Sie erhalten dadurch eine Rückmeldung über den landesweiten, jahrgangsbezogenen Leistungsstand ihrer Schule, ihrer Klasse und ihrer Schülerinnen und Schüler.

Diese Ergebnisse können für die Lehrpersonen eine Grundlage darstellen, um daraus spezifische Maßnahmen abzuleiten, mit dem Ziel die Qualität zu verbessern beziehungsweise bestehende Qualität zu sichern.

Im Zentrum dieser Dissertation steht die Fragestellung, welche subjektive Einstellung Lehrerinnen und Lehrer im Zuge der Implementierung von Vergleichsarbeiten entwickeln konnten. Des Weiteren stellen sich die Fragen, wie Lehrpersonen die Rückmeldungen aus Vergleichsarbeiten rezipieren und ob diese einen Beitrag zur Qualitätsverbesserung des Unterrichts oder der Einzelschule leisten können.

Den Fragestellungen wird mit Hilfe qualitativer Leitfadeninterviews, die mit 59 baden-württembergischen Lehrpersonen der Sekundarstufe I geführt wurden, nachgegangen. Die Auswertung der problemzentrierten Interviews erfolgt mit den Verfahren der Qualitativen Inhaltsanalyse nach Mayring.

Ziel dieser Arbeit ist es, den komplexen Prozess der Rezeption und Nutzung von Vergleichsarbeiten aus Sicht der Lehrpersonen zu durchleuchten und die einzelnen Phasen des Umgangs mit Vergleichsarbeiten transparent zu machen. Zudem sollen verschieden Typen gebildet werden, die sich sowohl in ihrer Einstellung zu Vergleichsarbeiten, als auch in ihrer Nutzung und im Umgang damit voneinander unterscheiden.

Mit den Ergebnissen aus dieser Forschungsarbeit ist die Hoffnung verbunden, dass sie vertiefende Erkenntnisse liefern, wie der Prozess der Implementierung gezielt und entsprechend der schulischen und akteursspezifischen Bedürfnisse ausgerichtet werden sollte, um tatsächlich einen Beitrag zur Qualitätssicherung und Qualitätsentwicklung leisten zu können.

Aufbau der Arbeit

Die Arbeit ist in einen theoretischen und einen empirischen Teil unterteilt. Der Theorieteil umfasst Kapitel 2 und 3 und gliedert sich wie folgt:

Kapitel 2 beschäftigt sich mit der Qualität, den Steuerungsmodellen und den einzelnen Maßnahmen, die zur Qualitätssicherung im Bildungswesen beitragen sol-

len. Nach der Beschreibung der Ausgangssituation (2.1) wird zunächst versucht, eine Definition des Qualitätsbegriffs zu formulieren (2.2). Im Anschluss daran werden Modelle von Steuerungskonzepten vorgestellt, die zeigen, wie Qualität in einem hochkomplexen Bildungssystem hergestellt werden kann (2.3). Den Abschluss des Kapitels bildet die Darstellung und Beschreibung konkreter Maßnahmen, die der Qualitätsentwicklung und –sicherung des deutschen Bildungssystems dienen sollen (2.4).

Das dritte Kapitel setzt sich intensiv mit der Implementierung von Vergleichsarbeiten als Instrument für die Schul- und Unterrichtsentwicklung auseinander. Am Anfang steht die Erläuterung der Konzeption und Funktion von Vergleichsarbeiten. Eine kritische Betrachtung schließt das erste Unterkapitel (3.1) ab. Kapitel 3.2 beschäftigt sich mit der praktischen Umsetzung von Vergleichsarbeiten am Beispiel des Landes Baden-Württemberg. Im Grundschulbereich nehmen mittlerweile alle Schülerinnen und Schüler an dem länderübergreifenden Projekt VERA (Vergleichsarbeiten in der Grundschule) teil. In der Sekundarstufe werden in den einzelnen Bundesländern jeweils länderspezifische Lernstandserhebungen durchgeführt, die unterschiedlich bezeichnet werden. In Baden-Württemberg heißen diese Instrumente Vergleichsarbeiten, deren Konzeption und Umsetzung in der vorliegenden Arbeit detailliert beschrieben wird. Die Implementierung von Vergleichsarbeiten, insbesondere die Formulierung von Gelingensbedingungen im Mehrebenensystem, stehen im Zentrum des Kapitels 3.3. Die Befunde aus wissenschaftlichen Projekten zur Implementierung und Rezeption von Lernstandserhebungen schließen sich daran an (3.4). Am Ende dieses Kapitels werden mögliche Modelle vorgestellt, die als Grundlage für Rezeptionsstudien geeignet erscheinen.

Der empirische Teil der vorliegenden Dissertationsschrift behandelt neben den methodischen Erläuterungen (Kapitel 4) auch die Ergebnisse der qualitativen Interviewstudie, die die Rezeption und Nutzung von baden-württembergischen Vergleichsarbeiten untersucht (Kapitel 5).

Die Arbeit schließt mit der Diskussion der Befunde und einem Ausblick auf zukünftige Forschungsprojekte ab (Kapitel 6).

Theoretischer Teil:

2 Qualität, Steuerungsmodelle und Maßnahmen der Qualitätssicherung im deutschen Bildungswesen

2.1 Die Ausgangssituation im Rückblick

Das Jahr 1997 stellt in der jüngeren Geschichte des deutschen Bildungswesens einen zentralen Zeitpunkt dar, da es als Ausgangspunkt für einschneidende bildungspolitische Änderungen und Neuerungen betrachtet werden kann (vgl. Steffens 2007). Anlass für diesen Umbruch waren die Reaktionen auf die Veröffentlichungen der Ergebnisse einer internationalen Schulleistungsstudie, der TIMS-Studie, die deutschlandweit für gehöriges Aufsehen sorgten. TIMSS steht für "Third International Mathematic and Science Study"[1] und wurde von der "International Association for the Evaluation of Educational Achievement" (IEA) organisiert. Bis dahin hatte Deutschland, im Vergleich zu anderen Ländern, im Bereich einer systematischen Evaluation des Bildungssystems durch die Teilnahme an sogenannten internationalen „Large Scale Assessments[2]" wenig Erfahrung gesammelt und noch keine lange Tradition aufgebaut. Vor allem im Zeitraum zwischen den frühen 1970er und den frühen 1990er Jahren nahm Deutschland weder an flächendeckenden internationalen Vergleichsuntersuchungen teil, noch wurden innerhalb der Landesgrenzen groß angelegte, länderinterne oder länderübergreifende Vergleichsuntersuchungen durchgeführt (vgl. Schwippert & Goy 2008, S. 392). Zuvor fanden erste deutsche Beteiligungen an den IEA-Studien ‚First International Mathematics Study' (FIMS) im Schuljahr 1963/64 und im Jahr 1970 an der ‚First International Science Study' (FISS) statt (vgl. van Ackeren 2002, S. 159ff). Eine mögliche Erklärung, warum sich Deutschland im Anschluss an diese

[1] Nachdem TIMSS weitergeführt wird, wurde das Akronym seit dem Jahr 2003 verändert. Es steht nun für „Trends in International Mathematics and Science Study"
[2] Eine ausführliche Beschreibung zur Bedeutung und Wirkungsweise von Large Scale Assessments" findet sich bei van Ackeren und Klemm 2011, S. 126f

beiden Studien an keiner internationalen Studie mehr beteiligte, sahen einige Autoren in der „eher geisteswissenschaftlichen Tradition deutscher Pädagogik“ (vgl. Bos et al. 2010, S. 253; Steffens 2007; van Ackeren & Klemm 2011, S. 137).

Mit der Veröffentlichung der Ergebnisse aus der TIMS-Studie erfuhren internationale Schulleistungsstudien in Deutschland eine größere Beachtung. Ziel der im Jahre 1994/1995 durchgeführten TIMS-Studie war es, die Schülerleistungen in Mathematik und den Naturwissenschaften in 41 Ländern der Welt miteinander zu vergleichen und zu bewerten. Die Ergebnisse fielen für Deutschland nur mittelmäßig aus. So wurde in Mathematik lediglich der 23. Platz erreicht (vgl. Baumert und Lehmann 1997). Die Nachricht über das schlechte Abschneiden bewirkte den sogenannten „TIMS-Schock“ und löste bundesweit eine Vielzahl an Reaktionen aus. Die Frage, warum die Ergebnisse so überraschend schlecht ausfallen konnten und welche Maßnahmen ergriffen werden sollten, um schlussendlich schulische Qualität zu verbessern, wurde vielfach diskutiert: eine wesentliche Erkenntnis bestand darin, dass in Deutschland bisher eine systematische und regelmäßige Überprüfung von Erträgen schulischer Bildungsprozesse fehlte (vgl. Köller 2008).

Im Oktober desselben Jahres befasste sich die Ständige Konferenz der Kultusminister und Kultussenatoren der Länder (KMK-Konferenz) auf ihrer 280. Plenarsitzung in Konstanz mit den Befunden der TIMS-Studie und beschloss Maßnahmen zur Qualitätssicherung an Schulen. Neben der Entwicklung länderspezifischer Evaluationsinstrumente sollte Deutschland in Zukunft regelmäßig an länderübergreifenden Vergleichsuntersuchungen zum Lern- und Leistungsstand von Schülerinnen und Schülern teilnehmen [Konstanzer Beschlüsse] (vgl. KMK 1997). Die erste internationale Schulleistungsstudie, an der Deutschland daraufhin teilnahm, war die PISA-Studie im Jahr 2000. PISA steht für „Program for International Student Assessment“ und wird unter der Ägide der OECD (Organisation for Economic Cooperation and development = Organisation für wirtschaftliche Zusammenarbeit und Entwicklung) im Abstand von drei Jahren mit wechselnden inhaltlichen Schwerpunkten durchgeführt. Auch die Ergebnisse aus dieser Studie bescheinigten dem deutschen Schulsystem erhebliche Defizite bei den erreichten Standards und ließen gravierende strukturelle Probleme des deutschen Bildungswesens vermuten.

Um die Qualität und Leistungsfähigkeit des deutschen Bildungssystems zu verbessern, so der allgemein vorherrschende Konsens, war man dazu bereit, grundlegende und einschneidende Änderungen des Bildungssystems vorzunehmen (vgl. Baumert 2001).

Jahre später wurden diese Entwicklungen und Entscheidungen die auf TIMSS und PISA folgten als „Revolution, die als Sattelzeit[3] in die Geschichte des Bildungswesens eingehen wird“ bezeichnet (Fend 2008b, S. 33). Mit diesem Begriff beschreibt Fend den Richtungswechsel in der Bildungsplanung, der auch „empirische Wende“ (Bos et al. 2010, S. 279) bezeichnet wird. Grundlage für bildungspolitische Entscheidungen sollten von nun an empirische Befunde liefern, die die Lernergebnisse der Schülerinnen und Schüler als den Ertrag (Output und Outcome[4]) des Bildungssystems erfassen. Bis dahin lag der Fokus auf den Input-Faktoren. Darunter versteht man die materialen und personalen Ressourcen, wie z.B. Lehrpläne oder Stundentafeln.

Alle von der KMK verabschiedeten Programme und Maßnahmen beinhalten als gemeinsame Zielsetzung die Verbesserung und Sicherung der Qualität des Bildungswesens. Somit drängt sich die Frage auf, was genau mit dem Begriff Qualität gemeint ist und wie sich Qualität definiert. Aus diesem Grund setzt sich das folgende Kapitel (2.2) mit dem Thema „Der Qualitätsbegriff im Bildungswesen“ auseinander.

Einen weiteren Schwerpunkt des zweiten Kapitels bildet die Frage, wie Qualität in einem hochkomplexen System wie dem Bildungswesen, in welchem eine Vielzahl von Akteuren auf unterschiedlichen Ebenen agiert, hergestellt werden kann. Um diesen Sachverhalt verstehen zu können, muss man die Steuerungskonzepte und –modelle des deutschen Bildungssystems ins Zentrum einer kritischen Betrachtung rücken (vgl. Oelkers und Reusser 2008, S. 18). Kapitel 2.3 stellt verschiedene Modelle zur Steuerung des Bildungswesens vor.

---

[3] Der Begriff Sattelzeit lehnt sich an die Metapher des Bergsattels an. Er beschreibt die Übergangszeit in eine neue Epoche.

[4] Unter Output versteht man die kurzfristigen Ergebnisse, mit Outcome werden langfristige und nachhaltige Resultate bezeichnet.

2.2 Der Qualitätsbegriff im Bildungswesen

Die der angestrebten Verbesserung von Qualität zugrunde liegende Definition von Qualität im Bildungssystem ist unscharf und uneinheitlich. Eine klare allgemein gültige und einheitliche Definition existiert nicht, wie die folgende Auswahl zeigt:

Eine schlichte und pragmatische Definition stammt aus Japan. Sie besagt, dass Qualität alles ist, was man verbessern kann (vgl. Steffens 2007, S. 46f).

Etwas weiter fasst die OECD diesen Begriff: „Qualität wird in einem doppelten normativen Sinn erfahren: einerseits kennzeichnet der Begriff einen Zustand im Hinblick auf die Vorstellung von Perfektion. Zweitens erlaubt er eine ordinale Attribuierung (besser, sehr gut…)“ (ebd., S. 46).

Heid postuliert, dass Qualität „*keine beobachtbare Eigenschaft oder Beschaffenheit eines Objektes*, sondern das *Resultat einer Bewertung der Beschaffenheit eines Objekts ist*“ (Heid 2000, S. 41, [Hervorhebungen im Original])

Diesen Ansatz vertreten auch Posch & Altrichter (1997). Neben einer Begriffsdefinition begründen die Autoren, worin die Schwierigkeiten bestehen, einen einheitlichen Qualitätsbegriff zu definieren. „Es wäre vergebliche Mühe, ein einheitliches Qualitätskonzept zu definieren, weil Qualität ein relativer Begriff ist und nur im Hinblick auf die Werte der verschiedenen Interessengruppen (…) näher bestimmt werden kann“ (ebd.).

Fend knüpft im Sinne der ordinalen Attribuierung an die Definition der OECD an, indem er sagt, dass der Qualitätsbegriff in den pädagogischen Diskurs lediglich eine formale Differenz einfügt. Er besagt nicht mehr, als die Unterscheidung von etwas Höherwertigem gegenüber etwas Geringerwertigem, die zunächst höchst formal bleibt. Daher ist es erforderlich, „zu einer inhaltlichen Formulierung dieser Differenz zu kommen“ (Fend 2000, S. 56).

Schließlich bietet Heid (2007) eine weitreichende Definition des schulischen Qualitätsbegriffs an, die einige der vorher genannten Überlegungen aufgreift. Er gliedert sein Resümee, das seine Auseinandersetzung zum Thema ‚Qualität von Schule’ abschließt in fünf Punkte. In seinem ersten Punkt betont er, dass „Qualität (…) das Resultat einer Sachverhaltsbeurteilung“ ist. Zu diesen „Sachverhalten gehören

auch Sachverhaltsbeschreibungen und Sachverhaltsinterpretationen“. Daran anknüpfend setzt er sich im zweiten Punkt seiner Abhandlung mit den kriterialen Bezugspunkten auseinander, die die Grundlage einer jeden Sachverhaltsbeurteilung bilden. Diese Beurteilungskriterien entstehen aus Diskussionen und Auseinandersetzungen über die Bestimmung des Schul- und Unterrichtszwecks, sie sind demnach entscheidungsabhängig. Dass sich in Beurteilungsprozessen die Einstellungen und Haltungen der verschiedenen Akteure widerspiegeln, führt er in seinem dritten Punkt aus. Demnach kommen in Beurteilungskriterien unterschiedliche Interessen zum Ausdruck. Diese können sowohl individuell als auch kollektiv formuliert sein, sie können „in der von gesellschaftlicher Praxis abhängigen soziokulturellen Tradition begründet und in ‚der Schule‘ institutionalisiert sein“ (ebd., S. 66). Als Beispiel für zwei konträre Positionen im Bereich Qualität von Schule führt er auf der einen Seite das wirtschaftliche Interesse an, nachdem Schule einen wesentlichen „Beitrag zur Sicherung oder Verbesserung des Wirtschaftsstandorts Deutschland leisten soll“ (ebd., S. 59). Auf der anderen Seite sieht er die Persönlichkeitsbildung der einzelnen Schülerinnen und Schüler. Wie sich die jeweiligen Interessen auf das unterrichtliche Handeln auswirken, behandelt Heid in seinem vierten Punkt. Er beschreibt darin, dass diese Interessen sofern sie als Bedingungen formuliert sind, die unterschiedlichen Akteure nicht zu bestimmten inhaltlichen Zweckbestimmungen zwingen können. In seinem letzten Punkt fasst er seine bisherigen Ausführungen zusammen und definiert Qualität der Schule wie folgt:

> „In pragmatischer Zuspitzung ließe sich sagen, dass ‚die Qualität der Schule‘ das messbare Maß ‚ist‘, in dem diejenige (Bildungs-)Praxis, durch die das mit ‚Schule‘ jeweils Gemeinte, bzw. Operationalisierte auch bewirkt, was sie bezweckt. Damit wird ‚Qualität‘ mehrstufig relativiert, und zwar auf die in der Regel strittigen Zweckbestimmung, mit der gesellschaftspolitisch und bildungspolitisch konkurrierende Subjekte dieser Zweckbestimmung ‚die Schule‘ jeweils definieren, praktizieren oder auch kritisieren“ (Heid 2007, S. 66, [Hervorhebungen im Original]).

An diese Definition von Heid schließt die Überlegung an, welche Subjekte im Bildungswesen agieren, bzw. nach Heid, miteinander konkurrieren.

Hier liefert Fend (2000) eine Antwort. Er unterscheidet in seiner Betrachtung des Bildungssystems mehrere Gestaltungsebenen, die in Abhängigkeit zueinander stehen. Er stellt fest, dass die Qualitätsfrage mehrebenenanalytisch differenziert werden muss und zwar auf der Ebene des Systems (Makroebene), der Einzelschule (Mesoebene) und des Unterrichts (Mikroebene). Dabei entscheidet nicht eine Ebene oder ein Segment allein, sondern deren Zusammenspiel. Die Akteure der einzelnen Ebenen haben „jeweils unterschiedliche Optimierungskriterien vor Augen“ (Fend 2000, S. 56). Die Produktivität des gesamten Systems hängt letztlich davon ab, wie weit diese Kriterien übereinstimmen bzw. sich entsprechen. Auf jeder Ebene findet eine Normierung der Bildungsqualität statt, wie im Folgenden exemplarisch dargestellt wird:

Qualität auf Systemebene: Die bildungspolitische und – administrative Strukturebene umfasst keine einheitliche nationale Ebene, sondern berücksichtigt föderale und regionale Ebenen, die im deutschen Bildungssystem aufgrund des Föderalismus existieren. Die Länder zeichnen sich für die Ausgestaltung des Schulsystems verantwortlich.

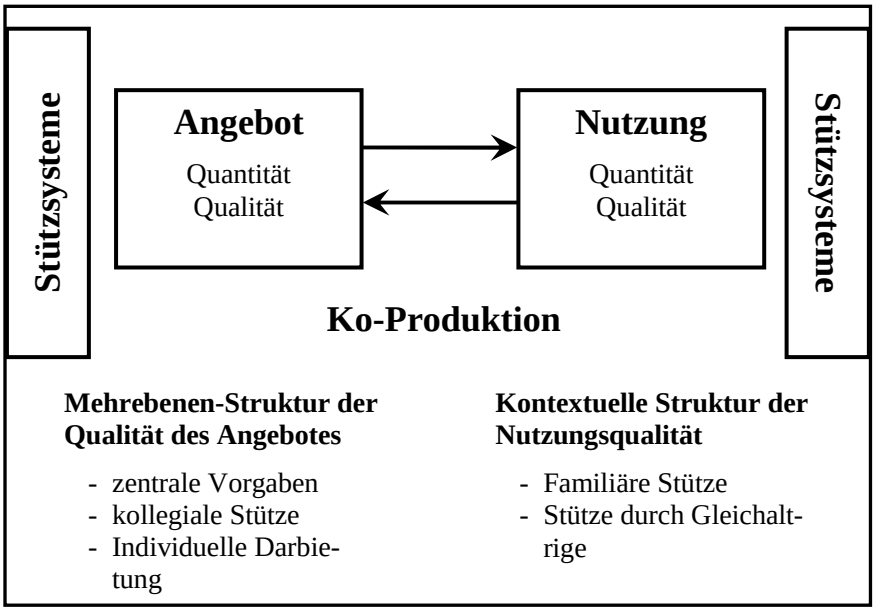
Auf dieser Ebene wird Qualität bisher z.B. über Lehr- und Bildungspläne, über erreichte Bildungsabschlüsse oder über die Ermittlung von Übertrittsquoten etc. normiert. Mittlerweile werden hier zusätzlich auch Instrumente eingesetzt, die die Leistungsfähigkeit eines Schulsystems im internationalen und nationalen Vergleich messen, indem die Erreichung der gesetzten Standards überprüft wird (vgl. Fend 2005, S. 15). Hierzu zählen z.B. die zahlreichen internationalen und nationalen Schulleistungsstudien an denen Deutschland teilnimmt.

Qualität auf Einzelschulebene: Auf der Ebene der Einzelschule als pädagogische Handlungseinheit liegen zahlreiche Erkenntnisse aus der internationalen Schuleffektivitätsforschung vor. In Anlehnung daran legen fast alle Bundesländer für die Schulen sogenannte Referenzrahmen fest, „die explizit formulieren, was Qualität in den Augen der für die Systemsteuerung Verantwortlichen heißt“ (Fend 2008a, S. 202). Diese Rahmen beinhalten unter anderem neben der Erstellung eines gemeinsamen Leitbildes auch verschiedene Evaluationsverfahren oder Handlungskataloge, die an die unterschiedlichen Akteure (Schulführung, Schulorganisation und Kollegium) adressiert sind.

Qualität auf Unterrichtsebene: Auf der Mikroebene des Lehrerhandelns und des Unterrichts hängt es vor allem von der Professionalität der Lehrpersonen ab, wie weit es ihnen gelingt, passende Lerngelegenheiten für ihre Schülerinnen und Schüler bereitzustellen. Aus der Unterrichtsforschung liegen Kriterien für guten Unterricht vor, die bei der Planung, Durchführung und Reflexion von Lehr-Lernsituationen zu beachten sind. (Vgl. Fend 2008b, S. 204ff; Helmke 2006; Helmke & Schrader 2006; Helmke & Weinert 2007, S. 63f).

Nach Fend (2000) kommt der Unterrichtsebene bei der Qualitätssicherung und –entwicklung eine herausragende Bedeutung zu, weil sich „Qualität letztlich in einer Verbesserung der Lern- und Entwicklungsprozesse von Kindern und Jugendlichen manifestieren muss“ (ebd., S. 56). Aus diesen Überlegungen heraus konzipiert er ein einfaches „Modell der Qualität im Bildungswesen“, das sich auf das Zusammenwirken (die Ko-Produktion) von Lehrenden und Lernenden fokussiert. Die Lehrenden unterbreiten ein Angebot, das die Lernenden für sich nutzen. Qualität kommt demnach immer nur durch eine optimale „Synchronisierung von Angebotsmerkmalen und Nutzungsmöglichkeiten“ zustande, wobei Stützsysteme und Kontextbedingungen vorausgesetzt werden müssen, die auch außerhalb des Unterrichts angesiedelt sind (vgl. ebd., S. 57).

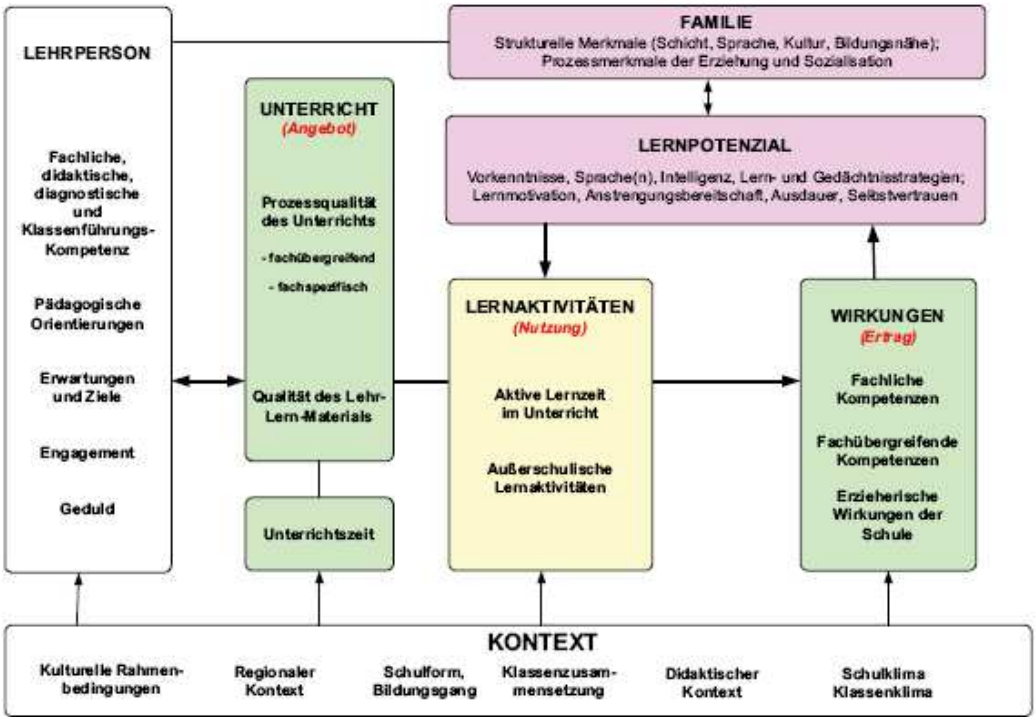
Abbildung 1: Modell der Qualität im Bildungswesen nach Fend (2000, S. 58)



Diese „Denkfigur“ (Helmke & Schrader 2006, S. 7) bietet die Ausgangsbasis weiterer Angebots-Nutzungs-Modelle, wie sie z.B. von Helmke (2006) vorliegen (Abbildung 2). Er präzisiert Fend’s Grundmodell, indem er Faktoren nennt, die für die Qualität und Quantität des Unterrichtsangebots, sowie des Lehr-Lern-Materials verantwortlich sind. Zum einen spielt in seinem Modell die Nutzerseite eine herausragende Rolle. Denn die Lernvoraussetzungen der Schülerinnen und Schüler lassen sich sowohl in kognitive als auch in motivationale, volitionale und soziale Merkmale und Kompetenzen untergliedern. Diese werden von familiären Bedingungen und dem häuslichen Umfeld der Schülerinnen und Schüler beeinflusst.

In einem weiteren Punkt berücksichtigt Helmke die Kontextabhängigkeit der Unterrichtsqualität, die einerseits von der Professionalität der Lehrperson und andererseits von der Zusammensetzung der Klasse und den einzelnen Individuen abhängt. In „schwierigen“ Klassen benötigt man unter Umständen mehr Zeit für eine gelungene Klassenführung als in „einfachen“ Klassen (Helmke & Schrader 2006, S. 8).

Abbildung 2: Ein Angebots-Nutzungs-Modell des Unterrichts nach Helmke (Helmke 2006; Helmke & Schrader 2006)



Zwischenfazit

Wie gezeigt, stößt man bei dem Versuch, den schulischen Qualitätsbegriff zu definieren einerseits auf Grenzen in dem Sinne, dass er zu abstrakt bleibt und in kein konkret anwendbares Handlungsmuster eingebettet werden kann. Andererseits steht den handlungsorientierten Ansätzen ein derart komplexes und beinahe unüberschaubares Bildungssystem gegenüber, indem Qualität auf verschiedenen Ebenen spezifisch von den jeweiligen Akteuren normiert wird. Wie diese Ebenen zusammenspielen und welche Effekte beim Übergang von einer Ebene zur nächsten zu Tage treten, konnte bisher nicht abschließend geklärt werden. „Gute schulische Leistungen bei Gewährleistung einer ausgewogenen Verteilung der Bildungschancen sind vielmehr das *Zusammenspiel einer Reihe von Faktoren* außerhalb und innerhalb eines Systems sowie von Faktoren auf verschiedenen Ebenen eines Bildungssystems“ (van Ackeren & Klemm 2011, S. 142). Im Bereich Qualität und Schule stehen mittlerweile zahlreiche Veröffentlichungen zur Verfügung (vgl. Büeler 1998; Helmke et al. 2000; Eder 2002; Bülow-Schramm 2006; Buer & Wagner 2007).

In den vergangenen Jahren wurden teilweise auch Qualitätssysteme aus der Wirtschaft in das System Schule übernommen. Hierzu zählen z.B. TQM (Total Quality Management), ISO (International Organization for Standardization) und EFQM (European Foundation for Quality Management). Diese Übernahme ist jedoch mit zahlreichen Übertragungsproblemen verbunden und führt zu Zweifeln an deren Gültigkeit und Eignung. Im berufsbildenden Sektor des Bildungssystems finden diese Qualitätssysteme am ehesten Anwendung, da dort die wirtschaftliche Nähe am stärksten gegeben ist (vgl. Schratz 2001). Arnold und Faber (2000) nehmen in ihrer Expertise „Qualitätssysteme und ihre Relevanz für Schule“ verschiedene Systeme, die außerhalb des Schulbereichs bestehen, in den Blick und untersuchen diese auf mögliche Impulse und Beiträge für das schulische Qualitätsmanagement. In der vorliegenden Arbeit finden diese Ansätze keine Verwendung.

Die bloße Betrachtung der Definition und Normierung von Qualität im Bildungswesen blendet weitere wichtige Aspekte aus, die zum vertieften Verständnis der Wirkungsweise eines Bildungssystems beitragen können. Hierzu zählt zum Beispiel die Auseinandersetzung mit der Funktionsweise und Steuerung von Bildungssystemen.

Um diese Funktionsweise analysieren zu können, benötigt man zunächst einmal theoretische Rahmenkonzeptionen in Form von heuristischen Modellen, mit denen man die Strukturen und Prozesse der Steuerung des vielschichtigen Bildungssystems darstellen kann. Mit ihrer Hilfe können Abläufe und Zusammenhänge zwischen den einzelnen Elementen und Ebenen dargestellt und erklärt werden. Aus diesem Grund folgt im nächsten Kapitel die Vorstellung von System-Steuerungsmodellen, die auch in aktuellen Studien zur Untersuchung der Qualität in Bildungssystemen verwendet werden.

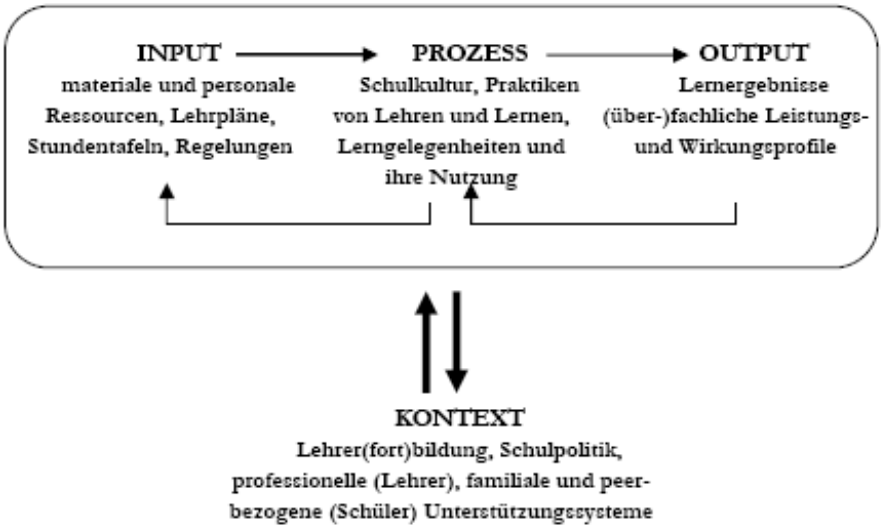
2.3 Modelle zur Steuerung des Bildungswesens

Bei der Betrachtung von Steuerungsmodellen des Bildungssystems lassen sich zwei grundsätzliche Typen unterscheiden. Dem einen Modell liegt ein produktions- und prozessorientierter Ansatz zugrunde, der die Qualitätsentwicklung des Bildungssystems aus ökonomischer Sicht beschreibt. Es unterscheidet Input, Prozess und Output/Outcome und wird in Kapitel 2.3.1 erläutert. Das zweite Modell umschreibt einen mehrperspektivischen Ansatz, der die Zusammenhänge zwischen den einzelnen Systemebenen des Bildungswesens und der darin agierenden Akteure in den Blick nimmt (siehe Kapitel 2.3.2).

2.3.1 Das Input-Prozess-Output-Modell

Dieses Steuerungsmodell wird in den internationalen Schulleistungsstudien der OECD „als einfaches Grundmodell“ (Fend 2008b, S. 21) verwendet und ermöglicht dadurch den Vergleich der Schulsysteme zwischen den unterschiedlichsten Ländern. Es lehnt sich an das CIPP-Modell von Stuffelbeam an, das bereits Ende der 1960er Jahre veröffentlicht wurde. CIPP steht für **C**ontext, **I**nput, **P**rocess, **P**roduct. Das Modell unterscheidet also Input-, Prozess- und Output-, sowie Kontextfaktoren, wie Abbildung Nr.3 auf der folgenden Seite zeigt.

Abbildung 3: Input-Prozess-Output-Modell nach Oelkers & Reusser (2008, S. 19)



Das zugrunde liegende Input-Prozess-Outputmodell ermöglicht eine lineare Beschreibung der Funktionsweise des Bildungssystems, indem es den Produktionsprozess schrittweise fokussiert. Jeder Schritt kann für sich evaluiert und somit die Qualität in jeder Phase des Prozesses kontrolliert werden.

Bis in die 1990er Jahre hinein dominierte im deutschen Bildungssystem die sogenannte Inputsteuerung, die durch die Zuteilung sachlicher, personeller und finanzieller Ressourcen und rechtlicher Verordnungen die Qualität sichern sollte (vgl. Galiläer 2005, S. 38ff; Köller 2007, S. 93; Zlatkin-Troitschanskaia 2007a, S. 187). Die Qualität der Prozesse, Kontexte und Resultate und die damit zusammenhängende systematische Überprüfung der Bildungserträge wurden kaum beachtet (vgl. Kuper 2002, S. 536f; Oelkers & Reusser 2008, S. 17).

Erst durch die aufgezeigten Mängel des deutschen Bildungssystems durch internationale Schulleistungsstudien geriet die lediglich auf die Input-Faktoren reduzierte Betrachtung in Kritik (vgl. Galiläer 2005, S. 18). Bei der Entwicklung eines ‚neuen' Steuerungsmodells wurde auf Erfahrungen der Länder zurückgegriffen, die in den PISA-Studien erfolgreich abgeschnitten hatten. Die Studie „Vertiefender Vergleich der Schulsysteme ausgewählter PISA-Teilnehmerstaaten" wurde vom Bundesministerium für Bildung und Forschung in Auftrag gegeben (vgl. Döbert et al. 2003).

Alle Länder wiesen in ihren Steuerungskonzepten deutliche Elemente der Outputorientierung auf, wie z.B. die Entwicklung von Bildungsstandards, eine verstärkte Autonomie der Einzelschule, den Ausbau schulinterner, differenzierter Bildungsangebote und eine verstärkte Outputkontrolle mit Hilfe von Evaluationsverfahren (vgl. Zlatkin-Troitschanskaia 2007b, S. 67 f). Maier (2009b) fasst diese Elemente unter dem Begriff testbasierte Schulreform (TBSR) zusammen.

In der Literatur wird häufig vom Paradigmenwechsel in der Steuerung des Bildungssystems gesprochen; von der Input- hin zur Outputsteuerung. Es fand jedoch kein Steuerungswechsel statt. Das bisherige Input-Modell wurde lediglich um Elemente der Outputsteuerung erweitert (vgl. Hoffmann-Göttig et al. 2005, S. 32; Köller 2007, S. 93; Zlatkin-Troitschanskaia 2007b, S. 81;). Berkemeyer et al. (2008, S. 153) sprechen hierbei von einer outputorientierten Inputsteuerung.

Das Input-Prozess-Outputmodell bleibt in der Beschreibung der Steuerungsfunktion relativ abstrakt. Fend hält das Modell für Ländervergleiche geeignet, für die Entwicklung länderspezifischer Konzepte fordert er jedoch Modelle, die stärker handlungsorientiert sind. Seiner Meinung nach muss die Produktionsfunktion des Bildungswesens handlungstheoretisch rekonstruiert werden, um sie aus der abstrakten Konzeption von Input und Output, die als Schema den OECD Studien zugrunde liegt, herauszuführen (vgl. Fend 2008a, S. 192).

Hier knüpft er an Ergebnisse an, die er bereits im Jahre 1986 bei seinen Untersuchungen zur Schulqualität im Rahmen der Konstanzer Schulforschung veröffentlichte. Sie beschreiben die Einzelschule als „pädagogische Handlungseinheit“, die sich aus verschiedenen Handlungsebenen zusammensetzt und bereits im Qualitätskapitel (2.2) erläutert wurde (vgl. Fend 1986). Weitere Forschungen belegen, dass vor allem die Einzelschule als Gestaltungseinheit und eigentlicher „Motor der Entwicklung“ verstanden werden muss (vgl. Rolff 1995, S. 293; Rolff 1998, S. 296). Die verschiedenen Akteure, die in Einzelschulen auf den unterschiedlichen Ebenen agieren, erhalten durch eine Ausweitung der Autonomie ihrer Institution erweiterte Entscheidungs- und Handlungsspielräume. Dies eröffnet ihnen einen gewissen Gestaltungsspielraum, der im mehrperspektivisch angelegten Konzept und Steuerungsansatz der Educational Governance Berücksichtigung findet.

2.3.2 Educational Governance

„Governance“ wird seit den 1980er Jahren im Sinne eines integrativen Ansatzes der Systemsteuerung in den Politik- und Sozialwissenschaften und in letzter Zeit zunehmend häufig in der Bildungsforschung unter dem Namen „Educational Governance“ verwendet. (Vgl. Altrichter et al. 2007a; Altrichter et al. 2007b; Altrichter & Maag Merki 2010; Berkemeyer 2010; Brüsemeister & Eubel 2008a; Heinrich 2007; Kussau & Brüsemeister 2007; Wissinger 2007).

Hinter dem Begriff verbirgt sich ein Steuerungsmodell, welches das gesamte Bildungswesen in den Blick nimmt und die Gestaltungsverantwortung auf den verschiedenen Ebenen des Bildungssystems ins Zentrum der Betrachtung rückt. Dies erfolgt derart, dass die Gestaltungsinstrumente auf der bildungspolitischen, der schulischen und der unterrichtlichen Ebene ausgearbeitet und miteinander verbunden werden (vgl. Fend 2008b, S. 13). Mit diesem Modell sollen Zusammenhänge und Strukturen komplexer gesellschaftlicher Systeme erklärt und analysiert werden können (vgl. Zlatkin-Troitschanskaia 2007b, S. 77; Fuchs 2008, S. 19ff; Langer 2008, S. 8;). Das Schulsystem wird in der Governance - Perspektive als eine solche komplex gestaltete, soziale Ordnung mit bestimmten Leistungsmerkmalen angesehen. Reformen im Schulsystem sind Versuche, diese soziale Ordnung zu transformieren (vgl. Bähr 2003; Altrichter 2009, S. 244). Mit Blick auf die Akteure stehen bei Governance das Verhältnis von staatlichem, gesellschaftlichem, privatem und privatwirtschaftlichem Handeln und die damit verbundenen Steuerungsprozesse, Verflechtungen, Wirkungen und Einflussmöglichkeiten im Vordergrund (vgl. Fuchs 2008, S. 20).

Fend (2008a) beschreibt den Wert und Nutzen des Educational Governance – Ansatzes als Beitrag zur Erklärung der Funktionsweise von Bildungssystemen wie folgt:

> „Die neue handlungstheoretische, institutionsorientierte und akteursensitive Konzeption des Bildungswesens blendet Strukturen nicht aus, sondern integriert sie in der Gestalt institutioneller Regelzusammenhänge. Handeln von Akteuren in Institutionen ist danach regelgeleitetes Zusammenhandeln. Im Bildungswesen ist es auf die Erfüllung der zentralen Aufgabe ausgerichtet, die das Bildungswesen leitet.

> Sie besteht darin, ein großes kulturelles Inhaltsprogramm an die kognitiven und emotionalen Strukturen heranwachsender Kinder und Jugendlicher anschlussfähig zu machen, letzteren also zu einem kulturgeleiteten Lern- und Aufbauprozess von Denken, Handeln und Empfinden zu verhelfen. Dieses komplexe Unternehmen wird auf mehreren Handlungsebenen realisiert und gestaltet; auf der Ebene von Bildungspolitik und Verwaltung, der einzelnen Schule und des Unterrichts von Lehrpersonen“ (ebd, S. 192).

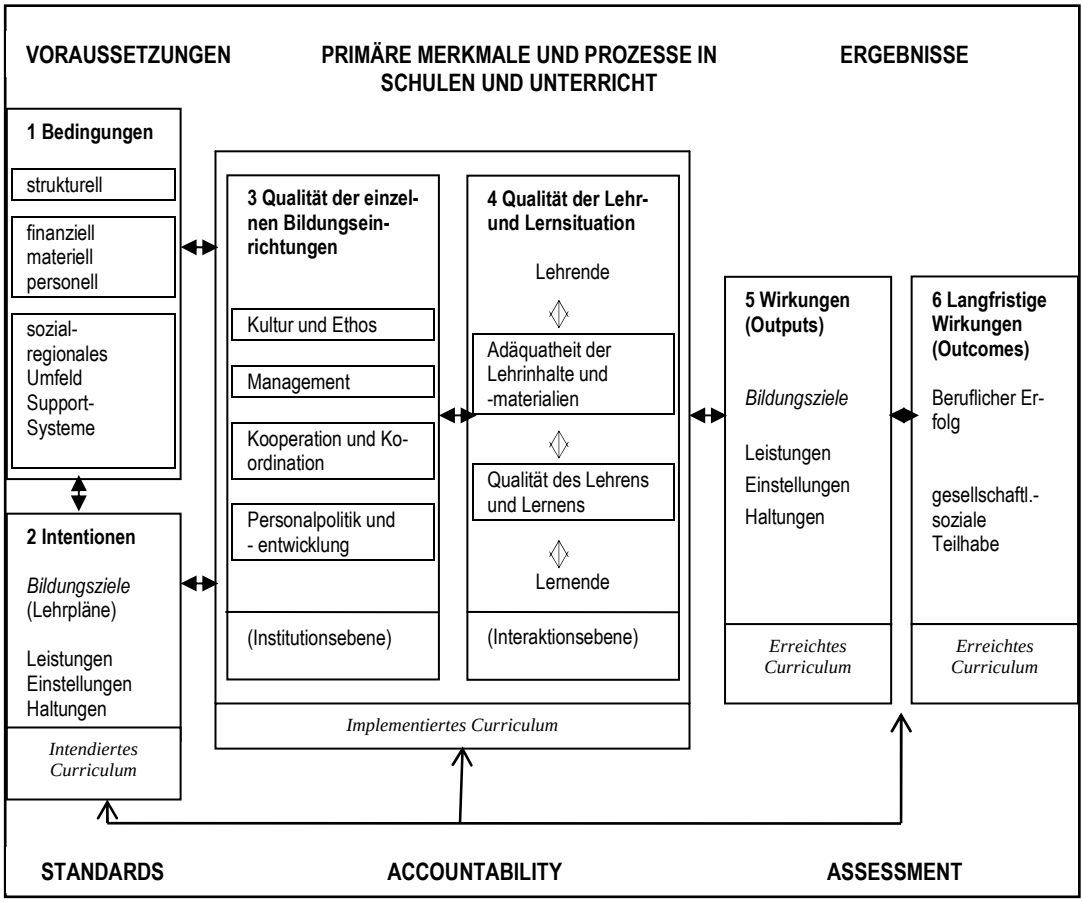
Welches der beiden Modelle sich zur Erläuterung der „Steuerung des Bildungssystems“ am besten eignet ist schwer zu beurteilen. Jedes berücksichtigt einen wichtigen Aspekt bei der Erklärung der Wirkungsweise des Bildungssystems. Mittlerweile liegen auch Modelle vor, die beide Ansätze miteinander kombinieren. Ein Beispiel für ein „kombiniertes Steuerungsmodell“ stellt das Modell von Ditton (2007) dar.

2.3.3 Kombiniertes Steuerungsmodell

Ditton skizziert ein Modell zur Steuerung des Bildungswesens, das sich aus der Kombination beider oben beschriebener Strukturmodelle des Bildungswesens herleiten lässt. Einerseits verwendet es den prozessorientierten Ansatz, der die Transformation von Eingangsbedingungen und Intentionen in Lernergebnisse, also die Abfolge von Inputs (Voraussetzungen), Prozessen (primäre Merkmale und Prozesse in Schulen und Unterricht) und Outputs, als kurzfristige und Outcomes als langfristige Wirkungen (Ergebnisse) beschreibt. Andererseits integriert es die Mehrebenensicht, die besagt, dass bei der Betrachtung von Bildungsprozessen mehrere beteiligte Ebenen einzubeziehen und zu unterscheiden sind. Bei den einzelnen Ebenen werden analog zu Fend Mikro-, Meso- und Makroebene unterschieden. Zur Mikroebene zählen hierbei die Ebene der einzelnen Individuen (Lernende und Lehrende) und die Ebene der Interaktionen zwischen Lehrenden und Lernenden bzw. die Ebene des Unterrichts. Auf der Mesoebene lassen sich die einzelnen Bildungseinrichtungen verorten, die ihrerseits in einen sozialregionalen Kontext eingebettet sind. Die Makroebene beschreibt diesen sozialräumlichen bzw. gesellschaftlich-kulturellen Kontext, zu dem auch die Unterstützungssysteme zählen. „Diese Struktur kann als vier-plus-Modell verstanden wer-

den, da die Kontextebene weiter differenziert werden könnte, z.B. nach regionalen Strukturen oder Merkmalen des Schulsystems“ (Ditton 2007, S. 84).

Abbildung 4: Modell zur Qualität und Qualitätsentwicklung im Bildungsbereich (Ditton 2000, S. 79)



Das Kernstück des Modells, das zur eigentlichen Sicherung oder Verbesserung der schulischen Qualität beiträgt, bildet die Wechselwirkung zwischen Schul- und Unterrichtsqualität (Faktoren 3 und 4). Unter Schulqualität werden Aspekte zur Schulkultur, zum Schulmanagement, zur Kooperation und Koordination und zur Personalentwicklung zusammengefasst. Die Unterrichtsqualität bezieht sich auf die Interaktionsebene zwischen Lehrenden und Lernenden. Die Voraussetzungen umfassen einerseits die Bedingungen (Faktor 1) und andererseits die Intentionen (Faktor 2). Das Modell unterscheidet kurzfristige (Outputs) und langfristige Wirkungen (Outcomes) (Faktoren 5 und 6).

Zusätzlich integriert Ditton das Steuerungspotential von Lehrplänen in seinem Modell, da Lehrpläne als zentrales Instrument der Qualitätssicherung aufgefasst werden. Die Lehrplanforschung untersucht die Wirksamkeit von Lehrplänen und unterscheidet die Begriffe intendiertes, implementiertes und erreichtes Curriculum. Das **intendierte Curriculum** beschreibt zunächst die erwarteten, geplanten Inhalte, die erreicht werden sollen. Das **implementierte Curriculum** umfasst die tatsächlich ausgewählten Inhaltsbereiche, die im Unterricht behandelt werden. Das **erreichte Curriculum** kann mittels Evaluationsverfahren getestet werden. Somit kann überprüft werden, ob die geplanten Absichtserklärungen bloßes Wunschdenken bleiben, oder ob tatsächlich Lehrplanwirkungen verzeichnet werden können (vgl. Fend 2008b, S. 81ff.).

Welches der beschriebenen Modelle letztendlich zur Erklärung der Funktionsweise und Steuerung von Bildungssystemen herangezogen wird, muss jeweils spezifisch begründet werden. Bevor die Wahl auf das eine oder andere fällt, sollten sich die Forscherinnen und Forscher auch mit den Grenzen und Problemen der unterschiedlichen Modelle auseinandersetzen. Im folgenden Kapitel werden sowohl mögliche Schwachpunkte des prozessorientierten, als auch des Mehrebenenmodells beschrieben.

2.3.4 Grenzen dieser Steuerungsmodelle

Oelkers und Reusser (2008) zeigen in ihrer Expertise zur Implementierung von Bildungsstandards mögliche Schwierigkeiten des Input-Prozess-Output-Modells auf. Ihrer Meinung nach liegt das Kernproblem des Modells bei der implizierten Rückübersetzung von Output in Input und Prozesse, die sie als „Rückverflüssigung“ bezeichnen. Bei diesem Rückverflüssigungsprozess ist einerseits die einzelne Lehrperson gefordert, aus den erzielten Ergebnissen (Outputs bzw. Outcomes) Rückschlüsse zu ziehen und diese in wirksame Input- und Prozesselemente umzuwandeln. Diese Übersetzungsleistung stellt eine professionelle Kernkompetenz von Lehrerinnen und Lehrern dar. Andererseits spielen hierbei nicht nur die einzelnen Lehrpersonen eine zentrale Rolle, sondern es müssen bei diesem Übersetzungsprozess weitere Akteure des Systems berücksichtigt werden. Hier rückt also die Governanceperspektive ins Zentrum der Betrachtung.

Die Lehrerinnen und Lehrer werden in der ersten Perspektive als die zentralen Akteure der Systemreform angesehen. Bei dieser Betrachtungsweise droht die Gefahr, Lehrpersonen nahezu ausschließlich die Gesamtverantwortung für eine gelingende Reform zu übertragen.

Die zweite Perspektive öffnet den Blick für weitere Mitspieler und Mitspielerinnen des Systems, die zum Teil auf unterschiedlichen Ebenen agieren, und deren Handeln für eine gelingende Umsetzung von Systemveränderungen ebenfalls einen wichtigen Beitrag darstellen. In dieser Mehrebenenbetrachtung kommt den Übergängen zwischen verschiedenen Ebenen und Akteuren eine besondere Bedeutung zu, die Altrichter dergestalt beschreibt:

> „In der Governance-Perspektive ist das Schulwesen als ein Mehrerbenensystem bezeichnet worden. Ein Charakteristikum solcher Gebilde ist, dass die ‚Übergänge' zwischen verschiedenen Ebenen und Akteuren prinzipiell prekär sind, weil sie gleichzeitig Übergänge zwischen verschiedenen Wissensformen und Handlungslogiken darstellen. Für die Koordination (‚Steuerung') eines Systems sind diese Übergänge ganz entscheidend. Relativ stabile Systeme entwickeln dafür spezielle ‚Brücken' wie Routinen, Umgangsformen, Verfahrensvorschriften, Instrumente, Traditionen oder Prüfverfahren, die der Brüchigkeit der ‚Brücken' und der Abschottung der Akteure entgegen arbeiten sollen" (Altrichter 2008, S. 101f).

Fend nimmt genau diese Übergänge ins Visier und entwickelt die Theorie der Rekontextualisierung, die in Abbildung 5 dargestellt ist. Mit diesem Konzept der Rekontextualisierung im Bildungswesen soll der aktive Gestaltungsanteil von Akteuren auf der jeweiligen Ebene betont werden. Dabei werden gleichzeitig die Rahmenvorgaben einer übergeordneten Ebene und die spezifischen Handlungsbedingungen auf einer untergeordneten berücksichtigt. Dies führt zu einer Theorie der Verschränkung von Handlungsebenen, die für das Verständnis von Gestaltungsmöglichkeiten einen zentralen Stellenwert bekommen wird (vgl. Fend 2008b, S. 26).

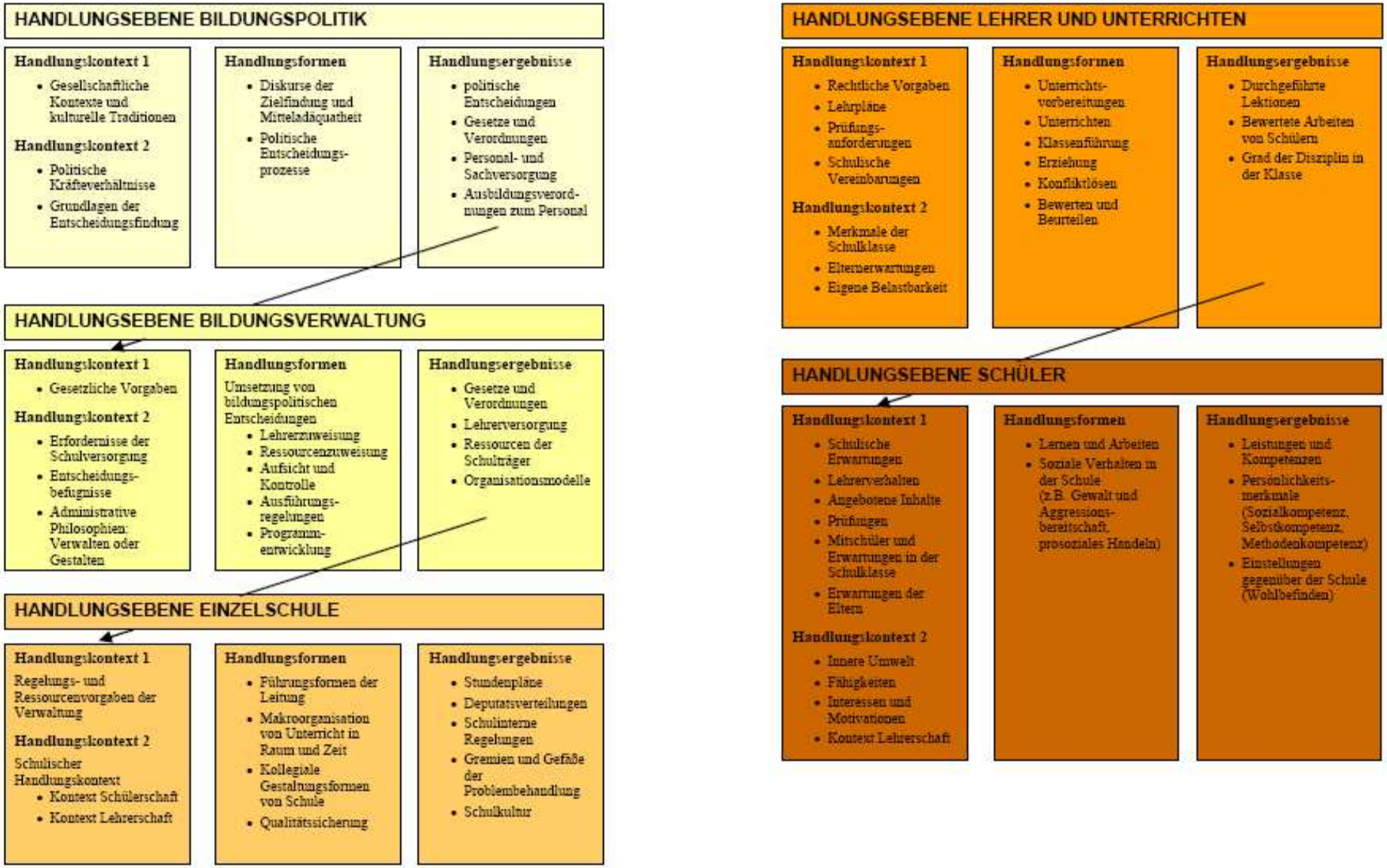
Den Zusammenhang zwischen den unterschiedlichen Handlungsebenen und ihren Kontexten beschreibt er wie folgt:

> „Auf jeder Handlungsebene (…) sind die Vorgaben auf der jeweils übergeordneten Ebene Umwelten des Handelns, die eine Umsetzung an die jeweils neuen, ebenenspezifischen Besonderheiten der Umwelten des Handelns erfordern“ (Fend 2008a, S. 195).

Fend unterscheidet hierbei zwei unterschiedliche Handlungskontexte. Der Handlungskontext 1 wird immer in Form von Handlungsergebnissen durch die übergeordneten Ebenen vorgegeben (siehe Pfeile). Sie bilden demnach den vorgegebenen Rahmen, den es auf der jeweiligen Ebene zu interpretieren und zu gestalten gilt. Am Beispiel der Handlungsebene „Lehrer und Unterrichten“ bestimmen z.B. die rechtlichen Vorgaben, Lehrpläne etc. den Handlungskontext 1 der Lehrerinnen und Lehrer. Der Handlungskontext 2 umfasst die spezifischen Bedingungen der jeweiligen Ebene. In unserem oben genannten Beispiel (Ebene der Lehrpersonen) bezieht er sich z.B. auf die Schulklasse, die Erwartungen der Schüler- und der Elternschaft, aber auch auf die individuelle Belastbarkeit des Einzelnen.

> „Auf jeder Handlungs- bzw. Gestaltungsebene richtet sich der Blick der Akteure einmal nach ‚oben’, einmal nach ‚unten’. Der Blick nach ‚oben’ bedeutet, dass die Vorgaben der übergeordneten Ebene rekontextualisert werden. Der Blick nach ‚unten’ bedeutet, dass die Folgewirkungen und Konsequenzen des Handelns berücksichtigt werden müssen. Der Blick in die Horizontale wiederum bedeutet, dass die genuinen Handlungserfordernisse auf der jeweiligen Handlungsebene berücksichtigt werden“ (Fend 2008b, S. 28).

Abbildung 5: Rekontextualisierung im Bildungswesen nach (Fend 2008b)



Auch Altrichter (2009) beleuchtet in der Perspektive der Educational Governance die Übergänge zwischen den verschiedenen Ebenen und Akteuren. Er führt vier Argumente an, die die Schwierigkeiten und Grenzen eines komplexen Steuerungsmodells aufzeigen:

Zum einen wirken bei der Steuerung des Bildungswesens sehr viele Akteure mit, die multiple Interessen verfolgen. Demnach ergibt sich die Steuerung des Systems „aus den Transaktionen aller relevanten SystemmitspielerInnen, die sich mit unterschiedlichen Interessen und durchaus unterschiedlichen Gestaltung- und Verhinderungsmacht ‚an der Gestaltung und an der Leistung ihres Systems' beteiligen" (ebd., S. 242).

In einem zweiten Punkt führt er die Eigenlogiken und Eigendynamiken von Akteuren und Systemen an. Jede Vorstellung von Steuerung muss davon ausgehen, dass sich bei Reformen automatisch Entwicklungen vollziehen können, die nicht der intendierten Zielsetzung entsprechen.

In seinem dritten Punkt bezieht er sich auf Fend und erläutert den Prozess der Rekontextualisierung, der die Zusammenhänge der einzelnen Ebenen des Bildungssystems thematisiert.

Den vierten Punkt beschreibt er als „Intentionale Gestaltung mit teilweise transintentionalen Ergebnissen". Demnach handeln Steuerungsakteure zielgerichtet und rational, indem sie versuchen, die Entwicklungen gemäß ihrer spezifischen Absichten zu beeinflussen. Allerdings stellt er fest, dass in komplexen Situationen „eine Vielzahl solcher Steuerungsakteure mit mehr oder weniger divergierenden Intentionen" agieren und somit die „Möglichkeit ihrer ‚Rationalität' (…) begrenzt" bleibt (ebd., S. 242).

2.3.5 Zwischenfazit

In den beiden vorangegangenen Kapiteln „Qualitätsbegriff im Bildungswesen“ und „Modelle zur Steuerung des Bildungswesen“ konnte gezeigt werden, dass die Herstellung, Verbesserung und Sicherung von Qualität im Bildungssystem ein vielschichtiger Prozess ist, der von vielen Faktoren moderiert und beeinflusst wird. Den Ausgangspunkt eines solchen Qualitätsprozesses bilden konkrete Maßnahmen oder Instrumente, die von Entscheidungsträgern der Bildungsadministration und Bildungspolitik entwickelt und verabschiedet werden. Diese Maßnahmenpakete beinhalten bestimmte Vorstellungen und Zielsetzungen von Qualität und sind an unterschiedliche Akteure im Bildungswesen adressiert. Setzt man sich mit der Qualitätsentwicklung des gesamten Bildungssystems auseinander, dann muss man zum einen alle Maßnahmen berücksichtigen, die hierzu einen Beitrag leisten sollen. Zum anderen müssen mögliche Zusammenhänge zwischen den einzelnen Maßnahmen betrachtet werden, um ein Verständnis für deren Zusammenwirken zu entwickeln und eventuell beabsichtigte Synergieeffekte erkennen zu können.

Des Weiteren ist es notwendig, die jeweiligen Adressaten und deren Handlungslogiken zu bedenken, da sie als relevante Systemmitspieler unterschiedliche Interessen verfolgen können und verschiedene Gestaltungsmöglichkeiten bei der Umsetzung der spezifischen Maßnahmen haben.

Aus diesen Überlegungen erfolgt im nächsten Kapitel zunächst die Vorstellung und Beschreibung aller relevanten Maßnahmen, die zur Qualitätssicherung des deutschen Bildungswesens entwickelt wurden und ihren Beitrag leisten sollen. Sie sind in der „Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring“ (KMK 2006) zusammengefasst.

2.4 Maßnahmen zur Qualitätssicherung im deutschen Bildungswesen

2.4.1 Überblick

Wie bereits in Kapitel 2.1 dargestellt, bildeten die moderaten Leistungsergebnisse der deutschen Schülerinnen und Schüler bei der TIMS- und PISA-Studie die Ausgangsbasis für einschneidende bildungspolitische Maßnahmen mit dem Ziel, die Qualität des deutschen Bildungswesens weiterzuentwickeln und zu sichern. Bemerkenswert dabei ist, dass darüber zwischen den 16 Bundesländern ein breiter Konsens herrscht, trotz der Kulturhoheit der Länder im Bildungsbereich. Bereits kurz nach der Veröffentlichung der PISA-Ergebnisse im Jahr 2001 verständigte sich die KMK auf sieben zentrale Handlungsfelder, in denen die Länder und die Kultusministerkonferenz vorrangig tätig werden sollen. Das fünfte Handlungsfeld beschreibt die Forderung nach „Maßnahmen zur konsequenten Weiterentwicklung und Sicherung der Qualität von Unterricht und Schule auf der Grundlage von verbindlichen Standards sowie einer ergebnisorientierten Evaluation“ (KMK 2001). Aus diesem Handlungskatalog wurden bis zum heutigen Zeitpunkt zahlreiche und vielfältige Maßnahmen beschlossen, angefangen von Initiativen einzelner Schulen in Eigenregie, über Modellversuche und Wettbewerbe bis hin zu landesweit und länderübergreifend vereinbarten Vorgaben (vgl. Ditton 2008, S. 38).

Welche konkreten Programme und Maßnahmen zur Qualitätsentwicklung und –sicherung in Deutschland bisweilen länderübergreifend bzw. länderintern beschlossen und umgesetzt wurden, wird in den folgenden Kapiteln aufgegriffen. Die Beschreibung der Maßnahmen erfolgt adressatenorientiert in der Systematik des Mehrebenensystems des Bildungssystems. Zunächst wird der Evaluationsbegriff erläutert (2.4.2). Danach werden die Programme beschrieben, die auf Systemebene, also der Makroebene des Bildungssystems ansetzen: die Einrichtung von Qualitätsagenturen (2.4.3), die Bildungsberichterstattung (2.4.4) und die Teilnahme an internationalen und nationalen Schulleistungsstudien (2.4.5).

Im Anschluss daran wird das Konzept der Bildungsstandards in den Blick genommen (2.4.6). Diese beziehen sich auf die Meso- und Mikroebene des Bildungssystems, demnach auf die Ebenen der Einzelschule und des Unterrichts. Da Lernstandserhebungen in Form von Vergleichsarbeiten im Zentrum der Arbeit stehen, erfolgt deren ausführliche Beschreibung in einem eigenständigen Kapitel (Kapitel 3).

2.4.2 Evaluation als Verfahren der Qualitätsüberprüfung

Der Begriff Evaluation stammt aus dem lateinischen und bezeichnet die Bewertung, die Bestimmung des Wertes (vgl. Wermke et al. 2007, S. 295). Nach Hanft. (2008, S. 287) stellt das Bewerten im Sinne der Einschätzung oder Klassifizierung eines Gegenstandes oder einer Leistung den grundlegenden Bestandteil jeder Evaluation dar. Evaluationen spielen in den europäischen Bildungssystemen eine zunehmende Rolle (vgl. Brüsemeister und Eubel 2008b).

Es lässt sich ein sehr weit gefasster Evaluationsbegriff von einem engeren unterscheiden. In den weit gefassten Begriff fließen alle Formen der Bewertung mit ein. Bauer (2004) schlägt deshalb vor, „den Begriff der Evaluation einzugrenzen und nur dann von Evaluation im engeren Sinn zu sprechen, wenn es sich um eine empirisch wissenschaftlich gestützte Form der Erfolgskontrolle (Wirksamkeitsüberprüfung) und Bewertung handelt. Evaluationen beruhen auf vorab festgelegten Kriterien, die aus Zielen abgeleiten werden. Für die Kriterien wiederum werden Indikatoren ausgewählt. Diese Indikatoren sind empirisch erfassbare, messbare oder auf andere Weise feststellbare Tatbestände, die auf die meist latenten Kriterien zurück schließen lassen. Für alle anderen Formen der Bewertung gibt es in der Alltagssprache genügend Begriffe, so dass der Begriff der Evaluation für eine Normierung im vorgeschlagenen Sinn zur Verfügung steht“ (ebd., S. 164).

Peek (2007) formuliert ebenfalls ähnliche Anforderungen an Evaluationsvorhaben. Zunächst erfolgt jede Evaluation zielgerichtet und systematisch, indem sie vorab festgelegten Fragestellungen nachgeht und zu deren Klärung beiträgt. Des Weiteren basieren Evaluationen auf Datengrundlagen, die in weiteren Analyseschritten zu Bewertungen führen.

Die Bewertungskriterien müssen vorab festgelegt werden und sich auf die Fragestellung beziehen. Evaluation kann einerseits zur Überprüfung der Praxis genutzt werden als auch zu deren Veränderung beitragen (vgl. ebd., S. 143).

Ditton (2000) knüpft an den letzten Punkt an und konkretisiert die Zielsetzung und Wirkungsweise von Evaluation wie folgt:

> „Das Ergebnis einer Evaluation sollten praktisch verwertbare Einzelinformationen zu effektiven und realisierbaren Verfahren der Verbesserung des Unterrichts sein und außerdem Anregungen dazu, wie durch Maßnahmen auf der schulischen Ebene und im schulischen Kontext ein effektiver Unterricht besser gewährleistet und abgesichert werden kann“ (ebd., S. 75).

Evaluation und Reflexion können in Schulentwicklungsprozessen einen wichtigen Beitrag leisten (vgl. Altrichter 1998; Böttcher et al. 2006). Bei Evaluationsverfahren lassen sich verschiedene Formen unterscheiden, die im Folgenden näher erläutert werden:

- Interne Evaluation, externe Evaluation und Selbstevaluation
- Formative und summative Evaluation
- Einmalige, mehrfache und kontinuierliche Evaluation
- Interindividuelle, intraindividuelle und kriteriale Bezugsnorm der Evaluation
- Metaevaluation

Interne, externe und Selbstevaluation

Das entscheidende Kriterium für interne Evaluation und Selbstevaluation ist, dass die zentrale Verantwortung für die Gestaltung und Durchführung bei der einzelnen Schule liegt und auch von Personen durchgeführt wird, die in der Schule arbeiten. Der Unterschied zwischen Selbstevaluation und interner Evaluation ist in der Person zu sehen, die die Evaluation durchführt. Evaluiert sich eine Person selbst, dann spricht man von Selbstevaluation. Ist die Person die evaluiert Mitglied derselben Institution, handelt es sich um interne Evaluationsverfahren.

Demgegenüber bedeutet externe Evaluation oder Fremdevaluation, dass die zentrale Verantwortung außerhalb liegt und sie von Personen durchgeführt wird, die nicht in der Schule arbeiten (vgl. Peek 2007, S. 143).

Sowohl externe als auch interne Evaluationsverfahren kommen im Bildungswesen mittlerweile häufig zur Anwendung. Die aktuellen Entwicklungen in den einzelnen Bundesländern lassen eine ‚Doppelstrategie' der Qualitätsentwicklung und Qualitätssicherung erkennen: Es werden schulinterne Maßnahmen in Form von Schulprogrammen und schulinterne Evaluationen durchgeführt, die durch schulexterne Maßnahmen in Form von standardisierten Lernstandserhebungen, zentralen Abschlussprüfungen und Schulinspektionen bzw. Qualitätsanalysen begleitet werden. Damit wird zugleich das Beziehungsgefüge zwischen dezentraler Steuerung (Schulebene) und zentraler Steuerung (Systemebene) neu justiert. Bei dieser Neubestimmung erhält die Systemebene wieder einen konstitutiven Stellenwert im Geflecht der Handlungsebenen (vgl. Steffens 2007, S. 43).

Artelt (2007) versteht unter Kriterien für eine gelingende und anschlussfähige externe Evaluation jene Faktoren, „die das Verhältnis und die Zusammenarbeit zwischen externer und interner Evaluation erleichtern" (ebd., S. 136). So können einzelschulische Evaluationsprozesse mit externen Evaluationsmaßnahmen verknüpft werden. Als Voraussetzung für eine gelingende Zusammenarbeit listet sie folgende Bedingungen auf:

- **Transparenz der Bewertungen und der Evaluationsverfahren:** Die Bewertungen müssen allen Beteiligten ebenso bekannt sein wie die angewandten Evaluationsverfahren.
- **Verbindliche und geteilte Maßstäbe:** Es besteht Einigkeit bezüglich der Bewertungsmaßstäbe, die ggf. auch durch Metaevaluationen nachgewiesen werden.
- **Kontextsensitivität**: Die externe Evaluation muss die individuellen Rahmenbedingungen der einzelnen Schule berücksichtigen.
- **Ergebnisoffenheit der Evaluation**: Erst am Ende einer Evaluation wird entschieden, wer zur Sicherung der Qualität welche Verantwortung übernehmen muss und welche Unterstützung erforderlich ist.

- **Klarheit und Verbindlichkeit:** Der Evaluationsbericht, in dem die evaluierten Bereiche präzise und nachvollziehbar bewertet werden, ist so abgefasst, dass sich aus ihm klar formulierte und verbindliche Vereinbarungen zur Qualitätsverbesserung ableiten lassen, die Konsequenzen (z.B. durch follow-up Arrangements) haben.
- **realistische Zielvereinbarungen:** Ziele sind Ziele, wenn sie erreicht werden können, d.h. realistische Erwartungshorizonte darstellen. Zielvereinbarungen als Folge von Evaluationen sollten anspruchsvoll, jedoch für die Schule auch realisierbar sein. (Vgl. ebd., S. 136).

Formative versus summative Evaluation: Formative Evaluationen beziehen sich auf Prozesse oder Phasen der Durchführung einer Maßnahme und dienen vor allem der Förderung und Optimierung solcher Prozesse. Die Wirkung der formativen Evaluation wird erhöht, wenn alle beteiligten Akteure, Lehrpersonen, Schüler und ggf. auch Eltern, in einem hohen Maß an allen Phasen der Durchführung beteiligt werden. Zahlreiche Metaanalysen bestätigen die Effektivität formativer Leistungsmessung und auf den Lehr-Lernprozess bezogene Rückmeldungen (vgl. Maier 2010). Summative Evaluationen beziehen sich auf die Ergebnisse und dienen vor allem der abschließenden Bilanzierung am Ende eines Bildungsabschnittes.

Einmalige, mehrfache, kontinuierliche Evaluationen: Die meisten Evaluationsverfahren sind singulär, einige sind jedoch auch zyklisch angelegt, so dass Veränderungen überprüft und bewertet werden können. Ein Beispiel für eine kontinuierliche Evaluation stellen die PISA-Erhebungen dar, die im Abstand von drei Jahren durchgeführt werden und immer dieselben Kompetenzbereiche überprüfen. Durch den Ergebnisvergleich der unterschiedlichen Erhebungswellen liegen mittlerweile bilanzierende Auswertungen zur Entwicklung des deutschen Schulsystems vor, die bis zu einer Dekade umfassen (vgl. Klieme et al. 2010).

Interindividuelle, intraindividuelle und kriteriale Bezugsnorm der Evaluation: Hier werden Verfahren nach ihrer Bezugsnorm unterschieden. Finden die Vergleiche zwischen verschiedenen Individuen (interindividuell) oder Vergleiche mit Ergebnissen der einzelnen Individuen zu verschiedenen Zeitpunkten (intraindividuell) statt. Oder liegen der Evaluation Kriterien zur Bewertung zugrunde.

Metaevaluation: Darunter versteht man die methodenkritische Überprüfung der Wirksamkeit von durchgeführten Evaluationsverfahren durch externe Evaluatoren. Denn „Selbsteinschätzungen sind nahezu zwangsläufig einseitig und möglicherweise zu wenig selbstkritisch. (…) Wenn die Selbstbeobachtungen nicht durch einen fremden Blick ergänzt werden, dann wird die eigene Wahrnehmung zu selektiv“ (Kempfert & Rolff 1999, S. 137).

In folgender Tabelle werden nochmals alle Evaluationsformen und -typen, die im Bildungswesen angewendet werden, aufgeführt:

Tabelle 2.4.1: Typisierung von Evaluationsformen nach van Ackeren & Klemm (2011, S. 160)

Wozu?	Formative Evaluation:	Evaluation während Lernphasen zur Förderung
	Summative Evaluation:	Evaluation nach Lernphasen zur abschließenden Bewertung
Wann?	Prozessbezogene Evaluation Produktbezogene Evaluation	
Durch wen?	Selbstevaluation als interne Evaluation Metaevaluation Fremdevaluation als externe Evaluation	
Wie oft?	Einmalige Evaluation Mehrfache Evaluation Kontinuierliche Evaluation	
Welcher Maßstab?	Individuelle Norm: Vergleich mit früheren Erhebungen Soziale Norm: Vergleich mit anderen Evaluierten Kriteriale Norm: Vergleich mit einem vorher festgelegten Standard	
Was?	Programm-Evaluation Unterrichtsevaluation Lernstandsevaluation	

2.4.3 Einrichtung von Qualitätsagenturen

Um eine systematische und regelmäßige Überprüfung der Bildungsstandards mit Hilfe von Testverfahren sicherzustellen, wurde in Berlin im Dezember 2003 die Einrichtung des Instituts zur Qualitätsentwicklung im Bildungswesen (IQB) als An-Institut der Humboldt-Universität zu Berlin eingerichtet (vgl. KMK 2004; KMK 2008). Das IQB arbeitet in enger Abstimmung mit den Ländern sowie mit nationalen und internationalen Forschungseinrichtungen, Verbänden und Institutionen. Die Arbeitsfelder und -bereiche des IQB umfassen u. a. die Generierung von Aufgabensammlungen zur Operationalisierung der Bildungsstandards, die Formulierung von Kompetenzmodellen, in denen für die einzelnen Fächer beschrieben wird, welche Kompetenzen die Schüler zu einem definierten Zeitpunkt erreicht haben sollen. Außerdem werden geeignete Vergleichsaufgaben formuliert und Tests entwickelt, die den Kompetenzerwerb und die Kompetenzentwicklung der Schülerinnen und Schüler erfassen. Des Weiteren befasst sich das IQB mit Fragen der Evaluation und Implementation von Maßnahmen und deren Wirksamkeit. Es generiert Steuerungswissen, indem es Indikatoren zur Bildungsqualität analysiert. Nicht zuletzt erfolgt am IQB die Einrichtung eines Forschungsdatenzentrums, indem computergestützte Test-, Auswertungs- und Rückmeldesysteme erarbeitet werden (vgl. Ditton 2008, S. 39).

Landesweit wurden in den vergangenen Jahren ebenfalls Qualitätsagenturen gegründet, um die vielfältigen Qualitätsinitiativen auf Länderebene zu koordinieren. Diese Qualitätsagenturen „sind nicht Teil der ‚klassischen' Schulaufsicht, sondern von dieser institutionell und personell unabhängig" (van Ackeren und Klemm 2011, S. 169). Ihre Aufgabe besteht darin, auf Landesebene systematisch zur Verbesserung der schulischen Bildung beizutragen. Als Arbeitsgrundlage dient ihnen ein mehr oder weniger differenzierter und konkretisierter Qualitätsrahmen, der sich an der Systematik des „Input-Prozess-Output bzw. Outcome - Modells" orientiert.

Die Agenturen beschäftigen sich u. a. mit folgenden Themenbereichen:

- Entwicklung und Überprüfung von Qualitätsstandards für Schulen,
- Konzeption, Durchführung und Weiterentwicklung von Vergleichsarbeiten,
- Gewinnung und Auswertung von Befunden der Schul- und Unterrichtsforschung, auch im Hinblick auf Handlungskonsequenzen,
- Berichterstattung zu Entwicklungsständen im Schulwesen,
- Konzeption von Instrumenten und Verfahren der Qualitätssicherung,
- Durchführung von Modellprojekten zur Weiterentwicklung des Schulwesens und an der
- Durchführung der externen Evaluation (‚Schulinspektion') in Zusammenarbeit mit den Staatlichen Schulämtern.

Bisher ungeklärt ist, ob sich die Qualitätsagenturen der Bundesländer im Ländervergleich ähneln und ob unter gleicher Begrifflichkeit der jeweiligen Einrichtung auch das Gleiche gemeint ist. Eine Übersicht über die länderspezifischen Institute zur Qualitätssicherung findet sich im Anhang bei Ditton (2008, S. 52).

2.4.4 Bildungsberichterstattung

Ein weiteres Element der Qualitätssicherung stellt die regelmäßige Bildungsberichterstattung dar, die seit 2003 auf nationaler Ebene durchgeführt wird und sich mittlerweile auch in den Ländern etabliert hat (vgl. Ditton 2008, S. 48). Der erste Bildungsbericht wurde von der KMK in Auftrag gegeben und konzentrierte sich ausschließlich auf den schulischen Bereich. Mittlerweile umfasst die Berichterstattung das gesamte Bildungswesen und erstellt Indikatoren zu zentralen Bereichen und Entwicklungen im Bildungssystem [z.B. Frühkindliche Betreuung, Schule und non-formale Lernwelten im Schulalter, Berufliche Ausbildung, Hochschulen etc.] (vgl. Konsortium Bildungsberichterstattung 2005, Döbert 2008). Dadurch liegen periodisch erfasste Daten sowohl über die einzelnen Bildungseinrichtungen als auch über das gesamte Bildungssystem vor. „Deutschland hat sich somit auf den Weg gemacht, in dessen Verlauf die Entwicklung des Bildungssystems auf einer breiten empirischen Basis evidenzbasiert (d.h. ausdrücklich auf der

Grundlage von empirisch nachgewiesener Wirksamkeit) gesteuert werden soll“ (van Ackeren & Klemm 2011, S. 171).

Die Bildungsberichterstattung stellt auch für die Entwicklung regionaler Bildungslandschaften ein wichtiges Instrument der Qualitätsentwicklung dar (vgl. Döbert 2008). Bundesweit beschäftigen sich seit dem Jahr 2008/09 zahlreiche Institutionen und Einrichtungen mit der Erstellung regionaler bzw. städtischer Bildungsberichterstattung. Im Land Baden-Württemberg liegen mittlerweile Bildungsberichte beispielsweise für die Städte Freiburg, Ravensburg und Schwäbisch Gmünd vor (vgl. Stadt Schwäbisch Gmünd 2010; Landratsamt Ostalbkreis 2010).

Eine Schwierigkeit der Bildungsberichterstattung betrifft die Generierung der Daten und deren Qualität.

> „Oftmals sind Daten, die von Interesse wären, entweder gar nicht verfügbar oder nicht in dem gewünschten Differenzierungsgrad (z.B. nach regionalen Einheiten oder Subgruppen). Außerdem liegen nicht alle nötigen Daten auf einer über die Zeit oder Länder hinweg einheitlichen Erhebungsbasis vor und können als ausreichend belastbar angesehen werden“ (Ditton 2008, S. 49).

2.4.5 Teilnahme an internationalen und nationalen Schulleistungsstudien

Ziel der internationalen Schulleistungsstudien (z.B. TIMSS, PISA und IGLU[5]) ist es, Informationen bereitzustellen, die es erlauben, die Leistungsfähigkeit von Bildungssystemen zu vergleichen und zu beurteilen. Dazu werden auf internationaler Ebene die Bereiche, Inhalte und Verfahren der Leistungsmessung festgelegt und alle teilnehmenden Nationen auf einer einheitlichen Datengrundlage beschrieben. Sie zielen damit primär auf die Beobachtung von Bildungssystemen als Ganzes (Systemmonitoring) und liefern bildungspolitisch relevantes Zustandswissen (vgl. Dedering et al. 2008; Peek 2002). Böttcher et al. (2008) erläutern den Begriff des „Systemmonitoring“ und dessen Funktionen dergestalt:

> „Systemmonitoring, allgemein gesprochen, meint die dauerhafte, datengestützte Beobachtung der Entwicklung der Gesellschaft insgesamt wie ihrer Teil- bzw. Subsysteme – und damit auch des Bildungswesens. Es ist also ein institutionalisierter Beobachtungs- und Analyseprozess auf Basis empirisch gesicherter Daten. Es hat im Wesentlichen drei Funktionen: die Beobachtung, Analyse und Darstellung wesentlicher Aspekte eines Systems, verbunden mit der Funktion der Systemkontrolle einschließlich der Angleichung von Leistungsmaßstäben (Benchmarks), sowie die Funktion, ‚Steuerungswissen’ zu generieren bzw. zu erweitern und ‚Steuerungshandeln’ begründbarer und zielgerichteter zu gestalten. Systemmonitoring enthält vor allem durch die Angleichung von Leistungsmaßstäben im Verhältnis zu anderen Staaten eine international vergleichende Komponente“ (ebd., S. 8).

Für PISA liegen mittlerweile die Befunde aus vier Erhebungen, und somit aus einem Jahrzehnt, vor (PISA 2000, 2003, 2006 und 2009). Dadurch wird es möglich, Trends und Bilanzen aus kontinuierlichen Schulleistungsstudien zu erstellen (vgl. Ehmke und Jude 2010; Klieme et al. 2010). Außer der Teilnahme an internationalen Schulleistungsstudien partizipieren die einzelnen Bundesländer, mit unterschiedlicher Beteiligung, auch an länderübergreifenden Vergleichsuntersuchungen, ebenfalls mit dem Ziel, Daten für ein System-Monitoring zu generieren. Im Rahmen nationaler Erweiterungen der internationalen Vergleichsstudien wird da-

[5] IGLU steht für **I**nternationale **G**rundschul-**L**ese-**U**ntersuchung

durch auch grundlagenorientierte, empirische Bildungsforschung geleistet (vgl. Baumert und Neubrand 2002). Zu länderübergreifenden Schulleistungsstudien zählen neben PISA-E u. a. die Ergänzungsstudie IGLU-E (**I**nternationale **G**rund-schul-**L**ese-**U**ntersuchung), DESI (**D**eutsch **E**nglisch **S**chülerleistung **I**nternatio-nal) oder BIJU (**Bi**ldungsverläufe und psychosoziale Entwicklung im **Ju**gendal-ter).

Im Bereich der länderinternen Erhebungen liegen mittlerweile zahlreiche Befunde aus einer Vielzahl von nationalen Schulleistungsstudien vor. In Hamburg wurde seit dem Jahr 1996 die Längsschnittstudie LAU (Aspekte der **L**ern**au**sgangslage) durchgeführt. In Rheinland-Pfalz gibt es die MARKUS-Studie (**Ma**thematische Gesamterhebung **R**heinland-Pfalz: **K**ompetenzen, **U**nterrichtsmerkmale, **S**chul-kontext). In Brandenburg fand die Studie QuaSUM (**Qua**litätsuntersuchung an **S**chulen zum **U**nterricht in **M**athematik) statt. Ein Überblick über die Beteiligung Deutschlands an internationalen, länderübergreifenden und –länderspezifischen Schulleistungsstudien findet man bei zahlreichen Autoren (vgl. Bos et al. 2010; Radisch 2008; van Ackeren 2002; van Ackeren 2003; van Ackeren und Klemm 2011).

2.4.6 Einführung von Bildungsstandards

Im Dezember 2001 beschlossen die Kultusminister der Länder die verbindliche Einführung von nationalen Bildungsstandards (vgl. KMK 2009). Diese legen fest, „über welche Kompetenzen Schülerinnen und Schüler am Ende eines bestimmten Bildungsabschnittes verfügen sollen“ (Klieme et al. 2003, S. 21). Sie formulieren folglich „verbindliche Anforderungen an das Lehren und Lernen in der Schule und stellen damit innerhalb der Gesamtheit der Anstrengungen zur Sicherung und Steigerung der Qualität schulischer Arbeit ein zentrales Gelenkstück dar. Bildungsstandards nennen präzise, verständlich und fokussiert die wesentlichen Ziele der pädagogischen Arbeit, ausgedrückt als erwünschte Lernergebnisse der Schülerinnen und Schüler“ (ebd., S. 9).

Die Entwicklung der Bildungsstandards erfolgte auf Bundesebene in den darauffolgenden Jahren sukzessive: Eine Arbeitsgruppe begann zuerst mit der Entwicklung nationaler Bildungsstandards für den Mittleren Schulabschluss in den Fächern Deutsch, Mathematik und Erster Fremdsprache, die im Jahr 2003 der Öffentlichkeit vorgestellt wurden. Mittlerweile liegen nationale Bildungsstandards für alle Schulformen vor (vgl. Böttcher & Dicke 2008, S. 103f; KMK 2010a; KMK 2010c).

Um die Länder bei der Entwicklung von länderspezifischen Bildungsstandards zu unterstützen, gab das Bundesministerium für Bildung und Forschung (BMBF) eine Expertise ‚Zur Entwicklung nationaler Bildungsstandards’ (vgl. Klieme et al. 2003) in Auftrag, die den Ländern als Orientierungsrahmen dienen soll (vgl. Ditton 2008, S. 39f).

Bildungsstandards stellen ein Instrument der Output-Orientierung dar, da sie Ziele formulieren, die von den Schülerinnen und Schülern erreicht werden sollen. Die Erwartungen an das neue Qualitätsinstrument sind hoch, wie im Zitat von Böttcher und Dicke (2008) deutlich wird:

> „Die an dieses Instrument geknüpften Hoffnungen sind vielfältig und tiefgreifend. Mittels Bildungsstandards soll Schule im Kontext von Outputsteuerung neu gesteuert werden.

> Ohne direkten Einfluss auf Lehrmethoden zu nehmen […] sollen sie den Unterricht so verändern, dass Schülerleistungen sich verbessern und Bildungsbenachteiligungen reduziert werden. Schließlich sollen sie helfen, […] Schülerleistungen vergleichbar zu machen. Grundsätzlich lässt sich feststellen: Bildungsstandards benennen die Kompetenzen, die Schule vermitteln muss. Insofern sind sie ergebnisorientiert“ (ebd., S. 104).

Allerdings führen die Autoren weiter an, dass Bildungsstandards als Instrument der Outputorientierung nur dann nützlich sind, wenn auf ihrer Grundlage standardbasierte Tests entwickelt werden. Denn Standards, die nicht ‚messbar‘ sind, wären logischerweise für eine Outputsteuerung unbrauchbar. Die Meinung, dass das Erreichen von Bildungsstandards mittels geeigneter spezifischer Testverfahren überprüft werden kann, teilen auch andere Autoren. Klieme (2005) sieht den Vorteil von Bildungsstandards für Schulen darin, dass sie Zielmarken vorgeben und somit den Output konkretisieren. Dadurch werden einerseits die erwarteten Schülerkompetenzen klar benannt und andererseits werden die erzielten Leistungen der Schülerinnen und Schüler überprüfbar. Dadurch, dass Bildungsstandards den Output beschreiben, sind die Lehrerinnen und Lehrer in der Gestaltung und Umsetzung ihrer Unterrichtspraxis relativ autonom. „In einem solchen Kontext müssen Standards keineswegs […] den ‚Eigensinn‘ von Lehrenden und Lernenden unterbinden. Sie können vielmehr zu einem der Motoren für die Steigerung der Bildungsqualität von Schulen werden“ (ebd., S. 7).

Köller (2007) gibt zu Bedenken, dass „Bildungsstandards als Steuerungsinstrumente im Bildungswesen nur wirksam [werden], wenn sie auf konkrete Aufgaben bzw. Testitems herunter gebrochen werden, die reliabel und valide die in den Standards formulierten Kompetenzen erfassen“ (ebd., S. 97). Hierzu liegen inzwischen einige Veröffentlichungen vor, die sich sowohl mit Bildungsstandards und Kompetenzmodellen (vgl. Gehrmann et al. 2010) als auch mit der Rezeption von Bildungsstandards (vgl. Pant et al. 2008; Tillmann 2007) beschäftigen. Erfahrungen anderer Länder zur ‚Standards-Based Reform’ (O'Day 2008) können ebenfalls bei der Implementierung herangezogen werden.

2.4.7 Überlegungen zu einer gelingenden Evaluation im Bildungssystem

In den vorangegangenen Kapiteln wurden die einzelnen Maßnahmen der „Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring“ vorgestellt und beschrieben. Diese Strategie ist Ausdruck für einen „Paradigmenwechsel in der Bildungspolitik in Deutschland im Sinne von Ergebnisorientierung, Rechenschaftslegung und Systemmonitoring“ (KMK 2006, S. 5). Die verschiedenen Maßnahmen sollen zur Verbesserung des Bildungswesens beitragen, indem „Prozesse zur Qualitätsentwicklung und Standardsicherung auf allen Ebenen, von der einzelnen Schule bis zum gesamten Bildungssystem“ (ebd., S. 6) systematisch umgesetzt und miteinander verbunden werden. Sie richten sich demnach an eine Vielzahl von Akteuren auf den verschiedensten Ebenen des Bildungssystems. Ob die intendierten Steuerungsvorstellungen der Programme auch tatsächlich ihre Wirkung entfalten, muss bei deren Implementierung ins System überprüft werden. An die Implementierung der spezifischen Maßnahmen schließt sich die Fragestellung an, ob die zuständigen Akteure die jeweiligen Zielsetzungen teilen, oder ob sie versuchen, die Entwicklungen gemäß ihrer spezifischen Absichten zu beeinflussen (vgl. Altrichter 2009, S. 242; Lehmann 2008).

Im Sinne der Educational Governance müssen insbesondere die Übergänge zwischen den einzelnen Ebenen des Bildungswesens in den Blick genommen werden, um deren Zusammenspiel zu beleuchten. Es würde den Rahmen dieser Arbeit sprengen, die Implementierung all dieser Instrumente detailliert zu beschreiben.

Im Zentrum der vorliegenden Arbeit steht eine weitere Maßnahme der Gesamtstrategie, die zur Qualitätssicherung im Bildungssystem beitragen soll; die Durchführung und Implementierung von Vergleichsarbeiten.

3 Vergleichsarbeiten als Instrument der Qualitätssicherung im deutschen Bildungssystem

Im vorhergehenden Kapitel wurden unterschiedliche Reformmaßnahmen vorgestellt, die alle einen Beitrag zur Qualitätssicherung und Qualitätsentwicklung des deutschen Bildungswesens leisten sollen. In Zusammenhang mit zentralen Bildungsstandards wurden auf der Ebene der Einzelschule und des Unterrichts Vergleichsarbeiten eingeführt, die der Überprüfung der standardbasierten Kompetenzen dienen sollen. Dieses Kapitel beschäftigt sich ausführlich mit der Implementierung von Vergleichsarbeiten als Instrument der Qualitätssicherung.

3.1 Zur Entstehungsgeschichte von Vergleichsarbeiten

Die Kultusministerkonferenz (KMK) hat nach der Veröffentlichung der PISA-Ergebnisse zur Weiterentwicklung der Qualität von Unterricht und Schule beschlossen, Standards zu erarbeiten. Damit sollen länderintern und länderübergreifend die Kompetenzen von Schülerinnen und Schülern zu einem bestimmten Zeitpunkt ihrer Bildungskarriere überprüft werden. Somit beziehen sich Lernstandserhebungen auf langfristige und nachhaltige Prozesse der Lern- bzw. Kompetenzentwicklungen (erreichte Kenntnisse, Fertigkeiten und Fähigkeiten) und nicht wie Klassenarbeiten auf das unmittelbar vorangegangene Unterrichtsgeschehen. Damit steht - wie es in der Expertise ‚Zur Entwicklung nationaler Bildungsstandards' (vgl. Klieme u.a. 2003, S 17ff) gefordert wird -, die Überprüfung kumulativen, systematisch vernetzten Lernens im Mittelpunkt (vgl. Bonsen et al. 2004a, S. 125).

Auf der 298. und 299. Sitzung der KMK in Eisenach und Würzburg im Jahre 2002 einigten sich die Kultusminister darauf, dass landesweite Vergleichs- bzw. Orientierungsarbeiten zur Überprüfung dieser Standards eingerichtet werden sollen.

> „Um die Einhaltung dieser Standards zu überprüfen, sollen in den Ländern landesweit Orientierungs- und Vergleichsarbeiten geschrieben werden. Ziel einer solchen Überprüfung muss es nach Ansicht der KMK sein, dass möglichst viele Schülerinnen und Schüler durch gezielte Förderung die gesetzten Ziele erreichen. Die Vergleiche sollen in der Primarstufe beginnen und auch in den weiterführenden Schulen ab Jahrgangsstufe 5 bzw. 7 durchgeführt werden. Auf der Grundlage der noch zu formulierenden Standards sollen für alle Fächer Aufgabenpools entwickelt werden, die den Ländern und ihren Schulen dann zur Verfügung stehen“ (KMK 2002a).

> „Darüber hinaus werden die Länder in landesweiten bzw. länderübergreifenden Orientierungs- oder Vergleichsarbeiten überprüfen, in welchem Umfang die vereinbarten Standards tatsächlich erreicht werden. Ziel dieses Verfahrens soll es sein, eine Qualitätssicherung zu gewährleisten, sich darüber länderübergreifend auszutauschen und es den Schülerinnen und Schülern in allen Ländern der Bundesrepublik Deutschland zu ermöglichen, in allen Bildungsgängen über individuelle Förderung die gesetzten Ziele zu erreichen“ (KMK 2002b).

Demnach werden Vergleichsarbeiten zur landesweiten Evaluation an Schulen in unterschiedlichen Jahrgangsstufen verwendet. Sie dienen insbesondere der flächendeckenden, jahrgangsbasierten Evaluation der einzelnen Schule und Klasse als Standortbestimmung vor dem Hintergrund der länderübergreifenden Bildungsstandards. Sie tragen somit dazu bei, als landesweite Projekte der Standardüberprüfung empirisch fundierte Beiträge zum Systemmonitoring, zur Unterrichtsentwicklung und zur Individualbewertung zu leisten. Damit ermöglichen Lernstandserhebungen den Schulen eine orientierende Standardüberprüfung in vergleichender Perspektive (vgl. Bonsen et al. 2004a; Kühle & Peek 2007, S. 429; Peek 2007). Außerdem kommt in den Ausführungen der KMK deutlich zum Ausdruck, dass möglichst viele Schülerinnen und Schüler die geforderten Bildungsstandards erreichen sollen. Dieses Ziel soll durch gezielte, individuelle Förderung sowohl in der Primarstufe, als auch in weiterführenden Schulen angestrebt und erreicht werden.

Seit einigen Jahren werden in allen Ländern der Bundesrepublik sowohl länderspezifische als auch länderübergreifende Tests durchgeführt. Allerdings existiert in den Bundesländern bisher keine einheitliche Begrifflichkeit für standardorien-

tierte Überprüfungsverfahren. So trifft man auf zahlreiche Namen wie die folgende Auswahl zeigt: „Lernstandserhebung“, „Orientierungsarbeit“, „Parallelarbeit“, „Diagnosearbeit“, „Jahrgangsstufentest“, „Kompetenztest“ oder „Vergleichsarbeit“. (Vgl. Bonsen et al. 2004a, S. 125; Hovestadt und Keßler 2005). Problematisch dabei ist, dass unter den verschiedenen Bezeichnungen oftmals ähnliche Verfahren verstanden werden. Sind ihnen jedoch unterschiedliche Ziele und Funktionen inhärent, so sind sie „oft nur noch für Insider verständlich. Ihre Vielfalt erschwert der breiteren Öffentlichkeit das Verständnis“ (Avenarius et al. 2003, S. 268).

Auch die Erfahrungen der einzelnen Bundesländer mit Vergleichsarbeiten sind sehr different. Die Bundesländer haben hier eine unterschiedlich lange Tradition mit teilweise verschiedenartiger Gewichtung der Funktionen und Verfahrensweisen (vgl. Bonsen et al. 2004a, S. 125; Herrmann 2007; Netzwerk Empiriegestützte Schulentwicklung (EMSE) 2006). Auf die Darstellung der heterogenen Konzeptionen der länderübergreifenden und länderspezifischen Verfahren in 16 Bundesländern soll in dieser Arbeit nicht eingegangen werden. Es folgt zunächst eine allgemeine Beschreibung über die Konzeption, Zielsetzung und Funktionen von Vergleichsarbeiten (3.1.1), sowie eine kritische Auseinandersetzung mit diesem Evaluationsinstrument, das zur Qualitätssicherung und zur Qualitätsentwicklung beitragen soll (3.1.2).

3.1.1 Konzeption, Zielsetzung und Funktionen von Vergleichsarbeiten

3.1.1.1 Konzeption

Die Grundlage der testbasierten Schulreform in Deutschland bilden folglich Bildungsstandards, deren Erreichung mit Hilfe von zentralen Leistungsüberprüfungen in Form von Lernstandserhebungen bzw. Vergleichsarbeiten in ausgewählten Jahrgangsstufen gemessen werden. Die Aufgaben orientieren sich an den curricularen Vorgaben. Da Vergleichsarbeiten auch Rückmeldungen über einzelne Schülerinnen und Schüler enthalten, müssen alle Schülerinnen und Schüler einer

Schulform bzw. eines Kurstyps landesweit die gleichen Aufgaben bearbeiten (vgl. Bonsen et al. 2004a, S. 130).

Die Lehrerinnen und Lehrer erhalten auf dieser Basis zu bestimmten Zeitpunkten des Bildungsgangs systematisch Rückmeldungen, die ihnen in der Form bisher nicht zugänglich waren und die sie ohne dieses Instrument nur schwer einschätzen konnten. Hierbei handelt es sich um Informationen über den Leistungsstand und das Leistungsprofil ihrer Schülerinnen und Schüler im Vergleich zu den in den Bildungsstandards enthaltenen Leistungserwartungen und im Vergleich zu den in anderen Klassen und Schulen erreichten Lernergebnissen. Diese Standortbestimmung ist nicht nur für ein Bildungssystem, sondern auch für die einzelne Schule als Schulevaluation bedeutsam (vgl. Orth 2005, S. 64; Ditton 2008, S. 42). Die Kolleginnen und Kollegen erhalten somit Hinweise, inwieweit Optimierungsbedarf besteht bzw. „ob man im Soll ist“ (Köller 2007, S. 100).

Die Vergleichsmöglichkeiten hierbei sind vielfältig. Die Lehrpersonen erhalten Rückmeldungen über die Leistungen ihrer Schülerinnen und Schüler, die den Vergleich einer einzelnen Klasse mit einheitlichen bundesweiten Standards ermöglichen. Die Rückmeldungen erlauben auch den Vergleich mit Parallelklassen und/oder Klassen mit ähnlicher Schülerschaft und ähnlichem Einzugsgebiet. Außerdem lassen sie die Verteilung der eigenen Schülerinnen und Schüler auf unterschiedliche Kompetenzniveaus zu, ermöglichen die Identifikation von Stärken und Schwächen der Klasse bzw. der Schülerinnen und Schüler und schließlich erlauben sie eine Trendanalyse, da sich die Vergleiche im zeitlichen Verlauf abbilden lassen. Somit erhalten Schulen also die Möglichkeit, Optimierungsprozesse in Ergebnissen sichtbar zu machen (vgl. ebd., S. 100 ff).

Die Basis für die Vergleiche bildet die Auswertung einer repräsentativ angelegten Zentralstichprobe. Dadurch wird der Bezug zu einer lerngruppen- bzw. schulübergreifenden Norm möglich. Bei Vergleichsarbeiten lassen sich drei Vergleichsmaßstäbe unterscheiden: Die **kriteriale Bezugsnorm** bezieht sich auf den erreichten Lernstand der Schülerinnen und Schüler in Bezug auf fachliche Ansprüche; die **soziale Bezugsnorm** orientiert sich am Vergleich mit dem durchschnittlichen Niveau der Klasse (arithmetisches Mittel der Leistungen aller Schülerinnen und Schüler) und die **individuelle Bezugsnorm** fokussiert die Lernentwicklung eines

einzelnen Schülers oder einer einzelnen Schülerin über eine gewisse Zeitspanne hinweg. (Vgl. Peek & Dobbelstein 2006a, S. 181ff; Steffens 2008, S. 3).

Neben dem Vergleich mit den Ergebnissen im Landesdurchschnitt wird den Schulen ein weiterer Referenzwert im Sinne eines ‚fairen' Vergleichs gegeben: Mit der Kontextualisierung von Schulen durch die Zuordnung zu so genannten Standorttypen erhalten die Einzelschulen zusätzlich eine Vergleichsmöglichkeit mit Schulen, die unter ähnlichen Bedingungen und mit vergleichbaren Ausgangslagen arbeiten (vgl. Bonsen et al. 2004a, S. 126; Kohler & Schrader 2004, S. 8).

> „Um Schulen eine ‚faire' Standortbestimmung und eine realistische Einschätzung ihrer Ergebnisse in den Lernstandserhebungen zu ermöglichen, reicht es also nicht aus, als Referenzwerte repräsentative Ergebnisse der unterschiedlichen Schulformen zur Verfügung zu stellen. Schulen mit ungünstigen Lernvoraussetzungen ihrer Schülerschaft ständen damit in Gefahr, die Qualität ihrer Förderbemühungen tendenziell zu unterschätzen, Schule mit günstiger Zusammensetzung dagegen zu überschätzen" (Peek & Dobbelstein 2006a, S. 181ff).

Nachtigall & Kröhne (2006) stellen Strategien aus bisherigen Schulleistungsstudien vor, die zeigen, wie ‚faire' Vergleichswerte ermittelt werden können. Als Grundlage des Vergleichs wird immer der Klassen- bzw. Schulmittelwert mit dem Mittelwert einer Population (Bsp. Landesmittelwert) herangezogen. Der alleinigen Heranziehung dieses Mittelwertes im Hinblick auf geplante Schul- und Unterrichtsentwicklungen stehen sie sehr kritisch gegenüber. Denn Schulen, die unter sehr schwierigen Bedingungen hervorragend arbeiten, werden schlechter beurteilt als Schulen in Akademikerwohnbezirken, die möglicherweise trotz schlechter schulischer Arbeit deutlich überdurchschnittlich abschneiden können. Daher schlägt er Vorgehensweisen vor, die aus der Erfahrung mit bisherigen Schulleistungsstudien abgeleitet sind. So werden z.B. bei der IGLU-Studie sowohl die Kontextbedingungen der Schule als auch der Schülerhintergrund berücksichtigt. In Ergänzung zu den Populationsmittelwerten wurden den Schulen auch die Mittelwerte von vier Vergleichsklassen als Referenzwert zurückgemeldet. In den PISA-Erhebungen wird mit Hilfe eines linearen Regressionsmodells ein theoretischer Vergleichswert ermittelt, wie er für die spezifisch vorliegenden Kontextbedingungen der jeweiligen Schule zu erwarten ist.

Die Auswertung der Vergleichsarbeiten wird in den Bundesländern unterschiedlich gehandhabt; einerseits kann sie zentral erfolgen, andererseits wird sie von den Lehrpersonen vor Ort mit Hilfe von Auswertungsanleitungen durchgeführt.

3.1.1.2 Zielsetzung und Funktionen von Vergleichsarbeiten

Primäre Zielsetzung von Vergleichsarbeiten ist der Wunsch, Schul- und Unterrichtsentwicklung anzustoßen (vgl. Ditton 2008, S. 42; Fleischer et al. 2008, S. 195ff).

> „Lernstandserhebungen zielen primär auf die Unterrichtsentwicklung in den Schulen bzw. in den Klassen. Durch die Bereitstellung von zusätzlichen Informationen über Stärken und Schwächen und durch die Unterstützung der Lehrerinnen und Lehrer bei der Beurteilung und Auswertung der Ergebnisse sind sie als ein zusätzliches Instrument zur Reflexion und zur Verbesserung des Unterrichts konzipiert“ (Peek & Dobbelstein 2006a, S. 183).

Die wesentlichen Aussagen in der aktuellen Literatur zum Thema Ziele und Funktionen von Vergleichsarbeiten finden sich in den Veröffentlichungen des länderübergreifend arbeitenden Netzwerks „Empiriegestützte Schulentwicklung - EMSE“. Dieses Netzwerk wurde im Jahre 2004 gegründet (vgl. Institut für Qualitätsentwicklung im Bildungswesen (IQB) 2010a). Es setzt sich aus Wissenschaftlerinnen und Wissenschaftlern, sowie Mitarbeiterinnen und Mitarbeitern der Kultus- bzw. Schulministerien, der Landesinstitute oder Qualitätsagenturen aller 16 Bundesländer zusammen. Die Mitglieder befassen sich mit empirischen Verfahren der Schulentwicklung und Bildungsplanung. Im Rahmen ihrer fünften Tagung in Berlin im Dezember 2006 erstellten sie zum Thema „zentrale standardisierte Lernstandserhebungen“ ein Positionspapier zur Funktionsweise und Verfahrensweisen im Umgang mit Vergleichsarbeiten. Das Positionspapier enthält die Ergebnisse aus Diskussionen um die Implementierung der bundesweiten Maßnahmen und die praktischen Erfahrungen aus bundeslandspezifischen Projekten. Die Funktionen von Vergleichsarbeiten lassen sich einerseits auf die Ebene der Schu-

len, andererseits auf Systemebene verorten (vgl. Peek & Dobbelstein 2006b; Netzwerk Empiriegestützte Schulentwicklung (EMSE) 2006).

Auf den Ebenen der Einzelschule und des Unterrichts können Vergleichsarbeiten Lehrenden zusätzliche Informationen über die erreichten Lernstände ihrer Schülerinnen und Schüler bereitstellen. Sie erhalten dadurch eine Rückmeldung über die Wirksamkeit ihrer schulischen Arbeit. Darüber hinaus erhalten Lehrpersonen einen Hinweis, wo ihre Schüler im Hinblick auf die zentral festgelegten, an der kriterialen Bezugsnorm orientierten Bildungsstandards und im Vergleich zu anderen Bezugsgruppen stehen. Dies kann den Lehrerinnen und Lehrern zur Verortung und Rechenschaftslegung ihrer Arbeit dienen.

Weiterhin werden die Stärken und Schwächen jedes Einzelnen in der Klasse sichtbar. Aus dieser Rückmeldung können die Lehrpersonen spezifische und individuelle Fördermaßnahmen ableiten. Ebenso können diese Informationen als Grundlage für Beratungsgespräche mit Schülerinnen, Schülern und Eltern genutzt werden.

Eine weitere Chance zur Nutzung von Vergleichsarbeiten stellt die Möglichkeit dar, dass Lehrerinnen und Lehrer ihren Unterricht weiterentwickeln können, indem sie die zurückgemeldeten, empirisch ermittelten Ergebnisse aus Vergleichsarbeiten als Ausgangsbasis ihrer Reflexionen nutzen. Auch auf der Ebene der Einzelschule werden Möglichkeiten gesehen, die Qualität der Zusammenarbeit und des Austausches zu erhöhen. Je nachdem, wie die Ergebnisse der gesamten Schülerschaft einer Schule ausfallen, können z.B. in das Schulcurriculum geeignete Schwerpunkte aufgenommen werden, die den Schülerinnen und Schülern gerecht werden. Zudem kann ein systematischer Austausch über die Ergebnisse in Fachgruppen und Fachkonferenzen erfolgen und diese neu beleben. Hieraus könnten z.B. gemeinsam Testaufgaben für schulinterne Prüfungs- und Klassenarbeiten generiert werden.

Neben dieser Vielzahl an Chancen und Möglichkeiten sieht das Positionspapier weitere Verwendungsmöglichkeiten der Vergleichsarbeiten vor, die auf der Systemebene angesiedelt sind. Durch die Rückmeldung der Ergebnisse an diese Ebene stehen detaillierte Informationen über die Leistungen des Bildungssystems im

Hinblick auf bestimmte Lernbereiche, Schulformen, Schulstufen etc. zur Verfügung.

An die Rückmeldung von Vergleichsarbeiten sind auch die Erwartungen geknüpft, die Erkenntnisse zur Generierung von Planungs- und Steuerungswissen zu verwenden. Außerdem könnte man die Daten dazu verwenden, ein ausdifferenziertes Unterstützungssystem aufzubauen, das je nach Bedarf spezielle Ausbildungs- und Fortbildungsprogramme anbietet.

Vergleichsarbeiten werden nicht nur positiv wahrgenommen. In der Literatur finden sich einige kritische Beiträge, die sich neben der Verwendung eines eingeschränkten Qualitätsbegriffs und einer mangelnden Testgüte auch auf das Ranking und das Problem des ‚teaching to the test' beziehen. Die einzelnen Kritikpunkte werden im folgenden Kapitel aufgeführt.

3.1.2 Kritik an Vergleichsarbeiten

3.1.2.1 Kritikpunkt: Überfrachtung der Zielsetzungen und Funktionen von Vergleichsarbeiten

Die aufgezeigte Vielfalt an Zielsetzungen und Funktionen von Vergleichsarbeiten, die auf mehreren Ebenen des Bildungssystems ihre Wirkungen entfalten sollen, machen eine Funktionsüberfrachtung deutlich. Es fällt auf, dass dem Instrument zwei sich grundsätzlich unterscheidende Funktionen inhärent sind; einerseits das Messen, Prüfen und Sichern von Kompetenzen im Sinne einer summativen Evaluation – sowie andererseits das Entwickeln, Fördern und Unterstützen beim Erreichen der Kompetenzen als formative Evaluationsform. Diese konträren Positionen werden in dem EMSE-Positionspapier „als zwei Seiten ein und derselben Medaille betrachtet; beide Seiten sind aufeinander zu beziehen und bilden einen unauflöslichen Zusammenhang" (Peek et al. 2006, S. 1).

Je nach Anlage und Konzeption von Lernstandserhebungen wird die jeweilige Position betont. Werden Vergleichsarbeiten z.B. am Ende eines Bildungsabschnittes bilanzierend durchgeführt und fließen die Ergebnisse in die Notengebung ein, so

liegt der Schwerpunkt auf einer summativen Evaluationsform und dient der jeweiligen Lehrperson als Rechenschaftslegung und Verortung.

Bei formativen Evaluationsverfahren stehen die individuellen Stärken und Schwächen der Schülerinnen und Schüler im Vordergrund. Die Ergebnisse aus Vergleichsarbeiten sollen der Feststellung von Lern- und Förderbedarf dienen oder zur Stärkung der Diagnosekompetenz von Lehrpersonen beitragen.

Da sich die beiden Evaluationsformen in ihrer Zielsetzung deutlich voneinander unterscheiden, kann es bei gleichzeitiger Verwendung zu Zielkonflikten führen (vgl. Maier 2010, S. 48).

Die bisherige Konzeption der Vergleichsarbeiten sieht vor, dass lediglich zu bestimmten Zeitpunkten (mindestens im 2-Jahres-Abstand), der Lernstand gemessen wird; also somit eine summative Evaluationsform vorliegt. Auch Maier kommt zu dem Schluss, dass die Vergleichsarbeiten, wie sie in die KMK- Gesamtstrategie eingebettet sind, ihren summativen Charakter nicht leugnen können (vgl. ebd., S. 48ff). Welche der beiden Evaluationsformen sich für die pädagogische Alltagspraxis von Lehrerinnen und Lehrern besser eignet bewerten Bonsen & Gathen (2004) wie folgt:

> „Für die pädagogische Alltagspraxis von LehrerInnen sind summative Leistungsmessungen (…) wenig konstruktiv. Zur tatsächlichen Optimierung ihres Unterrichts in einem formativen Sinne benötigen Lehrpersonen vielmehr ein regelmäßiges Feedback in geringeren Abständen. Für die täglich oder wöchentlich notwendigen Entscheidungen zur Unterrichtgestaltung sind regelmäßige, detaillierte Informationen über die Lernprozesse der SchülerInnnen notwendig“ (ebd., S. 238f).

In wie weit die bestehenden Konzepte von Lernstandserhebungen formative Elemente enthalten, muss jeweils genau überprüft werden. Schymala liefert hierfür mit seiner Arbeit einen wichtigen Baustein zur Untersuchung der unterschiedlichen Konzeptionen von Lernstandserhebungen und deren Zielsetzungen. Er erstellt ein Analyseraster, das Merkmale formativer Leistungsmessung enthält. Im Einzelnen nennt er 16 Aspekte zentraler Leistungsmessung, die den formativen Charakter unterstützen. Dazu zählen u. a. die Ausrichtung am Lernprozess des Individuums und daraus abgeleitete Fördermaßnahmen mit darauf abgestimmten

Fördermaterialien. Des Weiteren betont er die Bedeutung der intrinsischen Motivation des Schülers und dessen aktive Einbeziehung in den Lernprozess (vgl. Schymala 2010, S. 47).

Ein weiterer Aspekt der Funktionsüberfrachtung betrifft die unterschiedlichen Arbeitsfelder von Lehrerinnen und Lehrern. So sollen Vergleichsarbeiten einerseits deren Diagnosekompetenz verbessern. Andererseits sollen sie die Weiterentwicklung sowohl des Unterrichts als auch der Einzelschule anregen. Zuletzt sind Vergleichsarbeiten auch als Ausgangsbasis für Kommunikation mit den unterschiedlichsten Akteuren des Bildungssystems vorgesehen. Die Ergebnisse sollen sowohl mit den Schülerinnen und Schülern intensiv besprochen werden, als auch in Beratungsgesprächen mit den Eltern als Grundlage herangezogen werden. Nicht zuletzt sollen die Ergebnisse mit Kolleginnen und Kollegen, in allen Gremien der Einzelschule und mit dem Schulleiter ausführlich besprochen werden, um sie als Impuls für Veränderungen nutzen zu können.

Hier stellt sich die Frage, welche Ziele tatsächlich durch Vergleichsarbeiten erreicht werden, welche Funktionen im Vordergrund der Nutzung stehen und welche Prioritäten Lehrerinnen und Lehrer bei der Durchführung von Vergleichsarbeiten setzen.

3.1.2.2 Kritikpunkt: eingeschränkter Qualitätsbegriff

Kritisiert wird auch, dass in den Vergleichsarbeiten „Schulqualität“ in erster Linie mit guten fachlichen Lernleistungen assoziiert wird. Dazu kommt, dass eine Prozessorientierung bei der Frage nach der Schulqualität so gut wie gar nicht berücksichtigt wird. Hieraus wird gefolgert, dass Ansätze einer ausschließlichen und restriktiven ‚Output‘-Steuerung zu kurz greifen. Sie sind zwar populär, in pädagogischer Hinsicht jedoch nicht tragfähig. So wird in Expertenkreisen postuliert, dass vor allem externe Tests, die zur Kontrolle schulischer Arbeit eingesetzt werden, innovationshemmend wirken. Wenn der Unterricht sich primär nach ihnen ausrichtet, können sich neue, weiterführende Unterrichtskonzepte im Sinne einer Methodenvielfalt kaum entfalten (vgl. Steffens 2007, S. 38f).

Als Abhilfe wird vorgeschlagen, Lernstandserhebungen so anzulegen, dass sie nicht nur verschiedene Kompetenzen der Schülerinnen und Schüler erfassen, sondern auch differente Lernkulturen und Lernstile berücksichtigen.

Als weiterer Punkt wird die Notwendigkeit intensiver Lehrerfortbildung betont, ohne die eine erfolgreiche Handhabung dieser neuartigen Verfahren der Leistungsbeurteilung nicht gelingen kann. Schließlich darf nicht übersehen werden: Auch die besten Verfahren der Leistungsbeurteilung können von sich aus keine Verbesserung von Lernresultaten bewirken. Sie können lediglich auf Probleme aufmerksam machen und Anstöße zu deren Bewältigung geben (ebd., S. 39f).

3.1.2.3 Kritikpunkt: Testgüte

Vergleichsarbeiten bzw. Lernstandserhebungen variieren hinsichtlich ihrer inhaltlichen und psychometrischen Güte erheblich. Die eingesetzten Tests orientieren sich bislang an den Standards bzw. daran angelehnten Lehrplänen der Länder. Die Items sollen hinsichtlich der jeweils getesteten Kompetenzen spezifiziert werden, wobei offen bleibt, wie dies genau geschehen soll. Das Kernproblem aller flächendeckenden Vergleichsarbeiten besteht bisher darin, dass sie keinen expliziten Bezug zu nationalen, standardisierten Bildungsstandards haben. Dies ist noch nicht möglich, da die entsprechenden Skalen und Kompetenzmodelle durch das IQB noch nicht bereitgestellt werden können. „Stattdessen behilft man sich mit rudimentären Kompetenzstufenmodellen […] oder mit sozialen Vergleichen, in denen die einzelne Schule/Klasse am Landesdurchschnitt gemessen wird“ (Köller 2008, S. 67).

Auch wird kritisiert, dass all diese Testuntersuchungen, nicht durch eigens ausgebildetes Personal administriert und ausgewertet werden, sondern vor Ort durch die Lehrpersonen der jeweiligen Schule. Anders wären flächendeckende Erhebungen gar nicht realisierbar. Diese Vorgehensweise entspricht der Funktion der Vergleichsarbeiten als einer Art formativer Evaluation; impliziert aber, dass wissenschaftlich seriöse Vergleiche zwischen Schulen oder gar Ländern auf dieser Basis nicht vorgenommen werden können (vgl. ebd.).

3.1.2.4 Kritikpunkt: Rankings als unfaire Vergleiche

Einen weiteren Punkt des Anstoßes stellt das Ranking dar. Denn der eigentliche Sinn und Nutzen von Vergleichsarbeiten wird darin gesehen, Hinweise auf Stärken und Schwächen der Schülerinnen und Schüler zu erhalten sowie zum Nachdenken über mögliche Verbesserungen angeregt zu werden. Dies könnte bei einer sich epidemieartig ausbreitenden Rankingmentalität allzu sehr in den Hintergrund treten. Zudem könnten auch hierbei wiederum die Schwächeren die Leidtragenden sein. Den Schulen, die in den Rankings schlecht abschneiden, ist mit der Veröffentlichung oder Rückmeldung niederschmetternder Befunde nicht gedient (vgl. Ditton 2007, S. 91).

> „Öffentliche Schulrankings, die ein differenziertes und stark von den Bedingungen der jeweiligen Schule abhängiges Schulergebnis auf einen leicht kommunizierbaren Punktwert reduzieren, können zu Prozessen eines öffentlichen ‚naming and blaming' führen, die ungerecht und unfruchtbar für die pädagogische Arbeit in den Schulen sind" (Orth 2005, S. 64f).

3.1.2.5 Kritikpunkt: Kontrolle durch externe Evaluationsverfahren

Lehrerinnen und Lehrer fühlen sich teilweise durch Vergleichsarbeiten und andere externe Evaluationsverfahren kontrolliert. Der Grund hierfür liegt darin, dass die Ergebnisse nicht allein der Schule bzw. den Lehrpersonen, sondern auch der Schulaufsicht zur Verfügung gestellt werden. Oftmals bleibt es intransparent, welchen Zweck diese Rückmeldung an die Aufsicht hat (vgl. Köller 2007, S. 100).

Dadurch wird Evaluation oft mit Kontrolle und Fremdbestimmung assoziiert und weniger damit, dass durch empirische Daten Hinweise für Verbesserungen gewonnen werden können. Auch wenn sich in den vergangenen Jahren im Zuge der sogenannten ‚empirischen Wende' einiges geändert hat, ist eine Evaluationskultur im deutschen Schulwesen gerade erst dabei sich zu entwickeln (vgl. Ditton und Arnoldt 2004).

Weniger belastend wären dementsprechend Formen der Selbstevaluation, bei denen die gesammelten Daten in der Schule bleiben und zur Schul- und Unterrichtsentwicklung genutzt werden können, im Sinne formativer Evaluationsverfahren.

3.1.2.6 Kritikpunkt: "teaching to the test"

Das Problem des ‚teaching to the test' beschreibt eine pädagogische Praxis, die in erster Linie auf die Vermittlung der vorher bekannten Testinhalte abzielt. Die gesamte Unterrichtsarbeit wird zur permanenten Testvorbereitung genutzt (vgl. Orth 2005, S. 64). Dies kann dazu führen, dass Vertiefungsthemen, Projekte, Museumsbesuche, Gruppenunterricht nicht mehr berücksichtigt werden können (vgl. Herzog 2010, S. 41). Die Folgen aus dieser Einengung des Unterrichts auf die erwarteten Testbereiche liegt auf der Hand: „Die Tests büßen an Validität ein, die erzielten Testergebnisse sind höchstwahrscheinlich überschätzt und führen möglicherweise in die Irre" (Bonsen und von der Gathen 2004, S. 246).

Dieses Phänomen wird vor allem dann beobachtet, wenn mit den Testresultaten negative Konsequenzen im Sinne des high-stakes testing drohen, wie Untersuchungen zu standardbasierten Schulrefomen in angloamerikanischen Ländern zeigen. Wenn also die vergleichende Leistungstestung die Grundlage für die Vergabe von Bildungsabschlüssen oder zur Basis für mittelfristige oder langfristige Selektionsentscheidungen genutzt wird.

Wie in diesem Kapitel aufgezeigt wurde, existieren mittlerweile in allen 16 Bundesländern Vergleichsarbeiten in Form von Lernstandserhebungen. Die Konzeptionen, Zielsetzungen und Durchführungsmodalitäten unterscheiden sich in den Ländern teilweise erheblich voneinander. Ebenso werden in der Literatur zahlreiche Kritikpunkte genannt, die in Zusammenhang mit Vergleichsarbeiten gebracht werden. Im Folgenden sollen nun konkrete Umsetzungsbeispiele aus der Praxis vorgestellt werden: Beim ersten Beispiel handelt es sich um das Bundesländer übergreifende Projekt VERA, das in den Grundschulen aller 16 Bundesländer unter der Ägide der Universität Landau durchgeführt wird (3.2.1). Als länderspezifisches Exempel für Vergleichsarbeiten in der Sekundarstufe des allgemein bildenden Schulsystems wurden die baden-württembergischen Vergleichsarbeiten gewählt (3.2.2).

3.2 Vergleichsarbeiten in der Praxis

3.2.1 Länderübergreifende Projekte in der Grundschule – das Projekt VERA - ein Kooperationsprojekt der Bundesländer

In der Primarstufe wird in allen Bundesländern der Bundesrepublik Deutschland das Projekt VERA (**VER**gleichs**A**rbeiten) durchgeführt. Es umfasst eine flächendeckende Lernstandserhebung, die inzwischen am Ende der 3. Klassenstufe in den Fächern Mathematik und Deutsch durchgeführt wird (vgl. Universität Landau 2006). Entwickelt wurde es im Jahre 2002 an der Universität Landau von Andreas Helmke und Ingmar Hosenfeld in Kooperation mit dem rheinland-pfälzischen Ministerium für Bildung, Wissenschaft, Jugend und Kultur. Zwischen den Jahren 2002 und 2006 wurden die Vergleichsarbeiten zunächst in sieben Bundesländern noch zu Beginn der 4. Klassenstufe geschrieben. Seitdem im Jahr 2007 der Termin für die Durchführung der Vergleichsarbeiten auf das Ende der 3. Klassenstufe umgestellt wurde, beteiligen sich alle 16 Bundesländer an dem Programm (vgl. Ditton 2008, S. 41). Seit dem Jahr 2009 findet VERA unter der Ägide des länderübergreifenden Instituts für Qualitätsentwicklung in Berlin (IQB) statt.

In die Vergleichsarbeiten fließen normierte, auf die Bildungsstandards bezogene Aufgaben ein. Im Bereich der Aufgabenentwicklung, besteht beim Projekt VERA eine Besonderheit: Hier haben die Schulen selbst die Möglichkeit, neue Aufgaben zu entwickeln und für die jeweils folgende Normierungsstichprobe vorzuschlagen. Die Arbeiten beinhalten bei VERA zentrale und dezentrale Aufgabenstellungen, d.h. 50% der Aufgaben entstammen direkt dem zentralen Aufgabenpool, die restlichen Aufgaben werden von den Lehrpersonen der jeweiligen Schule aus dem selben Pool unter Berücksichtigung der pädagogischen Schwerpunktsetzung und des sozialen Umfelds der Schule ausgewählt (vgl. van Ackeren & Bellenberg 2004, S. 139).

Die Ziele von VERA umfassen u.a. die Schul- und Unterrichtsentwicklung, die Erfassung und Verbesserung der diagnostischen Kompetenzen der Lehrerinnen und Lehrer, die Standardsicherung und –entwicklung. Des Weiteren soll VERA einen Beitrag zur Effizienzsteigerung bei der Nutzung des Internets für die schuli-

sche Qualitätssicherung leisten und die Umsetzung moderner Kerncurricula, Lehr- und Rahmenpläne erleichtern und beschleunigen. Nicht zuletzt liefert das Evaluationsprogramm ergänzende Informationen über die Leistungen von Schülerinnen und Schülern, die auch für Elterngespräche genutzt werden können. (Vgl. Groß Ophoff et al. 2006, S. 20–22).

VERA verwendet sowohl verteilungsbezogene soziale Bezugssysteme, die durch die Erhebung von Kontextmerkmalen einen fairen Vergleich ermöglichen. Ebenso lässt sich durch den Bezug zu den Bildungsstandards ein kriterialer Vergleich durchführen. Außerdem können Vergleiche im zeitlichen Verlauf dargestellt werden. Die Konzeption sieht eine Unterscheidung in drei Fähigkeitsniveaus vor (elementare – erweiterte – fortgeschrittene Fähigkeiten), die in Kooperation mit fachdidaktischen Experten ermittelt wurden (vgl. ebd., S. 23–25).

VERA stellt im Bereich der Lernstandserhebungen in der Grundschule dadurch, dass sich alle Bundesländer daran beteiligen, eine Ausnahme dar. In der Sekundarstufe gibt es keine bundesübergreifenden Konzepte. Die Konzeptionen der Vergleichsarbeiten werden in den Ländern unterschiedlich umgesetzt und unterscheiden sich somit von Bundesland zu Bundesland.

3.2.2 Länderspezifische Projekte in der Sekundarstufe – am Beispiel: Vergleichsarbeiten in Baden-Württemberg

3.2.2.1 Einführung in länderspezifische Unterschiede

Vergleichsarbeiten werden mittlerweile in allen Bundesländern nicht nur in der Grundschule, sondern auch in der Sekundarstufe als Evaluationsinstrumente eingesetzt. Einbezogen sind stets die Fächer Deutsch und Mathematik, dazu kommen je nach Bundesland und Schulart weitere Fächer. Die Durchführung wird zum größten Teil von der Bildungsadministration des jeweiligen Bundeslandes verwaltet und beaufsichtigt. An der Generierung von Aufgaben wirken in jeweils unterschiedlichen Konstellationen Fachlehrpersonen, Mitarbeiter der Landesinstitute, die Schulaufsicht, das Ministerium und Wissenschaftler mit (vgl. van Ackeren & Bellenberg 2004, S. 139).

In der Praxis hat sich herausgestellt, dass aufgrund länderspezifischer Bildungstraditionen, wegen aktueller gesetzlicher bzw. schulrechtlicher Vorgaben oder aus anderen Gründen teilweise deutliche Unterschiede im Hinblick auf Vorbereitung, Durchführung und Weiterarbeit mit den Ergebnissen existieren. So werden die ermittelten Schülerleistungen in einigen Ländern benotet oder werden bei Übertrittsentscheidungen mit herangezogen; in anderen Ländern werden die Ergebnisse klassen- bzw. schulintern als Diagnoseinstrument und damit als Grundlage für eine individuelle Förderung genutzt, ohne dass dies unmittelbare schulrechtliche Konsequenzen hätte.

Ein weiterer Unterschied zwischen den Bundesländern zeigt sich bei der Korrektur der Vergleichsarbeiten: Der überwiegende Teil der Länder lässt die Arbeiten zum Teil klassenintern korrigieren, ein anderer Teil der Länder hingegen lässt zunächst eine wissenschaftliche, externe Auswertung der Arbeiten erfolgen. In einigen Bundesländern spielt die Schulaufsicht bei der Ergebnissicherung und -auswertung der Vergleichsarbeiten eine zentrale Rolle: In Baden-Württemberg müssen ausgewählte Schulen ihre Ergebnisse an das zuständige Landesinstitut zurückmelden, welches dann landesweite Vergleichsergebnisse zur Verfügung stellt. Ähnliche Verfahren gibt es auch in Bremen und Hamburg. Im Saarland und in Thüringen übernehmen Wissenschaftler die erste Datenauswertung und stellen diese der Schulaufsicht zur Verfügung, welche sie für die Schulen als Vergleich nutzbar machen muss. Uneinheitlich wird auch gehandhabt, ob die Ergebnisse aus Vergleichsarbeiten in die Zeugnisnote einfließen und somit als Selektionskriterium dienen.

In allen Bundesländern sind Bemühungen erkennbar, die individuellen Einzelschulergebnisse durch landesweite, regional bzw. sozial vergleichbare Ergebnisse differenzierbar und beurteilbar zu machen (vgl. ebd., S. 133ff).

Eine weitere Gemeinsamkeit in allen Bundesländern ist, dass zentrale Lernstandserhebungen das Ziel verfolgen, Lehrerinnen und Lehrer darin zu unterstützen, die fachlichen Leistungen ihrer Schülerinnen und Schüler an Kriterien und Standards zu messen. Damit soll eine schulübergreifende Standortbestimmung vorgenommen und Ansatzpunkte für die Weiterentwicklung des Unterrichts bereitgestellt werden.

Leitend ist dabei der Gedanke einer Professionalisierung des fachlichen Diskurses innerhalb einer Schule im Kontext ergebnisorientierter Schul- und Unterrichtsentwicklung (vgl. Peek & Dobbelstein 2006b, S. 177).

3.2.2.2 Vergleichsarbeiten in Baden-Württemberg

In Baden-Württemberg werden seit dem Schuljahr 2005/2006 in allen Schularten verbindlich Vergleichsarbeiten durchgeführt, nachdem erste Probedurchgänge bereits seit dem Jahr 1999 stattfanden (vgl. Maier 2008a). Während sie in der Grundschule anfangs Orientierungsarbeiten genannt wurden, nennt man sie in der Sekundarstufe Vergleichsarbeiten[6].

Das Landesamt für Schulentwicklung zeigt sich für die Konzeption, Aufgabenentwicklung, Erstellung der Vergleichsarbeiten, Durchführung und Auswertung verantwortlich. Bis zum Schuljahr 2007/2008 wurden die Vergleichsarbeiten am Ende eines vorgegebenen Bildungsabschnittes durchgeführt und benotet.

Im Oktober 2008 gab das Kultusministerium Änderungen zu der bisherigen Durchführung der Vergleichsarbeiten bekannt. Die Änderungen betreffen einerseits die Benotung, die ab dem Schuljahr 2009/10 abgeschafft wird. Andererseits wird der Termin der Vergleichsarbeiten vom Ende der Klassen sechs, acht und zehn auf den Beginn der Klassen sieben, neun und elf verschoben. Mit diesem Wechsel soll deutlich werden, dass es sich bei den Arbeiten nicht um Klassenarbeiten im klassischen Sinne handelt, auf die man wochenlang lernt, sondern um eine Überprüfung der Kompetenzen. Der vorausgegangene Bildungsabschnitt, der jeweils zwei Schuljahre umfasst, soll überprüft werden. Die Verlagerung der Vergleichsarbeiten an den Beginn eines neuen Lernabschnitts eröffnet nach Meinung des Philologenverbandes Baden-Württemberg (PHV BW) neue Perspektiven. Die Lehrpersonen könnten zu Beginn des neuen Lernabschnitts mit Hilfe der Vergleichsarbeiten eine Lernstandserhebung durchführen. So könnten eventuelle Lerndefizite erkannt und bei der Konzeption des Unterrichts berücksichtigt werden (vgl. Philologenverband Baden-Württemberg 2008).

---

[6] http://www.schule-bw.de/schularten/gymnasium/zentralepruefungen/diagnosevergleichsarbeiten/index.html

3.2.2.3 Konzeption und Zielsetzung der baden-württembergischen Vergleichsarbeiten

Konzeption

Zunächst wird die Konzeption der Vergleichsarbeiten im allgemein bildenden Schulsystem in Baden-Württemberg beschrieben. Neben allgemeinen Vorgaben und rechtlichen Grundlagen werden schulartspezifische Unterschiede erläutert und diejenigen Fächer genannt, die in der jeweiligen Schulart involviert sind. Die Grundschulen, die sich seit dem Jahr 2008 am länderübergreifenden VERA-Projekt beteiligen, sind nicht Gegenstand der Untersuchung (vgl. Landesinstitut für Schulentwicklung 2007, 2009).

Vergleichsarbeiten in Baden-Württemberg sind in einen vorgegebenen rechtlichen Rahmen eingebettet, der durch das Schulgesetz und Verwaltungsvorschriften reglementiert wird[7]. § 114 des baden-württembergischen Schulgesetzes regelt die Evaluation der Schulen. Diese müssen regelmäßig an Selbst- und an Fremdevaluationen teilnehmen, die der Bewertung ihrer Schul- und Unterrichtsqualität dienen. Des Weiteren sind Schulen verpflichtet an Lernstandserhebungen (international, national oder landesweit) teilzunehmen. Die Verfahren werden durch Rechtsverordnungen des Kultusministeriums flankiert.

Eine Neuerung betrifft ab dem Schuljahr 2008/2009 den Zeitpunkt der Durchführung und die Benotung der Vergleichsarbeiten. Die Verwaltungsvorschrift vom 17. Mai 2009 schreibt die Durchführung von Vergleichsarbeiten in Zukunft zu Beginn eines Schuljahres in folgenden Klassenstufen und Fächern fest:

- in den Klassen 7 der Hauptschulen, Realschulen und Gymnasien in den Fächern Deutsch und Mathematik,
- in den Klassen 9 der Realschulen in den Fächern Deutsch, Mathematik und der Pflichtfremdsprachen,
- in den Klassen 9 der Gymnasien in den Fächern Deutsch, Mathematik sowie nach Entscheidung der Gesamtlehrerkonferenz in einer der Fremdsprachen[8].

[7] http://www.schule-bw.de/entwicklung/dva/vadva/konzeption-dva/vorgaben/
[8] (http://www.schule-bw.de/entwicklung/dva/dva_docs/docs4dva/VwV_09_05_17.pdf)

Im Schuljahr 2006/2007, also dem zweiten Jahr seit der verpflichtenden Einführung von Vergleichsarbeiten und Gegenstand der vorliegenden Studie, wurden die Vergleichsarbeiten noch am Ende des 6. bzw. 8. Schuljahrs durchgeführt. In allen drei Schularten wurden die Vergleichsarbeiten in den Fächern Mathematik und Deutsch geschrieben. Je nach Schulart wurden die Kompetenzen der Schülerinnen und Schüler in weiteren Fächern erhoben: In der **Hauptschule** war die Durchführung von zwei Vergleichsarbeiten verpflichtend. Die Lehrpersonen konnten innerhalb der Fächer Mathematik, Deutsch und Englisch wählen. In der **Realschule** wurden im Schuljahr 2006/2007 Vergleichsarbeiten in den Klassen 6 und 8 durchgeführt. Neben den Fächern Deutsch und Mathematik hatten die Lehrpersonen zusätzlich die Wahl zwischen dem Fach Geschichte oder dem Fächerverbund Erdkunde-Wirtschaftskunde-Gemeinschaftkunde (EWG). In den baden-württembergischen **Gymnasien** waren Vergleichsarbeiten in den Klassen 6, 8 und 10 ebenfalls in den Fächern Deutsch und Mathematik verpflichtend. Die Wahl für ein zusätzliches Fach bestand für die Lehrenden zwischen dem Fach Biologie und dem Fächerverbund GWG (Geographie – Wirtschaft – Gemeinschaftskunde).

Zielsetzungen

Auf Grundlage des Schulgesetzes § 114 Evaluation und des Bildungsplans 2004 ergeben sich folgende Zielsetzungen der baden-württembergischen Vergleichsarbeiten[9].

Qualitätsentwicklung der Schule: Vergleichsarbeiten helfen zu überprüfen, inwieweit es den Schulen gelungen ist, die Erwartungen der baden-württembergischen Bildungsstandards zu erreichen und dienen damit der empirisch gesicherten, zielgerichteten und systematischen Qualitätsentwicklung. Die Arbeiten liefern landesweite Vergleichswerte für die Selbstevaluation der Schulen. Der Umgang der Schulen mit den Ergebnissen ist Gegenstand der Fremdevaluation.

Feststellung des Lernstands der Schülerinnen und Schüler: Die Vergleichsarbeiten vermitteln den Lehrpersonen, den Schülerinnen und Schülern sowie deren Eltern in bestimmten Fächern objektive Informationen über den individuellen Lernstand im Hinblick auf ausgewählte Schwerpunktbereiche des Bildungsstandards.

[9] (http://www.schule-bw.de/entwicklung/dva/vadva/konzeption-dva/ziele/)

Die Informationen zu den Lernvoraussetzungen der Klasse und einzelner Schülerinnen und Schüler, die zu Beginn eines neuen Bildungsabschnittes festgestellt werden, können bei der gezielten Planung des weiteren Unterrichts helfen.

3.2.2.4 Verwendung der Ergebnisse aus Vergleichsarbeiten für die Selbstevaluation

Das Landesinstitut für Schulentwicklung in Stuttgart gibt den Lehrerinnen und Lehrern für die zielgerichtete und systematische Nutzung der Daten aus Vergleichsarbeiten drei Arbeitsschritte vor: In einem ersten Schritt soll die Analyse der Ergebnisse erfolgen. Diese Analyse bildet dann die Grundlage für die Interpretation. In einem letzten Punkt werden die Lehrerinnen und Lehrer aufgefordert, entsprechende Konsequenzen daraus zu ziehen.

1. Schritt: Ergebnisse analysieren: Für die Durchführung und Auswertung der Vergleichsarbeiten werden den Lehrpersonen spezielle Manuale, sog. „Testspezifische Hinweise“ ausgehändigt. Die Korrektur der gelösten Vergleichsarbeiten der Schülerinnen und Schüler erfolgt anhand dieser explizit formulierten Vorgaben. Die Lehrpersonen sind angehalten, diesen Lösungsvorschlägen unbedingt Folge zu leisten. Die richtigen Ergebnisse der Schülerinnen und Schüler werden mit einem „X“ in einer Lösungsmatrix markiert. Daraus berechnet das Programm die Noten automatisch. Als Referenzwert gelten die Ergebnisse aus Pilotierungsstichproben, die an ausgewählten Schulen des Landes im Vorfeld der Vergleichsarbeiten ermittelt wurden. Der Bewertungsmaßstab richtet sich an einem ‚mittleren Schüler’ aus, dessen Testergebnis der Note 3,0 zugeordnet wird. Die Daten werden den Lehrerinnen und Lehrern auf Item-, Schüler- und Klassenebene zur Verfügung gestellt.

2. Schritt: Ergebnisse interpretieren: Das Manual enthält zum Thema „Umgang mit den Testergebnissen“ keinerlei Anleitungen, lediglich einen Link zu einer Internetseite. Dort findet man die 6-seitige Handreichung „Umgang mit den DVA-Ergebnissen im Rahmen der Selbstevaluation“ zum Herunterladen, die zur Analyse der Ergebnisse und zur Unterrichtsarbeit genutzt werden soll. Sie enthält Vor-

schläge und Anregungen, die den Schulen helfen sollen, die Ergebnisse der Vergleichsarbeiten für die Weiterentwicklung erfolgreich zu nutzen.[10]

Unter dem Punkt „Ergebnisse interpretieren“ findet man den Hinweis, dass nach der Datenauswertung erste Überlegungen angestellt werden sollen, die das Abschneiden der Schülerinnen und Schüler erklären. Dabei solle man die Ergebnisse nicht nur als Einzeldaten interpretieren sondern auch im Gesamtzusammenhang sehen; gemeint ist hier das Ergebnis der einzelnen Schüler, die Klassenergebnisse, der Unterricht und die Schule.

3. Schritt: Konsequenzen ziehen

Aufgrund der vorangegangenen Interpretation der Daten sollen passende schulische und unterrichtliche Maßnahmen eingeleitet werden. Der Erfolg von Maßnahmen sollte im Rahmen der Selbstevaluation regelmäßig an den Schulen reflektiert und evaluiert werden. Als mögliche Bereiche werden genannt: Diagnose und Förderung, Unterrichtsqualität und der Aufbau von Kooperations- und Teamstrukturen. Ab dem Jahr 2009 steht den Lehrpersonen eine ausführlichere Handreichung mit einem Umfang von 28 Seiten zur Verfügung.[11]

Auf Grundlage der beschriebenen Verordnungen und Vorgaben sind die Lehrerinnen und Lehrer der baden-württembergischen Sekundarstufe verpflichtet, die Vergleichsarbeiten in ihren Klassen durchzuführen. Ob mit diesen Verfahren und Vorgehensweisen die Ziele automatisch erreicht werden, die Vergleichsarbeiten intendieren, kann auf dieser theoretischen Basis nicht geklärt werden.

Um die tatsächliche Wirkung von Vergleichsarbeiten im Sinne einer Qualitätssicherung oder Qualitätsentwicklung müssen zunächst also alle Bedingungen, Ebenen und Akteure des Bildungswesens berücksichtigt werden, die bei der Implementierung von Vergleichsarbeiten in irgendeiner Form betroffen sind. Im Anschluss daran kann empirisch überprüft werden, welche Handlungslogiken die betroffenen Steuerungsakteure bei der Implementierung der Vergleichsarbeiten entwickeln und ob diese den offiziellen Zielsetzungen Landes entsprechen.

---

[10] (http://www.eltern.schule-bw.de/protokolle/07nov24datenUmgangLS.pdf)
[11] (http://www.schule-bw.de/entwicklung/dva/dva_docs/docs4dva/Umgang_mit_Ergeb_09.pdf)

3.3 Implementierung von Vergleichsarbeiten als Instrument der Qualitätsentwicklung und -sicherung

3.3.1 Allgemeine Überlegungen und Begriffsbestimmungen

„Implementierung“ ist ein Begriff, der ursprünglich aus den Ingenieurwissenschaften stammt und später von der Computertechnologie übernommen wurde. Im politischen System steht Implementierung für eine „politik- und verwaltungswissenschaftliche Bezeichnung für den Prozess der inhaltlichen Umsetzung von Gesetzen, Verordnungen, etc. in konkrete Tatsachen und materielle Leistungen. Die Implementierung politischer Entscheidungen ist in erster Linie eine Aufgabe der verschiedenen Ebenen staatlicher und kommunaler Verwaltung“ (Schubert & Klein 2006).

Im Fall der Implementierung von Vergleichsarbeiten als Instrument zur Qualitätsentwicklung und –sicherung handelt es sich um eine politisch intendierte Vorgabe, die im komplexen Bildungssystem verankert werden soll. Dabei ist es wichtig, die unterschiedlichen sozialen Kontexte, Ebenen und Prozesse im Sinne der „Educational Governance“ zu unterscheiden und zu berücksichtigen. Im Bildungswesen sind dies Akteure, die auf den unterschiedlichen Ebenen des System agieren; auf der Makro-, der Meso- und Mikroebene.

Aus den Erfahrungen, die man bei der Implementierung von Bildungsstandards gewonnen hat, weiß man, dass sich komplexe schulische Neuerungen nicht einfach als Steuerungsinstrument in das komplexe Bildungssystem integrieren lassen, sondern „dass man auch mit Problemen der Übersetzung rechnen muss“ (Oelkers & Reusser 2008, S. 47).

> „Adressat von Implementierungsstrategien sind nicht nur Lehrpersonen und die Bildungsverwaltung, sondern auch die Eltern, die Schülerinnen und Schüler, die Medien und nicht zuletzt auch die lokale Öffentlichkeit. Die Kommunikation mit diesen verschiedenen Akteursgruppen bringt divergierende Interpretationen mit sich, die sich auf ein bestimmtes Konzept von Bildungsstandards beziehen müssen. Dieses Konzept muss überzeugen und kann nicht einfach verordnet werden“ (ebd.).

Zwischen diesen einzelnen Ebenen muss es Verknüpfungen geben, die den Implementierungsprozess unterstützen und ihm nicht entgegenwirken. Diese Verknüpfungen lassen sich nicht verordnen, sondern müssen sich einspielen.

Mit dem „Educational Governance“ – Ansatz lassen sich auf allen Ebenen des Systems Bedingungen formulieren, die zu einer gelingenden Implementierung der Vergleichsarbeiten beitragen.

3.3.2 Gelingensbedingungen für die Implementierung von Vergleichsarbeiten im Mehrebenensystem

Inwieweit Lernstandserhebungen bzw. Vergleichsarbeiten positive Impulse in die Schulen bringen und dazu beitragen, fachdidaktische und pädagogische Zielperspektiven weiterzuentwickeln, hängt von mehreren Faktoren ab: Zum Beispiel spielt neben der Qualität der Testaufgaben auch die Auseinandersetzung und Akzeptanz von Lernstandserhebungen in den Kollegien eine wichtige Rolle. Des Weiteren stellt die Art und Weise des Umgangs mit den Ergebnissen sowohl in den Schulen als auch auf Seiten der Bildungspolitik einen wichtigen Faktor dar (vgl. Peek & Dobbelstein 2006b, S. 192). Im Folgenden wird jede Ebene des Schulsystems betrachtet, um deren Einfluss auf eine gelingende Implementierung darzustellen.

3.3.2.1 Gelingensbedingungen auf Systemebene

3.3.2.1.1 Anforderungen an Tests

Vergleichsarbeiten, als Instrument der Bildungspolitik, können aufgrund ihres Durchführungsdesigns als Vollerhebung ein bedeutsames Potential hinsichtlich der Schul- und Unterrichtsentwicklung bieten. Die Grundlage hierfür bildet die Qualität der Daten. Es muss sichergestellt werden, dass die Tests durch gewissenhafte Entwicklungs- und Pilotierungsarbeiten und durch ein angemessenes System zur Qualitätssicherung insgesamt tragfähig und belastbar sind (vgl. Bonsen et al. 2004a, S. 134).

Für die Nutzung dieses Potenzials formuliert Popham (2003, S. 51) notwendige Mindestanforderungen an die Durchführung von Vergleichsarbeiten. Sie müssen zum einen bedeutsame und gleichzeitig konkret beschriebene Fähigkeiten und Kenntnisse erfassen, die auch tatsächlich durch den schulischen Fachunterricht gefördert werden können. Die Testdurchführung muss sowohl für die Lehrerinnen und Lehrer als auch für die Schülerinnen und Schüler vom Aufwand her verkraftbar bleiben. Außerdem müssen die Ergebnisse verständlich und im Hinblick auf den Effekt des eigenen Unterrichts bedeutsam sein. Zuletzt müssen grundlegende methodische Anforderungen erfüllt sein, damit Rückmeldungen ihr Potenzial entfalten können. (Vgl. Bonsen & von der Gathen 2004, S. 239f; Nachtigall & Kröhne 2006, S. 59f).

Weitere Anforderungen werden an die Erhebung und Auswertung der Lernstandserhebungen gestellt. Sie sollten von Institutionen verantwortet werden, die von Politik, Verwaltung und Schulfeld unabhängig sind. Zudem sollen sie den Akteuren im Bildungswesen gezielt jene Daten zukommen lassen, die diese für ihren Handlungsbereich benötigen. Damit die Akteure mit den Ergebnissen aus Schulleistungstests arbeiten können, muss die Auswertung der Daten in einem Rückmeldeformat aufbereitet sein, das deren Handlungslogik entspricht. Je nach Adressat, z.B. Administration, Schulen oder Klassen sollen die Rückmeldungen in der Form präsentiert werden, die von der jeweiligen Ebene auch verstanden wird (vgl. Bähr 2006, S. 135ff).

3.3.2.1.2 Wandel der Rolle der Schulaufsicht

Bedingt durch die empirische Wende in der Bildungspolitik und der größeren schulischen Eigenständigkeit wird auch die Rolle der Schulaufsicht zwischen Beratung und traditioneller Schulaufsicht neu diskutiert. Einigkeit besteht darüber, dass es Auftrag der Schulaufsicht ist, dafür Sorge zu tragen, dass die Einhaltung entsprechender Vorgaben zur Qualitätssicherung und Qualitätsentwicklung in der Einzelschule vor Ort erfolgt (vgl. Rekus 2007; Stryck 2000). Dies setzt voraus, dass es innerhalb der Schulaufsicht vereinbarte Standards für die Begleitungs- und Beratungsarbeit gibt und dass diese zwischen Schulaufsicht und Fortbildung abgestimmt sind und gegenüber den Schulen offengelegt werden (vgl. Hoffmann-Göttig et al. 2005, S. 35f).

Unterstützung der Einzelschule im Prozess der Qualitätssicherung und Qualitätsentwicklung stellt eine Hauptaufgabe der Schulaufsicht dar. Dies wird von Artelt (2007) in drei Punkten folgendermaßen formuliert:

- Qualitätsentwicklung und Qualitätssicherung der einzelnen Schule muss angestoßen und begleitet werden. Hierzu muss die Schulaufsicht vor allem Schulleitungen und Kollegien bei internen Prozessen zur Qualitätssicherung und Qualitätsentwicklung beraten und unterstützen.
- Verbindliche Verfahren der Qualitätsentwicklung und Prüfung der Qualitätssicherung müssen in Schulen implementiert werden. Der Schulaufsicht kommt dabei auch die Aufgabe zu, Kontrasterfahrungen und schulübergreifenden Austausch zu ermöglichen.
- Es gilt, schulübergreifend Vergleichbarkeit und Standards zu sichern und zugleich auf Kontinuität und Verbindlichkeit bei der Umsetzung von Konsequenzen aus der schulinternen Evaluation zu achten. Dabei hilft die Schulaufsicht Steuerungswissen zu generieren. (Vgl. ebd., S. 134)

3.3.2.1.3 Unterstützungssysteme

Die Wirkung von Evaluationsmaßnahmen hängt entscheidend von unterstützenden Maßnahmen und Angeboten zur Datennutzung ab, über die bislang wenig systematisch nachgedacht wurde bzw. für die bislang kaum Erfahrungswerte vorliegen (vgl. van Ackeren & Bellenberg 2004, S. 158f). Schulen wünschen sich neben einer professionellen Unterstützung bei der Konzeption der Qualitätsentwicklung insbesondere eine professionelle Prozessbegleitung (vgl. Peek 2004b, S. 112). Außerdem benötigen sie Hilfe bei Grundfragen der Personalentwicklung und beim Aufbau unterstützender Strukturen.

> „Der Glaube, allein die Testung würde zu einer stetigen Unterrichtsentwicklung führen […] ist unrealistisch, vielmehr können Standards und das an sie angebundene System der flächendeckenden Vergleichsarbeiten und stichprobenbasierten *Assessments* nur dann Wirkung entfachen, wenn sie durch entsprechende Unterstützungssysteme flankiert werden“ (Köller 2008, S. 68), [Hervorhebung im Original).

Nach Steffens (2007, S. 39f) sollte sich der Auftrag der Unterstützungssysteme einerseits auf die Beratung und Hilfe sowie andererseits zur Sicherung der Standards auf ein Konzept konzentrieren, das folgenden Gesichtspunkten Rechnung tragen soll. Es soll einerseits Informationen liefern über nachhaltige Ergebnisse des Unterrichts und Bedingungen und Wege aufzeigen, wie Qualität verbessert werden kann. Andererseits soll es den Akteuren Hilfestellung bieten bei der Interpretation der Daten. Außerdem soll das Konzept einen fundierten Überblick in die Schullandschaft der Region bereitstellen. Und zuletzt soll es unter ‚Normalbedingungen‘ praktikabel und in seinen Zielsetzungen und Verfahren nachvollziehbar sein.

Wie ein solches professionelles und integriertes Unterstützungssystem aufgebaut werden kann wurde in Nordrhein-Westfalen anhand zweier Projekte (‚Schule & Co‘ und ‚Selbständige Schule‘) ermittelt. Es umfasst vier Bereiche, die sich alle auf die Ebene der Einzelschule beziehen (vgl. Bastian 2007, S. 59f). Neben der Unterstützung einer systematischen und teamorientierten Unterrichtsentwicklung des kompletten Kollegiums sowie der Schulleitung sieht es die Unterstützung ei-

nes systematischen Managements und zuletzt auch die Förderung einer Evaluationskultur an den Schulen vor.

3.3.2.2 Ebene der Einzelschule

Wie schon in dem Unterkapitel ‚Anforderungen an Tests' beschrieben wurde, ist es für eine wirksame Schul- und Unterrichtsentwicklung zunächst einmal erforderlich, dass die Rückmeldungen und Informationen aus Lernstandserhebungen in einem Format erstellt sind, die den Handlungslogiken der einzelnen Akteure entspricht (vgl. Bähr 2006, S. 135ff). Sie müssen also in einer Sprache und einer Form bereitgestellt werden, die es den Adressaten erlaubt, sich selbst das nötige Verständnis zu verschaffen (vgl. Peek 2004b, S. 112). Zudem ist es von Vorteil, wenn Daten nicht nur aus einer einzigen sondern aus verschiedenen Untersuchungen vorliegen.

Bei der Implementierung von Vergleichsarbeiten auf der Ebene der Einzelschule kommt der Schulleitung, den Fachkonferenzen und Fachgruppen eine besondere Bedeutung zu.

3.3.2.2.1 Rolle der Schulleitungen

Schulleitungen nehmen im Implementierungsprozess eine Schlüsselrolle ein. In erster Linie sind die Schulleitungen gefordert dafür zu sorgen, dass die schulbezogenen Ergebnisse in den Gremien, wie z.B. der Schul- und Gesamtlehrerkonferenz genügend Beachtung finden und besprochen werden (vgl. Rolff 2002b, S. 349f). Insbesondere gilt, dass die Schulleitungen selbst das Konzept von Standardsetzung im Sinne der Bildungsstandards kennen und dem Kollegium vermitteln können. Außerdem sollten sie in der Lage sein, den Akteuren ihrer Schule den Zusammenhang von Bildungsstandards, Schulleistungsstudien, Lernstandserhebungen und schulischem Qualitätsmanagement transparent machen zu können. Und schließlich sollten sie die Möglichkeiten und Grenzen der pädagogischen Nutzung von Lernstandserhebungen bzw. Vergleichsarbeiten kennen, um deren Potentiale in den Schulalltag einfließen zu lassen (vgl. Bonsen et al. 2004a, S. 135f).

Schwierigkeiten könnten auf der Ebene der Schulleitung dann entstehen, wenn die Prozesse der Selbstevaluation nicht professionell koordiniert und strukturiert sind. Ein weiteres Problem kann der Wegfall von follow-up Arrangements darstellen. Das bedeutet, dass im Anschluss an die Durchführung des Evaluationsverfahrens keine weiteren Schritte folgen, die sich aufgrund der gewonnen Erkenntnisse anböten. Ein Vergleich zwischen anfänglich formulierten Zielvereinbarungen und evaluierten Ergebnissen wird dann nicht thematisiert und bleibt aus (vgl. Artelt 2007, S. 134).

3.3.2.2.2 Rolle der Fachgruppen bzw. Fachkonferenzen

Die jeweiligen Fachgruppen und Fachkonferenzen sollen die Ergebnisse ihrer Schule interpretieren, Ursachen benennen und Konsequenzen für die Unterrichtsentwicklung ziehen. Dies erfordert eine fachliche Kommunikations- und Kooperationskultur an den Schulen, die sich realistisch gesehen in vielen Schulen erst entwickeln muss (vgl. Bonsen et al. 2004a, S. 135; Thürmann 2002). Wenn die Lehrpersonen auf Daten zurückgreifen können, die aus unterschiedlichen Evaluationsverfahren generiert und in einem Datenpool an der Schule systematisch gesammelt wurden, kann dies die Schul- und Unterrichtsentwicklung unterstützen (vgl. Köller 2007, S. 99).

3.3.2.2.3 Datenpools – Data-Richness

Neben der Deskription valide gewonnener Leistungsdaten müssen diese zu den Schulkontexten in Beziehung gesetzt werden. Um dies gewährleisten zu können, benötigen Schulen unterschiedliche Arten von Daten. Bonsen und Gathen (2004) fordern für eine effektive Nutzung von Evaluationsverfahren die zur Qualitätssicherung und -entwicklung der Einzelschule beitragen sollen, vier verschiedene Arten von Daten:

1. Demografische Daten zu Schülern, zum Kollegium und zum Kontext der Schule
2. Daten zu Einschätzungen und Wahrnehmungen unterschiedlicher Gruppen, z.B. Schüler, Lehrer und Eltern

3. Schulische Prozessdaten, die sich z.B. auf das Schulprogramm oder das Lehrer- und Schülerhandeln im Unterricht beziehen
4. Die Leistungsdaten der Schüler selbst, die das Zentrum einer output-orientierten Evaluation von Schule und Unterricht darstellen

Mit Hilfe dieser vier Datenarten, können erstens Analysen über Zeitverläufe und zweitens Analysen unter Einbeziehung verschiedener Variablen durchgeführt werden. Allerdings benötigt eine Schule für die Erhebung und Organisation solcher unterschiedlicher, sich gegenseitig ergänzender Daten und Datenbestände ein umfassendes System des Datenmanagements (vgl. ebd., S. 240f).

Eine derart umfassende Datensammlung wird in der Literatur „data-warehouse“ oder „data-richness“ genannt. Darunter versteht man eine über die Zeit gewachsene Datenbank, die mit Hilfe einer anwendungsfreundlichen und einfachen Benutzeroberfläche den gezielten Zugriff auf unterschiedliche Daten-Gruppierungen ermöglicht. Sie soll den Verantwortlichen zu einer datenbasierten Entscheidungsfindung verhelfen mit dem Ziel der Optimierung der individuellen und gemeinsamen pädagogischen Praxis.

3.3.2.3 Ebene des Unterrichts

3.3.2.3.1 Lehrerinnen und Lehrer

Die Lehrerinnen und Lehrer stellen auf der Ebene des Unterrichts die zentralen Akteure bei der Implementierung von Vergleichsarbeiten dar. Sie sind in ihren Klassen nicht nur verantwortlich für die Vorbereitung, Durchführung und Auswertung der Vergleichsarbeiten, sondern auch für die Umsetzung daraus resultierender Maßnahmen. Neben der Akzeptanz gegenüber den Verfahren erfordert der Umgang mit den Daten auch ein gewisses statistisches und computertechnisches Grundverständnis der Lehrpersonen. Denn alle Daten der Schülerinnen und Schüler müssen am Computer in eine Datenmaske eingetragen werden. Dabei stehen den Lehrpersonen keine Ansprechpartner zur Verfügung, sondern alle fachliche Hilfestellungen und sämtliche Erklärungen und Anleitungen müssen über das Internet besorgt werden (vgl. Bonsen et al. 2004a, S. 133).

Neben dieser Erweiterung der Methodenkompetenz der Lehrerinnen und Lehrer können Lernstandserhebungen auch den Professionalisierungsprozess der Lehrpersonen in Gang setzten; zum einen müssen die Rückmeldungen professionell beurteilt werden, zum anderen können die Lehrerinnen und Lehrer in den Bereichen Diagnostizieren, Beurteilung, Beratung und Evaluation ihre Kompetenzen erweitern (vgl. Weinert 2002).

Ängste und Sorgen von Lehrpersonen vor verstärkter Kontrolle ihrer Arbeit im Gefolge von Vergleichsarbeiten bedürfen besonderer Beachtung bei der Implementierung dieses Instruments (vgl. van Ackeren & Bellenberg 2004, S. 158).

Wie Lehrpersonen dieses Instrument zur Qualitätssicherung und Qualitätsentwicklung ihrer Arbeit tatsächlich nutzen, steht im Zentrum der vorliegenden Arbeit und wird im empirischen Teil ausführlich untersucht.

3.3.2.3.2 Schülerinnen und Schüler

Die Schülerinnen und Schüler bilden die Basis der Datengewinnung, da mit Hilfe von Lernstandserhebungen deren Kompetenzen in den unterschiedlichen Domänen erfasst werden. Für die Reliabilität der Daten ist es entscheidend, dass die Schülerinnen und Schüler die Tests angstfrei, motiviert und engagiert durchführen.

3.3.2.4 Zusammenfassung

Eine gelingende Implementierung ist von vielen Faktoren, Akteuren und deren subjektiver Einstellungen innerhalb der unterschiedlichen Ebenen des Bildungssystems abhängig. Es kommt auf die Ausgestaltung, auf den Umgang in den Schulen und in der Bildungsadministration und auf ihre Einbettung in ein Gesamtkonzept der Qualitätssicherung schulischer Arbeit an. Dabei darf nicht vergessen werden, dass die realen Effekte eines solchen Systems letztlich von den in der Schule Arbeitenden und Lebenden abhängen (vgl. Orth 2005, S. 65).

> „Standards und verpflichtende Leistungsüberprüfungen dürfen somit nicht als zentral installierbare ‚Selbstläufer' zu besserer Schulqualität und höherer Leistung (miss-)verstanden werden. Sie entfalten ihre Wirkung vielmehr nur dann, wenn dadurch Lehrpersonen ‚vor Ort' zu engagierter Reformarbeit animiert werden. Ob dies gelingt, hängt jedoch in starkem Maße von der weiteren (…) Eigenlogik ab, die in der jeweiligen Schule das Handeln anleitet: Bringt es der Schule einen Gewinn, der zentralen Vorgabe zu folgen? Oder ist es sinnvoller, die Forderungen nur formal zu erfüllen, um die eigenen Kräfte woanders einzusetzen? Genau dieser Punkt erweist sich als die entscheidende Klippe für alle zentralen Steuerungsbemühungen" (Tillmann 2006, S. 32).

Altrichter (2008) fasst die Gelingensbedingungen der Implementierung von Reformen auf der System-, Schul- und Unterrichtsebene zusammen. Er stuft die Verfahren und Instrumente nur dann als stützend ein, wenn die Akteure zum einen die normativen Einstellungen und die sachlichen und persönlichen Ressourcen haben, um diese ‚Strukturen' auch kontinuierlich zu praktizieren. Wenn die verschiedenen Akteure im Hinblick auf die zu praktizierenden ‚Strukturen' sich einigermaßen gleichsinnig verhalten und nicht - bewusst oder unbewusst - einander entgegen arbeiten. Als dritten Punkt einer gelingenden Implementierung nennt er die verschiedenen, neu eingeführten Verfahren, die aufeinander abgestimmt sein sollten und einander gleichsam zuarbeiten, und nicht durch interne Widersprüchlichkeit oder durch mangelnde Passung zwischen Verfahren Friktionen erzeugen. Sollen die Reformen langfristig Verbesserungen bringen, dann ist es erforderlich, wenn diese Strukturen institutionalisiert sind, d.h. durch formelle oder informelle Übereinkunft oder durch Routine ‚verstetigt' werden (vgl. ebd., S. 107).

3.3.3 Schwierigkeiten und Probleme der Implementierung

Von der Gathen (2006) formuliert einige Schwierigkeiten, die bei der Implementierung von Lernstandserhebungen auftreten können:

Der erste Punkt betrifft die Qualität der Daten, die teilweise nicht nutzerorientiert aufbereitet sind und dadurch nicht zufriedenstellend rezipiert werden können. In

einem weiteren Punkt bemängelt er die Diskrepanz zwischen Datenwert und Botschaft. Dieses ‚Juxtapositionsverhältnis' beschreibt das unverbundene Verhältnis zwischen wissenschaftlichem Wissen und Handlungswisssen (vgl. Rolff 2002a).

Das Problem der Projektion kann entstehen, wenn die Lehrperson die gewonnenen Daten ausschließlich auf das eigene professionelle Handeln überträgt. Hierbei spielt die emotionale Ebene eine entscheidende Rolle. Falls der von Daten erzeugte Druck auf die Lehrperson zu groß wird, reagiert sie mit Ablehnung. Dies erstickt jede Weiterentwicklung im Keim.

Evaluationskonzepte können die Implementierung neuer Instrumente unterstützen. Denn nur wenn Schulen eine überlegte Strategie einer intensiven Datennutzung entwerfen, kann diese auch zur Unterrichtsentwicklung systematisch herangezogen werden.

Als letzten Punkt kritisiert er, dass viele Rückmeldekonzepte noch zu ‚kurzsichtig' angelegt sind. Tatsächliche Effekte einer datenbasierten Unterrichtsentwicklung entstehen in einem langwierigen Prozess, die ohne Prozessvariablen nicht auskommen. (Vgl. ebd., S. 85ff).

Als problematisch stellt sich oftmals auch der Wissenstransfer innerhalb der Kollegien dar: Die von einzelnen Kollegen auf Fortbildungen erworbenen Kenntnisse und Kompetenzen werden den anderen Kollegen häufig nicht vermittelt. Diskussionen über konkrete Aufgaben und Fehlertypen könnten einen solchen schulinternen Wissenstransfer befördern (vgl. Hosenfeld 2005, S. 114).

Kohler (2005) benennt sieben mögliche Hindernisse, die der Etablierung einer professionellen Evaluationskultur an der Einzelschule im Weg stehen könnten:

1. **Berufsbild und Arbeitszeit:** Lehrerinnen und Lehrern fehlt es zunächst „an der notwendigen Wissensbasis für ein wirkliches Verständnis geeigneter Evaluationsverfahren" (ebd., S.57). Sie wehren diese Verfahren wegen erhöhter Arbeitsbelastung durch Einführung von Evaluationen ab.
2. **Das Fehlen geeigneter Unterstützungssysteme:** Schulen benötigen Unterstützung bei Schulentwicklungsprozessen. Wird diese von externen, außerschulischen Beratern angeboten, so fehlt jenen häufig das pädagogische Vokabular und sie werden dadurch von den schulischen Akteuren nicht ernst genommen. Eine Unterstützung durch die Schulaufsicht wird kritisch gesehen, da zwischen Schulen und

Schulaufsicht ein hierarchisches Verhältnis besteht. Lehrerinnen und Lehrer „können es sich vermutlich eher wenig vorstellen, mit Vertretern der Schulaufsicht in einen konstruktiven Dialog zu treten und sich, von ihnen unterstützt, im Kollegium um Schulentwicklung zu bemühen“ (ebd., S. 58f.).

3. **Probleme mit Kollegium und Schulleitung:** An Schulen herrscht keine ausgeprägte kollegiale Kommunikations- und Kooperationskultur, sie ist eher unterentwickelt. Die Forschung zur Schulentwicklung und Schulqualität sieht jedoch einerseits die kollegiale Kooperation als zentralen Gelingensfaktor von Evaluationen an. Andererseits wird engagierten Schulleiterinnen und Schulleiter eine tragende Rolle bei Schulentwicklungsprozessen beigemessen.
4. **Das Autonomie-Paritäts-Muster:** Dieses Muster beschreibt eine in traditionellen Schulkulturen verbreitete Vorstellung, dass der Unterricht ausschließlich in der Verantwortung der Lehrpersonen fällt und keine andere Person sich darin einmischen darf ("Autonomie"), sowie dass alle Lehrerinnen und Lehrer in ihrer Arbeitsqualität gleich zu behandeln sind ("Parität"). Bezogen auf externe Evaluationsverfahren kann dieses Muster ein potentielles Problem oder eine mögliche „Innovationsbarriere“ darstellen (vgl. Altrichter und Eder 2004), weil Lehrerinnen und Lehrer es gewohnt sind, autonom, nach ihren subjektiven Zielsetzungen und Entscheidungen zu agieren.
5. **Unzureichende Selbstwirksamkeitserwartungen** können bei Lehrerinnen und Lehrern zu einer geringen Handlungsbereitschaft und Einbringung in Schulentwicklungsprozesse führen.
6. **Objektiv hohe berufliche Belastung** und ein hohes subjektives Empfinden dieser Belastung führen zu Verweigerung, sich an Evaluationsprozessen zu beteiligen. „Interne und externe Evaluationen bedeuten ja zunächst einmal ein Mehr an objektiven Anforderungen und eine Zunahme der schulischen Konfliktpotentiale, ohne gleichzeitig Entlastungen bieten zu können“ (Kohler 2005, S. 67).
7. **Attributionsvoreingenommenheiten** könnten zu einer Ablehnungshaltung der Vergleichsarbeiten und anderer Evaluationsverfahren führen. Vor allem negative, unerwünschte Evaluationsergebnisse könnten darin münden, dass sich innerhalb des Kollegiums gegenseitig die Schuld zugewiesen wird.

3.4 Ergebnisse aus der Rezeptionsforschung zur Implementierung von Vergleichsarbeiten

Um zu erfahren, welche Schwierigkeiten und Probleme bei der Implementierung von Vergleichsarbeiten im Bildungssystem tatsächlich auftreten können, benötigt man empirische Forschungsprojekte, die den Prozess der Implementierung durchleuchten. Aus den zahlreichen und vielfältigen Erfahrungen, die man im Zuge der Beteiligung Deutschlands an internationalen und nationalen Schulleistungsstudien gewonnen hat, entwickelte sich ein neues Forschungsfeld, die Rezeptionsforschung. Diese Forschungsrichtung behandelt die Rezeption und Nutzung von Rückmeldungen aus Leistungserhebungen. Die bisher in Deutschland im Rahmen von Schulleistungsstudien realisierten Rückmeldungen unterscheiden sich in ihrer inhaltlichen und methodischen Ausgestaltung teilweise erheblich voneinander.

So richten sich Schulleistungsvergleichsstudien je nach Konzeption und Anlage an unterschiedliche Adressaten im Bildungswesen. Stichprobenbasierte internationale und länderübergreifende Studien dienen dem Systemmonitoring. Sie stellen somit der Makroebene (Bildungsadministration, Bildungspolitik, etc.) Informationen zur Verfügung, aus dem die Verantwortlichen Steuerungswissen generieren können (vgl. Altrichter & Maag Merki 2010; Arnold 2002). Die internationalen Schulleistungsstudien TIMSS, PISA, IGLU oder die länderübergreifende DESI-Studie sind Beispiele hierfür (vgl. van Ackeren 2003; van Ackeren & Klemm 2011). Als Antwort auf TIMSS und PISA wurde von der Kultusministerkonferenz z.B. die Gesamtstrategie der KMK zum Bildungsmonitoring verabschiedet. Die einzelnen Maßnahmen wurden als neue Steuerungselemente sukzessive im Bildungswesen implementiert. Die Ergebnisse aus Rezeptionsstudien zu Schulleistungsstudien auf Ebene des Systemmonitorings stehen in dieser Arbeit nicht im Vordergrund. Eine übersichtliche Darstellung finden sich bei einigen Autoren, z.B. Hartung-Beck (2009), Kohler (2005) oder van Ackeren (2003).

Für die Meso- und Mikroebene des Bildungssystems (Einzelschulen, Unterricht, Lehrerinnen und Lehrer, Schülerinnen und Schüler) lässt sich aus diesen Studien kein direkter Nutzen ableiten. Dazu benötigt man „Schulmonitoring“-Studien, die flächendeckend, in Vollerhebungen Daten ganzer Klassen, bzw. Jahrgänge erhe-

ben. Zu dieser Art von Schulleistungsstudien zählen z.B. die nationalen Schulleistungsstudien, wie LAU, QuaSUM, und MARKUS, sowie alle Formen der Lernstandserhebungen, die in den letzten Jahren in den Bundesländern verpflichtend eingeführt wurden.

Zunächst werden sowohl die Ergebnisse aus verschiedenen Rezeptionsstudien im Bereich des „Schulmonitorings“ vorgestellt (3.4.1), als auch aus zwei weiteren Untersuchungen, die sich mit der Generierung bestimmter Rezeptions- und Nutzungstypen beschäftigen (3.4.2). Im Anschluss daran erfolgt die Vorstellung einiger Modelle, die als theoretische Grundlage für Rezeptionsforschungsdesigns geeignet erscheinen. Das eine benennt alle wesentlichen Grundanforderungen und –bedingungen, die an Rückmeldungen gestellt werden. Das andere beschreibt den idealtypischen mehrstufigen Prozessablauf, der sich an Rückmeldungen von Daten anschließen kann.

3.4.1 Ausgewählte Rezeptionsstudien zum Schulmonitoring

Rezeptionsstudien liefern zum einen Hinweise zur „Gestaltung von Rückmeldungen“, zum anderer erhält man Erkenntnisse über die Rezeption und Nutzung der Ergebnisse der jeweiligen Schulleistungsstudie. Unterschiedliche Autoren kamen zu dem Ergebnis, dass die bloße Bereitstellung von Rückmeldedaten in den Schulen nicht automatisch zu Veränderungen im Sinne von Verbesserungen führt , sondern dass zahlreiche Faktoren bei der Rezeption und Nutzung von Rückmeldungen eine Rolle spielen (vgl. Kuper und Schneewind 2006; Kohler 2005; Rolff 2002b; Schneewind 2006, 2007; Terhart 2002; Tresch 2007; van Ackeren 2003; Watermann & Stanat 2004). In Kapitel 3.3 „Implementierung von Vergleichsarbeiten“ wurde aufgezeigt, dass die Rezeption der Ergebnisrückmeldungen einen vielschichtigen Prozess darstellt, in den mehrere Ebenen des Bildungssystems involviert sind. Wie diese Ebenen zusammenspielen und wie die daran beteiligten Akteure bei der Umsetzung der Maßnahme agieren, kann durch diese Rezeptionsstudien evaluiert werden.

Mittlerweile liegen zu einigen nationalen, flächendeckend durchgeführten Studien Rezeptionsbefunde vor, deren Ergebnisse im Folgenden kurz vorgestellt werden.

Um einen Überblick über die Zuständigkeit, Zielsetzung und erfassten Kompetenzbereiche der jeweiligen Studie zu erhalten, sind diese in Tabelle 2 dargestellt. Die Beschreibung der Studien erfolgt in zeitlicher Reihenfolge, beginnend mit den ersten Studien im Bereich des Schulmonitorings, die seit Mitte der 1990er Jahre in Deutschland durchgeführt wurden: LAU, QuaSUM und MARKUS. Außerdem werden drei länderspezifische Lernstandserhebungen vorgestellt; Kompetenztests in Thüringen, Lernstandserhebungen in Nordrhein-Westfalen und zuletzt Vergleichsarbeiten in Baden-Württemberg.

Im Anschluss an die Tabelle, die sowohl einen knappen Überblick über die Studien als auch über die Rezeptionsstudien bietet, werden die Ergebnisse aller Rezeptionsstudien und deren Ergebnisse systematisch zusammenfassend präsentiert.

Tabelle 3.4.1: Überblick über zentrale nationale Vergleichsstudien zum „Schulmonitoring“ in Anlehnung an van Ackeren (2003, S. 52

Name/Akronym, Bundesland, Jahr, Altersstufe	Bezeichnung der Studie	Zuständigkeit für Durchführung	Erhobene Kompetenzen	Ziele der Schulleistungsstudie	Rezeptionsstudie Erhebungsverfahren & -methoden
LAU 5-13[12] Hamburg 1996-2005 5., 7., 9., 11. und 13. Klassen	„Aspekte der Lernausgangslage und der Lernentwicklung“ (Längsschnittstudie)	Hamburger Schulbehörde Humboldt-Universität Berlin	Deutsch, Mathematik, 1. Fremdsprache Fachübergreifende Kompetenzen	Informationen zur Einstellung zu Schule und Unterricht: Leistungsvermögen, Kooperationsfähigkeit, soz. Integration	a) Leitfadeninterviews mit Schulleitungen an acht Hamburger Gymnasien (Klug & Reh (2000), b) Maritzen 2001
QuaSUM[13] Brandenburg 1999 5. und 9. Klassen	„Qualitätsuntersuchung an Schulen zum Unterricht in Mathematik“	Ministerium für Bildung, Jugend und Sport Brandenburg Humboldt-Universität	Mathematik	Informationen zu: Unterrichtsmerkmalen, Schulalltag Lernumwelt	QuaSUM 2 und 3: Fragebogenstudie an beteiligten Schulen und qualitative Fallstudie an sechs Gymnasien
MARKUS[14] Rheinland-Pfalz 2000 8. Klassen	„Mathematik-Gesamterhebung Rheinland-Pfalz: Kompetenzen, Unterrichtsmerkmale, Schulkontext	Ministerium für Bildung, Frauen und Jugend Rheinland-Pfalz, Universität Koblenz-Landau	Mathematik	Mathematikleistungen, Unterrichtsbedingungen, Qualität von Unterricht, Einstellungen der SuS, Zusammenhänge	WALZER-Studie[15] Befragung der Schulleitungen und Mathematiklehrkräfte.
Kompetenztests Thüringen[16] Seit 2002/03 3., 6. und 8. Klassen		Kultusministerium Thüringen; ThILLM[17]. Projekt „kompetenztest.de“ der Uni Jena	Deutsch Mathematik & Englisch (ab Klasse 6)	Diagnostische Daten zur individuellen Förderung und Weiterentwicklung des Unterrichts	Internetbefragung der Schulleitungen bzgl. Rezeption und Akzeptanz, ab 2007 zusätzlich Fachlehrpersonen [18]
Lernstand 9 bzw. 8[19] Nordrhein-Westfalen Seit 2004/05 9., dann 8. Klassen		Schulministerium Nordrhein-Westfalen in Zusammenarbeit mit Landesinstitut für Schule	Deutsch Mathematik 1. Fremdsprache	Vergleichende Überprüfung der Standards zur pädagogischen und didaktischen Nutzung	3 Online-Befragungen zu Rezeption und Durchführung
Vergleichsarbeiten Baden-Württemberg Seit 2005/06 6. und 8. Klassen,(später 7.,9.,11.)		Ministerium für Kultus, Jugend und Sport in Zusammenarbeit mit dem Landesinstitut für Schulentwicklung	Deutsch, Mathematik, 1. Fremdsprache und ausgewählte Fächer (siehe Kapitel 3.2)	Qualitätsentwicklung der Schule, Feststellung des Lernstandes der SuS	Quantitative Befragung der Lehrpersonen 2006 und 2007 zur Rezeption und Nutzung von VA (Maier 2009b)

[12] http://www.arge.schule-hamburg.de/Inhalt/LAU.html
[13] http://www.mbjs.brandenburg.de/media_fast/5527/Schulforschung_Heft1_quasum.pdf
[14] http://www.uni-koblenz-landau.de/landau/fb8/entwicklungspsychologie/Projekte/markus
[15] **W**irkungs**a**nalyse der **L**eistungsevaluation: **Z**ielerreichung, **E**rtrag für die Bildungsqualität der Schule und die **R**ückmeldung von Evaluationsergebnissen
[16] http://www.thueringen.de/de/tmbwk/bildung/informationen/vergleiche/kompetenztest/content.html
[17] ThILLM – Thüringer Institut für Lehrplanentwicklung, Lehrerfortbildung und Medien
[18] Ergebnisse zu Thüringer Kompetenztests unter *www.kompetenztests.de*
[19] http://www.standardsicherung.schulministerium.nrw.de/lernstand8/aktuelles/

3.4.1.1 Zur Methodik der ausgewählten Rezeptionsstudien

LAU 7

Die Rezeptionstudie zur LAU 7 - Studie wurden von Klug und Reh (2000) durchgeführt und beschrieben. Um Erkenntnisse zum Rezeptionsverhalten von Schulen und Lehrerinnen und Lehrern zu erhalten, führten sie Leitfadeninterviews mit den Schulleitungen an acht Hamburger Gymnasien durch.

QuaSUM 2 und QuaSUM 3

Zur QuaSUM-Studie[20] (vgl. Lehmann et al. 1999; Ministerium für Schule Jugend und Sport des Landes Brandenburg 2004), die die erreichten Lernstände von brandenburgischen Fünft- und Neuntklässlern in Mathematik erhob, wurden zwei Rezeptionsstudien durchgeführt, QuaSUM 2 und QuaSUM 3.

An QuaSUM 2, einer Fragebogenstudie, beteiligten sich ca.74 % der Schulen, wobei sich die Anteile der einzelnen Gruppen (Mathematiklehrpersonen, Klassenlehrerinnen und –lehrer, Fachkonferenzleitung und Schulleitung) unterscheiden (vgl. Peek 2004a). Im Vordergrund dieser Rezeptionsstudie stand die Fragestellung, wie sich Schulen mit externen Evaluationsdaten unter formalen, thematischen und handlungsorientierten Gesichtspunkten auseinandersetzen und wie sie diese für interne Evaluationsprozesse verwenden. Untersuchte Aspekte waren: 1) Einstellung der Lehrpersonen und Schulleitung zu externen Rückmeldeverfahren; 2) Auseinandersetzung mit Ergebnissen; 3) Intensität der Nutzung und 4) außerschulische Auseinandersetzung.

Bei der QuaSUM 3 - Studie (vgl. Nilshon 2004) handelt es sich um eine qualitative Fallstudie, an der sich unterschiedliche Akteure aus sechs Gymnasien beteiligten. Die Daten wurden mittels Gruppengesprächen erhoben und mit der verstehenden Diskurs-Methode ausgewertet. Die Zielsetzung der Studie betrafen folgende Punkte: 1) Was verbinden die Pädagoginnen und Pädagogen der ausge-

[20] Zur Durchführung der QuaSUM-Studie sowie deren Rezeption liegt mittlerweile eine Vielzahl an Veröffentlichungen vor. Eine Publikationsliste findet sich bei Institut für Qualitätsentwicklung im Bildungswesen (IQB) (2010b).

wählten Gymnasien mit dem QuaSUM-Projekt; 2) Welche Beziehung besteht zwischen den erhaltenen Rückmeldungen und den auf die Entwicklung der eigenen Schule bezogenen Erwartungen; 3) Welche Maßnahmen wurden ergriffen.

MARKUS- und WALZER

Die Rezeption der MARKUS-Studie erfolgte mittels einer von Andreas Helmke initiierten Studie namens WALZER (**W**irkungs**a**nalyse der **L**eistungsevaluation: **Z**ielerreichung, **E**rtrag für die Bildungsqualität der Schule und die **R**ückmeldung von Evaluationsergebnissen), die von 2001-2002 durchgeführt wurde (vgl. Schrader & Helmke 2004). Als Modell der Rezeptionsstudie wurde das „Rahmenmodell für die Bedingungen erfolgreicher Rezeption von Evaluationsergebnissen“ zugrunde gelegt (vgl. Hosenfeld et al. 2006, S. 296). Die leitenden Fragestellungen der Studie richteten sich an die Schulleitungen und die Mathematiklehrpersonen und betrafen zum einen die „Einsichten und Erfahrungen“ die aus WALZER gewonnen wurden. Zum anderen wurden Probleme und Konsequenzen erörtert, die für die Gestaltung künftiger Rückmeldungen wichtig sein könnten.

Thüringer Kompetenztests

Die jährlichen Lernstandserhebungen bzw. Vergleichsarbeiten in Thüringen nennt man Kompetenztests. Diese gehen über einen reinen Leistungsvergleich der Schülerinnen und Schüler hinaus. Mit ihrer Hilfe soll rechtzeitig Förderbedarf erkannt werden, um entsprechende Fördermaßnahmen durchzuführen. Des Weiteren sind sie in ein breites Fortbildungs- und Weiterbildungskonzept eingebettet (vgl. Arbeitsgruppe Kompetenztests 2010; Maier 2008b).

Die Rückmeldung der Ergebnisse erfolgt in Thüringen stufenweise. Sofort nach der Dateneingabe erhalten die Lehrpersonen eine erste Rückmeldung. Daran schließt sich nach ca. acht Wochen ein Klassen- und Schulbericht an. Den Abschluss bildet der Landesbericht, der nach einem Vierteljahr vorliegt. Die Rückmeldeformate unterscheiden sich je nach Adressat der Rückmeldung. Während die Lehrpersonen und Schulleiter Einblick in die vollständigen Ergebnisse erhalten, werden auf administrativer Ebene anonymisierte Daten rückgemeldet (vgl. Arbeitsgruppe Kompetenztests 2010).

Im Rahmen der Rezeptionsstudie von Nachtigall und Jantowski (2007) wurden in den Jahren 2006 und 2007 die Schulleitungen in einer quantitativen Studie via In-

ternet auf ihre Rezeption und Akzeptanz der Kompetenztests befragt. Diese Stichprobe wurde im Jahr 2007 noch um die Gruppe der Fachlehrpersonen erweitert (vgl. ebd., S. 408f). Eine ausführliche Ergebnisdarstellung findet sich sowohl bei Nachtigall (2007) als auch in Form der Landesberichte auf der Internetseite www.kompetenztest.de.

Lernstandserhebungen in Nordrhein-Westfalen

Im Bereich der Schulleistungsforschung liegen im Bundesland Nordrhein-Westfalen zahlreiche Forschungsprojekte vor, die sich mit der Entwicklung von Qualität in Schule und Unterricht befassen (vgl. Bonsen et al. 2004b; Bos 2008; Dobbelstein & Peek 2008; Landesinstitut für Schule und Weiterbildung Nordrhein-Westfalen 1995).

Im Bereich der Lernstandserhebungen Nordrhein-Westfalen zeichnen sich die Lehrpersonen für die Durchführung, Eingabe und Auswertung der Instrumente mit Hilfe internetbasierter Auswertungsmasken verantwortlich (vgl. Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen 2011).

Die Ergebnisse der Rezeptionsstudie stammen aus drei Online-Befragungswellen mit freiwilliger Teilnahme: Befragung I (Juni 2005) und II (September 2005) wurden im Kontext der ersten Lernstandserhebung mit unterschiedlicher Schwerpunktsetzung durchgeführt. So standen bei den ersten Befragungen die Rezeption der Datenrückmeldungen und die Beurteilung der Test- und Auswertungspraxis im Vordergrund. Die Befragung III Ende des Schuljahres 2005/2006 beschäftigte sich nochmals mit spezifischen Fragen zur Testdurchführung und Rezeption. Für die erste und zweite Erhebung lagen 2670, für die Dritte 768 Fragebögen vor. Die Ergebnisse aus den Untersuchungen findet man detailliert bei Kühle und Peek (2007).

Vergleichsarbeiten in Baden-Württemberg

Die Konzeption und Zielsetzung der baden-württembergischen Vergleichsarbeiten wurde in Kapitel 3.2.2 ausführlich dargestellt. Seit der verbindlichen Durchführung der Vergleichsarbeiten im Schuljahr 2005/2006 wurde die Implementierung dieses Instruments an der Pädagogischen Hochschule Schwäbisch Gmünd von Uwe Maier empirisch begleitet. Von ihm liegen Ergebnisberichte u. a. aus den ersten beiden Jahrgängen vor, die mit Hilfe quantitativer (2006 und 2007) Verfah-

ren erhoben wurden. Zusätzlich wurden nach der ersten Durchführung im Jahre 2006 auch qualitative Erhebungsverfahren in Grund-, Haupt- und Realschulen angewandt (vgl. Maier 2008a; Maier 2009a; Maier 2009b, Maier 2010).

Die Ergebnisse aus diesen sechs Beispielen für Rezeptionsstudien im Bereich des Schulmonitorings lassen sich nach folgenden Kriterien zusammenfassen: Zunächst werden die Befunde zu den Einstellungen und Haltungen der Lehrerinnen und Lehrer beschrieben, die diese im Zuge der Durchführung der Studien entwickeln konnten (3.4.1.2). In einem zweiten Schritt werden die Ergebnisse aus den Studien berichtet, wie dieser Personenkreis die zurückgemeldeten Daten rezipiert und nutzt (3.4.1.3). Daran schließt sich die Vorstellung der Erkenntnisse an, wie Rückmeldungen gestaltet sein sollten, um einen höheren Nutzen zu erzeugen (3.4.1.4). Im letzten Ergebniskapitel werden die kooperativen und kommunikativen Strukturen der Einzelschule beleuchtet und deren Auswirkungen auf die Verwendung des Evaluationswissens (3.4.1.5).

3.4.1.2 Ergebnisse der Rezeptionsstudien

3.4.1.2.1 Einstellung und Haltung der Lehrpersonen zu Schulmonitoringstudien

Die verschiedenen Rezeptionsstudien liefern im Bereich der Einstellungen und Haltung der Lehrerinnen und Lehrer zu den jeweiligen Schulmonitoring-Studien teilweise unterschiedliche Ergebnisse. Die Autoren der Rezeptionsstudie zu LAU-7 berichten z.B., dass das Interesse der Lehrerinnen und Lehrer an den Rückmeldungen zur LAU-7-Studie nicht besonders groß ausfiel. Als einen möglichen Grund hierfür wird diskutiert, dass die Schülerinnen und Schüler besser abschnitten als erwartet. Bei zu guten Ergebnissen geben die Lehrpersonen an, nichts Neues zu erfahren. So wurde auch keine Notwendigkeit gesehen, Konsequenzen aus den Rückmeldungen abzuleiten (vgl. Klug & Reh 2000).

In Bezug auf die Einstellung und Haltung der Lehrkräfte fallen die Ergebnisse bei QuaSUM 2 etwas differenzierter aus. Dort wurde berichtet, dass die Lehrpersonen und die Schulleitungen standardisierten Schulleistungsstudien gespalten gegenüber stehen. Das breite Antwortspektrum der Befragten reicht von Befürwortung

bis zu Ablehnung. In der Summe wird die Rückmeldung eher positiv angenommen. Insgesamt weisen alle Gruppen eine hohe Bereitschaft auf, sich mit den zurückgemeldeten Daten auseinanderzusetzen. Dies wird von den Autoren als Akzeptanz für externe Evaluationsverfahren gewertet (vgl. Peek 2004a, S. 80ff).

In den Gruppengesprächen, die im Rahmen der QuaSUM 3-Studie geführt wurden, zeigten sich die Lehrerinnen und Lehrer vor allem über die Verwendung der Rückmeldedaten enttäuscht. Sie gaben an, dass ihre Erwartungen im Sinne einer Ableitung von Konsequenzen und Maßnahmen zur Qualitätsverbesserung nicht erfüllt wurden (vgl. Nilshon 2004).

Diese kritische Haltung teilen Thüringer Lehrerinnen und Lehrer nicht. Im Gegenteil - dort stufen 90 % aller Schulen die Tests inklusiv ihrer Begleitinstrumente und der Rückmeldeverfahren als sehr positiv ein. Die Akzeptanz der Lehrerkräfte gegenüber Kompetenztests als auch gegenüber externen Evaluationsverfahren in Thüringen ist gegeben (vgl. Nachtigall & Jantowski 2007).

Auch die Lehrpersonen in Nordrhein-Westfalen teilen diese positive Einstellung. Sie verknüpfen positive Erwartungen mit Lernstandserhebungen im Sinne einer Bestandsaufnahme von Stärken und Schwächen ihrer Schülerinnen und Schüler, als Anregung für die Unterrichtsentwicklung und als Verbesserung der schulinternen Kommunikation. Die Ergebnisse eröffnen demnach Perspektiven für die Förderung, die Schul- und Unterrichtsentwicklung sowie für das Systemmonitoring.

Bezüglich der Einstellung und Haltung der Lehrpersonen im Umgang mit baden-württembergischen Vergleichsarbeiten skizziert Maier (2009b) ein differenziertes, schulartspezifisches Bild. Hauptschullehrkräfte akzeptieren zu Beginn die Vergleichsarbeiten eher als ihre kritischeren Kolleginnen und Kollegen an Realschulen und Gymnasien. Sie fühlen sich kaum belastet und betonen die Übereinstimmung der Testaufgaben mit Bildungsstandards. Im Laufe der Zeit stimmen auch die Lehrpersonen der Realschulen und Gymnasien dieser curricularen Validität zu. Die Hauptschullehrerinnen und –lehrer entwickeln zunehmend eine kritischere Einstellung zu Vergleichsarbeiten.

Vergleichsarbeiten, als externes Kontrollinstrument, werden in den beiden Jahren zunehmend als Belastung wahrgenommen, weil dadurch z.B. zusätzlich Druck auf Schulen und ihre Akteure ausgeübt wird. Dafür stimmt im Jahr 2007 jede zweite Lehrperson.

3.4.1.2.2 Ergebnisse zur Rezeption und Nutzung der Daten

Eine allgemeine Erkenntnis zur Rezeption und Nutzung von Daten aus Vergleichsstudien lässt sich aus der WALZER-Studie ableiten: Um Rückmeldungen rezipieren zu können, sollten die Lehrerinnen und Lehrer sowohl über die notwendigen technischen Voraussetzungen, als auch über die erforderlichen Fähigkeiten, Fertigkeiten und Kompetenzen verfügen. Ein Ergebnis dieser Studie lautet, dass nicht alle Lehrerinnen und Lehrer z.B. die erforderlichen Kenntnisse im Umgang mit dem Computer besitzen.

Bei der Rezeption der Lernstandserhebungen werden ebenfalls Probleme berichtet. Ein Großteil der Lehrpersonen hat Verständnisschwierigkeiten bei der Interpretation der Ergebnisse, die neben den Angaben zur statistischen Signifikanz auch die Referenzwerte „Schulform“ und „Schultyp“ betreffen.

Des Weiteren fand man durch diese Studie heraus, dass die Akzeptanz und grundsätzliche Einstellung der Lehrpersonen zu externen Evaluationsverfahren die Rezeption und Nutzung mehr bestimmen, als die Testergebnisse, die die Schülerinnen und Schüler in den Lernstandserhebungen erreichen.

Aus LAU 7 konnte abgeleitet werden, dass es für eine Nutzung externer Daten der Akzeptanz der Akteurinnen und Akteure bedarf, die gewonnenen Daten als relevant anzusehen. Wenn Rückmeldungen genutzt werden, dann zur Bestätigung bereits angestoßener schulischer Projekte, jedoch nicht zur Ableitung neuer Maßnahmen. Um aus LAU methodische und didaktische Schlussfolgerungen ziehen zu können, reicht die Datenqualität nicht aus. Hierfür benötige man zusätzliche fallorientierte Unterrichts- und Schulforschung. (Vgl. Klug & Reh 2000, S. 17ff.).

Die Auseinandersetzung mit den Ergebnissen aus QuaSUM erfolgt primär fachbezogen. Laut QuaSUM 2 setzen sich sowohl die Mathematiklehrpersonen als auch die Lehrerinnen und Lehrer der Fachkonferenz am intensivsten damit aus-

einander. Nur wenige Personen leiten aus den Rückmeldungen Konsequenzen für ihre pädagogische Arbeit ab.

Bei QuaSUM 3 berichten vier der sechs Gymnasien, dass sie sich durch die Ergebnisse in ihrer Arbeit bestätigt sehen. Sie beklagen den hohen zeitlichen Aufwand, der ihrer Meinung nach in keinem Verhältnis zur Nutzung steht. Nach ihren Angaben haben die Rückmeldungen keinen direkten Einfluss auf die Schulentwicklungsprozesse. Die Untersuchung ergab auch, dass die Ableitung von Maßnahmen und Konsequenzen aus Schulleistungsstudien und deren Rückmeldungen eher vom Professionalisierungsgrad der Akteure abhängt, als vom Leistungsstand der jeweiligen Klasse. Den Professionalisierungsgrad bestimmt die Autorin durch folgende Faktoren: a) Kooperative, kollegiale, regelmäßige und reflexive Zusammenarbeit der Beteiligten; b) Grad der deligierten Verantwortung (Welche aktive, steuernde Funktionen wird den Lehrerinnen und Lehrer innerhalb des Kollegiums zugesprochen). Dies ist von der jeweiligen Schul- und Lernkultur der Schule mitbestimmt (vgl. Nilshon 2004).

Die WALZER-Studie lieferte als ein Ergebnis zur Nutzung von MARKUS, dass eine größere zeitliche Verzögerung zwischen Untersuchung und Ergebnisrückmeldung ein Problem darstellen kann. Entsteht ein großer zeitlicher Abstand zwischen Durchführung der Studie und der Rückmeldung der Ergebnisse, so verringert sich der Nutzen für die schulischen Akteure erheblich, weil sie nur noch in sehr beschränktem Umfang in unterrichtliche Handlungen einfließen können. Die Autoren schlagen vor, dass in geeigneten Bereichen (z.B. Bewertung des Materials) die Bewertung bereits mit der Durchführung erhoben wird. Im Anschluss daran können dann weitere Punkte evaluiert werden.

Ein weiteres Problem der Ableitung von Konsequenzen für die Veränderung des Unterrichts stellt die Tatsache dar, dass Schulleistungen ein Produkt zahlreicher Bedingungsfaktoren darstellen. Welche Faktoren jeweils verändert werden müssen, lässt sich teilweise aus den Rückmeldungen nicht ableiten. Je differenzierter die Daten erhoben werden, desto eindeutiger kann der jeweilige Adressat bestimmt werden.

Die Thüringer Kompetenzstests werden primär zur Schulleistungsdiagnostik und als Leistungsvergleichsmöglichkeit verwendet. Je stärker die Rückmeldungen

Hinweise für die einzelne Schule enthalten bzw. diese sich auf konkrete unterrichtliche Leistungen der Schülerschaft beziehen lassen, desto intensiver fällt die Auseinandersetzung mit den Daten aus.

Außerdem liefert die Rezeption zu den Thüringer Kompetenztests den Befund, dass ein vertraulicher Datenumgang und die Schwerpunktlegung auf interne gegenüber externe Evaluationsverfahren weitere unterstützende Momente für die Nutzung darstellen.

Bei den nordrhein-westfälischen Lernstandserhebungen ist die Bereitschaft der Lehrpersonen besonders hoch, die Ergebnisse zum Beispiel für die Unterrichtsentwicklung zu nutzen (75 %). Fragt man jedoch nach konkreten Umsetzungsmöglichkeiten, so unterscheidet sich die Bereitschaft signifikant zwischen den Funktionsträgern. Während knapp die Hälfte der Schulleitungen motiviert sind konkrete Konsequenzen abzuleiten, teilen diese Motivation nur ein Viertel der Lehrerinnen und Lehrer.

Auf die Frage, welche Konsequenzen auf die Ergebnisrückmeldungen folgten und welche Maßnahmen umgesetzt wurden, gaben die Befragten folgende vier Bereiche an: Konzentration auf Testinhalte bzw. –aufgaben, intensivierte Förderung durch zusätzlichen Unterricht, der sich z.B. an bestimmten Kompetenzbereiche oder an bestimmten Schülerinnen und Schülern ausrichtet. Des Weiteren wurden die Erhebung zusätzlicher Informationen und zuletzt eine veränderte Unterrichtsmethodik bzw. –gestaltung genannt.

In Baden-Württemberg stehen vor allem die Gymnasial- und Realschullehrenden einer möglichen Nutzung sehr kritisch gegenüber, Hauptschullehrerinnen und –lehrer nähern sich im zweiten Jahr dieser reservierten Einstellung an. Dass Vergleichsarbeiten diagnostische Hinweise liefern, wird lediglich von 30% der Befragten genannt. Nur ein Viertel der Stichprobe gibt an, dass die Ergebnisse für Elterngespräche verwendet werden.

3.4.1.2.3 Ergebnisse zur Gestaltung von Rückmeldungen

Einige Studien liefern Erkenntnisse zur Gestaltung von Rückmeldungen und der damit in Verbindung stehenden Nutzung. Ein Problem kann der verwendete Vergleichsmaßstab darstellen. Im Projekt MARKUS wurden soziale Vergleichsmaß-

stäbe zugrunde gelegt, die über den Klassen- und Schulrahmen hinaus wichtige Informationen bezüglich der Leistungsverteilung der jeweiligen Klasse liefern können. Allerdings lassen sich daraus keine Ansatzpunkte für Änderungsmaßnahmen ableiten. Hierfür wären kriteriale Bezugsnormen erforderlich. Ein weiteres Problem betrifft die fehlende Fokussierung auf Einzelwerte: Die Lehrerinnen und Lehrer erhielten bei MARKUS lediglich den Mittelwert der Klasse zurückgemeldet und somit keine Auskunft über die Leistungsheterogenität.

Im Zusammenhang mit der Berücksichtigung des Kontextes steht ein weiterer Befund: Obwohl im Projekt MARKUS Kontextmerkmale berücksichtigt wurden (Kulturbesitz im Elternhaus, elterliche Lernunterstützung etc.) um einen fairen Vergleich zu ermöglichen, empfanden die Lehrpersonen mit günstigen Kontextbedingungen die Adjustierung nach unten ungerecht. Die Autoren schlagen für zukünftige Vergleichsstudien die Berechnung von Erwartungswerten oder die Bildung von Kontextgruppen vor.

3.4.1.2.4 Ergebnisse zur innerschulischen Kommunikation und Kooperation

Die Schulleitungen bewerten die Rückmeldungen zu LAU 7 u. a. als nützlich, weil dadurch die Diskussionen innerhalb der Kollegien in Gang gesetzt wurden und ein Austausch über Ziele, Inhalte, Methoden und die Wirksamkeit von Unterricht stattfindet. Allerdings werden die Ergebnisse aus Datenrückmeldungen in Fachkonferenzen kaum diskutiert. Ein weiterer Befund aus LAU 7 ergab, dass je auffälliger die Leistungsunterschiede zwischen den Parallelklassen ausgeprägt sind, desto intensiver finden die Auseinandersetzungen zwischen den beteiligten Lehrpersonen aus.

Eine weitere Studie zur Nutzung der LAU- Rückmeldedaten stellt die von Maritzen (2001) dar. Er stellt ebenfalls fest, dass bei Leistungsdifferenzen innerhalb der Schülerschaft teilweise innerschulische Maßnahmen angeregt wurden. Dies erfolgte im Sinne einer systematischen Nutzung der Daten sowohl für die Klassenbildung als auch für eine verstärkte Kooperation.

Ergebnisse aus Thüringen bestätigen die zentrale Rolle der Schulleitungen bei der Nutzung von Evaluationswissen. Dieser kommt bei der Implementierung und

Nutzung von Kompetenztests für die Schul- und Qualitätsentwicklung eine Schlüsselrolle zu. Sie sind „die Steuerzentralen für Schulentwicklung und beeinflussen mit ihrer Haltung gravierend das Verhalten der Lehrpersonen der Schule“ (Nachtigall & Jantowski 2007, S. 409). Die Autoren sehen sowohl in der verstärkten Zusammenarbeit mit der Schulleitung als auch in Fortbildungsangeboten für Fachlehrer Möglichkeiten zur Erhöhung der Akzeptanz der Kompetenztests.

Zu ähnlichen Befunden bezüglich kooperativer Strukturen kommen die Evaluatoren der Lernstandserhebungen in NRW. Verfügen Schulen über eine gemeinsame fächerübergreifende Auswertungsstrategie, dann werden die Ergebnisrückmeldungen stärker rezipiert. An diesen Schulen haben sich die einzelnen Akteure auch über gemeinsame Ziele und Entwicklungsaufgaben verständigt. Teilweise gibt es dort auch Qualitäts- oder Evaluationsteams. Daraus kann man ableiten, dass eine bestehende gute Kooperations- und Evaluationskultur bei der Rezeption und Nutzung externer Daten eine entscheidende Rolle spielt (vgl. Kühle & Peek 2007).

Die WALZER-Studie liefert im Bereich Kooperation die Erkenntnis, dass viele Lehrpersonen nicht in kooperativen Strukturen zusammenarbeiten, sondern Einzelkämpfer mit externen Rückmeldungen nicht vertraut sind. „Es ist nicht verwunderlich, dass in dieser Situation Ängste und Abwehrhaltungen aufkamen, die einer Kooperation im Kollegium entgegenstanden“ (Hosenfeld et al. 2006, S. 300).

In Baden-Württemberg findet die Kommunikation über die Ergebnisse aus Vergleichsarbeiten nur in wenigen Schulen im kollegialen Austausch statt. Informiert werden dort vor allem die Schulleitung und die Lehrenden, die in Parallelklassen Vergleichsarbeiten durchführten. In innerschulischen Gremien werden die Ergebnisse nur wenig ausgetauscht (20 – 25%). Allenfalls in Fachkonferenzen werden systematische Diskussionen geführt. Die Weitergabe der erzielten Ergebnisse der Schülerschaft werden nur von einem Viertel der Befragten an die nachfolgenden Kolleginnen und Kollegen weitergeleitetet.

In Thüringen finden weder Diskussion der Ergebnisse aus den Kompetenztests mit außerschulischen Beratern noch schulübergreifend statt. Bei QuaSUM waren die Eltern kaum in den Rückmeldeprozess eingebunden.

3.4.2 Befunde zu Rezeptionstypen

Neben den zahlreichen Befunden zur Rezeption und Nutzung von Rückmeldungen aus Studien zum Schulmonitoring liegen auch Untersuchungen vor, die in der Gruppe der Lehrpersonen bestimmte Nutzungstypen generieren. Im Folgenden sollen zwei empirische Forschungsprojekte vorgestellt werden, die mittels Typen bildender Verfahren, Lehrerinnen und Lehrer jeweils spezifischen Nutzungsprofilen zuordnen. Beim ersten Beispiel handelt es sich um eine Untersuchung, die im Rahmen von VERA 2004 durchgeführt wurde. Ein Schweizer Projekt, das die Daten aus PISA zur Typengenerierung verwendete, stellt das zweite Beispiel dar.

3.4.2.1 Rezeptionstypen im Rahmen der VERA-Studie

Die Autoren Groß Ophoff, Hosenfeld und Koch (2007) stellen in ihrer Rezeptionsstudie im Anschluss an VERA 2004 unterschiedliche Nutzungsprofile vor, die sich in der Auseinandersetzung mit den Ergebnisrückmeldungen voneinander unterscheiden. Ihr methodisches Vorgehen lehnt sich an das Modell von Helmke & Hosenfeld (2004) an, das in Kapitel 3.4.2.2 näher erläutert wird.

Als Grundlage der Typenbildung wurden zum einen die Antworten verwendet, die sich auf die Fragenbereiche Verständlichkeit, Intensität der Auseinandersetzung und Nützlichkeit der Rückmeldedaten beziehen. Es werden drei verschiedene Rezeptionstypen beschrieben: Rezeptionstyp 1, Rezeptionstyp 2 und Rezeptionstyp 3. Zum anderen wird untersucht, ob sich diese Typen in der Ableitung spezifischer Maßnahmen im Bereich der Unterrichtsentwicklung unterscheiden. Zu diesen Maßnahmen zählen u. a. Wiederholungen bzw. Vertiefungen des Stoffgebietes, Übungen zu bestimmten Aufgabentypen, Elterngespräche mit Schwerpunkt Fördermöglichkeiten, Aufgaben die sich am VERA-Format orientieren, Entwicklung bzw. Überarbeitung eigener Unterrichtsmaterialien und Entwicklung von an VERA angelehnten Übungsaufgaben. Es folgt die Beschreibung der drei Typen sowohl Im Hinblick auf die Auseinandersetzung mit den Daten als auch die Ableitung typenspezifischer Maßnahmen im Sinne der Qualitätssicherung.

Rezeptiontyp 1:

Dieser Typus setzt sich vor allem mit den individuellen und schulklassenspezifischen Ergebnissen auseinander, die sie als verständlicher bewerten, als die Rückmeldungen auf Landesebene, bzw. die adjustierten Werte. Im Vergleich zu den anderen beiden Rezeptionstypen fällt zum einen die intensive, positive Auseinandersetzung mit den Daten auf. Zum anderen bewertet dieser Typus die Verständlichkeit der Rückmeldungen als besonders hoch.

Auswirkungen auf Unterrichtsentwicklung: Dieser Personenkreis zeigt sich in allen oben genannten Bereichen sehr aktiv und verwendet jede der oben aufgeführten Maßnahmen.

Rezeptiontyp 2:

Im Vergleich zum ersten Typus setzt sich dieser nicht so intensiv mit den zurückgemeldeten Daten auseinander. Beide Typen schätzen die Nützlichkeit der Daten ähnlich hoch ein. Nur in der Bewertung der Verständlichkeit unterscheidet sich Typ 2 von Typ 1, weil er diese nicht so hoch einstuft wie Typ 1. Über alle drei Bereiche hinweg (Intensität, Verständlichkeit und Nützlichkeit) fällt deren Beurteilung durch Typ 2 homogen aus, ohne extreme Abweichungen nach oben oder unten.

Auswirkungen auf Unterrichtsentwicklung: Dieser Typ ist nicht so aktiv wie der erste. Er wendet vor allem Unterrichtsmethoden an, die der Wiederholung und Einübung dienen.

Rezeptiontyp 3:

Typ 3 setzt sich intensiv mit den Daten auseinander. Bezüglich der Verständlichkeit und der Datennutzung fällt die Intensität im Vergleich zu den beiden anderen Typen geringer aus. Die Ergebnisse aus VERA werden als kaum nützlich bewertet. Die Autoren interpretieren diesen Typ als Personen mit eher kritischer Einstellung zu den Vergleichsarbeiten.

Die Auswirkungen auf die Unterrichtsentwicklung fallen bei diesem dritten Rezeptionstypen eher gering aus. Er leitet aus den Rückmeldungen selten Maßnahmen ab, was nach Ansicht der Autoren als Indiz für die kritische, zurückhaltende Einstellung interpretiert werden könnte.

3.4.2.2 Typologie der Nutzung von Evaluationswissen nach Margrit Stamm (2003)

Die Schweizer Erziehungswissenschaftlerin der Universität Fribourg, Margrit Stamm, setzt sich in ihrem Forschungsprojekt mit der Frage auseinander, ob „der forcierte Einsatz von Evaluation automatisch zur Qualitätssteigerung von Bildungsleistungen und Bildungangeboten“ führt (Stamm 2005, S. 2; 2003b). Sie diskutiert am Beispiel von PISA, ob sich durch die Nutzung von Evaluation eher eine ‚grande idée’ verwirklichen lässt oder ob sich gar eine ‚grande illusion’ dahinter verbirgt (vgl. Stamm 2003a). Bei ihrer Untersuchung richtet sie ihren Fokus auf alle beteiligten Akteure und deren unterschiedlicher Verwendungsinteressen von Evaluationen: So fragen z.B. Schulbehörden nach der Tauglichkeit der Evaluationsverfahren, während Akteure der Einzelschule an der Verwirklichung ihrer pädagogischen Vorstellungen interessiert sind. Die Eltern haben vor allem die Förderung ihrer Kinder im Blick und die Bildungsverwaltungen hoffen, dass sich die Ergebnisse in die organisatorischen Strukturen einbetten lassen. Bereits hier wird deutlich, dass Evaluationsverfahren, je nach Sichtweise und Erfahrungen, sowohl als Enttäuschung und Bedrohung als auch als Bereicherung erlebt werden können.

Die Autorin generiert unter Verwendung von fokussierten Interviews und quantitativen Fragebögen bei einer Stichprobe von ca. 40 Probanden vier verschiedene Rezeptions- und Nutzungstypen, die sich bei der Verwendung von Evaluationswissen in einigen Bereichen voneinander unterscheiden (vgl. Stamm 2003a). Auf die eine Seite ihrer ausgewählten Vergleichsdimensionen stellt sie die Art und Weise der Nutzung von Evaluationswissen: 1.) direkte Nutzung nachweisbar, 2.) direkte Nutzung unklar und 3.) direkte Nutzung nicht vorhanden. Auf der anderen Seite werden die unterschiedlichen Ziele und Verwendungszwecke aufgelistet, die mittels Evaluation verfolgt werden können. Hierzu führt sie neben der Kontrolle sowohl die Legitimation als auch die Entwicklung bzw. Optimierung an. Aus der Kombination dieser Merkmale ergeben sich vier unterschiedliche Typen, die in der folgenden Tabelle dargestellt sind. Im Anschluss an die Tabelle erfolgt die Beschreibung diese vier Nutzungstypen.

Tabelle 3.4.2: Typologie der Nutzung von Evaluationswissen nach Stamm (2005, S.20)

<table>
<tr><th rowspan="2">Rezeptions- und Nutzungsyp</th><th colspan="3">Zweck/Ziel</th></tr>
<tr><th>Kontrolle</th><th>Legitimation</th><th>Entwicklung / Optimierung</th></tr>
<tr><td>direkte Nutzung nachweisbar</td><td colspan="2">REAKTION</td><td>INNOVATION</td></tr>
<tr><td>direkte Nutzung unklar</td><td rowspan="2">BLOCKADE</td><td rowspan="2">ALIBI</td><td></td></tr>
<tr><td>direkte Nutzung nicht vorhanden</td><td></td></tr>
</table>

Der Typus der Innovation

Bei diesem Typus handelt es sich um idealtypische Nutzer von Evaluationsergebnissen. Er betrachtet Evaluationsverfahren als persönliche Angelegenheit und bringt eine enorme Anstrengungsbereitschaft mit, um die Schul- und Unterrichtsentwicklung voranzubringen. Die Auseinandersetzung mit den Ergebnissen erfolgt intensiv und systematisch. Häufig werden im Anschluss daran neue Entwicklungsziele formuliert und neue Maßnahmen abgeleitet, die im ungünstigen Fall zur Überlastung und Überanstrengung führen können, mit der Folge, dass der immense Arbeitsaufwand und alle Anstrengungen im Nichts enden.

Der Typus der Reaktion

Auch der Reaktions-Typus zeichnet sich durch eine schnelle Nutzung der Befunde aus. Allerdings werden die Ergebnisse nicht als Anlass für Veränderungen interpretiert, sondern werden als Bestätigung der eigenen Arbeit betrachtet, wenn diese positiv ausfallen. Auf negative Evaluationsergebnisse folgt Kritik, auf die mit schnellen Entscheidungen reagiert wird. Insgesamt hat bei diesem Typus eine rasche Umsetzung der Ergebnisse Priorität, die Generierung eines Erkenntnisinte-

resses ist sekundär. Die methodischen Ansprüche an Evaluation sind bei dieser Gruppe hoch. Die Autorin nennt den Reaktion-Typus als Beispiel für „eine mit Unruhen begleitete Evaluation, deren Nutzung rasch vor sich geht, aber häufig als unüberlegte Notmassnahme“ (Stamm 2005, S. 24).

Der Alibi-Typus

Bei diesem Typus besitzen Evaluationsverfahren einen Rechtfertigungscharakter der mit einem großen Legitimationsdruck einhergeht. Das Wissen aus Evaluationsverfahren wird nicht verwendet um z.B. Maßnahmen für die Schul- und Unterrichtsentwicklung abzuleiten. Bei diesen Personen besteht demnach weder ein Erkenntnis- oder Verwendungs-, noch ein Handlungsinteresse.

Der Blockade-Typus

Nach Meinung des Blockade-Typus dienen Evaluationen zu gleichen Teilen der Überprüfung und Kontrolle, als auch der Entwicklung. Es erfolgt kaum eine direkte Nutzung von Evaluationswissen. Dieser Typus verfügt über einen erziehungswissenschaftlichen-pädagogischen Hintergrund und sieht sich in der Lage, das Wissen autonom zu bewerten. Allerdings blockiert die vorwiegend politisch gefärbte Einstellung eine unmittelbare Nutzung der Befunde, da sie sich vornehmlich gegen die Makroebene des Bildungssystems, mit ihren hierarchischen Strukturen richtet (Stamm 2003a, S. 191). Die gewonnene Information wird zunächst auf ihre politische Verwendungstauglichkeit hin überprüft. Im Anschluss daran wird eine passende Rezeptionsvariante gewählt. Nach Stamm (2005) handelt es sich bei diesem Typus um ein „Paradebeispiel der Frage, ob Evaluationen den Vorstellungen der Bildungspolitik überhaupt genügen können“ (ebd., S. 22).

3.4.2.3 Zusammenfassung der Befunde zur Typenbildung

Während die erste Typologie die im Rahmen der VERA-Studie generiert wurde, relativ grob und eindimensional (Intensität der Auseinandersetzung) drei mögliche Rezeptionstypen beschreibt, bietet die Typologie nach Margrit Stamm eine differenzierte, mehrdimensionale und dennoch übersichtliche Einteilung möglicher Rezeptions- und Evaluationstypen.

Ein Grund für die unterschiedliche Wertigkeit und Güte der beiden Typologien liegt möglicherweise darin, dass im Rahmen der VERA-Studie die Einstellung und Haltung der Lehrpersonen zu Schulleistungsstudien nicht erhoben wurde. Erkenntnisse aus Rezeptionsstudien haben jedoch u. a. gezeigt, dass die Akzeptanz und grundsätzliche Einstellung der Lehrpersonen zu externen Evaluationsverfahren die Rezeption und Nutzung mehr beeinflussen, als die Testergebnisse der Schülerinnen und Schüler (vgl. Kapitel 3.4.1.2.1). Stamm's Typologie enthält neben der Erfassung der Nutzungsintensität auch die Zweck- und Zielbestimmung von Evaluationen. Des Weiteren werden mehrere Akteure mit unterschiedlichen Aufgabenbereichen auf verschiedenen Ebenen des Bildungswesens zu ihrer Einstellung gegenüber Evaluationsverfahren befragt. Dass diese aufgrund ihrer Position im Bildungssystem unterschiedliche Perspektiven und differente Sichtweisen bezüglich des Evaluationsziels und –zwecks einnehmen, liegt auf der Hand. Schließlich verbinden und verfolgen diese Akteure ganz unterschiedliche Handlungsinteressen mit externer Evaluation. Somit erscheint es auch logisch, dass sich aus dieser Befragung verschiedenen Rezeptionstypen abbilden lassen, die sich im Umgang mit Evaluationsverfahren zum Teil erheblich unterscheiden.

Zudem wendet die Forscherin Stamm neben quantitativen Erhebungsmethoden auch qualitative an, die ihre Stärke in der Erfassung subjektiver Theorien haben. Bei VERA wurden die Daten ausschließlich quantitativ erhoben und wie bereits erwähnt, wurden die persönlichen Einstellungen und Haltungen zu Evaluationsmaßnahmen nicht erhoben.

Es bleibt dennoch ungeklärt, ob sich auch innerhalb einer spezifischen Akteursgruppe, wie sie auch die Gruppe der Lehrerinnen und Lehrer darstellt, verschiedenen Rezeptionstypen bilden lassen, die sich bezüglich ihrer Einstellung und Haltung zu externen Evaluationsverfahren unterscheiden lassen. Dieser Frage soll im Rahmen der vorliegenden Arbeit nachgegangen werden.

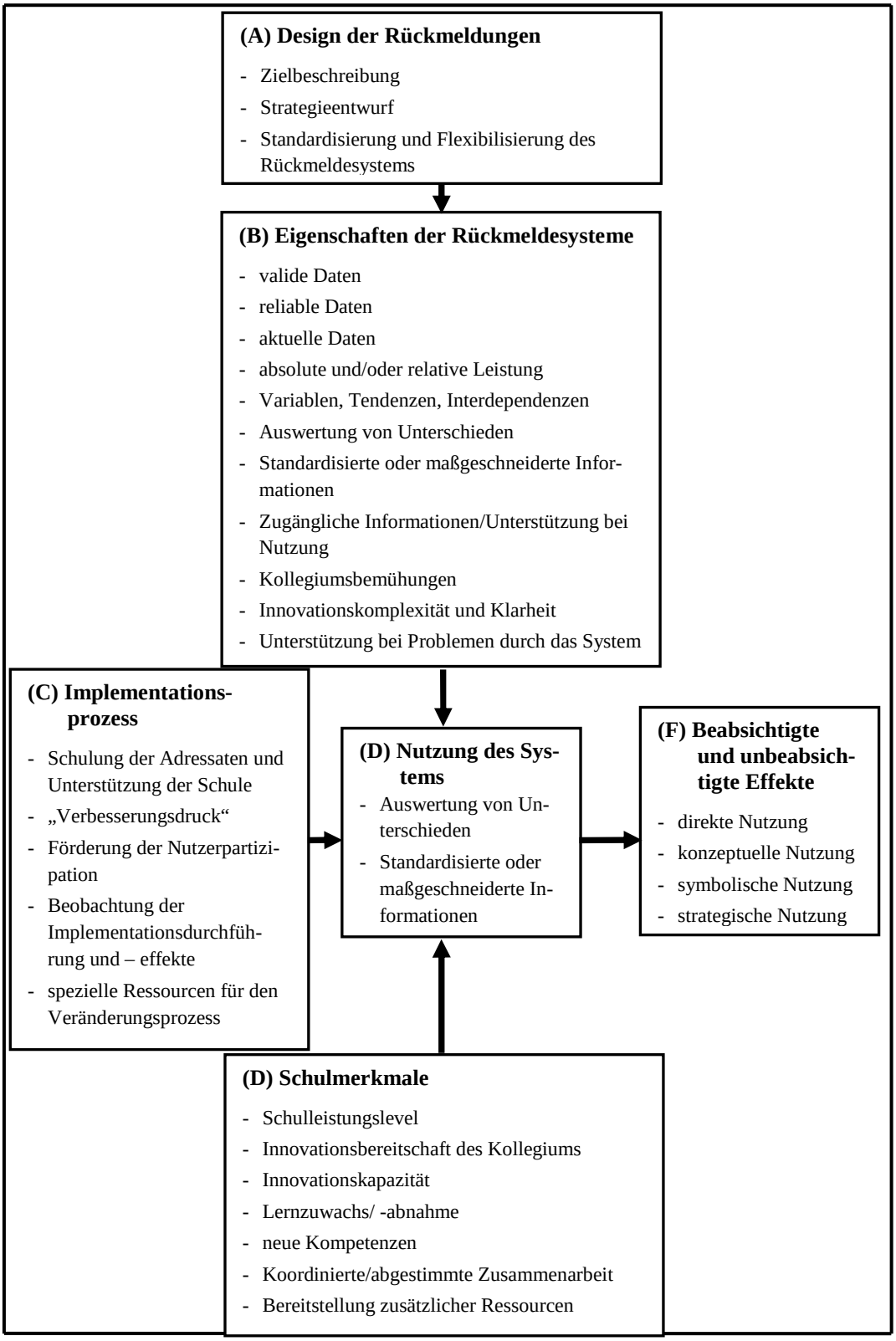
3.4.3 Modelle zur Rezeption von Rückmeldedaten

Die dargestellten Befunde aus den unterschiedlichen Forschungsprojekten zur Rezeption und Nutzung von Schulmonitoring-Studien beziehen sich auf verschiedene Bereiche und Akteure des Bildungssystems. Die Ebene der Einzelschule steht bei der Umsetzung dieser Evaluationsmaßnahmen im Vordergrund und stellt somit ein zentrales Moment dar. Um jene Elemente und Faktoren zu berücksichtigen, die an einer gelingenden Implementierung beteiligt sind, können Modelle verwendet werden, die sich als empirische Forschungsgrundlage eignen. In der vorliegenden Arbeit werden zwei relevante theoretische Ansätze ausgewählt und näher betrachtet. Zunächst wird das Modell von Visscher und Coe (2002) zum „School Performance Feedback System“ dargestellt, daran schließt sich die Vorstellung und Beschreibung des Modells von Helmke und Hosenfeld (2005) „Von der Evaluation zur Innovation“ an.

3.4.3.1 Modell Visscher & Coe (2002)

In anglo-amerikanischen Ländern werden Systeme des Bildungsmonitorings auf der Ebene der Einzelschule als „School Performance Feedback Systems“ (SPFS) bezeichnet. Im Zentrum dieser Rückmeldesysteme steht die regelmäßige Erhebung von Daten über Schülerleistungen und zentrale Kontextfaktoren. Diese Rückmeldungen zielen darauf ab, den Lehrpersonen wichtige Informationen zur Verbesserung ihres Unterrichts zur Verfügung zu stellen. Die Autoren Visscher und Coe erarbeiteten auf der Basis einer intensiven Literaturrecherche ein SPFS-Rahmenmodell, das die entscheidenden Faktoren für die Rückmeldung von Leistungsdaten enthält.

Abbildung 6: School Performance Feedback Systems nach Visscher & Coe (2002)



Nach Visscher und Coe (2002) wird die Intensität und Art der Nutzung von Feedback durch Anlage und Inhalt der Rückmeldungen sowie durch schulische Bedingungen beeinflusst.

Den Ausgangspunkt des Modells bildet das Design der Rückmeldungen (A). Hierzu zählen neben den Zielsetzungen und –beschreibungen des Feedbacks zum Beispiel auch die Hinweise zur strategischen Nutzung der Rückmeldungen. Außerdem finden sich auch Angaben über den Grad der Standardisierung, bzw. wie flexibel die Daten für spezifische Anforderungen verwendet werden können. Die Merkmale der konkreten Inhalte der Rückmeldungen (B) werden im nächsten Schritt zusammengefasst. Hier stehen die Gütekriterien der Leistungsrückmeldungen im Vordergrund. Des Weiteren werden Aspekte aufgeführt, die die Nutzung der Daten beeinflussen können. Dazu zählen u.a. die Zugänglichkeit zu den Daten, die Unterstützung durch Kollegen oder weitere Akteure des Systems oder auch die Bereitschaft für Innovationen. Diese Punkte lehnen sich bereits an Punkt (C) an, der die Bedingungen des Implementationsprozesses auflistet. Wie bereits in Kapitel 3.3 ausführlich beschrieben, spielt die Implementierung für die Nutzung der Daten (E) im schulischen Alltag eine entscheidende Rolle. Nach Visscher und Coe regt die alleinige Ausübung von Druck oder das Angebot von Unterstützung nicht zwangsläufig zu entsprechenden Veränderungsprozessen an. Wesentlich ist vielmehr eine Kombination dieser Strategien z.B. in Form eines umfassenden Unterstützungssystems und der Verpflichtung, aus den Ergebnissen Veränderungsmaßnahmen abzuleiten. Mit dem Bereich Schulmerkmale (D) führen die Autoren Aspekte an, die sich auf der Ebene der Einzelschule verorten lassen. Hierzu zählen der Leistungsstand der Schule, die schulische Bereitschaft und Kapazität für Neuerungen, die Kooperation im Kollegium sowie die Kompetenzen in der Analyse der zurückgemeldeten Ergebnisse. In Anlehnung an Rossi, Lipsey und Freeman (vgl. Rossi et al. 2004) werden im Modell drei Formen der Nutzung unterschieden; die instrumentelle, konzeptuelle und symbolische Nutzung. Unter instrumenteller Nutzung versteht man die direkte Verwendung von Evaluationsergebnissen: Die Ergebnisse werden analysiert und daraus Entscheidungen und Maßnahmen/Aktivitäten abgeleitet. Als konzeptuelle Nutzung wird dagegen die indirekte Verwendung von Ergebnissen bezeichnet, wodurch bestimmte Einstellungen oder Haltungen beeinflusst werden und darüber auf die abgeleiteten Akti-

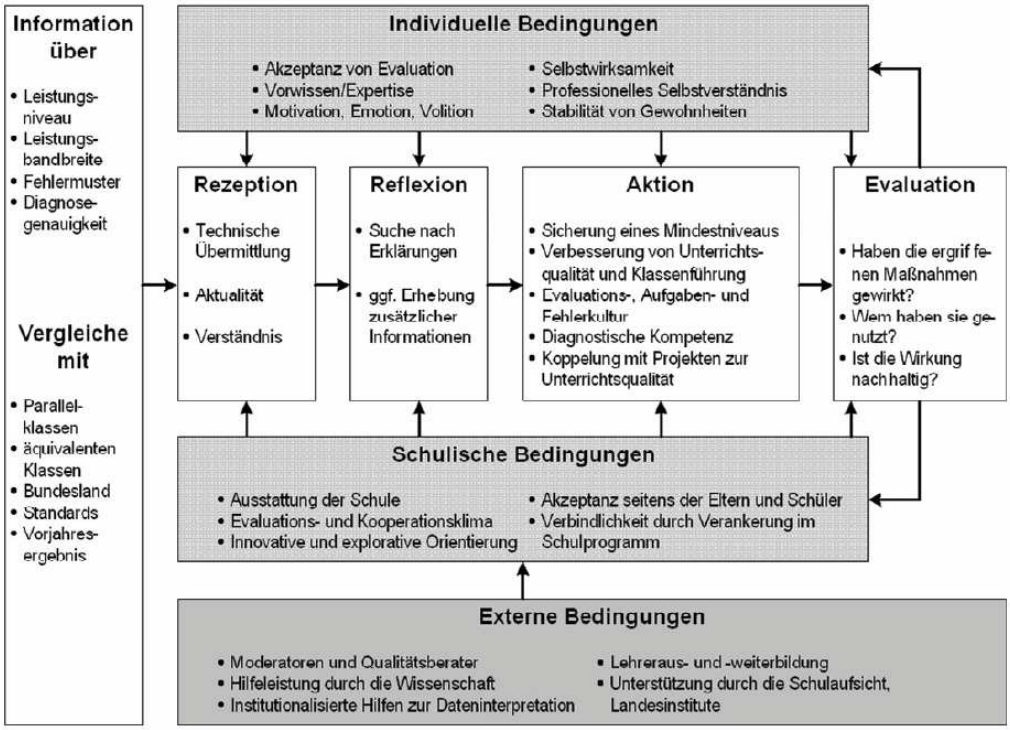
vitäten wirken. In der symbolischen Nutzung werden Informationen als Unterstützung des eigenen Standpunktes in Diskussionen herangezogen oder selektiv eingesetzt, um schon vorhandene Einstellungen oder getroffene Entscheidungen zu legitimieren. Visscher (2002, S. 57ff) ergänzt diese drei Nutzungsformen um strategische Nutzung, wonach Testergebnisse eher missbräuchlich verwendet werden, um strategische Ziele zu erreichen. Die Nutzung der Rückmeldungen mündet schließlich in beabsichtigte und unbeabsichtigte Effekte. Dabei entscheidet die einzelne Schule darüber, wie die Rückmeldungen verarbeitet werden (vgl. Bonsen & von der Gathen 2004, S. 247ff; Hosenfeld & Groß Ophoff 2007).

3.4.3.2 Modell von Helmke & Hosenfeld

Das von Helmke entwickelte Modell „*Von der Evaluation zur Innovation*“ skizziert den idealtypischen Verlauf, wie aus Evaluationsrückmeldungen mögliche Verbesserungen bzw. Innovationen abgeleitet werden können. Neben der prozessualen Achse, die die Schritte „Rezeption“, „Reflexion“, „Aktion“ und eine erneute „Evaluation“ vorsieht, werden mögliche Kontextfaktoren beschrieben, die den gesamten Prozess auf jeder einzelnen Stufe moderieren. Hierzu zählen neben den Individuellen Bedingungen auch schulische und außerschulische Kontexte (vgl. Hosenfeld & Groß Ophoff 2007, S. 357).

> „Der Wert dieses Modells liegt in seiner heuristischen Kraft, die verdeutlicht, dass eine Rückmeldung alleine keineswegs ausreicht, sondern spezifische Bedingungen des Gelingens erfüllt sein müssen um (insbesondere auch kontinuierlich) Veränderungen zu bewirken“ (Hosenfeld & Groß Ophoff 2007, S. 358).

Abbildung 7: „Von der Evaluation zur Innovation“ nach Helmke & Hosenfeld (2005)



Den Ausgangspunkt des Modells bilden die Rückmeldungen (in der Abbildung ganz links), die unterschiedliche Informationen z.B. über das Leistungsniveau der Klasse oder die Fehlermuster enthalten. Im Idealfall kommt es zunächst zur Rezeption der Rückmeldungen, gefolgt von der Reflexionsphase, in der nach möglichen Erklärungen für die Rückmeldeergebnisse gesucht wird. Daran schließt sich die Phase der Aktion an. Hier werden spezifische Maßnahmen zur Verbesserung und Optimierung abgeleitet und durchgeführt. Diese sollten in einem letzten Schritt auf ihre Effektivität und Effizienz hin evaluiert werden, wobei das daraus gewonnene Ergebnis wieder als Ausgangspunkt eines neuen Zyklus verstanden wird.

Dieser Gesamtprozess von der Rezeption zur Evaluation erfolgt schrittweise. Jeder Schritt gilt als Vorstufe für den nächsten. Jede der Phasen wird durch Kontextfaktoren beeinflusst. Diese sind einerseits die individuellen Bedingungen, also beispielsweise die Einstellungen der Lehrpersonen zu den Vergleichsarbeiten. Andererseits beeinflussen schulische und externe Faktoren den oben skizzierten Prozess.

Beitrag der Modelle zur Rezeptionsforschung:

Beide Modelle leisten einen wichtigen Beitrag für die Rezeptionsforschung. Das Modell von Visscher und Coe beinhaltet die maßgeblichen Faktoren, die für die Rückmeldung von Leistungsdaten relevant sind. Diese Faktoren wurden aus der intensiven Analyse zahlreicher Forschungsprojekte im Bereich der Datenrückmeldung generiert und stellen somit eine verlässliche Grundlage zur Bewertung zukünftiger Forschungen dar.

Die Stärke des Modells von Helmke und Hosenfeld liegt auf der Darstellung des Prozesses bei der Rezeption und Nutzung von Rückmeldedaten. Jede einzelne Phase der Verwendung von Rückmeldungen wird hierbei berücksichtigt.

Die Verwendung beider Modelle bietet somit den Vorteil, dass sowohl Prozessstrukturen, als auch inhaltliche Normierungen, die die Rezeption und Reflexion von Rückmeldedaten beeinflussen, berücksichtigt werden. Somit kann zum einen klar aufgezeigt werden, ob in bestimmten Phasen die Rezeption und Nutzung besonders gut gelingt oder ob es zu Brüchen kommt. Zum anderen liegen umfassende Kriterien vor, die zur Bewertung von Rezeptionsstudien herangezogen werden können.

Empirischer Teil:

Eine qualitative Interviewstudie zur Rezeption und Nutzung von Vergleichsarbeiten an baden-württembergischen Haupt-, Realschulen und Gymnasien.

4 Fragestellung und Forschungsdesign

Der empirische Teil der vorliegenden Dissertation knüpft an die theoretischen Ausführungen der vorhergehenden Kapitel an. Darin wird deutlich, dass bereits zahlreiche Ergebnisse zur Rezeption und Nutzung von Vergleichsarbeiten vorliegen, die vor allem aus quantitativen Erhebungen stammen (siehe Kapitel 3.4.1). Der Vielschichtigkeit des Bildungssystems geschuldet, existieren jedoch immer noch zu wenige Erkenntnisse darüber, wie die unterschiedlichen Faktoren, die die Rezeption und Nutzung von Leistungsrückmeldungen beeinflussen, zusammenwirken. Hier sind weitere systematisch gewonnene Erkenntnisse notwendig. Sie sollen dazu beitragen, diese Komplexität zu erfassen, die nicht zuletzt aufgrund der Vielzahl der daran beteiligten Akteure in Schule und Unterricht zustande kommt (vgl. Hosenfeld & Groß Ophoff 2007, S. 363). Eine Möglichkeit einen Beitrag zur Klärung dieses Sachverhalts zu liefern, liegt in der Anwendung qualitativer Forschungsmethoden. Diese Verfahren fokussieren die subjektive Perspektive der einzelnen Akteure. Im Zentrum der hier vorgestellten qualitativen Interviewstudie stehen die Fragen, wie Lehrerinnen und Lehrer in Baden-Württemberg die Vergleichsarbeiten in der Sekundarstufe I des allgemein bildenden Schulsystems wahrnehmen und nutzen und welchen Beitrag diese zur Qualitätsentwicklung und Qualitätssicherung beitragen können. Mit diesem Vorgehen ist die Hoffnung verbunden, vertiefende Erkenntnisse über den Umgang mit Vergleichsarbeiten zu erhalten

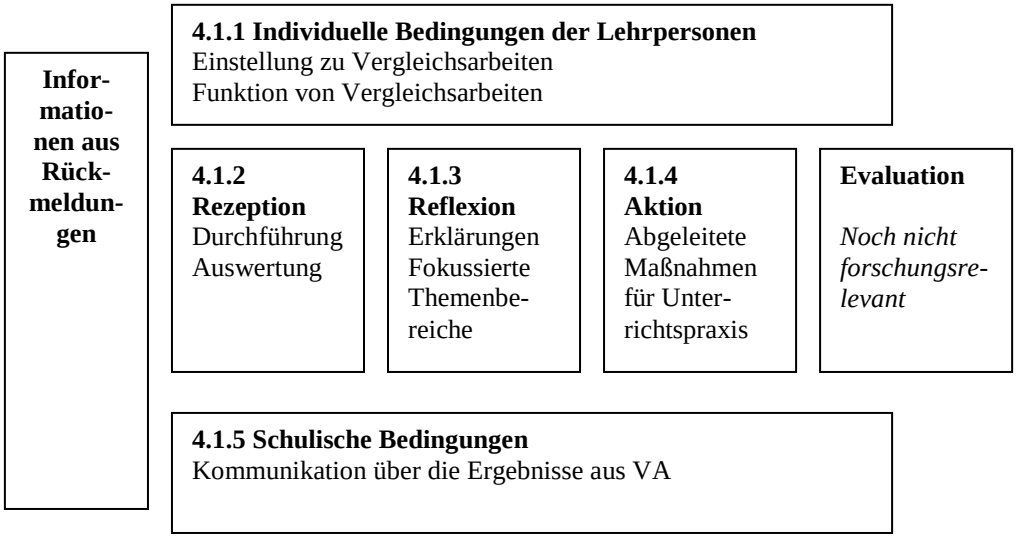
Das Modell von Helmke und Hosenfeld (2005) liefert die theoretische Grundlage im gesamten Projekt: bei der Fragestellung, der Datengewinnung, der Auswertung der Daten und bei der Präsentation der Ergebnisse (vgl. Kap. 3.4.3.2).

Bevor das Forschungsdesign und -vorgehen detailliert dargestellt und erläutert wird, erfolgt zunächst die Präzisierung der Fragestellung.

4.1 Fragestellung

Entsprechend des idealtypischen Verlaufs der Rezeption und Nutzung von Rückmeldedaten lassen sich die einzelnen Fragebereiche der Studie den verschiedenen Bereichen des Modells zuordnen (siehe Abbildung 8):

Abbildung 8: Übersicht über die Fragebereiche in Anlehnung an Helmke & Hosenfeld (2005)



Für jeden Bereich 4.1.1 bis 4.1.5 erfolgt nun die differenzierte Fragestellung. Der letzte Prozessschritt „Evaluation“, findet in dieser Studie keine Berücksichtigung. Da er an mögliche „Aktionen“ anschließt, bezieht er sich demnach auf die gewählten Maßnahmen der Lehrenden, die zur Verbesserung der Unterrichtsqualität einen Beitrag leisten sollen. Da diese jedoch erst nach einer gewissen Dauer auf ihre Wirkung hin überprüft werden können, können die Lehrerinnen und Lehrer zu diesem Zeitpunkt zu deren Effektivität noch keine Aussagen machen.

4.1.1 Individuelle Bedingungen der Lehrpersonen

Dieser Bereich befasst sich mit den subjektiven Einstellungen und Haltungen der Lehrerinnen und Lehrer zu Vergleichsarbeiten. Die konkreten Fragestellungen lauten:

1	Welche subjektive Einstellungen und Haltungen zu Vergleichsarbeiten konnten die Lehrerinnen und Lehrer im Zuge der Einführung entwickeln?
2	Welche Funktion haben Vergleichsarbeiten nach Meinung der Lehrpersonen?

4.1.2 Rezeption der Vergleichsarbeiten

Den Anfang des idealtypischen Prozesses von der Rezeption bis zur Evaluation bildet zunächst die Wahrnehmung der Rückmeldungen aus Vergleichsarbeiten. Im Vordergrund stehen die technischen Aspekte, insbesondere die Durchführung und Verständlichkeit der Daten:

3	Wie ließen sich die Vergleichsarbeiten durchführen?
4	Waren die Rückmeldungen verständlich?

4.1.3 Reflexion der Ergebnisse aus Vergleichsarbeiten

Der zweite Schritt setzt die Reflexionsprozesse der Lehrpersonen ins Zentrum des Interesses.

5	Welche Erklärungen leiten die Lehrerinnen und Lehrer aus den Ergebnissen aus Vergleichsarbeiten ab?
6	Welche Themenbereiche treten durch die Überlegungen und Bewertungen der Lehrpersonen besonders ins Zentrum der Reflexion?
7	Welche Gründe für die entsprechenden Ergebnisse werden von den Lehrpersonen angeführt?

4.1.4 Mögliche Aktionen, Maßnahmen

Im dritten Schritt wird nach konkreten Maßnahmen gefragt, die die Lehrerinnen und Lehrer aufgrund der ersten beiden Schritte (Rezeption und Reflexion) in Zukunft ergreifen werden:

8	Werden die Lehrpersonen aufgrund der Rückmeldung aus Vergleichsarbeiten eine Änderung ihrer Unterrichtspraxis vornehmen?
9	Welche Maßnahmen setzen die Lehrerinnen und Lehrer zur Optimierung ihres Unterrichts ein?

4.1.5 Schulische und Kontextbedingungen

Dieser Bereich fokussiert das schulische Umfeld der Lehrpersonen. Im Zentrum stehen die Kommunikationsstrukturen, die zwischen den verschiedenen Akteuren und Gremien der Einzelschulen bestehen.

10	Findet an den Schulen eine systematische Kommunikation über die Ergebnisse aus Vergleichsarbeiten statt?
11	Mit welchen Personengruppen bzw. in welchen Gremien werden die Ergebnisse systematisch diskutiert?
12	Werden die Ergebnisse aus Vergleichsarbeiten mit den Eltern besprochen?

4.1.6 Verdichtung der Kategorien, Typenbildung

Dieser Fragebereich bezieht sich erneut auf den Fragebereich 4.1.1. Einige Befunde aus bisherigen Studien weisen darauf hin, dass die Rezeption und Nutzung der Ergebnisse aus Schulleistungsstudien von der jeweiligen subjektiven Einstellung der einzelnen Lehrperson abhängig ist. Mit dieser Frage ist die Hoffnung verbunden, dass aus den Ergebnissen der qualitativ gewonnenen Daten typische Muster im Rezeptions- und Nutzungsverhalten dargestellt werden können.

13	Lassen sich innerhalb der befragten Lehrpersonen unterschiedliche Rezeptions- und Nutzungstypen generieren?

4.2 Forschungsdesign - Material und Methoden

Die vorliegende qualitative Interviewstudie 2007 wurde zu Beginn des Schuljahres 2007/2008 mit dem Ziel durchgeführt, Prozesse der Rezeption und Nutzung von Rückmeldungen aus Vergleichsarbeiten zu analysieren. Die Lehrerinnen und Lehrer beteiligten sich an der zweiten verpflichtenden Durchführung von Vergleichsarbeiten. Bereits ein Jahr zuvor wurden Lehrpersonen direkt im Anschluss an die ersten Vergleichsarbeiten mit der gleichen Fragestellung interviewt. Es zeigte sich jedoch, dass der Zeitpunkt der Datenerhebung, der direkt im Anschluss an die erste verbindliche Durchführung der Vergleichsarbeiten im Juni/Juli 2006, also unmittelbar vor den Sommerferien stattfand, einige Schwierigkeiten mit sich brachte. Die Lehrpersonen hatten in der kurzen Zeitspanne zwischen Korrektur und Ferienbeginn nicht ausreichend Möglichkeiten, um sich mit den Ergebnisrückmeldungen vertiefend auseinanderzusetzen und daraus weitere Maßnahmen abzuleiten. Die Qualität der Interviewaussagen war dadurch unbefriedigend. Besonders im Fragebereich „Reflexion der Ergebnisse“ und „Mögliche Aktionen“ waren die Aussagen ungenügend. Aus diesem Grund wurde bei der Interviewstudie 2007 der Zeitpunkt der Datenerhebung an den Beginn des darauffolgenden Schuljahres gelegt; alle Interviews fanden im September/Oktober 2007 statt. Somit hatten die befragten Lehrerinnen und Lehrer ausreichend Zeit sich mit der Rezeption und Nutzung von Vergleichsarbeiten auseinander zu setzen.

Bei der Interviewstudie 2007 stand die Sekundarstufe I im Zentrum der Untersuchung. Neben Haupt- und Realschulen wurden zum ersten Mal Gymnasien in die Stichprobe mit aufgenommen. An Grundschulen wurden keine Interviews durchgeführt, da sich diese ab dem Schuljahr 2006/07 an der bundesweiten VERA-Studie beteiligten.

4.2.1 Ziehung und Umfang der Stichprobe – Beschreibung der Stichprobe

4.2.1.1 Stichprobenauswahl

Bei der Fragestellung, wie Lehrerinnen und Lehrer die Ergebnisse aus Vergleichsarbeiten rezipieren und nutzen, interessiert zunächst nicht die Meinung einer einzelnen Lehrperson, sondern die Meinung unterschiedlicher Personen, die zufällig ausgewählt werden und an den verschiedenen Schularten in der Sekundarstufe Erfahrungen mit der Durchführung von Vergleichsarbeiten sammeln konnten. Die Auswahl einer größeren Stichprobe ermöglicht die Darstellung typischer Umgangsmuster mit Rückmeldedaten und in einem weiteren Analyseschritt auch die Beschreibung von bestimmten Rezeptions- und Nutzungstypen. Die Fallauswahl erfolgt mit Hilfe von Schullisten, aus denen zufällig einzelne Schulen generiert wurden. Die Anzahl der befragten Lehrpersonen orientierte sich an folgenden theoretischen Vorüberlegungen:

In der Stichprobe sollten Lehrerinnen und Lehrer aller Schularten des allgemein bildenden Schulsystems der Sekundarstufe I vertreten sein. Bei der Auswahl sollte darauf geachtet werden, dass die Anzahl der jeweiligen Schulen und die der jeweiligen Vertreter der Haupt-, Realschulen und Gymnasien ausgewogen sind und die Verteilung auf die unterschiedlichen Schularten annähernd gleich ist.

Außerdem sollten bei der Stichprobenauswahl folgende Punkte berücksichtigt werden (vgl. Merkens 2003, S. 294):

Die Personen sollten alle über ausreichend Wissen zu Vergleichsarbeiten verfügen. Das bedeutet, dass sie im Schuljahr 2006/2007 mindestens in einem Fach an der Durchführung einer Vergleichsarbeit beteiligt waren.

Des Weiteren sollten sie einerseits die Bereitschaft aufbringen, sich für die Befragung Zeit zu nehmen. Andererseits sollten die Beteiligen über die Fähigkeiten verfügen zu reflektieren und sich zu artikulieren. Von der Erfüllung der beiden Bedingungen kann in der Berufsgruppe von Lehrerinnen und Lehrern ausgegangen werden.

Als letzten Punkt listet Merkens die generelle Bereitschaft der Interviewpartnerinnen und –partner auf, an der Befragung teilzunehmen. Auch diese Bedingung wurde bei der Stichprobenziehung berücksichtigt. So hatte jeweils die Schulleitung als auch jede einzelne Lehrperson die Möglichkeit, eine Teilnahme an der Untersuchung abzulehnen.

Sampling durch Gatekeeper

So genannte Gatekeeper sind Experten, die einen Überblick haben über Personen, die in einem bestimmten sozialen Feld tätig sind - beispielsweise eine Streetworkerin, die mit einer bestimmten Gruppe von Jugendlichen arbeitet (vgl. Reinders 2005, S. 139ff; Merkens 1997, S. 101f). In der vorliegenden Untersuchung wurden Vertreter der Schulämter als Gatekeeper genutzt, um die Forscher bei der Erstellung der Schullisten zu unterstützen und mit Hilfe eines kurzen Schreibens die Schulen zur Mitarbeit zu motivieren. Zwar stehen bei der vorliegenden Studie die Lehrerinnen und Lehrer und nicht die Einzelschule im Zentrum des Forschungsinteresses, dennoch erhofft man sich durch dieses Vorgehen, unterschiedliche Typen zu erreichen. Würde man sich direkt an Lehrpersonen wenden, so erklärten sich womöglich nur besonders motivierte bzw. besonders kritische Personen zum Interview bereit.

Dokumentation der Stichprobenziehung

Die Stichprobenziehung erfolgte in folgenden Schritten:

1. Zuerst wurden die Regionen festgelegt, in welchen die Untersuchung durchgeführt werden sollte. In der Stichprobe sollten sowohl städtische, als auch ländliche Schulen vertreten sein, um regionale Differenzen abzubilden. Aus forschungsökonomischen Überlegungen, sollten die Regionen für die Interviewer gut erreichbar sein.

2. Nachdem die Regionen festgelegt waren, wurden dort für alle Schularten der Sekundarstufe I vollständige Schullisten erstellt, woraus per Zufall die Stichprobe gezogen wurde. Mit Unterstützung der zuständigen staatlichen Schulämter als Gatekeeper wurden die Schulen schriftlich über die Studie informiert und gebeten, an der Untersuchung teilzunehmen. Daraufhin wurde mit der Schulleitung telefonisch Kontakt aufgenommen und nachgefragt, ob die Schule Interesse an einer Beteiligung hätte. Bei diesem Gespräch stand die Information zum Projekt, zur Durchführung und zum Datenschutz im Vordergrund. Wurde die Teilnahme abgelehnt, so wurde die nächste Schule der Liste angerufen.

3. Stimmte die Schule einer Teilnahme zu, so wurde versucht, möglichst alle Lehrpersonen zu befragen, die im vorhergegangenen Schuljahr Vergleichsarbeiten durchgeführt hatten. Mit den einzelnen Lehrpersonen wurden daraufhin persönlich Termine für die Befragung an der Schule vereinbart.

4.2.1.2 Die Stichprobe der Interviewstudie 2007

Die Interviews wurden in allen drei Schulformen der Sekundarstufe I des baden-württembergischen allgemein bildenden Schulsystems durchgeführt. Die Haupt- und Realschulen wurden im Stadtbereich Stuttgart als auch in der Region Ostwürttemberg in den Landkreisen Aalen (Ostalbkreis), Göppingen und Heidenheim) per Zufall gezogen. Die beteiligten Gymnasien befinden sich in den Landkreisen Göppingen und Heidenheim.

Stichprobe an Hauptschulen

Für die Untersuchung an Hauptschulen wurden an acht Schulen der Region Stuttgart und Ostwürttemberg neun Lehrerinnen und sechs Lehrer befragt, die in insgesamt 14 Klassen Vergleichsarbeiten durchgeführt hatten: neun im Fach Deutsch, vier in Mathematik und zwei in Englisch. Die Klassengröße betrug im Durchschnitt 19,7 Schülerinnen und Schüler. Davon stammten durchschnittlich neun Kinder aus Familien mit Migrationshintergrund.

Stichprobe an Realschulen

Analog zur Hauptschulstichprobe wurden in der gleichen Region acht Realschulen gezogen. 23 Personen erklärten sich zu einem Interview bereit (zehn männliche Kollegen und 13 weibliche Kolleginnen). In 16 Klassen mit durchschnittlich 29 Schülern wurden im Fach Deutsch elf, im Fach Mathematik neun, im Fächerverbund Erdkunde-Wirtschaft-Gesellschaft zwei und im Fach Geschichte eine Vergleichsarbeit durchgeführt. Der Migrantenanteil lag im Durchschnitt bei fünf Schülern pro Klasse und damit deutlich niedriger als in der Hauptschulstichprobe.

In der Realschule stehen 23 Lehrpersonen – somit acht mehr als in der Hauptschule - für die Befragung zur Verfügung. Der Grund hierfür liegt einerseits darin, dass in der Realschule in vier Fächern Vergleichsarbeiten geschrieben wurden und wegen des Fachlehrerprinzips mehr Lehrer in die Durchführung von Vergleichsarbeiten involviert waren. Andererseits sind Realschulen im Vergleich zu den ein- oder zweizügigen Hauptschulen häufig zwei- oder dreizügig, folglich stehen insgesamt mehr Klassen und Lehrer zur Verfügung.

Stichprobe an Gymnasien

Die vier Gymnasien an denen die Interviews stattfanden, befinden sich in den Landkreisen Göppingen und Heidenheim. Jedes der vier Gymnasien ist dreizügig mit einer durchschnittlichen Schülerzahl von 29,9 Schülern und einem Migrantenanteil von 2,7 Schülern pro Klasse. Insgesamt wurden 21 Lehrer, davon zwölf Lehrer und neun Lehrerinnen, befragt. Während im Fach Deutsch zehn und im Fach Mathematik acht Vergleichsarbeiten geschrieben wurden, lag die Anzahl im Fach Biologie bei zwei und im Fächerverbund GWG[21] bei einer Vergleichsarbeit.

[21] Fächerverbund im Gymnasium: Geschichte – Wirtschaft - Gemeinschaftskunde

Gesamtstichprobe

Das beschriebene Verfahren der Stichprobenziehung liefert für die Untersuchung ein breites Spektrum an Interviews. Mit insgesamt 59 befragten Personen, 28 Lehrern und 31 Lehrerinnen aus 20 verschiedenen Schulen stehen für die Untersuchung der Fragestellungen ausreichend Fälle zur Verfügung. Dadurch dass in mehreren Schulen und Schularten der Sekundarstufe I (Haupt-, Realschule und Gymnasium) eine Vielzahl von Lehrerinnen und Lehrern befragt werden konnte, wird angenommen, dass das ganze Spektrum möglicher Rezeptions- und Nutzungstypen in der Stichprobe hinreichend enthalten ist und sie somit als repräsentativ bezeichnet werden kann.

Tabelle 4.2.1: Gesamtstichprobe der Interviewstudie 2007

	Einzugsgebiet	Anzahl Schulen	Anzahl Lehrpersonen (m/w)	Klassen
Hauptschulen	Region Stuttgart und Ostwürttemberg	8	15 (6/9)	14 Klassen (9 D, 4 M, 2 E) durchschnittlich 19,7 Schüler und durchschnittlich 9 Migranten pro Klasse (Lehrerangabe)
Realschulen	Region Stuttgart und Ostwürttemberg	8	23 (10/13)	16 Klassen (11 D, 9 M, 2 EWG, 1 G) durchschnittlich 29 Schüler und durchschnittlich 5 Migranten pro Klasse (Lehrerangabe)
Gymnasein	LKR Göppingen und LKR Heidenheim	4	21 (12/9)	12 Klassen (10 D, 8 M, 2 Bio, 1 GWG) durchschnittlich 29,9 Schüler und durchschnittlich 2,7 Migranten pro Klasse (Lehrerangabe)

4.3 Erhebungsmethoden

Im Fokus stehen subjektive Sichtweisen, Einstellungen und Haltungen von Lehrerinnen und Lehrern, die sie im Zuge der Einführung von Vergleichsarbeiten entwickeln konnten. Mit Hilfe qualitativer Forschungsmethoden können diese „Lebenswelten, << von innen heraus >>, aus der Sicht der handelnden Menschen“ (Flick et al. 2003, S. 14 [Hervorhebung im Original]) beschrieben werden. Qualitative Forschung „will zu einem besseren Verständnis sozialer Wirklichkeiten beitragen und auf Abläufe, Deutungsmuster und Strukturmerkmale aufmerksam machen“ (ebd., S. 14). Eine Stärke der qualitativen Forschung im Vergleich zu quantitativen, standardisierten Verfahren liegt in ihrer Offenheit und der damit in Zusammenhang stehenden Möglichkeiten, die Perspektive der einzelnen Personen differenziert abzubilden. In der Schulpädagogik kommen qualitative Forschungsmethoden zur zahlreichen Anwendung (vgl. Ackermann & Rosenbusch 1995).

Unter dem Oberbegriff „Qualitative Forschung“ werden zahlreiche unterschiedliche Forschungsansätze subsumiert, die sich zu drei Hauptzweigen zusammenfassen lassen (siehe Tabelle bei Flick et al. (2003), S. 19). Diese unterscheiden sich hinsichtlich ihrer theoretischen Position, ihrem zugrundeliegenden Gegenstandsverständnis und ihrem methodischen Schwerpunkt. Einer dieser Ansätze stellt die Sicht des Subjekts in den Vordergrund der Betrachtung. Er lehnt sich an die theoretische Position des Symbolischen Interaktionismus an (vgl. Denzin 2003) und beinhaltet zahlreiche Erhebungs- und Auswertungsverfahren, die je nach Fragestellung und Forschungsdesign passend ausgewählt werden müssen.

Zur Erforschung subjektiver Theorien eignen sich unterschiedliche Erhebungsmethoden (vgl. König 1995, S. 14–19; König & Zedler 1995). Zum einen bietet sich die Methode der freien Beschreibung an, in der der Interviewpartner möglichst frei und offen seine subjektive Sichtweise mitteilt. Als geeignete Verfahren zur Erhebung der Daten stehen dem Forscher neben dem freien Assoziieren, das narrative Interview oder die schriftliche Beschreibung zur Verfügung.

Beim sog. Grid-Verfahren werden aus einer Liste von zehn Personen drei ausgewählt und danach beurteilt, welche Gemeinsamkeiten zwei Personen im Unterschied zur dritten gemeinsam haben. Bei dieser Technik stehen die Konstrukte der Personen im Zentrum des Interesses.

Die Methode des Lauten Denkens (Selbskonfrontations-Interview) bietet eine weitere Möglichkeit der Erfassung subjektiver Theorien. Diese werden von den Befragten in Bezug auf konkrete Handlungssituationen artikuliert. Im Anschluss an durchgeführte Interviews bietet das Struktur-Lege-Verfahren, das in einem zweiten Schritt zu einem späteren Zeitpunkt mit den Befragten durchgeführt wird, die Möglichkeit, die Struktur der jeweiligen subjektiven Theorie zu explizieren und dadurch abzusichern.

Zudem stellt das Leitfaden-Interview oder halb-strukturierte Interview eine klassische Erhebungsmethode für die Erfassung subjektiver Theorien dar. Bei diesem Verfahren geht es darum, die Sichtweise des Gesprächspartners zu bestimmten Leitfragen möglichst unverzerrt zu erfassen. Im Forschungsdesign der vorliegenden Arbeit wurde diese Methode zur Erhebung der Daten favorisiert, da der Studie ein umfangreiches theoretisches Modell zugrunde liegt auf das sich die einzelnen Fragenbereiche beziehen. Das Leitfaden-Interview wird im folgenden Kapitel detailliert dargestellt.

4.3.1 Das Leitfaden-Interview zur Erforschung subjektiver Theorien

Im Bereich der qualitativen Befragungs- bzw. Interviewtechniken wird eine Vielzahl von Varianten unterschieden (vgl. Bortz & Döring 2006, S. 315; Friebertshäuser & Langer 2010; Friebertshäuser et al. 2010; Hopf 2003; Lamnek 2005). Welches Interview gewählt wird resultiert aus dem zugrundeliegenden Forschungsdesign: dem Erkenntnisinteresse, dem theoretischen Vorwissen, der Fragestellung und Zielsetzung als auch der methodischen Anlage der Studie. In der vorliegenden Untersuchung wurde die Methode des Leitfadeninterviews gewählt.

Das Leitfadeninterview - auch halbstrukturiertes Interview genannt- zählt zu den gängigsten Interviewtechniken der qualitativen Forschungsmethoden (vgl. Bock

1992; Bortz & Döring 2006, S. 314). Im Vorfeld der Datenerhebung wird ein Leitfaden entwickelt, der für die Datenauswertung und Datenanalyse ein Gerüst bereitstellt und so die Vergleichbarkeit zwischen verschiedenen Interviews erleichtert. Trotz der Vorstrukturierung durch den Leitfaden bleibt den Befragten während des Interviews die Offenheit der Themenwahl und die Gestaltungsmöglichkeit des Gesprächs bestehen (vgl. Marotzki 2003).

Je nachdem, welche Vorüberlegungen für die Entwicklung des Leitfadens zu Grunde liegen, lassen sich weitere Befragungstechniken unterscheiden. In der vorliegenden Untersuchung wird aus dem Bereich der Leitfadeninterviews das Problemzentrierte Interview verwendet.

4.3.1.1 Das Problemzentrierte Interview

Das problemzentrierte Interview wurde von Andreas Witzel entwickelt (vgl. Friebertshäuser & Langer 2010, S. 442). Dieser Interviewform liegt ein Leitfaden zugrunde, der aus einem bereits bestehenden wissenschaftlichen Konzept generiert wird. Die Forschenden gehen somit nicht ohne jegliches Vorwissen in die Interviewsituation, sondern kennen die wesentlichen und bedeutenden Elemente des Forschungsgegenstandes. In Anlehnung an diese Expertise werden die Inhalte der Befragung festgelegt und passende Leitfragen formuliert. Dabei werden Elemente aus der leitfadenorientierten und aus der offenen Befragung miteinander kombiniert.

Neben dem Leitfaden umfasst das problemzentrierte Interview zwei weitere Elemente der Datenerhebung. Zum einen werden in einem Kurzfragebogen zum Einstieg in das Interviews bedeutsame demografische Daten der Interviewpartner erfasst. Zum anderen wird im Anschluss an das Interview ein Postscriptum erstellt, das Informationen zur Interviewsituation, zum Interviewverlauf oder auch erste Eindrücke und Interpretationsideen enthält.

Der Interviewleitfaden

Im vorliegenden Forschungsprojekt existiert ein umfangreiches wissenschaftlich fundiertes Theoriekonzept zur Rezeption von Rückmeldungen aus Leistungsstudien und deren mögliche Verwendung für die Qualitätsentwicklung und Qualitäts-

sicherung an Schulen. Dem Leitfaden liegt das Zyklenmodell „Von der Rückmeldung zur Innovation“ (Hosenfeld & Groß Ophoff 2007) zugrunde. In diesem Modell wird der idealtypische Verlauf skizziert, wie Lehrpersonen Rückmeldungen aus Vergleichsstudien für die Qualitätsentwicklung und –sicherung ihres Unterrichts, bzw. ihrer Schule nutzen können (siehe Kapitel 3.4.2.2).

Der Leitfaden ist in vier Abschnitte unterteilt, wobei der erste Abschnitt im Sinne des problemzentrierten Interviews dem Kurzleitfaden entspricht, der demographische Daten der Befragten erfasst. Die anderen Abschnitte beinhalten die eigentliche Thematik; die Rezeption und Reflexion von Ergebnissen aus Vergleichsarbeiten. Der erste thematische Bereich bezieht sich auf die Lehrperson und enthält Fragen zur Einstellung der Lehrpersonen zu Vergleichsarbeiten. Die Rezeption und Nutzung der Ergebnisse stehen im Zentrum des zweiten Themenkomplexes. Der letzte Abschnitt fokussiert den schulischen Kontext der Kommunikation der Ergebnisse. Eine detaillierte Beschreibung der einzelnen Fragebogen und eine Zuordnung zur Fragestellung schließt sich nun an:

Der Kurzfragebogen: Demografische Angaben zur Lehrperson, zur Klasse und Schule

Der Kurzfragebogen enthält Fragen zu folgenden Bereichen:

- Geschlecht der interviewten Lehrperson
- Mitglied der Schulleitung
- Deputatsumfang
- Angaben in welchen Fächern an der Schule Vergleichsarbeiten durchgeführt wurden
- Angaben zur Klassengröße
- Informationen, wie viele Schüler der Klasse einen Migrationshintergrund aufweisen
- Jahres-Leistungsstand der Klasse
- Notendurchschnitt der Vergleichsarbeit

Teil A: Individuelle Bedingungen und Rezeption – Lehrerebene

Der erste thematische Teil des Interviews beschäftigt sich mit Fragen, die einerseits die subjektive Einstellung und Haltung der einzelnen Lehrperson zu den Vergleichsarbeiten und andererseits Informationen zur Rezeption der Ergebnisse erheben möchte. Diese Bereiche beziehen sich auf die „Individuellen Bedingungen der Lehrpersonen“ (Forschungsfrage 1 und 2 aus Kapitel 4.1.1) und auf die „Rezeption der Vergleichsarbeiten“ (Forschungsfragen 3 und 4 aus Kapitel 4.1.2):

1. **Funktion der Vergleichsarbeit:** Die erste Frage dient als offener Einstieg ins Interview, der die Interviewpartner dazu anregen soll, sich frei zu äußern. Auf die Fragen „Warum das Ministerium Vergleichsarbeiten eingeführt hat“ und „Wie der Sinn der Vergleichsarbeiten den Eltern und Schülern erklärt wird“ fühlen sich die Lehrpersonen nicht persönlich angesprochen, da der Fokus beim Ministerium liegt.
2. **Durchführung und Auswertung der Vergleichsarbeiten:** In diesem Fragekomplex geht es um die Praktikabilität der gesamten Testanweisung. Neben Fragen, ob sich die VA wie geplant durchführen ließ wurde auch gefragt, ob die Zeit, die die Schülerinnen und Schüler für die Aufgabenbearbeitung zur Verfügung hatten, ausgereicht hatte.
3. **Verständlichkeit der Rückmeldungen:** Hier wurden die Lehrpersonen dazu aufgefordert darüber zu berichten, welche Datenblätter der Rückmeldung sie besonders intensiv verwendet und wie gut sie diese verstanden haben. Außerdem wurde gefragt, auf welchen Ebenen die Vergleiche für die Lehrpersonen am interessantesten waren (Schüler-, Klassen- oder Landesvergleich).

Teil B: Reflexion – Klassenebene

Im zweiten Teil des Leitfadens steht die professionelle Nutzung der Rückmeldungen auf Klassenebene im Vordergrund der Untersuchung („Reflexion“ und „Aktion“). Neben den drei Bereichen Förderdiagnostik, Selektionsdiagnostik und zukünftige Unterrichtsgestaltung werden in einem vierten Bereich Daten erhoben, wie die Lehrpersonen die Ergebnisse aus den Vergleichsarbeiten für die Eltern- und Schülerkommunikation verwenden.

1. **Förderdiagnostische Hinweise durch Vergleichsarbeit:** Hierunter fallen Fragen, die darüber Auskunft geben sollen, ob die Lehrpersonen durch die Rückmeldung Anhaltspunkte für Stärken und Schwächen der Schüler erhielten und ob sich daraus Fördermaßnahmen planen und ableiten ließen. Außerdem wurde gefragt, ob die Informationen an nachfolgende Kollegen weitergegeben werden oder ob sie als Argumentationsbasis für Beratungsgespräche mit Eltern verwendet werden.
2. **Selektionsdiagnostische Hinweise:** Bei diesem Punkt wird gefragt, ob die Vergleichsarbeit einen Beitrag zur Begründung der Jahresnote liefert oder ob sie bei unsicheren Versetzungsentscheidungen ein zusätzliches Entscheidungskriterium darstellt. Des Weiteren sollten die Lehrpersonen angeregt werden, sich mit ihrer eigenen Benotungs- und Bewertungspraxis auseinanderzusetzen.
3. **Hinweise für die zukünftige Unterrichtsgestaltung:** Setzt man sich mit den möglichen Anwendungsgebieten der Rückmeldungen aus Vergleichsarbeiten für die zukünftige Unterrichtsgestaltung auseinander, so steht der Lehrperson eine Vielzahl an Gestaltungsmöglichkeiten zur Verfügung: Sie könnten zum Beispiel aufgrund der Rückmeldungen die Auswahl der Unterrichtsinhalte abändern, mehr Wiederholungen einfügen oder eine Neuanordnung der Stoffgebiete vornehmen. Es besteht auch die Möglichkeit neue Unterrichtsmethoden oder Lernmaterialien für die Schüler zu verwenden.
4. **Eltern- und Schülerkommunikation über Ergebnisse aus Vergleichsarbeiten:** Aus den Erfahrungen mit der Befragung aus dem Jahre 2006 zeigte es sich dass es vorteilhaft ist, diesen Punkt bereits im Vorfeld als geschlossene Aufgabe zu formulieren. In einer Multiple-Choice Matrix wurden den Personen alle möglichen Kommunikationspartner dargeboten. Somit konnten die Befragten aus einer Vielzahl an möglichen Items ihr individuelles Vorgehen bei der Kommunikation mit Eltern und Schülern markieren. Unter forschungsökonomischem Aspekt wird die Auswertung der Interviews dem Forscher dadurch vereinfacht und erleichtert.

Teil C: Reflexion - Schulebene

Der Umgang mit Rückmeldungen aus Vergleichsarbeiten auf der Ebene der Schule steht im Mittelpunkt des Frageteils C. Ermittelt werden einerseits die Informationsabläufe und -wege, wie die Ergebnisse aus Vergleichsarbeiten auf Schulebene kommuniziert werden. Andererseits steht die Frage im Zentrum, ob Diskussionen über die Ergebnisse der Vergleichsarbeiten und die Einbindung der Ergebnisse in die Evaluationsverfahren der Schulen. Diese Themen lassen sich im Modell dem Bereich „schulischen Bedingungen und Kontextbedingungen zuordnen (Forschungsfragen 10 – 11).

1. **Information über die VA-Ergebnisse auf Schulebene**: Hier geht es darum, ob die Einsichtnahme in die VA-Ergebnisse festgelegt wurde und wer sie festgelegt hat. Zwei weitere Unterpunkte beschäftigen sich mit der Frage, wer an der Schule den Klassendurchschnitt bzw. alle Schülerergebnisse einsehen darf. Analog zum Fragebereich B4 wurden den Befragten geschlossene Aufgaben zum Ankreuzen vorgelegt.
2. **Diskussion über die Ergebnisse der Vergleichsarbeit:** Im Zentrum dieser Fragestellung steht die Tiefe der Auseinandersetzung über die Ergebnisse der Vergleichsarbeit. Wieder erhielt der Befragte eine geschlossene Aufgabenmatrix, in der angekreuzt werden konnte mit welchen Personen bzw. in welchen Gremien (Parallelkollegen, Schulkonferenz…) und in welcher Tiefe die Diskussionen (keine Gespräche, informeller Austausche, systematische Diskussion) stattfanden.
3. **Einbindung der VA-Ergebnisse in die Selbstevaluation der Schule:** In dem letzten Punkt des Interviews geht es um den bisherigen Stand der Entwicklung eines Evaluationskonzeptes der Schule, das seit der Bildungsplanreform 2004 als verpflichtendes Modul von jeder Schule umgesetzt werden muss. Wenn dieses bereits existiert, dann sollen die Lehrer Erhebungsinstrumente und Daten nennen, die dazu verwendet werden. Die Frage, ob die Daten der Vergleichsarbeiten auch für die Selbstevaluation genutzt werden, bildet den Abschluss des Interviews.

4.3.2 Durchführung der Interviews - Datenerhebung und Transkription

4.3.2.1 Vorbereitung und Durchführung der Interviews

Die Interviews wurden wie in Kapitel 4.2.1.2 beschrieben in allen drei Schulformen der Sekundarstufe I des baden-württembergischen allgemein bildenden Schulsystems durchgeführt. Die Befragungen wurden jeweils an den Schulen durchgeführt. Einerseits um ein vergleichbares Setting zu gewährleisten, andererseits konnten alle Unterlagen der Vergleichsarbeiten an der Schule archiviert und bei Bedarf herangezogen werden. Außerdem ließen sich alle Interviews, die an einer Schule geführt wurden, an einem oder zwei Tagen organisieren. Schließlich hatte dieses Vorgehen zur Folge, dass sich alle Aspekte der Befragung auf dieses Umfeld bezogen. Des Weiteren war mit diesem Vorgehen die Vorstellung verbunden, dass die Äußerungen der Befragten in dieser authentischen, realen und offenen Befragungsatmosphäre deren subjektive Sichtweise und Einstellung zu den Vergleichsarbeiten widerspiegelt.

Für die Dauer des Interviews wurde höchstens eine Schulstunde, also 45 min anberaumt. Die Interviews konnten somit auch in einer Hohlstunde oder Randstunde durchgeführt werden und brachten keine außerordentliche zeitliche Belastung für die Lehrpersonen mit sich.

Im Anschluss an jedes Interview wurden die Fragebögen eingesammelt und für die Auswertung in einem Ordner abgeheftet. Dadurch liegt von jeder befragten Person sowohl der Kurzfragebogen als Teil des problemzentrierten Interviews vor als auch die quantitativen Befragungselemente, die den Interviewten während des Interviews im multiple choice Format vorgelegt wurden. Teilweise wurden erste Eindrücke des Interviews als post scriptum kurz notiert.

Die Interviews wurden von mehreren Personen der Pädagogischen Hochschule Schwäbisch Gmünd durchgeführt. Vor der Erhebung war eine intensive Vorbereitung der Interviewer notwendig, worüber die folgenden Abschnitte berichten:

Anforderungen an die Interviewer

Die Anforderungen an den Interviewer sind bei qualitativen Verfahren höher als bei quantitativen. Denn er nimmt keine distanzierte Rolle ein wie in standardisierten, geschlossenen Interviews, sondern ist im Gespräch eher engagiert, wohlwollend und emotional beteiligt mit einem ehrlichen Interesse an den Äußerungen des Gesprächspartners. Dadurch wird er selbst zum „Erhebungsinstrument“ (vgl. Bortz und Döring 2006, S. 308). Er übernimmt eine Doppelrolle im Erhebungsprozess: während er im Interview eine empathischen Haltung einnimmt muss er sich als Interviewer selbst zurücknehmen, um sich ganz auf den Gesprächspartner zu konzentrieren und um die Fragestellung nicht aus den Augen zu verlieren. Er sollte die Technik der offenen Fragestellung beherrschen und einen guten Überblick über den Leitfaden und dessen theoretischen Hintergrund besitzen, um gezielt nachfragen zu können (vgl. Hermanns 2003, S. 364).

Sind mehrere Interviewer tätig, so müssen diese im Vorfeld der Datenerhebung intensiv geschult werden. Nur so werden die Qualität der Interviews und die Durchführungsobjektivität gesichert. Im vorliegenden Forschungsprojekt waren mehrere Personen an der Erhebung beteiligt. Im Rahmen eines Seminars zur Interviewerschulung, die zwei Semesterwochenstunden umfasste, wurden diese Personen auf ihre Rolle als Interviewer vorbereitet.

Die Schulung enthielt folgende Elemente (vgl. Bortz & Döring 2006, S. 247 f):

1. **Vermittlung inhaltlicher Kenntnisse:** Intensive Auseinandersetzung mit den theoretischen Grundlagen zur Funktion, Zielsetzung und Durchführung von Vergleichsarbeiten. Insbesondere wurde die Durchführung der baden-württembergischen Vergleichsarbeiten der Sekundarstufe I beleuchtet. Durch diese weitreichende Informiertheit ist es den Interviewern während des Interviews möglich, auf Rückfragen professionell und kompetent antworten zu können.
2. **Aufbau des Fragebogens:** Die funktionale und inhaltliche Struktur des Fragebogens muss dem Interviewer vertraut sein. Da die Interviews möglichst offen gestaltet sind muss davon ausgegangen werden, dass die Antworten sich nicht stringent und sukzessive dem Fragebogen entlang zuordnen lassen. Daher ist es erforderlich, dass der Interviewer während des Gesprächs maximal flexibel ist und sich dem Redefluss seines Gesprächspartners anpasst. Um alle Fragenbereiche zu erheben, sollte der Interviewer geschickt auf fehlende Themen eingehen.

3. **Dokumentation der Antworten:** Die Interviews wurden mit digitalen Aufnahmegeräten aufgezeichnet und standen als mp3-Dateien zur Transkription zur Verfügung. Die Interviewenden lernten in der Schulung den Umgang mit den technischen Geräten, so dass ein reibungsloser Interviewverlauf gewährleistet ist.
4. **Verweigerungen:** Ein Element der Schulung befasste sich mit dem professionellen Verhalten der Interviewer in schwierigen Interviewsituationen. Es wurden einige Standardregeln vermittelt, wie z.B. mit Antwortverweigerungen umgegangen werden kann, indem die Frage zu einem späteren Zeitpunkt nochmals gestellt wird.
5. **Probeinterviews:** Um die Rolle des Interviewers zu verinnerlichen wurden Probeinterviews durchgeführt, die von den Anwesenden kritisch beobachtet, analysiert und diskutiert wurden. In die fiktiven Interviews wurden spezifische Schwierigkeiten und Probleme eingebaut (zu schnelles Reden, lange Schweigephasen etc.).

4.3.2.2 Transkription der Interviews

Transkription (lat. trans-scribere = umschreiben) bedeutet das Übertragen einer Audio- oder Videoaufnahme in eine schriftliche Form. Mittlerweile steht im Internet für die Transkription von Interviews kostenfreie Transkriptionssoftware zum Download zur Verfügung. In diesem Forschungsprojekt wurde mit der Software f4 transkribiert[22].

Transkriptionsregeln

Jede wissenschaftliche Transkription erfordert ein festes Regelsystem - Transkriptionsregeln. Eine ausführliche Auseinandersetzung mit den Grundlagen und Regeln des Transkribierens findet sich im Handbuch Qualitative Forschung (vgl. Friebertshäuser und Langer 2010) im Beitrag von Antje Langer (2010).

Zu jedem Transkript muss ein Transkriptkopf mit dem Postskriptum des Interviews, der den Kontext des Gesprächs festhält, erstellt werden. Dazu zählen neben Allgemeinen Angaben zur Aufnahme des Gesprächs (Pseudonym, Datum, Dauer und Ort der Aufnahme) eventuelle Angaben zur interviewten Person und die Schilderung der Interviewsituation. Die Interviews wurden mit einem dreistelligen Zahlencode anonymisiert: Die erste Ziffer (Hunderter) gibt Auskunft über die

[22] (http://www.audiotranskription.de/f4.htm)

Schulart. Hauptschulen wurden mit 100, Realschulen mit 200 und Gymnasien mit 300 codiert. Die einzelnen Interviews wurden durchnummeriert, in jeder Schule mit der Zahl 50 beginnend. Die hohe Ausgangszahl für das erste Interview der Teilstichproben (150, 250, 350) liegt darin begründet, dass im Schuljahr 2005/2006 bereits eine erste Stichprobenerhebung stattfand, die im gleichen System beginnend mit 100, 200, 300 codiert ist. Diese Stichprobe wurde in der vorliegenden Arbeit nicht berücksichtigt. Es liegen demnach für die jeweilige Schulart folgende Interviews vor: Hauptschulen: Nr. 150 – Nr. 164; Realschulen: Nr. 250 – Nr. 272, Gymnasien; Nr. 350 – Nr. 370

Das eigentliche Interview- bzw. Gesprächstranskript wird dergestalt verfasst, dass es mit fortlaufenden Zeilennummern versehen wird, um die entsprechenden Passagen schnell wiederzufinden. Danach wird der jeweilige Sprecher mit Kürzel gekennzeichnet (PH – für den Interviewer, I – für die interviewte Lehrperson). Die folgende Tabelle gibt die wichtigsten Transkriptionsregeln wieder, die in diesem Projekt zur Anwendung kamen:

Tabelle 4.3.1: Transkriptionsregeln nach Langer (2010)

( )	Unverständliche Passage; die Länge der Klammer entspricht etwa der Dauer
(schwer zu verstehen)	Unsichere Transkription; vermutete Äußerung in der Klammer
(.)	Sehr kurze Pause
(3)	Pause in Sekunden
LAUT	Laut gesprochen
’leise’	Leise gesprochen
Betont	Betont gesprochen
g e d e h n t	Gedehnt gesprochen
((lacht))	Para- oder nonverbaler Akt, steht vor der entsprechender Stelle
Da sagt der: „Komm her“	Zitat innerhalb der Rede
gegan-	Wortabbruch
[	Überlappung von Redebeiträgen bzw. direkter Redeanschluss, wird hervorgehoben, indem genau an der Stelle weitergeschrieben wird

4.4 Datenauswertung

4.4.1 Auswertung der Interviews mit der Methode der Qualitativen Inhaltsanalyse

Im Bereich der qualitativen Auswertungstechniken gibt es zahlreiche Varianten, die für die Analyse von Interviews angewendet werden können, zum Beispiel die Globalauswertung nach Legewie (1994), den Grounded Theory-Ansatz nach Glaser und Strauss (1967) und die Qualitative Inhaltsanalyse nach Mayring (2003, 2008; Gahleitner 2005). In der vorliegenden Arbeit wird die Methode der Qualitativen Inhaltsanalyse nach Mayring angewandt, da diese für die systematische Analyse von Texten konzipiert wurde und daher besonders geeignet erschien. Bei dieser Auswertungstechnik wird am Datenmaterial theoriegeleitet ein Kategoriensystem entwickelt. Außerdem bietet die Qualitative Inhaltsanalyse je nach Ausgangsmaterial zusätzlich verschiedene Grundformen qualitativer Techniken an (siehe unten), die passend zum Material ausgewählt werden können.

4.4.1.1 Grundlagen der Qualitativen Inhaltsanalyse nach Mayring

Mayrings Anliegen ist es, mit der Qualitativen Inhaltsanalyse ein Interpretationsverfahren zu entwickeln, das die einzelnen Analyseschritte und Analyseregeln systematisiert und überprüfbar macht. Bei dem Verfahren der Qualitativen Inhaltsanalyse sollen folgende Grundsätze befolgt werden (vgl. Mayring 2003; Steinke 2003):

- **Einbettung des Materials in den Kommunikationszusammenhang:** Der Text steht nicht isoliert da, sondern soll innerhalb seines Kontextes interpretiert werden. „Die Inhaltsanalyse will nie etwas über den Text an sich aussagen (…), sondern vom Text Schlussfolgerungen auf andere Bestandteile des umfassenden Kommunikationsmodells ziehen“ (Mayring & Brunner 2010, S. 325). Bei Interviews also über die Interviewpartner.

- **Strenges systematisches und regelgeleitetes Vorgehen:** Die einzelnen Schritte, die für die Textanalyse angewendet werden, sollen vorab in einem konkreten Ablaufmodell festgelegt werden. Dieses Ablaufmodell soll eng am Ausgangsmaterial entwickelt und festgelegt werden.
- **Kategorien stehen im Zentrum der Analyse:** Die Kategorie bzw. das Kategoriensystem bildet das zentrale Instrument der Analyse. Besondere Bedeutung erfährt die Konstruktion und Begründung der Kategorien, die präzise erläutert werden muss und somit intersubjektiv nachvollziehbar wird. Dieses Kategoriensystem wird auf den Text angewandt, indem es wie ein Rechen durch das Material gezogen wird und entsprechende Textstellen selektiert. Es steht also nicht die ganzheitliche Erfassung des Materials im Vordergrund.
- **Gegenstandsbezug statt Technik:** Nicht das technische Verfahren der Inhaltsanalyse steht im Mittelpunkt, sondern die Anbindung an den konkreten Gegenstand der Analyse. Die Wahl des Verfahrens muss vor Beginn festgelegt werden und wird dann an das Material herangetragen und dokumentiert.
- **Überprüfung der spezifischen Instrumente durch Pilot-Studien:** Die geplanten Verfahren sollen zuerst in Probedurchläufen an mehreren Beispielen angewendet werden. Dies ermöglicht eine Überprüfung und Optimierung des Vorgehens und stellt ein Mindestmaß an Standardisierung und Gewährleistung sicher.
- **Theoriegeleitetheit der Analyse:** In der Qualitativen Inhaltsanalyse orientieren sich alle Verfahrensentscheidungen am Stand der Forschung, der zur Begründung des Vorgehens herangezogen wird.
- **Einbezug quantitativer Analyseschritte:** Quantitative Analyseschritte können vor allem im Bereich von Häufigkeitsangaben mit qualitativen Schritten verknüpft werden.
- **Gütekriterien:** Die Interkoderreliabilität entspricht der eigentlichen Objektivität und stellt in der Qualitativen Inhaltsanalyse das wichtigste Gütekriterium dar. Hierbei müssen verschiedene Personen, die das gleiche Material analysieren, möglichst zu übereinstimmenden Ergebnissen kommen. Das Material sollte bereits in der Pilotphase auf dieses Gütekriterium hin untersucht werden, um mögliche Fehlerquellen zu beheben. Zur Überprüfung der Interkoderreliabilität können auch nur Materialausschnitte einer zweiten Person vorgelegt werden. Im Falle, dass vereinzelt keine Übereinstimmung herrscht, können mit dem Verfahren der konsensuellen Validierung im diskursiven Verfahren klare Kodierregeln festgelegt werden.

4.4.1.2 Grundformen Qualitativer Inhaltsanalyse

In der Literatur gibt es zahlreiche Interpretationsverfahren, mit ganz unterschiedlichen Charakteren (vgl. Schmidt 2010). Mayring (2008) erläutert diese in seinem Buch und listet in einer Tabelle 14 gängige Interpretationsarten auf, darunter z.B. Häufigkeitsanalysen, die Objektive Hermeneutik, Kontingenzanalysen usw. In einem weiteren Schritt arbeitet er deren jeweilige Besonderheit und Eignung heraus und zeigt auf, dass sich jedes der Verfahren einer der drei Grundformen der Qualitativen Inhaltsanalyse zuordnen lässt; dem Verfahren der Zusammenfassung, dem der Explikation oder dem der Strukturierung (ebd., S. 57). In der vorliegenden Arbeit werden als Auswertungsverfahren die Zusammenfassung und die Strukturierung verwendet, die im folgenden Abschnitt erläutert werden. Die Methode der Explikation, bei der unklare Textstellen im Zentrum der Analyse stehen und durch Hinzunahme erklärender Materialien interpretiert werden, kommt nicht zum Tragen. In der vorliegenden Arbeit waren die Aussagen der Lehrpersonen verständlich und es waren keine weiteren Quellen zur Klärung unklarer Passagen nötig.

Die Bildung der Kategorien bzw. Subkategorien kann auf zwei verschiedenen Wegen erfolgen. Entweder geht man fallvergleichend vor, indem zuerst auf der Ebene von Einzelfällen Textpassagen verglichen werden und die dadurch gefundenen Dimensionen und Ausprägungen in Subkategorien zusammengefasst werden. Im Anschluss werden diese dann mit den Subkategorien verglichen, die bei der Analyse der anderen Einzelfälle entstanden sind. Oder man wählt, wie in vorliegender Studie praktiziert, die thematisch vergleichende und fallübergreifende Analyse. Bei dieser Vorgehensweise wird zunächst die Kodierung des gesamten Datenmaterials vorgenommen. Im Anschluss daran wird für jede Kategorie das gesamte Textmaterial über alle Fälle hinweg extrahiert und untersucht. Diese Textausschnitte werden dann fallübergreifend und in einer Synopse zusammengestellt und analysiert (vgl. Kelle & Kluge 2010, S. 56).

Zusammenfassende Qualitative Inhaltsanalyse

Bei diesem inhaltsanalytischen Verfahren werden die Kategorien bzw. das Kategoriensystem aus dem Material heraus entwickelt und mit elaborierten Makrooperatoren der Zusammenfassung auf ein höheres Abstraktionsniveau aggregiert. Hierbei wird das Gesamtmaterial auf die wesentlichen Aussagen reduziert, die nur die wichtigsten Inhalte wiedergeben. Zur Reduktion können die Makrooperatoren Auslassung, Generalisation, Konstruktion, Integration, Selektion und Bündelung angewandt werden (vgl. Mayring & Brunner 2010). Der idealtypische Ablauf dieser Auswertungstechnik sieht sieben aufeinanderfolgende Schritte vor; Bestimmung der Analyseeinheiten, Paraphrasierung der inhaltstragenden Textstellen, Bestimmung des Abstraktionsniveaus, Reduktionsschritte, Zusammenstellung der neuen Aussagen. Das Verfahren steht in engem Zusammenhang mit der Induktiven Kategorienbildung.

Induktive Kategorienbildung

Diese Technik ist eine der wichtigsten qualitativen Analysetechniken (vgl. ebd., S. 327). Ihr liegt kein fertiges Kategoriensystem zugrunde, sondern dieses muss erst systematisch aus dem Textmaterial generiert werden. Dem gesamten Generierungsprozess liegt ein theoretisches Vorverständnis zugrunde, auf das sich die schrittweise durchgeführte Analyse bezieht. Wie bei der oben beschriebenen Zusammenfassung muss das Abstraktionsniveau der Kategorien festgelegt werden und kann bei Bedarf „nach oben“ gesetzt werden.

In der vorliegenden Arbeit werden dann diese beiden Techniken zur Auswertung einzelner Fragestellungen verwendet, wenn im Vorfeld der Analyse keine eindeutige Kategoriendefinition erfolgen kann. Liegt ein theoretisch fundiertes Kategoriensystem vor, so wird damit das Material deduktiv, strukturierend ausgewertet.

Strukturierende Qualitative Inhaltsanalyse und deduktive Kategorienbildung

So wie die induktive Kategorienbildung eng mit der Technik der Zusammenfassung verflochten ist, so bilden auch die Technik der Strukturierung mit dem Verfahren der deduktiven Kategorienbildung ein Paar. Bei diesem Verfahren wird das Kategoriensystem bereits vor der Analyse theoriegeleitet entwickelt und mit klaren, verständlichen Kodierregeln versehen. In einem zweiten Schritt wird das Material mittels des bestehenden Kategoriensystems ausgewertet. Ankerbeispiele präzisieren die Kategorien, indem prototypische Aussagen gesammelt werden und als Musterbeispiel dienen. Weiter lässt sich die Strukturierende Qualitative Inhaltsanalyse in vier Untergruppen ausdifferenzieren, nach einem formalen, inhaltlichen, typisierenden oder skalierenden Format (vgl. Mayring 2008, S. 59 ff).

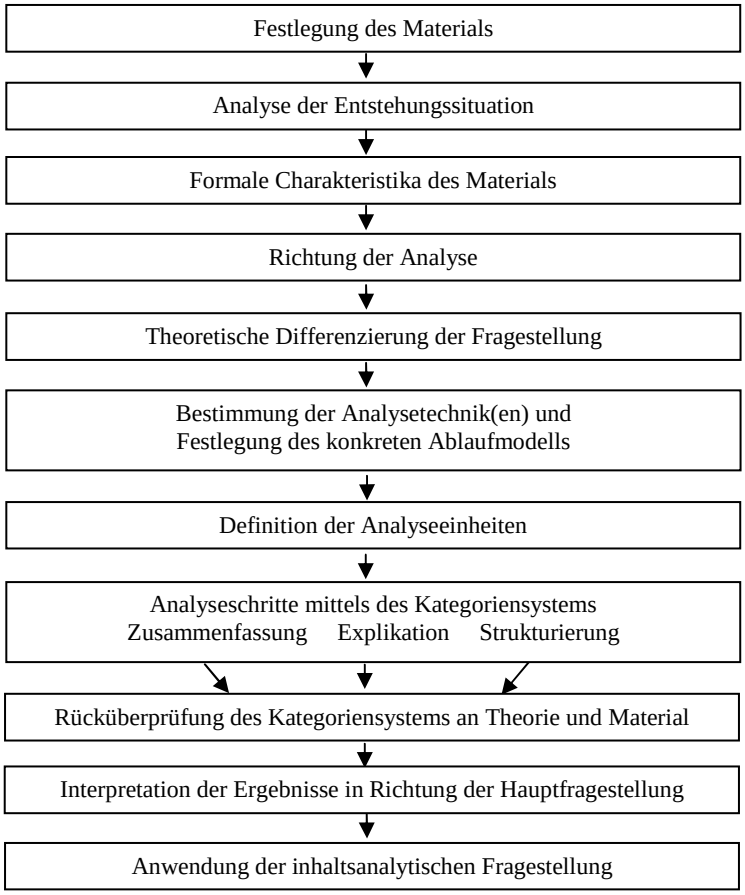
Welche Technik jeweils zur Auswertung des Textmaterials herangezogen wird, muss je nach Forschungsdesign entschieden werden. Es ist auch die Verwendung unterschiedlicher Verfahren denkbar, allerdings sollte im Sinne der Nachvollziehbarkeit und Regelgeleitetheit jede Interpretationstechnik auf das gesamte Material bzw. den jeweiligen thematischen Abschnitt immer abgeschlossen werden, bevor ein neuer Analyseschritt begonnen wird.

Bevor die Auswertung und Interpretation des Textmaterials durchgeführt wird, muss vorab im Ablauf der Analyse festgelegt werden, welche einzelnen Auswertungsschritte aufeinander folgen. Dazu hat Mayring sowohl ein allgemeines Ablaufmodell, als auch spezifische Modelle entwickelt, die auf die jeweilige Grundform der Interpretation passend zugeschnitten sind.

4.4.1.3 Allgemeines Ablaufmodell der Inhaltsanalyse

Das Ablaufmodell einer Qualitativen Inhaltsanalyse ist bei Mayring sehr komplex und besteht aus einer Vielzahl von aufeinander folgenden Schritten, die vom Planungsbeginn, über die Auswahl der Technik bis zur Überprüfung und Bewertung des Verfahrens reichen. Die Festlegung der einzelnen Interpretationsschritte muss vor der Analyse festgelegt werden, da sie somit für andere nachvollziehbar und der Prozess intersubjektiv überprüfbar wird. Für jede Inhaltsanalyse muss je nach Material und Fragestellung ein eigenes Ablaufmodell erstellt werden. Mayring (2000) stellt fest, dass die Inhaltsanalyse kein Standardinstrument sei, das immer gleich aussieht; sie muss an den konkreten Gegenstand, das Material angepasst sein und auf die spezifische Fragestellung hin konstruiert werden. Ein Beispiel für ein allgemeines inhaltsanalytisches Ablaufmodell zeigt die folgende Abbildung:

Abbildung 9: Allgemeines inhaltsanalytisches Ablaufmodell nach Mayring (2008, S. 54)

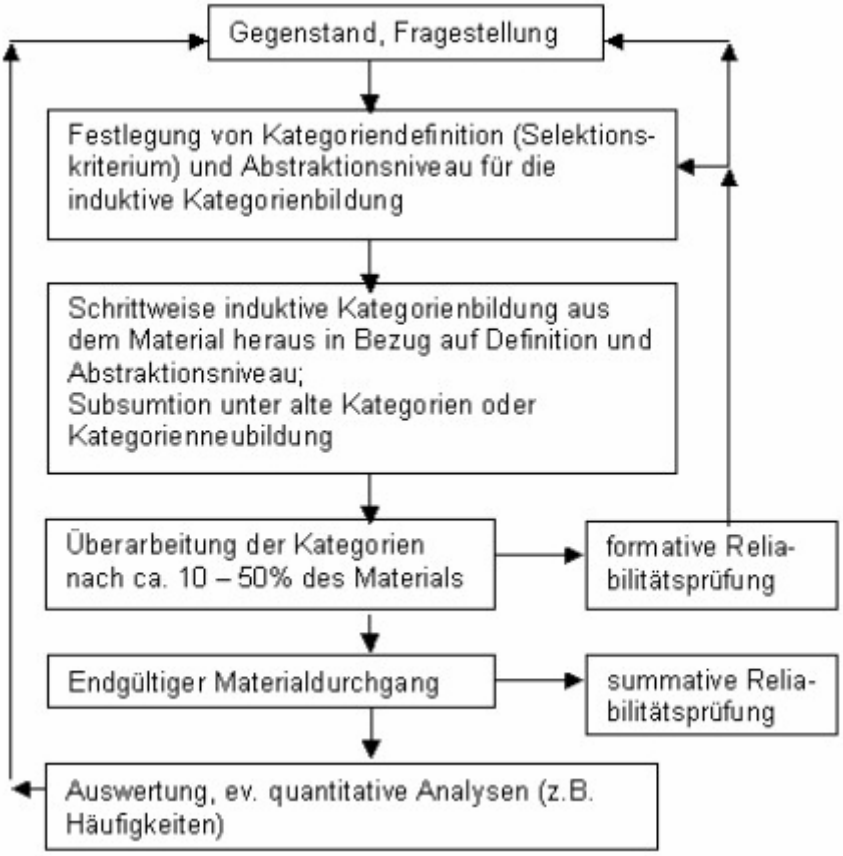


Die Festlegung, welche der drei Grundformen der Datenanalyse gewählt wird erfolgt in Schritt sechs des allgemeinen Ablaufmodells: „Bestimmung der Analysetechnik(en) und Festlegung des konkreten Ablaufmodells“.

In dieser Forschungsarbeit werden Kategorien einerseits induktiv, mit dem Verfahren der Zusammenfassung, und andererseits deduktiv, mittels der Strukturierung gewonnen. Im folgenden Kapitel wird die Grundform der Zusammenfassung am konkreten Material an einer Fragestellung exemplarisch erläutert und dargestellt. Die erste Frage, die den Lehrpersonen im Leitfadeninterview gestellt wurde, war die Frage nach der Funktion von Vergleichsarbeiten. Sie lautete konkret: „Was denken Sie, aus welchem Grund werden in Baden-Württemberg Vergleichsarbeiten durchgeführt?“ Die Antworten der 59 Lehrpersonen fielen ganz unterschiedlich aus und konnten nicht vorab deduktiv aus der Theorie heraus abgeleitet und ans Material herangetragen werden. Um das ganze Antwortspektrum aller Befragten zu erfassen, stellt für die Analyse dieser Interviewabschnitte das induktive Verfahren das Mittel der Wahl dar. Zu Beginn muss das passende Ablaufmodell für die induktive Kategorienbildung erstellt werden:

4.4.1.4 Induktive Kategorienbildung – die Technik der Zusammenfassung

Abbildung 10: Ablaufmodell induktive Kategorienbildung



Die entsprechenden Antwortpassagen der Lehrpersonen zu der Fragestellung „Funktion von Vergleichsarbeiten (VA) werden beim induktiven Interpretationsverfahren zunächst wortwörtlich aus dem Text in eine Auswertungstabelle übertragen (Schritt 1 und 2). Außer den wortwörtlich übernommenen Textpassagen enthält die Tabelle Informationen, aus welchem Interview die Passage entnommen wurde (**Fall**), in welchem Absatz das Zitat zu finden ist (**Absatz**) und die Nummerierung (**Nr.**) der sukzessiv aufgenommenen Textstellen. Dann wird auf die entsprechenden Textstellen das Verfahren Zusammenfassung angewendet: In der Spalte **Generalisierung** wurde die Kernaussage formuliert und dann in eine passende Kategorie (**Reduktion**) überführt (siehe folgende Tabelle):

Tabelle 4.4.1: Induktive Kategorienbildung am Beispiel "Funktionen der VA"

Fall	Absatz	Nr	Paraphrase	Generalisierung	Reduktion = Kategorien (K)[23]
150	3	1	Ich war schon immer dafür, dass Klassen möglicherweise auch Lehrer, dass was man als Lehrer leistet, vergleichbar gemacht werden muss, oder vergleichbar sein kann.	Vergleichbarkeit von Klassen, Lehrern	**K1**: Vergleichsarbeiten (VA) dienen landesweitem Vergleich • von Schulen, Klassen, Schülerinnen und Schülern, Lehrern… • um Standards zu überprüfen und zu sichern • als Rückmeldung für Verbesserungen **K2**: Vergleichsarbeiten (VA) sind ein Kontrollinstrument und Druckmittel **K3**: Vergleichsarbeiten (VA) als Folge der PISA-Studie **K4**: Vergleichsarbeiten (VA) haben keine Funktion **K5**: Vergleichsarbeiten (VA) sind purer Aktionismus
150	5	2	Natürlich um eine Vergleichbarkeit zu erreichen, um zumindest von der Statistik her festzustellen, wie verteilen sich die vielleicht leistungsstärkeren, leistungsschwächeren Schüler, Personen übers Land.	Landesweite Vergleichbarkeit von Schülerleistungen	
150	7-7	3	Dass das eben eine Vergleichsarbeit ist, das Land möchte damit vergleichen können und damit fertig.	Vergleichbarkeit erreichen	
150	5-5	4	Man könnte jetzt hinterhältig sein und sagen sie wollen den Lehrern eine reinwürgen. Sie wollen die Lehrer überprüfen, wie faul denn die Beamten so sind.	Lehrer ärgern, indem man ihre Leistung kontrolliert, weil Beamte faul sind	
151	3-3	5	Grundsätzlich denke ich, dass es schon sinnvoll ist, im Zuge des neuen Bildungsplans hier so eine Vergleichsarbeit zu machen, um mal einen Stand festzustellen.	Vergleich, um Leistungsstand zu ermitteln	
151	3-3	6	Gleichzeitig stellt sich für mich natürlich die Frage, in welcher Form das macht. Ob das dann wirklich auch die angebrachte Form ist, um so was festzustellen.	Zweifelt an Form, wie verglichen wird.	
152	7-7	7	Ich als Lehrer find es eigentlich eine gute Möglichkeit, einfach mal eine vorgefertigte Form zu haben, wie ich die Schüler kontrollieren kann. Es ist also nicht unbedingt meine erstellte Arbeit, sondern einfach von außen eine herangebrachte Arbeit, um einfach mein Tun so ein bisschen vergleichen zu können mit (einfach) den Vorgaben, eigentlich, sag ich jetzt mal, vom Bildungsplan.	VA als gute Möglichkeit für externen Vergleich des eigenen Handelns	
152	13-13	8	Dass man einfach die Leistung des Sohnes/ der Tochter einfach vergleichen kann mit anderen Schulen, eventuell auch mit der Parallelklasse.	Vergleich von Schülern, Schulen und Parallelklassen	
152	23-23	9	Ich denke, so ist es die passende Form, dass man einfach, dass das Ministerium gewisse Vergleichsarbeiten vorgibt. Und dass man einfach die dann auch zu machen hat. Einfach, um auch den besseren Vergleich zu haben.	Für Vergleichbarkeit	
153	3-3	10	Um Schülerleistung vergleichen zu können.	Vergleichbarkeit	
153	3-3	11	Auch als Hilfe für die Lehrer, um zu sehen, ob die Klasse entsprechend vorbereitet ist, und um das Ziel der Hauptschule dann irgendwann mal gut erreichen zu können.	Vergleich dient der Rückmeldung für Lehrer für Verbesserungen	
153	122-123	12	Haben Sie da Angst, ein Stück weit kontrolliert zu werden, persönlich, Ihre Arbeit? I: Dann, wenn das Ergebnis schlecht gewesen wäre, dann würde ich diese Angst haben.	Angst vor Kontrolle, wenn Ergebnisse schlecht ausfallen	
154	36-36	13	Es hat Stress verursacht, also auch hier an der Schule. Das habe ich mit Erstaunen beobachtet.	VA verursachen Stress	
154	34-35	14	Es gibt Lehrer die das als Kontrolle sehen.	VA als Kontrollinstrument	

[23] In der Spalte „Reduktion“ werden alle Kategorien aufgeführt, die aus den Kernaussagen (Spalte „Generalisierungen“) gebildet werden konnten. Wiederholungen werden durchgestrichen, neue Kernaussagen formulieren eine neue Kategorie.

Fall	Absatz	Nr	Paraphrase	Generalisierung	Reduktion
154	34-35	15	Ich sehe das relativ, ich sehe das gelassener. Ich fühle mich deswegen nicht kontrolliert. Das jetzt auf Grund der Ergebnisse Rückschlüsse auf meinen Unterricht geschlossen werden. Da habe ich keine Bedenken.	Fühlt sich selbst nicht kontrolliert	Kategorien stehen auf der ersten Seite der Tabelle
154	48-48	16	Da habe ich den Druck in der Schule als unmenschlich empfunden, wie die nur auf dieses Arbeiten hin getrimmt worden sind. Da sehe ich diese Arbeiten, ob Vergleichsarbeit oder Diagnosearbeit oder sonst was als sehr störend für den Schulbetrieb an, weil sie eben nicht als maßische Prüfung angesehen werden, Überprüfung, sondern die stören wirklich den Betrieb.	VA als Druckmittel, die Schulbetrieb stören	
154	49-50	17	I: Aber das zeigt ja dann schon, dass die meisten dann die Arbeiten definitiv als Kontrolle angesehen haben. L: Ja, denke ich schon auch. In der Schule läuft das auch über die Schulleitung so	VA als Kontrollinstrument, auch über Schulleitung	
154	54-54	18	Aber es kann einen Druck aufbauen, also einfach ein Leistungsdruck. Und das sehe ich schon als negativ.	VA als negatives Druckmittel	
155	26-26	19	L: Da sage ich nur Pisa dazu.	Als Folge von PISA	
155	30-32	20	L: Ich denke, Kontrolle auf jeden Fall, denke ich schon. I: Von den Lehrern dann? L: Nein, jetzt einfach vom Ministerium.	VA als Kontrolle des Ministeriums	
155	32-32	21	L: Nein, jetzt einfach vom Ministerium. Die bekommen die Ergebnisse ja irgendwann mal mitgeteilt und ja, weil sie einfach unter Druck sind. Gerade das Land Deutschland steht da ja unter großen Zwängen und Druck, aber auch als Übersicht dann einfach	Als Folge von PISA, Ministerium steht unter Druck	
158	10-10	22	L. Also eine spontane Antwort, die mir dazu einfällt, dass es eine Beschäftigungstherapie für Lehrer. Und da sind irgendwelche Statistikfanatiker drinnen im Club ja? Also einen vernunftbegabten Grund kann ich mir da nicht denken.	Keine Funktion, nur Beschäftigungstherapie	
⋮					
256		23	Aber das kam für mich eben auch durch die PISA-Studie und alle möglichen anderen Studien - als da die großen Lücken aufkamen und dass da eben gedacht wurde, da muss man etwas tun…	Schulleistungsstudien (PISA) als Ursache	
256		24	Die Vergleichbarkeit ist dann ja auch ganz wichtig, damit man sieht, wo sind die Lücken groß und klein	Vergleichbarkeit	
260	26-26	25	Kontrolle von Schülern und Lehrern	Kontrollmittel	
260	26-26	26	Und ja Aktionismus. Das ist in meinen Augen purer Aktionismus.	VA als purer Aktionismus	
260	60-61	27	Das Kreuzle da eintippen ist Beschäftigungstherapie.	Beschäftigungstherapie	

⋮

Die oben dargestellte zweiseitige Auswertungstabelle enthält die Antwortpassagen zu einem Fragebereich, nämlich nach der Funktion der Vergleichsarbeiten. Es wurden zunächst neun von insgesamt 59 Interviews ausgewertet, was ca. 15 % des Gesamtmaterials entspricht. An dieser Stelle des Auswertungsprozesses ist es sinnvoll die Kategorienbildung von mehreren Personen überprüfen zu lassen, bevor das gesamte Material interpretiert wird. Mit der Berechnung der Intercoderreliabilität wird die Übereinstimmung zwischen unterschiedlichen Personen berechnet, die dasselbe Material unabhängig voneinander kodieren. Ein Übereinstimmungskoeffizient gibt Auskunft darüber, in welchem Maße „einheitlich“ kodiert wurde. Sollte es in einigen Fällen zu keiner Übereinstimmung kommen, so werden von den Beteiligten mit Hilfe der konsensuellen Validierung klare Kodierregeln formuliert, die eine eindeutige Zuordnung der entsprechenden Textstellen gewährleisten. In qualitativen Auswertungs- und Interpretationsverfahren stellen diese Konsenskriterien wichtige Gütekriterien dar (siehe Kap. 4.4.1.1.).

4.4.2 Computergestützte Inhaltsanalyse und Darstellung der Ergebnisse

Würde man die Auswertung des Gesamtmaterials nach dem oben genannten tabellarischen Verfahren vornehmen, würde man vermutlich einen so umfangreichen und unübersichtlichen Anhang produzieren, dass die Nachvollziehbarkeit und intersubjektive Überprüfung der Auswertung nicht mehr gegeben wäre. Für die Interpretation von Interviews aus großen Stichproben empfiehlt sich ein pragmatisches Vorgehen, z.B. die Verwendung einer computerunterstützten Auswertungssoftware MAXQDA®[24].

Qualitative Daten mit der Unterstützung von Computerprogrammen zu analysieren, ist eine relativ neue Entwicklung im Bereich der Forschungsmethodik. Die analytischen Möglichkeiten computergestützter Auswertungsverfahren (QDA-Software) reichen vom Datenmanagement und der Datenexploration bis hin zur Kategorienbildung, zur kategorienbasierten Auswertung und zu komplexen Analysemethoden (vgl. Kuckartz und Grunenberg 2010).

[24] MAXQDA® wurde von Udo Kuckartz entwickelt. Detaillierte Informationen finden sich im Internet unter http://www.maxqda.de/

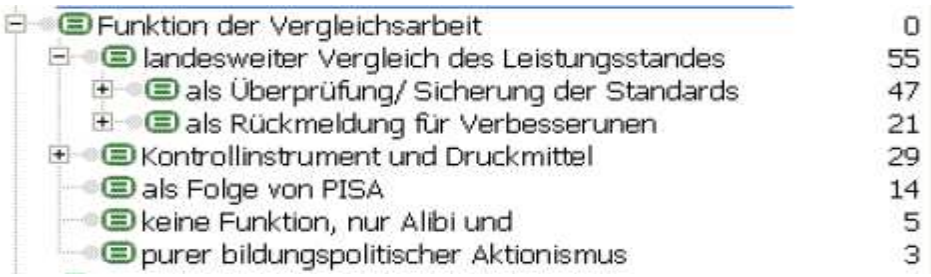
Alle Interviews wurden mit der „MAXQDA® -Software“ erfasst und für die Auswertung vorbereitet. MAXQDA® ist ein Programm, das Unterstützung bei Textanalysen bietet und die Übersichtlichkeit von umfangreichem Textmaterial erleichtert. Mit Hilfe dieses Programms können qualitative Datenanalysen systematisch durchgeführt werden. Voluminöses Datenmaterial kann mit vergleichsweise geringem Arbeitsaufwand strukturiert und in einer nachvollziehbaren Weise dargestellt werden.

Eine der zentralen Funktionen von MAXQDA® ist die Möglichkeit, Codes (Kategorien) zu Textpassagen zuzuordnen. Das Kategoriensystem ist jederzeit im Fenster CODESYSTEM zugänglich. Zu Beginn der Arbeit ist dieses Fenster noch leer. Das Kategoriensystem entsteht sukzessive, während der regelgeleiteten Durcharbeitung des Materials (induktive Kategorienbildung). Es kann beliebig ausdifferenziert und gestaltet werden, je nachdem, ob neue Aspekte aus dem Material heraus entwickelt werden (neue Kategorie anlegen) oder eine Textpassage einer bereits bestehende Kategorie zugeordnet werden kann.

Auf dem Bildschirm wird das Codesystem als Baumstruktur dargestellt. Es ähnelt der bekannten Darstellung der Dateiordner im Windows Explorer. Codes, die untergeordnete Subcodes besitzen, sind an einem kleinen Plus- bzw. Minuszeichen vor dem Code zu erkennen. Durch Klicken auf dieses Zeichen werden die Subkategorien aus- bzw. eingeblendet, wenn sie vorher nicht sichtbar waren.

Am Ende des Auswertungsprozesses entsteht ein komplexer und ausdifferenzierter Kategorienbaum inklusive Quantifizierungen. In dem oben aufgeführten Beispiel zur Frage nach der Funktion von Vergleichsarbeiten sieht der spezifische MAXQDA® - Kategorienbaum wie folgt aus:

Abbildung 11: Beispiel für Kategorienbaum zu „Funktion von Vergleichsarbeiten“



Dieser Kategorienbaum wird für die Präsentation der Ergebnisse (Kapitel 5) in einer Tabelle noch differenzierter dargestellt.

Tabelle 4.4.2: Beispiel für Präsentation der induktiv gewonnenen Ergebniskategorien

Individuelle Bedingungen				
Hauptkategorie I: Funktion der Vergleichsarbeiten (VA)				
Subkategorien	**Ges (n=59)**	**HS (n=15)**	**RS (n=23)**	**Gym (n=21)**
1. Vergleichsarbeiten (VA) dienen landesweitem Vergleich	55	14	22	19
Um Standards zu überprüfen und zu sichern	47	8	20	19
Als Rückmeldung für Verbesserungen	21	4	10	7
2. VA als Kontrollinstrument und Druckmittel	29	7	12	10
3. VA als Folge der PISA-Studie	14	3	4	7
4. VA haben keine Funktion	5	1	2	2
5. VA als purer Aktionismus	3	0	0	3

Der Tabelle können folgende Informationen entnommen werden:

Die erste Zeile bezieht sich auf die einzelnen Auswertungsschritte des zugrundeliegenden Modells von Helmke und Hosenfeld zur Rezeption und Nutzung von Daten aus Rückmeldungen (siehe Kapitel 3.4.2.2). Im oben dargestellten Beispiel handelt es sich demnach um Ergebnisse zu den Individuellen Bedingungen. Die zweite Zeile enthält den Titel der zu diesem Auswertungsschritt generierten Hauptkategorie und dessen Nummerierung. Die Hauptkategorien lassen sich in weitere Unter- bzw. Subkategorien gliedern, die am Ende des Quantitativen Inhaltsanalyseverfahrens gezählt und nach der Häufigkeit ihrer Nennungen gelistet werden. Hier gilt zu beachten, dass Mehrfachnennungen einer einzelnen Lehrperson innerhalb einer Subkategorie nur einfach gewertet werden. Wenn jedoch Antworten einer Lehrperson unterschiedlichen Subkategorien zugeordnet werden können, so fließen diese jeweils in die entsprechende Quantifizierung mit ein.

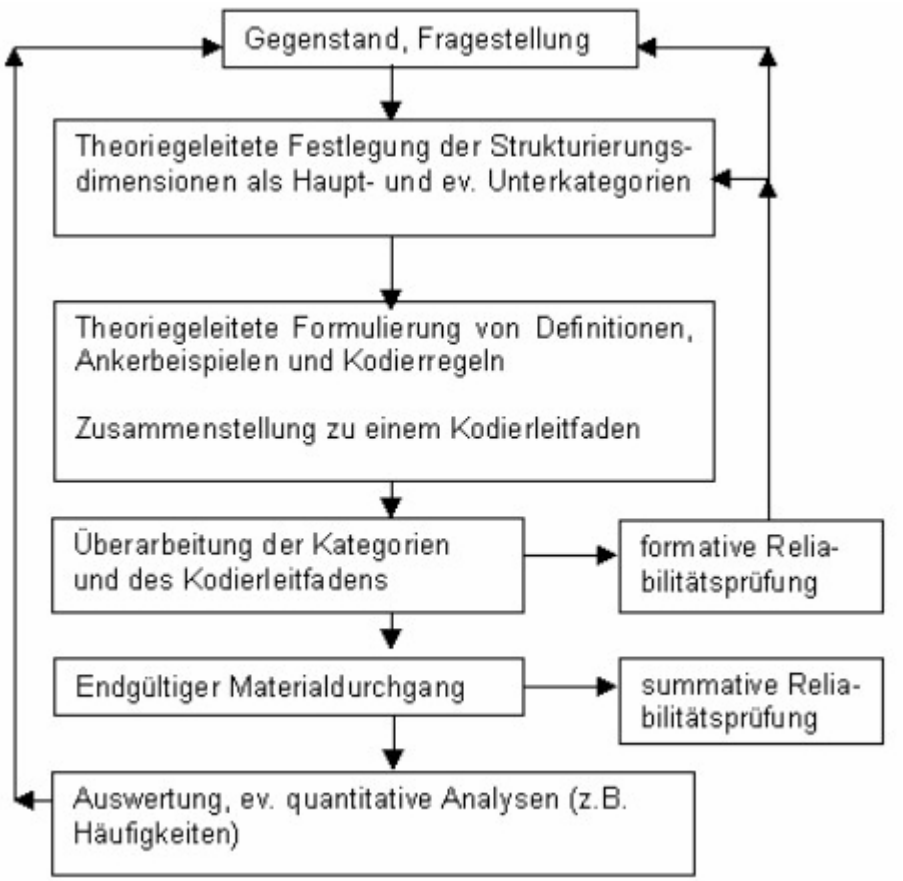
Die Spalte Ges (n=59) bezieht sich auf die Gesamtstichprobe des Materials. Sie beinhaltet die Interviewaussagen der 59 Lehrpersonen der Sekundarstufe I des allgemein bildenden baden-württembergischen Schulsystems zu den entsprechenden Subkategorien. In den drei weiteren Spalten (HS=15), (RS=23) und (Gym=21)

werden die Aussagen der Lehrerinnen und Lehrer den drei unterschiedlichen Schularten zugeordnet: es wurden also 15 Hauptschul-, 23 Realschul- und 21 Gymnasiallehrpersonen interviewt.

4.4.2.1 Deduktive Analyseschritte – die Technik der Strukturierung

Neben induktiven Verfahren werden bei der Auswertung des Gesamtmaterials vor allem deduktive Analysetechniken durchgeführt. Die Technik der Strukturierung stellt die zentralste Auswertungs- und Interpretationstechnik dar (Mayring 2008, S. 82). Auch bei diesem Verfahren muss vorab ein Ablaufmodell entwickelt werden:

Abbildung 12: Ablaufmodell deduktive Kategorienbildung



Am Beispiel der Fragestellungen zur Rezeption von Vergleichsarbeiten wird der Prozess der deduktiven Kategorienbildung demonstriert. Bereits vor der Analyse werden die Kategorien formuliert und definiert, mit einem Ankerbeispiel versehen und die Kodierregeln bestimmt. Auch dies geschieht im klassischen methodischen inhaltsanalytischen Vorgehen nach Mayring mit Hilfe einer Tabelle (vgl. Mayring 2008):

Tabelle 4.4.3: Beispiel für deduktive Kategorienbildung zu „Verständlichkeit der Rückmeldungen“

Kategorie	Definition	Ankerbeispiel	Kodierregeln
Rückmeldungen waren verständlich und unproblematisch.	Die Lehrpersonen erhalten Rückmeldungen über die Ergebnisse. Diese sind verständlich aufbereitet.	„Für mich waren die Auswertungsdaten verständlich und nachvollziehbar“ (150).	Die Rückmeldungen aus den Vergleichsarbeiten werden von den Lehrpersonen angeschaut und verstanden.
Rückmeldungen waren unverständlich.	Die Rückmeldungen wurden angeschaut aber es konnten keine Informationen daraus entnommen werden.	„Und dann muss ich sagen, mit dieser Tabelle hier, kann ich gar nichts anfangen, die sagt mir gar nichts“ (162).	Die Rückmeldungen werden angeschaut. Es können keine Informationen entnommen werden.
Rückmeldungen wurden nicht angeschaut.	Die Lehrpersonen verwenden die Rückmeldungen aus Vergleichsarbeiten nicht.	Ich habe sie mir bis jetzt noch nicht angeschaut. Da kann ich jetzt nichts dazu sagen“ (351).	Die Lehrpersonen schauen die Rückmeldungen aus Vergleichsarbeiten nicht an.

Bei der computerunterstützten Datenanalyse werden diese vorformulierten theoriegeleiteten Kategorien „im Codebaum angelegt“. Darauf folgt die entsprechende Zuordnung der Interviewabschnitte. Die Ergebnisdarstellung erfolgt analog zum induktiven Vorgehen in Form einer Tabelle. Das Beispiel bezieht sich auf die Rezeption der Ergebnisrückmeldungen, dem ersten Schritt im Modell von Helmke & Hosenfeld (2005).

Die Tabelle enthält die Ergebnisse zur Hauptkategorie II: „Verständlichkeit der Rückmeldungen“:

Tabelle 4.4.4: Beispiel für Präsentation der deduktiv gewonnenen Ergebniskategorien

Rezeption von Vergleichsarbeiten				
Hauptkategorie II: Verständlichkeit der Rückmeldungen				
Subkategorien	**Ges (n=59)**	**HS (n=15)**	**RS (n=23)**	**Gym (n=21)**
1. Rückmeldungen waren verständlich	16	2	6	8
2. Rückmeldungen wurden nicht angeschaut	13	3	6	4
3. Rückmeldungen waren unverständlich bzw. unpraktisch	5	2	3	0

4.4.2.2 Festlegung induktiver bzw. deduktiver Analyseverfahren

Welche Fragestellungen mit deduktiven bzw. induktiven Auswertungsverfahren ausgewertet wurden wird in den beiden folgenden Tabellen aufgeführt:

Tabelle 4.4.5: Übersicht über die Anwendung deduktiver Analyseverfahren bzgl. Fragestellungen

Deduktive Analyseverfahren	
Kategorien	**Subkategorien**
Durchführung der VA Rezeption (Frage 3)	keine Probleme; Zeit ausreichend/zu kurz/zu lang; Kritik an Durchführung
Verständlichkeit der VA Rezeption (Frage 4)	verständlich; nicht verständlich; nicht angeschaut
Zeitpunkt der VA Reflexion (Fragen 5 – 7)	zu spät; zu früh; Zeitpunkt verschieben
Niveau der VA Reflexion (Fragen 5 – 7)	Niveau angemessen, Niveau zu hoch, Niveau zu niedrig
Ableitung von Fördermaßnahmen Reflexion (Fragen 5 – 7)	Anhaltspunkte für Fördermaßnahmen erhalten; keine Hinweise erhalten; Fördermaßnahmen sind möglich; keine Fördermaßnahmen möglich

Tabelle 4.4.6: Übersicht über die Anwendung induktiver Analyseverfahren

Induktive Analyseverfahren	**Kategorien**
Individuelle Bedingungen (Fragen 1 und 2)	Bei induktiven Analyseverfahren werden die Kategorien aus dem Interviewmaterial entwickelt (vgl. Kap. 4.4.1.4).
Rezeption der Vergleichsarbeiten (Fragen 3 und 4)	
Weitere Reflexionen zu VA Reflexion (Fragen 5 – 7)	
Abgeleitete Aktionen (Forschungsfragen 8 und 9)	

4.4.2.3 Auswertungsverfahren der quantitativen Teilbefragungen zur Kommunikation der Ergebnisse

Um zu erfahren, in welcher Form das Thema Vergleichsarbeiten und deren Ergebnisse im schulischen Kontext diskutiert wird (Forschungsfragen 10 – 12), erhielten die Befragten während des Interviews folgenden Fragebogen zum Kommunikationsverhalten mit der Bitte, die entsprechenden Punkte anzukreuzen (Abbildung 13). Die Lehrpersonen hatten somit die Möglichkeit, die Qualität und Intensität der Gespräche zu bewerten. Unter dem Begriff ‚informeller Austausch' werden oberflächliche Gespräche verstanden, die beiläufig und nebenbei geführt wurden. ‚Systematische Diskussionen' beschreiben Dialoge, in denen sich die Gesprächspartner mit dem Thema Vergleichsarbeit intensiver und vertieft auseinandersetzen. Für die weitere Auswertung des schulischen Kommunikationsverhaltens war die rechte Spalte „systematische Diskussion" von Interesse.

Abbildung 13: Fragebogen zum schulischen Kommunikationsverhalten

C2) Diskussion über VA-Ergebnisse

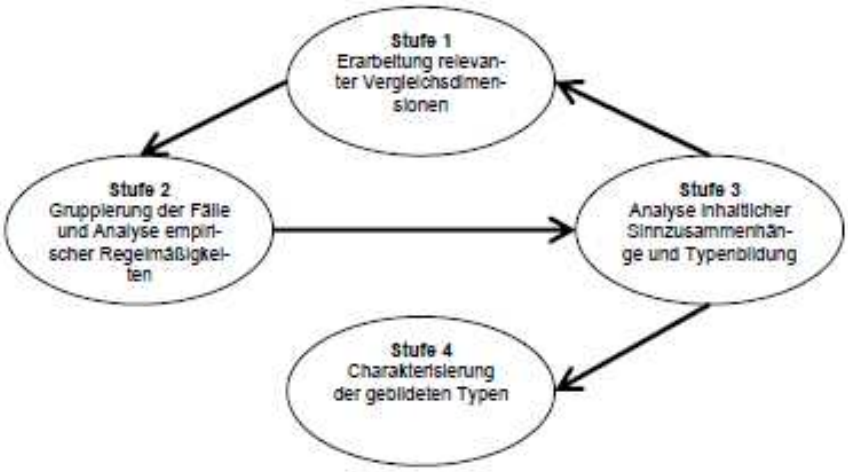
Wo wird an Ihrer Schule über die Ergebnisse der Vergleichsarbeiten diskutiert? (Mehrfachnennungen möglich)

	keine Gespräche	informeller Austausch	systematische Diskussion
Mit ausgewählten Kollegen.	○	○	○
mit Parallelkollegen	○	○	○
In der Klassenkonferenz	○	○	○
In der Jahrgangsstufenkonferenz	○	○	○
In der entsprechenden Fachkonferenz	○	○	○
Mit allen an den VA beteiligten Kollegen	○	○	○
In der Gesamtlehrerkonferenz.	○	○	○
In der Schulkonferenz	○	○	○
Mit der Schulleitung	○	○	○
Mit den Schülern dieser Klasse	○	○	○
Mit den Eltern dieser Klasse	○	○	○

4.5 Typenbildung

In der Literatur zu qualitativen Forschungsmethoden finden sich zahlreiche Beispiele und Verfahren der Typenbildung (vgl. Flick 2010; Kelle und Kluge 1999; Lamnek 2005;). Je nach Autoren existieren ganz unterschiedliche Typenbegriffe. Kluge (1999) beschreibt in ihrem Buch einerseits verschiedene Arten von Typen und Typologien. Sie unterscheidet Real-, Ideal-, Extrem-, Durchschnitts- und Prototypen. Andererseits stellt sie drei spezifische Verfahren vor, wie empirisch begründete Typen gebildet werden können. Neben dem Konzept nach Barton und Lazarsfeld und der Prozesstrukturanalyse nach Gerhardt, erläutert sie die „Typologische Analyse“ nach Kuckartz. Basierend auf diese theoretischen Vorüberlegungen und Konzeptionen skizziert sie ein vierstufiges Modell der empirisch begründeten Typenbildung. Es erläutert in allgemeiner Form den Prozess der Typenbildung (Abbildung 14) und wird im folgenden Textabschnitt beschrieben (vgl. Kelle & Kluge 1999, S. 81f; Kluge 1999, S. 260ff).

Abbildung 14: Prozess der Typenbildung nach Kluge (1999)



1. Schritt: Erarbeitung relevanter Vergleichsdimensionen

Zunächst werden diejenigen Kategorien und Merkmale der Gesamtauswertung definiert, anhand derer mögliche Ähnlichkeiten oder Unterschiede im Material sichtbar werden.

2. Schritt: Gruppierung der Fälle und Analyse empirischer Regelmäßigkeiten

Die Fälle werden anhand der definierten Kategorien oder Vergleichsdimensionen zugeordnet und nach empirischen Regelmäßigkeiten in Gruppen zusammengefasst. Nach der Zuordnung erfolgt eine differenzierte, vergleichende Analyse der jeweiligen Gruppierungen. Bei dieser Kontrastierung ist darauf zu achten, dass die Fälle innerhalb einer Gruppe eine hohe interne Homogenität aufweisen. Das bedeutet, dass die ausgewählten Fälle sich weitgehend ähneln und somit ein bestimmter Typus vorliegt.

Werden verschiedene Gruppen kontrastierend gegenübergestellt, so muss überprüft werden, ob zwischen ihnen eine genügend hohe Heterogenität vorliegt. Werden Unterschiede sichtbar, so bilden diese auf der Ebene der Typologie eine Kontrastierung des Datenmaterials ab.

3. Schritt: Analyse inhaltlicher Sinnzusammenhänge und Typenbildung

Mit diesem Schritt sollen die empirisch gebildeten Gruppen innerhalb und zwischen den Gruppen auf ihre inhaltlichen Sinnzusammenhänge untersucht werden. Dieses Verfahren trägt dazu bei, dass die bis dahin kausal ermittelten Zusammenhänge auch verstanden und erklärt werden. Hierdurch kann es zu einer Reduktion des Merkmalsraums und damit der Gruppen kommen. Die Analyse der Gruppen kann dazu führen, dass weitere Kategorien oder Subkategorien in den Fokus rücken und die bisher erläuterten Schritte 1-3 erneut durchgeführt werden müssen. Der Pfeil von Stufe 3 auf Stufe 1 veranschaulicht dieses Prozedere (Abbildung 14).

4. Schritt: Charakterisierung der gebildeten Gruppen

Am Ende des Verfahrens steht die umfassende Charakterisierung der gebildeten Typen. Dafür werden die Merkmalskombinationen und inhaltlichen Sinnzusammenhänge verwendet, die das Typische und Charakteristische erfassen und abbilden.

4.6 Inhaltsanalytische Gütekriterien

Bei der vorliegenden Inhaltsanalyse wurden die erforderlichen Gütekriterien beachtet und eingehalten (vgl. Mayring 2007). Zum einen wurde das gesamte Verfahren dokumentiert. Die Dokumentation umfasst neben der Explikation des Vorverständnisses auch die Zusammenstellung des Analyseinstrumentariums. Des Weiteren wird die Durchführung der Datenerhebung und die Auswertung der Daten erklärt.

Bei den Interpretationen wird argumentativ begründet. Voraussetzung hierfür ist das adäquate Vorverständnis des Interpreten und die Kenntnis über den theoretischen Stand der Nutzung von Vergleichsarbeiten. Die Grundlagen für die Interpretationen werden vor allem in Kapitel 3 gelegt. Die Widerlegung von Alternativdeutungen kann ein wichtiges Argument der Geltungsbegründung von Interpretationen sein.

Ein spezifisches Merkmal qualitativer Forschung ist die Offenheit des Gegenstands und die damit einhergehende Gefahr eines unsystematischen Vorgehens. Aus diesem Grund ist es wichtig, vor der Analyse mittels Ablaufschemata ein strenges und regelgeleitetes Vorgehen sicherzustellen. In der vorliegenden Arbeit wurden bereits im Vorfeld Ablaufschemata erstellt, sowohl für das induktive, als auch für das deduktive Vorgehen (siehe Kapitel 4.4.1).

Ein weiteres Gütekriterium der Inhaltsanalyse stellt die Nähe zum erforschten Gegenstand dar. In der vorliegenden Arbeit werden die Lehrerinnen und Lehrer nach ihrem Rezeptions- und Nutzungsverhalten von Vergleichsarbeiten befragt. Die Befragten erhalten die Möglichkeit ihre Erfahrungen mitzuteilen, die sie mit den kurz davor durchgeführten und ausgewerteten Instrumenten sammeln konnten. Der Forschungsgegenstand knüpft also an deren schulische Realität an und steht folglich im Interesse der Lehrpersonen.

Eine Schwierigkeit bei qualitativen Erhebungs- und Auswertungsverfahren stellt sicher die Kodierung und Interpretationsart der Forscherinnen und Forscher dar. Um hierbei einen Konsens zwischen unterschiedlichen Personen herzustellen, wurden die ersten neun Interviews jeweils von drei unabhängigen Personen aus-

gewertet und kodiert. Im Anschluss daran wurden deren Ergebnisse und Interpretationen miteinander verglichen und in der Gruppe diskutiert. Bei Unstimmigkeiten legte man jeweils einheitliche Kodierregeln fest. Erst danach erfolgte die Auswertung des Gesamtmaterials. Im Anschluss daran wurde durch dieses vergleichende Verfahren die Interkoderreliabilität bestimmt. Die Zuordnung der Textsegmente zu den verwendeten Kategorien wurde von allen Beteiligten ähnlich vollzogen. Die Übereinstimmung zwischen den Mitarbeiterinnen und Mitarbeitern betrug ca. 97 %.

Eine weitere Möglichkeit der Gewähr von Gütekriterien stellt das Verfahren der Triangulation dar. Darunter versteht man den Versuch, für die Fragestellung unterschiedliche Lösungswege zu finden, Stärken und Schwächen aufzuzeigen und die Ergebnisse zu vergleichen (vgl. Mayring 2002). Zu der vorliegenden Fragestellung wurden bereits quantitative Studien durchgeführt (vgl. Maier 2009). Somit besteht die Möglichkeit, diese Erkenntnisse mit den eigenen Forschungsergebnissen zu vergleichen und zu diskutieren.

Generell erscheint die gezielte, professionelle Schulung des Interviewpersonals im Vorfeld der Erhebung, Auswertung und Interpretation der Daten als notwendiges Verfahren, um eine hohe Durchführungsobjektivität zu gewährleisten (vgl. Kap. 4.3.2.1).

5 Ergebnisse

Auch Darstellung der Ergebnisse erfolgt schrittweise anhand des Modells von Helmke/Hosenfeld „Von der Innovation zur Evaluation“ (Kap. 3.4.2.2) und in der Reihenfolge der Fragestellungen (Kap. 4.2), beginnend mit der subjektiven Einstellung, also den individuellen Bedingungen und Einstellungen der Lehrpersonen zur Funktion von Vergleichsarbeiten (5.1). Die ausführliche Darstellung der induktiven Auswertung dieser Fragestellung erfolgte im Methodenkapitel (Kap. 4.4.1.4).

In den Kapiteln 5.2 und 5.3 werden jeweils die Ergebnisse zur Rezeption und Reflexion der Datenrückmeldungen vorgestellt. Daran schließen sich die Befunde zu möglichen Aktionen (Kap. 5.4) an, die die Lehrpersonen aus den Ergebnissen ableiten. Die Aussagen der Lehrerinnen und Lehrer zu den schulischen und externen Bedingungen sind in Kapitel 5.5 zusammengefasst. Am Schluss werden alle Teilergebnisse in einem tabellarischen Gesamtüberblick dargestellt (Kap. 5.6).

Nachdem das gesamte Material ausgewertet ist und alle Kategorien vorliegen, können mit Hilfe typenbildender Verfahren spezifische Rezeptions- und Nutzungstypen gebildet werden. Die Typenbildung und die Beschreibung der fünf verschiedenen Typen werden in Kapitel 5.7. ausführlich erläutert und dargestellt.

5.1 Individuelle Bedingungen – Funktion der VA

Tabelle 5.1.1: Individuelle Bedingungen

Hauptkategorie I: Funktion der VA				
Subkategorien	**Ges (n=59)**	**HS (n=15)**	**RS (n=23)**	**Gym (n=21)**
1. Vergleichsarbeiten (VA) dienen landesweitem Vergleich	55	14	22	19
um Standards zu überprüfen und zu sichern	47	8	20	19
als Rückmeldung für Verbesserungen	21	4	10	7
2. VA als Kontrollinstrument und Druckmittel	29	7	12	10
3. VA als Folge der PISA-Studie	14	3	4	7
4. VA haben keine Funktion	5	1	2	2
5. VA als purer Aktionismus	3	0	0	3

In der Tabelle sind alle Antworten enthalten, die die befragten Personen zur Frage nach der vermuteten „Funktion“ von Vergleichsarbeiten äußerten. Die Antworten sind in einzelnen Subkategorien zusammengefasst und werden nach der Häufigkeit ihrer Nennungen gelistet.

Die am häufigsten genannte Meinung der Lehrpersonen (55 Personen) bezüglich der intendierten Funktion von Vergleichsarbeiten bezieht sich auf die Möglichkeit, dass durch Vergleichsarbeiten Schülerinnen und Schüler, Lehrerinnen und Lehrer, einzelne Klassen oder auch ganze Schulen sowohl innerhalb der Schule als auch landesweit miteinander verglichen werden können. Diese fast einstimmige Antwort verwundert nicht, schließlich ist bereits im Wort „Vergleichs-Arbeiten“ die genannte Funktion enthalten. Einige Personen dieser Gruppe bewerten diesen Vergleich sehr positiv und betonen, dass sie schon immer dafür waren, „dass Klassen möglicherweise auch Lehrer, dass was man als Lehrer leistet, vergleichbar gemacht werden muss, oder vergleichbar sein kann“ (151).

Des Weiteren lässt sich diese erste Subkategorie „Vergleichsarbeiten dienen dem landesweiten Vergleich“ in zwei weitere Unterkategorien ausdifferenzieren, je nachdem, welchem Zweck der Vergleich dient.

Die größere Anzahl der Lehrpersonen, 47 Personen, ist der Meinung, dass der Vergleich im Sinne einer summativen Evaluation verwendet wird. Am Ende eines Bildungsabschnittes sollen also bilanzierend die erreichten Kompetenzen und damit die Bildungsstandards überprüft und gesichert werden. Vergleichsarbeiten werden von diesem Personenkreis als kleine Prüfung angesehen, mit der man überprüfen will, was die Schülerinnen und Schüler in den vergangenen zwei Schuljahren gelernt haben. „Dass man eben einfach wissen will, was man am Ende von der Orientierungsstufe, weil das ja so ein Block ist, was man da in den zwei Schuljahren gelernt hat“ (156). Diese Einstellung teilen alle Haupt-, Realschul- und alle Gymnasiallehrpersonen mit je einer Ausnahme.

22 der befragten Lehrerinnen und Lehrer sehen den Nutzen und die Funktion von Vergleichsarbeiten in formativen Evaluationsverfahren. Sie geben an, dass durch das differenzierte Vergleichen Stärken und Schwächen einzelner Schülerinnen und Schüler zurückgemeldet werden, die das Handeln der Lehrpersonen in dem Sinne bestimmen, dass gezielt Verbesserungen angestrebt werden. Ihrer Meinung nach möchte das Ministerium den Schulen und den darin agierenden Personen durch Vergleichsarbeiten eine Rückmeldung geben, wo diese stehen. Diese Rückmeldung wird demnach als Ausgangspunkt für Reflexionen betrachtet. Zum einen können die Lehrerinnen und Lehrer dadurch ihr unterrichtliches Vorgehen kritisch beleuchten, um z.B. einschätzen zu können, „wie weit der Unterricht gut verlaufen ist, verbesserungsfähig ist, wo vielleicht falsche Schwerpunkte gesetzt wurden oder die Richtigen“ (164). Neben der Überprüfung der Nachhaltigkeit was im Unterricht erreicht wird, können aber auch die Schwächen und die Stärken einzelner Schüler herausgefunden werden (vgl. 253). Diese förderdiagnostische Sichtweise wird von der Hälfte der Realschullehrpersonen und von jeweils einem Drittel der Befragten aus Hauptschulen und Gymnasien eingenommen.

Im Gegensatz zu dieser positiven Perspektive sieht die Hälfte der Befragten aller drei Schularten Vergleichsarbeiten als Kontrollinstrument an und erlebt diese als Druckmittel. Diese neuen Evaluationsinstrumente werden also nicht wie oben beschrieben als Vergleichsmaß für angestrebte Verbesserungen angesehen, sondern „man möchte, ich denke auch dass man die Lehrer dazu zwingen möchte, einen bestimmten Standard zu erreichen, oder vielleicht auch gewissenhafter zu arbeiten“ (250). Vergleichsarbeiten werden folglich als Druckmittel erlebt, um den

Lehrerinnen und Lehrern „nicht ein völlig freies Schaffen zu ermöglichen. Zumal es sind ja auch gar keine Kontrollen mehr da“ (ebd.). Dieser Aspekt wird von den Lehrenden angeführt, weil im Zuge der Bildungsplanreform 2004 die Einzelschule in Baden-Württemberg mehr Gestaltungsfreiräume erhielt und die Schulaufsichtsbehörde ihre „Kontrollfunktion“ in Form von Unterrichtsbesuchen reduzierte. Die Frage, ob sich die Lehrperson durch Vergleichsarbeiten kontrolliert fühle, wird von der Hälfte bejaht. Die meisten geben an, dass sie aufgrund des guten Ergebnisses keinen Druck empfinden und sich durch Vergleichsarbeiten nicht kontrolliert fühlen. Sie räumen jedoch ein, dass sie ein Problem damit hätten, wenn ihre Klasse schlechter abgeschnitten hätte.

Das Argument, dass Vergleichsarbeiten als Folge der PISA - Studie implementiert wurden, teilen insgesamt 14 Lehrerinnen und Lehrer. Anlass für die Maßnahme war „erst einmal der PISA - Schock, der das Ganze in die Wege geleitet hat“ (156). Durch diese internationale Schulleistungsstudie wurden zunächst Mängel des deutschen Schulsystems sichtbar und durch flächendeckende Vergleichsarbeit erhält man nun spezifische Hinweise, wo diese Lücken groß und klein sind (vgl. 256). Mit dieser Aussage weisen die Lehrpersonen darauf hin, dass groß angelegte, stichprobenbasierte, internationale Schulleistungsstudien wie PISA, TIMSS etc. nicht auf die Ebene der Einzelschule bzw. der Unterrichtsebene übertragen werden können, sondern dem Systemmonitoring dienen. Vergleichsarbeiten hingegen, die an allen Schulen des Landes in ausgewählten Klassenstufen flächendeckend durchgeführt werden, lassen eher Rückschlüsse auf die Qualität der Schule bzw. des Unterrichts zu.

Insgesamt geben fünf Personen an, dass Vergleichsarbeiten keine Funktion erfüllen. Sie erleben das ganze Verfahren als zusätzlichen, sinnlosen Arbeitsaufwand, als Papierkrieg und als Beschäfigungstherapie, ohne jegliche Zweckbestimmung. Dementsprechend erwarten diese Personen nichts von diesen in ihren Augen unnötigen Instrumenten. Sie vermuten, dass diese von fanatischen Statistikern entwickelt wurden und werten sie extrem ab.

> „Also eine spontane Antwort, die mir dazu einfällt, dass es eine Beschäftigungstherapie für Lehrer ist. Und da sind irgendwelche Statistikfanatiker drinnen im Club, ja! Also einen vernunftbegabten Grund kann ich mir da nicht denken.“ (158)

Drei Befragte erleben Vergleichsarbeiten als puren Aktionismus. Als Auslöser wird auch die PISA-Studie angeführt. Doch im Vergleich zu oben beschriebener „als Folge von PISA“ - Kategorie sollen nicht Daten erhoben werden, die ein differenziertes Aggregationsniveau erfassen, im Sinne eines Schulmonitorings. Die Aussage eines Realschullehrers beschreibt diese Subkategorie passend. Er kritisiert, dass mit dem Verfahren nicht die Ursachen für die schlechten Ergebnisse behoben werden. Denn dazu wären individuelle Feinanalysen notwendig, die differenziert über die jeweiligen Stärken oder Schwächen der Schülerinnen und Schüler informieren und aufklären. Anstatt diagnostisch vorzugehen, beklagt dieser Lehrer das oberflächliche, pauschale und undetaillierte Vorgehen bei der willkürlichen Produktion weiterer Output-Ergebnisse. Er ist überzeugt, dass die tatsächlich vorherrschenden Probleme durch dieses Vorgehen nicht gelöst werden können.

> „…und ja Aktionismus. Das ist in meinen Augen purer Aktionismus. Pisa-auslöser und jetzt muss man schauen, dass man um sich schlägt, dass man zu Ergebnissen kommt. An den Ursachen wird nicht gearbeitet, Einzelanalysen werden keine betrieben. Das geht pauschal, die Schüler bekommen eine Nummer, ich bekomme eine Nummer. Der Lösung der Detailfragen und Probleme kommt man so nicht näher, absolut nicht.“ (260)

5.1.1.1 Fazit: individuelle Bedingungen –Funktion der Vergleichsarbeit

Zusammenfassend lässt sich festhalten, dass die Antworten der Lehrerinnen und Lehrer zur Frage nach der „Funktion von Vergleichsarbeiten“ sehr unterschiedlich ausfallen. Die neutralen Positionen („VA dienen dem Vergleich, um Standards zu überprüfen“ oder „VA als Folge der PISA-Studie“) und positiven Einstellungen („VA als Rückmeldung für Verbesserungen“) stimmen mit der intendierten Zielsetzung des Ministeriums aus § 114 des Schulgesetzes überein. Demnach können Vergleichsarbeiten nach Meinung der Lehrenden in ihrer Funktion sowohl für die Qualitätsentwicklung der Einzelschule, als auch für die Feststellung des Lernstands der Schülerinnen und Schüler verwendet werden (Landesinstitut für Schulentwicklung).

Auffallend ist, dass jede zweite Person, über alle drei Schularten hinweg, Vergleichsarbeit als Instrument der Kontrolle erlebt, mit dem Druck auf die Akteure

erzeugt werden kann. Diese Anzahl ist erschreckend hoch, wenn man bedenkt, dass Druck und negative Assoziationen Entwicklungen eventuell verhindern und somit den erhofften Verbesserungen im Wege stehen.

Vor allem finden sich unter den Gymnasiallehrpersonen kritische Äußerungen, die Vergleichsarbeiten als funktionslos und als puren Aktionismus attribuieren. Auch bei diesen Personen muss damit gerechnet werden, dass mit einer derartigen Einstellung keine Maßnahmen im Sinne einer Qualitätsverbesserung vorgenommen werden können, sondern alles „beim Alten“ bleibt.

5.2 Ergebnisse zur Rezeption

Der erste Auswertungsschritt im Modell von Helmke/Hosenfeld der den idealtypischen Verlauf der Nutzung von Rückmeldungen aus Leistungserhebungen skizziert, beschreibt die Rezeption der Vergleichsarbeiten. Die Aussagen der interviewten Lehrerinnen und Lehrer lassen sich den beiden folgenden Hauptkategorien zuordnen:

I. Durchführung der Vergleichsarbeit

II. Verständlichkeit der Rückmeldungen

5.2.1 Rezeption - Hauptkategorie I: Durchführung der Vergleichsarbeit

Tabelle 5.2.1: Rezeption – Hauptkategorie I				
HK I: Durchführung der Vergleichsarbeit				
Subkategorien	**Ges (n=59)**	**HS (n=15)**	**RS (n=23)**	**Gym (n=21)**
1. Keine Probleme bei der Durchführung der Vergleichsarbeit	54	14	20	20
2. Zur Verfügung stehende Zeit war für Schüler				
…ausreichend	39	8	13	18
…zu kurz	10	1	7	2
…zu lang	5	4	0	1
3. Uneinheitliches Vorgehen bei der Aushändigung der VA	5	1	1	3

5.2.1.1 keine Probleme bei der Durchführung (insg. 54; HS 14, RS 20, Gym 20)

Insgesamt geben beinahe alle Interviewpartner (54 von 59) an, mit der Durchführung der Vergleichsarbeiten keinerlei Probleme gehabt zu haben. Als Grund hierfür wird angeführt, dass die Vorgaben und Lehrerhandreichungen zur Durchführung der Vergleichsarbeit klar und verständlich formuliert waren. Die Lehrerinnen und Lehrer erhielten sowohl klare Hinweise in Bezug auf die Terminierung als auch auf die Art und Weise der Durchführung der Vergleichsarbeiten. Ebenso lobten die Befragten die Verständlichkeit des gesamten Informationsmaterials, sowie das rechtzeitige Bereitstellen der Lösungen (vgl. 151). Dies galt auch dann, wenn die Erhebungen von fachfremden Lehrpersonen durchgeführt wurden.

> „Also, ich habe es ja gar nicht selber gemacht. Ich war im Schullandheim. Der Fremdlehrer hat da aber kein Problem gehabt, damit zu Recht zu kommen. Und diese Anweisungsblätter sind eigentlich sehr eindeutig. So dass man auch, wenn man es auch zum ersten Mal und fachfremd macht, kann man das durchführen.“ (362)

Dass sich Vergleichsarbeiten unproblematisch durchführen lassen, selbst wenn fachfremde Personen diesen Part übernehmen, kann als wichtige Voraussetzung für deren breite Verwendung gewertet werden. Würden sich bereits hier Probleme und Hindernisse aufzeigen, so müsste damit gerechnet werden, dass einige Lehrerinnen und Lehrer womöglich die Arbeiten ablehnen würden.

5.2.1.2 Zur Verfügung stehende Zeit war für Schüler

5.2.1.2.1 …ausreichend (insg. 39; HS 8, RS 13, Gym 18)

Etwa zwei Drittel aller Lehrpersonen sind der Meinung, dass die zur Verfügung stehende Zeit für die Bearbeitung der Aufgaben ausreichend war. Diese Einstellung kommt in folgenden Aussagen zum Ausdruck:

> „So weit ich mich erinnern kann, hat die [Zeit] gut gereicht. Also wirklich, zeitlich gut machbar“ (161). „Der Zeitrahmen war in Ordnung, meiner Meinung nach. Also ist jeder auch fertig geworden. Von dem her war es nicht das Problem“ (266). „Zeit war ordentlich bemessen und ausreichend bemessen.“ (355)

5.2.1.2.2 …zu kurz (insg. 10; HS 1, RS 7, Gym 2)

Zehn Personen sind der Ansicht, dass den Schülerinnen und Schülern für die Bearbeitung der Vergleichsarbeit zu wenig Zeit zur Verfügung stand. Betrachtet man die einzelnen Schularten und -fächer, so fällt auf, dass vor allem ein Drittel der Realschulehrpersonen diese Aussage teilt. Die meisten dieser Personen beziehen sich in ihren Angaben auf das Fach Deutsch, Klassenstufe 6. Als Grund für den Zeitdruck werden ein erhöhter Umfang im Vergleich zu den Vergleichsarbeiten vom Vorjahr und eine komplexe Aufgabenstellung angeführt. Dies hatte zur Folge, dass sich die Schülerinnen und Schüler bei der Bearbeitung sehr beeilen mussten und teilweise auch nicht fertig wurden (vgl. 257). Die Lehrerinnen und Lehrer machen die Aufgabenstellung dafür verantwortlich, weil die Schülerinnen und Schüler aufgefordert waren einen halben Aufsatz zu verfassen. Zuvor benötigten sie viel Zeit, um sich in das Thema einzulesen.

> „Und der Ausgangstext, das waren ja zwei DIN A 4 Seiten. Und da brauchten sie erst mal Zeit das zu lesen. Sich in die Geschichte rein zu versetzten, sich zu überlegen wie könnte die Geschichte ausgehen usw. Also, das braucht Zeit. Und da haben manchen schon ein bisschen ein Problem mit der Zeit gehabt. Das war wirklich knapp bemessen. Ich habe festgestellt im Vorjahr war so eine große Aufgabenstellung nicht dabei gewesen.“ (259)

5.2.1.2.3 …zu lang (insg. 5; HS 4, RS 0, Gym 1)

Dass die Schülerinnen und Schüler für die Bearbeitung der Vergleichsarbeit zu viel Zeit hatten meinen, im Kontrast zu den oben beschriebenen Realschullehrkräften, insgesamt fünf Lehrerinnen und Lehrer, darunter vier Hauptschullehrpersonen. „Die haben es früher abgegeben, eigentlich, als sie Zeit hatten“ (156). Dass ausgerechnet die schwächsten Schülerinnen und Schüler schneller mit den Tests fertig sind als die Vergleichsgruppen an Realschulen und Gymnasien fällt auf und verlangt nach Erklärungsgründen. Eine Lehrerin führt als möglichen Grund die oberflächliche Bearbeitungsweise sowie die fehlende Geduld und Ausdauer der Schülerinnen und Schüler an.

> „Ja, die Schüler waren früher fertig. Die Zeit brauchte ich gar nicht. Vielleicht liegt es auch an der Geduld und Ausdauer, dass die Schüler die nicht haben. Sie können die Zeit teilweise schon brauchen. Manche Aufgaben wurden oberflächlich gemacht, aber da fehlt dann einfach auch die Ausdauer.“ (154)

5.2.1.3 Uneinheitliches Vorgehen bei der Ausgabe der Vergleichsarbeiten an die Lehrerinnen und Lehrer (insg. 5; HS 1, RS 1, Gym 3)

Als großen Kritikpunkt bei der Durchführung der Vergleichsarbeiten führen die Lehrpersonen den Zeitpunkt der Aushändigung der Vergleichsarbeit an, da dieser im Vergleich zum Vorjahr und im Vergleich mit anderen Schulen teilweise uneinheitlich festgelegt wurde. Eine Lehrperson wundert sich über die sehr frühe Ausgabe und die damit verbundene Möglichkeit, Schülerinnen und Schüler noch gezielt auf den Test vorzubereiten.

> „Ich fand es sehr seltsam, dass die Aufgaben schon eine Woche vorher bei uns waren.“ (162)

Dadurch erhalten diese Lehrerinnen und Lehrer die Möglichkeit die Aufgaben und das Niveau der Vergleichsarbeiten im Vorfeld zu erfahren. Sie könnten dieses Wissen dazu verwenden, ihre Schülerinnen und Schüler in der verbleibenden Zeit auf bestimmte Themen vorzubereiten. Eine Hauptschullehrerin berichtet von einem Kollegen des Faches Englisch, der die frühzeitige Aushändigung der Vergleichsarbeit dafür nutzte, mit seiner Klasse bestimmte Schwierigkeitsbereiche der Vergleichsarbeit eine Woche lang gezielt einzuüben, bis diese beherrscht wurden. Dass dieser Kollege nicht an derselben Schule tätig ist, betont die Lehrerin ausdrücklich:

> „Denn ich kenne einen Kollegen von mir, nicht an unserer Schule, aber an einer anderen, die haben dann wirklich gezielt darauf hin geübt. Und das sollte ja nicht Sinn der Sache sein. Die haben sich das angeschaut und haben gesehen, aha, bei Aufgabe so und so da kommt dann was dran mit dem „listening“ oder so, das haben wir schon ewig nicht mehr geübt, das üben wir halt noch eine Woche lang. Dann können es die Kinder natürlich.“ (162)

Eine weitere Lehrperson bekundet, dass sie die Vergleichsarbeiten bereits am Vortag nach Unterrichtsschluss erhalten hat. Somit hatte diese die Gelegenheit, sich auf die Vergleichsarbeit am nächsten Tag vorzubereiten.

Fünf Personen, denen die Vergleichsarbeit am Morgen der Durchführung zugestellt wurde, beklagen sich teilweise massiv darüber, dass zwischen dem Erhalt der Unterlagen und der Bearbeitung durch die Schüler zu wenig Zeit verbleibt, um sich auf die Durchführung der Vergleichsarbeit ausreichend und gründlich vorbereiten zu können. Durch diese Regelung geraten die Lehrenden unter Zeitdruck

und sie berichten von der Sorge, nicht adäquat und professionell handeln zu können. Sie empören sich darüber, dass diese strenge Regelung im Vergleich zum Vorjahr eingeführt wurde und dass mit einem Mal nur noch bestimmte Personen der Schule, in diesem Fall handelt es sich um die Schulleitung, die Vergleichsarbeiten erhalten, diese kopieren dürfen und bis kurz vor der Durchführung unter Verschluss halten müssen.

> „Das Einzige, was ich als Mangel empfand, dass man beim ersten Mal, wir waren ja jetzt die zweite Generation, da durften es die Lehrer vorher kopieren. Jetzt hat es bei uns eine große Empörung gegeben, als dann bekannt wurde, ich denke auch an anderen Schulen, das dürfen nur ausgewählte, also A15-Studiendirektoren, die das vorher kopieren durften, unter Verschluss halten dürfen und uns Lehrern dann am Morgen der Arbeit erst ausgeben dürfen.“ (350)

In diesem Zusammenhang unterstellen sie der Schulleitung mangelndes Vertrauen in Bezug auf einen korrekten, vorschriftsmäßigen Umgang des Lehrpersonals mit den Vergleichsarbeiten. Die kurzfristige Ausgabe der Vergleichsarbeit hat zur Folge, dass die durchführenden Lehrpersonen enorm unter Zeitdruck geraten, wenn sie die Aufgaben erst morgens, kurz vor der Durchführung erhalten. Denn zum einen müssen sie die Anzahl der Arbeiten kontrollieren, ob sie für die Klassengröße ausreicht. Zum anderen möchten sie sich vor der Bearbeitung durch die Schülerinnen und Schüler einen Überblick über die geforderten Kompetenzen verschaffen und dafür vor allem ausreichend Zeit und Ruhe zur Verfügung zu haben. Doch die Realität sah für einige anders aus, wie folgende Aussagen deutlich machen:

> „Das hat nämlich ganz pragmatisch dann zur Folge gehabt, ich war schon sehr früh an diesem Wochentag hier. Und als dann, das ist jetzt kein Vorwurf gegen den Kollegen, aber als dann der Kollege kam und den ganzen Pack mir übergeben hat, ich habe mich schnell in ein Zimmer rein, wo einigermaßen Ruhe war, und habe dann ganz schnell, für mich war das ja alles Neuland, dann da durch die Aufgabenstellung hindurchgearbeitet. Und musste mich erst einmal informieren und die nachzählen, ob alles vollständig ist. Dann bin ich gerade so zwei, drei Minuten vorher fertig geworden. Dann musste ich schon mit dem Packen ins Zimmer reinstürmen und weitermachen bei den Schülern. Das habe ich als sehr inhuman und inadäquat empfunden, muss ich sagen. Beim ersten Mal konnte man es gemütlicher machen, ich weiß nicht, ob Sie das wissen, da durfte es eben jeder selber kopieren. Und jetzt hat man das so verfügt, dass das unter Verschluss gehalten wird bis kurz vorher. Diese Modalität des Ablaufs fand ich schlecht.“ (350)

> „Was halt schwierig war, ein kleines Problem war, dass wir die ja nur ganz kurz vorher bekommen haben. Die haben wir morgens abgeholt und dann ging es los. Das wäre für die Lehrer geschickt gewesen, wir hätten die eine Stunde vorher gehabt, um die mal selber anzuschauen, weil das ging ja dann los mit einem Diktat und da musste ich selber erst mal kurz schauen. Das war für den Lehrer (…) schwierig zu praktizieren.“ (358)

Wie gezeigt, würden sich die Lehrerinnen und Lehrer im Vorfeld der Durchführung etwas mehr Vorbereitungszeit wünschen. Dadurch entfiele der als unangenehm erlebte Zeitdruck der sich auf die Lehrpersonen negativ auswirkt, weil er zur Verunsicherung beiträgt. Davon sind vor allen Dingen diejenigen betroffen, die es gewohnt sind, sehr gut vorbereitet in den Unterricht zu gehen. Auf der anderen Seite birgt eine frühzeitige Aushändigung zum einen die Gefahr, dass Lehrerinnen und Lehrer ihre Klassen auf die Vergleichsarbeiten vorbereiten im Sinne eines „teaching to the test“, wie es weiter vorn beschrieben wurde. Zum anderen könnten sie ihr Vorwissen in Form von Hilfestellungen während der Durchführung ihren Schülerinnen und Schülern anbieten. Diese Unterstützung ist bei externen Evaluationen nicht erwünscht, denn Vergleichsarbeiten sollen den tatsächlichen Kenntnisstand der Schülerinnen und Schüler ermitteln und nicht den der Lehrpersonen. Wie weit bei dieser Form der ‚Unterstützung’ die Angst der Lehrerinnen und Lehrer mitspielt, dass schlechte Schülerergebnisse auf ihre Arbeit zurückgeführt werden können, kann nur vermutet werden. Es ist gewiss nicht einfach für Lehrerinnen und Lehrer, die für gewöhnlich in der Rolle sind, ihre Schülerinnen und Schüler aktiv zu fördern, zu fordern und zu unterstützen, dass sie bei Vergleichsarbeiten lediglich die passive Rolle der „Testdurchführung“ einnehmen sollen. Auch für die Klasse ist dieser plötzliche Rollentausch nur schwer nachzuvollziehen, wenn sie in sonstigen Lehr-Lernsituationen häufig mit Unterstützung rechnen können.

> „Ja, also es trägt auch nicht zur Vertrauensbildung bei, dass wir das hier morgens kriegen und dann damit los marschieren sollen ohne uns vorher damit beschäftigen zu können. Das verunsichert allgemein, und dann weißt du wirklich nicht, wo sind wirklich irgendwo Fallen. Und dann geht es um Folgeprobleme, dann gibt es einen Schüler der blickt wirklich ab einer Stelle etwas wirklich nicht, wo du aber weißt, der hat es normalerweise drauf. Und das verunsichert dich, wenn du dann das Diktum hast, es darf keine Hilfe usw. [gegeben werden]. Das ist ein bisschen Käse. Gut, es mag sein das ist beim ersten Mal so. Für mich war es absolute Premiere dieses Mal. Und dann diese Art, ich kriege es gerade mal zehn Minuten oder eine viertel Stunde oder was weiß ich vorher, marschiere damit in den Klassenraum.

> Kann mich überhaupt nicht damit auseinandersetzen. Was erwartet mich da. Kapier praktisch im Beisein von 32 Schülern, oder 31 waren es, weil einer krank war, realisiere ich dann auch was hier abläuft. Das ist ein bisschen unglücklich.“ (363)

5.2.2 Rezeption – Hauptkategorie II: Verständlichkeit der Rückmeldung

Tabelle 5.2.2: Rezeption – Hauptkategorie II

HK II: Verständlichkeit der Rückmeldungen				
Subkategorien	**Ges (n=59)**	**HS (n=15)**	**RS (n=23)**	**Gym (n=21)**
1. Rückmeldungen waren verständlich	16	2	6	8
2. Rückmeldungen wurden nicht angeschaut	13	3	6	4
3. Rückmeldungen waren unverständlich / unpraktisch	5	2	3	0

5.2.2.1 Rückmeldungen waren verständlich (insg. 16; HS 2, RS 6, Gym 8)

16 Lehrpersonen geben an, dass die Rückmeldungen für sie verständlich aufbereitet und dadurch auch nachvollziehbar sind. Die Verständlichkeit kommt u. a. dadurch zustande, dass die Auswertungsdaten klar und nachvollziehbar aufbereitet waren (vgl. 150). Die Auswertungsmatrix bestand darin, in einer vorbereiteten Excel-Tabelle Kreuze für richtige Aufgaben zu setzen, aus der das Programm die Auswertungsdaten in Prozentangaben für die richtigen Antworten generierte.

> „Verständlich waren sie. Einfach von den Prozentzahlen.“ (152) „Also diese Items praktisch und diese Kreuzchen, die ich da machen musste? Das war für mich schon verständlich. Also das war nicht das Problem.“ (266)

Vor allem Mathematiklehrer betonen, dass sie mit der Darstellung der Rückmeldung keine Schwierigkeiten hatten. Schließlich gehört sowohl der Umgang mit Tabellen, Statistiken und Grafiken als auch die Nutzung des Computers zum Themenbereich der Mathematik.

> „Hat mir jetzt als Mathematikerin keine Probleme bereitet (lacht)“ (361). „Ich hatte da natürlich kein Problem damit, da der Umgang mit Tabellen in Mathematik sowieso Stoff ist.“ (354)

Drei Personen dieser Gruppe berichten, dass sie die Rückmeldungen nicht intensiv genutzt hatten, obwohl sich diese, aufgrund der übersichtlichen Aufbereitung und Darstellung, sehr gut für die Ableitung weiterer Maßnahmen und Schritte eignen würden. Bei dieser Aussage wird deutlich, dass Rückmeldungen, auch wenn sie sehr differenziert die Stärken und Schwächen der Schülerinnen und Schüler aufzeigen, nicht automatisch zu Änderungen im Sinne von Verbesserungen führen. Die Gründe, warum diese Lehrpersonen die Daten trotz deren Eignung nicht nutzen, werden im Kapitel „Reflexion“ ausführlich dargestellt.

> „Ich hab es angeschaut, nicht intensiv, aber ich hab es angeschaut und es ist nachvollziehbar.“ (353)

> „Ich muss zu meinem eigenen Leidwesen eingestehen, ich hätte damit viel anfangen können. Es war ja wunderschön aufgegliedert mit diesem Streuungsdiagramm, konnte man sogar sehen, wo bei der Klasse jetzt Schwierigkeiten immer noch sind oder Lücken sind und wo man gut gearbeitet hat.“ (368)

Eine Realschullehrperson bewertet die Darstellungsebene der Rückmeldung ebenfalls als verständlich, sie wundert sich jedoch darüber, warum bei einer angeblich anonymen Lernstandserhebung die Namen der Schülerinnen und Schüler angegeben werden müssen. Diese Frage ist in dem Fall berechtigt, wenn die Daten der Schülerinnen und Schüler im Sinne eines Systemmonitorings an übergeordnete Ebenen – z.B. Schulleitung, Schulamt, Landesinstitut für Schulentwicklung oder Ministerium – zurückgemeldet werden sollen, denn diese sind nicht an individuellen Leistungsergebnissen interessiert. Im Sinne einer Verbesserung auf Individual- oder Klassenebene benötigt die Lehrperson jedoch die Namen zur Ableitung spezifischer Fördermaßnahmen. In dieser Aussage wird nochmals deutlich, dass Vergleichsarbeiten unterschiedliche Funktionen und Zielsetzungen beinhalten können, je nachdem, welche Akteure die Daten nutzen sollen.

> „Also für mich waren die Grafiken und Tabellen schon verständlich, was ich mich nur gefragt habe ist, warum man da den Namen vom Schüler angeben muss, wenn es doch anscheinend anonym ist. Das fand ich ein bisschen seltsam.“ (264)

5.2.2.2 Rückmeldungen wurden nicht angeschaut (insg. 13; HS 3, RS 6, Gym 4)

Auf die Frage, ob Lehrerinnen und Lehrer die Rückmeldungen aus den Vergleichsarbeiten verstanden hätten, gaben 13 Lehrpersonen an, diese nicht angeschaut oder verwendet zu haben.

> „Nein. Ich habe sie mir bis jetzt noch nicht angeschaut. Da kann ich jetzt nichts dazu sagen.“ (351)
>
> „Die Auswertung von dem Schaubild - Diagramm besser/schlechter - hab ich gar nicht benutzt.“(256)

Bei einigen der Befragten klangen die Antworten wie Eingeständnisse:

> „Ehrlich gesagt, ich habe es gar nicht angeguckt.“ (151)
>
> „Da muss ich jetzt gestehen, ich habe relativ wenig angeschaut.“ (366)

Ein Lehrer begründet sein Desinteresse an den Rückmeldungen mit dem Lehrerwechsel nach Klasse 6. Im Anschluss an die Vergleichsarbeiten gab er die Klasse ab und hatte auch kein Interesse mehr daran, Rückschlüsse aus den Ergebnissen zu ziehen. Er erwägt, dass sich vielleicht der Nachfolgeklassenlehrer interessieren könnte (vgl. 157). Vier weitere geben ihr Desinteresse an den dargebotenen Rückmeldedaten der Vergleichsarbeiten offen zu (vgl. 159, 254). Eine erklärt ihre Einstellung damit, dass sie in ihrer Klasse weit wichtigere Probleme zu bearbeiten hätte, als Ergebnisse aus Vergleichsarbeiten zu besprechen.

Als weitere Gründe für die fehlende Auseinandersetzung mit den Daten werden einerseits mangelnde Zeit und Skepsis gegenüber der Aussagekraft der Vergleichsarbeit angeführt.

> „Ich war froh, dass ich fertig geworden bin (…). Das war es eigentlich auch. Für die Auswertung auf Klassenebene und auf Schulebene - dazu habe ich gar keine Zeit. Man könnte das eventuell hinterher machen, wenn alle drei, wir haben drei Parallelklassen, wenn alle drei das in Computer eingegeben haben, dann könnte man vielleicht schauen.“ (250)

Andererseits werden auch Erfahrungen aus dem Vorjahr berichtet, die für eine reduzierte Auseinandersetzung und Verwendung bei der aktuellen Vergleichsarbeit verantwortlich sind. Bei der ersten Durchführung der Vergleichsarbeiten wurden alle Daten intensiv angeschaut und alle Eingaben vorschriftsmäßig erledigt. Danach wurden alle Grafiken gründlich studiert und diskutiert. Als Grund für die in-

tensive Nutzung wird angeführt, dass die Klasse eher schlecht abgeschnitten hatte und die Lehrkraft sich bemühte, die Ergebnisse einschätzen und erklären zu können. Dieses Jahr kamen zufriedenstellende Ergebnisse zustande und die Lehrperson verzichtet auf eine Analyse der Daten. Außerdem wurde schulintern der Beschluss gefasst, auf die Eingabe der Daten zu verzichten.

> „Dazu muss ich sagen, dass ich sie mir beim ersten Mal sehr intensiv angeschaut habe und alles eingegeben habe. Und dann wurden die Grafiken ausgeworfen - im Landesvergleich usw. Beim zweiten Mal hab ich eigentlich darauf verzichtet, das überhaupt alles einzugeben, weil wir es nicht mehr mussten. Das haben wir schulintern beschlossen. (…). Dafür gab es ja einen Grund: Beim ersten Mal war natürlich alles noch neu, man wollte alles durchspielen. Das ging damals auch nicht besonders gut aus mit der Klasse und da wollte ich wissen, wie das einzuschätzen ist. Diesmal war ich dann zufrieden mit den Ergebnissen und dann war es mir nicht mehr so wichtig. Ich habe schon vermutet, dass alles im Mittelfeld liegt und es bei einigen hin und her geht.“ (255)

5.2.2.3 Rückmeldungen waren unverständlich (insg. 5; HS 2, RS 3, Gym 0)

Zwei Hauptschul- und drei Realschullehrpersonen bemängeln, die verwendeten Grafiken und Darstellungen der Rückmeldungen seien unverständlich, zu undifferenziert und umständlich zu handhaben und daher unbrauchbar. Trotz des Versuchs, Informationen aus den Rückmeldungen zu gewinnen, gelang es den Lehrpersonen nicht. Sie sahen auch keinen Sinn in der Art und Weise des Rückmeldeformates. Schlechte Ergebnissen wurden mit einem Fragezeichens, gute Resultate mit einem Ausrufezeichen gekennzeichnet. Die Bedeutung und Interpretation der Kennzeichnung bleibt diesen Personen verborgen und sie können keinen Nutzen daraus ziehen. Weitere Schwierigkeiten bereiteten den Lehrerinnen und Lehrern teilweise die Prozentangaben und die Balkendiagramme. Obwohl sie sich damit beschäftigt hatten, können sie keinen direkten Bezug zu den Aufgaben- und Kompetenzbereichen herstellten.

> „Und dann muss ich sagen, mit dieser Tabelle hier, kann ich gar nichts anfangen, die sagt mir gar nichts. Wir haben es versucht herauszufinden, was es sein könnte, aber mit diesen Strichen und diesen Prozentzahlen, das sind ja wohl die Items hier und hier die Prozente, aber daraus lese ich nichts ab. Sage ich ehrlich. Wenn Sie etwas erkennen wunderbar, aber ich nicht.“ (162)

> „Was ich auch sehr seltsam finde ist dieser Prozentsatz von der Klasse und von der Vergleichsstichprobe und dann wenn die Klasse entweder, wenn die Klasse besser war, ein Ausrufezeichen darunter oder ein Fragezeichen. Und da kann ich eigentlich... Ausrufezeichen heißt sie waren besser, Fragezeichen heißt sie waren um einiges schlechter als der Vergleichswert und ... das ist ja gut und recht, aber ich für mich weiß da nicht welchen Sinn das dann haben soll, (lachen) dieses Fragezeichen.“ (264)

> „Diese Darstellung mit diesem Balkendiagramm fand ich ein bisschen blöd, weil da nicht drauf zu erkennen war, welche Kompetenzen da jetzt abgefragt wurden. Da musste man dann immer in den Aufgabenblättern wieder hinterher blättern. Das wäre natürlich eine super Erleichterung, oder wenn da wirklich so eine Zuordnung stattfinden würde, dass da irgendwie steht Zeichensetzung, Alarm, oder. Alles andere fand ich dann, diesen Klassendurchschnitt fand ich dann irgendwie doch etwas zu schwammig.“ (164)

Eine Realschullehrerin berichtet von der Erfahrung, dass insbesondere bei älteren Kolleginnen und Kollegen Verständnisschwierigkeiten beim Lesen der Diagramme auftreten können und sie auf die Unterstützung und Hilfe angewiesen sind. In den zur Verfügung gestellten Materialien findet diese Personengruppe keine Unterstützung. Die meisten werden innerhalb des Kollegiums um Hilfestellungen bitten. Voraussetzungen dafür sind zum einen, dass innerhalb der Schule ein gutes kollegiales und kooperatives Klima herrscht und sich die Betroffenen trauen, ihre Schwierigkeiten bei der Dateninterpretation und –nutzung zuzugeben. Zum anderen könnten an der Einzelschule spezielle Fachleute die Lehrpersonen während der Durchführung und Auswertung von Vergleichsarbeiten betreuen. Eine weitere Unterstützung könnte durch die Schulleitung angeboten werden, die sich dafür verantwortlich zeigt.

> „Also ich habe älteren Kollegen geholfen - mit denen die Grafik angeschaut und erklärt, was das alles bedeutet. Wer im Diagramme lesen fit ist, kann damit etwas anfangen. Aber die Älteren hatten schon eher Probleme damit, was das alles bedeuten soll? Es war relativ kurz erklärt in der Rückmeldung, aber man ist nicht besonders schlau daraus geworden. Wer überhaupt Probleme damit hat, muss bei Null anfangen - was ist die x-Achse, was die y-Achse.“ (269)

Aufgrund der beschriebenen Probleme beschränken sich diese Personen bei den Rückmeldedaten lediglich auf den Notendurchschnitt der Klasse. Eine intensive Nutzung im Sinne der intendierten Zielsetzung bleibt in diesen Fällen aus.

> „Ja, bei uns haben viele ältere Kollegen Schwierigkeiten gehabt. Die haben auch die Diagramme nicht ausgedruckt. Die haben sich dann nur auf den Notendurchschnitt der Vergleichsarbeit bezogen.“ (269)

5.3 Ergebnisse zur Reflexion

Der Schritt „Reflexion über die Vergleichsarbeiten“ schließt sich direkt an die Rezeption an. In dieser Phase setzen sich die Lehrpersonen mit den Ergebnissen der Vergleichsarbeit auseinander und suchen nach möglichen Erklärungen für das entsprechende Abschneiden ihrer Schülerinnen und Schüler.

In diesem Bereich lassen sich die Aussagen der Lehrerinnen und Lehrer fünf Hauptkategorien zuordnen, die folgende Themengebiete umfassen:

I. Auswertung der Vergleichsarbeiten
II. Aufgaben der Vergleichsarbeit
III. Niveau der Vergleichsarbeit
IV. Zeitpunkt der Vergleichsarbeit
V. Förderdiagnostischer Nutzen der Vergleichsarbeit

5.3.1 Reflexion – Hauptkategorie I: Auswertung der Vergleichsarbeit

Tabelle 5.3.1: Reflexion – Hauptkategorie I

HK I: Auswertung der Vergleichsarbeit				
Subkategorien	**Ges (n=59)**	**HS (n=15)**	**RS (n=23)**	**Gym (n=21)**
1. Kritik an Bewertungspraxis	39	4	15	11
Punktevergabe unpräzise und ungerecht	30	4	15	11
Notenschlüssel ungerecht, fehlende lineare Skalierung	10	0	9	1
generelle Kritik an Bewertung der VA	2	0	1	1
2. VA sind einfach und schnell zu korrigieren	9	3	2	4

Etwa 2/3 der befragten Lehrerinnen und Lehrer (n=39) äußert sich in den Interviews kritisch zu der Auswertung und Bewertung der Vergleichsarbeiten. Die Hälfte aller Lehrerinnen und Lehrer kritisiert die der Auswertung zugrunde gelegte Punkteverteilung, die nach deren Meinung unpräzise und dadurch ungerecht

sei. In engem Zusammenhang damit stehen die Kritikpunkte am Notenschlüssel und an der generellen Bewertung von Vergleichsarbeiten. Des Weiteren äußern sich die Lehrpersonen zu der Praktikabilität der vorgeschriebenen Auswertungstechnik.

5.3.1.1 Kritik an Bewertungspraxis

5.3.1.1.1 Punktevergabe ist unpräzise und ungerecht (Insg. 30; HS 4, RS 15, Gym 11)

Die meisten Interviewpartner, vor allem in der Realschule und im Gymnasium, bemängeln eine ungerechte und ungenaue Punktevergabe bei der Korrektur und stellen die Vergleichsarbeiten in Frage. Sie geben an, dass in der Korrekturanweisung weder der Schwierigkeitsgrad der Aufgabe noch die Teilleistungen Berücksichtigung finden. Die Lehrpersonen berichten auch darüber, ihre eigenen Klassenarbeiten um einiges differenzierter und gerechter zu bewerten.

> „Ich habe den Anspruch von der Diagnosearbeit in Frage gestellt, in vielen Punkten. (…). Da hat jemand der irgendetwas hin gemurkst hat, trotzdem die volle Punktzahl bekommen nach der Bewertung. Da habe ich jetzt nicht so sehr mich in Frage gestellt, wie die Vergleichsarbeit.“ (250)

> „Da wäre vielleicht anzumerken, dass die Items nicht ganz gerecht verteilt sind. Manchmal bekommt man einen Punkt, sag ich mal, für eine ganz, ganz leichte Aufgabe und für eine wirkliche schwierigere Leistung bekommt man den gleichen Punkt. Oder was auch ein Problem war, das weiß ich noch, da musste man zwei Dinge richtig haben, jetzt war ein Teil richtig und der andere nicht und es gab zusammen nur einen Punkt, oder ein Item. Das fand ich nicht ganz gerecht, also die Punkteverteilung innerhalb der Aufgaben daran erinnere ich mich noch.“ (161)

In diesen Zitaten kommen Willkür und Unprofessionalität beim vorgegebenen und somit festgeschriebenen Auswertungsprocedere eindeutig zum Ausdruck. Schließlich zählen Bewerten und Beurteilen zu den zentralen Arbeitsfeldern der Profession der Lehrerinnen und Lehrern, mit denen sie in ihrer täglichen Unterrichtspraxis sehr vertraut sind. Von daher ist es nachvollziehbar, dass sie diese undifferenzierte, nicht nachvollziehbare Bewertungspraxis stark kritisieren, weil sie in keiner Weise deren Handlungslogik und Vorgehensweise entspricht und sie als ungerecht erleben. Wie die einzelnen Lehrpersonen mit dieser Diskrepanz um-

gehen, ist unterschiedlich. Die einen befolgen die zentralen Anweisungen strikt und betonen ausdrücklich, dass sie dies nur im Rahmen der Vergleichsarbeiten, aber niemals in Klassenarbeiten anwenden würden. Schließlich schadet man mit diesem undifferenzierten Vorgehen den Schülerinnen und Schülern und bestraft sie mit schlechten Noten unverhältnismäßig.

> „So stand es in der Anweisung und die habe ich befolgt. Ich hätte es in meinen Klassenarbeiten einfach anders gemacht. Da hätte es eine andere Punkteverteilung gegeben und da wäre nur das eine Wort als falsch gerechnet worden und die drei anderen als richtig. Ich will ja auch niemanden bestrafen, wenn er einen kleinen Fehler macht und dann plötzlich alle Punkte flöten gehen.“ (162)

Bei einigen Lehrpersonen kommen Ärger und Wut in den Aussagen deutlich zum Ausdruck. Sie fühlen sich durch die strikten, undifferenzierten Vorgaben gegängelt und regen sich darüber auf, dass Vergleichsarbeiten somit ihr Potential nicht entfalten können. Würde man differenzierter bewerten, so könnten die Ergebnisse eindeutige Hinweise auf die Stärken und Schwächen der Schülerinnen und Schüler liefern und sich somit als Diagnoseinstrument eignen.

> „Es ist manchmal einfach ärgerlich sie zu korrigieren. Weil ich mich gegängelt fühle, ich fühl mich z.B. gegängelt wenn ich ein Item habe, das mir vier Rückmeldungen über Schülerleistung gibt und ich soll das als eine Rückmeldung werten. Also wenn z.B. drei Fragen gestellt sind und es gab einen Fehler, dann wird dieses Item nicht gewertet, dann ärgert mich das, weil ich eine Differenzierungsmöglichkeit, die ich für Billiggeld in der Hand habe nicht nutzen kann“ (365).

Die unklaren Vorgaben bei der Bewertung der Aufgaben führen teilweise zu Diskussionen innerhalb des Kollegiums, weil die einzelnen Lehrpersonen durch die unklaren Lösungsvorschläge verunsichert sind. So kam es in einigen Kollegien bezüglich der Bewertungen zu Absprachen, die sich von den zentralen Vorgaben unterscheiden.

> „Ich habe da gewisse Probleme muss ich sagen bei der Korrektur von den Vergleichsarbeiten, in der Hinsicht, weil die Lösungsvorschläge, die dabei waren, meiner Meinung nach teilweise bisschen schwammig waren. Das heißt, wir haben schon im Kollegium schon darüber diskutiert und zwar ging es einfach darum, dass, wenn da eine Lösung dran steht oder ähnliches beispielsweise, dann weiß ich nicht genau, kann ich das andere durchgehen lassen? Ja oder nein?“ (266)

> „Manchmal haben wir uns auch abgesprochen. Wir waren im Zwiespalt, was machen wir da? Der hat sich nur ein bisschen verrechnet, hat halt nicht genau die Lösung. (…). Dass ich gerne einen Punkt geben würde, wenn ich se-

> he, derjenige hat es verstanden, oder zwei Zahlen vertauscht oder so. Ist halt schade dann.“ (364)

Dass sich nicht alle Lehrerinnen und Lehrer an die Vorgaben halten, sondern individuell über die Bewertung einzelner Aufgaben entscheiden, zeigt auch das folgende Zitat. Diese Lehrperson gibt offen zu, dass sie sich bei der Bewertung der Vergleichsarbeiten nicht an die strengen Vorschriften hält, sondern auch dann Punkte vergibt, wenn Schülerinnen und Schüler nur Teilleistungen erbringen. Sie begründet ihr Handeln zum einen damit, dass die Vergleichsarbeit nicht nur als Rückmeldung dient, sondern als Note ins Zeugnis mit einfließt. Zum anderen kann sie es nicht mit ihrem Gewissen vereinbaren, bei der Bewertung nicht zum Wohle der Schüler beizutragen. Aus diesem Grund gleicht sie bei der Punktevergabe individuell aus und folgt somit ihren Überzeugungen und paidotropen Einstellungen.

> „Bei der Auswertung kam ich manchmal ein bisschen ins Schleudern, nicht ins Schleudern direkt, sondern dass ich da – Menschenskinder - jetzt von der Anweisung her soll ich so bewerten, da heißt also, wenn er das sagt dann das und dann das. Da hätte ich manchmal bei allen Sachen null Punkte geben sollen, wenn er aber in Wirklichkeit dann doch Sachen richtig getroffen hat. Und dann habe ich an der Stelle halt dann doch ausgeglichen, weil ich sage, das ist jetzt nicht nur die Rückmeldung für das. Der Schüler bekommt es ins Zeugnis rein und da stehe ich auf Seiten des Schülers, weil der kriegt es ins Zeugnis rein. Dass ich das ausgeglichen habe, Mensch, die da oben sagen ich muss es jetzt so und so machen. Das ist mein eigenes Gewissen da.“ (366)

5.3.1.1.2 Notenschlüssel ungerecht, da fehlende lineare Skalierung (Insg. 10; HS 0, RS 9, Gym 1)

Zehn Lehrerinnen und Lehrer, neun davon unterrichten an einer Realschule, kritisieren bei der Bewertung der Vergleichsarbeiten die nicht linearen Notenskalen, die von den Verantwortlichen so festgelegt wurden. Im Gegensatz zu diesen vorgegebenen Skalen müssen die Lehrpersonen bei eigenen Klassenarbeiten linear skalierte Punkte-Noten-Tabellen verwenden, die der Hälfte der Punktzahl die Note 3,5 zuordnet. Sie ärgern sich darüber, dass extern festgelegte, sonst geltende Maßstäbe in Klassenarbeiten eingehalten werden müssen und im Allgemeinen bei Nichtbeachtung von der Schulaufsicht beanstandet und verfolgt werden, während in Vergleichsarbeiten diese Standards nicht gelten. Dieses Vorgehen ist für die Lehrerinnen und Lehrer weder nachvollziehbar noch einleuchtend und wird als

Widerspruch erlebt. Die Äußerungen in den Interviews fallen in diesem Bereich durchgängig abwertend aus und beziehen sich auf diejenigen Akteure, die für die Gestaltung und Erstellung der Vergleichsarbeiten verantwortlich sind. Eine Lehrperson betitelt die Verantwortlichen als ‚Schloris' und bringt dadurch ihre Geringschätzung zum Ausdruck, weil ihrer Meinung nach die Vergleichsarbeiten nicht professionell sondern willkürlich und nicht nachvollziehbar konzipiert wurden. Eine Anhebung des Notenschnittes um eine ganze Note, birgt die Gefahr, dass die Akzeptanz und Wertschätzung der Vergleichsarbeit als externes Evaluationsinstrument sinkt. Dass dieser zentrale Punkt negative Auswirkungen auf die Einstellung und Haltung einiger Lehrerinnen und Lehrer haben könnte, in dem Sinne, dass Vergleichsarbeiten in ihrer Aussagekraft an Bedeutung verlieren, schwingt in den Äußerungen mit.

> „Ich habe bloß gedacht, das sind schon Schloris da irgendwie. Ich werde als Lehrer gezwungen Arbeiten zu schreiben, wo 3,5 die Hälfte der Punktzahl ist und anders darf es nicht bewertet werden. Und hier ist plötzlich 2,5 die Hälfte der Punktzahl. Also das hat mich schon sehr verblüfft." (265)

> „Aber, der Bewertungsmaßstab, der hier angewendet wurde, den dürfen wir ja im Grund genommen bei Klassenarbeiten gar nicht anwenden. Uns wird immer wieder gesagt, wir müssen eine lineare Bewertung anlegen, jeder Punkt muss eigentlich gleich viel wert sein und das ist ja hier total daneben." (252)

> „Ich versteh sowieso nicht, warum die Notenskala hier nicht proportional ist und ich muss das immer machen. (…). Was natürlich auch falsch war, das hätte ich nie als Benotung machen dürfen, es war keine lineare Notenskala. Das war eine ganz brutale. Hätte ich als Lehrer das so früher gemacht, hätte mir die Behörde, das Schulamt, die Aufgabe, die Arbeit gestrichen. So darf ich nicht benoten. Das ist ein bisschen Widerspruch für mich gewesen." (265)

> „... und warum nicht halbe Punktzahl eine 3-4 ist, sondern eine 2-3, das kapiert auch niemand. Also ich nicht zumindest." (256)

Einige Personen suchen nach Erklärungen, warum die Verantwortlichen bei der Bewertung der Vergleichsarbeiten von geltenden Standards abweichen. Eine Lehrperson bringt ebenfalls ihren Unmut zum Ausdruck. Sie vermutet den Grund darin, dass diese Notenskala absichtlich so festgelegt wurde mit dem Ziel, eine Normalverteilung zu erzwingen.

> „(Die) erzwingen das praktisch, dass es diese Normalverteilung gibt. Das ist hanebüchen, da gibt es Aufgaben, wo ich durch, was weiß ich, durch Glück oder Raten einen Punkt bekommen kann. Mit dem Punkt habe ich dann

> schon die Fünf oder so. Das ist einfach die Skala, die ist einfach nicht linear. Und das ist das, was die (Vergleichsarbeit) einfach zu gut macht. Von den Aufgaben her, wenn man die Aufgaben korrigieren würde, dann an eine gerade Verteilung hin, dann würden die Schnitte ganz anders aussehen, denk ich mal“ (251).

Welche Konsequenzen sich aus nicht linearen und geschönten Notenskalen noch ergeben können, zeigen die weiteren Äußerungen der Befragten: Eine Lehrperson berichtet von glücklichen Schülern, die sich über die vergleichsweise guten Ergebnisse freuen. Denn obwohl die Schülerinnen und Schüler angaben, in der Vergleichsarbeit wenig gewusst zu haben, erzielten sie mit der Note 3.0 noch verhältnismäßig gute Noten. In vergleichbaren Klassenarbeiten würde dieser Leistungsstand mit der Note 4,0 bewertet werden.

> „Die (Schüler) haben sich gefreut. Die waren relativ begeistert, als sie herauskamen, weil sie beim Schreiben schon gespürt haben, ich habe nicht so viel gewusst und trotzdem habe ich eine 3. Klasse! (…) Aber da die Benotung nach oben dermaßen locker ist, haben meine Schüler zwar im Durchschnitt nur 40% der Arbeit gelöst, haben aber einen Schnitt von 3, 0. Bei meinen normalen Noten wäre das eine 4,0. Mit einem richtigen Notenschlüssel wäre diese Arbeit den Leistungen der Schüler gerecht gewesen. So haben sie zu gute Noten bekommen.“ (256)

Diese ‚geschönten’ Ergebnisse wirken sich insbesondere für jene Schülerinnen und Schüler aus, deren schlechte Jahresleistung die Versetzung in die nächste Klasse gefährden. Realistische Bewertungsmaßstäbe würden die Nichtversetzung bestätigen. Im folgenden Fall wurde eine Schülerin, der die Nichtversetzung drohte, auf Grund des angehobenen Notenspiegels versetzt.

> „Und deshalb ist z.B. eine Schülerin, die vor der Vergleichsarbeit auf 4,5 stand, und die bei einer regulären Arbeit vielleicht sich hätte nicht verbessern können, die hat sich durch diese Arbeit verbessert und weil wir die mit gezählt haben, hat sie die Vier erreicht und konnte somit versetzt werden. In einem anderen Fall hätte es vielleicht eine Nichtversetzung bedeutet.“ (252)

Generell sorgen sich einige Lehrerinnen und Lehrer auch um ihre „Außenwirkung“ und um das Bild, das z.B. bei den Eltern entsteht. Wenn die Kinder bei landesweiten, standardisierten Tests plötzlich um einiges besser abschneiden, als der Gesamtdurchschnitt aller Klassenarbeiten, die von den Lehrpersonen konzipiert und bewertet wurden, kann man sich vorstellen, dass manche Eltern irritiert sind, wenn die Noten weit auseinander klaffen. Die Jahresnoten werden durch die veränderte Skala unglaubwürdig. Dies könnte einen Grund darstellen, warum vor al-

lem die Lehrerinnen und Lehrer an Realschulen diesen Punkt beklagen. Die Eltern wundern sich zu Recht, wenn staatliche Vorgaben im Durchschnitt die Leistungen der Schülerinnen und Schüler um einen Notenpunkt besser bewerten. Sie stellen die Bewertungspraxis der Lehrpersonen in Frage und vermuten, dass ihr Kind eventuell an ein Gymnasium wechseln könnte, da es im landesweiten Vergleich viel besser abschneidet und sie den Grund hierfür bei den Lehrpersonen suchen.

Einige Lehrpersonen berichten von den problematischen Auswirkungen der geschönten Vergleichsarbeitsergebnisse auf die Eltern.

> „Den Eltern wird da ein anderes Bild vorgespielt und wir stehen da als Lehrer eigentlich immer ein bisschen dumm da, wenn dann Schüler die bei uns über das Jahr eine Fünf haben, dann plötzlich eine Drei schreiben. Und das kam ja von der Punkteverteilung und wie die Noten zugeordnet waren.“ (252)

Nicht zuletzt aus diesem Grund lehnen die Lehrerinnen und Lehrer die ‚nivellierten' Ergebnisse aus Vergleichsarbeiten als Basis für Eltern-Lehrer-Gespräche ab, weil sie bezüglich der tatsächlichen Schülerleistungen keine Aussagekraft haben.

> „Eigentlich (eignet sich die VA) nicht, weil das von der Nivellierung her, ich hatte ja fast kaum einen Schüler unter 5, wo das dann in den kritischen Bereich gehen würde und deshalb nehme ich für ein Gespräch niemals als Grundlage das Ergebnis von einer Vergleichsarbeit.“ (258)

Zwei Lehrpersonen empören sich über den Notenspiegel, der ihrer Ansicht nach nicht legitimiert werden kann und sogar juristisch anfechtbar ist.

> „Was mir da natürlich aufgefallen ist, was wir so lernen, (die) Hälfte der Punktzahl muss 3,5 sein, das entspricht dem ja teilweise nicht. Das ist von Jahr zu Jahr unterschiedlich und das ist für mich uneinsichtig. Das ist für mich irgendwo willkürlich. Manchmal war die Hälfte Punktzahl schon drei manchmal gewesen, und manchmal war Hälfte der Punktzahl weiter unten gewesen da dann. Und das kapiere ich nicht. Das könnte ich beispielsweise bei einer normalen Klassenarbeit nicht machen, sonst rücken mir die Eltern auf das Dach, oder auch der Chef. Das wird bei uns verkündet, das wird auch vom Verwaltungsgericht so und so muss das sein. Das kapiere ich nicht, dass das bei den Vergleichsarbeiten so nicht gemacht wird.“ (259)

> „Das dürfen wir gar nicht, das ist ja keine Frage zur Anregung. Da kriege ich vom Schulamt eins auf den Deckel. Die Arbeit wird eingezogen. ‚Das können Sie so nicht machen'. Dann ist für mich gar kein Reflektionsgrund da. Das darf ich nicht machen. Ich muss auf 3,5 die Mitte legen und ich muss linear benoten. (…) Wie machen die das? Wieso wagen die das? Juristisch könnte man diese Vergleichsarbeiten, wenn sie jetzt einen entsprechenden Juristen an der Hand hätten, sicherlich anklagen, gell?“ (265)

5.3.1.1.3 Generelle Kritik an Bewertung der VA (Insg. 2; HS 0, RS 1, Gym 1)

Ein Realschul- und ein Gymnasiallehrer lehnen eine Bewertung von Vergleichsarbeiten grundsätzlich ab, weil dadurch Notendruck erzeugt wird. Sie begründen ihre Position damit, dass in der Schule auch ohne die Vergleichsarbeiten genügend Druck und Stress herrscht und dieser jeglicher Förderung entgegen steht. Die Schülerinnen und Schüler müssen am Ende eines anstrengenden Schuljahres innerhalb eines kurzen Zeitraums noch einmal drei Arbeiten schreiben, was eine große Belastung sowohl für diese, als auch für die Lehrerinnen und Lehrer darstellt.

> „Warum müssen die bewerten, warum müssen die überhaupt bewertet werden. Warum kann man sie nicht schreiben ohne den Druck der Noten im Hintergrund.“ (268)

> „Warum verknüpft man das mit einer Notengebung? Es ist schon ohnehin genug Stress in der Schule da, wenn Sie ein klassisches Buch nehmen von Meister ‚Die Förderung der schulischen Lernmotivation’. Man sollte eher auf Neugier und andere Motive achten, denke ich. Lernen und Leistungsdruck ist schon genügend vorhanden. Warum also hier noch mal am Ende von einem ohnehin stressigen Schuljahr in einem Zeitrahmen von, muss man sich mal klar machen, von neun Tagen für diese Kleinen, drei Arbeiten - Freitag, Mathe, dann Montag Deutsch, oder Bio, und dann ein paar Tage darauf Deutsch, die Reihenfolge ist ja jetzt egal - also, wer sich mal im schulischen Alltag auskennt, der wird sehen, das stellt für zwei Seiten eine große Belastung dar!“ (350)

5.3.1.2 VA sind einfach und schnell zu korrigieren (Insg. 9; HS 3, RS 2, Gym 4)

Dass die Vergleichsarbeiten am Computer einfach und schnell zu korrigieren sind, erfreut insgesamt neun Personen. Vor allem im Vergleich zu eigenen Arbeiten stellt dieses Auswertungsverfahren eine Erleichterung dar.

> „War von der Auswertung wunderbar, ging sehr schnell, das Eingeben in den Computer, das Programm war klasse.“ (161); „Mit der Eingabe, das ist ja relativ einfach.“ (154); „(Die Vergleichsarbeit ist) leichter zu korrigieren als jede andere Arbeit.“ (263); „Ich habe noch nie so eine schöne Arbeit zu korrigieren gehabt. (Lachen). Auch wenn es mit einer der schlechteren

> Schnitte in der Klasse war, aber die Korrekturzeit hat sich auf eine halbe Stunde oder so reduziert durch die Items.“ (368)

Den Grund, warum die Auswertung so zeitsparend durchgeführt werden kann und welche Probleme damit zusammenhängen könnten, nennen ein Gymnasial- und ein Hauptschullehrer. Die Lehrpersonen erhalten durch die eindeutigen Vorgaben keinen Bewertungsspielraum oder Interpretationsmöglichkeiten, sondern können stur nach Vorlage die entsprechenden Kreuze in die Excel-Tabelle eingeben, was ein rasches Vorgehen ermöglicht. Dennoch schwingt in den Beiträgen dieser Personen Kritik an der ungleichen und undifferenzierten Bewertungsvorgabe mit.

> „Die Auswertung, da ist man ja auch ziemlich streng reglementiert, sag ich mal, und streng geführt, da bleiben einem ja eigentlich kaum Freiräume. Und das ist insofern gesehen manchmal sogar einfach, dadurch dass man relativ schnell abhaken kann. Ja.“ (367)

> „Klar, es ist vom Korrigieren her leichter, richtig. Ein Kreuz, falsch, keines, also es ist von der Auswertung her leichter. Aber manchmal müsste man überlegen, ob man da nicht zwei Punkte für die schwierigere oder die größere Leistung gibt und dann halt einen Punkt nur. Und dann wirklich nicht zwei Dinge mit einem Punkt abhakt.“ (161)

Ein Hauptschullehrer lobt die Qualität der Auswertung der Vergleichsarbeit: „Also ich finde die Auswertung recht differenziert und umfassend.“ (151)

5.3.2 Reflexion – Hauptkategorie II: Aufgaben und Zusammenstellung der Vergleichsarbeit

Tabelle 5.3.2: Reflexion – Hauptkategorie II

HK II: Aufgaben und Zusammenstellung der Vergleichsarbeit				
Subkategorien	**Ges (n=59)**	**HS (n=15)**	**RS (n=23)**	**Gym (n=21)**
1. Kritik an einzelnen Aufgaben- und Fragestellungen	26	8	9	9
Kritik am multiple choice Format	6	4	0	2
Sprache schülerfremd	6	0	3	3
Sonstige Kritik	11	3	4	4
2. Kritik an Zusammenstellung der VA und Aufgabenauswahl	20	6	7	7

5.3.2.1 Kritik an einzelnen Aufgaben- und Fragestellungen (insg. 26; HS= 8, RS=9, Gym = 9)

Die Kritik an den einzelnen Aufgaben bezieht sich in den häufigsten Fällen auf das multiple choice – Format der meisten Aufgaben. Vier Hauptschullehrpersonen und zwei des Gymnasiums wünschen weniger Ankreuz-Aufgaben, sondern Formate, die mehr produktive Leistungen erfordern und Prozesse abbilden. Denn dadurch kann ihrer Meinung nach eher abgelesen werden, ob die Schülerinnen und Schüler ein Verständnis des Lerngegenstandes entwickeln konnten oder an welchen Punkten Schwierigkeiten herrschen. Möchte man Vergleichsarbeiten im Sinne eines Diagnoseinstruments nutzen, ist dieser Kritikpunkt zentral. Je differenzierter die Arbeiten die Leistungen der Schülerinnen und Schüler abbilden, desto präziser können die Lehrpersonen geeignete individuelle Fördermaßnahmen ableiten. Leisten Vergleichsarbeiten diese Präzisierung und Differenzierung nicht, so liegt es auf der Hand, dass die Arbeiten eher summativ, also leistungsbilanzierend verwendet werden als formativ und somit als Diagnose- und Förderinstrument.

> „Ich hätte mir gewünscht, dass eigene produktive Aufgaben stattgefunden haben. Also dass der Schüler zum Beispiel auch mal ganz konkret ein Dreieck zeichnen muss oder Parallelen einzeichnen muss. Also gerade im Bereich Geometrie wäre das sicherlich gut möglich gewesen, so dass ich auch mal schauen kann, ob da ein Verständnis vorliegt. Und das gleiche gilt auch für Deutsch.“ (151)

> „Ja, ich fand die Art und Weise, man musste ja fast nichts schreiben. Und für mich in Deutsch ist das einfach nicht akzeptabel. Wenn man nur ankreuzen muss, klar Textverständnis, man muss verstehen. Aber, meines Erachtens mussten sie zu wenig schreiben. Und sobald unsere Hauptschüler schreiben müssen dann ist es aus. Bei den Meisten halt.“ (155)

> „Ja das ist die Art Aufgabenstellung, ja, denk ich. Also, weil sie auf der einen Seite, wie soll man sagen, sehr guten Schülern nicht die Möglichkeit gibt, alles zu zeigen, was sie können, ja? Und auf der anderen Seite so billige Aufgaben dabei sind, dass halt auch die schlechten Schüler vielleicht auf einen Vierer wieder hochkommen.“ (356)

Einen weiteren Kritikpunkt den jeweils drei Realschul- und Gymnasiallehrpersonen nennen, betrifft die schülerfremde Sprache mancher Aufgabenstellungen. Die Schülerinnen und Schüler verstehen die Aufgabe nicht und können sie, trotz vorhandener Fähigkeiten und Fertigkeiten, nicht lösen.

> „Bei der ersten Frage zum Beispiel gab es Probleme im Verständnis bei der Formulierung der Frage. Die Fragen sind für uns Lehrer klar, aber für Schüler nicht immer ganz klar formuliert. Da müsste man vielleicht ein bisschen daran feilen, dass diese Fragen wirklich auch schülergerecht formuliert werden.“ (351)

Von den Lehrpersonen werden im Bereich der Aufgaben weitere Punkte genannt, die nach Meinung der Lehrpersonen Mängel aufweisen. Sie beziehen sich u. a. auf die Komplexität der Aufgabenstellung, auf die sowohl fehlende Eindeutigkeit als auch auf Fehler bei der Aufgabenstellung, auf die unübersichtliche Aufgabendarstellung und die schülerfremde Sprache bzw. Themenbereiche. Diese Kritikpunkte sind insgesamt betrachtet sehr spezifisch, da sie sich auf ganz konkrete Aufgaben der Vergleichsarbeiten aus dem Jahr 2007 beziehen und somit nicht generalisiert werden können. Im Folgenden werden die einzelnen Kritikpunkte kurz aufgeführt und zur Verdeutlichung mit einem passenden Interviewzitat belegt:

- zu komplexe Aufgabenstellungen

> „Ich glaube, die Art der Fragestellung. Also, nicht die Inhalte die abgefragt wurden, sondern die Art der Fragestellung, die Komplexität mancher Fragestellung.“ (253)

- mangelnde Komplexität der Aufgaben

> „Aber Dinge wie Grammatik sind völlig ausgegrenzt worden. Natürlich fragt man sich dann, warum mache ich das dann überhaupt noch mit meinen Schülern? Wenn man dann wiederum sagt, um die Sprache besser zu beherrschen, dann wird man das wiederum nur feststellen, wenn man Textproduktion rein bringt. Textproduktion gab es ja aber auch nicht. Ich würde mir nur wünschen, dass die Aufgabentypen, besonders in Deutsch, anders gestellt werden, umfangreicher gestellt werden.“ (153)

- nicht eindeutige Aufgabenstellung

> „Das war auch dieses Mal in dieser Pilotierungsarbeit, das Geheimnis kann ich verraten. Da waren auch Dinge drin, wo ich mir selber überlegen muss, worauf will die Aufgabe eigentlich hinaus. Dass war in ein oder zwei Fragestellungen dermaßen unverständlich oder missverständlich und wenn ich den Kindern keine Hilfeleistung geben kann. Also, so würde ich eine Aufgabe niemals formulieren. Und ich darf ja keine Hinweise liefern. In einer Klassenarbeit kann ich mal einen Kommentar dazu geben.“ (260)

> „Also jetzt noch mal zu dieser einen Zeichnung mit diesen Blüten und der Bienen, weiß ich noch, die war ein bisschen verwirrend. Die fand ich nicht ganz eindeutig. Das weiß ich noch, dass wir da auch untereinander, die Bio-Kollegen, dann diskutiert haben. Das war irgendwie nicht ganz eindeutig.“ (352)

„Also, gerade z.B. diejenigen die schlechter abgeschnitten haben als sonst. Es gab so ein paar Fragen, die nicht ganz eindeutig waren für die Schüler. Also, nicht weniger, aber da hingen gleich eine ganze Punkte mit zusammen. Wenn die jetzt einmal falsch verstanden haben, bei einer Aufgabe wo man fünf Items haben konnte, dann waren gleich vier weg oder so. Das hatte natürlich gleich immense Auswirkungen.“ (358)

- unübersichtliche Darstellung der Aufgabe und oberflächliches Bearbeiten

„Also, teilweise war es natürlich nicht übersichtlich, das Ganze. Zum Beispiel gerade bei Aufgabe 11, wenn ich da den Flächeninhalte sehe. Viele sehen nur "Flächeninhalt" und kreisen dann schon unten eigentlich das Ganze ein. Klar, das ist ein Fehler von den Schülern, weil sie nicht genau richtig lesen. Aber das war vielleicht ein bisschen umständlich das Ganze gemacht.“ (152)

„Mir fehlte so ein bisschen mehr dieses übergreifende Textverständnis, also dass man auch Fragen darüber hinaus stellt. Es war doch sehr am Text und sehr, ja... Ich hatte dann zum Teil Kinder, die dann einfach die Aufgabenstellung nicht gelesen (…). Also, die haben einfach nicht richtig gelesen. Und die dachten, ja klar, da muss man Fragen dazu finden. Aber die waren ja vorgegeben.“ (163)

- Fehler in den Aufgabenstellungen

„Da sind teilweise auch Aufgabenstellungen drin gewesen, die falsch waren, schlichtweg falsch. (…) Teilweise mit falschen Formulierungen, mit falschen Fragestellungen. Ich habe das im Übrigen auch mitgeteilt, dann auch geschrieben. Weil ich das einfach auch so nicht akzeptieren will.“ (260)

- Themenwahl benachteiligt einige Kinder

„Ich fand, bei uns in der Klasse, wir haben Streitschlichter selber an der Schule, dann fand ich es natürlich dann wieder ein Vorteil für die Kinder, die selber am Streitschlichter teilnehmen und sich damit besser auskennen, wie Kinder die jetzt nicht aktiv darin involviert sind. Von dem her fand ich es da eine Problematik, von daher hätte man da eher einen Text wählen sollen, was jetzt vielleicht außerhalb der Schule, wo vielleicht jedes Kind einen ähnlichen Zugang hat, oder ja wo einfach die Vorraussetzungen vielleicht ähnlicher sind.“ (163)

5.3.2.2 Kritik an Zusammenstellung der VA und Aufgabenauswahl (insg. 13; HS 4, RS 5, Gym 4

Insgesamt 13 Personen bewerten die Zusammensetzung der Vergleichsarbeiten und die Aufgabenauswahl als zu einseitig. Viele wichtige Themen würden in den Vergleichsarbeiten nicht berücksichtigt. Diese Argumente finden sich in allen Fächern wieder, was folgende Auswahl zeigt:

> „Was ich jetzt wirklich als Kritik, und nicht nur ich, sondern die Deutschkollegen und zwar alle haben uns etwas mokiert darüber, das ist die Einseitigkeit. Also, mit Abstand die meisten Aufgaben bezogen sich dann auf das Textverstehen. Einerseits schon okay, ja das ist ein wichtiger Punkt, aber die Grammatik beispielsweise ist viel zu kurz gekommen.“ (367)

> „Es ist natürlich schwierig, vom ganzen Ablauf her, wird vieles nicht berücksichtigt, was auch viele machen. Zum Beispiel Körperberechnung oder so. In die Richtung war in Mathematik zum Beispiel gar nichts.“ (152)

Diese einseitige Themenauswahl führt auf Lehrer- und Schülerseite zu Unzufriedenheit und Enttäuschung. So beklagen einige, dass bei der Themenauswahl zentrale Inhalte nicht abgefragt wurden, obwohl ein Großteil des Unterrichts für die Erarbeitung dieser Themen verwendet wurde. Enttäuscht stellen sie fest, dass anstelle dieser wichtigen Themen ihrer Meinung nach unwichtige, „winzige“ Kapitelchen abgefragt wurden. Diese Auswahl veranlasste auch die Schülerinnen und Schüler zu Fragen, die sie vornehmlich an ihre Lehrerinnen und Lehrer richteten, warum denn jene Themen nicht in der Vergleichsarbeit abgeprüft wurden, die man so intensiv gelernt hätte.

> „Es war sehr, sehr enttäuschend und auch die Schüler waren enttäuscht, ja.“ (367)

> „Ich war ein bisschen unzufrieden, dass wichtige Punkte überhaupt nicht abgeprüft worden sind. Von der Grammatik her, waren wichtige Dinge gar nicht drin.“ (159)

> „Und ich war von den Vergleichsarbeiten in Klasse 6 maßlos enttäuscht, dass viel zu wenig Bruchrechnung und Dezimalrechnung dran kam. Das haben wir das gesamte Jahr durchtrainiert. Stattdessen kam ein ganz winziges Kapitelchen draus dran, diese Sachen mit dem Zahlenstrahl und so. Und bei den 8-ern, da habe ich auf Teufel komm raus Formelumformungen, binomische Formeln und alles trainiert. Das kam fast gar nicht dran oder überhaupt nicht dran, gell? Also insofern waren die Schüler auch enttäuscht, die haben

> dann gesagt: „Herr <xxx>, wieso haben Sie das ganze Zeug so häufig trainiert? Das kam ja fast nicht dran. (…). Die ganzen Basisaufgaben kamen nicht dran. Es war für mich etwas frustrierend, muss ich sagen.“ (265)

Des Weiteren wird beklagt, dass die eingeschränkte Auswahl der Aufgaben die geforderten Kompetenzen des Bildungsplans nicht valide abbildet. Am Beispiel des Faches Deutsch argumentiert die Lehrperson wie folgt:

> „Die Deutsch-Vergleichsarbeit müsste meiner Meinung nach überarbeitet werden, denn der Deutsch-Bildungsplan, der gibt doch viel viel mehr an Kompetenzen vor als das, was in Deutsch dann letzten Endes abgefragt wurde.“ (153)

Eine andere Hauptschullehrperson kritisiert, dass die einseitige Zusammenstellung der Vergleichsarbeit nichts über die im jeweiligen Fach erworbene Kompetenz der Schüler aussagt. Würde die Aufgabenauswahl anders getroffen, könnten sich ihrer Meinung nach die Ergebnisse der Schülerinnen und Schüler sogar ins Gegenteil umkehren. Diese Lehrperson zweifelt auch an der Fähigkeit der Verantwortlichen bei der Erstellung der Vergleichsarbeit.

> „Weil, wie gesagt vorher, das sind so Unterschiede, da legt jemand einfach jetzt sehr viel Wert auf andere Dinge, und die werden hier auf einmal nicht abgefragt. Obwohl die Schüler vielleicht, wenn es eine andere Vergleichsarbeit wäre, vielleicht das Ganze sich umdreht. Ich bin mir absolut überzeugt, wenn man es anders stellt und andere Aufgabenstellungen hat, dann könnte sich das ganze drehen. Und es heißt ja nicht, dass die Leute, die die Aufgabenstellungen gemacht haben, dass das auch gut ist. Man kann es auch gerade andersrum sehen.“ (152).

Eine Gymnasiallehrperson ärgert sich sehr darüber, dass in den Vergleichsarbeiten nicht konkrete Themengebiete abgefragt werden, die auch im Unterricht behandelt wurden. Ihre Ablehnung wird auch durch ihre despektierliche Wortwahl deutlich, da sie Vergleichsarbeiten als „Müll“ bezeichnet.

> „Das ist ja das Aberwitzige mit diesen Vergleichsarbeiten, dass sie jetzt an der Basis dazu führen, dass man diesen Müll teilweise, einübt.
> I: Finden Sie nicht gut, was die da fragen oder warum Müll?
> L: Wieso fragen die keine konkreten grammatischen Sachen. Satzglieder, Wortarten. Man macht das in der fünften und sechsten Klasse. Es wird nicht abgefragt korrekt. Da werden nur so wischi waschi Fragen beantwortet und gefragt.“ (359)

Den Wunsch nach mehr Transparenz, wie z.B. die Aufgabenauswahl von den Verantwortlichen getroffen wurde, äußert diese Lehrperson. Sie hätte sich gerne am Auswahlprozess und der Zusammenstellung der Vergleichsarbeit beteiligt, um

u. a. die Gründe und Ziele zu erfahren, die bei der Zusammenstellung der Vergleichsarbeiten leitend waren. Denn nur, wenn man die Ziele kennt, sich darüber intensiv im Diskurs auseinandersetzt und klärt, kann das Instrument zur Evaluation und somit zur Qualitätsverbesserung oder Qualitätssicherung einen Beitrag leisten. Diese Aussage umfasst die Erkenntnisse, die im Theorieteil im Kapitel „Qualität“ erörtert wurden. Alle Beteiligten sollten an der Zielklärung beteiligt werden, bzw. darüber umfassend informiert werden, damit die Ergebnisse aus Vergleichsarbeiten und die gewonnenen Erkenntnisse in Maßnahmen umgesetzt werden können.

> „Ich hätte mir eher gewünscht, dass man erfahren hätte warum die dieses Diktat ausgesucht haben. Das war ganz unkindgemäß. Ich hätte gerne gewusst, warum ausgerechnet so etwas, haben die damit etwas Besonderes beabsichtigt? Oder bei der Grammatik? Warum hat man sich jetzt spezialisiert auf diese und diese Art Fragen. Ich kann es echt nicht verstehen. (…) Ich würde mir lieber wünschen, dass man mal darüber diskutiert, wie die Aufgaben ausgewählt wurden, warum welche Aufgaben ausgewählt werden. Was man da dran jetzt evaluieren will, an der Art der Aufgabe. Und so weiter und so fort.“ (360)

5.3.3 Reflexion - Hauptkategorie III: Niveau der Vergleichsarbeit

Tabelle 5.3.3: Reflexion – Hauptkategorie III

HK III: Niveau der Vergleichsarbeit				
Subkategorien	**Ges (n=59)**	**HS (n=15)**	**RS (n=23)**	**Gym (n=21)**
1. Niveau der VA zu niedrig	29	13	13	3
2. Niveau der VA zu hoch	4	0	1	3
3. Niveau der VA angemessen	2	0	0	2

5.3.3.1 Niveau der VA zu niedrig (Insg. 29; HS 13, RS 13, Gym 3)

Das Niveau der Vergleichsarbeiten wird von der Hälfte der Befragten, jeweils 13 Lehrerinnen und Lehrern der Haupt- und Realschulen und drei aus dem Gymnasium, als zu niedrig eingestuft. Auffallend ist bei der Betrachtung der einzelnen

Schularten, dass beinahe alle Hauptschullehrpersonen diese Aussage teilen. Hier stellt sich die Frage, warum ausgerechnet in der Schulart mit der leistungsschwächsten Schülerschaft das Niveau der Vergleichsarbeiten als zu niedrig angegeben wird? Von welchen Ausgangsbedingungen gingen die Verantwortlichen bei der Konzeption aus? Wie setzte sich die Pilotierungsstichprobe der Hauptschule zusammen, dass in deren Folge landesweite Vergleichsarbeiten konzipiert wurden, die im Niveau derart einfach sind, dass keinerlei Aussagen zur Leistungsfähigkeit der Schülerinnen und Schüler abgeleitet werden können? Hier schließt sich für die Auswertung der Interviews eine zentrale Fragestellung an, nämlich die, wie weit die Aussagen der Lehrerinnen und Lehrer zur Nutzung von Vergleichsarbeiten verwendet und gewertet werden können, wenn davon ausgegangen wird, dass das Niveau in keiner Weise dem Niveau der Schülerinnen und Schüler entspricht? Diese Feststellung muss bei der Interpretation dieser Daten stets berücksichtigt werden.

> „Ich fand sie sehr einfach.“ (163)

> „Diese Aufgaben waren relativ einfach zu machen.“ (266)

> „Ich fand sie allerdings ein bisschen zu einfach, muss ich ehrlich sagen, weil ich habe Schüler, vor allem zwei, die waren wirklich nicht gut in Englisch und können wirklich auch wenig auch mal selber produzieren, und einen Satz sagen oder so, das klappt nicht. Das sollten sie eigentlich am Ende der sechsten Klasse schon mal können und die haben da wirklich, ich glaube die eine hat eine 3,0 geschrieben und die andere eine 2,5 oder so. Das fand ich dann wirklich seltsam, dass die so gut plötzlich abschneiden. Vielleicht mache ich sonst zu schwere Arbeiten, das kann auch sein. (LACHT).“ (162)

Welche Folgen Lernstandserhebungen mit einem zu niedrigen Niveau haben können, wird von den Lehrerinnen und Lehrern sehr unterschiedlich eingestuft. Einige führen die positiven Auswirkungen der guten Leistungsergebnisse auf die Lerneinstellung und Motivation ihrer Schülerinnen und Schüler an.

> „Die Schüler haben schon gesagt, dass es eine leichte Arbeit war. Sie haben sich dann letztendlich gefreut, über die gute Note. Das hat natürlich einen positiven Effekt, dass viele, die jetzt eigentlich eher schlechter waren in Mathe, da auch mal ein Erfolgserlebnis hatten.“ (151)

Dass die guten Ergebnisse auch trotz strenger Bewertungskriterien erzielt wurden, berichtet diese Hauptschullehrerin. Sie sorgt sich um die möglicherweise unrealistischen Selbsteinschätzungen ihrer Schülerinnen und Schüler als Folge der irrealen Rückmeldeergebnisse.

> „In Deutsch [wurde] das Bild einfach verfälscht, also, dass die Schüler zu gut weg gekommen sind. Also, ich habe schon sehr streng bewertet. Da war eine Aufgabe, da hat man so ein bisschen Freiheit gehabt und da war ich schon sehr streng. Und war dann wirklich dann entsetzt, wo ich dann, also es hat mich gefreut für die Schüler. Aber die Schüler denken dann auch, Mensch bin ich gut, und das ist ja eigentlich nicht der Fall, sie sind in Deutsch ja nicht gut.“ (155)

Auf die Frage, warum die Ergebnisse in ihren Klassen so gut ausfielen, machen die Lehrpersonen sowohl die Art und Weise der Aufgabenstellung als auch das Antwortformat verantwortlich. Diese genannten Punkte beziehen sich demnach auf die Ergebnisse, die bereits in den Kapiteln zur Kritik an einzelnen Aufgaben als auch an der Zusammensetzung der Vergleichsarbeiten (5.3.2.1. und 5.3.2.2) ausführlich erörtert wurden.

> „Wir haben, wie man sieht, viel Schüler mit Migrationshintergrund, die haben einfach im sprachlichen Bereich unwahrscheinlich viele Probleme. Und wenn ich dann sehe, das ist eine Vergleichsarbeit mit gut, das ist einfach nicht gut. (…) Ja, ich fand die Art und Weise, man musste ja fast nichts schreiben. Und für mich in Deutsch ist das einfach nicht akzeptabel. Wenn man nur ankreuzen muss, klar Textverständnis, man muss verstehen. Aber, meines Erachtens mussten sie zu wenig schreiben. Und sobald unsere Hauptschüler schreiben müssen dann ist es aus. Bei den meisten halt. (155)

Der Kritik der undifferenzierten Betrachtung schließt sich auch dieser Lehrer an, der der Vergleichsarbeit eine mangelnde Aussagekraft anlastet:

> „Diese Aufgaben sind für einen starken Schüler keine Herausforderung. Die differenzieren meines Erachtens nicht. Die differenzieren nicht zwischen ausreichend, befriedigend, gut, sondern die differenzieren zwischen ausreichend und nicht ausreichend.“ (272)

Neben der Freude über die guten Noten, finden sich Lehrpersonen, die sich einerseits um den sinkenden Leistungsanspruch der Hauptschule sorgen, andererseits befürchten sie, dass aufgrund der abweichenden Ergebnisse vom zu erwartenden Leistungsstand die Professionalität der Lehrenden in Frage gestellt wird. Auch dieser Punkt wurde bereits im vorangegangenen Kapitel detailliert beschrieben.

> „Ich glaube nicht, dass ich zu streng bin, ich glaube eher, dass die Arbeit zu leicht war. (…) Eine gewisse Anforderung, meine ich, brauchen wir schon. Bloß die Anforderungen den Schülerleistungen anzupassen finde ich auch nicht so ganz richtig.“ (157)

> „Ich denke gerade unsere Hauptschüler, die müssen gefordert werden und das Niveau der Vergleichsarbeit 2007 war meines Erachtens zu gering. Wie gesagt der Vergleich, normalerweise ein Schüler bei mir mit fünf, schreibt

> eine Vergleichsarbeit mit zwei. Da stehe ich natürlich als Lehrerin blöd da, weil es ja eigentlich nicht dem Leistungsstand des Schülers entspricht.“ (155)

Wieder andere werten die Vergleichsarbeit aufgrund des niedrigen Niveaus als nichtssagend und setzen sich mit den Ergebnissen nicht weiter auseinander. Sie begründen diese Entscheidung damit, die sich ergebenden Konsequenzen, z.B. die Empfehlung zum Wechsel der Schülerinnen und Schüler auf die Realschule, nicht tragen zu können.

> „Ich habe halt gemerkt, dass es massenhaft Aufgaben gab, wo der Reihe nach jeder Schüler lösen konnte. Oder fast jeder. Und da war für mich der Inhalt dieser Rückmeldung, dass es einfach zu nichtssagend ist. (…) Wenn ich aufgrund dessen Zeugnisnoten verteilen müsste, sollte, mich trauen würde, ich weiß gar nicht wie ich das ausdrücken sollte, ja? Ja dann müsste ich ja die Hälfte meiner Mannschaft auf die Realschule schicken.“ (158)

> „Den Eltern wird da ein anderes Bild vorgespielt und wir stehen da als Lehrer eigentlich immer ein bisschen dumm da, wenn dann Schüler die bei uns über das Jahr eine Fünf haben, dann plötzlich eine Drei schreiben.“ (252)

Eine Realschullehrerin vermutet, dass das niedrige Niveau der Vergleichsarbeiten bildungspolitisch gewollt und gezielt festgelegt wurde. Sie begründet ihre Vermutung damit, dass ein niedriges Niveau auf weniger Widerstand stößt und in den Folgejahren sukzessive erhöht werden kann.

> „Ich glaube, dass die [VA] nur vorsichtig gestellt sind. Weil, von dem was die können müssten, am Ende von Klasse sechs, kann man noch mehr raus holen. Ich glaube, dass jetzt am Anfang die Vergleichsarbeiten noch einfacher sind, damit man es sich nicht verscherzt. Und dass die dann jetzt langsam anziehen. Ich glaube, ich erwarte, dass diese Vergleichsarbeit schwieriger wird als die vom letzten Jahr. Und dass es sukzessive dann auf ein Niveau kommt, wo man sie dann auch haben will. Aber im Moment sind die, meiner Meinung noch nicht adäquat für das was die am Ende von Klasse acht oder sechs müssen.“ (251)

Ein Realschullehrer sieht im niedrigen Niveau die Gefahr, dass die Schülerinnen und Schüler dazu verleitet werden, weniger zu arbeiten, anstatt dass die Leistungsbereitschaft erhöht wird.

> „Und von daher gesehen, habe ich jetzt aus den Vergleichsarbeiten eigentlich nur den Rückschluss gezogen, dass die mir zu human sind. Die Kinder eher verleitet, also wenn man das jetzt im Prinzip erklären würde, würde das die Kinder eher verleiten weniger zu tun, als mehr.“ (250)

5.3.3.2 Niveau der VA zu hoch (Insg. 4; HS 0, RS 1, Gym 3)

Nur drei Gymnasiallehrpersonen und eine aus der Realschule stufen das Niveau der Vergleichsarbeiten und den Umfang als zu hoch ein. Zu ihrem Urteil kommen sie, weil sie zum Vergleich Arbeiten aus dem Vorjahr heranziehen. Einen weiteren Grund warum sie das Niveau der Vergleichsarbeiten höher einschätzen als das Niveau sonstiger Klassenarbeiten, liegt in der Komplexität der Aufgabenstellung. Einige Aufgaben in Vergleichsarbeiten sind umfangreicher und komplexer konzipiert, so dass sie den Schülerinnen und Schülern größere Schwierigkeiten bereiten.

> „Es war zu umfangreich. Wir haben die anderen Vergleichsarbeiten ja jetzt auch schon vorliegen. Das war wesentlich mehr, als die anderen Jahre.“ (257)

> „Das Niveau, war ja, weil die Aufgaben oft etwas anders sind als die, die man so in üblichen Klassenarbeiten stellt. Es ist schon für manche Kinder schwerer gewesen.“ (361)

> „Wobei ich sie schwieriger fand als die letztes Jahr, die Vergleichsarbeit, aber ... Ja nicht unbedingt schwieriger. Also auf jeden Fall im Niveau angezogen? Ja, hm hm.“ (364)

5.3.3.3 Niveau der VA angemessen (Insg. 2; HS 0, RS 0, Gym 2)

Dass das Niveau der Vergleichsarbeit angemessen war, äußern lediglich zwei Lehrpersonen im Gymnasium

> „(Das Niveau) war in Ordnung. Das zeigt ja auch der Durchschnitt, dass die normal war.“ (352)

> „Vom Schwierigkeitsgrad, könnte ich pauschal nicht sagen, ob zu leicht oder zu schwer. Andere Aufgabentypen, ich könnte mich da nur wiederholen und vom zeitlichen Umfang her ein bisschen knapp bemessen. War so unser erster Eindruck, aller Kollegen. Wir haben ja morgens die Aufgaben bekommen und haben gedacht, ob die Schüler das schaffen, weil die Texte doch umfangreich waren. Da musste mit den Texten gearbeitet werden und wir hatten so den Eindruck, dass es vom zeitlichen Umfang ein bisschen knapp bemessen ist, hat sich aber hernach als grundlos herausgestellt. Wir sind ganz gut mit der Zeit raus gekommen. Also, Schwierigkeitsgrad angemessen.“ (355)

5.3.4 Reflexion - Hauptkategorie IV: Zeitpunkt der Vergleichsarbeit

Tabelle 5.3.4: Reflexion – Hauptkategorie IV				
HK IV: Zeitpunkt der Vergleichsarbeit				
Subkategorien	**Ges (n=59)**	**HS (n=15)**	**RS (n=23)**	**Gym (n=21)**
1. Zeitpunkt zu spät, keine Aktionen mehr möglich	17	2	4	11
Zeitdruck, viele andere Termine	9	0	3	6
2. Zeitpunkt zu früh	5	0	5	0
3. Zeitpunkt verschieben	4	1	3	0

5.3.4.1 Zeitpunkt zu spät, keine Aktionen mehr möglich (Insg. 17; HS 2, RS 4, Gym 11)

Insgesamt sind 17 Personen davon überzeugt, dass aus den Ergebnissen der Vergleichsarbeiten keine Maßnahmen mehr abgeleitet werden können. Die größte Gruppe bilden die Lehrerinnen und Lehrer an Gymnasien (11 von 21). Als Grund führen sie den späten Zeitpunkt der Durchführung der Vergleichsarbeiten an, der kurz vor dem Schuljahresende anberaumt war. Selbst diejenigen, die sich für die Ergebnisse interessieren und die sich mit den Rückmeldungen intensiv auseinander gesetzt haben, sehen aufgrund der späten Durchführung keine Möglichkeit, Maßnahmen im Sinne von Verbesserungen abzuleiten. De Facto bleiben den Lehrerinnen und Lehrern nach der Durchführung und Auswertung der Vergleichsarbeit nur noch drei Wochen am Ende des Schuljahres, in denen sie konkrete Maßnahmen durchführen könnten. Und dieser Zeitraum ist viel zu kurz, um alle erforderlichen Erkenntnisse im Unterricht nachhaltig umzusetzen.

> „Das Interesse war schon da, aber es ist ein kleines Problem dadurch, dass die so spät geschrieben werden, im Juni. Da sage ich bestimmt was, was sie nicht das erste Mal hören. Es bleibt ja die Zeit gar nicht das richtig auszuwerten. Ich kriege zwar eine Rückmeldung in dem Sinn, dass ich es schön im Balkendiagramm sehe, ob ich besser oder schlechter bin wie der Landesdurchschnitt, ob meine Klasse gut oder schlechter abgeschnitten hat. Dann werden mir viele Prozentzahlen nur so um die Ohren gehauen, wie hier. Und

> (ich) sehe dann immer anhand der Prozentzahlen ob ich besser oder schlechter bin. Aber bis ich das erfahre ist Mitte Juni und dann bleiben mir drei Wochen ungefähr, theoretisch rein drei Wochen, um jetzt das zu evaluieren. Sprich, um aus dem irgendwelche Vorteile zu ziehen und sagen ‚das hat meine Klasse noch nicht richtig behandelt'; ‚das habe ich zu wenig behandelt'; ‚das haben sie nicht verstanden', und so weiter." (268)

Dieses Zitat zeigt, dass sich selbst für interessierte Lehrerinnen und Lehrer die Frage stellt, wie Erkenntnisse aus Vergleichsarbeiten in diesem kurzen Zeitraum noch genutzt werden können. Die Antwort darauf erwarten sie von den Verantwortlichen der Vergleichsarbeiten, wie die Fortsetzung des oben genannten Zitats zeigt:

> „Und jetzt soll mir die Landesregierung bitte klar machen wie das jetzt umgesetzt werden kann. Denn dann fangen die großen Ferien an und im nächsten Jahr geht es wieder ganz normal mit dem neuen Schuljahr weiter. Das heißt, WER soll bitte in dieser kurzen Zeit noch mal das nachholen oder vertiefen, wenn irgendwelche Standards nicht da sind?" (ebd.)

Der Zeitpunkt der Vergleichsarbeiten kurz vor den Sommerferien birgt für die Lehrpersonen noch weitere Probleme. An vielen Schulen werden in den letzten Schulwochen, die theoretisch für die Umsetzung von Maßnahmen zur Verfügung stünden, viele außerschulische Veranstaltungen, wie z.B. Klassentage, Schullandheimaufenthalte oder auch Projekttage durchgeführt. Dadurch entsteht ein enormer Zeitdruck, der sich teilweise sehr negativ auf die Einstellung der Lehrpersonen zu den Vergleichsarbeiten auswirken kann. Diese Gymnasiallehrperson benennt deutlich, dass durch den Zeitdruck bei ihr die Aggression steigt und dies zur Ablehnung von Vergleichsarbeiten führt. Ihrer Meinung nach sollte das Schuljahr mit entspannenden und dadurch motivierenden Elementen abgeschlossen werden. Mit Vergleichsarbeiten tritt jedoch das Gegenteil ein. Stress und Druck wird erzeugt und die Durchführung von langfristig geplanten Unternehmungen, wie z.B. Schullandheimaufenthalte, gerät u. a. durch kurzfristige Terminänderungen in Gefahr. In Äußerungen wie „Dinge, die wir dann an der Front erleben" und ironischen Aussagen, wie „Da kommt Freude auf" wird die negative Haltung und Einstellung der Lehrperson besonders deutlich.

> „Das heißt mit diesen Terminen für die Vergleichsarbeiten kommt man in einem ganz normalen Schuljahr schon ins Schleudern. Normal heißt, dass diese Terminierungen genau eine Woche vor unseren angesetzten Zeugniskonventen waren. Und wir sind doppelt ins Schleudern gekommen, denn wir machen ein Schullandheim Ende Klasse sechs. Das ist wirklich ein Spießru-

> tenlauf insofern, das ist nicht dazu angetan, da jetzt irgendwie Entspannung oder Motivation kommen zu lassen. Sondern diese Dinge, die wir dann an der Front erleben, die steigern natürlich die Aggression und die Ablehnung von den Dingen. (…) „Z. B. es wurde letztes Jahr ja, was weiß ich und da rege ich mich und haben wir uns alle hier bis zur Schulleitung, wahnsinnig darüber aufgeregt, drei Termine waren in einer Woche angesetzt und dann wurde kurzfristig der Mittwochstermin umverlegt auf den Montag in der folgenden Woche. Wir konnten nur darüber spekulieren warum. Wir wollten da schon längst im Schullandheim sein. Wir wollten am Wochenende, wenn diese drei Arbeiten rum sind, wollten wir ins Schullandheim fahren. Und das hat man normalerweise ein Jahr im Voraus zu buchen. So hatte ich die wunderbare, angenehme Aufgabe, das Ganze wieder umzubuchen. Da kommt Freude auf.“ (363)

Der späte Zeitpunkt wirkt sich nicht nur auf die Lehrpersonen, sondern auch auf die Schülerinnen und Schüler, sowie deren Eltern ungünstig aus. Es steht nicht einmal genügend Zeit zur Verfügung, um mit allen Beteiligten eine ausführliche Rückmeldung in Form von Besprechungen durchzuführen. Das bedeutet, dass die Schülerinnen und Schüler nachdem sie die Vergleicharbeiten durchführten, keine detaillierten Rückmeldungen erhielten. Selbst für die Rückgabe der Ergebnisse war keine Zeit vorhanden. Teilweise mussten die Lehrpersonen ihre Kolleginnen und Kollegen, meist die Klassenlehrerinnen und –lehrern, um zusätzliche Stunden bitten. In den folgenden Äußerungen kommt dieser Sachverhalt deutlich zum Ausdruck:

> „Die Terminierung war ganz schlecht, weil die Kinder danach ins Schullandheim gegangen sind. Ich hatte keine Stunde mehr, um mit denen die Vergleichsarbeit zu besprechen. Ich musste die ausgeben, so zwischendrin mal. Ich habe mir Zeit geholt, bei der Klassenlehrerin, habe die Arbeit zurückgeben, konnte gar nicht viel dazu sagen, gar nichts. So unter den Tisch gefallen, das war ganz blöde... bei der Auswertung.“ (360)

> „Das Kriterium fehlende Zeit schließt letztendlich auch jegliche Beratung der Eltern aus. Man gibt sie ja ab. Es findet ja kein Elternabend mehr statt, oder irgendwie eine Möglichkeit, genau.“ (363)

Neben der fehlenden Zeit wurde von den Lehrerinnen und Lehrern des Weiteren beanstandet, dass Ende der Klassen sechs und acht in fast allen Schulen ein Lehrerwechsel stattfindet. Somit können die Lehrpersonen die gewonnen Erkenntnisse, auch nicht im folgenden Schuljahr umsetzen, weil sie die Klasse in den meisten Fällen abgeben. Hier stellt sich die Frage, was mit den Daten geschieht. Man hält letztlich „eine farbige Statistik“ in den Händen, mit der man jedoch nichts mehr anfangen kann, weil man die Klasse an eine andere Lehrperson abgibt.

> „Also, ich könnte mir schon vorstellen, dass man daraus einen Gewinn ziehen kann. Aber nicht zu diesem Zeitpunkt, weil man die Klasse ja abgibt.“ (156)

> „Sie werden am Ende von Klasse 6 geschrieben. Häufig gibt man ja nach der Klasse 6 auch die Klassen ab. Und ich krieg zwar eine schöne farbige Statistik. Aber dann hat es sich.“ (354)

Um dieses Problem des Lehrerwechsels zu umgehen erwägen einige, die Ergebnisse aus den Vergleichsarbeiten an die zukünftigen Lehrerinnen und Lehrer weiterzugeben. Gleichzeitig wird in den Zitaten deutlich, dass es sich hierbei eher um eine theoretische Möglichkeit handelt, die in der Praxis nicht nur auf Begeisterung stößt. Es findet sich in den Interviews auch der Vorschlag, dass die Vergleichsarbeiten von vornherein von den aufnehmenden Lehrpersonen durchgeführt werden, um das ganze Potential, das Vergleichsarbeiten bieten, nutzen zu können.

> „Wobei ich gleich sagen muss, dass ich es totalen Blödsinn finde, dass man es Ende Klasse 6 und Klasse 8 schreibt, weil ja dann der Lehrer wechselt. Also ich kann jetzt schon meine Erkenntnisse den Lehrern weitergeben, aber ich finde es ein bisschen witzlos (lachen). Na, ja.“ (264)

> „Weil man meistens die Schüler ja dann auch abgibt. Ist dann ja so. Im Prinzip müssten ja dann, müsste der aufnehmende Lehrer immer diese Vergleichsarbeit und die Ergebnisse machen, damit er dann, könnte man ja machen, damit er dann weiß, wo sind die Stärken und die Schwächen von der Klasse.“ (251)

5.3.4.2 Zeitpunkt zu früh, 8 Wochen vor Schuljahresende (Insg. 5; HS 0, RS 5, Gym 0)

Dass der Zeitpunkt der Vergleichsarbeiten zu früh liegt wird lediglich von Realschullehrpersonen beklagt. Insgesamt führen fünf Personen dieses Argument an. Der frühe Zeitpunkt hat zur Folge, dass einige Themenbereiche, die Gegenstand der Vergleichsarbeit sein könnten, noch nicht behandelt wurden. Diese Lehrperson ist auch der Überzeugung, dass alle Themenbereiche vor der Vergleichsarbeit geübt werden müssen und setzt sich dadurch zusätzlich unter Druck.

> „Wo halt das Problem ist, ist halt der Zeitpunkt. Man hat immer noch einen Themenbereich den man noch machen muss. Und dann hat man ja diese Übungsphase für die Vergleichsarbeit, die man machen muss, und von daher ist man halt immer ein bisschen am Rödeln um eigentlich diesen letzten Themenbereich noch abhaken.“ (251)

Eine Weitere findet den Zeitpunkt ebenfalls unsinnig und stellt die Konzeption der Vergleichsarbeit in Frage. Denn acht Wochen vor Schuljahresende wurde noch nicht der ganze Unterrichtsstoff behandelt. Aus diesem Grund können Vergleichsarbeiten nach Meinung dieser Lehrperson nicht bewertet werden, weil hierzu die rechtlichen Grundlagen nicht vorhanden sind.

> „Es ist ja Unsinn. Es soll der Stoff dieses Schuljahres abgefragt werden und wenn ich acht Wochen vorher schon fertig wäre, dann heißt die Frage, was mache ich dann in der Zwischenzeit. Da kann ja irgendetwas von der Konzeption her nicht stimmen. Das kann ja nicht sein, das soll ja heißen ich bin immer schon acht Wochen vorher fertig und dann frage ich mich, was soll ich in der restlichen Zeit unternehmen. Das kann ja nicht stimmen. Das heißt da wird ein Stoff abgefragt, das ist übrigens auch sehr problematisch, ich gebe eine Note die wird ja als Arbeit bewertet. Ich gebe eine Note drauf für einen Stoff den ich abfrage, den ich noch gar nicht behandelt habe. Ist sowieso die Frage, wie weit das überhaupt rechtlich akzeptabel ist oder abgesichert ist. Da stimmt die ganze Logik nicht.“ (260)

5.3.4.3 Zeitpunkt der VA verschieben (Insg. 4; HS 1, RS 3, Gym 0)

Vier Personen schlagen vor, den Zeitpunkt der Vergleichsarbeit an den Anfang des folgenden Schuljahrs zu verschieben. Dadurch könnte sie zwar nicht mehr als Klassenarbeit gewertet werden, aber der als stressig empfundene Zeitraum zwischen Ostern und Sommerferien wäre entlastet.

> „Unter Umständen wäre das eventuell zu überlegen, ob man nicht die Arbeit am Anfang vom nächsten Schuljahr schreibt, wenn gar nicht so viele Sachen da sind. Das ist aber auch schwierig, weil eigentlich sollte sie ja in die Notengebung mit einfließen. Es ist halt, das wissen die auf dem Ministerium auch, dass die Phase nach Ostern keine gute Phase ist, um solche Sachen zu machen.“ (250)

Ein weiteres Argument für die Verlegung der Vergleichsarbeit an den Beginn des nächsten Schuljahres betrifft die Möglichkeit ihrer Nutzung als Rückmeldeinstrument. Davon könnten nach Meinung einer Lehrperson auch Eltern profitieren. Da Schüler zur Vorbereitung von Vergleichsarbeiten nicht viel lernen müssen, stellt die Verschiebung des Zeitpunkts auch kein Problem dar. Im Gegenteil; dadurch hätte man noch mehr Zeit, um Defizite abzubauen.

„Wenn es vom Zeitpunkt am Anfang des Schuljahres wäre, dann wären auch mehr Rückmeldungen auch von Eltern oder dass dann mehr darüber diskutiert wird.“ (259)

„Also die Schüler müssen auf diese Vergleichsarbeit ja nicht viel lernen. Von dem her könnte man die gut am Anfang vom Schuljahr schreiben. Ich glaube schon, dass es möglich wäre. Man hätte ja können auch sagen: „Wir entflechten den ganzen Bildungsplan dahingehend, dass wir statt noch mehr reinzupacken bisschen was rausnehmen. (…) Nach 6 schreiben, am Anfang von 7 und dann noch Zeit geben bis was weiß ich vier Woche lang, je nachdem gewissen Zeitraum, um das was an Defiziten rauskommt nachzuholen. Wäre das nicht? Das wäre dann eigentlich doch effektiv.“ (268)

5.3.5 Reflexion – Hauptkategorie V: Förderdiagnostischer Nutzen der Vergleichsarbeit

Tabelle 5.3.5: Reflexion – Hauptkategorie V

HK V: Förderdiagnostischer Nutzen der Vergleichsarbeit				
Subkategorien	**Ges (n=59)**	**HS (n=15)**	**RS (n=23)**	**Gym (n=21)**
1. Durch VA keine neuen Anhaltspunkte erhalten	34	10	9	15
2. Anhaltspunkte erhalten	30	7	13	10
Ausreißer im Fokus	12	4	4	4
Anhaltspunkte für bestimmte Inhalte	11	0	9	2
Anhaltspunkte erhalten, wenn man Klasse weiterführt	4	2	0	2
Anhaltspunkte aufgrund der übersichtlichen Auswertung	3	0	0	3
3. Keine Fördermaßnahmen möglich	42	10	17	15
VA ist ungeeignetes Diagnoseinstrument	22	8	7	7
Keine Zeit für Fördermaßnahmen	20	3	7	10
Fehlendes schulisches Angebot	12	2	9	1
4. Fördermaßnahmen möglich	7	2	2	3

5.3.5.1 Durch VA keine neuen Anhaltspunkte erhalten (Insg. 34; HS 10, RS 9, Gym 15)

Insgesamt geben über die Hälfte der Befragten an, durch die Vergleichsarbeiten keine neuen Anhaltspunkte über die Stärken bzw. Schwächen ihrer Schüler erhalten zu haben. In der Hauptschule und dem Gymnasium liegt der Anteil sogar bei 2/3 der Befragten. Die meisten Lehrerinnen und Lehrer begründen ihre Meinung damit, dass sie ihre Schülerinnen und Schüler in zwei Schuljahren so gut kennenlernen, dass sie für die Einschätzung deren Leistungsfähigkeit keine Vergleichsarbeiten benötigen. Somit erhalten sie keine neuen Informationen und Erkenntnisse. Einige führen in ihrem Unterricht bereits regelmäßige Leistungsanalysen durch und können daraus Maßnahmen ableiten. Folgende Aussagen geben einen Einblick in das Antwortspektrum der Lehrerinnen und Lehrer:

> „Ich habe da jetzt keine anderen Stärken oder Schwächen der Schüler erkannt wie ich es im normalen Unterricht auch sehe. Dazu brauche ich diese Vergleichsarbeiten nicht. Ich weiß wer nicht lesen kann, ich weiß wer keinen Satz formulieren kann, ich weiß wer keine zwei Sätze fehlerfrei abschreiben kann und das reicht mir erstmal, da habe ich genug zutun die zu fördern.“ (158)

> „Nein. Ich kenne die Schwächen meiner Schüler und ich mache Aufsatzanalyse, ich mache Rechtschreibanalyse. Ich weiß, dass sie Problem haben. Brauche ich nicht.“ (254)

> „Oft ist es allerdings auch so, dass das halt Sachen sind, die vorher schon allen klar sind. Dass der Schüler an der Stelle Schwächen hat. Dass man das auch schon angesprochen, geübt hat. Aber dass es vom Schüler selber nicht umgesetzt wird oder dann auch zum Teil von den Eltern nicht mit umgesetzt wird.“ (357)

5.3.5.2 Anhaltspunkte erhalten (Insg. 30; HS 7, RS 13, Gym 10)

30 Personen haben durch die Vergleichsarbeiten Anhaltspunkte zu Stärken und Schwächen ihrer Schüler erhalten. Vor allem jene Schüler rückten ins Zentrum der Aufmerksamkeit, bei denen das Ergebnis der Vergleichsarbeit von der erwarteten Leistung enorm abwich, sogenannte Ausreißer. Außerdem erhielten die Lehrpersonen auch Hinweise zu spezifischen Themenbereichen, die im Unterricht entweder besonders intensiv oder kaum behandelt wurden.

Ausreißer im Fokus

Jeweils vier Lehrpersonen aus jeder Schulart berichten davon, dass sie sich vor allem mit denjenigen Schüler intensiver auseinander gesetzt haben, deren Ergebnisse der Vergleichsarbeit von ihrer sonstigen Leistung extrem abgewichen sind.

> „Aber wer jetzt ganz nach oben oder unten Ausreißer war, habe ich mir genauer alles angeschaut, das schon. Das interessiert mich ja schon, wenn man es mal so großartig auswertet. ( ) Also, da habe ich schon so bei einzelnen Schülern auch geschaut, wer halt normal auch auf seiner 2,0 die er sonst auch hatte liegt, mit dem habe ich mich nicht so großartig beschäftigt. Immer nur so die Ausreißer.“ (162)

> „Also, es waren ein paar Ausreißer dabei. Bei manchen hat es mich gewundert, dass sie gerade bei dieser oder jener Aufgabe so gut waren. Und bei der anderen, wo ich es eher erwartet hätte nicht. Aber es waren vielleicht ein oder zwei Schüler.“ (271)

> „Bei den Schülern, die jetzt in der Vergleichsarbeit gut abgeschnitten haben und die sonst verhältnismäßig schwach waren, da konnte ich dann schon teilweise erkennen, da haben sie vielleicht mal mitgedacht.“ (155)

Anhaltspunkte für bestimmte Inhalte

Neun Lehrpersonen an Realschulen und zwei an Gymnasien konnten aus den Rückmeldungen Hinweise zu bestimmten Unterrichtsinhalten ableiten. Die Aussagen die dieser Kategorie zugeordnet werden können, beziehen sich auf ganz spezifische Aufgaben, die in den Vergleichsarbeiten 2007 bearbeitet werden mussten. In den meisten Aussagen lässt sich ein Zusammenhang erkennen zwischen den „im Unterricht behandelte Themengebiete“ und den darin „erzielten Ergebnissen“.

> „Und bei einigen Fragen war ein riesiger Unterschied. Das lag in meinen Augen da dran, dass ich das Thema intensivst bearbeitet habe und manches Thema gar nicht bearbeitet hatte. Ja in EWG stehen ja nur Bereiche und wenn es ein spezielles Thema dann war, dann konnten sie es teilweise nicht beantworten. Da war es dann ziemlich abweichend.“ (262)

In einigen Zitaten finden sich auch Aussagen der Lehrerinnen und Lehrer wieder, welche Maßnahmen diese aus den Rückmeldungen ableiten. Im ersten Zitat berichtet die Gymnasiallehrkraft davon, ganz deutlich Anhaltspunkte über Schwächen ihrer Schülerinnen und Schüler erhalten zu haben. Als Folge daraus nutzte sie den darauf folgenden Unterricht dazu, diese geforderten Kompetenzbereiche nochmals zu behandeln.

Im zweiten Beispiel zeigt eine andere Lehrperson ihrer Schülerin auf, wo diese ihre Schwächen hat und fordert sie auf, außerhalb des Unterrichts dafür Sorge zu tragen, ihre Lücken zu schließen, indem sie z.B. den Lernstoff wiederholt oder andere Personen bittet, ihr den Sachverhalt zu erklären.

> „Man hat ja ganz klar gesehen, auf welche Fragestellung konnten sie gut und vollständig antworten, wo liegen Schwächen der Schüler. Bei uns muss ich schon auf die Thematik der Schwächen eingehen hier in der Frage. Also, für mich war auf jeden Fall für den folgenden Unterricht klar, was ich jetzt verstärkt üben muss, damit sie am Ende der sechsten Klasse das Ziel der Vergleichsarbeit, das ich jetzt als Ziel setzte, oder auch angesehen habe, erreiche. Und zwar in allen geforderten Kompetenzen. (…). Ja, ich habe natürlich erkannt, dass die größte Schwierigkeit war, bei meiner Klasse jetzt, bei der Mehrheit, vollständige Sätze bei den Antworten zu bilden. Das habe ich durchweg erkannt.“ (369)

> „… wenn ich sehe, überall wo Brüche waren, hat sie quasi nichts gewusst, weil sie die Brüche nicht weiß - dann habe ich Einzelne, vor allem die schwächeren Schüler, darauf hingewiesen: 'Schau, alles was der Bereich war, da war gar nichts da bei dir. Wiederhol das in den Ferien oder such dir jemanden, der dir das erklärt.“ (256)

Anhaltspunkte erhalten, wenn man die Klasse weiterführt

Eine weitere Subkategorie im Bereich „Anhaltspunkte erhalten“ umfasst jene Aussagen der Lehrerinnen und Lehrer, die sich für Rückmeldungen und Hinweise aus Vergleichsarbeiten einen Sinn sehen, wenn sie die jeweilige Klasse ein weiteres Schuljahr weiterführen.

> „Das Ganze, wenn das überhaupt einen Wert haben sollte, müsste ich kurz gesagt eigentlich die Klasse noch weiter haben. Es kann sein, ich habe die Klasse gar nicht mehr im nächsten Jahr, irgend ein anderer, der wieder heilfroh ist, wenn er jetzt zum Beispiel in Klasse 7 einigermaßen, mit diesen überladenen neuen Lehrplänen von G8, ja heilfroh ist, wenn er sein eigenes Zeug unterkriegt.“ (350)

Anhaltspunkte erhalten aufgrund der übersichtlichen Auswertung

Dass Schwächen bzw. Stärken erkannt werden, liegt nach Meinung von drei Gymnasiallehrpersonen auch an der übersichtlichen Darstellung der Auswertungsunterlagen.

> „Also, der erste Punkt, es ist natürlich relativ leicht Schwächen zu erkennen, einfach auf Grund der übersichtlichen Auswertungsinstrumente, die bei dem Programm dabei sind.“ (357)

> „Ich muss zu meinem eigenen Leidwesen eingestehen, ich hätte damit viel anfangen können. Es war ja wunderschön aufgegliedert mit diesem Streuungsdiagramm, konnte man sogar sehen, wo bei der Klasse jetzt Schwierigkeiten immer noch sind oder Lücken sind und wo man gut gearbeitet hat.“ (368)

5.3.5.3 Keine Fördermaßnahmen möglich (Insg. 42; HS 10, RS 17, Gym 15)

Insgesamt sind 42 Lehrerinnen und Lehrer der Meinung, dass in Folge der Vergleichsarbeiten keine Fördermaßnahmen möglich sind. Diese Aussage wird von den Lehrerpersonen unterschiedlich begründet. Die größte Gruppe bewertet Vergleichsarbeiten als ungeeignetes Diagnoseinstrument. Eine weitere Anzahl beklagt die fehlende Zeit, um Maßnahmen umsetzen zu können. Ein Fünftel bemängelt das fehlende schulische Förderangebot. Diese drei Begründungen werden in folgenden Unterkapiteln ausführlich dargestellt.

VA ist ungeeignetes Diagnoseinstrument

22 Personen halten die Vergleichsarbeit als Diagnoseinstrument für ungeeignet. Sie kritisieren an der Vergleichsarbeiten, dass es sich nicht um Förderdiagnostik handelt, wenn man lediglich ableiten kann, wo Schülerinnen und Schüler Fehler machen. Sie hätten sich differenziertere, tiefer greifende Aufgaben gewünscht, die mithilfe einer Handreichung eindeutige diagnostische Hinweise enthalten.

> „Was ich rauslesen kann ist, ich kann bei den Matheaufgaben sehr deutlich ablesen, wo meine Schüler Schwächen haben. Also, ich kann genau dem Aufgabentyp das zuordnen, kann also sagen, er oder sie kann Bruchrechnen, Addition beherrscht er nicht. Aber ist das Förderdiagnostik? (…) Es wäre vielleicht sinnvoll, wenn einige Aufgaben tiefer gehend wären, oder da wiederum eine Handreichung dazu man kriegen könnte, könnte ich mir vorstellen, dass man speziell differenziert Aufgaben stellt, die der Diagnostik dienen. Und das vielleicht auch noch benannt wird, genauer. Ja.“ (153)

Des Weiteren finden sich in dieser Subkategorie auch Aussagen der Lehrpersonen, in denen sie die Nichteignung als Diagnoseinstrument begründen. Zum einen führen die Personen an, dass man am Ende von zwei Jahren die Klasse so gut kennt und man mit Vergleichsarbeiten keinen weiteren Erkenntnisgewinn erzielt (siehe Kapitel 5.3.5.1). Zum anderen differenzieren die Aufgaben nicht, so dass man durch das Anforderungsniveau und die zugrunde liegenden Bewertungsmaßstäbe keine spezifischen, differenzierten Förderaussagen treffen lassen.

> „Aber, zur Diagnose hat die Vergleichsarbeit, wie gesagt, echt nichts mehr beigetragen, weil man macht es ja, ich hatte die ja über zwei Jahre, d.h. ich weiß wie das Textverständnis ist, aus dem langen Arbeiten im Unterricht, und Rechtschreibung und Grammatik auch. Das sehe ich ja auch wenn die Aufsätze schreiben, jedes mal wieder, wo von den Einzelnen die Probleme sind.“ (358)

> „Diese Aufgaben sind für einen starken Schüler keine Herausforderung. Die differenzieren meines Erachtens nicht. Die differenzieren nicht zwischen ausreichend, befriedigend, gut, sondern die differenzieren zwischen ausreichend und nicht ausreichend.“ (272)

In der folgenden Aussage eines Gymnasiallehrers kommt die Enttäuschung über den förderdiagnostischen Nutzen der Vergleichsarbeit deutlich zum Ausdruck. Zunächst betont dieser seine Hingabe zu seinem Beruf, besonders im Hinblick auf die individuelle Förderung und Unterstützung der einzelnen Schülerinnen und Schüler. In Erwartung, mit Vergleichsarbeiten ein Instrument zu erhalten, das zur Diagnose und somit der gezielten Förderung der Schüler eingesetzt werden kann, zeigt er sich bereit, Vergleichsarbeiten dafür zu nutzen. Die Erfahrung lehrte ihn jedoch, dass es sich seiner Meinung nach nicht um ein nützliches Diagnoseinstrument handelt, sondern lediglich um eine bilanzierende, unbrauchbare Ergebnisdarstellung, ohne jeglichen diagnostischen Hinweis.

> „Also, ich muss sagen, ich wurde ganz groß enttäuscht. Entgegen Verlautbarungen, dass man jetzt hier nützliche Diagnoseinstrumente kriegt, dass man was erfährt über seine Schüler. (…). Und ich bin eigentlich leidenschaftlich gern Lehrer. Ich beschäftige mich mit meinen Schülern. Ich gehe mit jedem innerlich auch ein bisschen mit. Deshalb dachte ich auch zunächst einmal, na ja, wenn das wirklich dann an Einsicht bringt, um zum Beispiel jemand gezielt auch helfen zu können für die Zukunft hin, dann könnte man vielleicht doch noch ein Fünkchen Verstand an dem ganzen komischen System finden. Aber, es war eigentlich so. Und ich habe das dann in allen Gesprächen mit Kollegen, auch von anderen Fächern, so gehört: Sie sehen es ja vor sich liegen. Man gibt das dann ein über Excel. Und dann ist das Programm. Das wirft sich dann selber ja aus. Man kriegt schön geordnete Diagramme, man kriegt Übersichten, man kriegt farbige Übersichten, man kriegt mit Kommastellen, also Aufgabe, Item, eigene Klasse, Vergleichsstichprobe, Differenz und so weiter. (…) Das habe ich auch meinen Kindern ausgedruckt und habe es meinen Kindern brav gezeigt. Schaut mal da sind eure Schwächen!“ (350)

Keine Zeit für Fördermaßnahmen

1/3 der Befragten klagen, dass im Anschluss an die Vergleichsarbeit keine Zeit mehr für Fördermaßnahmen bleibt. Diese Subkategorie lehnt sich mit ihren Argumentationssträngen an Kapitel 5.3.4.1 „Zeitpunkt der VA zu spät, keine Aktionen mehr möglich“. Teilweise ziehen einige Lehrpersonen die Möglichkeit in Betracht, die Erkenntnisse an die nachfolgenden Kolleginnen und Kollegen weiterzugeben, so dass diese gezielt auf die Defizite und Schwächen eingehen kann, um den Schülern die maximale Förderung zu bieten.

> „Nein! (Fördermaßnahmen) waren auch eigentlich zeitlich gar nicht mehr möglich, das war ja kurz vor Schluss.“ (157)

> „Dazu haben wir gar keine Zeit gehabt. Wann hätte ich da noch Fördermaßnahmen einleiten sollen? Man muss mal überlegen, wann die Vergleichsarbeiten geschrieben worden sind, das war ja kurz vor Schuljahresende.“ (250)

> „Fördermaßen kann man daraus leider keine groß noch ableiten, weil es einfach am Ende von Schuljahr Sechs geschrieben wird. Man müsste es an einen Kollegen übergeben, der dann die Klasse weiter führt, wenn man nicht die Chance hat das selber noch zu machen. Das hatte ich in Klasse, also bei der ersten Vergleichsarbeit. Da habe ich dann auch gezielt darauf geachtet eigentlich. Jetzt bei der zweiten Klasse war es einfach so, dass ich die abgegeben habe zur Sieben. Da ist es nicht möglich persönlich irgendwelche Fördermaßnahmen noch einzuleiten.“ (357)

Fehlendes schulisches Förderangebot

12 Lehrpersonen berichten von einem fehlenden schulischen Förderangebot. Den Grund hierfür sehen die Befragten u. a. bei Einsparungsmaßnahmen und bei fehlendem Lehrpersonal. In den Aussagen der Lehrerinnen und Lehrer kommt deutlich zum Ausdruck, dass sich Förderung in den meisten Schulen auf Fördermaßnahmen einer Lehrperson im Unterricht beschränkt. Es existiert keine schulische Förderstruktur die ausreichend Kapazitäten und Angebote bietet, um Schülerinnen und Schüler aller Altersstufen mit ihren spezifischen Leistungsproblemen unterstützend zu fördern und somit auch zur Entlastung der einzelnen Lehrerinnen und Lehrer beiträgt. Einige berichten, dass es an der Schule eine kostenlose Hausaufgabenbetreuung gibt, die wenigstens dafür sorgt, dass die Schülerinnen und Schüler regelmäßig ihre Hausaufgaben erledigen, deren Lösungen im Unterricht dann besprochen werden können. Wenn es Fördermaßnahmen gibt, dann beschränken sich diese maximal auf die Klassen 5 und 6. Diese stellen jedoch insgesamt betrachtet eine absolute Ausnahme dar.

> „Das ist schwierig. Es gibt wenig was man bei uns als Förderung machen kann. Wir haben kaum Lehrerstellen, wir haben sowieso eine Unterversorgung. (…)Wir bräuchten dringend Lehrer auch für Förderkurse. Es gibt einfach nichts, was man machen kann. Ich habe manchen, die da so schlecht waren noch gesagt, sie sollen doch bitte in die Hausaufgabenbetreuung gehen, das ist etwas, was ich empfehlen kann, was halt auch kostenlos angeboten wird. Das ist ein Tropfen auf den heißen Stein, ich weiß. Aber, dann haben die wenigstens die Hausaufgaben regelmäßig und können da wenigstens mit machen und nicht: „Ich habe gar nichts. Ich weiß von nichts". Förderkurse speziell auch in Englisch, gibt es bei uns gar nichts.“ (162)

> „Aber so eine spezielle Fördergeschichten werden bei uns momentan personell einfach nicht, einsparen.“ (251)

> „Da haben wir überhaupt keine Stunden zur Verfügung. Da haben wir gar keine Chance. Da brauchen wir überhaupt gar nicht darüber nachdenken.“ (258)

> „Fördemaßnahmen gibt es nicht mehr. Es ist bei uns an der Schule nicht möglich noch Fördermaßnahmen über die Klasse Sechs hinaus überhaupt anzubieten.“ (253)

> „Stützkurs gibt’s aber nur in Klasse fünf. In Klasse sechs und sieben nicht mehr. Also, d. h. die da jetzt schlecht raus gekommen sind, die erfahren in Klasse sieben keine Förderung deswegen.“ (259)

In der folgenden Aussage eines Gymnasiallehrers werden die Schwierigkeiten der Nutzung von Vergleichsarbeiten und die Umsetzung von Fördermaßnahmen noch einmal zusammengefasst:

> „Die Schulwirklichkeit sieht halt ganz anders aus. Das wird jetzt irgendwo archiviert, zu den Akten gelegt. Entweder hat man die Klasse gar nicht mehr oder wenn man sie hat, dann hat man gar nicht die zeitliche Möglichkeit und die Ressourcen und die Lehrerstunden, um jetzt daraus ein, sage ich mal, Ergänzungsprogramm, Förderprogramm (zu machen). Außerdem müsste noch mal ein Schritt weiter gedacht das Förderprogramm dann so differenziert sein, dass ich Montag nachmittags mit dem einen eine Stunde lang Grammatik mache, mit dem anderen, der hier defizitär war im Bereich Lesen, mit dem anderen, der hier defizitär war im Bereich Textanalyse, dass ich da ganz gezielt dann arbeiten müsste.“ (350)

Fördermaßnahmen möglich

Die Meinung, dass die Ableitung von Fördermaßnahmen aus den Rückmeldungen generell möglich ist, teilen sieben Lehrpersonen. Vor allem in Bezug auf spezifische Themengebiete werden die Inhalte im Unterricht oder mit einzelnen Schülern nochmals vertieft und eingeübt. Teilweise passen die Lehrerinnen und Lehrer ihre Aufgabenstellungen und Vorgehensweisen an die Aufgaben aus den Vergleichsarbeiten an.

> „Ja, genau. Ja, die Rückmeldung hat zum Teil schon etwas gebracht. Man übt vielleicht noch mal die eine oder andere Sache auf die Vergleichsarbeit hin. Da konnte man dann sehen, welcher Schüler hat z. B. das mit den Sprichwörtern und Redensarten verstanden und welcher nicht. Und es festigt einfach noch mal den Eindruck, wo die Schüler Schwächen haben oder wo die Klasse eine Schwäche hat und dass man da einfach noch mal daran arbeiten kann.“ (269)

> „Ein Stück weit sicher. Wenn es jetzt gerade in Richtung Bruchrechnen, was eh drankommt, und man sieht einfach, die Schüler haben jetzt hier extreme Schwierigkeiten beim Bruchrechnen, dann denke ich schon, dass man einfach vielleicht im Durchgang sich überlegt: Okay, da habe ich jetzt ein Stück weit vielleicht etwas falsch gemacht. In der Vergleichsarbeit wird eher jetzt so nach diesen Brüchen gefragt. Versuche ich es vielleicht, mein zukünftiges Bruchrechnen eher so ein bisschen an die Vergleichsarbeit zu orientieren.“ (152)

> „Ich habe konkret in der Klasse drei Schüler gehabt, die weit unter dem Schnitt lagen. Dann habe ich von mir aus da jeweils die Eltern kommen lassen und gesagt: "Gut, ich habe die Klasse noch mal ein Jahr. Wir werden das entsprechend beobachten." Ja, das einzige, was ich machen konnte. Andere Fördermaßnahmen habe ich an der Stelle überhaupt nicht gesehen.“ (354)

5.4 Ergebnisse zur Aktion

5.4.1 Auswirkungen der Vergleichsarbeiten auf die Unterrichtspraxis

Nach den beiden Schritten der Rezeption und Reflexion von Leistungsrückmeldungen erfolgt im idealtypischen Modell von Helmke & Hosenfeld der Schritt „Aktion“, also die Ableitung von Maßnahmen zur Verbesserung auf die Unterrichtspraxis.

In diesem Bereich konnten 2 Subkategorien gebildet werden, die wiederum nach der Häufigkeit ihrer Nennungen aufgelistet sind. Die erste Subkategorie „VA liefern keinen Anlass für Änderungen der Unterrichtspraxis“ wird mit wenigen Ausnahmen von allen Lehrerinnen und Lehrern genannt. Die zweite Kategorie bezieht sich auf mögliche „Änderungen der Unterrichtspraxis“, die die Lehrpersonen im Zusammenhang mit Vergleichsarbeiten ableiten.

Tabelle 5.4.1: Aktion – Hauptkategorie I

HK I: Auswirkungen auf die Unterrichtspraxis				
Subkategorien	**Ges (n=59)**	**HS (n=15)**	**RS (n=23)**	**Gym (n=21)**
1. VA liefern keinen Anlass für Änderung der Unterrichtspraxis	50	14	19	17
2. Änderungen der Unterrichtspraxis	36	7	16	13
Spezielle Themen der VA als Schwerpunkt wählen	19	4	11	4
Mehr Wiederholungen	13	4	7	2
Komplexere Aufgaben ab Klasse 5	12	0	6	6
Neue Unterrichts- bzw. Lehr-Lernmethoden	8	0	4	4

5.4.1.1 VA liefern keinen Anlass für die Änderung der Unterrichtspraxis (Insg. 50; HS 14, RS 19, Gym 17)

Beinahe alle Lehrerinnen und Lehrer geben an, dass sie aufgrund der Vergleichsarbeiten keine Änderungen ihrer Unterrichtspraxis vornehmen werden. Als ein Hauptgrund wird angeführt, dass sie sich bei der Planung und Durchführung ihres Unterrichts auf die neuen Bildungspläne beziehen. Und diese bilden schließlich die Basis für Vergleichsarbeiten.

> „Ja, so von den Aufgaben mache ich eigentlich sehr vieles auch schon so ähnlich. Es gibt jetzt eigentlich nichts Neues, was ich sagen kann, das hätte mich jetzt vom Hocker gerissen.“ (152)

> „Nein, weil die Inhalte der Vergleichsarbeiten sind ja schon etwas abgestimmt auf das was in den neuen Bildungsplänen gefordert wird und das was auch als unser Ziel hier Hauptschulabschlussprüfung, was da gemacht wird.“ (154)

> „Die Unterrichtsmethoden werde ich wohl kaum ändern, weil, es hat immer gut geklappt bisher.“ (156)

Wenn Änderungen vorgenommen werden, dann liefern nach Angabe dieser Lehrperson nicht die Vergleichsarbeiten den Anlass dazu. Sie ändert dann Lernstrategien ab oder nimmt neue auf, wenn sie darüber gelesen hat, von Kolleginnen und Kollegen inspiriert wird oder eigene Ideen zur Änderung entwickelt.

> „Das hat mit den Vergleichsarbeiten sicher nichts zu tun. Wenn ich eine Lernstrategie verändere, dann habe ich sie vielleicht in der Literatur gelesen oder bei einem Kollegen gesehen und dann hat mir die vielleicht gefallen. Oder ich habe selber eine Idee, dann mache ich die. Aber das würde ich jetzt nicht ändern, weil mir die Vergleichsarbeit Erkenntnisse gegeben hätte.“ (250)

5.4.1.2 Änderung der Unterrichtspraxis (Insg. 36; HS 7, RS 16, Gym 13)

Auf die Frage, welche zukünftigen Maßnahmen die Lehrpersonen planen, um eine Verbesserung des Unterrichts zu erzielen, antworten insgesamt 36 Lehrpersonen. Ihre Aussagen beziehen sich einerseits auf die Einübung spezieller Themenbereiche, die in den Vergleichsarbeiten abgefragt wurden. Andere planen für die Zukunft, dass sie mehr Wiederholungen einbauen oder den Schülerinnen und Schü-

lern bereits ab Klasse 5 komplexere Aufgaben stellen werden. Eine geringe Anzahl von Lehrpersonen gibt an, neue Unterrichtsmethoden anzuwenden.

Spezielle Themen der Vergleichsarbeit als Schwerpunkt wählen

Der größte Anteil der Lehrerinnen und Lehrer (n=20) greift spezifische Themen aus der Vergleichsarbeit heraus und plant, diese im zukünftigen Unterricht zu erarbeiten und zu vertiefen. Ganz sicher sind sich die Lehrpersonen jedoch nicht, ob sich dadurch bei der nächsten Vergleichsarbeit die Leistungen verbessern werden. Einige räumen ein, dass man damit rechnen müsse, dass in der folgenden Vergleichsarbeit neue Themenbereiche behandelt, bzw. die Aufgaben anders gestellt werden. In diesem Fall würde dann eine vertiefte Behandlung spezieller Themen der Vergleichsarbeiten nicht zu verbesserten Leistungen führen.

Weiterhin klingt in einigen Interviews durch, dass die Schwächen der Schüler in einzelnen Bereichen nicht erst durch die Vergleichsarbeit sichtbar wurden, sondern schon vorher bekannt waren.

> „Wenn es jetzt gerade in Richtung Bruchrechnen, was eh drankommt, und man sieht einfach, die Schüler haben jetzt hier extreme Schwierigkeiten beim Bruchrechnen, dann denke ich schon, dass man einfach vielleicht im Durchgang sich überlegt: ‚Okay, da habe ich jetzt ein Stück weit vielleicht etwas falsch gemacht. In der Vergleichsarbeit wird eher jetzt so nach diesen Brüchen gefragt. Versuche ich es vielleicht, mein zukünftiges Bruchrechnen eher so ein bisschen an die Vergleichsarbeit zu orientieren.' Wobei man jetzt auch nicht weiß, ob das natürlich das Richtige ist. Nicht alles, was hier drinnen steht in einer Vergleichsarbeit sollte gleichzeitig heißen, dass es richtig ist. Es kann sein, dass in der nächsten Vergleichsarbeit, sag ich jetzt einmal, vom Schema her anders das Ganze aufgearbeitet wird. Und dann kann das schon wieder falsch laufen." (152)

> „Was ich für mich jetzt mitgenommen habe ist, dass ich jetzt eher weiß was in den Vergleichsarbeiten dran kommt, das habe ich vorher noch nicht so gewusst. Beispielsweise, wenn ich Deutsch nehme, dass eben der Text, oder Textanalyse oder Textverarbeitung, dass man da vielleicht noch mehr Zeit widmen müsste, weil die Kinder da ganz schlecht sind. Im Lesen, im Auswerten von Texten, das habe ich aus dieser Vergleichsarbeit entnommen, aber das hätte ich natürlich auch vorher schon gewusst." (250)

Mehr Wiederholungen

13 Personen werden mehr Wiederholungen in ihrem Unterricht einplanen, damit Themen, die schon länger zurückliegen, nicht in Vergessenheit geraten. Ihnen wurde deutlich, dass Aufgaben, die sie als gut lösbar einschätzten, den Schülerinnen und Schülern schwer fielen. Den Grund sehen sie darin, dass die Themen vor längerer Zeit im Unterricht behandelt wurden und ihre Schüler diese nicht mehr abrufen konnten. Von zahlreichen und regelmäßigen Wiederholungen versprechen sie sich eine Verbesserung der Ergebnisse.

> „Und zwar habe ich in Mathe gemerkt, dass Dinge, die lange zurückliegen, also ganz konkret mal am Anfang des Schuljahres gemacht hatte, dann eine Zeit lang nicht mehr trainiert hatte, in Vergessenheit geraten waren. Und es waren dann eben auch Sachen, von denen man dachte: ‚Kein Problem, zu einfach, machen meine Schüler mit links.' Und genau das dann eben nicht. Also, es hat gezeigt, dass einfach alles, auch Dinge, die ich für einfach halte, ich immer wieder wiederholen muss. Das hat es mir dann schon gezeigt. (…) Vielleicht die Anregung, dass, wenn man wiederum Aufgaben wiederholt, Themenbereiche wiederholt, dass man auch mit relativ wenigen Beispielen schon überprüfen kann, ob die Schüler was verstanden haben, ein Themenfeld verstanden haben oder nicht." (153)

> „Ja mehr wiederholen. Mehr, klar. Man soll es eigentlich nicht, aber jeder wiederholt, weil ja der Stoff auch aus Klasse 5 ist und die Schüler, die wissen zum Beispiel römische Zahlen, wissen die in Klasse 6 Ende nicht mehr so und dann glaubt man zwar weniger, dass es dran kommt, aber es könnte dran kommen und der Wunsch besteht, dass man es noch mal wiederholt von den Schülern. Dann mache ich das." (261)

> „Ja, Wiederholungen mehr, dass man einfach Stoff Klasse 5 wiederholt, weil es ja dann am Ende von Klasse 6 ist und dann viel vergessen wird. Also wir wiederholen mehr." (264)

Komplexere Aufgaben ab Klasse 5

Jeweils sechs Realschul- und Gymnasiallehrer würden den Schülern bereits ab dem fünften Schuljahr komplexere Aufgaben stellen. Sie sind sich allerdings auch der Tatsache bewusst, dass für diese Art von Aufgaben viel mehr Unterrichtszeit benötigt würde und diese im Schulalltag nicht zur Verfügung steht.

> „Ich würde stärker, von Fünf angefangen, diese Art von Aufgaben, also diese Komplexität in Sinne von einem Text, sowohl Grammatik, als auch sprachliche, als auch stilistische und ähnliche Mittel versuchen zu erarbeiten. Das habe ich eigentlich erst in Sechs so intensiv gemacht. Das war so meine Idee, das für Fünf anzulehnen." (253)

Eine Realschullehrperson berichtet darüber, dass sie sich im Team Gedanken darüber machten, die zukünftigen Arbeiten an das Format von Vergleichsarbeiten anzupassen und somit verschiedene Kompetenzen in einem Test zu behandeln. Zum einen wird dadurch die Komplexität des Verfahrens erhöht, zum anderen bietet es den Kindern die Möglichkeit ihre Leistungen zu verbessern, da unterschiedliche Kompetenzen aus verschiedenen Themengebieten verlangt werden und sie ihre Schwächen in manchen Bereichen mit vorhandenen Stärken ausgleichen können.

Eine andere Lehrperson aus dem Gymnasium nimmt sich vor, in Zukunft kognitiv anspruchsvollere Aufgaben ihren Schülerinnen und Schülern anzubieten, damit diese ihre Kompetenzen in komplexeren Dimensionen erweitern und ausbauen können. Allerdings reflektiert und nennt sie mögliche Schwierigkeiten komplexe Aufgaben im Unterricht zu behandeln. Für deren Lösung benötige man weit mehr Unterrichtszeit, die jedoch aufgrund der Stofffülle nicht zur Verfügung steht.

> „Wir haben da zusammen darüber gesprochen und haben uns überlegt, dass wir jetzt eine Kombinationsarbeit, z.B. die wir jetzt dieses Jahr machen wollen so in diesem Bereich machen. So ähnlich aufbauen wie die Vergleichsarbeit, dass eben verschiedene Punkte angesprochen werden und damit auch verschiedene Kompetenzen. Für die Schüler natürlich eine Chance einfach besser zu sein, weil die ja dann ihre Kompetenzen in verschiedenen Gebieten haben. Und wenn sie in einem Gebiet nicht so gut sind, können sie das durch das andere wieder ausgleichen.“ (257)

> „Ja gut, ich denk, dass man grundsätzlich auch bei der Stoffvermittlung grad was die Bereiche, die man vielleicht früher nicht so betrachtet hat, also bei den Kompetenzen, dass man die mehr in den Fokus nimmt. Gerade Problemlösen, so vernetztes Denken, Modellieren sind solche Punkte, die das Schulbuch auch wiederum wenig vorgibt. Dass man die halt auch üben sollte und thematisieren sollte. Aber Problemlösen, das ist schwierig, weil das Problemlösen braucht Zeit und Zeit hat man keine. Insofern zieht man sich dann wieder zurück auf den Standard. Wenn man viel Zeit hätte, könnte man auch mit Schülern Probleme lösen und auch komplexere Aufgaben lösen und andere und verschiedene Herangehensweisen mal ausprobieren. Aber auf Grund der Zeit und der Stofffülle ist das nicht möglich oder schwierig möglich.“ (353)

Neue Unterrichts- bzw. Lehr-Lernmethoden

Die Frage nach neuen Unterrichts- bzw. Lehr- und Lernmethoden beantworten die Lehrpersonen unterschiedlich. Neben der Einübung spezifischer Methoden und der Verwendung neuer Medien werden auch neue Lernstrategien für die Schüler und die Umgestaltung des Unterrichts aufgeführt:

Verstärktes Methodentraining

„Ja, was ich schon mehrfach gesagt habe. Die Arbeit mit den Texten. Da hatten wir auch Stunden, wo man Methodentraining machen kann - hier habe ich verstärkt noch mal so etwas gemacht. Wie gehe ich an Texte heran, wie werte ich die aus, 5-Schritte-Lesemethode. Also das habe ich in vielen Bereichen verstärkt gemacht.“ (255)

Neue Medien verwenden, zum Beispiel Karteikarten

„Ja das habe ich natürlich auch gemacht wie Kärtlerechnen und solche Sachen. Und auf farbiges Papier große Zahlen, wenn wir Zahlen bilden, also groß geschrieben die 2 und so mit Magneten an der Tafel. Die Spielereien habe ich früher nicht gemacht. Für mich gibt es Spielereien, die auch nett sind und die Kinder begeistern, aber man kann das auch mit Folienschnipseln auf dem Tageslichtprojektor auch machen. Wie gesagt, da habe ich ein paar Dinge geändert, aber nichts Wesentliches.“ (263)

„Was ich jetzt in diesem Schuljahr ausprobiere ist, mit einer fünften Klasse, das ist ein Karteikartensystem in Deutsch. Deswegen Deutsch ist auch genauso ein Fach wie Englisch, Latein, wo man Fachbegriffe lernen muss und die auch parat haben muss. Man muss wissen, was ist ein Substantiv, was ist ein Subjekt und versuche dadurch halt ein bisschen schon die Grundlage zu legen für eine stabile Grundgrammatik sozusagen.“ (357)

Neue Lernstrategien für die Schüler

„Lernstrategien der Schüler. Ja, ich denke da ändert sich schon was, weil sie wissen auch sie müssen schon auch zusätzlich zu Hause hinsetzen, wiederholen und üben, auf diese Arbeit, weil die eben den Stoff von zwei ganzen Schuljahren umfassen kann. Deshalb ändert das auch etwas an den Lernstrategien der Schüler. Das finde ich aber sehr positiv. Das hat mir auch schon immer an den Zentralen Klassenarbeiten gefallen.“ (361)

„Aber was wir uns ausgemacht haben in der letzten Klassenkonferenz war, dass wir einfach auch vermehrt Wert darauf legen, dass die Schüler selbständig werden und das bedeutet natürlich, dass sie am Ende von einem Unterrichtsstoffgebiet z. B. sich selber Spiele überlegen können und machen, egal ob das ein Würfelspiel ist..“ (267)

Umgestaltung des Unterrichts in Richtung Projektunterricht

„Also, die DVA ist jetzt sicherlich ein Anlass dazu. Grundsätzlich Umgestaltung des Unterrichts ist sowieso mein Ziel. Ich möchte ganz gerne in Richtung Projektunterricht rein gehen, um es irgendwie ansprechender zu gestalten. Das ist eine Frage der Zeit, habe mal die Hoffnung, dass ich es in diesem Schuljahr in einer Klasse hin kriege. Dass ich so die Päckchen packe, dass ich dann irgendwann so einen Block habe, dass ich dann ein schönes Projekt machen kann. Da macht auch mir Unterricht mehr Spaß.“ (366)

5.5 Ergebnisse zur Kommunikation im schulischen Kontext

Die Auswertung des quantitativen Teilfragebogens ergab eine differenzierte Darstellung welche Lehrerinnen und Lehrer mit den unterschiedlichen Akteuren innerhalb des schulischen Kontextes „keine Gespräche“ oder „einen informellen Austausch“ führen und mit wem die Ergebnisse aus Vergleichsarbeiten „systematisch“ und dadurch vertieft diskutiert werden.

In der ersten Tabelle 5.5.1 erfolgt zunächst die Auswertungsdarstellung der Kategorie „Keine Gespräche“. Demnach kann festgestellt werden, dass in keiner Schule die Ergebnisse aus Vergleichsarbeiten in der Schulkonferenz besprochen werden. Außerdem findet in den meisten Schulen weder in der Gesamtlehrerkonferenz noch in den Jahrgangsstufenkonferenzen ein Austausch darüber statt. Über 2/3 der Lehrerinnen und Lehrer führen auch keine Gespräche in der Klassenkonferenz und mit den Eltern. Etwa die Hälfte tauscht sich auch nicht in der Fachkonferenz oder mit der Schulleitung aus. Selbst mit ausgewählten Kolleginnen und Kollegen oder jenen, die an den Vergleichsarbeiten beteiligt waren, spricht etwa 1/3 der Lehrpersonen nicht.

Tabelle 5.5.1: Schulischer Kontext – Keine Gespräche

Keine Gespräche über die Ergebnisse aus Vergleichsarbeiten wurden geführt …	**Ges (n=59)**
in der Schulkonferenz	59
in der Gesamtlehrerkonferenz	51
in der Jahrgangsstufenkonferenz	48
in der Klassenkonferenz	44
mit den Eltern der Klasse	44
in der entsprechenden Fachkonferenz	31
mit der Schulleitung	28
mit allen an den VA beteiligten KollegInnen	21
mit ausgewählten Kolleginnen und Kollegen	19
mit den Schülerinnen und Schülern dieser Klasse	8
mit Parallelkolleginnen und –kollegen	6

Acht Lehrerinnen und Lehrer teilen selbst ihren Schülerinnen und Schülern nach der Durchführung und Auswertung der Vergleichsarbeit keine Informationen über die erzielten Ergebnisse mit. Mit den Parallelkolleginnen und –kollegen führen sechs Interviewpartner keine Unterhaltung darüber.

Informelle Gespräche sind nicht geplant, sondern finden eher unsystematisch und ‚nebenbei' statt. Im Vergleich dazu geht es in systematischen Gesprächen um einen intensiven, diskursiven Austausch über Vergleichsarbeiten. Die Auswertung der Matrix zu „informelle Gespräche werden geführt mit…" ergab folgende Reihenfolge: etwa die Hälfte der Lehrerinnen und Lehrer praktizieren mit ihren Schülerinnen und Schülern einen Austausch auf informeller Ebene. Mit den Kolleginnen und Kollegen der Parallelklassen und jenen, die an den VA beteiligt waren liegt der Anteil der Gespräche etwas geringer, bei ca. 22-25 Personen. Ein Drittel tauscht sich zum Thema Vergleichsarbeiten eher oberflächlich mit der Schulleitung und in der entsprechenden Fachkonferenz aus. Und bei einem Fünftel werden die Arbeiten in der Klassen- und Jahrgangsstufenkonferenz und mit den Eltern angesprochen.

Tabelle 5.5.2: Schulischer Kontext – Informelle Gespräche

Informelle Gespräche über Vergleichsarbeiten wurden geführt …	**Ges (n=59)**
mit den Schülerinnen und Schülern dieser Klasse	29
mit Parallelkolleginnen und –kollegen	25
mit ausgewählten Kolleginnen und Kollegen	24
mit allen an den VA beteiligten KollegInnen	22
mit der Schulleitung	19
in der entsprechenden Fachkonferenz	16
in der Klassenkonferenz	11
mit den Eltern der Klasse	11
in der Jahrgangsstufenkonferenz	10
in der Gesamtlehrerkonferenz	6
in der Schulkonferenz	0

Für die weitere Auswertung des schulischen Kommunikationsverhaltens war vor allen Dingen die „systematische Diskussion“ von Interesse. Tabelle 5.5.3 enthält die Angaben zu den Häufigkeiten, mit welchen Personen bzw. in welchen Gremien systematische Gespräche zu den Vergleichsarbeiten geführt wurden.

Tabelle 5.5.3: Schulischer Kontext – Systematische Diskussionen

Systematische Diskussionen über Vergleichsarbeiten wurden geführt …	**Ges (n=59)**
mit Parallelkolleginnen und –kollegen	28
mit den Schülerinnen und Schülern dieser Klasse	22
mit ausgewählten Kolleginnen und Kollegen	16
mit allen an den VA beteiligten KollegInnen	16
in der entsprechenden Fachkonferenz	12
mit der Schulleitung	12
in der Klassenkonferenz	4
mit den Eltern der Klasse	4
in der Gesamtlehrerkonferenz	2
in der Jahrgangsstufenkonferenz	1
in der Schulkonferenz	0

Am häufigsten diskutierten die Lehrerinnen und Lehrer mit ihren Parallelkolleginnen und -kollegen (n=28) und mit den betroffenen Schülerinnen und Schülern (n=22). Jeweils 16 Personen gaben an, mit ausgewählten Kolleginnen und Kollegen und mit jenen, die an den Vergleichsarbeiten beteiligt waren, ausführlich gesprochen zu haben. In der entsprechenden Fachkonferenz und mit der Schulleitung tauschten sich jeweils 12 Lehrerinnen und Lehrer über die Ergebnisse und die gesammelten Erfahrungen im Umgang mit Vergleichsarbeiten aus. Die unteren fünf aufgelisteten Personen, Gruppen und Gremien werden aufgrund ihrer geringen Anzahl an Nennungen im folgenden Auswertungsprozess nicht mehr berücksichtigt.

Geht man der Frage nach, ob sich in Bezug auf die systematische Diskussionen schulartspezifische Unterschiede nachweisen lassen, kann dies mit Hilfe einer Kreuztabelle analysiert werden.

Tabelle 5.5.4: Schulischer Kontext – Systematische Diskussionen aufgelistet nach Schulart

	HS (n=15)	**RS (n=23)**	**Gym (n=21)**	**Gesamt**
Mit ParallelkollegInnen	9	10	9	**28**
Mit den SchülerInnen dieser Klasse	3	11	8	**22**
Mit ausgewählten KollegInnen	3	7	6	**16**
mit allen an den VA beteiligten Kollegen	4	7	5	**16**
in der entsprechenden Fachkonferenz	0	9	3	**12**
mit der Schulleitung	4	8	0	**12**
Summe, schulspezifische Diskussionen	**23**	**52**	**31**	

Demnach lässt sich insgesamt festhalten, dass sich Lehrpersonen an Realschulen im Vergleich zu den Kolleginnen und Kollegen aus den beiden anderen Schularten intensiver mit den Ergebnissen der Vergleichsarbeiten auseinandersetzen (vgl. untere Zeile). Sowohl mit den Parallelkolleginnen und –kollegen, mit den Schülerinnen und Schülern als auch in der entsprechenden Fachkonferenz tauscht sich etwa die Hälfte der Realschullehrenden systematisch über die erzielten Ergebnisse aus. Etwa ein Drittel dieses Personenkreises führt auch systematische Diskussionen mit ausgewählten Kolleginnen und Kollegen, die sich vornehmlich an der Durchführung der Vergleichsarbeiten beteiligen. Im Vergleich zu den anderen beiden Teilstichproben finden in der Realschule auch systematische Auseinandersetzungen über das Abschneiden der Schülerinnen und Schüler bei Vergleichsarbeiten mit der Schulleitung statt.

Die Vergleichsstichprobe der Gymnasiallehrpersonen schneidet in den oberen vier Bereichen etwa gleich ab wie die der Realschule. Auffallend in dieser Gruppe ist jedoch, dass die Ergebnisse aus Vergleichsarbeiten in der entsprechenden Fachkonferenz nur von drei Personen besprochen werden. Ein extremes Ergebnis liefert die Befragung dieses Personenkreises im Hinblick auf systematische Gespräche mit der Schulleitung: keine Lehrperson tauscht sich mit der Schulleitung aus.

In der Hauptschulstichprobe fällt auf, dass die Ergebnisse aus Vergleichsarbeiten vor allem mit Parallelkolleginnen und –kollegen diskutiert werden, denn von 4/5 der Gruppe wird dies bestätigt. Ein Viertel der Stichprobe behandelt das Thema ausführlich mit den Kolleginnen und Kollegen, die ebenfalls an der Durchführung beteiligt waren und mit der Schulleitung. Nur jede/r Fünfte führt mit den Schülerinnen und Schülern oder mit ausgewählten KollegInnen ausführliche Gespräche über die erzielten Resultate.

5.6 Gesamtübersicht aller Kategorien

Tabelle 5.6.1: Gesamtübersicht aller Ergebnistabellen

Tabelle 5.1.1: Individuelle Bedingungen				
Hauptkategorie I: Funktion der VA				
Subkategorien	**Ges (n=59)**	**HS (n=15)**	**RS (n=23)**	**Gym (n=21)**
1. Vergleichsarbeiten (VA) dienen landesweitem Vergleich	55	14	22	19
um Standards zu überprüfen und zu sichern	47	8	20	19
als Rückmeldung für Verbesserungen	21	4	10	7
2. VA als Kontrollinstrument und Druckmittel	29	7	12	10
3. VA als Folge der PISA-Studie	14	3	4	7
4. VA haben keine Funktion	5	1	2	2
5. VA als purer Aktionismus	3	0	0	3

Rezeption der Vergleichsarbeiten Tabellen 5.2.1 und 5.2.2				
Tab. 5.2.1: Hauptkategorie I: Durchführung der Vergleichsarbeit				
1. keine Probleme bei der Durchführung der Vergleichsarbeit	54	14	20	20
2. Zur Verfügung stehende Zeit war für Schüler				
…ausreichend	39	8	13	18
…zu kurz	10	1	7	2
…zu lang	5	4	0	1
3. Uneinheitliches Vorgehen bei der Aushändigung der VA	29	7	12	10
Tab. 5.2.2: Hauptkategorie II: Verständlichkeit der Rückmeldungen				
1. Rückmeldungen waren verständlich	16	2	6	8
2. Rückmeldungen wurden nicht angeschaut	13	3	6	4
3. Rückmeldungen waren unverständlich / unpraktisch	5	2	3	0

Reflexion der Vergleichsarbeiten Tabellen 5.3.1 – 5.3.5				
Tab. 5.3.1: Hauptkategorie I: Auswertung der Vergleichsarbeit				
Subkategorien	**Ges (n=59)**	**HS (n=15)**	**RS (n=23)**	**Gym (n=21)**
1. Kritik an Bewertungspraxis				
Punktevergabe unpräzise und ungerecht	30	4	15	11
Notenschlüssel ungerecht, fehlende lineare Basis	10	0	9	1
generelle Kritik an Bewertung der VA	2	0	1	1
2. VA sind einfach und schnell zu korrigieren	9	3	2	4
Tab. 5.3.2: Hauptkategorie II: Aufgaben der Vergleichsarbeit				
1. Kritik an einzelnen Aufgaben- und Fragestellungen	26	8	9	9
Kritik am multiple choice Format	6	4	0	2
Sprache schülerfremd	6	0	3	3
Sonstige Kritik	11	3	4	4
2. Kritik an Zusammenstellung der VA und Aufgabenauswahl	20	6	7	7
Tab. 5.3.3: Hauptkategorie III: Niveau der Vergleichsarbeit				
1. Niveau der VA zu niedrig	29	13	13	3
2. Niveau der VA zu hoch	4	0	1	3
3. Niveau der VA angemessen	2	0	0	2
Tab. 5.3.4: Hauptkategorie IV: Zeitpunkt der Vergleichsarbeit				
1. Zeitpunkt zu spät, keine Aktionen mehr möglich	17	2	4	11
Zeitdruck, viele andere Termine	9	0	3	6
2. Zeitpunkt zu früh	5	0	5	0
3. Zeitpunkt verschieben	4	1	3	0

Reflexion (Fortsetzung)				
Tab. 5.3.5: Hauptkategorie V: Förderdiagnostischer Nutzen der Vergleichsarbeit				
Subkategorien	**Ges (n=59)**	**HS (n=15)**	**RS (n=23)**	**Gym (n=21)**
1. Durch VA keine neuen Anhaltspunkte erhalten	34	10	9	15
2. Anhaltspunkte erhalten	30	7	13	10
Ausreißer im Fokus	12	4	4	4
Anhaltspunkte für bestimmte Inhalte	11	0	9	2
Anhaltspunkte erhalten, wenn man Klasse weiterführt	4	2	0	2
Anhaltspunkte aufgrund der übersichtlichen Auswertung	3	0	0	3
3. Keine Fördermaßnahmen möglich	42	10	17	15
VA ist ungeeignetes Diagnoseinstrument	22	8	7	7
Keine Zeit für Fördermaßnahmen	20	3	7	10
Fehlendes schulisches Angebot	12	2	9	1
4. Fördermaßnahmen möglich	7	2	2	3

Tabelle 5.4.1: Aktion				
Hauptkategorie I: Auswirkungen auf die Unterrichtspraxis				
Subkategorien	**Ges (n=59)**	**HS (n=15)**	**RS (n=23)**	**Gym (n=21)**
1. VA liefern keinen Anlass für Änderung der Unterrichtspraxis	50	14	19	17
2. Änderungen der Unterrichtspraxis	36	7	16	13
Spezielle Themen der VA als Schwerpunkt wählen	19	4	11	4
Mehr Wiederholungen	13	4	7	2
Komplexere Aufgaben ab Klasse 5	12	0	6	6
Neue Unterrichts- bzw. Lehr-Lernmethoden	8	0	4	4

5.7 Typenbildung und typologische Analyse – zur Beschreibung spezifischer Rezeptions- und Nutzungstypen von Vergleichsarbeiten

Die Auswertung des gesamten Materials mittels induktiver und deduktiver Verfahren ergab zahlreiche Kategorien und Subkategorien, die einen Gesamtüberblick liefern, wie Lehrpersonen an Haupt-, Realschulen und Gymnasien die Ergebnisse aus Vergleichsarbeiten rezipieren und für sich und ihre schulische Arbeit nutzen. Der Einzelfall, im Sinne einer ausführlichen hermeneutischen Interpretation jedes einzelnen Interviews, fand bisher keine gesonderte Berücksichtigung, sondern es sollte fallübergreifend eine Gesamtschau über alle befragten Personen erstellt werden (vgl. Kelle und Kluge 2010, S. 76). Im folgenden Auswertungsverfahren werden nun spezifische Kategorien betrachtet, die im Sinne einer Typenbildung geeignet erscheinen, um in der Gruppe der interviewten Lehrerinnen und Lehrer der Sekundarstufe I bestimmte Rezeptions- und Nutzungstypen unterscheiden zu können.

Im Rahmen eines Forschungsseminars zur Anwendung qualitativer Forschungsmethoden, das von mir an der Pädagogischen Hochschule angeboten wurde, beteiligten sich auch Studierende höheren Semesters an der Analyse und Auswertung des Datenmaterials mit dem Ziel, geeignete Rezeptions- und Nutzungstypen zu generieren. Die Ergebnisse und Zuordnungen wurden innerhalb der Gruppe jeweils intensiv diskutiert, ausgetauscht und empirisch begründet. Auf die Einhaltung der Gütekriterien wurde geachtet (siehe Kapitel 4.6).

Die Typenbildung nach Kluge erfolgt, wie im Methodenteil (Kapitel 4.5) erläutert, in vier Stufen. In der vorliegenden Analyse werden zunächst die ersten beiden Schritte der Typenbildung zusammenfassend dargestellt. Da für die visualisierende Darstellung dieser beiden Schritte die Gesamtstichprobe mit 59 Fällen zu groß erscheint, wird auf den nächsten drei Seiten exemplarisch, an der Teilstichprobe der Realschullehrpersonen, die Verfahrensweise erläutert. Für die Typenbildung insgesamt wurde das Gesamtmaterial aus allen Interviews herangezogen was ab Schritt drei dargestellt ist.

Schritt 1 und 2: Erarbeitung relevanter Vergleichsdimensionen, Gruppierung der Fälle und Analyse der empirischen Regelmäßigkeiten

Betrachtet man die Gesamtübersicht aller Kategorien, die in Kapitel 5.6 dargestellt sind, so fällt auf, dass sich einige Kategorien zur Kontrastierung im Sinne einer Typenbildung nicht eignen, weil sie nicht genügend differenzieren. Das hängt damit zusammen, dass sich die Aussagen von Lehrerinnen und Lehrern zu bestimmten Themenbereichen kaum unterscheiden, weil darüber bei nahezu allen Befragten eine einhellige Meinung herrscht. Ein Beispiel stellt die Subkategorie „Vergleichsarbeiten (VA) dienen dem landesweiten Vergleich“ dar. 22 von 23 Realschullehrpersonen geben diese Antwort auf die Frage nach möglichen Gründen für die Einführung der Vergleichsarbeiten (Tabelle 5.1.1 im Kapitel 5.6). Desgleichen verhält es sich mit den Kategorien und Subkategorien „problemlose Durchführung“, „Niveau der Vergleichsarbeit“ und „Kritik an Bewertung“.

Allerdings liegen in der Gesamtauswertung genügend Kategorien und Subkategorien vor, die sich - kontrastierend gegenübergestellt - für eine Typenbildung eignen. Anhand eines Kategorienbeispiels zum Fragebereich „Funktion von Vergleichsarbeiten“ soll die Vorgehensweise der Kontrastierung exemplarisch veranschaulicht werden. So beurteilt ein Teil der Realschullehrerinnen und –lehrer die Funktion von Vergleichsarbeiten positiv. Sie betrachten die Vergleichsarbeiten „als Rückmeldung für Verbesserungen“, im Sinne eines Diagnoseinstruments. Ein anderer Teil hingegen bewertet Vergleichsarbeiten als „Kontrollinstrument und Druckmittel“ sehr negativ.

Mit Hilfe eines speziellen Features der „MAXQDA® -Software“, dem Code-Matrix-Browser, kann das Antwortverhalten der Interviewten im Gesamtüberblick visualisiert werden. Somit können für die Auswertung sowohl Unterschiede als auch Gemeinsamkeiten zwischen den verschiedenen Interviews auf einen Blick erkannt werden. Die Abbildung 15 zeigt eine solche grafische Darstellung. Auf der linken Seite sind die Kategorien und Subkategorien abgebildet, die im Zentrum des Interesses stehen und im Vorfeld der Analyse aktiviert werden müssen. Rechts daneben sind diejenigen Interviews aufgeführt, die für die Auswertung relevant sind. Im speziellen Fall wurden sowohl die Kategorien „Vergleichsarbeiten als Rückmeldung für Verbesserungen“ mit Subkategorie „Diagnoseinstrument“,

als auch die Kategorie „Kontrollinstrument und Druckmittel“ mit ihren Subkategorien aufgeführt.

Abbildung 15: VA als Diagnose- versus Kontrollinstrument

An dieser Matrix können nun Ähnlichkeiten, Übereinstimmungen, Unterschiede oder Auffälligkeiten markiert werden. Um diese Auswertungs- und Analyseschritte nun exemplarisch anhand der Stichprobe der Realschullehrpersonen zu zeigen, werden farbige Markierungen verwendet.

Der ersten und zweiten Zeile der Matrix kann entnommen werden, dass sich die Aussagen von insgesamt zehn Realschullehrpersonen zur Funktion von Vergleichsarbeiten den Kategorien und Subkategorien „als Rückmeldung für Verbesserungen“ und/oder „als Diagnoseinstrument“ zuordnen lassen. Neun Realschullehrerinnen und Realschullehrer äußern sich im Rahmen der Antwortkategorien „Kontrollinstrument“ und/oder „Druckmittel“. Die Personen, die sich eindeutig den beiden erstgenannten Kategorien zuordnen lassen sind rot umkreist (Nr. 250, 251, 261, 266 – 272). Mit einer schwarzen Markierung wurden diejenigen gekennzeichnet, die Vergleichsarbeiten als Kontroll- und Druckmittel ansehen (Nr. 253, 254, 260, 263, 270 u. 271). Zwei Lehrpersonen, in der Abbildung grün markiert, nennen im Interview beide Aspekte (257 u. 264). Sie können mit dieser Form der Visualisierung nicht eindeutig der einen oder der anderen Gruppe zugeordnet werden. Um eine Zuordnung zu einer Gruppe zu ermöglichen, muss der jeweilige Einzelfall herangezogen werden, um die inhaltlichen Sinnzusammenhänge im spezifischen Fall zu erfassen (Schritt 3). Das „Stufenmodell empirisch begründeter Typenbildung“ (Kapitel 4.5) enthält dieses stufenweise Vorgehen. Es berücksichtigt den Fall, dass man „vor allem bei der Analyse der inhaltlichen Sinnzusammenhänge auf weitere relevante Merkmale stoßen [wird], die zu einer

Erweiterung des Merkmalsraums und damit zu einer neuen Gruppierung der Fälle führen“ (Kluge 1999, S. 261).

Wendet man dieses visualisierende Verfahren auf das Gesamtmaterial an, so finden sich weitere Kategorien und Subkategorien, anhand derer Unterschiede und Ähnlichkeiten sichtbar werden. Bei der vorliegenden Arbeit betrifft es folgende Themenbereiche, Kategorien und Subkategorien, deren Merkmalsausprägungen sich für die Typenbildung nach Kluge (1999) eignen:

1. Einstellung zu VA
2. Funktion von Vergleichsarbeiten (VA)
3. Durchführung der VA
4. Auseinandersetzung mit Rückmeldedaten
5. Anhaltspunkte für Schwächen und Stärken der Schüler
6. Bewertung einzelner Aspekte der VA
7. Innerschulische Kommunikation der Ergebnisse

Nachdem die „differenzierenden“ und kontrastierenden Themenbereiche genannt sind, erfolgt zunächst die Gruppenbildung, die im agglomerativen Verfahren erfolgt.

> *„Beim **agglomerativen Verfahren** wird von den Einzelfällen ausgegangen und untersucht, welche sich jeweils am stärksten ähneln, indem sie entweder in möglichst vielen Merkmalen übereinstimmen oder eine möglichst geringe Distanz aufweisen: die beiden Fälle, die sich am ähnlichsten sind, werden zunächst in einer Gruppe zusammengefasst. Im nächsten Schritt wird wieder geprüft, welche Fälle sich am stärksten ähneln, und es wird entweder ein neues Cluster gebildet oder der entsprechende Fall einem bereits gebildeten Cluster zugewiesen, wenn der Einzelfall und dieses Cluster die geringste Distanz aufweisen. Diese Schritte wiederholen sich, bis alle Fälle zugeordnet worden sind.“ (Kluge 1999, S. 270, [Hervorhebung im Original])*

Wie in Abb. 15 dargestellt, betrachtet man dabei Kategorien und Subkategorien der einzelnen Interviews und deren Merkmalsausprägungen und –dimensionen. Mit Hilfe des Code-Matrix-Browsers sucht man nach Fällen, die möglichst ähnlich gelagert sind oder sich auch kontrastierend gegenüberstehen. In der vorliegenden Arbeit führt diese Einzelfallanalyse zu fünf Typen, die sich bezüglich der Merkmalsausprägungen innerhalb der sieben oben aufgeführten relevanten Vergleichsdimensionen unterscheiden. Sie sind in Tabelle Nr. 5.7.1 dargestellt.

Tabelle 5.7.1: Typenbildung zur Rezeption und Nutzung von Vergleichsarbeiten (VA)

Typen Merkmals kategorien	Der begeisterte, überzeugte und befürwortendeTyp	Der positiv mitspielende Typ	Der desinteressierte, gelassene Typ	Der negativ kritische Typ	Der ablehnende, belastete Typ
Funktion der VA	Vergleichsarbeiten (VA) eignen sich bestens als Diagnose und Förderungsinstrument. VA dienen nicht als Kontrollinstrument, sondern als ‚positives Druckmittel'. VA geben objektive, kriteriale Zielmarken vor, die als Vergleichsgrundlage dienen.	VA dienen der Überprüfung und der Orientierung an Standards; summative Evaluation. VA als Konkretisierung der Bildungspläne. Lehrpersonen fühlen sich nicht kontrolliert.	VA dienen dem Vergleich. VA als Verpflichtung, kein Nutzen erkennbar. Lehrpersonen fühlen sich persönlich nicht kontrolliert. Druckmittel für diejenigen, die sich bisher nicht an Vorgaben gehalten haben.	Lehrende erleben VA als Kontrollinstrument. VA dienen als Druckmittel, Ranking. VA als Verpflichtung, als purer Aktionismus. Förderung steht nicht im Vordergrund.	VA sind ein Kontrollinstrument und Druckmittel. VA dient als Alibi. VA als purer Aktionismus, ohne Funktion. Kein pädagogischer Nutzen erkennbar.
Einstellung zu VA	Lehrpersonen betonen ihre positive Einstellung zu VA und zu Evaluation. Sie bringen ihre Freude zum Ausdruck. Sie befürworten VA uneingeschränkt. Sie sehen viele Chancen und Nutzungsmöglichkeiten in VA.	Lehrende erachten VA im Allgemeinen als sinnvoll. Positiver Nutzen erkennbar, jedoch keine Zeit für Umsetzung.	Gelassene, desinteressierte Einstellung zu VA. Zurückhaltende Äußerungen, halten persönliche Meinung zurück. Nehmen VA als gegeben hin.	Lehrpersonen können keinen Nutzen erkennen, wenn nicht wesentliche Änderungen erfolgen. Betonen ihre äußerst kritische Haltung. Generelle Kritik gegen Vergleiche.	VA sind sehr schlecht und bringen keinen Nutzen. VA als sinnlose Arbeitsbeschaffungsmaßnahme. Stark belastet, weil sie zu VA gezwungen werden.
Durchführung der VA	Problemlose Durchführung, schnelle Korrektur.	Problemlose Durchführung, Zeit teilweise knapp.	Problemlose Durchführung, halten sich nicht an Vorgaben	Im Allg. durchführbar, uneinheitliche Durchführung.	Uneinheitliche Durchführung.
Auseinandersetzung mit Rückmeldedaten	Großes Interesse an Rückmeldedaten. Intensive Auseinandersetzung mit Grafiken und Datenblättern. Freude an der Auseinandersetzung.	Schauen sich Rückmeldedaten an. Diese sind verständlich und Lehrperson zeigt teilweise Interesse. Eher oberflächliche Auseinandersetzung, bei „Ausreißern" intensiver.	Rückmeldedaten werden nicht angeschaut, aufgrund von Desinteresse oder fehlender Zeit. Teilweise werden VA nicht ausgewertet.	Setzen sich teilweise sehr kritisch mit Rückmeldedaten auseinander. Kritische Haltung wird durch Auseinandersetzung verstärkt. Teilweise keine Auseinandersetzung.	Kein Interesse an Rückmeldedaten. Grafiken werden nicht angeschaut.

Typen Merkmals kategorien	Der begeisterte, überzeugte und befürwortendeTyp	Der positiv mitspielende Typ	Der desinteressierte, gelassene Typ	Der negativ kritische Typ	Der ablehnende, belastete Typ
Anhaltspunkte für Schwächen und Stärken der SuS	Erhalten Anhaltspunkte über Stärken und Schwächen ihrer Schüler. Teilweise bessere Rückmeldung als in Klassenarbeiten.	Fühlen sich bestätigt. Erhalten teilweise neue Anhaltspunkte und Anregungen. Regt Eigenreflexion an. Sind bereit Dinge zu ändern (Unterricht, Themen, Förderunterricht…).	VA liefern keine neuen Anhaltspunkte zu Schwächen oder Stärken.	VA liefert keine neuen Hinweise oder Anhaltspunkte – fühlen sich bestätigt.	Keine Hinweise aus VA.
Bewertung einzelner Aspekte der VA	Nutzen auf allen Ebenen der Einzelschule. Einzige Kritik/Probleme: keine Förderangebote und zuviel zusätzliche organisatorische Aufgaben. Erkennen mögliche Probleme die einer gelingender Evaluation im Wege stehen. Nutzen dennoch ihre restliche Zeit für Förderung.	Lehrpersonen nennen einige positive Aspekte der VA. Negative Punkte betreffen Zeitpunkt und/oder Niveau und/oder Bewertungspraxis Kritisieren generell die Benotung und fehlende Unterstützung. Verbesserungsvorschläge werden formuliert. VA sind noch in der Entwicklungsphase – keine große Bedeutung.	Kritik an Objektivität. Kritisieren Aussagekraft.	Lehrpersonen nennen ausschließlich negative Punkte von VA – Zusammenstellung der Aufgaben- oder Themenwahl. Bemängeln fehlende Zeit , fehlendes schulisches Angebot. Große Enttäuschung. Wut über Bewertung. Legitimationsproblem.	Äußern ausschließlich Kritik. Nennen gravierende Fehler: Willkürliche, undifferenzierte Benotung und Bewertung. Niveau etc…
Auswirkungen auf Änderungen der Unterrichtspraxis	Erweiterung um komplexere Aufgaben. Mehr Individualisierung – z.B. Gruppenarbeitsphasen. Veränderte Auswahl von Unterrichtsmaterialien.	Kaum Auswirkungen, da Zeit für Umsetzung von Maßnahmen fehlt. Wiederholen vor allem Themen der VA.	Keine Änderungen, teilweise wird distanziert von Änderungen bei KollegInnen berichtet.	Keine Auswirkungen im positiven Sinn. Teaching-to-the-test. Mehr Hausaufgaben. Mehr lehrerzentrierte Unterrichtsformen.	Keine Auswirkungen.
Innerschulische Kommunikation der Ergebnisse	Intensive, systematische Diskussion über Ergebnisse mit allen beteiligten Akteuren (Schulleitung, Gremien, Kollegium, Schülerinnen und Schüler, Eltern)	Systematische Diskussionen werden teilweise geführt. Wünschen sich intensiveren Austausch.	Keine systematischen Diskussionen über Vergleichsarbeiten, nur informell.	Vielfältige systematische Diskussion über Ergebnisse mit unterschiedlichen Akteuren.	Kein Austausch über Vergleichsarbeiten.

Schritt 3: Analyse der inhaltlichen Sinnzusammenhänge und Typenbildung,

Im vorangegangenen Abschnitt wurden die Typen empirisch, anhand der Stichprobe der Realschullehrpersonen, ermittelt, ohne die Sinnzusammenhänge innerhalb eines Typs noch zwischen den fünf Typen zu berücksichtigen. Nun liegt aber die Stärke von qualitativen Forschungsmethoden genau darin, dass die einzelnen Gruppen nicht nur anhand empirisch gewonnener Kategorien und Subkategorien rekonstruiert, sondern auch erklärt und dadurch verstanden werden. Mit der Analyse der inhaltlichen Sinnzusammenhänge des gesamten Interviewmaterials (n=59) möchte man zeigen, „dass das Zusammentreffen einzelner Merkmale nicht zufällig ist, sondern daß sich hinter diesen "äußeren" Korrelationen "innere" Sinnzusammenhänge verbergen, mit deren Hilfe die einzelnen Typen (…) sowie das Gesamtgefüge einer Typologie (…) verstanden und erklärt werden kann" (Kluge 1999, S. 277, [Hervorhebungen im Original]).

Dieser Verstehensprozess wird dadurch ermöglicht, indem die aufgeführten relevanten Kategorien und deren Ausprägungen mit einschlägigen Textpassagen aus dem Interviewmaterial belegt oder auch widerlegt werden. Auf diese Weise können die vermuteten Zusammenhänge rekonstruiert werden. Mit Hilfe computerunterstützter Auswertungsverfahren, wie MAXQDA®, können die entsprechenden Textpassagen durch die Funktion des Text-Retrievals schnell gefunden und für die Analyse verwendet werden (vgl. Kuckartz 2010; Kelle & Kluge 2010; Kluge 1999). Die Analyse der inhaltlichen Sinnzusammenhänge erfolgt „typenweise", beginnend mit „dem begeisterten, überzeugten Typ".

5.7.1 Der begeisterte, überzeugte und befürwortende Typ

Die Personen, die diesem zugeordnet werden können, betonen im Interview ihre positive Einstellung zu Vergleichsarbeiten und bringen ihre Freude darüber zum Ausdruck. Sie sind von dem neuen Vergleichsinstrument überzeugt und befürworten die damit einhergehenden Neuerungen und Änderungen sehr.

> „Ich persönlich befürworte die Vergleichsarbeiten, bin ein absoluter Befürworter. Ich (…) finde es klasse.“(263)

> „Ich freue mich über neue Aufgaben, über neue Aufgabenformen die ich da bekomme.“ (251, 35)

Die durchweg positive, befürwortende Einstellung und Haltung zu Vergleichsarbeiten die bei diesem Typus zum Ausdruck kommt, lässt darauf schließen, dass sich diese Personen, mit besonderem Engagement, Fleiß und Elan den Vergleichsarbeiten und den damit in Zusammenhang stehenden Änderungen widmen. Die Offenheit und Bereitschaft gegenüber den Neuerungen und die damit verbundenen Chancen werden in den Interviews deutlich benannt. Dieser Personenkreis ist gespannt, neugierig, motiviert und gänzlich davon überzeugt, dass in diesem Instrument ein großes Potential liegt, das im Sinne der Qualitätssicherung und Qualitätsverbesserung von Schule und Unterricht einen wesentlichen Beitrag leisten kann.

Als Gründe für die positive Grundstimmung und befürwortende Haltung führen sie verschiedene Punkte an. Einerseits begegnen sie Leistungsvergleichen und Evaluationsverfahren im Schul- und Unterrichtswesen generell offen. Denn durch Vergleichsarbeiten werden Zielmarken vorgegeben, deren Erreichen angestrebt und am Ende eines Bildungsabschnitts überprüft werden. Somit ist der Leistungsvergleich zwischen einzelnen Schülerinnen und Schülern, zwischen Lehrerinnen und Lehrern und zwischen verschiedenen Schulen möglich und wird sichtbar. Diese „Benchmarks“ bieten den Lehrpersonen Orientierung und geben einen klaren Rahmen vor, der als Orientierung und Referenz dient und deren professionelles Handeln rahmt und bestimmt.

> „Ich war schon immer dafür, dass Klassen, möglicherweise auch Lehrer, dass das, was man als Lehrer leistet, vergleichbar gemacht werden muss, oder vergleichbar sein kann.“ (150)

> „Ich habe ein Niveau, das mir vorgegeben wird. Ich kann schon in Klasse 5 sagen, auf dieses Niveau muss ich die Kinder vorbereiten.“ (263)

Ein Lehrer dieser Gruppe zeigt sich gegenüber Vergleichsarbeiten besonders aufgeschlossen. Er kann sich vorstellen, diese jedes Schuljahr in bestimmten Fächern mit dem Ziel durchzuführen, von externer und objektiver Seite Vergleichskriterien zu erhalten, um daran sein Lehrerhandeln in unterschiedlichen Bereichen zu reflektieren und gegebenenfalls abzuändern oder neu auszurichten. Als Beispiel führt er die Beurteilungs- und Bewertungspraxis der Lehrenden an, die sich eher an der individuellen oder sozialen Bezugsnorm orientieren, als an kriterialen Normen. Durch Vergleichsarbeiten, die sich an Bildungsstandards orientieren, erhalten Lehrerinnen und Lehrer diese externen Referenzkriterien, mit denen die jeweiligen Leistungsergebnisse ihrer Schülerinnen und Schüler verglichen werden können.

> „Für mich wäre es interessant, wenn es nicht nur in Klasse sechs, sondern wenn es möglich wäre, zentral in jeder Klasse zu Schwerpunktfächern gegen Ende des Schuljahres eine Vergleichsarbeit durchgeführt werden könnte. In dem Fall wäre es für mich dann auch, wenn meine Klassen leistungsstärker und leistungsschwächer sind, in dem Fall wäre es für mich dann auch ein Hinweis darauf, wie weit habe ich möglicherweise versehentlich meine Bewertungen an eine schwächeren Klasse angepasst oder an eine stärkere Klasse angepasst. Also, ich wäre von dem her dem gegenüber recht aufgeschlossen, wenn von unabhängiger Seite her mir eine Vergleichsmöglichkeit an die Hand gegeben wird.“ (150)

Andererseits kommt die positive Grundstimmung und Zufriedenheit bei dieser Gruppe vor allem dann zum Ausdruck, wenn schwächere Schülerinnen und Schüler in den Vergleichsarbeiten besonders gute Ergebnisse erzielen oder wenn die ganze Klasse besser abschneidet als die Vergleichsgruppe. Die Lehrerinnen und Lehrer erleben das überdurchschnittliche Abschneiden ihrer Klasse als Bestätigung und freuen sich besonders über die Leistungszuwächse ihrer schwächsten Schülerschaft.

> „Es ist für mich, ich muss schon sagen, eine gewisse Befriedigung, wenn ich dann sehe, dass man doch an schwächere Schüler, doch gerade im Fach Mathe, Englisch geht auch noch, Deutsch ist dann schwieriger, etwas hinbringen kann und sie über den Durchschnitt `rausheben kann.“ (150)

> „Was für mich interessant war, war der Vergleich mit dieser Vergleichsgruppe dort, also mit der Vergleichsklasse. Und natürlich freut man sich da, wenn man da in der Vergleichsstichprobe, wenn da mit der eigenen Klasse gut ist.“ (267, 26-27)

Gleichzeitig sind sich die Befürworterinnen und Befürworter von Vergleichsarbeiten darüber bewusst, dass die meisten ihrer Kolleginnen und Kollegen diese positive und überzeugte Einstellung nicht teilen und sie mit dieser Meinung eher zu einer Minderheit gehören.

> „Von daher bin ich wahrscheinlich einer von den wahrscheinlich Wenigen, die Vergleichsarbeiten befürworten. (…) Aber, ich glaube da stehe ich einzeln, alleine, auf weiter Flur.“ (150)

In diesem Zusammenhang weisen sie ebenfalls ausdrücklich darauf hin, dass sie sich im Vergleich zu Manchen im Kollegium durch Vergleichsarbeiten nicht kontrolliert fühlen und durch dieses externe Evaluationsverfahren nicht unter Druck geraten. Im Gegenteil, eine Lehrerin bewertet Vergleichsarbeiten sogar „als schönes Druckmittel“, um bestimmte, weniger engagierte Kolleginnen und Kollegen zu mehr Mitarbeit zu bewegen.

> „Ich fühle mich da jetzt nicht da kontrolliert darüber, (…) nein ich fühle mich nicht kontrolliert.“ (251, 36-36)

> „Und allein das ist schon so ein schönes großes Druckmittel für manche Lehrer, dass die sich auch endlich mal auf den Hintern setzen (…).“ (368, 64-64)

Die Nutzungsmöglichkeiten, die diese Lehrpersonen im Vergleichen von Leistungsergebnissen vorwiegend sehen, beziehen sich auf mehrere Ebenen und auf unterschiedliche Akteure der Einzelschule. Auf der Individualebene können Vergleichsarbeiten einerseits als Diagnoseinstrument für die einzelnen Schülerinnen und Schüler und für deren Förderung verwendet werden. Andererseits betrachten einige Vergleichsarbeiten auch als Rückmeldung für sich selbst, um daraus Änderungen abzuleiten. Diese Sichtweise fokussiert demnach die Unterrichtsebene und umfasst alle möglichen Verbesserungen und Entwicklungen, sowohl die der Schülerinnen und Schüler, als auch die der Lehrerinnen und Lehrer.

> „Ich denke mir, das ist eine Rückmeldung für einen selber und dann kann man was draus machen oder nicht. Also ich hoffe, dass jeder was draus macht.“ (267, 15-15)

Auch auf der Ebene der Schule, insbesondere für die Schulleitung, wird ein möglicher Nutzen gesehen. Dadurch, dass durch die Rückmeldeinformationen auch einzelne Schulen miteinander verglichen werden können, erhält man Anhaltspunkte, wo die jeweilige Einzelschule ansetzen kann, um ihre Leistungsergebnisse insgesamt zu verbessern.

> „Natürlich sind die Schulen da vergleichbar. Damit die dann sehen, die schaffen so, die schaffen so und wie kann man das dann für eine bessere Rückmeldung für den Rektor, wo er dann anknüpfen kann.“ (267, 13-13)

Für diesen Typus stellt es eine Selbstverständlichkeit dar, dass die Rückmeldungen aus Leistungsvergleichen für alle Beteiligten im Bildungssystem ein Nutzungspotential bieten können. Sie sehen den positiven Nutzen sowohl auf Unterrichts- als auch auf Schulebene. In Anlehnung an das Modell von Ditton (2000) zur Qualität und Qualitätsentwicklung im Bildungsbereich, das in Kapitel 2.3.3 in Abbildung 4 dargestellt ist, beziehen sich die Aussagen dieser Lehrpersonen auf die Punkte 3 und 4 im Modell. Demzufolge leisten Vergleichsarbeiten sowohl einen Beitrag auf Institutionsebene, indem sie zur Qualität der einzelnen Bildungseinrichtung beitragen, als auch auf Individualebene zur Verbesserung der Qualität der Lehr-Lernsituation.

Wie die meisten Lehrpersonen aus der Gesamtstichprobe bewerten auch diese die Durchführung als unproblematisch. Sie heben darüber hinaus noch einen Vorteil von Vergleicharbeiten hervor. Sie sei leichter zu korrigieren als jede andere Arbeit und man benötige für die Korrektur kaum Zeit.

> „Leichter zu korrigieren als jede andere Arbeit.“ (263, 25-25)

> „Ich habe noch nie so eine schöne Arbeit zu korrigieren gehabt. (LACHEN). Auch wenn es mit einer der schlechteren Schnitte in der Klasse war, aber die Korrekturzeit hat sich auf eine halbe Stunde oder so reduziert durch die Items.“ (368, 16-16)

Analog zur überaus positiven Einstellung und Haltung zu Vergleichsarbeiten, setzt sich diese Personenkreis auch besonders intensiv, mit starkem Interesse, Freude und Engagement mit den Auswertungsinstrumenten auseinander. Die Grafiken und Tabellen werden als hilfreich, anregend und verständlich bewertet. Das Interesse der Lehrerinnen und Lehrer an den Ergebnissen aus Vergleichsarbeiten liegt auch hier sowohl auf der schulischen Ebene, indem die Klasse im Vergleich

zur landesweiten Stichprobe betrachtet wird, als auch auf Individualebene, bei der einzelne Schülerinnen und Schülern miteinander verglichen werden können.

> „Ja, das regt es an, das auf jeden Fall. Ich fand es interessant. Die Auswertung hat mir auch Spaß gemacht. Diese Diagramme anschließend mir angeschaut. Ich war ganz gespannt was da heraus kommt. Das ist mehr als nur Schnitte anzuschauen am Ende.“ (361, 43-43)

> „Das was mich an dem interessiert, das ist die Auflösung, wo die Aufgaben aufgeschlüsselt sind, weil ich ja die Aufgaben kenne und sehe, okay das müssten sie jetzt können und das müssen sie nicht können. Das finde ich immer hochinteressant für mich zu sehen, wie bin ich im Landesvergleich. Wenn ich eine Aufgabe, wo ich weiß, so was in der Art habe ich intensiver geübt gehabt, sind wir dann tatsächlich besser als der Landesvergleich oder wirkt sich das gar nicht aus. Und das ist das, was ich anschaue. Diese Boxplots, diese Verteilung die überfliege ich im Prinzip. Diese Aufgabenaufschlüsselung ist das, was wirklich total interessant ist, weil wenn man was intensiv im Unterricht bearbeitet hat, also ich bekommen die Vergleichsarbeit, denke, ach super, die machen das,(…) das ist kein Problem und dann machen die das und sind dann trotzdem im Landesschnitt schlechter, dann weiß ich, da ist irgendwas schief gelaufen.“ (251, 57-60)

Vor allem aufgrund der übersichtlichen Darstellung der Ergebnisse lassen sich ihrer Meinung nach Rückschlüsse auf die Stärken und Schwächen der Schülerinnen und Schüler ziehen. Die intensive Auseinandersetzung mit den Daten regt diesen Typus dazu an, nach möglichen Ursachen und Gründen, die für das jeweilige Abschneiden ihrer Klasse, ihrer Schülerinnen und Schüler verantwortlich sein könnten, zu suchen (vgl. 361).

> „Also jetzt im Nachhinein, nachdem ich sie vor mir liegen hatte, gelesen hatte und auch die Ergebnisse, die ja hier wunderbar schön rauskamen, gesehen habe, war mir das auf jeden Fall klar, dass ich dort die Lernschwierigkeiten meiner Schüler herauslesen kann und das war für mich super.“ (267, 10-10)

> „Ich fand es schon interessant. Man fängt dann schon an zu überlegen. Warum ist an dem Punkt jetzt hier so diese Differenz. Ich habe eine Zeitlang überlegt, sind die jetzt da so viel besser? Ich hab dann nach und nach so gedacht, wahrscheinlich habe ich an der Stelle vorsichtiger korrigiert oder großzügiger korrigiert. Das weiß ich jetzt aber nicht, was der Grund ist.“ (361, 39-39)

Eine Lehrerin ist der Ansicht, dass diese Arbeiten im Vergleich zu Klassenarbeiten viel deutlicher mögliche Anhaltspunkte über die Leistungsfähigkeit des einzelnen Kindes liefern können. Da Schülerinnen und Schüler im Gegensatz zu Klassenarbeiten auf Vergleichsarbeiten nicht lernen können, schneiden „clevere“ Kinder besser ab, selbst wenn sie nicht zu den „Fleißigen“ zählen. Hingegen er-

zielen Schwächere, die für Klassenarbeiten z.B. Inhalte auswendig lernen, in Vergleichsarbeiten schlechtere Ergebnisse. Sie ist davon überzeugt, dass es sich bei jenen vornehmlich um Mädchen handele, die weniger in der Lage sind, vernetzt zu denken.

> „Ein cleveres Kind, das aber stinkefaul ist, das schreibt im Vergleich zu den Klassenarbeiten relativ gut so eine Vergleichsarbeit. (…). Und ein Auswendiglerner, was oft Mädchen sind, die ein wenig schwach sind, da fällt es in der Vergleichsarbeit auf. Das habe ich auch hier deutlich gesehen.“ (263, 35-35)

Dadurch, dass die Lehrerinnen und Lehrer aufgrund der intensiven Auseinandersetzung mit den Rückmeldedaten Hinweise auf Stärken und Schwächen erhalten und mögliche Ursachen reflektieren, sehen sie auch Chancen, bestimmte Maßnahmen sowohl zur Verbesserung der Schülerleistungen als auch des Unterrichts zu ergreifen. Als einen möglichen Ansatzpunkt formulieren sie die Förderung der einzelnen Schülerinnen und Schüler. Sie räumen dabei allerdings sofort ein, dass dieser Wunsch an der Schule und im Unterricht häufig nicht realisiert werden kann. Einerseits stellt die fehlende Zeit im Anschluss an die Vergleichsarbeit ein Problem dar. Andererseits können keine Maßnahmen ergriffen werden, da an der Schule aufgrund unterschiedlicher Faktoren keine Förderangebote existieren.

Eine Lehrerin berichtet, dass es an ihrer Realschule neben dem Unterrichtsangebot eine breite Palette an Projekten und Initiativen gibt und dadurch kaum Stunden für Fördermaßnahmen zur Verfügung stehen. Und die eine Förderstunde, die den Schülerinnen und Schülern der sechsten Klassen angeboten wird, fällt die meiste Zeit krankheitsbedingt aus.

> „Bei uns gibt es insofern keine Fördermaßnahmen, weil wir keine Stunden mehr haben. Wir haben das „Jugend forscht“, wo viele Stunden drauf gehen, wir haben das „Rote Kreuz“, Schulsanitäter, Streitschlichter und alles Mögliche. Es gab auch in dem Schuljahr, wo die 6. waren, gab es einen Mathestütz- und Förderkurs, aber am Mittwoch in der 6. Stunde, das ist unsere Verfügungsstunde und wenn es so war, konnte die Kollegin die Stunde nicht halten und ab Mai ist die ausgefallen, weil sie schwerkrank geworden ist.“ (263, 37-37)

Über diese schulischen Gegebenheiten herrscht bei diesen Lehrerinnen und Lehrern Unzufriedenheit und es wird nach Lösungsmöglichkeiten gesucht, wie diese Defizite behoben werden könnten. Sie entwickeln teilweise klare Vorstellungen darüber, wie eine Optimierung der schulischen Situation zu erreichen wäre.

Gleichzeitig führen sie die Gründe an, die deren Realisierung verhindern. Die Lehrpersonen hätten zum Beispiel am liebsten eine wöchentliche Förderstunde mehr in ihrer Klasse, bei Bedarf würden sie diese auch auf freiwilliger Basis anbieten. Dies ist jedoch im Schulalltag aufgrund der zusätzlichen Arbeitsbelastung, die z.B. durch Kompetenzprüfungen und andere schulische Programme und Innovationen zustande kommt, nicht realisierbar. Sie beklagen massiv, dass in der Schule über Wochen hinweg viel Zeit für zusätzliche organisatorische Aufgaben verwendet wird, wovon die Schülerinnen und Schülern keinen direkten Nutzen haben, obwohl der Förderbedarf offensichtlich ist.

> „Und ich würde das sehr begrüßen, Förderunterricht. Ich würde es sogar sehr begrüßen der Lehrer fördert seine Klasse eine Stunde mehr. Wäre mir bedeutend lieber, als stundenlang irgendwelche Organisation für Kompetenzprüfungen und so was in meine Zeit reinzuhängen, die die mir auch keiner zahlt. Würde ich viel lieber auf freiwilliger Basis meine Wiederholungsstunden übernehmen. Und da die Zeit jetzt, wenn ich überlege seit Anfang des Schuljahrs, ich glaube ich habe jede Woche mindestens einen Mittag auf irgendwas schulisches, was aber meinen Schülern nicht direkt bekommt, habe ich investiert. Und in Mathe hätten die es so nötig. Eine freiwillige Stunde. „Ihr kommt, wenn ihr meint ihr habt Nachholbedarf.“ (ebd.)

Diesen Lehrerinnen und Lehrern ist durchaus bewusst, dass die vorherrschenden schulischen Bedingungen kontraproduktiv sind und einer gelingenden Evaluation im Wege stehen können. Sie sehen die Gefahr, dass alle Maßnahmen und Anstrengungen der Lehrpersonen, die mit sehr viel Aufwand und Engagement verbunden sind, nur dann Wirkung zeigen, wenn die ganze Schule dahinter steht und die Prozesse mit unterstützt. Bleibt diese Unterstützung aus, so bleibt ihrer Meinung nach das ganze Verfahren wirkungslos, trotz der enormen Anstrengungsbereitschaft einzelner Personen.

> „Von daher ist vieles nicht in Ordnung. Und wenn die ganze Evaluation einem nichts bringt, wenn eine Schule nicht sagt: „Ihr dürft jetzt fünf Stunden mehr verbraten“, dann ist es irgendwo sinnlos der ganze Aufwand. Und es ist ein Mordsaufwand.“ (ebd.)

Dennoch nutzen einige Lehrpersonen trotz des fehlenden schulischen Angebots die verbleibende Zeit, um Fördermaßnahmen, die aus den Rückmeldungen abgeleitet wurden, im Unterricht umzusetzen.

> „Aber ich habe dann die Zeit genutzt, ich habe es kopiert für die Schüler. Ich habe gesagt: „Schaut, das müssen wir aufarbeiten. Das sind die Dinge, die

> ihr auf die Reihe bringen müsst.“ (…) Das sind die Dinge, die mir wichtig sind.“ (263)

Im Unterschied zu den anderen Rezeptions- und Nutzungstypen, die in der vorliegenden Arbeit generiert werden konnten, stellt diese Gruppe im Bereich Auswirkungen und Änderungen der Unterrichtspraxis eine Ausnahme dar. Sie geben als einzige Gruppe an, bereits aufgrund der Rückmeldungen und Hinweise aus Vergleichsarbeiten konkrete Maßnahmen abzuleiten und in ihrem Unterricht anzuwenden. Die Änderungen beziehen sich dabei auf unterschiedliche Bereiche des Unterrichts. Einige haben sich in dieser Gruppe vermehrt mit ihren Kolleginnen und Kollegen über ihre bisherigen Arbeiten ausgetauscht und das Spektrum im Team um anspruchsvolle, schwierige Aufgaben erweitert, die die Schülerinnen und Schüler auch einmal an ihre Grenzen bringen. Sie wenden vermehrt komplexe Aufgabenstellungen an und fordern die Kinder heraus.

> „(Die Tests) sind anders geworden. Nicht mehr 5 gleiche Aufgaben, sondern jede Aufgabe anders. (…). Andere Aufgaben ja. Das haben wir besprochen. Auch meine Klassenarbeiten sehen inzwischen ganz anders aus. Ich drücke mich nicht mehr um schwere Aufgaben. (…). Es hat sich bei mir einiges geändert. Wir tauschen da immer die Arbeiten aus. ‚Und was hast du?’ So Dinge mit Zahlenstrahl einteilen habe ich früher nicht gemacht. Ich will nicht, dass ein Kind Angst kriegt vor Mathe. Aber es muss sein, diese Sache muss sein. (…) Und deshalb war es dieses Jahr drin. Und ich sehe halt manche packen es nicht. Aber das sind wieder die abstrakten Sachen.“ (263, 54, 63 und 66)

Manche planen im Zuge der Vergleichsarbeiten im Unterricht vermehrt Gruppenarbeitsphasen ein, die zwar mehr Zeit beanspruchen und sich nur für Doppelstunden eignen, aber den Vorteil bieten, dass die Selbstständigkeit und die Selbsttätigkeit der Schülerinnen und Schüler gefördert wird.

> „Ich mache in Mathe mehr Gruppenarbeit, aber ich habe oft abbrechen müssen die Gruppenarbeit. Ich habe Freitag immer eine Doppelstunde gehabt und da bietet es sich einfach an, weil man reagieren kann, wenn man nur 45 Minuten hat. Ich habe genaue Aufgabestellung gehabt, das ging mir immer um das Selbsterarbeiten. Das wollte ich denen beibringen, nicht das was der Lehrer vorne vorturnt ist richtig.“ (263, 67-67)

> „Was für mich jetzt wichtig war, was wir letztes Mal, ich meine bei 5. Klässlern ist es immer so, sie brauchen eine Anlaufzeit und arbeiten noch relativ langsam. Aber was wir uns ausgemacht haben in der letzten Klassenkonferenz war, dass wir einfach auch vermehrt Wert darauf legen, dass die Schüler selbstständig werden.“ (276, 59-59)

Teilweise wählen die Lehrerinnen und Lehrer im Zuge der Vergleichsarbeiten ihre Unterrichtsmaterialien nach anderen, spezifischeren Kriterien aus. Sie geben den Schülerinnen und Schülern auch vermehrt Gelegenheit, um geeignete Materialien und Methoden zur Sicherung des Lerninhalts selbst herzustellen bzw. auszuwählen. Sie sind auch gerne dazu bereit, die benötigte Zeit für aufwändigere Verfahren im Unterricht zur Verfügung zu stellen. Eine Lehrerin berichtet, dass von Seiten der Schulleitung für die neuen Lernmethoden Extrastunden im Stundenkontingent eingeplant wurden.

> „Man wählt vielleicht das Material, das man hat anders aus. Ja doch." (263, 77-77)

> „(…) und das bedeutet natürlich, dass sie am Ende von einem Unterrichtsstoffgebiet z. B. sich selber Spiele überlegen können und machen, egal ob das ein Würfelspiel ist. Wir Lehrer bereiten teilweise Kärtchentechnik vor, dass sie einfach sich selber Material herstellen, aber von uns auch welches bekommen. Dazu finden wir auf jeden Fall Zeit, ich kann es im Stundenplan mit einbauen, da habe ich Zeit bekommen. (…). Da hat sich sogar im Stundenplan was getan, wenn man so will. Von Schulleiters Seite her." (276, 57-59)

Die Kommunikation über die Ergebnisse aus Vergleichsarbeiten findet bei der Gruppe der Befürworterinnen und Befürworter vor allem auf systematische Art und Weise statt. Am häufigsten setzen sich diese Lehrpersonen intensiv mit ihren Parallelkolleginnen und Parallelkollegen, mit ausgewählten Personen ihres Kollegiums und mit ihren Schülerinnen und Schülern über die Ergebnisse auseinander. Eine Realschullehrperson erweitert diesen Kreis noch und diskutiert die Ergebnisse in nahezu allen Gremien, mit Ausnahme der Gesamtlehrer- und der Schulkonferenz. Mit den Eltern der Schülerinnen und Schüler reichte in den meisten Fällen die Zeit nicht, um in systematischen Diskussionen die Ergebnisse zu besprechen.

> „Also auf jeden Fall war es natürlich wichtig für uns, dass wir mit den Parallelkollegen sprechen. Fachkonferenz hatten wir keine mehr, in der GLK hat es der SL mitgeteilt, wie es gelaufen ist. Gut, wir haben natürlich untereinander auch fächerübergreifend geschaut, weil es in der Klasse war. Denn es ist ja so, dass in Deutsch ähnliche Probleme aufgetreten sind wie in EWG. Ja gut, wenn man es so will Jahrgangsstufenkonferenz. Schulleitung sowieso, hat den Zugang [zu den Ergebnissen und] reicht es weiter." (267, 67-73)

5.7.2 Der positiv mitspielende Typ

Die Lehrerinnen und Lehrer, die dem „positiv mitspielenden Typus“ zugewiesen werden können, sehen den Grund für Vergleichsarbeiten vornehmlich in der Überprüfung der Bildungsstandards, im Sinne einer Standardsicherung, die am Ende eines bestimmten Bildungsabschnittes als bilanzierendes und somit summatives Evaluationsverfahren durchgeführt wird.

> „Um sicherzustellen, dass am Ende der Klasse 6, 8 oder 10 die jeweils auf zwei Jahre festgeschriebenen oder ausgeschriebenen Standards auch eingehalten werden.“ (354, 26-26)

> „Eigentlich genau das, dass man sagt, man muss gewisse Standards erreichen, Ende der Klasse sechs. Und da gibt es Überprüfung und dass wir die erreicht haben, zeigt dann diese Arbeit.“ (271, 33-33)

> „Also, für die Schüler habe ich es einfach so erklärt, es ist noch mal eine Zusammenfassung der gesamten letzten beiden Schuljahre, das als Klassenarbeit gemacht wird. Dient dazu, dass wir einfach den ganzen Stoff noch mal wiederholen, dass ihr einfach gut für die nächsten beiden Schuljahre gerüstet seit.“ (272, 33-33)

Zum einen soll mit Hilfe von Vergleichsarbeiten überprüft werden, ob diese, vom Ministerium vorgegebenen Standards, überhaupt erreicht werden. Sie erachten es als sinnvoll, dass diese als Benchmarks und somit als festgelegte Vergleichskriterien genutzt werden. Denn dadurch bietet sich ihrer Meinung nach nicht nur ein Vergleich für die Leistungen der einzelnen Schülerinnen und Schüler und Klassen einer Schule, sondern auch ein Vergleich zwischen den Schulen einer Region oder des Landes an. Die Lehrpersonen sehen sogar die Möglichkeit, dass Vergleiche zwischen verschiedenen Bundesländern möglich sind, wenn man als Referenz die Bildungsstandards der Kultusministerkonferenz (KMK) zugrunde legt.

> „Diese Vergleichsarbeiten wurden eingeführt, um einen gemeinsamen Standard zu bekommen. Dass Lehrer sich an gewisse Vorgaben halten müssen, die dann überprüft werden. Gerade vielleicht auch, es gibt ja diese nationalen Bildungsstandards, dass man innerhalb Deutschlands vergleichbarer wird, dass Lehrer gewisse Vorgaben haben.“ (271, 27-27)

> „Darüber hinaus ist es natürlich sehr sinnvoll, wenn man versucht so etwas wie eine Benchmark anzusetzten. (…) Aber es ist auch für den Lehrer eine gute Handhabe wenn es eine Messlatte gibt. Weil allein durch die Konzeption der Schule, oder auch vom Standort her, kann das ja sein, dass man atypische Klassen hat und meint das ist normal und sich vom Niveau überhaupt nicht mehr vergleichen kann mit einem Durchschnittsniveau. Also eine

> Benchmark ist eigentlich sinnvoll, auch in der Schule. Das würde ich nicht negativ sehen.“ (362, 38-38)

Zum anderen führen sie als Erklärung an, dass durch Vergleichsarbeiten die neuen, relativ offen formulierten Bildungspläne konkretisiert werden. Im Vergleich zu den baden-württembergischen Lehrplänen, die bis zum Jahre 2004 galten und sowohl konkrete Inhalte als auch Lehr-Lernziele beinhalteten, wird der neue Bildungsplan 2004 als vage empfunden. Durch Vergleichsarbeiten, die das Erreichen der Bildungsstandards überprüfen, werden nun die Erwartungen und Ziele, die das Ministerium verfolgt, sichtbar und Lehrpersonen können sich daran orientieren und ihr Handeln entsprechend ausrichten.

> „Mittlerweile vermute ich, dass es einfach darum geht, einfach diesen neuen Bildungsplan ein bisschen konkreter werden zu lassen. Dass wir als Lehrer sehen, um was es denen eigentlich geht. Und dass wir einfach auch da eine gewisse Rückmeldung haben, um das was wir machen sich mit dem deckt, was sich das Ministerium darunter vorstellt. (…) Weil im Prinzip ist es so, dass der Bildungsplan in Mathematik total vage ist. Da gibt es manche Inhalte, da brauche ich keinen Matheunterricht dafür machen. Und wenn ich jetzt halt, um so mehr Vergleichsarbeiten ich sehe, um so mehr weiß ich um was es denen eigentlich geht, was für Aufgabentypen die haben wollen. Weil das teilweise schon von dem auseinander geht mit dem was man bisher im Matheunterricht gemacht hat.“ (251, 32-35)

Wie der Typus der überzeugten Befürworterinnen und Befürworter fühlen sich auch die Repräsentanten dieses Typs durch Vergleichsarbeiten nicht kontrolliert. Selbst diejenigen, die der Meinung sind, dass es als Kontrollinstrument verwendet werden könnte, betonen ausdrücklich, dass sie sich persönlich nicht kontrolliert fühlen. Sie betrachten Vergleichsarbeiten als externe Bestätigung über die Leistungsfähigkeit ihrer Schülerinnen und Schüler. Sie sehen es teilweise auch als positive Rückmeldung, um Anhaltspunkte zu erhalten, wo man ansetzen könnte, um Verbesserungen vornehmen zu können, räumen jedoch sofort ein, dass dies aufgrund des Zeitpunktes nicht möglich ist.

> „…dass man vergleichbar ist, aber auch eine gewisse Kontrolle über die Lehrer . I: Fühlen Sie sich persönlich kontrolliert? L: Nein.“ (271, 29-29)

> „Nein, ich habe da überhaupt kein Problem damit (mit Kontrolle). Ich weiß, was meine Schüler können und was nicht und ich bekomme es dadurch eigentlich noch mal dokumentiert. Es ist für mich eigentlich auch irgendwo eine positive Rückmeldung, dass ich überprüfen kann, wo könnte ich ansetzen, wo muss ich noch was besser machen. Das Problem ist halt, wie gesagt, wann. (271, 30-30)

Desgleichen bewerten auch sie die Durchführung generell als positiv und unproblematisch. Nur wenige Realschullehrpersonen kritisieren im Fach Deutsch in der sechsten Klasse, dass die zur Verfügung stehende Zeit, die für die Bearbeitung der Aufgaben eingeplant war, den Schülerinnen und Schülern nicht ausreichte.

> „Die Vergleichsarbeiten lassen sich so durchführen, wie sie geplant wurden. Und auch vom zeitlichen Aufbau her, ist es eigentlich kein Problem.“ (357, 30-30)

> „Also, es war ein bisschen ein Problem mit der Zeit, da die Sechstklässler, das war sehr viel und da war die Zeit sehr knapp. Die mussten sehr viel Text selber schreiben, im Gegensatz zu Klasse acht, da war es wieder sehr wenig. Die waren viel früher fertig und die Sechser, die mussten wirklich durch schreiben, dass die Zeit gereicht hat. (…). Aber es war zu umfangreich. Die Schüler mussten freie Texte schreiben und die Sechstklässler brauchen einfach noch länger.“ (271, 35-37)

Die meisten Personen dieses Typs schauen die Rückmeldedaten an, setzen sich jedoch nicht so intensiv damit auseinander, wie es die überzeugten Befürworterinnen und Befürworter praktizieren. Teilweise haben die Lehrerinnen und Lehrer bereits bei der Eingabe der Schülerlösungen in das Excel-Datenblatt Hinweise auf Stärken und Schwächen erhalten und zeigen sich wenig überrascht über die Ergebnisse aus den Rückmeldungen. Manche geben auch an, aufgrund der übersichtlichen Darstellung relativ leicht Schwächen und Stärken ihrer Schülerinnen und Schüler ablesen zu können.

> „Es hat mich nicht wirklich überrascht, was daraus kam, muss ich ehrlich sagen. (…). Hat mich im Prinzip nicht überrascht was da raus kam, war im Prinzip das was ich erwartet habe.“ (272, 42-44)

> „Also, (…) es ist natürlich relativ leicht Schwächen zu erkennen, einfach auf Grund der übersichtlichen Auswertungsinstrumente, die bei dem Programm dabei sind.“ (357, 37-37)

Dies stellt einen Grund dar, warum sie sich nicht intensiv mit den Schaubilder und Grafiken auseinandersetzen. Als weitere Gründe für die oberflächliche Betrachtung der Daten stellt zum einen die Tatsache dar, dass sich aufgrund des späten Zeitpunkts keine Maßnahmen ableiten lassen. Zum anderen begründet eine Lehrperson ihr geringes Interesse damit, weil sie als Schule nicht aufgefordert wurden, die Ergebnisse an das Institut für Schulentwicklung zu melden. Somit wird kein Handlungsbedarf gesehen, sich mit den Rückmeldedaten vertiefend auseinander zu setzen.

> „Wobei ich habe das schon beim Korrigieren eigentlich gesehen, wo viele die Fehler gemacht haben. Also, ich habe das halt gesehen bei welcher Aufgabe die meisten Fehler gemacht haben, man sieht ja dann den Durchschnitt. Ich habe insofern nichts damit anfangen können, weil ich die Klasse abgegeben habe, nach Klasse sechs. Der Kollege, die Kollegin, die jetzt die Klasse hat könnte wahrscheinlich eher was damit anfangen. (…). Da habe ich nicht hin geschaut. Wir mussten dieses Jahr nicht rein, also wir mussten nicht einreichen. Das habe ich nicht geschaut. Ich habe geschaut, wo befindet sich die Klasse auf welchem Stand.“ (271, 38-43)

Vereinzelt geben Lehrpersonen an, dass sie nur im Falle von starken Abweichungen der Schülerleistungen die Ergebnisse näher betrachten und gegebenenfalls mit den betreffenden Schülerinnen und Schülern besprechen. Sie würden daraus jedenfalls kaum Änderungen für ihr Handeln ableiten und bestätigen dadurch, dass die Ergebnisse aus Vergleichsarbeiten keine Auswirkungen auf die Zukunft ihrer Unterrichtsgestaltung haben. Sie vertreten die Position, dass wenn man sich bei der Planung und Gestaltung des Unterrichts an die Zielsetzungen des neuen Bildungsplans hält, den erforderlichen Anforderungen auf jeden Fall gerecht wird. Außerdem ist es in ihren Augen selbstverständlich, dass jede Unterrichtsstunde reflektiert und gegebenenfalls optimiert wird.

> „Also ich habe es hauptsächlich mal für mich angeschaut. Da sieht man dann die Vergleichswerte. Das habe ich den Schülern auch letztes Jahr gezeigt - wo sie stehen, wo die Vergleichsgruppe steht. Das habe ich gezeigt. Zwar nicht zu jeder Aufgabe, aber grad, wo sie stark abgewichen sind, habe ich es gezeigt und besprochen.“ (270, 26-26)

> „Da sind natürlich die Abweichungen besonders interessant, weicht das nach oben oder unten ab.“ (357, 34-34)

Trotz dieser etwas zurückhaltenden Auseinandersetzung mit den Rückmeldedaten, erkennen die Lehrerinnen und Lehrer dieser Gruppe in Vergleichsarbeiten ein positives Nutzungspotential, räumen aber ein, dass der bisherige Entwicklungsstand und die bisherige Verwendungspraxis noch verbesserungsfähig sind. Damit dieses Potential genutzt werden kann, nennen sie teilweise Verbesserungsvorschläge, die sich auf verschiedene Punkte beziehen. Die meisten Personen dieser Gruppe kritisieren, wie auch die erste Gruppe, den Zeitpunkt der Vergleichsarbeiten, der im Jahr 2007 noch am Ende des Schuljahres stattfand. Aufgrund der kurzen Zeitspanne, die zwischen der Durchführung der Vergleichsarbeiten und dem Ferienbeginn bleibt, können keine Maßnahmen mehr durchgeführt werden, um gezielt auf die Befunde der Auswertung reagieren zu können.

> „Es ist für mich eigentlich auch irgendwo eine positive Rückmeldung, dass ich überprüfen kann, wo könnte ich ansetzen, wo muss ich noch was besser machen. Das Problem ist halt, wie gesagt, wann. (…) Vergleichsarbeiten finde ich prinzipiell eine gute Sache, denn man muss schon gewisse Standards erreichen. Und dass nicht jeder immer in seinem eigenen Dunst umher wurstelt, das ist ganz wichtig.“ (271, 30-30 und 123-123)

> „Das Problem ist, bei uns passiert nichts, weil wir ja die Zeit gar nicht haben. Und das ist der ganze Knackpunkt an der Geschichte. Also, die Idee ist gar nicht schlecht, dass man solche Schwächen, Probleme und so was einfach über so eine Arbeit feststellt, aber der zweite Teil fehlt.“ (272, 28-28)

Des Weiteren schließen sich auch deshalb mögliche unterstützende Maßnahmen für die Schülerinnen und Schüler der Klassen aus, weil Vergleichsarbeiten 2007 am Ende der sechsten Klassen durchgeführt werden und danach in den meisten Schulen ein Lehrerwechsel stattfindet. Die Lehrerinnen und Lehrer haben somit keine Gelegenheit, die gewonnenen Erkenntnisse in der Klasse umzusetzen, da sie die Klasse abgeben und diese im darauf folgenden Schuljahr in den meisten Fällen nicht mehr unterrichten. Ein Realschullehrer schlägt vor, dass die aufnehmende Lehrperson die Vergleichsarbeiten selbst durchführt und dadurch Hinweise über Stärken und Schwächen der einzelnen Schülerinnen und Schüler erfährt und ihm genügend Zeit bleibt, daraus geeignete Maßnahmen abzuleiten. Eine andere räumt diese Möglichkeit auch ein, dass sie die Erkenntnisse, die sie aus den Rückmeldungen bezüglich der individuellen Leistungen ihrer Klasse gewonnen hat, weiterleiten könnte. Allerdings bewertet sie dieses Vorgehen als ‚witzlos’.

> „Ist dann ja so, im Prinzip, müssten ja dann, müsste der aufnehmende Lehrer immer diese Vergleichsarbeit und die Ergebnisse machen, damit er dann, könnte man ja machen, damit er dann weiß, wo sind die Stärken und die Schwächen von der Klasse.“ (251, 75-76)

> „Also ich habe es mir eher individuell angeguckt, weil ich ja gucken muss welche Schwäche hat mein Schüler und wo. Wobei ich gleich sagen muss, dass ich es totalen Blödsinn finde, dass man es Ende Klasse 6 und Klasse 8 schreibt, weil ja dann der Lehrer wechselt. Also ich kann jetzt schon meine Erkenntnisse den Lehrern weitergeben, aber ich finde es ein bisschen witzlos (LACHEN). Na, ja.“ (264, 36-37)

> „Man müsste es an einen Kollegen übergeben, der dann die Klasse weiterführt, wenn man nicht die Chance hat das selber noch zu machen. Das hatte ich in Klasse, also bei der ersten Vergleichsarbeit. Da habe ich dann auch gezielt darauf geachtet eigentlich.“ (357, 37-37)

Ein weiterer Punkt der von diesem Typus kritisiert wird betrifft generell die Bewertung von Vergleichsarbeiten als weitere Klassenarbeit. Sie sind der Meinung, dass Klassenarbeiten immer einen gewissen Notendruck auf die Schülerinnen und Schüler ausüben. Wenn aber Vergleichsarbeiten Hinweise auf Stärken und Schwächen in einem förderdiagnostischen Sinn aufzeigen sollen, müssen sie ihrer Meinung nach nicht bewertet werden, vor allem dann, wenn das Niveau der Vergleichsarbeiten so niedrig ist, dass die Schülerinnen und Schüler keine Gefahr laufen, sich derart zu verschlechtern, dass es Auswirkungen auf eine Nichtversetzung haben könnte.

> „Ja meine Frage ist, müssen die Vergleichsarbeiten unbedingt als Klassenarbeit gelten, ja? (…) Warum müssen die bewertet, warum müssen die überhaupt bewertet werden. Warum kann man sie nicht schreiben ohne den Druck der Noten im Hintergrund. So quasi, sie sind zwar nicht schwer, aber trotzdem sie gelten, als fünfte oder sechste Klassenarbeit und das zählt, egal wie sie rauskommt. Die wenigsten Schüler, ich gehe jetzt vom Fach Mathe aus, sowohl in der 6. als auch in der 8. Klasse durch die Vergleichsarbeit dermaßen reingeritten, dass sie haben müssen um die Versetzung kämpfen. Das Gegenteil ist der Fall. Ich habe in der 6.Klasse den besten Durchschnitt gehabt in der Vergleichsarbeit im Vergleich zu den anderen Mathearbeiten.“ (268, 44-44)

Das Problem, dass die Vergleichsarbeiten im Niveau insgesamt viel zu einfach waren teilen beinahe alle Hauptschullehrpersonen und etwa die Hälfte der Realschullehrerinnen und -lehrer, auch über die einzelnen Gruppen hinweg.

> „Teilweise ja, wobei ich mir das dann auch so erklärt habe, da waren auch Sachen auch dabei, also von zwei Schuljahren und die waren doch teilweise recht einfach. Also ich denke schon, dass wir da teilweise ein doch anderes Niveau haben.“ (267,54-54)

> „Jetzt kann man sagen, ich haben über dem Standard unterrichtet, ja? Dann war der Stoff von meinen ganzen normalen Klassenarbeiten war dann zu schwer, weil wenn für mich die Richtschnur die Vergleichsarbeit sein soll… ich sage aber die Vergleichsarbeiten waren zu leicht. Da bin ich bestimmt auch wieder nicht der erste Lehrer. Ich sage einfach, das war zu leicht und entspricht gar nicht dem Niveau einer 6. Klasse in Realschule in Mathe.“ (268, 46-46)

Die verschiedenen Kritikpunkte, die diese Gruppe der Lehrerinnen und Lehrer an Vergleicharbeiten auszusetzen haben, hält sie trotz des bestehenden Interesses an Vergleichsarbeiten letztlich davon ab, Maßnahmen zur Förderung ihrer Schülerinnen und Schüler oder zur Verbesserung des Unterrichts zu ergreifen. Neben dem geringen Niveau begründen sie dies auch mit dem Zeitpunkt der Vergleichs-

arbeiten und dem anstehenden Lehrerwechsel. Außerdem beklagen sie die mangelnde Unterstützung von Seiten der Bildungsadministration und -politik. In dem folgenden Zitat eines Realschullehrers kommt die ganze Problematik nochmals zum Tragen:

> „Das Interesse war schon da, aber es ist ein kleines Problem dadurch, dass die so spät geschrieben werden, im Juni. Da sage ich bestimmt was, was sie nicht das erste Mal hören. Es bleibt ja die Zeit gar nicht das richtig auszuwerten. Ich kriege zwar eine Rückmeldung in dem Sinn, dass ich es schön im Balkendiagramm sehe, ob ich besser oder schlechter bin wie der Landesdurchschnitt, ob meine Klasse gut oder schlechter abgeschnitten hat. Dann werden mir viele Prozentzahlen nur so um die Ohren gehauen, wie hier. Und sehe dann immer anhand der Prozentzahlen ob ich besser oder schlechter bin. Aber bis ich das erfahre ist Mitte Juni und dann bleiben mir drei Wochen ungefähr, theoretisch rein drei Wochen, um jetzt das zu evaluieren. Sprich, um aus dem irgendwelche Vorteile zu ziehen und sagen: „Das hat meine Klasse noch nicht richtig behandelt; Das habe ich zu wenig behandelt; Das haben sie nicht verstanden“, und so weiter. Und jetzt soll mir die Landesregierung bitte klar machen wie das jetzt umgesetzt werden kann. Denn dann fangen die großen Ferien an und im nächsten Jahr geht es wieder ganz normal mit dem neuen Schuljahr weiter. Das heißt wer soll bitte in dieser kurzen Zeit noch mal das nachholen oder vertiefen, wenn irgendwelche Standards nicht da sind?“ (268, 34-34)

> „Das Ganze soll doch einen Wert haben, wenn ich die Möglichkeit habe dann zu sagen: „Aha, das können sie noch nicht gut, das muss man eigentlich üben.“ Bloß für das haben wir ja hier keine Zeit mehr.“ (ebd., 37-37)

> „Ich glaube, dass es organisatorisch gar nicht machbar ist, wenn man es nicht innerhalb von dem Schuljahr macht. Das würde wieder bedeuten, man müsste die Dinger wieder schreiben. Weil einfach von der Organisation her ein Lehrerwechsel ist, dann ein anderer Lehrer. Dann ist auch nicht gesagt, dass der der eine 6.Klasse hat, dann wieder im nächsten Jahr die 6.Klasse hat.“ (ebd., 70-70)

In der Gesamtheit betrachtet äußern die „positiven Mitspielerinnen und Mitspieler“ ihre Bereitschaft, Vergleichsarbeiten in ihren Klassen durchzuführen. Sie sehen durchaus Entwicklungspotential darin, das in Zukunft effektiver genutzt werden könnte, um die Qualität von Schule und Unterricht zu verbessern. Zum derzeitigen Zeitpunkt äußern sie jedoch noch Bedenken, weil sie der Meinung sind, dass die Konzeption dieses Evaluationsinstruments noch nicht ausreichend entwickelt und ausgereift ist, also sozusagen noch in den Kinderschuhen steckt.

Sie sind zuversichtlich, dass sich diese Defizite in den folgenden Jahren beheben lassen und den Schulen mit diesem Testverfahren ein effektives Instrument zur Verfügung steht.

> „Auf der anderen Seite sind die Vergleichsarbeiten natürlich erst in den Anfängen und die Struktur dieser Vergleichsarbeiten, bzw. die Aufgabenstellungen die sind manchmal etwas schwierig, die sind weit her geholt.“ (271, 123-123)

Die Lehrerinnen und Lehrer dieser Gruppe sehen generell in Wiederholungen bestimmter Themenbereiche, bei denen ihre Schülerinnen und Schüler schlechter abschnitten, eine Möglichkeit, wie Leistungen in Vergleichsarbeiten verbessert werden könnten. Ein Gymnasiallehrer schlägt vor, ganz gezielt auf Vergleichsarbeiten zu trainieren, im Sinne des „teaching to the test“.

> „Nachdem ich jetzt für mich diese Vergleichsarbeit analysiert habe, könnte ich mir vorstellen, dass ich gezielt dafür trainiere. Aber das ist dann für mich kein Training für Mathematikunterricht im Allgemeinen. Sondern für den Test. Das gibt es ja für alles, bei Medizinertest gibt es auch ein Spezialtraining. Das macht die Schüler nicht wirklich besser, würde aber für das eine Ereignis einen Vorteil bringen. Finde ich aber nicht unbedingt richtig als Zielsetzung. Das ist ja auch nur einen Klassenarbeit. Da würde ich es in Kauf nehmen, dass man ein spezielles Lernziel hat, eine bestimmte Fragestellung, die ich auch mit rein nehme. Aber ich kann jetzt nicht den ganzen Unterricht darauf ausrichten. Ich kann auch einigermaßen sicher sein, dass es ja auch weiter entwickelt wird.“ (362, 71-71)

Im Vergleich zu der Gruppe der überzeugten Befürworterinnen und Befürworter führen diese Lehrerinnen und Lehrer in der Summe weniger „Systematische Diskussionen“ sondern viel eher einen „Informellen Austausch“ über die Ergebnisse, die ihre Schülerinnen und Schüler in Vergleichsarbeiten erzielten. Die meisten Personen besprechen die Ergebnisse intensiv im Unterricht mit der Klasse. Teilweise finden auch mit den Parallelkollegen oder mit der Schulleitung Gespräche statt. Sie erklären sich generell zum Austausch bereit, wenn die Kolleginnen und Kollegen den ersten Schritt tun und ihr Interesse bekunden.

> „Wenn das jemand will, dann kann er es haben. Gerade die übernehmenden Lehrer, das wäre ja was. Die könnten sagen, he ja.“ (251, 126-127)

> „Und dann habe ich mit den Schülern die Vergleichsarbeit sehr genau besprochen und habe dieses Auswertungsraster mitgegeben. Ja. Für was sie ein Item bekommen, für was nicht. Damit für Eltern transparent ist, warum gibt es hier einen Abzug, warum hier nicht. (…) Das finde ich, muss transparent gemacht werden. Und das habe ich mitgegeben, habe ich mit den Schülern besprochen, habe gesagt welche Lösungen möglich gewesen wären. Also, da

> habe ich ziemlich viel Zeit investiert. Und über die Schüler habe ich das dann den Eltern [mitgeteilt].“ (271, 71-71)

> „Wir haben die [Ergebnisse] im zentralen Computer. Wir, die parallel unterrichten, wir haben uns ausgetauscht, wir haben gesagt, wie ist es bei dir, wie sieht bei dir die Kurve aus. Wir haben uns das gemeinsam angeschaut, aber andere Kollegen haben da eigentlich wenig Interesse dran. Immer die, die es halt betrifft. Da ist nichts festgelegt.“ (271, 88-88)

Sie erwähnen, dass mit den Schulleitungen aufgrund einer notorischen Arbeitsüberlast und wegen Zeitmangels ein eher informeller Austausch geführt wird. Durch das gute Abschneiden der Schülerinnen und Schüler haben auch die Eltern kein gesteigertes Interesse an Gesprächen. Diese suchen immer erst dann ein Eltern-Lehrer-Gespräch, wenn die erzielten Resultate ihrer Kinder nicht deren Erwartungen entsprechen und schlecht ausfallen.

> „Aber, die Schulleitung hat immer wieder so viel zu tun. Ich stehe mit meiner Rektorin im relativ guten Kontakt, man tauscht sich aus, man fragt mal nach und fragt wie war es und das war es dann. Oder wie fanden Sie die Vergleichsarbeit und dann sagt man ja, das fand ich gut und das weniger. Aber das eher so eine informative Sache.“ (271, 104-104)

> „[Von den Eltern] wollte auch niemand was wissen. Eltern kommen grundsätzlich nur, wenn es extrem negativ ist. Bei positiven Erscheinungen wird man nie einen Elternteil sehen. Das ist selbstverständlich.“ (272, 101-102)

Eine Realschullehrperson würde sich einen intensiveren Austausch über die Ergebnisse wünschen. Sie sieht die Schulleitung in der Verantwortung, konkrete Maßnahmen zu initiieren und dafür Sorge zu Tragen, dass diese dann von den Lehrerinnen und Lehrern verbindlich durchgeführt werden müssten. Bleibt diese Anordnung und Aufforderung aus, so wird sich ihrer Meinung nach an der Schule nichts ändern.

> „Bei uns ist es halt so, dass ich sage, solange unser Rektor sich nicht rührt und uns in den Hintern tritt dann machen wir es erst mal so wie bisher.“ (251, 161-162)

5.7.3 Der desinteressierte, gelassene Typ

Dieser Typus von Lehrerinnen und Lehrer stehen den Vergleichsarbeiten von Anfang an relativ gelassen gegenüber. Auf die Frage, warum dieses neue Instrument eingeführt wurde, und wie sie es den Eltern oder Schülern erklären, antworten die meisten, dass es nichts zu erklären gäbe. Sie lassen sich durch dieses neue Evaluationsinstrument nicht aus der Ruhe bringen und führen diese in ihren Klassen durch, weil Vergleichsarbeiten eben verpflichtend sind und sie dieser Anordnung zunächst formal erfüllen. Sie distanzieren sich so weit wie möglich davon, setzen sich mit der Zielsetzung und Funktion von Vergleichsarbeiten nicht auseinander, sondern sehen es lediglich als verpflichtende Vorgabe an, die sie ausführen müssen.

> „Aber was gibt es da groß zu erklären? Es muss gemacht werden, weil es vorgeschrieben ist.“ (269, 5-5)

> „Also, da möchte ich sagen, da haben ich gar nicht viel erklärt, sondern das haben wir als gegeben erst mal hin genommen. Wir mussten es machen.“ (250, 35-35)

> „Aber, für uns war wichtig dass die geschrieben ist, dass der Form genüge getan wird, dass das irgendwie noch in die Note mit eingerechnet wird und für den Rest haben wir gar keine Zeit gehabt.“ (270, 83-83)

Warum Vergleicharbeiten verpflichtend durchgeführt werden begründen sie zum einen damit, dass dadurch auch diejenigen Kolleginnen und Kollegen gezwungen sind, sich am neuen Bildungsplan orientieren, die es bisher noch nicht gemacht haben. Dadurch wird auf diese Lehrerinnen und Lehrer ein Stück weit Druck ausgeübt gewissenhafter zu arbeiten, um ein völlig freies Schaffen zu unterbinden.

> „Ich denke, dass sich jetzt jeder Lehrer mehr an den Bildungsplan hält - weil es gibt doch so ein paar Kollegen, von denen man früher gehört hat, dass sie das durchgezogen haben, wovon sie der Meinung waren. Dadurch dass das jetzt abgefragt wird, ist jeder gezwungen, das durchzuführen.“ (270, 2-2)

> „Das würde mich auch interessieren (lacht). Natürlich möchte man, ich denke auch dass man die Lehrer dazu zwingen möchte einen bestimmten Standard zu erreichen, oder vielleicht auch gewissenhafter zu arbeiten. (…) Ich denke auch, dass es ein gewisses Druckmittel ist, um den Lehrern nicht ein völlig freies Schaffen zu ermöglichen.“ (250, 31-31)

In dieser Gruppe fällt auf, dass sie sich im Bereich Druck und Kontrolle durch Vergleichsarbeiten nicht auf sich selbst beziehen, sondern auf einige ihre Kolleginnen und Kollegen. Sie selbst sind mit ihrer eigenen Arbeitsweise zufrieden und von ihrem Tun überzeugt. Sie brauchen keine Vergleichsarbeiten, an denen sie sich orientieren und empfinden diese aus diesem Grund auch als unangenehm und ein Stück weit auch unnötig.

> „Also, prinzipiell finde ich das nicht schlecht, den Lehrer auch zu kontrollieren oder ihn zu zwingen, dass er seinen Unterricht ordentlich macht. Aber, ich denke schon, dass ich meinen Unterricht ordentlich mache, und dass ich mir Mühe gebe und vielseitig den Unterricht gestalte. Und deshalb ist so was für mich dann wirklich nur unangenehm.“ (352, 171-171)

Zum anderen nennen sie die Vergleichbarkeit der Schülerleistungen als weiteren Grund für die Einführung von Vergleichsarbeiten, dem sie jedoch generell gelassen begegnen.

> „Das soll einfach nur ein Vergleich sein, die sollen da jetzt nicht groß nervös werden.“ (370, 22-22)

Allerdings wird die Art und Weise des Vergleichs und dessen Aussagekraft von Anfang an in Frage gestellt. Denn wenn man nur den Notendurchschnitt der Vergleichsarbeit als Vergleichsmaß heranzieht, so liefert dieser keine Informationen über die tatsächlichen Ursachen, die für die Leistungsunterschiede verantwortlich sein könnten. Ihrer Meinung nach müssten auch Hintergrundvariablen erhoben werden, die neben den schulischen Bedingungen insbesondere den familiären und soziokulturellen Kontext der Schülerinnen und Schüler erheben. Und nachdem diese Faktoren in Vergleichsarbeiten bisher nicht erfasst werden, haben diese für diese Gruppe auch keine große Bedeutung.

> „… und einfach damit man einen Vergleich hat, wobei ich finde, dass man nicht nur die Notendurchschnitte vergleichen kann. (…).Man braucht auch den Zeugnisschnitt - ist es an sich eine schwache Klasse? (…) Dann hatte ich z. B. eine Klasse, die hatte in Klasse 5 ein halbes Jahr lang gar keine Mathematik - so etwas muss aufgearbeitet werden. (…) Und dann eben auch: Was ist es für ein Hintergrund von Zuhause? Helfen die Eltern mit? Stehen die Eltern dahinter? Ist es den Eltern egal? Sind die Eltern überhaupt nicht zu Hause und kümmern sich nicht darum? Ich denke, dass das alles mit dazu zählt. Dann kommt es auch noch auf den Einzugsbereich an. Wo kommen die Schüler her? Das sind alles Bereiche, die mit reinspielen.“ (270, 5-8)

Ein Realschullehrer begründet seine gelassene Haltung damit, dass er sich kurz vor dem Ruhestand befindet und auch ohne Vergleicharbeiten weiß, welche Bedürfnisse seine Schülerinnen und Schüler haben. Sein Handeln ist eher paidotrop ausgerichtet und er versucht, das Beste zu erreichen. Er weiß von vorne herein, dass er die staatlichen Vorgaben in Form des Lehr- bzw. Bildungsplans aus zeitlichen Gründen nicht umsetzen kann und auch nicht gewillt ist, weil es seiner Meinung nach den Kindern nichts bringt.

> „Wenn Sie meine Position ansehen und mein Alter, dann fühle ich mich nirgendwo mehr unter Druck gesetzt. Mir ist das ... Ich denke dass ich so viel Erfahrung habe und auch meinen Job so gerne mache, dass ich weiß, was die Schülerinnen und Schüler eventuell brauchen. Und meine Vorstellung ist die, oder ich richte mich eigentlich nur nach den Kindern, was ich vorgesetzt bekomme und versuche da das Beste daraus zu machen. Weil wenn ich den Lehrplan anschaue, denn bekomme ich sowieso nicht durch. Das nützt mir nichts, wenn ich den Lehrplan durch gemacht habe und die Kinder dann irgendwo völlig nebenher schweben und nichts davon mitbekommen haben.“ (250, 33-33)

Die technische Durchführung der Vergleichsarbeit wird auch von diesen Lehrerinnen und Lehrern nicht kritisiert.

> „Aber ansonsten war es eigentlich kein Problem bei der Durchführung.“ (266, 22-22)

> „Bei der Durchführung. Da gab es keine Schwierigkeiten.“ (355, 40-40)

Jedoch hält sich diese Gruppe nicht zwingend an die Einhaltung aller Testgütekriterien, indem sie die „Durchführungsobjektivität“ verletzt. Aufgrund der zugesicherten Anonymität gestehen diese Personen ein, dass sie u. a. mit der zur Verfügung stehenden Zeit ganz individuell umgehen, indem sie zum Beispiel ihre Klasse früher beginnen lassen, so dass jener mehr Zeit zur Bearbeitung bleibt. Sie begründen ihr Handeln damit, dass es im Ermessen der Lehrperson liegen sollte, ob sie ihren Schülerinnen und Schülern einen zeitlichen Spielraum von 15 min mehr oder weniger Zeit zur Verfügung zu stellt oder nicht.

> „Zum Glück ist es anonym. Wir hatten mehr Zeit, als wir gebraucht haben. Es hieß, man muss offiziell anfangen, dann und dann Schluss machen. Aber ich habe nicht eingesehen, dass ich noch zehn Minuten da sitzen soll, weil keine Fragen mehr kamen. Und sagen, so Kinder wir warten jetzt, in zehn Minuten ist offizieller Beginn. Ich habe die also zehn Minuten vorher anfangen lassen. Trotzdem sind nicht alle fertig geworden, aber es waren wenige, die nicht fertig geworden sind. Ich finde, man sollte die Zeit nicht so streng wie bei Prüfungsarbeiten, das sollte man einfach nicht so vorgeben. Das ist

> nicht gut. (…) Gut, man muss einen Zeitrahmen geben, aber da sollte doch so in etwa mindestens eine viertel Stunde im Ermessen des Lehrers liegen. Ob man da jetzt sagt, ich lasse die jetzt noch eine Viertelstunde, weil da gab es Probleme. Es kann ja immer mal was sein, einer muss mal schnell raus, weil dies und dies ist oder so. (360, 37-39)

Des Weiteren halten sie sich auch während der Bearbeitungszeit nicht an die staatlichen Vorgaben, die die Objektivität des Testverfahrens sichern sollen. Sie unterstützen ihre Schülerinnen und Schüler bei Problemen, indem sie ihnen Hilfestellungen geben und Lösungsvorschläge an die Tafel schreiben.

> „Obwohl ich muss sagen, ich habe denen auch zum Teil Hilfen gegeben, weil bestimmte Sachen für die unklar waren. Gerade bei diesem Test. (…). Dann habe ich einfach etwas vorgegeben, gezeigt an der Tafel, wie sie es machen sollten, sonst wären die total verwirrt gewesen.“ (360, 41-43)

Außerdem berichtet eine Person, dass ein Schüler angeblich die Arbeit schon im Vorfeld der Vergleichsarbeit von einem Cousin bekommen hatte. Wie weit diese Behauptung des Schülers der Wahrheit entsprach wurde von der entsprechenden Lehrperson nicht überprüft. Eine Möglichkeit der vorzeitigen Aushändigung besteht durchaus. Seit der verpflichtenden Einführung von Vergleichsarbeiten in Baden-Württemberg kam es immer wieder zu Vorfällen, dass Aufgaben aus Vergleichsarbeiten vor der Durchführung an die Öffentlichkeit gelangten. Einzelne Fälle wurden durch die Presse bekannt. So kam es in einem Fall dazu, dass die Vergleichsarbeit in einer Realschule auf dem Kopierer vergessen wurde und das Ministerium als Folge die Vergleichsarbeit in diesem Fach im ganzen Land aussetzte. In einem anderen Fall wurden Aufgaben aus Vergleichsarbeiten im Internet veröffentlicht und ebenfalls abgesagt.

> „Ein Schüler von mir, deswegen weiß ich es, kam auf mich zu und hat ganz freudestrahlend berichtet, dass er von seinem Cousin jetzt Aufgaben hat und die sind so, wie dann da überprüft wird. Es war klar, er konnte das zu dem Zeitpunkt eigentlich noch nicht wissen.“ (270, 23-23)

Bei der Auswertung und Eingabe der Lösungen in die Excel-Tabelle hält sich die Gruppe an die Vorgaben und führt es pragmatisch durch. Eine Lehrperson im Gymnasium berichtet, dass sie die Ergebnisse aus Vergleichsarbeit nicht auswerten musste und sich daran auch hielt.

> „Sagen wir mal so, ich habe mir bei der Auswertung an die Vorgaben gehalten. Ich war froh, dass ich fertig geworden bin. Und habe einfach die Aufgaben einfach anhand vom Lösungsblatt durch genommen und habe es bewertet. Das war es eigentlich auch.“ (250, 41-41)

> „Die [Graphen und Tabellen aus den Rückmeldungen] haben wir bekommen und wussten, dass wir das nicht auswerten müssen. Und dann haben wir das auch nicht gemacht.“ (352, 47-47)

Eine Realschullehrperson zweifelt an der Objektivität der Auswertung. Als Grund führt sie die unklaren Lösungsvorschläge an, die den Lehrerinnen und Lehrern einen relativ großen Beurteilungsspielraum eröffnen. Im Austausch mit ihren Kolleginnen und Kollegen kommt sie zu dem Schluss, dass durch den Interpretationsspielraum, das Gütekriterium der Objektivität verletzt wird und die Vergleiche somit ihre Aussagekraft verlieren.

> „Ich habe da gewisse Probleme muss ich sagen bei der Korrektur von den Vergleichsarbeiten, in der Hinsicht, weil die Lösungsvorschläge, die dabei waren, meiner Meinung nach teilweise bisschen schwammig waren. Das heißt, wir haben schon im Kollegium schon darüber diskutiert und zwar ging es einfach darum, dass, wenn da eine Lösung dran steht oder ähnliches beispielsweise, dann weiß ich nicht genau, kann ich das Andere durchgehen lassen? Ja oder nein? Und dann frage ich mich, in wieweit eine Vergleichsarbeit vergleichbar ist. Wenn da also ähnliches steht, dann lasse ich beispielweise eine andere Lösung auch durchgehen, die eventuell ein Lehrer in Heidenheim nicht mehr durchgehen lässt. Und dann stellt sich die Frage in wieweit ist es vergleichbar.“ (266, 12-12)

Die Gruppe unterscheidet sich von den anderen Gruppen bezüglich der Auseinandersetzung mit den Rückmeldedaten. Sie betonen, dass sie sich damit nicht beschäftigt hatten. Als Begründungen geben sie an, dass sie sich entweder dafür nicht interessierten, dass die Zeit dazu nicht zur Verfügung stand oder dass grundsätzlich die Struktur und Anlage der Vergleichsarbeit keine wirklichen Unterschiede herausstellt, die die Grundlage für weitere Analysen bietet.

> „Dazu habe ich gar keine Zeit. Man könnte das eventuell hinterher machen, wenn alle drei, wir haben drei Parallelklassen, wenn alle drei das in Computer eingegeben haben, dann könnte man vielleicht schauen. Aber ich denke, dass die Struktur der Arbeit, wenn man die vergleicht, gibt es keine großen Unterschiede.“ (250, 44-44)

Eine Lehrperson versteht die Frage nach der Verständlichkeit der Rückmeldungen nicht. Hier liegt die Vermutung nahe, dass sie sich mit der Konzeption der Vergleichsarbeit und deren Verwendung kaum auseinandergesetzt hat.

> I: „Bei der Rückmeldung gibt es ja bestimmte Grafiken und Tabellen. Wie intensiv schauen Sie diese an und wie verständlich ist es? (Pause) Oder benutzen Sie das gar nicht?“
> L.: „Was ...?“ [*Lehrperson weiß nicht, worum es geht?*]
> I.: „Das ist eine Excel-Tabelle, da kann man die Fehler von den Schülern eingeben.“ (270, 23-25)

Nach Meinung dieser Gruppe liefern die Vergleichsarbeiten keine neuen Anhaltspunkte über Schwächen und Stärken ihrer Schülerinnen und Schüler und bilden somit keinen Ausgangspunkt, um daraus Änderungen für ihre Unterrichtspraxis abzuleiten. Auch in diesem Bereich berichten die Lehrerinnen und Lehrer eher von ihren Kolleginnen und Kollegen, als von ihrem eigenen Handeln.

> „Ich weiß es von einer Kollegin, die nun vermehrt parallel Übungsaufgaben herausgibt - dass wochenmäßig immer Aufgaben von Klasse 5, wenn sie die VA in Klasse 6 schreiben, bearbeitet werden. Dass die einfach drin bleiben im Stoff. Das habe ich von einer Kollegin gehört, die das jetzt neu macht. Früher nicht so ausgiebig und heute macht sie es wirklich regelmäßig. Das ist aber das Einzige, was ich gehört habe.“ (270, 93-93)

In dieser Gruppe werden keine systematischen Diskussionen über Vergleichsarbeiten geführt. Einige berichten von wenigen informellen Gesprächen.

5.7.4 Der negativ kritische Typ

Die Personen, die der Gruppe der Kritikerinnen und Kritiker zugewiesen werden können, bekunden im Gespräch ihre kritische Einstellung zu Vergleichsarbeiten. „Also ich bin halt ein großer Kritiker dieser Arbeit. (…) Vielleicht zähle ich zu den schärfsten Kritikern sogar“ (350, 31-31, 35-35). Im Unterschied zu Gruppe 3 setzen sich diese Personen intensiv und kritisch mit zahlreichen Aspekten der Vergleichsarbeit auseinander. In Folge dieser Auseinandersetzung entwickeln sie eine kritische Haltung zu der derzeitigen Durchführungspraxis. Die Kritik reicht von Misstrauen gegenüber Vergleichsarbeiten bis zur Ablehnung, weil es sich ihrer Meinung nach um ein sinnloses, unnötiges und überflüssiges Instrument handelt. Ein Gymnasiallehrer betont, dass „es (ihm) ehrlich gesagt sowieso ein Rätsel“ ist, warum es Vergleichsarbeiten gibt (358, 31-31).

> „Obwohl, (…) es ist kein dummes Instrument, aber so wie es gehandhabt wird und wie es gekommen ist, habe ich großes Misstrauen.“ (365, 16-16)
>
> „Im Vorhinein hat man denen (den Eltern) das gesagt, was da an Unsinn oder, Entschuldigung, Sinn oder Unsinn auf sie zukommt.“ (350, 95-95)
>
> „Also, ich finde es so unnötig wie ein Kropf, weil diese DVAs sind im Grunde genommen nichts.“ (351, 110-110)

Diese Lehrerinnen und Lehrer erleben Vergleichsarbeiten als verpflichtendes Instrument, gegen das sie sich nicht wehren können. Es wird vom Ministerium vorgegeben, somit sind sie angehalten, diese bildungspolitischen Vorgaben umzusetzen.

> „Ich habe den Eltern auch gesagt, das wäre von oben halt vorgegeben.“ (257, 62 -62)
>
> „Also, zunächst mal natürlich, weil es Pflicht ist.“ (258, 58-58)
>
> „Ja, weil das Ministerium das eben will. Verstehen Sie, das kommt ja, ich kann mich nicht dagegen wehren.“ (262, 16 – 16)

Fragt man diese Gruppe nach den Gründen, warum das Ministerium Vergleichsarbeiten durchführt, so betrachten es die meisten Personen als Folge der schlechten PISA-Ergebnisse. Ihrer Meinung nach stellen Vergleichsarbeiten den Versuch des Ministeriums dar, der Öffentlichkeit eine aktive Schulpädagogik zu verkau-

fen. Sie sehen es teilweise als puren Aktionismus an, der Ausdruck ist für ein unüberlegtes, ergebnisloses Handeln der Bildungspolitik.

> „Verkauft wird es als eine Maßnahme gegen Pisa und eine Maßnahme im bundesweiten Vergleich und eine Maßnahme zur optimalen Förderung von schwachen Schülern.“ (253, 37-37)

> „Aber das kam für mich eben auch durch die PISA-Studie und alle möglichen anderen Studien - als da die großen Lücken aufkamen und dass da eben gedacht wurde, da muss man etwas tun und die Vergleichbarkeit ist dann ja auch ganz wichtig, damit man sieht, wo sind die Lücken groß und klein.“ (256, 40-40)

> „Meiner Meinung nach geht’s darum, der Öffentlichkeit eine aktive Schulpolitik zu verkaufen. Also ich denk man ist etwas unter Druck, weil die Pisastudie, ... und dergleichen. Wobei meiner Meinung nach nie sauber aufgearbeitet worden ist, wo jetzt das Abschneiden schlecht war und wo gut. Und was Pisa überhaupt gemessen hat. Das hat eigentlich kaum jemanden interessiert.“ (365 14-14)

> „Okay, ich sehe das als Reaktion auf die Pisastudie. Ja, das ist jetzt auf jeden Fall etwas, wo sie Aktionismus beweisen (lacht).“ (369, 7-7)

Die Hauptfunktion von Vergleichsarbeiten sehen sie nicht im Vergleichen von Schülerleistungen oder in der Förderung. Eine Realschullehrerin stellt klar heraus, dass ihrer Meinung nach „die Förderung der Schülerinnen und Schüler nicht im Vordergrund steht“ (253, 39-39). Sie äußern grundsätzlich Kritik am Vergleichen von Schülerinnen und Schülern, die sich in ihren Augen Menschen und Schulen aufgrund ihrer Unterschiedlichkeit und Vielfältigkeit nicht zum Vergleich anbieten.

> „Die möchten langfristig vielleicht einen Vergleich, das heißt an der Schule ist es so, an der Schule ist es anders, an einer anderen ist es wieder anders. Aber meiner Meinung nach sind das Kinder, das sind junge Menschen und die funktionieren nicht wie Maschinen. In Firmen kann man so was machen, aber in Schulen mit Kindern ist es mit einem Fragezeichen versehen die ganze Geschichte. (…). Kritisch.“ (262, 16-16)

Im Vordergrund sollen Vergleichsarbeiten dazu verwendet werden, um Lehrerinnen und Lehrer zu kontrollieren, auch wenn dieses Ziel von der Bildungspolitik nicht offiziell bekundet wird.

> „Um Möglichkeiten zu haben, doch zum Kontrollieren, im Sinne von, also das Wort Kontrolle fällt mir da ein. Ich weiß, dass es anders verkauft wird, aber ich denke das ist ein Wort von Kontrolle und von Ranking und ja auch (um) ein Mittel in der Hand zu haben, bestimmte Dinge durchzusetzen. (…). Ich persönlich fühle mich nicht kontrolliert, aber ich habe das Gefühl dass

> mit dem ganzen Wust von Zahlen die da entstehen, irgendeine Art Kontrolle ausgeführt werden soll.“. (253, 35-35 und 39-39)

Des Weiteren sorgt sich diese Gruppe, dass die Ergebnisse in einem Ranking aufgelistet und veröffentlicht werden. Dadurch geraten diejenigen unter Druck, die in Vergleichsarbeiten schlecht abschneiden und unter dem Durchschnitt liegen. Bei diesen Lehrerinnen und Lehrern überwiegt die Angst davor, als „schlechteste Lehrer“ dazustehen. Dieser Druck führt teilweise dazu, dass diese Personen die Aufgabentypen aus Vergleichsarbeit besonders intensiv mit ihren eigenen vergleichen und dementsprechend anpassen.

> „Ich glaube, die große Angst, die verbreitet ist, ist, dass jeder das Gefühl hat, dass er nachher mit einem schlechten Durchschnitt dasteht und er denkt, alle anderen sind der Meinung, er ist der schlechteste Lehrer. Diese Angst ist, glaube ich, schon da. Und dementsprechend fragen sie dann auch - wollen die Aufgaben sehen, um zu gucken, ob ihre Aufgaben, die sie im Unterricht haben, dem Stil passen oder ob die Schüler nachher nicht wissen, was sie damit anfangen, weil sie so einen Aufgabenstil gar nicht kennen. So war das Interesse dann da. (270, 91-91)

Nach Meinung dieser Gruppe gab es bei der Durchführung der Vergleichsarbeit keine Probleme. „Von den Regularien war es so durchführbar und so haben wir es auch gemacht“ (257, 67-67). Zwei Gymnasiallehrer bewerten jedoch die Aushändigung der Arbeiten an die Lehrperson im Vorfeld der Durchführung als inhuman und inadäquat. In ihrem Fall erhielten sie die Arbeiten morgens, kurz vor Beginn der Durchführung der Vergleichsarbeit. Sie hatten dadurch kaum Zeit und Ruhe, weder um die Anzahl der Arbeiten auf ihre Vollständigkeit hin zu kontrollieren noch um sich einen kurzen Überblick über die Aufgaben zu verschaffen. Sie beklagen die Hetze und den Druck und wünschen sich, dass es in Zukunft wie im Vorjahr praktiziert wird, da konnte jede Lehrperson die Arbeiten rechtzeitig selbst vervielfältigen.

> Und als dann, das ist jetzt kein Vorwurf gegen den Kollegen, aber als dann der Kollege kam und den ganzen Pack mir übergeben hat, ich habe mich schnell in ein Zimmer rein, wo einigermaßen Ruhe war, und habe dann ganz schnell, für mich war das ja alles Neuland, dann da durch die Aufgabenstellung hindurchgearbeitet. Und musste mich erst einmal informieren und die nachzählen, ob alles vollständig ist. Dann bin ich gerade so zwei, drei Minuten vorher fertig geworden. Dann musste ich schon mit dem Packen ins Zimmer reinstürmen und weitermachen bei den Schülern. Das habe ich als sehr inhuman und inadäquat empfunden, muss ich sagen. Beim ersten Mal konnte man es gemütlicher machen, ich weiß nicht, ob Sie das wissen, da durfte es eben jeder selber kopieren. Und jetzt hat man das so verfügt, dass

> das unter Verschluss gehalten wird bis kurz vorher. Diese Modalität des Ablaufs fand ich schlecht.“ (350, 56-57)

> „Was halt schwierig war, ein kleines Problem war, dass wir die ja nur ganz kurz vorher bekommen haben. Die haben wir morgens abgeholt und dann ging es los. Das wäre für die Lehrer geschickt gewesen, wir hätten die eine Stunde vorher gehabt um die mal selber anzuschauen, weil das ging ja dann los mit einem Diktat und da musste ich selber erst mal kurz schauen. Das war für den Lehrer war das schwierig zu praktizieren.“ (358, 43-44)

Nach der Eingabe der Lösungen in das Excel-Datenblatt erfolgt die Auswertung per Computer automatisch. Den Lehrpersonen stehen Grafiken und Schaubilder zur Verfügung, in denen die Ergebnisse der einzelnen Schülerinnen und Schüler, der ganzen Klasse und der Landesdurchschnitt dargestellt sind. Der größte Anteil in der Gruppe der Kritikerinnen und Kritiker setzt sich mit diesen Ergebnisdaten auseinander. Sie interessieren sich sowohl für die Einzelergebnisse ihrer Schülerinnen und Schüler, also auch für die gesamte Klasse und den Landesvergleich.

> „Ich habe mir schon die Vergleiche angeschaut. Und habe immer geschaut, wie steht denn dann die Klasse. Das war aber minimal unterschiedlich diese Vergleichsklasse da im Landesdurchschnitt, das war immer so wie etwas bei meiner Klasse.“ (257, 79-79)

> „Ja, schon sehr genau. So, dass ich bei jeder Aufgabe eben kontrollieren konnte, wie der Stand der Klasse ist, im Vergleich zu dieser Vergleichsklasse, die da ja mit gezeigt wird.“ (252, 58-58)

> „Im Klassenvergleich haben wir in der Schule ein bisschen angeschaut, ansonsten im Landesvergleich nicht.“ (253, 55-55)

Einige geben an, dass sie die Rückmeldungen beim ersten Durchgang, als die Vergleichsarbeiten noch Neuland waren, intensiver nutzten. Eine Lehrperson hat aufgrund eines Schulbeschlusses beim zweiten Durchgang darauf verzichtet, die Daten komplett einzugeben. Als weiteren Grund für die oberflächliche Nutzung führt sie die zufrieden stellenden Ergebnisse ihrer Klasse an.

> „Dazu muss ich sagen, dass ich sie mir beim ersten Mal sehr intensiv angeschaut habe und alles eingegeben habe. Und dann wurden die Grafiken ausgeworfen - im Landesvergleich usw. Beim zweiten Mal hab ich eigentlich darauf verzichtet, das überhaupt alles einzugeben, weil wir es nicht mehr mussten. Das haben wir schulintern beschlossen. (…) Dafür gab es ja einen Grund: Beim ersten Mal war natürlich alles noch neu, man wollte alles durchspielen. Das ging damals auch nicht besonders gut aus mit der Klasse und da wollte ich wissen, wie das einzuschätzen ist. Diesmal war ich dann zufrieden mit den Ergebnissen und dann war es mir nicht mehr so wichtig. Ich habe schon vermutet, dass alles im Mittelfeld liegt und es bei einigen hin und her geht“ (255, 68-69 und 73-73)

Die Lehrerinnen und Lehrer dieser Gruppe fühlen sich durch die Rückmeldungen in ihren Einschätzungen der Leistungen ihrer Schülerinnen und Schüler im Allgemeinen bestätigt. Sie erhalten durch Vergleichsarbeiten keine neuen Hinweise auf deren Stärken und Schwächen und zeigen deshalb auch kein großes Interesse an den Rückmeldungen. Nur einige von ihnen geben an, dass einzelne Schülerinnen und Schülern extreme Abweichungen zu ihren sonstigen Leistungen aufweisen. In diesem Fall setzen sich die Lehrerinnen und Lehrer mit den Ergebnissen ihrer „Ausreißer“ intensiver auseinander. Als Ursache für die große Abweichung betrachten sie u. a. die Schwierigkeiten, die diese Kinder mit manchen Aufgaben- und Fragestellungen hatten.

> „Das war aber so, wie ich mir das eigentlich erwartet hatte. Da ist mir nichts Besonderes aufgefallen, wie gesagt in der Gesamtheit der Gruppe nicht, aber in einzelnen Fällen halt.“ (253, 54-54)

> „Ja da muss ich sagen ist mir aufgefallen, dass einige Schüler da recht gut abgeschnitten haben. Da war ich sehr erstaunt. Im Vergleich zu den sonstigen Arbeiten.“ (262, 30-30)

> „Was mich überrascht hatte, in der Summe der Klasse waren, dass an einzelnen Aufgabenstellungen, die mir eigentlich ganz klar erschienen doch viele gescheitert sind. Und an anderen, die ich für kompliziert gehalten habe doch der Großteil der Klasse zurande gekommen ist.“ (253, 47-47)

In ihrer Gesamtbewertung zeigt sich diese Personengruppe von der Verwendung und der Nutzung von Vergleichsarbeiten enttäuscht. Ein Gymnasiallehrer, der vorgibt, leidenschaftlich gerne zu unterrichten und sich um jeden seiner Schülerinnen und Schüler zu kümmern, bringt seine Enttäuschung zum Ausdruck. Zu Beginn hatte er erwartet, mit Vergleichsarbeiten ein nützliches Diagnoseinstrument zu erhalten. In der Realität bekommt man jedoch lediglich Grafiken und Schaubilder, die keine Erkenntnisse über die Leistungen der Schülerinnen und Schüler enthalten. Die Lehrerinnen und Lehrer wissen nicht, was sie mit den Rückmeldedaten anfangen sollen. Selbst die Fachvorsitzenden, die man um Unterstützung bei der Interpretation der Daten und bei der Umsetzung bittet, können keine Hilfestellungen anbieten. Somit verfliegt bei dieser Gruppe selbst der letzte Funke Hoffnung, dass Vergleichsarbeiten im positiven Sinn als Instrument für die Verbesserung der Unterrichts- und Schulentwicklung genutzt werden können.

> „Also, ich muss sagen, ich wurde ganz groß enttäuscht. Entgegen Verlautbarungen, dass man jetzt hier nützliche Diagnoseinstrumente kriegt, dass man was erfährt über seine Schüler. (…) Und ich bin eigentlich leidenschaftlich

> gern Lehrer. Ich beschäftige mich mit meinen Schülern. Ich gehe mit jedem innerlich auch ein bisschen mit. Deshalb dachte ich auch zunächst einmal, na ja, wenn das wirklich dann an Einsicht bringt, um zum Beispiel jemand gezielt auch helfen zu können für die Zukunft hin, dann könnte man vielleicht doch noch ein Fünkchen Verstand an dem ganzen komischen System finden. Aber, es war eigentlich so. (…) Man kriegt schön geordnete Diagramme, man kriegt Übersichten, man kriegt farbige Übersichten, man kriegt mit Kommastellen, also Aufgabe, Item, eigene Klasse, Vergleichsstichprobe, Differenz und so weiter. Aber, im Prinzip ist das nur eine grafisch schöne Auflistung von Dingen, die ich eigentlich der Arbeit auch schon alleine hätte entnehmen können, oder die man natürlich weiß, wenn man seine Pappenheimer, sage ich jetzt mal, zwei Jahre kennt. Und insofern war ich eigentlich enttäuscht. Ich habe sogar noch dann hier einen Fachvorsitzenden angerufen: ‚Ja, was passiert jetzt eigentlich mit dem hier? Was machen wir da damit?' Weil ich noch ein Stück pädagogischen Idealismus vielleicht hatte, in Resterwartung, dass jetzt irgendwie sich das noch stärker für die Schüler auswirkt. „Ja, durchschauen, guck dir es an. Vielleicht bei der Fachkonferenz mal zehn Minuten drüber reden. Und dann wird es abgeheftet." Und ich muss sagen, da bin ich wirklich enttäuscht. (350, 59-59)

Als Folge für den Unterricht bringen Vergleichsarbeiten auch keinen Vorteil. Im Gegenteil, anstelle von Methodenvielfalt oder einer gezielten Umgestaltung des Unterrichts etc. bemängeln diese Lehrerinnen und Lehrer, dass für zeitintensive Unterrichtsverfahren keine Zeit mehr bleibt. Durch den Druck, den Vergleichsarbeiten auf Lehrpersonen auslösen, verwenden diese nur noch zeitsparende und frontale Vermittlungsmethoden. Somit bleiben Gruppenarbeit oder Aufgaben, die Problemlöseprozesse fokussieren, auf der Strecke. Es findet ein „teaching to the test“ statt, indem nur noch Inhalte und Aufgaben vermittelt werden, die auf die Vergleichsarbeiten vorbereiten, damit die Schülerinnen und Schüler gut abschneiden. Selbst für Wiederholungen reicht die Unterrichtszeit nicht und die Lehrpersonen teilen Arbeitsblätter aus, die von den Schülerinnen und Schülern zuhause bearbeitet werden.

> „Alles, was Zeit braucht - Gruppenarbeit, längere Problemlöseverfahren, all das kommt zu kurz. Man macht es eher nicht, weil man das Gefühl hat 'Oh, ich muss Gas geben, damit ich jedes Thema wenigstens noch so anschneide, dass die Schüler das in der Vergleichsarbeit verstehen.' Dass die Schüler das Grundverständnis haben. Daher führt es für mich eher ins Negative - ich mache mehr Frontal und mehr Hausaufgaben. Was Wiederholungen angeht, so reicht die Zeit nicht. Ich habe nicht wiederholt, aber ich habe (…) angefangen, Blätter auszuteilen - wöchentlich.“ (256, 134-135)

> „Als Lehrer fühlt man sich dann vielleicht schon auch noch mal natürlich unter Druck und guckt, dass man noch mal eine kleine Wiederholungsphase vorher machen kann. Habe ich dann auch gemacht, obwohl mir das ganze

> eigentlich widerstrebt. Und, als Lehrer hat man auch anderes zu tun.“ (350, 39-39)

Einige räumen ein, eventuell in Zukunft für kommende Jahrgänge bereits ab der fünften Klassenstufe komplexere Aufgabenformate im Unterricht zu bearbeiten. Andere werden vermehrt Arbeiten konzipieren, die sich an Vergleichsarbeiten orientieren und unterschiedliche Kompetenzbereiche beinhalten. Dazu zählen einerseits Aufgaben zum Problemlösen, zum vernetzten Denken oder zum Modellieren, die weder im bisherigen Unterricht noch in Schulbüchern verankert sind. Diese Lehrerinnen und Lehrer sind sich allerdings auch der Tatsache bewusst, dass für diese Art von Aufgaben viel mehr Unterrichtszeit benötigt würde und diese im Schulalltag nicht zur Verfügung steht.

> „Ich würde stärker, von Fünf angefangen, diese Art von Aufgaben, also diese Komplexität in Sinne von einem Text, sowohl Grammatik, als auch sprachliche, als auch stilistische und ähnliche Mittel versuchen zu erarbeiten. Das habe ich eigentlich erst in Sechs so intensiv gemacht. Das war so meine Idee, das für Fünf anzulernen.“ (253)

> „Wir haben da zusammen darüber gesprochen und haben uns überlegt, dass wir jetzt eine Kombinationsarbeit, z.B. die wir jetzt dieses Jahr machen wollen so in diesem Bereich machen. So ähnlich aufbauen wie die Vergleichsarbeit, dass eben verschiedene Punkte angesprochen werden und damit auch verschiedene Kompetenzen. Für die Schüler natürlich eine Chance einfach besser zu sein, weil die ja dann ihre Kompetenzen in verschiedenen Gebieten haben. Und wenn sie in einem Gebiet nicht so gut sind, können sie das durch das andere wieder ausgleichen.“ (257)

> „Ja gut, ich denk, dass man grundsätzlich auch bei der Stoffvermittlung grad was die Bereiche, die man vielleicht früher nicht so betrachtet hat, also bei den Kompetenzen, dass man die mehr in den Fokus nimmt. Gerade Problemlösen, so vernetztes Denken, Modellieren sind solche Punkte, die das Schulbuch auch wiederum wenig vorgibt. Dass man die halt auch üben sollte und thematisieren sollte. Aber Problemlösen, das ist schwierig, weil das Problemlösen braucht Zeit und Zeit hat man keine. Insofern zieht man sich dann wieder zurück auf den Standard. Wenn man viel Zeit hätte, könnte man auch mit Schülern Probleme lösen und auch komplexere Aufgaben lösen und andere und verschiedene Herangehensweisen mal ausprobieren. Aber auf Grund der Zeit und der Stofffülle ist das nicht möglich oder schwierig möglich.“ (353)

Die Kritikpunkte, die dieser Personenkreis im Rahmen der Reflexion anführt, sind zahlreich und beziehen sich auf verschiedene Aspekte. Verglichen mit den anderen Gruppen, bemängelt auch diese Gruppe der Kritikerinnen und Kritiker zunächst einmal den Zeitpunkt der Vergleichsarbeit. Dadurch dass der Termin am

Ende des Schuljahres stattfindet, bleibt den Lehrerinnen und Lehrern keine Zeit, um notwendige, langfristigen und nachhaltigen Maßnahmen abzuleiten. Zudem kritisieren auch sie das geringe Niveau der Vergleichsarbeiten.

> „Wissen Sie es geht am Ende vom Schuljahr so viel auf einen Schlag los. Ich habe diese 8. Klasse, die 6. Klasse, zwei Mal die Mathevergleichsarbeiten gemacht, dann ist gleichzeitig bei den Zehnern Schluss gewesen. (…) Ich habe mündliche Prüfung gemacht und so weiter. Also am Schluss konnte ich manches gar nicht mehr in Ruhe nachschauen. (…) Klar, das war am Ende vom Jahr, wo alle übrigen Sachen auch sind. Es haben die Zehner die Abschlussfahrten dann noch und ach…das war so viel durcheinander in dem Fall. (…) Dann war ich letztes Schuljahr noch krank, gell? Das ist praktisch das Ende vom letzten Schuljahr hier gewesen.“ (265, 65-71)

Dadurch, dass sich diese Lehrpersonen intensiv mit Vergleichsarbeiten und deren möglicher Auswirkungen auf die Unterrichtspraxis beschäftigen, weiten sie die Kritik noch auf weitere Bereiche aus. Zum einen ärgern sie sich über die Zusammenstellung der Vergleichsarbeiten. Bei diesem Punkt führen sie sowohl die mangelhafte Aufgaben-, als auch die einseitige Themenauswahl an:

> „Es gab so ein paar Fragen, die nicht ganz eindeutig waren für die Schüler. Also, nicht weniger, aber da hingen gleich eine ganze Punkte mit zusammen. Wenn die jetzt einmal falsch verstanden haben, bei einer Aufgabe wo man fünf Items haben konnte, dann waren gleich vier weg oder so.“ (358, 66-66)

> „Und ich war von den Vergleichsarbeiten maßlos enttäuscht. Dass bei den 6.ern …viel zu wenig Bruchrechnung und Dezimalrechnung dran kam. Das haben wir das gesamte Jahr durchtrainiert. Stattdessen kam ein ganz winziges Kapitelchen draus dran, diese Sachen mit dem Zahlenstrahl und so. Und bei den 8.ern, da habe ich auf Teufel komm raus Formelumformungen, binomische Formeln und alles trainiert. Das kam fast gar nicht dran oder überhaupt nicht dran, gell? Also in sofern waren die Schüler auch enttäuscht, die haben dann gesagt: „Herr <xxx>, wieso haben Sie das ganze Zeug so häufig trainiert? Das kam ja fast nicht dran. (…) Die ganzen Basisaufgaben kamen nicht dran. Es war für mich etwas frustrierend, muss ich sagen.“ (265, 30-32)

Zum andern bringt bei dieser Gruppe der Themenbereich „Bewertung der Vergleichsarbeiten“ Wut, Ärger und Enttäuschung hervor. Der in den Vergleichsarbeiten vorgeschriebene Bewertungsmaßstab unterscheidet sich so entscheidend von bisherigen offiziellen Vorgaben, an denen sich professionelle Lehrpersonen bei der Leistungsbeurteilung orientieren. Sie sind entsetzt und beklagen sich, dass die bildungspolitisch Verantwortlichen bei Vergleichsarbeiten diese grundsätzlichen Vorgaben nicht einhält und einen geschönten Notenspiegel zugrunde legt,

der auch das Maß der ‚Proportionalität' verletzt. Sie sind sich sicher, würden sie diese Form der Leistungsmessung und –bewertung in ihren Arbeiten anwenden, dass sie von Seiten der Schulaufsichtsbehörde massiv unter Druck gesetzt würden.

> „Es ist meiner Meinung nach so, dass ich dachte, ich muss proportional bewerten. Und die müssen das nicht. Mich hat dann eher irritiert und interessiert, warum die das nicht müssen, wenn ich das muss." (256, 111-113)

> „Aber da die Benotung nach oben dermaßen locker ist, haben meine Schüler zwar im Durchschnitt nur 40% der Arbeit gelöst, haben aber einen Schnitt von 3, 0. Bei meinen normalen Noten wäre das eine 4,0. Mit einem richtigen Notenschlüssel wäre diese Arbeit den Leistungen der Schüler gerecht gewesen. So haben sie zu gute Noten bekommen." (256, 58-58)

> „Ich versteh sowieso nicht, warum die Notenskala hier nicht proportional ist und ich muss das immer machen, was ich auch ganz gut find … und warum nicht halbe Punktzahl eine 3 - 4 ist sondern eine 2 - 3. Das kapiert auch niemand. Also ich nicht zumindest. (…) Das war für mich das Schlechteste an der Vergleichsarbeit. wenn diese Vergleichsarbeit normal, nach meinen Maßstäben benotet worden wäre, wären auch die Noten so gewesen, wie die Schüler waren. Für mich war es einfach ein Problem, dass ich Schülern Noten herausgegeben habe, die meinem Empfinden nach eigentlich eine Note schlechter sind." (256, 76-78)

> „Ich haben bloß gedacht: „Das sind schon Schloris da irgendwie." Ich werde als Lehrer gezwungen Arbeiten zu schreiben, wo 3,5 die Hälfte der Punktzahl ist und anders darf es nicht bewertet werden. Und hier ist plötzlich 2,5 die Hälfte der Punktzahl. Also das hat mich schon sehr verblüfft." (265, 23-23)

> „Was natürlich auch falsch war, das hätte ich nie als Benotung machen dürfen, es war keine lineare Notenskala. Das war eine ganz brutale...hätte ich als Lehrer das so früher gemacht, hätte mir die Behörde, das Schulamt die Aufgabe, die Arbeit gestrichen. So darf ich nicht benoten. Das ist ein bisschen Widerspruch für mich gewesen. Ja? (…) Das ist eine gewisse Irritation für mich, aber durch diese extrem gute 2,5 – mittige Skala war es dann natürlich so, dass das Ergebnis von den schlechten Aufgaben von den Schülern dadurch wieder ausgeglichen wurde. Also die sprachlich schwierigen Aufgaben, die manche oder ein Großteil gar nicht hingekriegt haben, sind praktisch unter den... sind weggefallen im Grunde genommen aus diesem Schnitt heraus. Und die haben dann viel besser abgeschnitten, zum Teil die Schwachen, als sie eigentlich sind in Mathe. Wegen diesem irrsinnigen Notendurchschnitt." (265, 23-26)

Die negativen Folgen und Konsequenzen, die sich daraus ergeben, sind zahlreich. Zum einen wird bemängelt, dass Schülerinnen und Schüler, deren Versetzung bei bisherigen Klassenarbeiten gefährdet ist, durch diese ‚unrealistische' Notengebung ihre Gesamtleistung so anheben, dass sie dadurch versetzt werden. Als wei-

teres Problem wird die Reaktion der Eltern auf die „verbesserten Ergebnisse“ ihrer Kinder betrachtet. Wenn staatlich vorgegebene, umfangreiche Kompetenzprüfungen ein positiveres Leistungsprofil der Schülerinnen und Schüler zeichnen, als es professionelle Lehrerinnen und Lehrer innerhalb eines längeren Zeitraums bestätigen, haben Eltern einige Gründe, die Professionalität der Lehrkräfte in Frage zu stellen und dagegen anzugehen. Das führt bei der Lehrerschaft zu erheblichen Legitimationsproblemen, denen sie im Schul- und Unterrichtsalltag zusätzlich ausgesetzt sind.

> „Aber der Bewertungsmaßstab, der hier angewendet wurde, den dürfen wir ja im Grund genommen bei Klassenarbeiten gar nicht anwenden. Uns wird immer wieder gesagt, wir müssen eine lineare Bewertung anlegen, jeder Punkt muss eigentlich gleich viel Wert sein und das ist ja hier total daneben. Und deshalb, z.B. ist eine Schülerin, die vor der Vergleichsarbeit auf 4,5 stand, und die bei einer regulären Arbeit vielleicht sich hätte nicht verbessern können, die hat sich durch diese Arbeit verbessert und weil wir die mit gezählt haben, hat sie die Vier erreicht und konnte somit versetzt werden. In einem anderen Fall hätte es vielleicht eine nicht Versetzung bedeutet. Also, dieser Bewertungsmaßstab im unteren Bereich damit, das haben auch die Kollegen so geäußert, damit sind wir eigentlich nicht zufrieden. Denn Eltern wird da eine anderes Bild vorgespielt und wir stehen da als Lehrer eigentlich immer ein bisschen dumm da, wenn dann Schüler die bei uns über das Jahr eine Fünf haben, dann plötzlich eine Drei schreiben. Und das war ja von der Punkteverteilung und wie die Noten zugeordnet waren.“ (252, 73-73)

In dieser Gruppe fällt auf, dass sie im Vergleich zu anderen Gruppen mit allen Akteuren und Gremien an ihrer Schule systematische und intensive Gespräche führen. Entsprechend der Gruppe der Befürworterinnen und Befürworter setzen auch sie sich mit den Ergebnissen aus Vergleichsarbeiten auseinander und tauschen ihre Erfahrungen und Meinungen vor allem mit den Parallelkolleginnen und –kollegen und mit ihren Schülerinnen und Schülern aus. Realschullehrpersonen diskutieren häufig in den entsprechenden Fachkonferenzen.

5.7.5 Der ablehnende, belastete Typ

Bei diesem Typus von Lehrpersonen, die die Vergleichsarbeiten massiv ablehnen, kommt die negative, belastete Einstellung zu Vergleichsarbeiten in allen Interviews deutlich zum Ausdruck. „Solche Lehrer braucht man, (…) solche Doofen, die alles mitmachen.“ (260, 51-52) Diese Lehrerinnen und Lehrer erkennen in Vergleichsarbeiten keinen Sinn und betonen ihre Wut. Der Grund für ihren Ärger liegt in einer extremen Überlastung, der durch die zahlreichen zusätzliche Arbeitsanforderungen, denen sie im Schul- und Unterrichtsalltag ausgesetzt sind, zustande kommt. Eine Teilzeitkraft beklagt die Mehrfachanforderungen, die trotz ihres reduzierten Lehrdeputats auf ihr lasten. Sie nahm im Vorfeld der Vergleichsarbeit neben all ihren Arbeitsbereichen noch an einer Fortbildung zu Vergleichsarbeiten teil, über die sie sich im Interview ebenfalls enorm ärgert.

> „Weil wir uns alle über diese ganze Zusatzarbeit, die relativ wenig bringt, nur ärgern. Sie müssen sich vorstellen, ich bin Teilzeitlehrer, muss zwei Vergleichsarbeiten, plus Prüfung, bin Verbindungslehrer, Klassenlehrer und, und, und. Wenn es mir wirklich was bringen würde, okay. Ich habe mich schon in der Fortbildung geärgert. Ich ging zu einer Fortbildung über die Vergleichsarbeit, die sehr schlecht war und dann sagt diese Kollegin, das Aha-Erlebnis war für sie, dass die Schüler der Klasse acht Schwierigkeiten mit dem Konjunktiv haben. Da brauche ich keine Vergleichsarbeit.“ (254, 146-147)

Ein Realschullehrer gibt ebenfalls an, am Ende seiner Kräfte zu sein. Bei ihm liegt der Grund für seine extreme Belastung unter anderem darin, dass seine Schule für die Pilotierung ausgewählt wurde und er sowohl für die Durchführung und Auswertung der Pilotierungsarbeiten als auch der Vergleichsarbeiten verantwortlich ist. Er ärgert sich darüber, dass er aufgrund des großen zeitlichen Arbeitsaufwandes zum eigentlichen Arbeiten nicht mehr kommt. Seine Kritik umfasst selbst das Interview, für das er sich extra Zeit nehmen muss, obwohl er diese nicht hat. Er kommt zu dem Schluss, dass er von dem ganzen Verfahren „die Schnauze voll hat“.

> „Ich kann das Ding (Pilotierung) nicht mehr hören, sage ich ihnen ganz klar. Ich mache das jetzt so mit, weil das so ist, aber ich stehe sehr distanziert dem Ganzen gegenüber, zwischenzeitlich. (…) Ich habe es Ihnen ja schon vorher gesagt, weil ich zum eigentlich Arbeiten nicht mehr komme. Ich mache nur

> noch Listen und hier Interviews jetzt, das gehört auch dazu, usw. Das kann es nicht sein. (…) Ich habe die Schnauze voll, ja.“ (260, 17-23)

Eine weitere Lehrperson beklagt die Belastung durch ihre extrem schwierige Klasse im Bereich des Sozialverhaltens. Sie legte mehr Wert auf erziehenden Unterricht und die Vermittlung sozialer Kompetenzen als auf die Vermittlung fachlicher Inhalte. Somit blieb ihr kaum Zeit für die Auseinandersetzung mit Vergleichsarbeiten, die für diese Lehrperson in dieser Klasse keine Rolle spielte.

> „Muss aber sagen, ich hatte eine extrem schwierige Klasse. Zwei Schüler hatten in der fünften Klasse schon eine Anzeige wegen Körperverletzung. Ich habe sicherlich mehr Wert auf soziale Kompetenzen, als auf fachliche Kompetenzen gelegt im ersten Schuljahr von Klasse fünf. Da ging also ein Großteil des Unterrichts nur mit solchen Dingen drauf, Polizei an der Schule, Sozialarbeiter an der Schule, Anti-Gewalttraining und so weiter. Und da ich als Klassenlehrer ja nur das Fach Deutsch und Sport bei den Mädchen hatte, können Sie sich vorstellen was dann auf der Strecke blieb.“ (254, 41-42)

Alle Lehrpersonen dieser Gruppe fühlen sich zur Durchführung von Vergleichsarbeiten gezwungen und können keinen pädagogischen Nutzen darin erkennen. Ihrer Meinung nach handelt es sich bei Vergleichsarbeiten um eine reine Beschäftigungstherapie, ohne Sinn und Zweck. Eine Lehrperson betont, dass mit Hilfe von Selbstevaluationen mögliche Schwächen erkennbar werden und darauf Konsequenzen folgen müssten. Bei dem vorliegenden Verfahren sei dies jedoch nicht gegeben und nicht möglich. Die Lehrpersonen bemängeln die fehlenden Analyseschritte und dass an den tatsächlichen Ursachen nicht gearbeitet werden würde. Nach Meinung dieser Gruppe haben Vergleichsarbeiten also keine Funktion sondern werden als reines Alibi verstanden. Die bildungspolitisch Verantwortlichen der Landesregierung beweisen mit der Einführung von Vergleichsarbeiten lediglich puren Aktionismus als Folge der moderaten PISA Ergebnisse.

> „Und da werde ich aber gezwungen dazu, ich werde gezwungen (…) um zu dokumentieren, das hast du gemacht. Und das kann es nicht sein. (…) Weil ich das muss. Weil ich da keinen pädagogischen Nutzen drin erkennen kann. Ich muss es, das ist eine Anordnung und deshalb müssen wir es. (…) das war nur, damit wir hier beschäftigt waren. (…) Das Kreuzle da eintippen ist Beschäftigungstherapie.“ (260, 32-33, 74-74)

> „Also, das ist für mich ein Stück weit Beschäftigungstheorie.“ (259, 130-130)

> „Ich meine Selbstevaluation wäre dann für mich gegeben, wenn ich so einen Unterschied merke, dass ich sage hoppla was ist eigentlich los. (…) L: Also, wenn eine Konsequenz daraus erwachsen soll, wenn ich sage der hat die und

> die Schwäche und das hat jetzt die und die Konsequenz, der würde jetzt so oder so gefördert, dann würde ich ja noch einen Sinn dahinter sehen. Aber das ist doch nicht. (…) I: Ja, es ist einfach bloß eine Überschüttung. (259, 126-130)

> „und ja Aktionismus. Das ist in meinen Augen purer Aktionismus. (…) Pisa-auslöser und jetzt muss man schauen, dass man um sich schlägt, dass man zu Ergebnissen kommt. An den Ursachen wird nicht gearbeitet, Einzelanalysen werden keine betrieben. Das geht pauschal, die Schüler bekommen eine Nummer, ich bekomme eine Nummer. Der Lösung der Detailfragen und Probleme kommt man so nicht näher, absolut nicht.“ (260, 26-28)

Diese Lehrerinnen und Lehrer erleben Vergleichsarbeiten also nicht als Vergleichsinstrument, das zum Beispiel zur Förderung von Schülerinnen und Schülern verwendet werden könnte. Sie bewerten es primär als „Kontrolle von Schülern und Lehrern“ (260, 26-26). Sie sind davon überzeugt, dass vor allem die Kontrolle der Lehrpersonen von den Eltern erwartet und erwünscht ist. Im folgenden Zitat schwingt die negative „Außenansicht“ des Lehrerberufs mit, der teilweise in der Gesellschaft vorherrscht. Faule Lehrer, die vormittags Recht und nachmittags frei haben, werden nun endlich, Dank der Vergleichsarbeiten kontrolliert und überprüft.

> „Ich sage den Eltern dann, ich glaube eher, dass die Lehrer doch durchaus mehr überprüft werden als die Schüler. Und von den Eltern wird es so gesehen, das hören die Eltern dann gern. Da reagieren sie sofort darauf. Denn ich habe das Gefühl dass die Elternschaft schon erwartet, dass die Lehrer von außen kontrolliert werden.“ (254, 54-55)

Wenn Vergleichsarbeiten als Kontrollinstrument verwendet werden, kann damit vor allem Druck auf die Lehrpersonen ausgeübt werden. Einige geben zu, dass sie sich durch dieses Evaluationsinstrument stark unter Druck gesetzt fühlen und betonen gleichzeitig, dass sich durch diesen Druck die Leistungen der Schülerinnen und Schüler nicht ändern werden. Eine Realschullehrperson beschreibt, dass Druck auch dann entstehen kann, wenn die Ergebnisse aus den Leistungsvergleichen als Rankingliste öffentlich gemacht werden.

> „Ja freilich hatte ich auch Druck. Das Ganze ist Kontrolle. Was soll das? Die Leistungen sind in keinster Weise verbessert.“ (260, 74-74)

> „Ich denke (Kontrolle und Druck) ist der nächste Hintergedanke. Ich denke irgendwann werden die Vergleichsarbeiten dafür benutzt um ein Ranking zu machen, und das Ranking dann zu veröffentlichen und damit einen gewissen Druck aus zu üben.“ (254, 39-39)

Ein Realschullehrer bewertet den Zeitpunkt der Durchführung als unsinnig, weil er in einem Fach acht Wochen vor Schuljahresende liegt und bis dahin noch nicht alle Themenbereiche im Unterricht behandelt sein können. Des Weiteren problematisiert sie, dass Vergleichsarbeiten auch in dem Fall bewertet werden, wenn einzelne Testinhalte noch nicht gelehrt wurden. Ihre Bedenken beziehen sich vor allem auf den rechtliche Rahmen der Notengebung.

> „Es ist ja Unsinn. Es soll der Stoff dieses Schuljahres abgefragt werden und wenn ich acht Wochen vorher schon fertig wäre, dann heißt die Frage, was mache ich dann in der Zwischenzeit. Da kann ja irgendetwas von der Konzeption her nicht stimmen. Das kann ja nicht sein, das soll ja heißen ich bin immer schon acht Wochen vorher fertig und dann frage ich mich, was soll ich in der restlichen Zeit unternehmen. Das kann ja nicht stimmen. Das heißt da wird ein Stoff abgefragt, das ist übrigens auch sehr problematisch, ich gebe eine Note die wird ja als Arbeit bewertet. Ich gebe eine Note drauf für einen Stoff den ich abfrage, den ich noch gar nicht behandelt habe. Ist sowieso die Frage, wie weit das überhaupt rechtlich akzeptabel ist oder abgesichert ist. Da stimmt die ganze Logik nicht. (…) Es geht einfach um die Frage der Notenfindung. Wenn ich eine Arbeit schreibe, eine reguläre Arbeit schreibe über einen Stoff den ich nicht behandelt habe, kriege ich Ärger logischerweise und berechtigterweise. Ich kann nicht eine Arbeit schreiben über einen Stoff der gar nicht behandelt wurde.“ (260, 68-70)

Die Personen, die dieser Gruppe zugeordnet werden können, betonen ihr Desinteresse an den Rückmeldedaten. Sie schenken ihnen keine Beachtung und befassen sich in keiner Weise damit. Zum einen interessieren sie sich nicht dafür, weil sie mit ganz anderen Problemen zu kämpfen hatten. Zum anderen geben sie an, die Klasse in zwei Jahren so gut zu kennen, dass Vergleichsarbeiten keine weiteren Hinweise für die Stärken und Schwächen ihrer Schülerinnen und Schüler liefern können.

> „Gar nicht intensiv, hat mich gar nicht groß interessiert. (…) Interessiert mich überhaupt nicht, weil ich ganz, ganz andere Probleme in dieser Klasse hatte.“ (254, 74 – 75)

> „Nein, das interessiert mich (…) nicht. Wenn ich jetzt meinen Schnitt nehme, also 3,1. Ich nehme mal an, die Arbeiten sind im Grunde so konzipiert, dass ich sage mal, dass es in der Mitte der Notenskala, 3,5 dass das machbar sein muss. Dann gehe ich nicht davon aus, sagen wir mal, dass ich jetzt völlig daneben liege. Ich kenne ja die Klasse zwei Jahre und ich weiß ja was für ein Potential drin ist.“ (259, 49-49)

Diese Lehrerinnen und Lehrer bemängeln, wie die Gruppe der Kritikerinnen und Kritiker, vor allem die Bewertungspraxis, die den Vergleichsarbeiten zugrunde

liegt und die Benotung. Zum einen beklagen sie, dass die Ergebnisse aus Vergleichsarbeiten, in der Summe betrachtet, die unterschiedlichen Fähigkeits- und Leistungsniveaus der Schülerinnen und Schüler nicht genügend differenziert widerspiegeln. So schneiden sowohl schlechte, als auch gute Schülerinnen und Schüler bei Vergleichsarbeiten nur mittelmäßig ab und landen auf einem mittleren Niveau.

> „Mir ist im Grund bloß aufgefallen, dass gute Schüler eher schlechter und schlechte Schüler eher etwas besser abgeschnitten haben. Also das heißt, es gab jetzt nicht ganz Gute und es gab auch nicht ganz Schlechte. Das ist mir aufgefallen. Also einfach durch die hohe Punktzahl und so weiter wird das im Grunde genommen einnivelliert.“ (259, 45-45)

Zum anderen wird die strenge Anleitung bei der Bewertung der Vergleichsarbeiten kritisiert. Sie sieht bei der Korrektur der Aufgaben keine Abstufung in halbe Punkte vor. Die Aufgaben werden entweder bei der richtigen Lösung mit voller Punktzahl bewertet oder mit null Punkten, auch wenn die Schülerinnen und Schüler nur minimale Fehler machen.

> „Aber dann halt doch durch die Anweisung, wenn das nicht dabei ist, dann darf gar kein Punkt gegeben werden. Sonst kann man bei Arbeiten mit einem Punktsystem auch mal einen halben Punkt geben, verstehen Sie? Das gibt es dann da ja nicht. Wenn die Anweisung ist, wenn von fünf Sachen drei da sein müssen, sonst gibt es gar nichts. Der hat zwei und dann hätte ich halt einen halben Punkt noch gegeben.“ (259, 46-46)

Auch in dieser Gruppe wird der Bewertungsmaßstab von Vergleichsarbeiten als willkürlich gewählt kritisiert, der jegliche Transparenz und Kontingenz vermissen lässt. Im Vergleich zur Leistungsmessung bei Klassenarbeiten sei dieser rechtlich nicht haltbar.

> „Was mir da natürlich aufgefallen ist, was wir so lernen Hälfte der Punktzahl, muss 3,5 sein, das entspricht dem ja teilweise nicht. Das ist von Jahr zu Jahr unterschiedlich und das ist für mich uneinsichtig. Das ist für mich irgendwo willkürlich. Manchmal war die Hälfte Punktzahl schon drei manchmal gewesen, und manchmal war Hälfte der Punktzahl weiter untern gewesen da dann. Und das kapiere ich nicht. Das könnte ich beispielsweise bei einer normalen Klassenarbeit nicht machen, sonst würden mir die Eltern auf das Dach, oder auch der Chef. Das wird bei uns verkündet, das wird auch vom Verwaltungsgericht so und so muss das sein. Das kapiere ich nicht, dass das bei den Vergleichsarbeiten so nicht gemacht wird.“ (259, 67 -69).

Diese Gymnasiallehrer bringt seinen Ärger über die missverständliche Bewertungspraxis im Interview deutlich zum Ausdruck. Er hatte große Mühen, mit den

Parallelkolleginnen und –kollegen eine einheitliche Bewertung innerhalb und zwischen den einzelnen Klassen durchzuführen. Als großes Manko führt er die intransparenten Bewertungskriterien an, die zu einem gewissen Teil auf Missverständnissen beruhen. Er verzweifelt beinahe an der Tatsache, dass öffentliche Vergleichsstudien, die ihren Ursprung in den Geisteswissenschaften haben, so große Mängel aufweisen. Aus diesem Grund lehnt er Vergleichsarbeiten für sich ab.

> „Ja, (LACHT) da kommt mir jetzt, indem ich diese Arbeit besprochen habe mit meinen Schülern, bin ich aus dem Ärger nicht mehr raus gekommen. Denn ich kriege natürlich an die [Schüler] nicht ran, die Logik die [sich] dahinter [verbirgt]. Im Übrigen gibt [es auch] missverständliche Bewertungsmaßstäbe. Wir haben uns, drei Kollegen, wir haben uns unterhalten, sehr lange darüber, um zu einer einheitlichen Bewertung zu kommen unseren Schülern gegenüber, der Jahrgangsstufe gegenüber, wir wissen aber nicht ob andere Schulen sich an das halten. (…) Also, das ist völlig intransparent. Ich habe selbstverständlich das durchgesprochen, die Ergebnisse ihre Arbeit zurückgegeben, das durchgesprochen von vorne bis hinten, auch meine Bewertungskriterien offen gelegt, aber letztendlich ist das nicht transparent. Weil teilweise, ich will sagen im Detail das ist nicht überwiegend aber es sind unter die 5% sage ich jetzt mal einfach so, sind Missverständnisse. (…) Also ich halte die Vergleichbarkeit in einem geisteswissenschaftlichen oder geistesfachlichen, also (…),…aber gut das ist mein eigenes Verzweifeln. Will ich nicht, muss ich ja auch nicht.“ (363, 62-66)

In dieser Gruppe findet so gut wie kein systematischer Austausch über die Ergebnisse aus Vergleichsarbeiten statt. Es wird vereinzelt von informellen Gesprächen berichtet.

5.7.6 Zuordnung der einzelnen Interviews zu den fünf Rezeptions- und Nutzungstypen

In der folgenden Tabelle ist die Aufteilung der Interviews in die entsprechende Typenspalte dargestellt. Bei einigen Interviews ist eine eindeutige Zuordnung zu einem Typus nicht möglich. Die Nummern der entsprechenden Interviews sind in diesem Fall nicht zentral angeordnet sondern wurden je nach Passung nach links oder rechts gerückt.

Eine Schwierigkeit stellt die Zuteilung der Interviews dar, die an Hauptschulen geführt wurden. Das extrem niedrige Niveau hatte zur Folge, dass sich die meisten Lehrerinnen und Lehrer für die Ergebnisse der Vergleichsarbeiten nicht interessierten und sich folglich auch nicht damit auseinandersetzten. Dementsprechend müssten die meisten Lehrpersonen in die Spalte des „desinteressierten, gelassenen Typs“ einsortiert werden, was aber nicht deren grundsätzlicher Einstellung entspricht.

Tabelle 5.7.2: Zuordnung der einzelnen Interviews zur Typologie

Der begeisterte, überzeugte und befürwortende-Typ	Der positiv mitspielende Typ	Der desinteressierte, gelassene Typ	Der negativ kritische Typ	Der ablehnende, belastete Typ
150	153	Restliche Interviews	157	158
251 263 267	255 261 264 268 270 271 272	250 252 253 266 269	256 257 258 262 265	254 259 260
361 368	354 357 362 364	353 355 356 360 366 367 370	350 351 352 358 365 369	359 363

6 Zusammenfassung der Ergebnisse, Diskussion und Ausblick

6.1 Qualitätsbegriff und Steuerungsmodelle

Seit dem Schuljahr 2005/06 werden in Baden-Württemberg Vergleichsarbeiten als Instrument der Qualitätsentwicklung und Qualitätssicherung verpflichtend durchgeführt. Sie stellen ein Element einer Gesamtstrategie dar, die von der Kultusministerkonferenz (KMK) als Antwort auf das moderate Abschneiden deutscher Schülerinnen und Schüler bei internationalen Schulleistungsstudien entwickelt wurde. Diese Strategie umfasst zahlreiche Maßnahmen und Instrumente, die sukzessive in allen Bundesländern, auf den unterschiedlichen Ebenen des Bildungssystems implementiert wurden; alle mit dem Ziel, Qualität zu entwickeln und zu verbessern, bzw. bestehende Qualität zu sichern (KMK 2006).

Somit rückt der Qualitätsbegriff zuallererst ins Zentrum einer kritischen Betrachtung. Bereits die theoretische Auseinandersetzung mit dem schulischen Qualitätsbegriff zeigt, dass keine allgemein gültige Definition des Begriffs existiert. Qualität stellt nach Heid (2007) das Ergebnis einer Sachverhaltsbeurteilung dar, die auf Kriterien basiert, die wiederum erst durch Aushandlungen der beteiligten Akteure festgelegt werden und deren Sichtweisen widerspiegeln. Die Normierung von Qualität stellt sich im Bildungswesen zudem problematisch dar, da sich auf allen Ebenen (Makro-, Meso- und Mikroebene) eine Vielzahl von unterschiedlichen Akteuren befindet. Ob sich all jene gleichermaßen an diesen Aushandlungen beteiligen und ob deren Ansichten und Handlungslogiken übereinstimmen und somit kongruent sind, muss bezweifelt werden. Die Realität zeigt, dass im Bildungssystem häufig Entscheidungen und Maßnahmen zunächst von der Ebene der Bildungspolitik und Bildungsadministration getroffen werden und somit von „oben herab“, als „Top-Down“-Steuerungsinstrumente in das Bildungssystem implementiert werden. Die Problematik dieses Vorgehens wird u. a. an zahlreichen Befunden aus der Lehrplanforschung deutlich. Mit Lehrplänen und Lernzielkatalogen als Elemente der Inputsteuerung wurde indirekt versucht, die Lerner-

gebnisse der Schülerinnen und Schüler zu verbessern. Dabei wurden jedoch die Relevanz der beteiligten Akteure, der Prozesse und Kontexte nicht berücksichtigt.

Genau hier setzen jene Steuerungsmodelle an, die umfassende Betrachtungsweisen ermöglichen, wie sie in Kapitel 2.3.2 und 2.3.3 beschrieben sind.

Das Modell der Educational Governance und das Fend'sche Modell der Rekontextualisierung liefern bei der Implementierung neuer Maßnahmen und Instrumente einen wichtigen Beitrag, weil sie zum einen alle beteiligten Akteure berücksichtigen und zum anderen mögliche Übersetzungsschwierigkeiten beleuchten, die bei den Übergängen von einer Ebene zur nächsten entstehen könnten. Neben die vertikale, hierarchische Steuerungsphilosophie, wie sie in Top-Down-Verfahren beschrieben sind, rückt durch die Akteursperspektive eine horizontale Gestaltungsmacht ins Blickfeld. Absprachen können auf den jeweiligen Ebenen mit den beteiligten Personen und Gruppen ausgehandelt werden und als gemeinsamer Handlungsrahmen formuliert und festgelegt werden. Diese Gestaltungsmöglichkeiten müssen bei der Umsetzung und Implementierung von Maßnahmen berücksichtigt werden.

Allerdings werden durch die Verwendung der neuen Steuerungsmodelle, neben der oben beschriebenen Umorientierung von einer Input- auf eine evaluationsbasierte Outputsteuerung und der Gestaltungsmöglichkeit einzelner Akteursgruppen, auch weitere Merkmale im Bildungssystem verankert, die auf den ersten Blick nicht deutlich zum Vorschein kommen aber dennoch ausgeprägte Wirkungen erzielen können und folglich berücksichtigt werden müssen:

So bildet die Erweiterung der Eigenverantwortung der Schule, die durchaus mit Wettbewerbseffekten einhergehen kann, ein weiteres Merkmal der neuen Steuerung. In direktem Zusammenhang damit steht die Rückbindung professionellen Handelns an ein Schulprogramm, das zum einen zentrale Zielsetzungen vorgibt und zum anderen auch der Rechenschaftslegung dienen kann. Die Einführung von Qualitätssicherungssystemen, wie z.B. Leistungsvergleichsstudien, Schulinspektion oder auch die Durchführung von Vergleichsarbeiten stellen weitere zentrale Steuerungselemente dar. Nicht zuletzt sind verschiedenste Akteure an der Umsetzung der Maßnahmen beteiligt, die je nach Formation in individuelle, kollektive oder kooperative Akteure eingeteilt werden können. Deren Handlungskoordinati-

on ist über Modi wie Beobachtung, Verhandlungen und Beeinflussung gekennzeichnet.

Aus diesen Merkmalen der neuen Steuerung lassen sich zahlreiche Folgen für das Bildungssystem ableiten: Für die schulischen Akteure bedeutet sie sowohl neue Arbeitsaufgaben als auch mehr Verantwortung. Teilweise kann man eine Hierarchisierung des Verhältnisses zwischen Schulleitung und Lehrerschaft beobachten. Ein Veränderungsdruck auf das berufliche Selbstverständnis der Lehr- und Leitungspersonen und auf die Bereitschaft zur individuellen Nutzung formaler, nonformaler und informeller Lernangebote ist erkennbar. Konkurrenz- und Legitimationsdruck unter Lehrpersonen kann entstehen, indem diese zum Beispiel Ergebnisse aus Vergleichsarbeiten weniger dafür nutzen, um den Unterricht weiterzuentwickeln, sondern um sich in der schulischen Öffentlichkeit zu rechtfertigen. Nicht zuletzt berichten Autoren als Folge der neuen Steuerung von einer strukturellen Entmachtung der Lehrerprofession durch andere Akteure, wie z.B. durch die Schulinspektion, durch Netzwerke, durch kooperative Bildungssettings oder auch durch Eltern. (Vgl. u. a. Altrichter 2009, Wissinger 2011)

Meines Erachtens bietet die Governance-Perspektive eine große Chance, dass sich die Akteure eben nicht entmachten lassen, sondern dass sie auf ihrer Ebene ihre Gestaltungs- und Handlungsmöglichkeiten dazu nutzen, um gezielt an der Verbesserung der schulischen und unterrichtlichen Qualität ebenenspezifisch mitzuwirken. Da Prozesse der Qualitätssicherung und Qualitätsentwicklung auf allen Ebenen des Bildungssystems angestrebt werden, sehe ich als wichtige Voraussetzung für das Gelingen, dass alle Akteure des Systems über die Konzeption und Zielsetzung der unterschiedlichen Maßnahmen und deren Zusammenwirken ausreichend informiert sind. Dies wird auch bereits in der Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring (2006) deutlich hervorgehoben:

> „Damit die verschiedenen Maßnahmen die notwendigen Impulse zur Verbesserung des Bildungswesens auch tatsächlich auslösen, ist es erforderlich, Prozesse der Qualitätsentwicklung und Standardsicherung auf allen Ebenen, von der einzelnen Schule bis zum gesamten Bildungssystem, systematisch umzusetzen und miteinander zu verbinden. Insbesondere muss sichergestellt werden, dass Informationen über die Qualität des Bildungssystems so weit wie möglich auch für die Entwicklung jeder einzelnen Schule genutzt werden können.“ (ebd.; S.6)

Ob es bisher tatsächlich gelungen ist, alle Beteiligten sowohl über die einzelnen Maßnahmen der KMK als auch über länderspezifische Programme, deren Konzeption und Zielsetzungen ausreichend zu informieren, muss aus meiner Sicht in Frage gestellt werden und könnte Gegenstand weiterer Untersuchungen darstellen. Ob den Akteuren deutlich wird, dass auf allen Ebenen systematisch Instrumente implementiert wurden, die sich teilweise aufeinander beziehen und zusammenspielen, darf bezweifelt werden. Der Grund für die mangelhafte Transparenz könnte u. a. an der föderalen Struktur des deutschen Bildungssystems liegen. Bildung ist Ländersache, die Strukturen des eigenen Bundeslandes stehen im Zentrum des Interesses. Ein Blick über die Landesgrenzen hinaus wird von der Mehrzahl der Beteiligten selten vorgenommen, somit werden auch die zentrale Rolle der Kultusministerkonferenz und ihre Gesamtstrategie womöglich von den Wenigsten wahrgenommen. Ebenso wenig ist den Akteuren die Bedeutung und Einflussnahme regionaler Strukturen bewusst, die einen wichtigen Beitrag bei der Einrichtung und Gestaltung spezifischer Bildungslandschaften leisten können.

Auch wenn den jeweiligen Akteuren möglicherweise der Gesamtüberblick fehlt, so ist es umso wichtiger, dass die Personen jene Maßnahmen kennen, die sich auf ihre Ebene beziehen. Im vorliegenden Forschungsprojekt zur Implementierung der baden-württembergischen Vergleichsarbeiten stellt nach offiziellen Vorgaben sowohl die Ebene des Unterrichts als auch die Ebene der Einzelschule die relevante Bezugsgröße dar. Vergleichsarbeiten und die daraus gewonnen Erkenntnisse richten sich vornehmlich an die Lehrerinnen und Lehrer als zentrale Hauptakteure und Hauptverantwortliche für eine gelingende Implementierung und Umsetzung.

Aus diesem Grund setzt sich diese Arbeit mit der empirischen Erforschung der Rezeption und Nutzung der Ergebnisse aus Vergleichsarbeiten auseinander. Mit Hilfe geeigneter Modelle kann der idealtypische Rezeptions- und Nutzungsprozess dargestellt und die tatsächliche Wirksamkeit empirisch überprüft werden. Das Modell von Helmke und Hosenfeld (2005) halte ich für geeignet, um den Prozess differenziert darzustellen, weil es die einzelnen Phasen schrittweise fokussiert. Dadurch können detaillierte Aussagen gemacht werden, an welchen Stellen die Nutzung gelingt, bzw. an welchen Punkten sich konkrete Problemstellen und Schwierigkeiten auftun.

Das Modell von Visscher und Coe (2002) enthält Ergebnisse aus zahlreichen Rezeptionsstudien und listet die wichtigsten Merkmale und Bedingungen auf, die für eine wirksame Nutzung von Rückmeldungen von Bedeutung sind. Aus diesem Grund liefert es meiner Meinung nach zentrale Kriterien, die sich zur Reflexion und Bewertung von Forschungsergebnissen aus Rezeptionsstudien eignen.

Die Diskussion und Darstellung der zentralen Ergebnisse erfolgt in Anlehnung an das Modell (Kap. 3.4.3.2) und in der Systematik des Ergebniskapitels (Kapitel 5).

6.2 Individuelle Bedingungen am Beispiel der Funktion von Vergleichsarbeiten

Zu Beginn des Interviews wurden die Lehrerinnen und Lehrer mit der Frage konfrontiert, aus welchen Gründen ihrer Meinung nach das Ministerium wohl die Vergleichsarbeiten (VA) durchgeführt hat. Ziel dieser Frage war, die Einstellungen und Haltungen der Lehrenden zum neuen Qualitätsinstrument zu erfassen. Aus den Antworten ließen sich insgesamt fünf Kategorien bilden, die im Folgenden nach Häufigkeit der Nennungen diskutiert werden.

Durch die ersten beiden Antwortkategorien „VA dienen landesweitem Vergleich um Standards zu überprüfen und zu sichern“ und „als Rückmeldungen für Verbesserungen“ wird deutlich, dass die Lehrenden in den ersten beiden Schuljahren, in denen Vergleichsarbeiten verpflichtend im baden-württembergischen Schulsystem eingeführt wurden, durchaus eine professionelle Perspektive im Hinblick auf die Funktion von Vergleichsarbeiten entwickeln konnten.

Beinahe alle Befragten geben an, dass Vergleichsarbeiten dazu dienen, als summatives, bilanzierendes Evaluationsverfahren zu überprüfen, ob die Bildungsstandards erreicht wurden. Diese Einstellung entspricht den offiziellen Zielsetzungen des Landes Baden-Württemberg, wie sie im Schulgesetz § 114 verankert sind. Demnach dienen Vergleichsarbeiten zum einen der empirisch gesicherten, zielgerichteten und systematischen Qualitätsentwicklung. Zum anderen stellen Vergleichsarbeiten allen Beteiligten (Lehrenden, Schülerinnen, Schülern und Eltern) objektive Informationen über den individuellen Lernstand der einzelnen Schülerinnen und Schüler in ausgewählten Kompetenzbereichen bereit. Die Erkenntnis-

se, die Lehrende aus den Rückmeldungen gewinnen, sollen demnach zur Förderung einzelner Schülerinnen und Schüler verwendet werden. Diese zweite „offizielle Zielsetzung“ teilt insgesamt nur ein Drittel der befragten Lehrpersonen. Sie sind der Ansicht, dass die Ergebnisse aus Vergleichsarbeiten als Rückmeldung verwendet werden können, um Verbesserungen, im Sinne von formativ-diagnostischen Evaluationsverfahren zu erzielen. Dieses Ergebnis wird durch quantitative Erhebungen gestützt. Maier (2009b; S. 249) berichtet: „Im Jahr 2007 haben nur etwa über 30% der befragten Lehrkräfte durch die Tests zusätzliche Diagnosehinweise erhalten“. Man kann sich vorstellen, dass sich die Verantwortlichen von Vergleichsarbeiten mit diesem moderaten Ergebnis nicht zufrieden geben können. Schließlich handelt es sich bei „Diagnose“ und „Förderung“ um zentrale schulische Bereiche, die einen wichtigen Beitrag zur Qualitätssicherung und Qualitätsentwicklung sowohl auf der unterrichtlichen Ebene als auch auf schulischer Ebene leisten können.

Wenn man die beiden Kategorien betrachtet, kommt hier bereits die doppelte Zielsetzung der Vergleichsarbeiten nämlich „Messen“ und „Entwickeln“ zum Ausdruck (vgl. Kap. 3.1.2.1; Kritikpunkt: Überfrachtung der Zielsetzungen und Funktionen von Vergleichsarbeiten). Die Lehrenden sehen den Schwerpunkt der Vergleichsarbeiten eindeutig im Bereich des Messens und Überprüfens der erreichten Bildungsstandards, als summative Evaluation, was sicher auf den gewählten Zeitpunkt, am Ende eines Bildungsabschnittes zurückzuführen ist. Nur ein geringer Anteil der Befragten bringt die Förderung der Schülerinnen und Schüler mit Vergleichsarbeiten in Zusammenhang.

Neben diesen „systemkonformen“ Einstellungen schreiben die Lehrenden Vergleichsarbeiten jedoch auch andere Funktionen zu, die nicht im Sinne der Verantwortlichen gedacht sind. Die Hälfte der Lehrerinnen und Lehrer ist der Meinung, dass Vergleichsarbeiten als Kontroll- und Druckmittel verwendet werden. Dieser Befund deckt sich mit den Ergebnissen aus der quantitativen Studie zur Rezeption und Nutzung baden-württembergischer Vergleichsarbeiten, die von Maier (2008a, 2008b, 2009b) durchgeführt wurde (siehe Kap. 3.4.1.2.1). Dass das Gefühl von Kontrolle und Druck eine Entwicklung und Verbesserung hemmen können und Lehrende mit Ablehnung reagieren, formuliert Gathen (2006). Auch van Ackeren und Bellenberg (2004, S. 158) weisen darauf hin, dass die Sorgen und Ängste der

Lehrenden vor Kontrolle durch Vergleichsarbeiten ernst genommen werden müssen. Hier sei bereits bei der Implementierung Rücksicht darauf zu nehmen. Allerdings liefern die Autorinnen hierauf keine konkrete Antwort, welche Vorkehrungen oder Maßnahmen im Vorfeld getroffen werden sollten, um die Ängste und Sorgen abzubauen bzw. gar nicht erst entstehen zu lassen.

Meiner Meinung nach muss bei der Suche nach einer Antwort das gesamte Bildungssystem in den Blick genommen werden. Die Implementierung von Vergleichsarbeiten bezieht alle Ebenen mit ein (vgl. Kap. 3.3.2, Gelingensbedingungen für die Implementierung). Somit kommen in der Mehrebenenperspektive weitere Akteure in Betracht, die der Entstehung von Druck und Kontrolle entgegenwirken können.

Auf der Ebene der Einzelschule nimmt z. B. die Schulleitung eine zentrale Rolle ein (vgl. Kohler 2005; Klug und Reh 2000). Mit zunehmender Eigenverantwortung der Schule steigt auch die Verantwortung der Schulleitung, indem sie die notwendigen organisatorischen, strukturellen und kommunikativen Voraussetzungen sicherstellt und für eine tragfähige Feed-back-Kultur sorgt (vgl. Kap. 3.3.2.2.1; Rolle der Schulleitungen). Sie muss die Zielsetzung der Vergleichsarbeiten und anderer Evaluationsverfahren zunächst in allen Gremien und bei allen Beteiligten transparent kommunizieren und ein Evaluationskonzept erarbeiten (vgl. Gathen 2006, Rolff 2002b). Sie kann Unterstützung bei der Nutzung der Rückmeldedaten anbieten und sollte sowohl gegenüber Fragen und Anregungen als auch gegenüber Sorgen und Ängste der Lehrenden offen sein. Außerdem könnte die Schulleitung eine Evaluationsstelle einrichten, deren Mitarbeiter als Ansprechperson/en und Prozessbegleitung alle Evaluationsverfahren professionell betreuen.

Des Weiteren können innerhalb des Kollegiums Teams gebildet werden, die sich gemeinsam mit den zurückgemeldeten Ergebnissen befassen und mögliche Schritte gemeinsam verantworten (vgl. Kapitel 3.3.2.2.2; Rolle der Fachgruppen bzw. Fachkonferenzen). Somit lastet die Erarbeitung von geeigneten Maßnahmen nicht auf einer einzelnen Lehrperson, die als Einzelkämpfer/in allein für die Umsetzung verantwortlich ist. Sondern es besteht die Möglichkeit mit beteiligten Akteuren im Rahmen des kollektiven und kooperativen Austausches mögliche Schritte gemeinsam auszuhandeln und Maßnahmen durchzuführen. Diese Einrichtung von Fach-

teams oder Fachkonferenzen nehmen bei der Umsetzung von Evaluationsergebnissen für die Unterrichtsentwicklung eine Schlüsselrolle ein. Auch hier trägt die Schulleitung eine große Verantwortung, da sich die Zusammenarbeit in den Fachteams an klaren Zielvorgaben ausrichten soll, deren Einhaltung regelmäßig überprüft werden muss.

Einige Autoren sehen die unterstützende Begleitung auch als zentrale Aufgabe der Schulaufsicht oder der Schulbehörde an, die mittels zahlreicher Unterstützungs- und Fortbildungsangeboten den Lehrenden mit Rat und Tat beiseite stehen (vgl. Hoffmann-Göttig, et. al., 2005), so dass diese sich weder unter Druck gesetzt, noch kontrolliert fühlen (vgl. Kapitel 3.3.2.1.3; Unterstützungssysteme). Insgesamt betrachtet kommen der administrativen Ebene bei der effektiven Umsetzung von Evaluationsergebnissen mehr beratende und qualitätssichernde, als kontrollierende Aufgaben zu (vgl. Kapitel 3.3.2.1.2; Wandel der Rolle der Schulaufsicht). Der Wunsch nach externer Unterstützung wird in den Interviews von einigen Lehrerinnen und Lehrern geäußert, vor allen Dingen dann, wenn es um konkrete Hilfestellungen bei der Auswertung von Vergleichsarbeiten geht. Außerdem wünschen manche Lehrpersonen ein breiteres Fortbildungsangebot zum Thema Umgang mit Rückmeldedaten. Zuletzt können von dieser Stelle auch unterstützende Angebote erwartet werden, die den Schulleitungen bei der Bewältigung ihrer Arbeitsaufgaben zu Gute kommen.

Ein Viertel der befragten Lehrpersonen gibt einen weiteren Grund für die Durchführung von Vergleichsarbeiten an. Sie sehen die Einführung des Instruments als direkte Reaktion auf die PISA-Studie. Diese Sichtweise bezieht sich auf die Entstehungsgeschichte von Vergleichsarbeiten. Sie wurden als eine Maßnahme der Qualitätssicherung und Qualitätsentwicklung im Zuge auf die moderaten Ergebnisse bei internationalen Schulleistungsstudien sukzessive im Bildungssystem implementiert und stellen somit ein Instrument der KMK- Gesamtstrategie zum Bildungsmonitoring dar (vgl. KMK 2006). Diese Lehrerinnen und Lehrer nehmen mit ihrer Aussage die geschichtlichen Hintergründe wahr und bleiben in ihren Erklärungsmustern relativ objektiv, weil sie ihre Aussagen kaum normieren.

Insgesamt fünf Personen schreiben Vergleichsarbeiten keine Funktion zu, sondern werten diese als „Beschäftigungstherapie“, der jegliche weitere Sinnhaftigkeit fehlt. Diese Perspektive steht der Entwicklung einer professionellen Evaluations-

kultur entgegen. Kohler (2005) führt in einem Punkt aus, dass Lehrende zum einen aufgrund erhöhter Arbeitsbelastung, zum anderen aufgrund des fehlenden professionellen Verständnisses für Evaluation, jegliche Verfahren ablehnen. Bei den Lehrpersonen, die Vergleichsarbeiten lediglich als Beschäftigungsmaßnahme betrachten, kommt eine generelle Ablehnung und negative Haltung zum Ausdruck. Sie führen keine konkreten Gründe an, was sie daran stört. Eventuell fühlen sie sich durch den Aufwand, den die Durchführung und Auswertung mit sich bringen, zusätzlich belastet. Ihnen fehlt das Verständnis für dieses Instrument, vermutlich ist ihnen auch die offizielle Zielsetzung nicht bekannt. Sie wurden nicht gefragt, ob sie in ihren Klassen Vergleichsarbeiten durchführen möchten, sondern sie wurden dazu verpflichtet, im Sinne eines Top-Down-Steuerungsinstruments. Man muss davon ausgehen, dass diese Lehrerinnen und Lehrer weder die Ergebnisrückmeldungen als Anlass für Verbesserungen verwenden noch irgendwelche Überlegungen zur Bedeutung und Verwendung von Vergleichsarbeiten anstellen werden.

Eine ähnliche abwertende Haltung kommt auch in der nächsten Kategorie zum Ausdruck. Drei Lehrpersonen betrachten Vergleichsarbeiten als puren Aktionismus und richten ihre Kritik an die Verantwortlichen der Bildungspolitik und Bildungsverwaltung. Sie bemängeln, dass ein effektives und umfangreiches Konzept bisweilen fehle, um die bestehenden Defizite und Mängel, die im Bildungssystem vorherrschen, nachhaltig beheben zu können. Sie wünschen sich detaillierte, ausdifferenzierte Instrumente, die Einzelanalysen ermöglichen. Diese Personen haben einen hohen Anspruch an ein Qualitätsmanagement, erwarten professionelle Arbeit und sind mit derzeitigen Verfahren in keiner Weise einverstanden. Um diesen Anforderungen gerecht zu werden, muss bei zukünftigen Vergleichsarbeiten sichergestellt werden, dass eine gewissenhafte, professionelle Entwicklungs- und Pilotierungsarbeit gewährleistet wird (vgl. Bonsen et al. 2004a; S. 134), die den Zielsetzungen von Vergleichsarbeiten entsprechen. Außerdem müssen grundlegende methodische Anforderungen erfüllt sein (vgl. Kapitel 3.3.2.1.1; Anforderungen an Tests). Meiner Meinung nach muss man diese Äußerungen dann ernst nehmen, wenn Vergleichsarbeiten tatsächlich den grundlegenden Anforderungen nicht genügen. Es muss sichergestellt werden, dass Vergleichsarbeiten valide, reliable und aktuelle Daten erzeugen und die Eigenschaften der Rückmeldesysteme

im Sinne des SPFS-Modells nach Visscher & Coe (vgl. Kpl. 3.4.3.1) unbedingt berücksichtigt werden.

6.3 Rezeption von Vergleichsarbeiten

Unter dem ersten Schritt „Rezeption von Vergleichsarbeiten“, der sich im idealen Fall an die Rückmeldungen anschließt, werden diejenigen Aussagen zusammengefasst, die sich den beiden Hauptkategorien „Durchführung der VA“ und „Verständlichkeit der Rückmeldungen“ zuordnen lassen.

Hauptkategorie I: Durchführung von Vergleichsarbeiten

Die Aussagen der Lehrpersonen, die der Kategorie „Durchführung von Vergleichsarbeiten“ zugeordnet werden, lassen sich in drei weitere Subkategorien unterteilen, die im Folgenden die Diskussionsgrundlage bilden: „Keine Probleme bei der Durchführung“, „zur Verfügung stehende Zeit“ und „uneinheitliches Vorgehen“.

Beinahe alle Lehrerinnen und Lehrer (54 von 59) geben an, dass sie bei der Durchführung der Vergleichsarbeiten keine Probleme hatten. Somit ist eine grundlegende Voraussetzung zur Nutzung der Ergebnisse gegeben. Würden bereits bei der Durchführung Schwierigkeiten und Probleme vorherrschen, dann wäre die Nutzung von vornherein gefährdet.

Bei der Bewertung und Einschätzung der zur Verfügung stehenden Zeit, fällt die Antwort der Befragten nicht mehr so eindeutig aus, sondern es lassen sich schulartspezifische Unterschiede erkennen. Während fast alle Lehrerinnen und Lehrer an Gymnasien der Ansicht sind, dass ausreichend Zeit für die Bearbeitung der Aufgaben eingeplant war, teilen diese Meinung nur jeweils die Hälfte der Realschul- und Hauptschullehrpersonen. Beinahe ein Drittel der Realschullehrenden gibt an, dass die Bearbeitungszeit für ihre Schülerinnen und Schüler zu knapp bemessen war. Im Gegensatz dazu meint fast ein Drittel der Hauptschullehrkräfte, dass die Zeit für die Aufgabenbearbeitung für ihre Schülerinnen und Schüler als zu lang empfunden wurde. Dieses Ergebnis stimmt nachdenklich. Denn spätestens seit den Befunden aus den großen Schulleistungsstudien, wie z.B. den PISA-Studien ist bekannt, dass vor allem die schwächeren Schülerinnen und Schüler bei

Vergleichsstudien deutlich schlechtere Leistungen erzielen als ihre Gleichaltrigen aus Realschulen und Gymnasien. Dass ausgerechnet in den Vergleichsarbeiten die schwächste Gruppe zu viel Zeit für die Bearbeitung der Aufgaben zur Verfügung hatte, kann sowohl als Hinweis auf eine mangelhafte Testkonzeption als auch auf eine fehlerhafte Pilotierung gedeutet werden. Auch dieser Aspekt richtet sich an Diejenigen, die für die Entwicklung von Vergleichsarbeiten verantwortlich sind, sowohl bei der Pilotierung, als auch bei der Erstellung des Testinstruments deutlich höhere Maßstäbe anzusetzen. Ein anderer Grund, der für den Zeitüberschuss bei der Durchführung der Vergleichsarbeiten in der Hauptschule in Betracht käme, könnte der Umstand sein, dass es den Schülerinnen und Schülern an nötiger Ausdauer und Geduld mangeln könnte. Die Aufgaben werden nur oberflächlich bearbeitet und die Zeit, die für eine vertiefte und gründliche Lösung eingeplant war, wird nicht benötigt. Ergebnisse zur Reflexion von Vergleichsarbeiten, die unter Punkt 6.4 diskutiert werden, versprechen hierzu mehr Aufklärung, welche Gründe die Lehrerinnen und Lehrer der Hauptschulen dafür als Erklärung anführen.

In einer dritten Subkategorie im Bereich „Durchführung der VA“ wird das teilweise uneinheitliche Vorgehen bei der Aushändigung der Vergleichsarbeiten durch die Schulleitung oder einer beauftragten Person massiv kritisiert. Die Angaben der Lehrerinnen und Lehrer streuen in diesem Punkt extrem. Einige berichten, dass die Vergleichsarbeiten bis zu einer Woche vor offizieller Durchführung ausgehändigt wurden und die Lehrenden somit ausreichend Zeit für eine spezifische Vorbereitung ihrer Schülerinnen und Schüler hatten. Andere hingegen beklagen den Zeitdruck der durch die knappe, kurzfristige Ausgabe der Arbeiten am Morgen der Durchführung entstand. So fehlte diesen Lehrkräften – um sich gut vorbereitet zu fühlen - die Zeit. Einerseits, um die Anzahl der Arbeiten zu kontrollieren, andererseits um sich einen kurzen Überblick über die Texte und das Anforderungspotential der Vergleichsarbeiten verschaffen zu können. Im Hinblick auf die Testgütekriterien wird bei diesem uneinheitlichen Vorgehen die Durchführungsobjektivität der Vergleichsarbeiten massiv verletzt und die Ergebnisse eignen sich nicht für objektive Leistungsvergleiche.

Im ersten Fall wurden die Vergleichsarbeiten bis zu einer Woche vorher an die Lehrenden ausgehändigt und es wird berichtet, dass die Schülerinnen und Schüler

gezielt spezifische Aufgaben eine Woche lang einüben konnten (vgl. Kapitel 3.1.2.6; Kritikpunkt: „teaching to the test“). Generell steht die beschriebene Vorgehensweise der eigentlichen Zielsetzung von Vergleichsarbeiten entgegen, nämlich die erreichten Kompetenzen der Schülerinnen und Schüler am Ende eines Bildungsabschnittes zu erfassen. Durch das Verfahren des ‚teaching to the test’ werden die Ergebnisse verfälscht und können für einen objektiven Leistungsvergleich nicht herangezogen werden.

Ein weiteres Problem des ‚teaching to the test’ stellt die Tatsache dar, dass alle Unterrichtszeit nur zur Testvorbereitung verwendet wird und wichtige pädagogische Maßnahmen in den Hintergrund rücken.

Im zweiten geschilderten Fall wurden die Vergleichsarbeiten so kurzfristig vor der Durchführung der Vergleichsarbeiten an die Lehrenden ausgehändigt, dass diese aufgrund der oben aufgeführten Gründe unter Druck gerieten. Sie waren teilweise extrem gestresst und gehetzt, was sie als inhuman und unverantwortlich bewerteten und zu einer negativen Einstellung gegenüber Vergleichsarbeiten führen könnte.

Ein weiterer kritischer Punkt der von diesen Personen genannt wird, betrifft die Hierarchisierung und erlebte Entmachtung, die von Seiten der Schulleitung bei der Ausgabe der Vergleichsarbeiten auf die Lehrenden einwirkt. Denn nur den Leitungspersonen war es erlaubt, die Vergleichsarbeiten im Vorfeld zu vervielfältigen und kurz vor der Durchführung an die zuständigen Lehrerinnen und Lehrer auszuhändigen. Dieser Punkt kann zu weitreichenden Problemen im Kollegium führen. Einerseits bewerten die Lehrerinnen und Lehrer dieses Vorgehen als mangelndes Vertrauen der Schulleitung, denn es kann dadurch der Eindruck entstehen, dass sich nur Leitungspersonen an zentrale Vorgaben halten, während alle anderen in diesem Bereich „unzuverlässig“ sind. Andererseits gilt es zu bedenken, dass Schulleiterinnen und Schulleiter über dieselbe Ausbildung verfügen wie alle Lehrenden einer Schulart und somit nicht professioneller handeln. Diese erlebte Hierarchie könnte zu äußerst negativen Stimmungen im Kollegium führen und dadurch das Entwicklungspotenzial der verschiedenen Akteure auf der Ebene der Einzelschule hemmen, da es zu keiner Aushandlung gemeinsamer Zielsetzungen kommt.

Die Gründe dafür, warum Vergleichsarbeiten im zweiten Jahr der verpflichtenden Durchführung nur von der Schulleitung ausgehändigt werden dürfen, könnten darin liegen, dass in Baden-Württemberg einige Pannen bei der Durchführung durch die Medien bekannt wurden. In einem Fall wurden im Vorfeld der Vergleichsarbeiten Aufgaben im Internet veröffentlicht. In einem anderen Fall wurde in einer Schule eine Vergleichsarbeit auf dem Kopierer vergessen. Beide Fälle führten landesweit zu einer offiziellen Absage der Vergleichsarbeit im betroffenen Fach und zu Negativschlagzeilen.

Als Lösung für die beiden beschriebenen Extrembeispiele der Aushändigung könnte meines Erachtens die Variante dienen, die von einer dritten Gruppe berichtet wird. Dort wurde die Vergleichsarbeiten am Vortag, nach Schulschluss, an die beteiligten Lehrerinnen und Lehrer ausgeteilt, so dass diese genügend Zeit zur Auseinandersetzung hatten und nicht das Gefühl des Drucks oder des „Nicht–ausreichend-vorbereit-seins“ vorherrschte. Meines Erachtens könnte man den Lehrenden zu dem Zeitpunkt Einsicht in die Vergleichsarbeiten gewähren, evtl. mit Unterstützung durch Evaluationsbeauftragte. Die Arbeit sollte jedoch danach bis zum Tag der Durchführung unter Verschluss an der Schule bleiben.

Eine weitere Alternative könnte auch die frühzeitige Einsicht am Morgen der Durchführung darstellen, damit den Lehrenden genügend Zeit zur Einstimmung bleibt und sie entspannt die Arbeit aufnehmen und nicht unter Druck geraten. Dieses Vorgehen trägt eher zu einer positiven Einstellung bei und sorgt dafür, dass keine negative Stimmung gepaart mit Misstrauen gegenüber der Schulleitung aufkommen kann.

Wie weit die Lehrerinnen und Lehrer jedoch ihre Vorinformationen dazu nutzen, während der Erarbeitung der Vergleichsarbeiten ihren Schülerinnen und Schüler Hilfestellungen zu geben, bleibt fraglich. Hier kann man nur auf die Professionalität der Lehrkräfte hoffen und im Vorfeld z. B. alle Beteiligten ausführlich über die Zielsetzung und Funktion von Vergleichsarbeiten aufklären. Außerdem müssen klare Absprachen getroffen werden, wie sich die Lehrenden während der Durchführung der Vergleichsarbeiten verhalten müssen, um u.a. die Durchführungsobjektivität nicht zu verletzen. Eine Möglichkeit um für vergleichbare Bedingungen zu sorgen sehe ich auch in diesem Punkt in Fortbildungen, die durch entsprechend geschultes Fachpersonal angeboten werden.

Hauptkategorie II: Verständlichkeit der Rückmeldung

Unter dieser Hauptkategorie lassen sich die Aussagen der interviewten Personen nach der Häufigkeit der Nennungen den drei Subkategorien „Rückmeldungen waren verständlich“, „Rückmeldungen wurden nicht angeschaut“ und „Rückmeldungen waren unverständlich/unpraktisch“, zuordnen.

Insgesamt äußert sich nur jeder / jede Vierte aller Befragten dazu, dass die Rückmeldungen verständlich waren. Ihrer Meinung nach hängt die Verständlichkeit zum einen an der einfachen Darstellung der Auswertungsmatrix und den Prozentangaben, die generell keine Probleme beim Verstehen bereiten. Zum anderen sind Lehrende sowohl den Umgang mit Tabellen und Grafiken als auch mit dem Computer gewohnt, was als wichtige Voraussetzung für die Nutzung von Rückmeldungen gilt (vgl. Schrader und Helmke 2004). Hier darf nochmals kritisch darauf hingewiesen werden, dass diese Position nur von ca. einem Viertel der Interviewten vertreten wird, was in der Gesamtbetrachtung einen sehr kleinen Teil ausmacht. Vor allem wenn man bedenkt, dass Rückmeldungen im optimalen Fall die Ausgangsbasis aller weiteren Schritte bildet. Weiterhin gilt zu bedenken, dass auch eine hohe Verständlichkeit von Rückmeldungen nicht automatisch zu deren Verwendung und Nutzung führt, wie zahlreiche Autoren belegen (vgl. Kap. 3.4.1). Die Ergebnisse dieser Interviewstudie können zur Konkretisierung dieser Befunde beitragen:

So berichten drei Lehrpersonen darüber, dass sie trotz der optimalen und übersichtlichen Darstellung der Rückmeldungen diese nicht benutzen, obwohl sie viel damit hätten anfangen können. Dieser Befund zeigt, dass die Nutzung sowohl von der individuellen Bereitschaft der einzelnen Person abhängt, als auch von strukturellen und unterstützenden Systemfaktoren, die dieses Potenzial aufgreifen müssen, um die Lehrerinnen und Lehrer bei weiteren Maßnahmen und Entwicklungsschritten anzuleiten und zu begleiten.

Die zweite Subkategorie „Rückmeldungen wurden nicht angeschaut“ wird ebenfalls von etwa einem Viertel der Lehrerinnen und Lehrer genannt. Als Grund für die Nichtbeachtung der Rückmeldedaten wird in allen Interviews das Desinteresse angeführt, wobei die Ursachen hierfür unterschiedlich angegeben werden.

Die einen beklagen die fehlende Zeit oder führen als Grund an, dass ihre Klasse demnächst von einer anderen Lehrperson weitergeführt wird und ihnen die zurückgemeldeten Daten nichts bringen. Die Ursachen sind demnach konzeptioneller Art und richten sich an die Bildungsadministration und Bildungspolitik. Andere klagen darüber, weitaus wichtigere Probleme in ihrer Klasse zu haben, so dass Vergleichsarbeiten für sie keine Relevanz besitzen. Oder es wird generell die Aussagekraft von Vergleichsarbeiten in Frage gestellt. Wieder andere machten im Vorjahr negative Erfahrungen bei der Nutzung von Vergleichsarbeiten, so dass dadurch ihr Desinteresse zustande kam. Diese Aussage belegt, dass negative Erfahrungen, die man bei der Durchführung von Vergleichsarbeiten evtl. sammelt, nachhaltige negative Auswirkungen auf die zukünftige Rezeption und Nutzung dieses Instruments haben können. Daher ist es von entscheidender Bedeutung, dass jede Durchführung durch ein breites Unterstützungssystem begleitet wird. Bei einigen Aussagen wird auch deutlich, dass die Nutzung in Abhängigkeit zu den erzielten Ergebnissen der Schülerinnen und Schüler steht. Fallen diese zufriedenstellend aus, so ist das Interesse an Rückmeldungen eher gering. Dieser Befund wird auch von den Ergebnissen der LAU-7-Studie gestützt (vgl. Klug & Reh 2000).

Ein eklatantes Ergebnis für die Nichtnutzung der Rückmeldungen stellt die Aussage einer Lehrkraft dar. Sie berichtet, dass an ihrer Schule ein Beschluss existiert, der besagt, dass die Vergleichsarbeiten weder eingegeben noch ausgewertet werden müssen. Demnach spielen die Ergebnisse aus Vergleichsarbeiten an dieser Schule keine Rolle, weil sie von vornherein nicht verwendet werden. Dass Schulen in dieser Form eigenmächtige Beschlüsse fassen, obwohl Vergleichsarbeiten landesweit gesetzlich verpflichtend implementiert wurden, kann meiner Meinung nach nur durch übergeordnete Ebenen z.B. durch die Schulaufsicht oder durch externe Evaluationsverfahren überwacht und reglementiert werden. Um herauszufinden, ob die Intensität der Nutzung und Verwendung von Qualitätsinstrumenten auf der Ebene der Einzelschule von externen Evaluationsverfahren abhängen, ist weitere Forschung nötig.

Erste empirische Untersuchungen, die von Studierenden aus höheren Semestern im Rahmen von Hauptseminaren unter meiner Leitung durchgeführt wurden, zeigen, dass eine bevorstehende Schulinspektion vor allem die Schulleitung dazu an-

regen kann, alle erforderlichen und verpflichtenden Maßnahmen zur Qualitätssicherung und Qualitätsentwicklung gewissenhaft und gründlich durchzuführen, zu dokumentieren und das Kollegium zur aktiven Mitarbeit zu verpflichten. In diesem aktuellen Forschungsbereich zu externen Verfahren, wie z.B. die „Schulinspektion“, können in nächster Zeit einige Forschungsergebnisse erwartet werden.

In der dritten Subkategorie zur Verständlichkeit der Rückmeldungen werden diejenigen Aussagen der Lehrerinnen und Lehrer gebündelt, die der Meinung sind, dass die Rückmeldungen unverständlich und unpraktisch sind. Diese Aussage teilen insgesamt nur fünf Lehrpersonen; drei aus der Real- und zwei aus der Hauptschule. Sie bewerten die Darstellung als unverständlich, umständlich, undifferenziert und dadurch unbrauchbar. Als konkrete Punkte werden die symbolische Verwendung von Ausrufezeichen und Fragezeichen als nichtssagend kritisiert: Ausrufezeichen weisen auf eine überdurchschnittliche, Fragezeichen auf eine unterdurchschnittliche Leistung hin. Außerdem kamen einige Lehrende nicht mit den prozentualen Angaben und den Balkendiagrammen zurecht.

Zuletzt wird berichtet, dass vor allem ältere Kolleginnen und Kollegen Verständnisprobleme aufweisen, weil sie die Diagramme nicht lesen können.

Bei diesen Punkten wird deutlich, dass dieser Personenkreis Hilfe und Unterstützung bei der Rezeption und Interpretation von Rückmeldungen aus Vergleichsarbeiten benötigt. Wie bereits mehrfach diskutiert wurde, böte sich auch für diese Problembereiche die Einrichtung eines breiten Unterstützungsangebotes an, das auf allen Ebenen des Bildungssystems die nötigen Ressourcen bereitstellt.

6.4 Reflexion von Vergleichsarbeiten

Den zweiten Schritt im Modell von Helmke & Hosenfeld bildet die Reflexion und somit die Suche nach möglichen Ursachen und Gründen, die die entsprechenden Vergleichsergebnisse erklären können. Damit ist die weitere Vorstellung verbunden, dass sich aus diesen Überlegungen geeignete Maßnahmen ableiten lassen, mit dem Ziel, Verbesserungen zu erreichen.

Die Reflexionen, die die befragten Lehrerinnen und Lehrer anstellen, lassen sich thematisch den fünf Hauptkategorien zuordnen; I) Auswertung der Vergleichsar-

beit, II) Aufgaben und Zusammenstellung der Vergleichsarbeit, III) Niveau der Vergleichsarbeit, IV) Zeitpunkt der Vergleichsarbeit und zuletzt V) Förderdiagnostischer Nutzen der Vergleichsarbeit.

Betrachtet man die Themengebiete der Hauptkategorien so stellt man fest, dass sich die ersten vier Bereiche auf die Konzeption und Konstruktion der Vergleichsarbeiten beziehen und somit an die übergeordnete Ebene im System, an die Verantwortlichen der Vergleichsarbeiten, adressiert sind. Lediglich die fünfte Kategorie „ Förderdiagnostischer Nutzen der Vergleichsarbeit“ betrifft den Gestaltungsbereich der Lehrenden auf Schul- und Unterrichtsebene und bezieht sich somit auf deren direkten Handlungskontext.

Im weiteren Vorgehen erfolgt nun die Diskussion der Ergebnisse der einzelnen Hauptkategorien:

Reflexion – Hauptkategorie I) Auswertung der Vergleichsarbeit

Etwa 2/3 aller Lehrerinnen und Lehrer üben Kritik an der Bewertungspraxis der Vergleichsarbeiten. In dieser Subkategorie sind die Lehrkräfte aus Realschulen und Gymnasien besonders stark vertreten.

Im Zentrum der Kritik stehen die ungenaue und nicht differenzierende Punkteverteilung und der ungerechte Notenschlüssel, der auf einer nichtlinearen Skalierung basiert und als Auswertungsvorschrift verpflichtend vorgeschrieben wird. In den meisten Aussagen kommen Unverständnis und Willkür, teilweise auch Ärger und Wut deutlich zum Ausdruck, weil die Lehrenden verpflichtet sind, sich bei der Auswertung an diese verbindlichen Vorgaben zu halten. Dabei gilt es zu bedenken, dass sich diese negative Gefühle ungünstig auf die Einstellung und Akzeptanz gegenüber Vergleichsarbeiten auswirken können.

Die Reaktionen der Lehrerinnen und Lehrer auf die ungenauen Punkteverteilungen fallen in diesem Bereich sehr unterschiedlich aus. Während sich einige Lehrpersonen bei der Korrektur strikt an die zentralen Vorgaben halten, obwohl sie das Vorgehen negativ bewerten, gibt es andere, die sich bewusst über die Vorgaben hinwegsetzen. Sie vergeben für Teilleistungen zusätzliche Punkte und bewerten die Leistungen individuell im Sinne ihrer Schülerinnen und Schüler, weil sich das vorgeschriebene strenge Verfahren nicht mit ihrem Gewissen vereinbaren lässt. Eine dritte Gruppe erarbeitet bei unklaren Vorgaben mit den Kolleginnen und

Kollegen gemeinsame, alternative Lösungsvorschläge, die dann für alle Beteiligten als Bewertungsgrundlage gelten.

Die Aussagen der Lehrenden zum ungerechten Notenschlüssel und der fehlenden linearen Skalierung bilden einen weiteren Kategorienbereich. Hier finden sich mit einer Ausnahme nur Beiträge von Realschullehrpersonen. Sie kritisieren die Notenskala, weil diese gültigen Standards der Leistungsmessung nicht entspricht. Die vorgegebene Skala ordnet der Hälfte der Punktzahl die Note 2,5 zu und liegt somit eine ganze Note über sonstigen Bewertungsskalen. Für die Lehrenden ist es absolut unverständlich, dass sich offizielle Vorgaben auf willkürlich festgelegte Grundlagen stützen.

Die Konsequenzen, die sich aus der geschönten Skala ergeben sind vielfältig. Positive Auswirkungen lassen sich vor allem für diejenigen Schülerinnen und Schüler festhalten, die für mittelmäßige Leistungen gute bis sehr gute Noten erzielen. Zu Legitimationsschwierigkeiten kann es in den Fällen kommen, wenn ‚unrealistisch' gute Noten der Vergleichsarbeit - bei drohender Nichtversetzung – doch noch zur Versetzung führen. Die Beurteilungs- und Bewertungskompetenz der Lehrkräfte wird dann grundsätzlich in Frage gestellt. Dies könnte auch erklären, dass sich vor allem Lehrpersonen, die an Realschulen unterrichten, zu diesem Aspekt äußern. An Realschulen findet man häufiger als z.B. an Hauptschulen engagierte Eltern, die sich für die Leistungsergebnisse ihrer Kinder interessieren. Manchen Eltern wurde eventuell auch der Wunsch verwehrt, dass ihrem Kind im Anschluss an die Grundschule eine Gymnasialempfehlung ausgestellt wurde. Wenn nun die Note in der staatlich vorgegebenen Vergleichsarbeit um mindestens eine Note besser ist, als die Jahresleistung, die durch Messverfahren der Lehrperson ermittelt wurde, so können Lehrende gegenüber Eltern in einen extremen Legitimationsdruck geraten. Viele Lehrende lehnen daher Vergleichsarbeiten als Basis für Eltern-Lehrer-Gespräche ab, weil sie aufgrund ihrer mangelhaften konzeptionellen Struktur auch juristisch anfechtbar sind.

Als weiteres Ergebnis in diesem Bereich kann zum einen die generelle Kritik an der Bewertung von Vergleichsarbeiten berichtet werden, weil dies als summative Bilanz dem formativen Nutzen entgegen läuft.

Als einzige positive Subkategorie zum Thema Bewertung von Vergleichsarbeiten können die Aussagen von insgesamt neun Lehrenden zusammengefasst werden. Sie betonen die Praktikabilität der Auswertung, die im Vergleich zu Klassenarbeiten einfacher und schneller von statten geht.

Die Ergebnisse zu der ersten Hauptkategorie „Auswertung der Vergleichsarbeit“ zeigen deutlich, dass die meisten Lehrerinnen und Lehrer hier massive Probleme benennen. Durch die willkürlichen, individuell festgelegten Bewertungsverfahren werden die Testgütekriterien, vor allem die Auswertungsobjektivität extrem verletzt und jegliche Vergleichbarkeit verhindert. Man darf zu Recht Zweifel an der Erstellung der Bewertungsvorgaben anmelden. Es stellen sich Fragen, ob den Verantwortlichen die Schwierigkeiten vor der flächendeckenden Durchführung bewusst waren. Ergaben die Rückmeldungen aus den Pilotierungsarbeiten keine Hinweise auf die mangelhafte, weil undifferenzierte und unrealistische Bewertungspraxis? Sind sich die Verantwortlichen über mögliche Folgen klar, die aus diesen Vorgaben resultieren können? Wenn Lehrerinnen und Lehrer zahlreiche negative Erfahrungen im Umgang mit Vergleichsarbeiten berichten, die auf eine ungründliche Testentwicklung hinweisen, muss in den Folgejahren damit gerechnet werden, dass sich sowohl die Einstellung und Haltung als auch die Akzeptanz gegenüber Vergleichsarbeiten ins Negative kehren. Diese Entwicklung könnte meines Erachtens durch professionelle Testkonzeptionen verhindert werden.

Einige Lehrkräfte stellen die Bewertung von Vergleichsarbeiten grundsätzlich in Frage. Dem Wunsch, die Benotung abzuschaffen, kommt die Landesregierung mit ihre Verwaltungsvorschrift vom Oktober 2008 nach. Seither werden Vergleichsarbeiten nicht mehr benotet.

Reflexion – Hauptkategorie II) Aufgaben der Vergleichsarbeit

Auch in diesem Bereich finden sich nur kritische Äußerungen, die sich an die Verantwortlichen von Vergleichsarbeiten richten.

So übt die Hälfte der Befragten Kritik an einzelnen Aufgaben- und Fragestellungen. 2/3 der Personen sind mit der Gesamtzusammenstellung der VA nicht einverstanden.

Im Bereich der Aufgaben- und Fragestellungen findet sich in den Interviews eine heterogene Vielfalt an Kritikpunkten. Einige beziehen sich auf das multiple-

choice-Format und auf die schülerfremde Sprache. Ferner finden sich Aussagen, die auf der einen Seite bemängeln, dass die Aufgabenstellungen zu komplex seien. Auf der anderen Seite wird im Gegensatz dazu von anderen beklagt, dass die Aufgaben eine mangelnde Komplexität aufweisen. Weitere Kritikpunkte betreffen u. a. die unübersichtliche oder fehlerhafte Darstellung der Aufgaben.

Bei der Gesamtzusammenstellung der Vergleichsarbeiten wird vor allem bemängelt, dass die Themenbereiche einseitig ausgewählt wurden und wichtige Bereiche größtenteils nicht vorkamen und somit die Bildungsstandards nicht valide abgebildet wurden. Dies führte bei allen Beteiligten (Lehrenden, Schülerinnen und Schülern) zu Unzufriedenheit und Enttäuschung.

Einige Lehrpersonen äußerten den Wunsch nach mehr Transparenz und nach Beteiligung bei der Erstellung der Testinstrumente. Diesem Wunsch wird z.B. bei Schulmonitoringstudien wie VERA, die im Grundschulbereich in allen Bundesländern durchgeführt werden, entsprochen. Dort haben Schulen die Möglichkeit, sich an der Entwicklung neuer Aufgaben zu beteiligen und für die Normierungsstichprobe vorzuschlagen (siehe Kap. 3.2.1; Länderübergreifende Projekte in der Grundschule – das Projekt VERA). Meiner Meinung nach könnte eine aktive Beteiligung der Lehrenden deren Akzeptanz gegenüber Vergleichsarbeiten erhöhen und somit einen wichtigen Beitrag zur Qualitätssicherung und Qualitätsentwicklung liefern. Außerdem könnten die Lehrerinnen und Lehrer im Bereich der Aufgabenentwicklung ihre Kompetenzen ausbauen, vertiefen und weiterentwickeln. Würden sich Vergleichsarbeiten in der Sekundarstufe z.B. aus 50% staatlicher Vorgaben und 50% eigener schulinterner Aufgaben zusammensetzen, würde sicherlich auch die Kritik an einzelnen Aufgaben und Fragestellungen geringer ausfallen. Um die Vergleichbarkeit zu sichern, könnte man sich nur auf die staatlichen Aufgaben beziehen. Die gemeinsame Konzeption mit Kolleginnen und Kollegen würde auch die Kooperation innerhalb der Einzelschule stärken und festigen.

Damit bei der Aufgabenentwicklung grundlegende Standards eingehalten werden, die auch unterschiedliche Kompetenzstufen und Schwierigkeitsstufen berücksichtigen, benötigen die Kolleginnen und Kollegen meiner Ansicht nach Unterstützung und Hilfestellungen durch geschultes Personal. Ob diese Fachpersonen an jeder Einzelschule oder schulübergreifend an den Staatlichen Schulämtern, bzw.

für Gymnasien an den Regierungspräsidien für die Betreuung und Beratung zur Verfügung stehen, muss konzeptionell und strukturell durchdacht werden.

Meines Erachtens müsste auch die Zusammenarbeit zwischen dem Landesinstitut für Schulentwicklung, das für die Konzeption und Entwicklung der baden-württembergischen Vergleichsarbeiten verantwortlich ist und dem Institut zur Qualitätsentwicklung in Berlin stärker forciert werden (vgl. Kap. 2.4.3; Einrichtung von Qualitätsagenturen). Das IQB befasst sich im großen Stil mit der Generierung von Aufgabensammlungen zur Operationalisierung von Bildungsstandards und der Testentwicklung. Durch intensive Zusammenarbeit mit allen Landesinstituten könnte ein reichhaltiges Aufgabenangebot zusammengestellt werden, das auch wichtige Kommentare zur Kompetenzentwicklung enthalten könnte.

Im Zeitalter des Internets dürfte es kein Problem darstellen, wenn die Nutzung dieses Aufgabenpools für die Lehrerinnen und Lehrer frei zugänglich wäre und für die tägliche Unterrichtspraxis verwendet werden könnte.

Reflexion – Hauptkategorie III) Niveau der Vergleichsarbeit

Insgesamt betrachtet kritisiert die Hälfte der Lehrenden das niedrige Niveau der Vergleichsarbeiten. Allerdings treten hier deutliche schulartspezifische Unterschiede hervor. Während beinahe alle Hauptschullehrkräfte dieser Aussage zustimmen, liegt der Anteil in der Realschule etwa bei der Hälfte und im Gymnasium lediglich bei 1/7 der Personen.

Ansonsten wird das „Niveau von Vergleichsarbeiten“ von vier weiteren Lehrerinnen und Lehrern (3 Gymnasial- und 1 Realschullehrperson) als zu hoch und von zwei Gymnasiallehrkräften als angemessen bewertet. Für das hohe Niveau werden nach Meinung der betroffenen Lehrkräfte einmal die umfangreiche Konzeption und ein anderes Mal ein erhöhter Schwierigkeitsgrad im Vergleich zum Vorjahr verantwortlich gemacht.

Das zentrale und auffälligste Ergebnis dieser Hauptkategorie stellt eindeutig die einhellige Meinung aller Hauptschullehrerinnen und –lehrer dar, die das Niveau der Vergleichsarbeit als zu niedrig angeben. Dieser Befund wirft einige Fragen auf, die sich vor allen Dingen an die zuständigen Entwickler der Vergleichsarbeiten richten. Welche Faktoren und Bedingungen führten dazu, dass bei der Konzeption der Tests das Niveau der Schülerinnen und Schüler nicht angemessen be-

rücksichtigt wurde? Welche Erhebungsverfahren und Messinstrumente wurden zur Ermittlung des Niveaus eingesetzt? Sind Fehler bei der Ermittlung und Berechnung des Niveaus zu verzeichnen? Wie setzte sich die Pilotierungsstichprobe zusammen? Bestand diese möglicherweise aus den leistungsschwächsten Klassen Baden-Württembergs und somit weit unterhalb des Leistungsniveaus durchschnittlicher Hauptschülerinnen und Hauptschüler? Über welche Kompetenzen im Bereich der professionellen Testentwicklung verfügten die Verantwortlichen? Sind sich diese darüber bewusst, welche Auswirkungen ein zu niedriges Niveau auf den Rezeptions- und Nutzungsprozess von Lehrerinnen und Lehrer haben kann? Die schulartspezifische Betrachtung der Ergebnisse zeigt an verschiedenen Stellen der Interviewauswertung eine verhaltene Beteiligung der Hauptschullehrkräfte. Die Vermutung liegt nahe, dass dies in Zusammenhang mit dem niedrigen Niveau diskutiert werden muss. Es muss davon ausgegangen werden, dass die Vergleichsarbeiten, die in Hauptschulen durchgeführt wurden, eine geringe Aussagekraft haben und sich nicht als zentrales Instrument der Qualitätsentwicklung und Qualitätssicherung eignen.

Reflexion – Hauptkategorie IV) Zeitpunkt der Vergleichsarbeit

Beinahe ein Drittel aller Befragten gibt an, dass der Zeitpunkt der Durchführung von Vergleichsarbeiten zu spät stattfand. Das führt dazu, dass im direkten Anschluss keine Zeit mehr vorhanden ist, um aus den Rückmeldungen geeignete Maßnahmen abzuleiten und diese auch durchzuführen. Dazu kommt, dass am Ende eines Schuljahres zahlreiche weitere Veranstaltungen, wie Schullandheimaufenthalte oder Projekttage stattfinden und folglich noch weniger freie Zeit für Vergleichsarbeiten zur Verfügung steht. Diesen Zeitdruck erleben einige Lehrkräfte als äußerst negativ und äußern sich dementsprechend wütend und aggressiv in den Interviews. Schließlich bleibt weder Zeit für eine persönliche intensive Auseinandersetzung mit den Daten, noch können die Ergebnisse mit den Schülerinnen und Schülern oder mit den Eltern besprochen werden. Und die Möglichkeit, dass man Maßnahmen oder Rückmeldungen auf den Anfang des nächsten Schuljahrs verlegt, besteht in den meisten Fällen auch nicht, weil die Lehrerinnen und Lehrer ihre Klasse abgeben.

In diesem Punkt wird wiederholt deutlich, dass die Konzeption und Zielsetzung der baden-württembergischen Vergleichsarbeiten nicht kongruent, sondern gegen-

läufig sind. Der Landesregierung kann es nicht um formative Evaluationsformen gehen, sondern lediglich um die Überprüfung und Standardsicherung, sonst würden Vergleichsarbeiten nicht am Ende von zwei Schuljahren durchgeführt werden. Und diese Zielsetzung müsste meiner Meinung deutlicher kommuniziert werden, so dass nicht der Eindruck bei den Lehrerinnen und Lehrern vorherrscht, Maßnahmen zur individuellen Förderung ableiten zu müssen, die konzeptionell nicht durchführbar sind. In meinen Augen ist es unverantwortlich gegenüber den Lehrenden, offizielle Zielsetzungen in die Schulen zu transportieren, die selbst bei höchstem Engagement und Reformwillen, den einige Kolleginnen und Kollegen besitzen, zeitlich nicht realisierbar sind. In diesem Punkt darf man sich nicht wundern, wenn dieser innere Widerspruch zu einer ablehnenden Einstellung und Haltung der Lehrkräfte führt.

Womöglich hat die verantwortliche bildungspolitische und –administrative Ebene diesen massiven Kritikpunkt realisiert, denn ein Jahr nach der Datenerhebung, im Oktober 2008 wurde der Zeitpunkt von Vergleichsarbeiten auf den Beginn eines Schuljahres verschoben; in Zukunft werden in den Klassen sieben, neun und elf Vergleichsarbeiten durchgeführt werden. Dieser Vorschlag wurde auch von vier Lehrpersonen in den Interviews geäußert.

Welche Wirkungen aus dieser revidierten Konzeption resultieren, muss Gegenstand neuer Forschung darstellen. Auch zu dieser Fragestellung wurden von Studierenden der Pädagogischen Hochschule im Rahmen eines Hauptseminars erste Studien durchgeführt. Eine Studentin führte im Anschluss an die Vergleichsarbeiten im Schuljahr 2010/11 Leitfadeninterviews mit drei Lehrenden eines Gymnasiums durch, das bereits an der Untersuchung 2007 teilnahm. Auf die Frage, wie die Lehrenden den neuen Zeitpunkt beurteilen, antworten die beiden Gymnasiallehrer und eine Lehrerin insgesamt positiv und nennen folgende Vorteile:

Durch den frühen Zeitpunkt, der vier Wochen nach Schuljahresbeginn stattfindet, lastet nicht der „wahnsinnige Druck das ganze Schuljahr“ (Lehrerin) auf den Schülerinnen und Schülern. Ihr Kollege findet den Zeitpunkt auch „okay“. Als Gründe führt er vor allem die Nachhaltigkeit der erlernten Kompetenzen und die Diagnosemöglichkeit an:

> „Das [die Vergleichsarbeit, Anmerkung S.J.] soll ja messen, was ein Schüler kann. Und was der Schüler kann, dass soll er nachhaltig können. Also es soll

> sozusagen die Nachhaltigkeit messen. (…). Denn die Nachhaltigkeit von dem, was wir unterrichten ist insgesamt zu schlecht, also bleibt wenig hängen. Und das kann man mit so einer Diagnosevergleichsarbeit eigentlich schon rauskriegen und es sollte eben dann auch den Lehrern und den Schülern helfen, also dass sie dann wissen, da muss man noch was machen.“ (Lehrer 3)

Der andere Kollege nimmt zu dem Thema eine etwas kritischere Position ein, weil er zwei Seiten der Medaille betrachtet. Den Vorteil sieht er darin, dass Vergleichsarbeiten dem neuen Lehrer hilfreich sein können, weil sie ihm Informationen über die Klasse liefern. Der Nachteil liegt seiner Meinung nach darin, dass die Erwartungen an die Ergebnisse zu hoch gesteckt sind. Denn in den ersten vier Wochen eines Schuljahres strömt so viel Neues auf die Schüler ein, außerdem liegen die großen Ferien dazwischen. Und so ist man zu dem Zeitpunkt „gerade mal warmgelaufen im neuen Schuljahr und da weiß ich um Gottes Willen nicht mehr, was am Ende der Klasse fünf an Satzgliedern dran war“ (Lehrer 2).

Die Interviewausschnitte zeigen, dass sich im Bereich „neuer Zeitpunkt“ einerseits ganz unterschiedliche Meinungen und Möglichkeiten der Nutzung abbilden lassen. Andererseits stellt sich die Frage, ob die Lehrerinnen und Lehrer aufgrund der zur Verfügung stehenden Zeit, vermehrt Maßnahmen zur individuellen Förderung ergreifen werden und ob sie aufgrund der Rückmeldungen ihre Unterrichtspraxis ändern werden.

Neben der Kritik am späten Zeitpunkt findet sich insgesamt in fünf Interviews von Realschullehrenden der Hinweis, dass die Vergleichsarbeiten zu früh stattfanden und zwar bis zu acht Wochen vor Schuljahresende. Kritisiert wird von diesen Lehrerinnen und Lehrern vor allem, dass zu dem Zeitpunkt noch nicht alle Unterrichtsinhalte behandelt sind. Aus diesem Grund ist es ungerecht, den gesamten Themenbereich in Vergleichsarbeiten zu behandeln, wenn nicht sichergestellt ist, dass alle Klassen alle Inhalte erarbeitet haben. Insgesamt betrachtet bildet diese Subkategorie die Meinung einer kleinen Gruppe von Lehrerinnen und Lehrern ab und stellt sicherlich eine Ausnahme dar. Durch die Verlegung des Zeitpunkts an den Anfang des folgenden Schuljahres wird dieses Argument in Zukunft keine Bedeutung mehr haben.

Reflexion – Hauptkategorie V) Förderdiagnostischer Nutzen der Vergleichsarbeit

In diesem Themenbereich lassen sich aus den Aussagen der Lehrerinnen und Lehrer insgesamt vier Subkategorien bilden; „Durch VA keine neuen Anhaltspunkte erhalten“, „Anhaltspunkte erhalten“, „Keine Fördermaßnahmen möglich“ und „Fördermaßnahmen möglich“.

Über die Hälfte der Befragten gibt an, dass sich aus den Vergleichsarbeiten keine neuen Hinweise über die Leistungen ihrer Schülerinnen und Schüler ergaben, die sie nicht bereits in den beiden Schuljahren erkannt haben. Schulartspezifische Unterschiede zeigen sich in dieser Subkategorie; bei den Gymnasial- und der Hauptschullehrenden teilen jeweils 2/3 diese Meinung. In der Realschule stimmt weniger als Hälfte der Interviewpartner dem Argument zu.

Dass die Lehrerinnen und Lehrer spezielle Anhaltspunkte aus Vergleichsarbeiten erhalten haben, gibt die Hälfte aller Befragten an, über alle Schularten hinweg. Die Anhaltspunkte lassen sich konkreten Situationen zuordnen. Die meisten Lehrpersonen haben extrem abweichende Leistungen der Schülerinnen und Schüler zur Kenntnis genommen, also standen sogenannte Ausreißer im Fokus ihrer Aufmerksamkeit. Vor allem Realschullehrende konnten Anhaltspunkte auf bestimmte Inhalte aus den Ergebnissen ableiten. Ein geringer Anteil Lehrender vertritt die Ansicht, dass man dann Anhaltspunkte aus Vergleichsarbeiten ableiten kann, wenn man die Klasse weiterführt. Die übersichtliche Darstellung des Auswertungsinstruments stellt nach Meinung der Lehrpersonen einen wichtigen Nutzungsfaktor dar.

Die Ergebnisse zeigen deutlich, dass Lehrerinnen und Lehrer in Vergleichsarbeiten kein Instrument sehen, das in der Lage ist, grundlegende, differenzierte Informationen über die spezifischen Leistungsstände ihrer Schülerinnen und Schüler zu liefern. Aus diesem Grund ist es nicht verwunderlich, dass über 2/3 der interviewten Personen die Meinung vertreten, dass aus Rückmeldungen von Vergleichsarbeiten keine Fördermaßnahmen abgeleitet werden können. Über ein Drittel aller Lehrpersonen betonen, dass die Aufgaben zu oberflächlich und die Rückmeldungen nicht genügend differenzieren. Man erhält lediglich „bunte Statistiken“ die jedoch keine Aussagekraft besitzen.

Auch unter dieser Kategorie wird die fehlende Zeit für Fördermaßnahmen im Anschluss an die Vergleichsarbeiten kritisiert. Somit besteht ein enger Zusammenhang zur Hauptkategorie IV) Zeitpunkt der Vergleichsarbeit. Den Lehrerinnen und Lehrern steht in den ersten Jahren der verpflichtenden Durchführung von Vergleichsarbeiten keine Zeit für die Verwendung der Erkenntnisse im Bereich der Diagnose und Förderung zur Verfügung. Dies kann als klarer Hinweis interpretiert werden, dass die Verantwortlichen formative Evaluationsverfahren zwar schriftlich festlegen, aber die grundlegenden Strukturen und Bedingungen nicht bereitstellen.

Dieselbe Kritik richtet sich auch an die Ebene der Einzelschule. Ein Fünftel der Lehrenden kritisieren das fehlende schulische Förderangebot. Trotz einer breiten Bereitschaft dieser Lehrerinnen und Lehrer, ihren Schülerinnen und Schülern die erforderlichen Fördermaßnahmen zu bieten, werden sie durch schulische Rahmenbedingungen eher daran gehindert als unterstützt. Wenn an einer Schule überhaupt Förderangebote existieren, dann beschränken sich diese auf die Klassenstufen fünf und sechs und finden relativ unstrukturiert und unspezifisch statt. Einige Lehrende klagen, dass die Schulwirklichkeit im Umgang mit Vergleichsarbeiten so aussieht, dass diese direkt nach der Durchführung archiviert werden und somit keine Konsequenzen abgeleitet werden.

Diese Befunde weisen deutlich auf die fehlenden Evaluationsstrukturen und -kultur hin, sowohl auf der Meso- als auch auf der Makroebene des Bildungswesens. Hier müssen meiner Meinung nach zahlreiche Verbesserungen angestrebt werden, die im Modell von Visscher und Coe (siehe Kap. 3.4.3.1) unter Punkt (C) Implementationsprozess (Makroebene) und (D) Schulmerkmale (Ebene der Einzelschule) angesiedelt sind. Insbesondere zählen auf der Makroebene die Unterstützung der Schule und die Beobachtung der Implementationsdurchführung und der Implementationseffekte, die entsprechend berücksichtigt werden müssen. Je nach Bedarf müssen dafür spezielle Ressourcen für den Veränderungsprozess bereitgestellt werden. Auf der Ebene der Einzelschule spielt die Bereitstellung zusätzlicher Ressourcen ebenso eine zentrale Rolle, wie auch eine koordinierte und aufeinander abgestimmte Zusammenarbeit. Nur wenn diese grundlegenden Voraussetzungen geschaffen werden, sehe ich eine Chance darin, dass Vergleichsar-

beiten tatsächlich einen Beitrag zur Qualitätsentwicklung und Qualitätssicherung leisten können.

6.5 Aktionen und abgeleitete Maßnahmen aus Vergleichsarbeiten

Nachdem sich die Lehrerinnen und Lehrer in allen Schularten in zahlreichen Themenbereichen vor allem kritisch gegenüber der Makroebene äußern, ist es nicht verwunderlich, dass beinahe alle Befragten (50 von 59) der Ansicht sind, dass Vergleichsarbeiten grundsätzlich keinen Anlass für Änderungen der Unterrichtspraxis liefern. Wenn aufgrund der Rückmeldungen aus Vergleichsarbeiten Änderungen vorgenommen werden, dann werden im Unterricht, sofern dafür Zeit vorhanden ist, spezielle Themen behandelt, in denen die Schülerinnen und Schüler schlechte Resultate erzielten. Ein Viertel der Lehrerinnen und Lehrer kann sich vorstellen, in Zukunft häufiger Themen zu wiederholen oder komplexere Aufgaben zu behandeln. Sowohl in der Realschule als auch im Gymnasium planen vier Personen, neue Unterrichts- bzw. Lehr-Lernmethoden einzuführen. Diese Lehrkräfte subsumieren unter dem Begriff neue „Unterrichts- bzw. Lehr-Lernmethoden“ ein verstärktes Methodentraining, die Verwendung neuer Medien (z.B. Karteikarten) oder neuer Lernstrategien, die spielerisch zur Festigung von Lerninhalten beitragen sollen. Eine Gymnasiallehrerin berichtet von ihrer Idee, in Zukunft verstärkt Projektunterricht anzubieten. Ob diese Beispiele einer ‚veränderten Lernkultur’ einen effizienten Beitrag im Hinblick der Qualitätssicherung und –entwicklung leisten wird, darf bezweifelt werden.

Insgesamt betrachtet decken sich die Ergebnisse dieser Interviewstudie zu ‚Auswirkungen der Vergleichsarbeiten’ mit den Befunden aus der quantitativen und qualitativen Studie von Maier (2009b). Auch in den meisten Rezeptionsstudien (siehe Kap. 3.4.1.2.2; Ergebnisse zur Rezeption und Nutzung der Daten) berichten die Autoren von abgeleiteten Maßnahmen, die sich auf spezielle Testaufgaben und –inhalte beziehen oder auf Wiederholungen oder Schwerpunktsetzung bestimmter Kompetenzbereiche.

Im Gegensatz dazu werden Thüringer Kompetenztests primär für die Schulleistungsdiagnostik eingesetzt. Ebenso stellt in Nordrhein-Westfalen, neben oben genannten Punkten, die intensive Förderung durch zusätzlichen Unterricht eine zentrale Maßnahme dar, die als Antwort auf die Ergebnisse aus Lernstandserhebungen abgeleitet werden. Diese förderdiagnostischen Bereiche spielen nach Einschätzung der Lehrerinnen und Lehrer in den baden-württembergischen Vergleichsarbeiten definitiv keine Rolle.

Betrachtet man das Modell von Helmke & Hosenfeld (2005) aus Kap. 3.4.3.2, so findet man dort im Schritt „Aktion“, die idealtypische Beschreibung, welche möglichen Maßnahmen von den Lehrerinnen und Lehrern als Folge der Reflexionen durchgeführt werden können. Die Autoren zählen dort mehrere möglichen Maßnahmen auf: „Die Sicherung eines Mindestniveaus“, „die Verbesserung von Unterrichtsqualität und Klassenführung“, „den Aufbau einer Evaluations-, Aufgaben- und Fehlerkultur“, „die Diagnostische Kompetenz“ und zuletzt „die Koppelung mit Projekten zur Unterrichtsqualität“.

Vergleicht man die Ergebnisse der vorliegenden Interviewstudie mit diesen aufgeführten Möglichkeiten, dann kann festgehalten werden, dass die Rückmeldungen aus Vergleichsarbeiten kaum einen Beitrag zu einer veränderten Unterrichtspraxis im Sinne einer Verbesserung und Sicherung schulischer Qualität leisten. Die Auswirkungen auf Unterricht und Schule sind nur marginal spürbar und beziehen sich in den meisten Fällen auf reine Wiederholungen und speziellen Übungseinheiten.

6.6 Innerschulische Kommunikation und Unterstützung

Um die innerschulische Kommunikationsstruktur der Einzelschule zu erfassen, erhielten die Lehrenden die Möglichkeit in einer mulitple-choice Matrix anzukreuzen, ob und mit welchen Personen und / oder Personengruppen die Ergebnisse aus Vergleichsarbeiten besprochen wurden.

Die Qualität und Intensität der innerschulischen Kommunikation konnten die Lehrenden bewerten, in dem ihnen eine dreistufige Skala vorgelegt wurde. Sie konn-

ten wählen, ob sie mit den jeweiligen Akteuren „keine Gespräche“, einen „informellen Austausch“ oder eine „systematische Diskussion“ geführt hatten.

Die Auswertung der innerschulischen Kommunikationsstrukturen zeichnet ein relativ nüchternes Bild. In keiner Schule wird der Themenbereich „Vergleichsarbeiten“ in der Schulkonferenz besprochen. Des Weiteren gibt 4/6 bis 5/6 der Befragten an, dass weder in der Gesamtlehrer-, der Jahrgangsstufen- oder Klassenkonferenz, noch mit den Eltern der Klasse Gespräche stattfinden.

Informelle Gespräche über die Ergebnisse aus Vergleichsarbeiten werden etwa von der Hälfte der Lehrpersonen mit ihren Schülerinnen und Schülern geführt. In einer vergleichbaren Qualität tauscht sich etwas mehr als 1/3 der Lehrerinnen und Lehrer mit den Kolleginnen und Kollegen aus, die entweder in der Parallelklasse oder in einem anderen Fach in die Durchführung von Vergleichsarbeiten involviert sind. Außerdem geben 19 der 59 Lehrkräfte an, sich mit der Schulleitung auszutauschen. In der entsprechenden Fachkonferenz liegt der Anteil bei 16 Personen.

Sowohl in der Klassen- und der Jahrgangsstufenkonferenz als auch „mit den Eltern der Klasse“ liegt der Anteil des informellen Austausches bei 1/6 der Lehrkräfte am niedrigsten.

Auf den ersten Blick erscheint dieser „informelle Austausch“ über die Ergebnisse von Vergleichsarbeiten für die innerschulische Kommunikationsstruktur nicht sehr bedeutend zu sein. Schließlich erfasst man mit dieser quantitativen Erhebungsmethode nicht, ob der Inhalt der Gespräche eher neutral, negativ oder positiv ausfällt und welche Auswirkungen diese auf die Organisation als Ganzes haben. Man sollte jedoch bedenken, dass in den Gesprächen „zwischen Tür und Angel“ eventuell wichtige Informationen und Meinungen ausgetauscht werden. Möglicherweise werden in diesem Bereich Aussagen gemacht, die für die Verwendung der Daten im Unterricht oder die Weiterentwicklung der Einzelschule einen wichtigen Beitrag leisten könnten. Aus diesem Grund halte ich es für wesentlich, dass an den jeweiligen Schulen für den Austausch genügend Zeit eingeplant wird und dass Strukturen geschaffen werden, die den systematischen Dialog fordern, damit alle Aspekte beim Aufbau einer Feedbackkultur berücksichtigt werden können.

Die Auswertungsmatrix der „systematischen Diskussionen“ zeigt, dass etwa bei der Hälfte der Personen ein Austausch mit den Parallelkollegen und –kolleginnen stattfindet. Mit den Schülerinnen und Schülern setzt sich 1/3 der Lehrerinnen und Lehrer über die Ergebnisse aus Vergleicharbeiten auseinander. Ein Viertel der Befragten spricht mit ausgewählten Kolleginnen und Kollegen oder mit jenen, die ebenfalls eine Vergleichsarbeit durchführten. Ein Fünftel führt Diskussionen mit der Fachkonferenz und mit der Schulleitung.

Betrachtet man die einzelnen Schularten, so treten in diesem Punkt schulspezifische Unterschiede besonders deutlich hervor. An den Realschulen finden insgesamt gesehen die meisten systematischen Diskussionen statt, denn der Austausch in den einzelnen Konferenzen bzw. mit den verschiedenen Akteuren, liegt immer zwischen einem Drittel und der Hälfte der Lehrenden. Dieser ‚verstärkte’ Austausch könnte darin begründet sein, dass die Realschullehrerinnen und –lehrer die Ergebnisse gegenüber Eltern, Schülerinnen und Schülern eher rechtfertigen müssen und sie unter einem gewissen Legitimationsdruck stehen, als Lehrende von Hauptschulen oder Gymnasien. Dass Rückmeldungen verstärkt Diskussionen innerhalb des Kollegiums anregen, wird u.a. auch durch die LAU 7 Studie belegt.

Die geringe und zurückhaltende Auseinandersetzung in den Hauptschulen könnte als Zeichen für das niedrige Niveau interpretiert werden, das alle Lehrende in den Interviews beklagen. Die überdurchschnittlichen und unrealistischen Leistungsergebnisse der Schülerinnen und Schüler werden von den Lehrenden nicht ernst genommen. Sie sehen deshalb auch keinen Anlass, sich über die Ergebnisse mit anderen auszutauschen.

An Gymnasien fällt der systematische Austausch mit den Kolleginnen und Kollegen aus den Parallelklassen und mit den Schülerinnen und Schülern am intensivsten aus. Außerdem werden die Ergebnisse auch von ca. einem Viertel der Lehrenden mit ausgewählten Kollegen und mit allen, die an der Vergleichsarbeit beteiligt sind, ausführlich diskutiert. Auffallend ist in dieser Schulart, dass die Fachkonferenz kaum eine Plattform für eine fachspezifische Analyse der Vergleichsarbeiten darstellt. Dieses Ergebnis wird ebenfalls durch die Befunde aus LAU 7 gestützt (vgl. Kap. 3.4.1.2.4).

Ein markantes Ergebnis liefert der Blick auf die Kommunikation zwischen Schulleitung und Lehrenden; im Gymnasium gibt es zwischen diesen beiden Akteursgruppen keine geplanten, systematischen Gespräche. Dies könnte daran liegen, dass es innerhalb des Kollegiums spezielle Evaluationsbeauftragte gibt, die für die Durchführung und Auswertung verschiedener Verfahren und Instrumente der Qualitätsentwicklung verantwortlich sind. Auf Rückfrage in einem Gymnasium wurde dies bestätigt. In dieser Schule gibt es einen Lehrer, der zunächst alle Evaluationsergebnisse sammelt. Diese werden einem schulischen Evaluationsteam, in dem Lehrende, Schülerinnen und Schüler vertreten sind, vorgestellt und diskutiert. Gemeinsam wird nach geeigneten Maßnahmen gesucht, die durchgeführt und im Anschluss daran auf ihre Wirksamkeit hin überprüft werden. Dieser Befund stützt die These, dass kooperative Strukturen, fächerübergreifende Evaluationskonzepte, gemeinsame Auswertungsstrategien die Rezeption und Nutzung von Evaluationswissen erhöhen und somit zum Aufbau einer Evaluationskultur beitragen.

6.7 Fünf unterschiedliche Rezeptions- und Nutzungstypen

In einigen Rezeptionsstudien wurde herausgestellt, dass sich die Akteure in der Rezeption und Nutzung von Rückmeldungen aus Evaluationsstudien bezüglich ihrer Einstellung und Haltung unterscheiden. So wurde z.B. in der QuaSUM 2 Studie berichtet, dass die Meinungen der Lehrerinnen und Lehrer zu externen Evaluationsverfahren von Befürwortung bis Ablehnung reichen (vgl. Peek 2004a). In Baden-Württemberg konnte auch Maier (2009b) in seiner quantitativen und qualitativen Studie unterschiedliche Einstellungen zu Vergleichsarbeiten nachweisen, deren Auftreten er jedoch schulartspezifisch begründet. Demnach fühlen sich Hauptschullehrkräfte durch Vergleichsarbeiten kaum belastet, während Lehrkräfte an Realschulen und Gymnasien von Anfang an eine kritischere Haltung einnehmen würden. In der hier vorliegenden Interviewstudie lassen sich ebenfalls unterschiedliche Einstellungen und Haltungen von Lehrenden nachweisen. Doch die schulartspezifischen Zusammenhänge die Maier berichtet, können mit den Interviewdaten in der Form nicht bestätigt, aber dennoch erklärt werden. Möglicherweise besteht ein Zusammenhang zwischen der gering empfundenen Belastung

der Hauptschulkolleginnen und -kollegen und dem niedrigen Niveau der Vergleichsarbeiten. Die Vergleichsarbeiten wurden an den Hauptschulen nicht ernst genommen, so dass sich möglicherweise dadurch die gelassene Haltung erklären lässt.

Diese qualitative Interviewstudie liefert im Bereich der unterschiedlichen subjektiven Einstellungen und Haltungen von Lehrerinnen und Lehrern zu Vergleichsarbeiten einen entscheidenden Beitrag, sowohl für die Rezeptions- als auch für die Typenforschung. Mit Hilfe eines differenzierten und detaillierten, empirisch begründeten Typen bildenden Verfahrens nach Kluge (1999), wurden fünf unterschiedliche Rezeptions- und Nutzungstypen gebildet:

- Der begeisterte, überzeugte und befürwortende Typ
- Der positiv mitspielende Typ
- Der desinteressierte, gelassene Typ
- Der negativ kritische Typ
- Der ablehnende, belastete Typ

Die Typen unterscheiden sich bezüglich ihrer Einstellung und ihrer Beurteilung zur Funktion von Vergleichsarbeiten. Des Weiteren finden sich Unterschiede bei der Durchführung der VA und der Auseinandersetzung mit Rückmeldedaten. Sie haben verschiedene Meinungen und Bewertungen zu einzelnen Aspekten der Vergleichsarbeit. Nicht zuletzt bestehen zwischen den fünf Gruppen große Differenzen in der innerschulischen Kommunikationsstruktur.

Nachdem die fünf Typen in Kap. 5.7 ausführlich beschrieben wurden, erfolgt hier eine kurze Zusammenfassung. Die Diskussion orientiert sich an der Fragestellung, welche Auswirkung der jeweilige Typus auf die Rezeption und Nutzung von Vergleichsarbeiten hat.

Der „begeisterte, überzeugte und befürwortende Typ“

Dieser Typ zeichnet sich durch seine extrem positive Einstellung gegenüber Vergleichsarbeiten und externer Evaluationsverfahren aus. Er sieht darin ein großes Nutzungspotential, das sowohl auf der Ebene des Unterrichts als auch auf der Ebene der Einzelschule einen wichtigen Beitrag zur Qualitätsentwicklung und Qualitätssicherung leisten kann. Mit Interesse und Freude setzt er sich intensiv mit den Rückmeldedaten auseinander. Er hält sie bestens geeignet für die Diagnose und Förderung der Schülerinnen und Schüler, weil er Anhaltspunkte über deren

Stärken und Schwächen erhält. Für die zukünftige Unterrichtspraxis plant er die Übernahme komplexerer Aufgaben aus den Vergleichsarbeiten und wendet sich verstärkt individualisierenden Unterrichtsformen zu. Die innerschulische Kommunikation findet bei diesem Typ auf allen Ebenen, mit allen Akteuren auf einer systematischen und intensiven Art und Weise statt. Die Umsetzung von Maßnahmen sieht er als persönliche Herausforderung und nützt dazu alle verfügbaren Ressourcen. Dabei kritisiert er die fehlenden strukturellen und organisatorischen Rahmenbedingungen, die einer gelingenden Evaluation im Wege stehen. Meiner Meinung nach ist die Gefahr bei diesem Typus groß, dass er sich überanstrengt und überlastet. Die enormen Anstrengungen könnten nicht zu den erwünschten Wirkungen führen, so dass am Ende die Enttäuschung und der Frust überwiegen.

Vergleicht man die Eigenschaften und Merkmale des begeisterten, überzeugten und befürwortenden Typs mit den Typologien, die aus der VERA-Studie und aus der Studie von Margit Stamm generiert wurden, so werden deutliche Zusammenhänge sichtbar (vgl. Kap. 3.4.2). Sowohl TYP 1 (VERA) und der Stamm'sche „Innovations"-Typ weisen klare Parallelen auf. Denn auch diese Typen bringen eine enorme Anstrengungsbereitschaft zur Datenauseinandersetzung mit. Sie stufen die Verständlichkeit der Rückmeldungen als hoch ein und sehen viele Nutzungsmöglichkeiten bei externen Evaluationsverfahren. Sie leiten aus den Ergebnissen neue Entwicklungsziele ab und formulieren konkrete Maßnahmen für die Unterrichtsentwicklung. Der „Innovations"-Typ betrachtet zudem externe Verfahren als persönliche Angelegenheit und setzt alles dran, möglichst zahlreiche geeignete Maßnahmen abzuleiten und umzusetzen. Auch Stamm beschreibt, dass der immense Arbeitsaufwand im ungünstigen Fall zur Überlastung und Überanstrengung führen könnte und alle Anstrengungen im Nichts enden.

Meines Erachtens stellt der begeisterte, überzeugte und befürwortende Typ einen wichtigen Motor der Schulentwicklung dar, weil er sich bereit erklärt, über das übliche Maß hinaus, aktiv daran mitzuwirken und mitzugestalten. Dieses Potential gilt es effektiv in Prozesse einzubinden und auf allen Ebenen die erforderlichen strukturellen und organisatorischen Rahmenbedingungen bereit zu stellen. Dieser Typ kann auch einen wesentlichen Beitrag beim Aufbau kooperativer und kommunikativer Elemente beisteuern. Die Schulleitung sollte alles daran setzen, dass Kolleginnen und Kollegen mit dieser positiven Einstellung eine geeignete und

passende Plattform für die Umsetzung ihrer Ideen zur Verfügung gestellt bekommen.

Der „positiv mitspielende Typ“

Im Vergleich zum ersten Typus fällt bei diesem ebenfalls eine grundsätzlich positive Grundhaltung zu Vergleichsarbeiten auf, die Begeisterung und Überzeugung hält sich jedoch in Grenzen. Der positiv mitspielende Typ erkennt im Allgemeinen einen positiven Ansatz in der Konzeption der Vergleichsarbeiten im Sinne summativer, Kriterien bezogener Evaluationsverfahren. Allerdings werden von diesem Typus einige Punkte bemängelt, z.B. die fehlende Zeit für die Umsetzung oder der anstehende Lehrerwechsel. Seiner Meinung nach befinden sich Vergleichsarbeiten noch in einem Entwicklungsstadium für welches er bereits Verbesserungsvorschläge parat hat. Dementsprechend moderat fällt die Auseinandersetzung und Nutzung der Rückmeldedaten aus. Lediglich die Ergebnisse unerwarteter ‚Ausreißer’ werden von ihm intensiver analysiert. Insgesamt fühlen sich die Personen, die diesem Typus zugeordnet werden können, durch die Ergebnisse ihrer Schülerinnen und Schüler bestätigt. Wenn Maßnahmen durchgeführt werden, was jedoch aus Zeitgründen kaum geschieht, dann werden insbesondere Themengebiete aus der Vergleichsarbeit wiederholt. Der kommunikative Austausch des „positiv mitspielenden Typs“ läuft am ehesten auf der Ebene von informellen Gesprächen statt.

Der beschriebene Typus erinnert an Typ 2 der VERA-Typologie, der sich ebenfalls durch ein moderateres Nutzungsverhalten vom ersten Typ unterscheidet. Es lassen sich auch Ähnlichkeiten mit dem von Stamm beschriebenen „Reaktions“-Typus herstellen. Und zwar in dem Sinne, dass Ergebnisse nicht als Anlass für Veränderungen interpretiert werden, sondern um damit die eigene Arbeit zu bestätigen. Während der Reaktionstyp jedoch schnell und häufig auch unüberlegt Maßnahmen ableitet, hält sich der „positive Mitspieler“ eher zurück. Er bemängelt den ‚unreifen’ Zustand von Vergleichsarbeiten und leitet kaum Maßnahmen ab. Für die Zukunft sieht er durchaus Entwicklungspotential, das durch geeignete Verbesserungsmaßnahmen sichergestellt werden könnte.

Auch bei diesem Typus kommt meiner Ansicht nach die grundsätzlich positive Einstellung und Haltung zu Vergleichsarbeiten deutlich zum Ausdruck. Insgesamt

entsteht der Eindruck, dass der Prozess der Rezeption und Nutzung im Vergleich zu den überzeugten Befürworterinnen und Befürwortern verhaltener abläuft. Dennoch ist ein breites Nutzungspotential vorhanden, das durch einige Verbesserungen noch effektiver genutzt werden könnte. Die kritischen Anmerkungen dieses Typus richten sich vor allen Dingen an die Makroebene, speziell an diejenigen, die für die Erstellung der Vergleichsarbeiten verantwortlich sind. Ebenso könnten auf der Ebene der Einzelschule Strukturen eingerichtet werden, damit sich dieser Typus vom Mitspieler zum „Akteur“ entwickeln kann.

Der desinteressierte, gelassene Typ

Für diesen Typ haben Vergleichsarbeiten und andere Evaluationsverfahren keine Bedeutung. Dementsprechend gering fällt das Interesse für die Zielsetzung von Vergleichsarbeiten und für die Durchführung aus. Teilweise werden die Vergleichsarbeiten nicht einmal ausgewertet. Wenn Rückmeldedaten vorliegen, dann werden diese in den meisten Fällen nicht zur Kenntnis genommen. Selbst in den Interviews entsteht der Eindruck, dass die Lehrerinnen und Lehrer die diesem Typ zugeordnet werden, ihre persönliche Meinung zurückhalten und sich nicht positionieren. Es werden auch keine systematischen Diskussionen geführt.

Dieser Typ wird bisher in dieser extremen Ausprägung in keiner Typologie erfasst. Stamm beschreibt zwar die Typik des Alibis, der sich ähnlich desinteressiert gibt. Was diesen jedoch vom „desinteressierten, gelassenen Typen“ unterscheidet ist, dass sich der Alibi-Typus einem großen Legitimationsdruck ausgesetzt fühlt und Evaluationsverfahren vor allem der Rechenschaftslegung dienen. Aber wie oben beschrieben zeigt der Desinteressierte auch in diesem Punkt äußerste Gelassenheit, er fühlt sich persönlich in keiner Weise kontrolliert.

Ob und wie dieser Typus zu einer gelingenden Evaluation beitragen kann, ist für mich schwierig zu beurteilen. Es kann davon ausgegangen werden, dass unter den gegebenen Bedingungen von ihm keine Initiative zu erwarten ist. Und wie weit z.B. die Schulleitung oder andere schulischen Akteure dazu beitragen können, ihn zur Mitarbeit bei Evaluationsverfahren zu aktivieren, lässt sich nicht beantworten, solange man nicht die jeweiligen individuellen Eigenheiten und Stärken der Person kennt.

Meines Erachtens dürfte sich das Desinteresse von Lehrenden nicht derart äußern, dass sie selbst wählen können, ob sie sich an einer mögliche Nutzung von Vergleichsarbeiten beteiligen oder nicht. In schulische Evaluationsverfahren sollten alle Akteure an einer gemeinsamen Zielsetzung und Umsetzung der Verfahren beteiligt und eingebunden werden. Die Intensität der Mitarbeit wird sich bei der heterogenen Zusammensetzung von Kollegien interindividuell immer unterscheiden lassen. Aber die grundsätzliche Nutzungsbereitschaft muss bei allen Beteiligten vorhanden sein. Die Schulleitung bzw. Steuerungsgruppe hat für die Bereitstellung aller notwendigen Bedingungen und Strukturen zu sorgen.

Der negativ kritische Typ

Dieser Typ erlebt Vergleichsarbeiten ausschließlich negativ; als Kontrollinstrument und Druckmittel, teilweise auch als puren Aktionismus. Er setzt sich ähnlich intensiv - wie der begeisterte und überzeugte Befürworter - mit Vergleicharbeiten auseinander. Allerdings fällt diese gründliche Auseinandersetzung um ein Vielfaches kritischer aus; denn es werden nur Mängel von Vergleichsarbeiten aufgeführt. Dadurch kommen Enttäuschung und Wut bei ihm deutlich zum Ausdruck. Die Auswirkungen auf die Unterrichtspraxis fallen bei diesem Typen eher in den Bereich des ‚teaching-to-the-test'. Außerdem führt er zahlreiche systematische Diskussionen über die Ergebnisse aus Vergleichsarbeiten mit unterschiedlichen Akteuren.

Insbesondere die intensive Auseinandersetzung mit den Daten und die grundsätzliche kritische Einstellung des negativ kritischen Typs erinnert an den Typ 3 der VERA-Studie und an den „Blockade"-Typen aus der Typologie von Stamm. Die Kritik richtet sich, wie bei Stamm, vornehmlich an die Makroebene und bezieht sich vor allem auf die Gütekriterien und die Konzeption der Vergleichsarbeit.

Meines Erachtens ist bei diesem Typen die grundsätzliche Bereitschaft zur kritischen Auseinandersetzung mit staatlichen Vorgaben hoch. Er verwendet viel Zeit darauf und hat einen hohen Anspruch an Vergleichsarbeiten. Der Grund, dass er diese ablehnt, liegt an der mangelhaften Konzeption und an den vielen Schwachstellen, die er an zahlreichen Kriterien festmacht. Die Vermutung liegt nahe, dass sich dieser Typ bei anspruchsvolleren und verbesserten Testinstrumenten durch-

aus einem der beiden ersten Typen „Begeisterte, Überzeugte und Befürwortende“ oder den „positiven Mitspielern“ zuordnen lässt.

Die Verantwortlichen der Bildungsadministration haben durch die mangelhafte Testkonstruktion ein großes Nutzungspotential ‚verspielt’. Es bleibt zu hoffen, dass bei der Entwicklung zukünftiger Tests zum einen höhere Standards angestrebt und erreicht werden. Zum anderen dass es gelingt, die Bereitschaft und Motivation derjenigen Lehrerinnen und Lehrer hochzuhalten, die von den bisherigen Vergleichsarbeiten enttäuscht wurden. Gelingt dies nicht muss damit gerechnet werden, dass sich die Anzahl des „desinteressierten und gelassenen Typs“ erhöht.

Der ablehnende, belastete Typ

Bei diesem Typen kommt die enorme Belastung deutlich zum Ausdruck unter der er als Lehrer leidet. Diese wird durch Vergleichsarbeiten noch erhöht, weil er dazu gezwungen wird. Er fühlt sich dadurch kontrolliert und gerät weiter unter enormen Druck. Vergleichbar mit dem desinteressierten, gelassenen Typus setzt sich auch dieser nicht mit den Ergebnissen auseinander, leitet keine Maßnahmen ab und tauscht sich mit niemandem aus. Während der Desinteressierte jedoch gelassen bleibt, sind der erlebte Druck und die Belastung bei diesem Typen enorm hoch.

Meiner Meinung nach ist der psychische Zustand, in dem sich die Personen dieses Typus befinden, äußerst Besorgnis erregend. Die Belastung hängt nicht mit einzelnen Evaluationsverfahren zusammen, sondern ist grundsätzlicher zu interpretieren. Für diesen Typus wäre es sicher hilfreich, wenn sie in den wesentlichen belastenden schulischen Situationen Unterstützung und Hilfestellungen erhalten würden. Liegen die Gründe, die zur hohen Belastung führen, jedoch im persönlichen Bereich, und können durch schulische Maßnahmen nicht minimiert werden, so ist eine professionelle Hilfe dringend anzuraten.

6.8 Ausblick

Mit den Ergebnissen aus dieser Interviewstudie konnte dargestellt werden, dass bei der Rezeption und Nutzung der baden-württembergischen Vergleichsarbeiten im Sinne der Unterrichtsentwicklung und Qualitätssicherung von Schule noch deutlicher Handlungsbedarf besteht. An zahlreichen Beispielen wurde aufgeführt, dass auf allen Ebenen des Bildungssystems weitere Anstrengungen und Verbesserungen nötig sind, um das Potential von Vergleichsarbeiten auszuschöpfen. Im Folgenden sind nur einige Möglichkeiten dargestellt, die dazu beitragen könnten, die Qualität des deutschen Bildungssystems kontinuierlich weiter zu entwickeln, zu verbessern und dadurch zu sichern.

Ebene	Mögliche Aufgabe zur Nutzung der Vergleichsarbeiten
Makroebene	• Aus- und Fortbildung der Lehrerinnen und Lehrer, • Bereitstellung von Testinstrumenten höchster Güte, • Begleitung und Unterstützung durch das Landesinstitut, • etc....
Schulleitung / Steuergruppen	• Schulprogramme entwickeln; • Lehrerfortbildungen anregen, • Kooperationen im Kollegium und in den Fachkonferenzen fördern,… • etc.
Fachkonferenzen	• Fachlichen Kompetenzerwerb in versch. Bereichen anstoßen, begleiten und überprüfen • Gemeinsame Ziele formulieren • …..
Lehrerin / Lehrer	• Kooperationen mit anderen Kolleginnen • Ausbau der Diagnosekompetenz • Fachspezifische Lehr-Lernprozesse in einer Lerngruppe anstoßen, begleiten und überprüfen. • Schülerinnen und Schüler zunehmend unterstützen bei ihrer Selbstständigkeit • …

7 Literaturverzeichnis

Ackermann, Heike; Rosenbusch, Heinz Stephan (1995): Qualitative Forschung in der Schulpädagogik. In: Eckard König und Peter Zedler (Hg.): Bilanz qualitativer Forschung. Band I: Grundlagen qualitativer Forschung. Weinheim: Deutscher Studien Verlag, S. 135–167.

Altrichter, Herbert (1998): Reflexion und Evaluation in Schulentwicklungsprozessen. In: Herbert Altrichter, Wilfried Schley und Michael Schratz (Hg.): Handbuch zur Schulentwicklung. Innsbruck: Studien Verlag, S. 263–335.

Altrichter, Herbert (2008): Veränderungen der Systemsteuerung im Schulwesen durch die Implementation einer Politik der Bildungsstandards. In: Thomas Brüsemeister und Klaus-Dieter Eubel (Hg.): Evaluation, Wissen und Nichtwissen. Wiesbaden: VS Verlag für Sozialwissenschaften , S. 75–115.

Altrichter, Herbert (2009): Governance - Schulreform als Handlungskoordination. In: *Die Deutsche Schule* 101 (3), S. 240–252.

Altrichter, Herbert; Eder, Ferdinand (2004): Das "Autonomie-Paritätsmuster" als Innovationsbarriere? In: Heinz Günter Holtappels (Hg.): Schulprogramme-Instrumente der Schulentwicklung. Konzeptionen, Forschungsergebnisse. Weinheim & München: Juventa, S. 195–221.

Altrichter, Herbert; Maag Merki, Katharina (Hg.) (2010): Handbuch. Neue Steuerung im Schulsystem. Wiesbaden: VS Verlag für Sozialwissenschaften.

Altrichter, Herbert; Brüsemeister, Thomas; Wissinger, Jochen (2007a): Einführung. In: Herbert Altrichter, Thomas Brüsemeister und Jochen Wissinger (Hg.): Educational Governance. Handlungskoordination und Steuerung im Bildungssystem. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 9–13.

Altrichter, Herbert; Brüsemeister, Thomas; Wissinger, Jochen (Hg.) (2007b): Educational Governance. Handlungskoordination und Steuerung im Bildungssystem. Wiesbaden: VS Verlag für Sozialwissenschaften.

Arbeitsgruppe Kompetenztests (2010): Thüringer Kompetenztests. Unter Mitarbeit von Christof Nachtigall. Hg. v. Lehrstuhl für Methodenlehre und Evaluationsforschung. Friedrich-Schiller-Universität Jena. Online verfügbar unter http://www.kompetenztest.de/kompetenztest1.html, zuletzt aktualisiert am 19.01.2010, zuletzt geprüft am 19.07.2010.

Arnold, Karl-Heinz (2002): Schulentwicklung durch Rückmeldung der Lernwirksamkeit an die Einzelschule: Möglichkeiten und Grenzen der Schuleffizienzforschung. In: *Zeitschrift für Pädagogik* 48 (5), S. 741–764.

Arnold, Rolf; Faber, Konrad (2000): Qualitätssysteme und ihre Relevanz für Schule. Einführung und Überblick. Seelze-Velber: Kallmeyer.

Artelt, Cordula (2007): Externe Evaluation und einzelschulische Entwicklung - Ein zukunftsreiches Entwicklungsverhältnis? In: Jürgen van Buer und Cornelia Wagner (Hg.): Qualität von Schule. Ein kritisches Handbuch. Frankfurt am Main: Lang, S. 131–140.

Avenarius, Hermann; Ditton, Hartmut; Döbert, Hartmut; Klemm, Klaus; Klieme, Eckhard; Rürup, Matthias; Tenort… (2003): Bildungsberichterstattung für Deutschland. Erste Befunde.

Bähr, Konstantin (2003): Die Rolle von Schulleistungstests für das Qualitätsmanagement im Bildungswesen, in Schulen und Klassenzimmern. Online verfügbar unter http://www.konstantinbaehr.ch/dl-files/Baehr_Leistungstest.pdf, zuletzt aktualisiert am 03.06.2003, zuletzt geprüft am 05.05.2010.

Bähr, Konstantin (2006): Erwartungen von Bildungsadministrationen an Schulleistungstests. In: Harm Kuper und Julia Schneewind (Hg.): Rückmeldung und Rezeption von Forschungsergebnissen. Zur Verwendung wissenschaftlichen Wissens im Bildungssystem. Münster: Waxmann, S. 127–141.

Bastian, Johannes (2007): Autonomie von Schulen ist kein Selbstzweck. Oder: Autonomiepostulate ohne Unterstützungssysteme sind hohl! In: Ulrich Herrmann (Hg.): In der Pädagogik etwas bewegen. Impulse für Bildungspolitik und Schulentwicklung. Weinheim: Beltz, S. 53–61.

Bauer, Karl-Oswald (2004): Evaluation - via regia zu besserer Qualität in Schule und Unterricht? In: Heinz Günther Holtappels, Klaus Klemm, Hermann Pfeiffer, Hans-Günter Rolff und Renate Schulz-Zander (Hg.): Jahrbuch der Schulentwicklung. Weinheim & München: Juventa (13), S. 161–185.

Baumert, Jürgen (2001): PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. Rev. Nachdr. der Erstausg. Opladen: Leske + Budrich.

Baumert, Jürgen; Lehmann, Rainer (1997): TIMSS - Third International Mathematics and Science Study. Dritte Internationale Mathematik- und Naturwissenschaftsstudie ; Anlage, Fragestellungen und Durchführung der TIMSS-Studie in der Bundesrepublik Deutschland. Berlin: Max-Planck-Institut für Bildungsforschung.

Baumert, Jürgen; Neubrand, Michael (2002): PISA 2000 - die Länder der Bundesrepublik Deutschland im Vergleich. Opladen: Leske + Budrich.

Berkemeyer, Nils (2010): Die Steuerung des Schulsystems. Wiesbaden: VS Verlag für Sozialwissenschaften.

Berkemeyer, Nils; Feldhoff, Tobias; Brüsemeister, Thomas [1](2008): Schulische Steuergruppen - ein intermediärer Akteur zur Bearbeitung des Organisationsdefizits der Schule? In: Roman Langer (Hg.): "Warum tun die das?". Governanceanalysen zum Steuerungshandeln in der Schulentwicklung. 1. Aufl. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 149–172.

Bock, Marlene (1992): Das halbstrukturierte-leitfadenorientierte Tiefeninterview. Theorie und Praxis der Methode am Beispiel von Paarinterviews. In: Jürgen H. P. Hoffmeyer-Zlotnik (Hg.): Analyse verbaler Daten. Über den Umgang mit qualitativen Daten. Opladen: Westdeutscher Verlag, S. 90–109.

Bonsen, Martin; Gathen, Jan von der (2004): Schulentwicklung und Testdaten. Die innerschulische Verarbeitung von Leistungsrückmeldungen. In: Heinz Günther Holtappels, Klaus Klemm, Hermann Pfeiffer, Hans-Günter Rolff und Renate Schulz-Zander (Hg.): Jahrbuch der Schulentwicklung. Weinheim & München: Juventa (13), S. 225–252.

Bonsen, Martin; Büchter, Andreas; Peek, Rainer (2004a): Datengestützte Schul- und Unterrichtsentwicklung. Bewertungen der Lernstandserhebungen in NRW durch Lehrerinnen und Lehrer. In: Wilfried Bos, Heinz Günther Holtappels, Hermann Pfeiffer, Hans-Günter Rolff und Renate Schulz-Zander (Hg.): Jahrbuch der Schulentwicklung. Daten, Beispiele und Perspektiven. Weinheim & München: Juventa (14), S. 125–148.

Bonsen, Martin; Büchter, Andreas; van Ophuysen, Stefanie (2004b): Im Fokus: Leistung. Zentrale Aspekte der Schulleistungsforschung und ihre Bedeutung für die Schulentwicklung. In: Heinz Günther Holtappels, Klaus Klemm, Hermann Pfeiffer, Hans-Günter Rolff und Renate Schulz-Zander (Hg.): Jahrbuch der Schulentwicklung. Weinheim & München: Juventa (13), S. 187–223.

Bortz, Jürgen; Döring, Nicola [4](2006): Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. Heidelberg: Springer Verlag.

Bos, Wilfried (Hg.) (2008): Schulische Qualitätsanalyse in Nordrhein-Westfalen. Konzepte, erste Erfahrungen, Perspektiven. Köln: Luchterhand.

Bos, Wilfried; Postlethwaite, Thomas Neville; Gebauer, Miriam M. [3](2010): Potentiale, Grenzen und Perspektiven internationaler Schulleistungsforschung. In: Rudolf Tippelt und Bernhard Schmidt

(Hg.): Handbuch Bildungsforschung. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 275–295.

Böttcher, Wolfgang; Dicke, Jan Nikolas (2008): Bildungsstandards und Controlling. In: Wolfgang Böttcher, Wilfried Bos, Hans Döbert und Heinz Günter Holtappels (Hg.): Bildungsmonitoring und Bildungscontrolling in nationaler und internationaler Perspektive. Dokumentation zur Herbsttagung der Kommission Bildungsorganisation, -planung, -recht (KBBB). Münster: Waxmann, S. 103–106.

Böttcher, Wolfgang; Bos, Wilfried; Döbert, Hans; Holtappels, Heinz Günter (Hg.) (2008): Bildungsmonitoring und Bildungscontrolling in nationaler und internationaler Perspektive. Dokumentation zur Herbsttagung der Kommission Bildungsorganisation, -planung, -recht (KBBB). Deutsche Gesellschaft für Erziehungswissenschaft. Münster: Waxmann.

Böttcher, Wolfgang; Holtappels, Heinz Günther; Brohm, Michaela (Hg.) (2006): Evaluation im Bildungswesen. Eine Einführung in Grundlagen und Praxisbeispiele. Weinheim: Juventa-Verlag

Brüsemeister, Thomas; Eubel, Klaus-Dieter (2008a): Evaluationsbasierte Steuerung, Wissen und Nichtwissen - Einführung in die Thematik. In: Thomas Brüsemeister und Klaus-Dieter Eubel (Hg.): Evaluation, Wissen und Nichtwissen. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 7–34.

Brüsemeister, Thomas; Eubel, Klaus-Dieter (Hg.) (2008b): Evaluation, Wissen und Nichtwissen. Wiesbaden: VS Verlag für Sozialwissenschaften.

Büeler, Xaver (1998): Schulqualität und Schulwirksamkeit. In: Herbert Altrichter, Wilfried Schley und Michael Schratz (Hg.): Handbuch zur Schulentwicklung. Innsbruck: Studien Verlag, S. 661–694.

Buer, Jürgen van; Wagner, Cornelia (Hg.) (2007): Qualität von Schule. Ein kritisches Handbuch. Frankfurt am Main: Lang.

Bülow-Schramm, Margret (2006): Qualitätsmanagement in Bildungseinrichtungen. Münster: Waxmann.

Dedering; Kathrin; Bos Wilfried (2008): Internationale Schulleistungsuntersuchungen und Schulleistungstests - eine Einführung. In: Wolfgang Böttcher, Wilfried Bos, Hans Döbert und Heinz Günter Holtappels (Hg.): Bildungsmonitoring und Bildungscontrolling in nationaler und internationaler Perspektive. Dokumentation zur Herbsttagung der Kommission Bildungsorganisation, -planung, -recht (KBBB). Münster: Waxmann, S. 177–182.

Denzin, Norman K. [2](2003): Symbolischer Interaktionismus. In: Uwe Flick, Ernst von Kardoff und Ines Steinke (Hg.): Qualitative Forschung. Ein Handbuch. Reinbeck bei Hamburg: Rowohlt, S. 136–150.

Deutscher Bildungsserver: Bildungsstandards. Online verfügbar unter http://www.bildungsserver.de/zeigen.html?seite=1584, zuletzt geprüft am 22.12.2009.

Deutscher Bildungsserver: Vergleichsarbeiten/ Lernstandserhebungen in den Ländern. Online verfügbar unter http://www.bildungsserver.de/zeigen.html?seite=4452, zuletzt geprüft am 22.12.2009.

Ditton, Hartmut (2000): Qualitätskontrolle und Qualitätssicherung in Schule und Unterricht. Ein Überblick zum Stand der empirischen Forschung. In: Andreas Helmke, Walter Hornstein und Ewald Terhart (Hg.): Qualität und Qualitätssicherung im Bildungsbereich: Schule, Sozialpädagogik, Hochschule: Beltz, S. 73–92.

Ditton, Hartmut (2007): Schulqualität - Modelle zwischen Konstruktion, empirischen Befunden und Implementierung. In: Jürgen van Buer und Cornelia Wagner (Hg.): Qualität von Schule. Ein kritisches Handbuch. Frankfurt am Main: Lang, S. 83–92.

Ditton, Hartmut (2008): Qualitätssicherung in Schulen. In: Eckhard Klieme und Rudolf Tippelt (Hg.): Qualitätssicherung im Bildungswesen: Eine aktuelle Zwischenbilanz. ZfPäd. 53. Beiheft: Beltz, S. 36–58.

Ditton, Hartmut; Arnoldt Bettina (2004): Schülerbefragungen zum Fachunterricht - Feedback an Lehrkräfte. In: *Empirische Pädagogik* 18 (1), S. 115–139.

Dobbelstein, Peter; Peek, Rainer (2008): Lernstandserhebungen und zentrale Prüfungen im Kontext der Qualitätsanalyse. In: Wilfried Bos (Hg.): Schulische Qualitätsanalyse in Nordrhein-Westfalen. Konzepte, erste Erfahrungen, Perspektiven. Köln: Luchterhand, S. 17–35.

Döbert, Hans; Klieme, Eckhard; von Kopp; Botho; Weiß, Manfred (2003): Vertiefender Vergleich der Schulsysteme ausgewählter PISA-Teilnehmerstaaten. Berlin: BMBF.

Döbert, Hartmut (2008): Indikatorengestützte Bildungsberichterstattung. eine Einführung. In: Wolfgang Böttcher, Wilfried Bos, Hans Döbert und Heinz Günter Holtappels (Hg.): Bildungsmonitoring und Bildungscontrolling in nationaler und internationaler Perspektive. Dokumentation zur Herbsttagung der Kommission Bildungsorganisation, -planung, -recht (KBBB). Münster: Waxmann, S. 13–18.

Eder, Ferdinand (Hg.) [1](2002): Qualitätsentwicklung und Qualitätssicherung im österreichischen Schulwesen. Unter Mitarbeit von Peter Posch, Michael Schratz, Werner Specht und Josef Thonhauser. Innsbruck: Studien-Verlag.

Ehmke, Timo; Jude, Nina (2010): Soziale Herkunft und Kompetenzerwerb. In: Eckhard Klieme, Cordula Artelt, Johannes Hartig, Nina Jude, Olaf Köller, Manfred Prenzel et al. (Hg.): PISA 2009. Bilanz nach einem Jahrzehnt: Waxmann, S. 231–253.

Fend, Helmut (1986): "Gute Schulen - schlechte Schulen". Die einzelne Schule als pädagogische Handlungseinheit. In: *Die Deutsche Schule* 78 (3), S. 275–293.

Fend, Helmut (2000): Qualität und Qualitätssicherung im Bildungswesen: Wohlfahrtsstaatliche Modelle und Marktmodelle. In: Andreas Helmke, Walter Hornstein und Ewald Terhart (Hg.): Qualität und Qualitätssicherung im Bildungsbereich: Schule, Sozialpädagogik, Hochschule: Beltz, J, S. 55–72.

Fend, Helmut (2005): Systemsteuerung im Bildungswesen - Anschlussfähigkeit an die Schulwirklichkeit. In: Katharina Maag Merki, Anita Sandmeier, Patricia Schuler und Helmut Fend (Hg.): Schule wohin? Schulentwicklung und Qualitätsmanagement im 21. Jahrhundert. Berichte aus dem Forschungsbereich Schulqualität & Schulentwicklung. Zürich: FS&S (Forschungsbereich Schulqualität und Schulentwicklung) Universität Zürich (Schriftenreihe zu "Bildungssystem und Humanentwicklung"), S. 15–27.

Fend, Helmut (2008a): Dimensionen von Qualität im Bildungswesen. Von Produktindikatoren zu Prozessindikatoren am Beispiel der Schule. In: Eckhard Klieme und Rudolf Tippelt (Hg.): Qualitätssicherung im Bildungswesen: Eine aktuelle Zwischenbilanz. ZfPäd. 53. Beiheft: Beltz, J, S. 190–209.

Fend, Helmut (2008b): Schule gestalten. Systemsteuerung, Schulentwicklung und Unterrichtsqualität. Wiesbaden: VS Verlag für Sozialwissenschaften.

Fleischer, Jens; Spoden, Christian; Wirth, Joachim; Leutner, Detlev (2008): Flächendeckende Lernstandserhebungen - spezifische Herausforderungen und Lösungsansätze. Das Beispiel lernstand 8 in Nordrhein-Westfalen. In: Wolfgang Böttcher, Wilfried Bos, Hans Döbert und Heinz Günter Holtappels (Hg.): Bildungsmonitoring und Bildungscontrolling in nationaler und internationaler Perspektive. Dokumentation zur Herbsttagung der Kommission Bildungsorganisation, -planung, -recht (KBBB). Münster: Waxmann, S. 195–207.

Flick, Uwe (2010): Qualitative Sozialforschung. Eine Einführung. Reinbeck bei Hamburg: Rowohlt-Taschenbuch-Verlag.

Flick, Uwe; Kardoff, Ernst von; Steinke, Ines [2](2003): Was ist qualitative Forschung? Einleitung und Überblick. In: Uwe Flick, Ernst von Kardoff und Ines Steinke (Hg.): Qualitative Forschung. Ein Handbuch. 2. Auflage. Reinbeck bei Hamburg: Rowohlt, S. 13–29.

Friebertshäuser, Barbara; Langer, Antje [3](2010): Interviewformen und Interviewpraxis. In: Barbara Friebertshäuser, Antje Langer, Annedore Prengel, Heike Boller und Sophia Richter (Hg.): Handbuch Qualitative Forschungsmethoden in der Erziehungswissenschaft. Weinheim: Juventa-Verlag, S. 437–455.

Friebertshäuser, Barbara; Langer, Antje; Prengel, Annedore; Boller, Heike; Richter, Sophia (Hg.) [3](2010): Handbuch Qualitative Forschungsmethoden in der Erziehungswissenschaft. Weinheim: Juventa-Verlag.

Fuchs, Hans-Werner [1](2008): Educational Governance und Neue Steuerung: Grundsätze - Beispiele - Erwartungen. In: Roman Langer (Hg.): "Warum tun die das?". Governanceanalysen zum Steuerungshandeln in der Schulentwicklung. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 19–30.

Gahleitner, Silke-Birgitta (2005): Die Qualitative Inhaltsanalyse als flexible Auswertungsmethode in der Sozialarbeitsforschung. In: Silke-Birgitta Gahleitner, Susanne Gerull, Begona Petuya Ituarte, Lydia Schambacher-Hardtke und Claudia Streblow (Hg.): Einführung in das Methodenspektrum sozialwissenschaftlicher Forschung. Uckerland: Schibri-Verlag, S. 53–63.

Galiläer, Lutz (2005): Pädagogische Qualität. Perspektiven der Qualitätsdiskurse über Schule, soziale Arbeit und Erwachsenenbildung. Weinheim: Juventa-Verlag.

Gathen, Jan von der (2006): Die innerschulische Rezeption von Leistungsrückmeldungen aus Large-Scale-Assessments - Grundlagen und Ziele von Fallstudien. In: Harm Kuper und Julia Schneewind (Hg.): Rückmeldung und Rezeption von Forschungsergebnissen. Zur Verwendung wissenschaftlichen Wissens im Bildungssystem. Münster: Waxmann, S. 77–88.

Gehrmann, Axel; Hericks Uwe; Lüders, Manfred (Hg.) (2010): Bildungsstandards und Kompetenzmodelle. Beiträge zu einer aktuellen Diskussion über Schule, Lehrerbildung und Unterricht. Bad Heilbrunn: Klinkhardt.

Glaser, Barney G.; Strauss, Anselm L. (1967): The discovery of grounded theory. Strategies for qualitative research. New York, NY: Aldine.

Groß Ophoff, Jana; Hosenfeld, Ingmar; Koch, Ursula (2007): Formen der Ergebnisrezeption und damit verbundene Schul- und Unterrichtsentwicklung. In: *Empirische Pädagogik* 21 (4), S. 411–427.

Groß Ophoff, Jana; Koch, Ursula; Hosenfeld; Ingmar; Helmke, Andreas (2006): Ergebnisrückmeldungen und ihre Rezeption im Projekt VERA. In: Harm Kuper und Julia Schneewind (Hg.): Rückmeldung und Rezeption von Forschungsergebnissen. Zur Verwendung wissenschaftlichen Wissens im Bildungssystem. Münster: Waxmann, S. 19–40.

Hanft, Anke (2008): Bildungs- und Wissenschaftsmanagement. Unter Mitarbeit von Heinke Röbken, Marco Zimmer und Frank Fischer. München: Vahlen.

Hartung-Beck, Viola [1](2009): Schulische Organisationsentwicklung und Professionalisierung. Folgen von Lernstandserhebung an Gesamtschulen. Wiesbaden: VS Verlag für Sozialwissenschaften.

Heid, Helmut (2000): Qualität: Überlegungen zur Begründung einer pädagogischen Beurteilungskategorie. In: Andreas Helmke, Walter Hornstein und Ewald Terhart (Hg.): Qualität und Qualitätssicherung im Bildungsbereich: Schule, Sozialpädagogik, Hochschule: Beltz, J, S. 41–51.

Heid, Helmut (2007): Qualität von Schule. Zwischen Bildungstheorie und Bildungsökonomie. In: Jürgen van Buer und Cornelia Wagner (Hg.): Qualität von Schule. Ein kritisches Handbuch. Frankfurt am Main: Lang, S. 55–66.

Heinrich, Martin (2007): Governance in der Schulentwicklung. Von der Autonomie zur evaluationsbasierten Steuerung. Wiesbaden: VS Verlag für Sozialwissenschaften.

Helmke, Andreas (2006): Was wissen wir über guten Unterricht? Über die Notwendigkeit einer Rückbesinnung auf den Unterricht als dem "Kerngeschäft" der Schule. In: *Pädagogik* 2, S. 42–45.

Helmke, Andreas; Hosenfeld, Ingmar (2003): Vergleichsarbeiten (VERA): eine Standortbestimmung zur Sicherung schulischer Kompetenzen. Online verfügbar unter http://vera-web.uni-landau.de/verapub/fileadmin/downloads/schulverwaltung_hrs.pdf, zuletzt aktualisiert am 19.05.2003, zuletzt geprüft am 03.05.2010.

Helmke, Andreas; Schrader, Friedrich-Wilhelm (2006): Lehrerprofessionalität und Unterrichtsqualität. In: *Schulmagazin 5 - 10* 38 (9), S. 5–12.

Helmke, Andreas; Weinert, Franz-Emanuel [6](2007): Unterrichtsqualität erfassen, bewerten, verbessern. Seelze: Klett Kallmeyer.

Helmke, Andreas; Hornstein, Walter; Terhart, Ewald (Hg.) (2000): Qualität und Qualitätssicherung im Bildungsbereich: Schule, Sozialpädagogik, Hochschule: Beltz, J.

Hermanns, Harry (2003): Interviewen als Tätigkeit. In: Uwe Flick, Ernst von Kardoff und Ines Steinke (Hg.): Qualitative Forschung. Ein Handbuch. 2. Auflage. Reinbeck bei Hamburg: Rowohlt, S. 360–368.

Herrmann, Ulrich (Hg.) (2007): In der Pädagogik etwas bewegen. Impulse für Bildungspolitik und Schulentwicklung. Weinheim: Beltz.

Herzog, Walter (2010): Besserer Unterricht dank Bildungsstandards und Kompetenzmodellen? In: Axel Gehrmann, Hericks Uwe und Manfred Lüders (Hg.): Bildungsstandards und Kompetenzmodelle. Beiträge zu einer aktuellen Diskussion über Schule, Lehrerbildung und Unterricht. Bad Heilbrunn: Klinkhardt, S. 37–46.

Hoffmann-Göttig, Joachim; Eschmann, Winfried; Daumen, Cäcilie (2005): Und sie bewegt sich doch… Vom Umgang mit den Ergebnissen externer Evaluation aus Sicht von Bildungspolitik und Schulaufsicht. In: Gerold Becker, Albert Bremerich-Vos, Marianne Demmer, Katharina Maag Merki, Botho Priebe, Knut Schwippert et al. (Hg.): Standards. Unterrichten zwischen Kompetenzen, zentralen Prüfungen und Vergleichsarbeiten. Velber: Friedrich Verlag (Friedrich Jahresheft XXIII), S. 32–36.

Hopf, Christel [2](2003): Qualitative Interviews - ein Überblick. In: Uwe Flick, Ernst von Kardoff und Ines Steinke (Hg.): Qualitative Forschung. Ein Handbuch. Reinbeck bei Hamburg: Rowohlt, S. 349–360.

Hosenfeld, Ingmar (2005): Rezeption - Reflexion - Aktion. Wie lassen sich Lernstandserhebungen und Vergleichsarbeiten pädagogisch nutzen? In: Gerold Becker, Albert Bremerich-Vos, Marianne Demmer, Katharina Maag Merki, Botho Priebe, Knut Schwippert et al. (Hg.): Standards. Unterrichten zwischen Kompetenzen, zentralen Prüfungen und Vergleichsarbeiten. Velber: Friedrich Verlag (Friedrich Jahresheft XXIII), S. 112–114.

Hosenfeld, Ingmar; Groß Ophoff, Jana (2007): Nutzung und Nutzen von Evaluationsstudien in Schule und Unterricht. In: *Empirische Pädagogik* (21 (4)), S. 352–367.

Hosenfeld, Ingmar; Schrader, Friedrich-Wilhelm; Helmke, Tuyet (2006): Von der Rezeption zur Ergebnisrückmeldung: Leistungsevaluation im Spannungsfeld von System-Monitoring und Schulentwicklung. In: Ingmar Hosenfeld und Friedrich-Wilhelm Schrader (Hg.): Schulische Leistung. Grundlagen, Bedingungen, Perspektiven. Münster: Waxmann, S. 289–313.

Hovestadt, Gertrud; Keßler, Nicole (2005): 16 Bundesländer. Eine Übersicht zu Bildungsstandards und Evaluationen. In: Gerold Becker, Albert Bremerich-Vos, Marianne Demmer, Katharina Maag Merki, Botho Priebe, Knut Schwippert et al. (Hg.): Standards. Unterrichten zwischen Kompeten-

zen, zentralen Prüfungen und Vergleichsarbeiten. Velber: Friedrich Verlag (Friedrich Jahresheft XXIII), S. 8–10.

Institut für Qualitätsentwicklung im Bildungswesen (IQB) (2010a): Das EMSE-Netzwerk. Übersicht.

Institut für Qualitätsentwicklung im Bildungswesen (IQB) (2010b): Publikationsliste QuaSUM. Online verfügbar unter http://www.iqb.hu-berlin.de/arbbereiche/fdz/Studien/stud_2/Publikationslis.pdf, zuletzt aktualisiert am 25.08.2010, zuletzt geprüft am 29.06.2011.

Kelle, Udo; Kluge, Susann [2](2010): Vom Einzelfall zum Typus. Fallvergleich und Fallkontrastierung in der qualitativen Sozialforschung. Wiesbaden: VS Verlag für Sozialwissenschaften

Kelle, Udo; Kluge, Susanne (1999): Vom Einzelfall zum Typus. Fallvergleich und Fallkontrastierung in der qualitativen Sozialforschung. Opladen: Leske & Budrich.

Kempfert, Guy; Rolff, Hans-Günter (1999): Pädagogische Qualitätsentwicklung. Ein Arbeitsbuch für Schule und Unterricht. Weinheim: Beltz.

Klieme, Eckhard (2005): Bildungsqualität und Standards. Anmerkungen zu einem umstrittenen Begriffspaar. In: Gerold Becker, Albert Bremerich-Vos, Marianne Demmer, Katharina Maag Merki, Botho Priebe, Knut Schwippert et al. (Hg.): Standards. Unterrichten zwischen Kompetenzen, zentralen Prüfungen und Vergleichsarbeiten. Velber: Friedrich Verlag (Friedrich Jahresheft XXIII), S. 6–7.

Klieme, Eckhard; Artelt, Cordula; Hartig, Johannes; Jude, Nina; Köller, Olaf; Prenzel, Manfred et al. (Hg.) (2010): PISA 2009. Bilanz nach einem Jahrzehnt: Waxmann.

Klieme, Eckhard; Avenarius, Hermann; Blum, Werner (2003): Zur Entwicklung nationaler Bildungsstandards. Eine Expertise. Bonn: BMBF

Klug, Christian; Reh, Sabine (2000): Was fangen die Schulen mit den Ergebnissen an? Die Hamburger Leistungsvergleichsstudie aus der Sicht "beforschter" Schulen. In: *Pädagogik* 52 (12), S. 16–21.

Kluge, Susann (1999): Empirisch begründete Typenbildung. Zur Konstruktion von Typen und Typologien in der qualitativen Sozialforschung. Opladen: Leske & Budrich.

KMK (1997): Grundsätzliche Überlegungen zu Leistungsvergleichen innerhalb der Bundesrepublik Deutschland. Konstanzer Beschluss. Online verfügbar unter http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/1997/1997_10_24-Konstanzer-Beschluss.pdf, zuletzt aktualisiert am 18.06.2009, zuletzt geprüft am 28.04.2010.

KMK (2001): Pressemitteilung der 296. Plenarsitzung der Kultusministerkonferenz am 05./06. Dezember in Bonn. Kultusministerium beschließt konkrete Maßnahmen zur Verbesserung der schulischen Bildung in Deutschland.

KMK (2002a): Ergebnisse der 298. Plenarsitzung der Kultusministerkonferenz am 23. und 24. Mai 2002 in Eisenach.

KMK (2002b): Ergebnisse der 299.Plenarsitzung: Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (KMK). Online verfügbar unter http://www.kmk.org/presse-und-aktuelles/pm2002/ergebnisse-der-299plenarsitzung.html, zuletzt geprüft am 02.05.2010.

KMK (2004): Gründung des Instituts zur Qualitätsentwicklung im Bildungswesen. Online verfügbar unter http://www.kmk.org/fileadmin/pdf/PresseUndAktuelles/2004/pm041207_227_1.pdf, zuletzt aktualisiert am 07.12.2004, zuletzt geprüft am 28.04.2010.

KMK (2006): Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring. Online verfügbar unter http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2006/2006_01_01-Gesamtstrategie-Endf.pdf, zuletzt aktualisiert am 13.10.2006, zuletzt geprüft am 03.05.2010.

KMK (2007a): Pressemitteilung - Aktuelle Forschungsinteressen der Länder im Rahmen der Gemeinschaftsaufgabe von Bund und Ländern gemäß Art. 91b Abs. 2. Online verfügbar unter

http://www.kmk.org/fileadmin/pdf/PresseUndAktuelles/2007/071113forschungsint-laender.pdf, zuletzt aktualisiert am 13.11.2007, zuletzt geprüft am 28.04.2010.

KMK (2007b): Pressemitteilung - Neue Steuerungselemente für weitere Verbesserung des Bildungswesens. Online verfügbar unter http://www.kmk.org/presse-und-aktuelles/pm2007/verbesserung-des-bildungswesens.html, zuletzt geprüft am 28.04.2010.

KMK (2008): Institut zur Qualitätsentwicklung im Bildungswesen (IQB) kann auf neuer Grundlage weiterarbeiten. Online verfügbar unter http://www.kmk.org/presse-und-aktuelles/meldung/institut-zur-qualitaetsentwicklung-im-bildungswesen-iqb-kann-auf-neuer-grundlage-weiterarbeiten.html, zuletzt geprüft am 28.04.2010.

KMK (2009): Konzeption der Kultusministerkonferenz zur Nutzung der Bildungsstandards für die Unterrichtsentwicklung. Online verfügbar unter http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2009/2009_12_10-Konzeption-Bildungsstandards.pdf, zuletzt aktualisiert am 09.03.2010, zuletzt geprüft am 28.04.2010.

KMK (2010a): Dokumente: nationale Bildungsstandards für alle Schularten. Online verfügbar unter http://www.kmk.org/bildung-schule/qualitaetssicherung-in-schulen/bildungsstandards/dokumente.html, zuletzt geprüft am 28.04.2010.

KMK (2010b): Überblick: Gesamtstrategie zum Bildungsmonitoring: Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (KMK). Online verfügbar unter http://www.kmk.org/bildung-schule/qualitaetssicherung-in-schulen/bildungsmonitoring/ueberblick-gesamtstrategie-zum-bildungsmonitoring.html, zuletzt geprüft am 28.04.2010.

KMK (2010c): Überblick: Qualitätssicherung in Schulen. Bildungsstandards. Online verfügbar unter http://www.kmk.org/bildung-schule/qualitaetssicherung-in-schulen/bildungsstandards/ueberblick.html, zuletzt geprüft am 28.04.2010.

Kohler, Britta (2005): Rezeption internationaler Schulleistungsstudien. Wie gehen Lehrkräfte, Eltern und die Schulaufsicht mit Ergebnissen schulischer Evaluationsstudien um? Münster: Waxmann.

Kohler, Britta; Schrader, Friedrich-Wilhelm (2004): Ergebnisrückmeldung und Rezeption: Von der externen Evaluation zur Entwicklung von Schule und Unterricht. In: *Empirische Pädagogik* 18 (1), S. 3–17.

Köller, Olaf (2007): Standards und Qualitätssicherung zur Outputsteuerung im System und in der Einzelinstitution. In: Jürgen van Buer und Cornelia Wagner (Hg.): Qualität von Schule. Ein kritisches Handbuch. Frankfurt am Main: Lang, S. 93–102.

Köller, Olaf (2008): Bildungsstandards in einem Gesamtsystem der Qualitätssicherung im allgemeinbildenden Schulsystem Deutschlands. In: Eckhard Klieme und Rudolf Tippelt (Hg.): Qualitätssicherung im Bildungswesen: Eine aktuelle Zwischenbilanz. Zeitschrift für Pädagogik. 53. Beiheft: Beltz, J, S. 59–75.

König, Eckard (1995): Qualitative Forschung subjektiver Theorien. In: Eckard König und Peter Zedler (Hg.): Bilanz qualitativer Forschung. Band II: Methoden. Weinheim: Deutscher Studien Verlag, S. 11–29.

König, Eckard; Zedler, Peter (Hg.) (1995): Bilanz qualitativer Forschung. Band I: Grundlagen qualitativer Forschung. Weinheim: Deutscher Studien Verlag.

Konsortium Bildungsberichterstattung (2005): Gesamtkonzeption der Bildungsberichterstattung. Online verfügbar unter http://www.bildungsbericht.de/daten/gesamtkonzeption.pdf, zuletzt aktualisiert am 31.05.2006, zuletzt geprüft am 26.04.2010.

Kuckartz, Udo [3](2010): Einführung in die computergestützte Analyse qualitativer Daten. Wiesbaden: VS Verlag für Sozialwissenschaften

Kuckartz, Udo; Grunenberg, Heiko [3](2010): Qualitative Daten computergestützt auswerten: Methoden, Techniken, Software. In: Barbara Friebertshäuser, Antje Langer, Annedore Prengel, Heike Boller und Sophia Richter (Hg.): Handbuch Qualitative Forschungsmethoden in der Erziehungswissenschaft. Weinheim: Juventa-Verlag, S. 501–514.

Kühle, Ben; Peek, Rainer (2007): Lernstandserhebungen in Nordrhein-Westfalen. Evaluationsbefunde zur Rezeption und zum Umgang mit Ergebnisrückmeldungen in Schulen. In: *Empirische Pädagogik* 21 (4), S. 428–447.

Kuper, Harm (2002): Stichwort: Qualität im Bildungssystem. In: *Zeitschrift für Erziehungswissenschaft* 5 (4), S. 533–551.

Kuper, Harm; Schneewind, Julia (Hg.) (2006): Rückmeldung und Rezeption von Forschungsergebnissen. Zur Verwendung wissenschaftlichen Wissens im Bildungssystem. Münster: Waxmann.

Kussau, Jürgen; Brüsemeister, Thomas (2007): Educational Governance: Zur Analyse der Handlungskoordination im Mehrebenensystem der Schule. In: Herbert Altrichter, Thomas Brüsemeister und Jochen Wissinger (Hg.): Educational Governance. Handlungskoordination und Steuerung im Bildungssystem. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 15–54.

Lamnek, Siegfried (2005): Qualitative Sozialforschung. Lehrbuch. 4. vollständig überarbeitete Auflage.

Landesinstitut für Schule und Weiterbildung Nordrhein-Westfalen (Hg.) (1995): Evaluation und Schulentwicklung: Ansätze, Beispiele und Perspektiven aus der Fortbildungsmaßnahme Schulentwicklung und Schulaufsicht. Bönen: Verlag für Schule und Weiterbildung.

Landesinstitut für Schulentwicklung (LS), Stuttgart: Vergleichsarbeiten "DVA". Hg. v. Landesbildungsserver Baden-Württemberg, Stuttgart.

Landesinstitut für Schulentwicklung (LS), Stuttgart (2007): Qualitätsentwicklung und Evaluation. Landesbildungsserver Baden-Württemberg, Stuttgart. Online verfügbar unter http://www.schule-bw.de/entwicklung/qualieval/, zuletzt aktualisiert am 08.01.2007, zuletzt geprüft am 22.12.2009.

Landesinstitut für Schulentwicklung (LS), Stuttgart (2009): Diagnose- und Vergleichsarbeiten. Landesbildungsserver Baden-Württemberg, Stuttgart. Online verfügbar unter http://lbsneu.schule-bw.de/entwicklung/dva, zuletzt aktualisiert am 05.03.2009, zuletzt geprüft am 22.12.2009.

Landratsamt Ostalbkreis (Hg.) (2011): Bildung im Ostalbkreis 2011. Erster indikatorengestützter Bericht.

Langer, Antje [3](2010): Transkribieren - Grundlagen und Regeln. In: Barbara Friebertshäuser, Antje Langer, Annedore Prengel, Heike Boller und Sophia Richter (Hg.): Handbuch Qualitative Forschungsmethoden in der Erziehungswissenschaft. Weinheim: Juventa-Verl. (Juventa-Handbuch), S. 515–526.

Langer, Roman (Hg.) (2008): "Warum tun die das?". Governanceanalysen zum Steuerungshandeln in der Schulentwicklung. Wiesbaden: VS Verlag für Sozialwissenschaften.

Legewie, Heiner (1994): Globalauswertung von Dokumenten. In: Andreas Boehm, Andreas Mengel und Thomas Muhr (Hg.): Texte verstehen. Kozepte, Methoden, Werkzeuge. Konstanz: Universitätsverlag Konstanz, S. 177–182.

Lehmann, Rainer H. (2008): Internationale Ansätze zu einer Strategie des Bildungsmonitoring - aktuelle Diskussion und mögliche Konsequenzen für Deutschland. In: Wolfgang Böttcher, Wilfried Bos, Hans Döbert und Heinz Günter Holtappels (Hg.): Bildungsmonitoring und Bildungscontrolling in nationaler und internationaler Perspektive. Dokumentation zur Herbsttagung der Kommission Bildungsorganisation, -planung, -recht (KBBB). Münster: Waxmann, S. 19–33.

Lehmann, Rainer H.; Peek, Rainer; Gänsfuß, Rüdiger Lutkat Sabine; Mücke, Stephan; Barth, Ingola (1999): QuaSUM. Qualitätsuntersuchung an Schulen zum Unterricht in Mathematik. Ergebnisse einer repräsentativen Untersuchung im Land Brandenburg. Potsdam

Maier, Uwe (2008a): Rezeption und Nutzung von Vergleichsarbeiten aus der Perspektive von Lehrkräften. In: *Zeitschrift für Pädagogik* 54 (1), S. 95–115.

Maier, Uwe (2008b): Vergleichsarbeiten im Vergleich - Akzeptanz und wahrgenommener Nutzen standardbasierter Leistungsmessungen in Baden-Württemberg und Thüringen. In: *Zeitschrift für Erziehungswissenschaft* 11 (3), S. 454–474.

Maier, Uwe (2009a): Professionelle Nutzung von Vergleichsarbeiten? Ergebnisse einer qualitativen Interviewstudie mit Lehrkräften in Baden-Württemberg. In: *Lernen aus Evaluationsergebnissen*, S. 131–144.

Maier, Uwe (2009b): Wie gehen Lehrerinnen und Lehrer mit Vergleichsarbeiten um? Eine Studie zu testbasierten Schulreformen in Baden-Württemberg und Thüringen. Baltmannsweiler: Schneider-Verlag Hohengehren.

Maier, Uwe (2010): Formative und summative Aspekte testbasierter Schulreformen. In: Axel Gehrmann, Hericks Uwe und Manfred Lüders (Hg.): Bildungsstandards und Kompetenzmodelle. Beiträge zu einer aktuellen Diskussion über Schule, Lehrerbildung und Unterricht. Bad Heilbrunn: Klinkhardt, S. 47–54.

Maritzen, Norbert (2001): Schulleistungsforschung und Schulentwicklung, zwei Seiten einer Medaille. In: *Journal für Schulentwicklung* 5 (2), S. 46–54.

Marotzki, Winfried (2003): Leitfadeninterview. In: Ralf Bohnsack, Winfried Marotzki und Michael Meuser (Hg.): Hauptbegriffe Qualitativer Sozialforschung. Ein Wörterbuch. Opladen: Leske + Budrich.

Mayring, Philipp [8](2003): Qualitative Inhaltsanalyse. Grundlagen und Techniken. Weinheim: Beltz.

Mayring, Philipp [9](2007): Qualitative Inhaltsanalyse. Grundlagen und Techniken. Weinheim: Beltz.

Mayring, Philipp [10](2008): Qualitative Inhaltsanalyse. Grundlagen und Techniken. Weinheim: Beltz.

Mayring, Philipp; Brunner, Eva [3](2010): Qualitative Inhaltsanalyse. In: Barbara Friebertshäuser, Antje Langer, Annedore Prengel, Heike Boller und Sophia Richter (Hg.): Handbuch Qualitative Forschungsmethoden in der Erziehungswissenschaft. Weinheim: Juventa-Verlag, S. 323–333.

Merkens, Hans (1997): Stichproben bei qualitativen Studien. In: Barbara Friebertshäuser und Annedore Prengel (Hg.): Handbuch Qualitative Forschungsmethoden in der Erziehungswissenschaft. Weinheim & München: Juventa, S. 97–106.

Merkens, Hans (2003): Auswahlverfahren, Sampling, Fallkonstruktion. In: Uwe Flick, Ernst von Kardoff und Ines Steinke (Hg.): Qualitative Forschung. Ein Handbuch. 2. Auflage. Reinbeck bei Hamburg: Rowohlt, S. 286–299.

Ministerium für Schule Jugend und Sport des Landes Brandenburg (Hg.) (2004): Schulrückmeldungen von Schulleistungsstudien am Beispiel des QuaSUM-Projektes. Zwei Untersuchungen zur Wirksamkeit. Teltow.

Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen (2011): Lernstandserhebungen in Klasse 8. Online verfügbar unter http://www.schulministerium.nrw.de/BP/Schulsystem/Qualitaetssicherung/Lernstandserhebungen/Lernstandserhebungen_Klasse_8/index.html, zuletzt aktualisiert am 28.03.2011, zuletzt geprüft am 28.03.2011.

Nachtigall, Christof; Jantowski, Andreas (2007): Die Thüringer Kompetenztests unter besonderer Berücksichtigung der Evaluationsergebnisse zum Rezeptionsverhalten. In: *Empirische Pädagogik* 21 (4), S. 401–410.

Nachtigall, Christof; Kröhne, Ulf (2006): Methodische Anforderungen an schulische Leistungsmessung - auf dem Weg zu fairen Vergleichen. In: Harm Kuper und Julia Schneewind (Hg.): Rückmeldung

und Rezeption von Forschungsergebnissen. Zur Verwendung wissenschaftlichen Wissens im Bildungssystem. Münster: Waxmann, S. 59–74.

Netzwerk Empiriegestützte Schulentwicklung (EMSE) (2006): Erstes Positionspapier zu: Zentral standardisierte Lernstandserhebungen, zuletzt aktualisiert am 11.07.2007, zuletzt geprüft am 03.05.2010.

Nilshon, Ilse (2004): Qualitätssicherung auf der Grundlage von System-Monitoring. Sechs Fallstudien an Brandenburger Gymnasien. (Projekt QuaSUM 3). In: Ministerium für Schule Jugend und Sport des Lndes Brandenburg (Hg.): Schulrückmeldungen von Schulleistungsstudien am Beispiel des QuaSUM-Projektes. Zwei Untersuchungen zur Wirksamkeit. Teltow.

O'Day, Jennifer (2008): Standards-Based Reform - Promises, Pitfalls, and Potential Lessons from the U.S. In: Wolfgang Böttcher, Wilfried Bos, Hans Döbert und Heinz Günter Holtappels (Hg.): Bildungsmonitoring und Bildungscontrolling in nationaler und internationaler Perspektive. Dokumentation zur Herbsttagung der Kommission Bildungsorganisation, -planung, -recht (KBBB). Münster: Waxmann, S. 107–127.

Oelkers, Jürgen; Reusser, Kurt (2008): Qualität entwickeln - Standards sichern - mit Differenz umgehen. Hg. v. Bundesministerium für Bildung und Forschung (BMBF). Bonn, Berlin.

Orth, Gerhard (2005): Eine Chance für die Qualität schulischer Arbeit. Standards und zentrale Leistungsüberprüfungen. In: Gerold Becker, Albert Bremerich-Vos, Marianne Demmer, Katharina Maag Merki, Botho Priebe, Knut Schwippert et al. (Hg.): Standards. Unterrichten zwischen Kompetenzen, zentralen Prüfungen und Vergleichsarbeiten. Velber: Friedrich Verlag (Friedrich Jahresheft XXIII), S. 63–65.

Pant, Hans Anand; Vock, Miriam; Pöhlmann, Claudia; Köller, Olaf (2008): Offenheit für Innovationen. Befunde aus einer Studie zur Rezeption der Bildungsstandards bei Lehrkräften und Zusammenhänge mit Schülerleistungen. In: *Zeitschrift für Pädagogik* 54 (6), S. 827–845.

Peek, Rainer [2](2002): Die Bedeutung vergleichender Schulleistungsmessungen für die Qualitätskontrolle und Qualitätsentwicklung von Schulen und Schulsystemen. In: Franz E. Weinert (Hg.): Leistungsmessungen in Schulen. Weinheim: Beltz, S. 323–336.

Peek, Rainer (2004a): Klassenbezogene Rückmeldungen aus Schulleistungsstudien und ihre Rezeption in beteiligten Schulen im Land Brandenburg. (Projekt QuaSUM 2). In: Ministerium für Schule Jugend und Sport des Landes Brandenburg (Hg.): Schulrückmeldungen von Schulleistungsstudien am Beispiel des QuaSUM-Projektes. Zwei Untersuchungen zur Wirksamkeit. Teltow, S. 9–111.

Peek, Rainer (2004b): Qualitätsuntersuchung an Schulen zum Unterricht in Mathematik (QuaSUM) -. Klassenbezogene Ergebnisrückmeldungen und ihre Rezeption in Brandenburger Schulen. In: *Empirische Pädagogik* 18 (1), S. 82–114.

Peek, Rainer (2007): Interne Evaluation und einzelschulische Entwicklung - Spagat zwischen Mindeststandards und Machbarem. In: Jürgen van Buer und Cornelia Wagner (Hg.): Qualität von Schule. Ein kritisches Handbuch. Frankfurt am Main: Lang, S. 141–149.

Peek, Rainer; Dobbelstein, Peter (2006a): Benchmarks als Input für die Schulentwicklung - das Beispiel der Lernstandserhebungen in Nordrhein-Westfalen. In: Harm Kuper und Julia Schneewind (Hg.): Rückmeldung und Rezeption von Forschungsergebnissen. Zur Verwendung wissenschaftlichen Wissens im Bildungssystem. Münster: Waxmann, S. 41–58.

Peek, Rainer; Dobbelstein, Peter (2006b): Zielsetzung: Ergebnisorientierte Schul- und Unterrichtsentwicklung. Potenziale und Grenzen der nordrhein-westfälischen Lernstandserhebungen. In: Wolfgang Böttcher, Heinz Günther Holtappels und Michaela Brohm (Hg.): Evaluation im Bildungswesen. Eine Einführung in Grundlagen und Praxisbeispiele. Weinheim: Juventa-Verlag, S. 177–193.

Peek, Rainer; Steffens, Ulrich; Köller, Olaf (2006): Positionspapier des Netzwerks Empiriegestützte Schulentwicklung (EMSE) zu: Zentrale standardisierte Lernstandserhebungen. 5. EMSE-Tagung, Berlin 08.12.2006.

Philologenverband Baden-Württemberg (PhV BW) (2008): Ein Wegfall der Benotung schmälert den Stellenwert von Vergleichsarbeiten nicht. Online verfügbar unter http://bildungsklick.de/pm/63772/bernd-saur-ein-wegfall-der-benotung-schmaelert-den-stellenwert-von-vergleichsarbeiten-nicht/, zuletzt geprüft am 10.05.2010.

Popham, W. James (2003): The seductive allure of data. In: *Educational Leadership* 60 (5), S. 48–51.

Posch, Peter; Altrichter, Herbert (1997): Möglichkeiten und Grenzen der Qualitätsevaluation und -entwicklung im Schulwesen. Innsbruck: Studien-Verlag.

Radisch, Falk (2008): Von FIMS bis PIRLS und PISA. Deutschlands Abschneiden bei internationalen Schulleistungsvergleichen. In: Wolfgang Böttcher, Wilfried Bos, Hans Döbert und Heinz Günter Holtappels (Hg.): Bildungsmonitoring und Bildungscontrolling in nationaler und internationaler Perspektive. Dokumentation zur Herbsttagung der Kommission Bildungsorganisation, -planung, -recht (KBBB). Münster: Waxmann, S. 183–194.

Reinders, Heinz (2005): Qualitative Interviews mit Jugendlichen führen. Ein Leitfaden. München: Oldenbourg.

Rekus, Jürgen (2007): Qualitätssicherung durch Bildungsstandards. Schulaufsicht vor neuen Aufgaben? In: Johann J. Beichel und Konrad Fees (Hg.): Bildung oder outcome? Leitideen der standardisierten Schule im Diskurs. Herbolzheim: Centaurus Verlag, S. 53–71.

Rolff, Hans-Günter (1995): Evaluation - ein Ansatz zur Qualitätsentwicklung von Schulen? In: Landesinstitut für Schule und Weiterbildung Nordrhein-Westfalen (Hg.): Evaluation und Schulentwicklung: Ansätze, Beispiele und Perspektiven aus der Fortbildungsmaßnahme Schulentwicklung und Schulaufsicht. Bönen: Verlag für Schule und Weiterbildung, S. 293–310.

Rolff, Hans-Günter (1998): Entwicklung von Einzelschulen: Viel Praxis, wenig Theorie und kaum Forschung. Ein Versuch, Schulentwicklung zu systematisieren. In: *Jahrbuch der Schulentwicklung* (10), S. 295–325.

Rolff, Hans-Günter (2002a): Rückmeldung und Nutzung der Ergebnisse von großflächigen Leistungsuntersuchungen. Chancen und Grenzen. In: Hans-Günter Rolff (Hg.): Jahrbuch der Schulentwicklung, Bd. 12. Weinheim & München: Juventa, S. 75–98.

Rolff, Hans-Günter [2](2002b): Was bringt die vergleichende Leistungsmessung für die pädagogische Arbeit in Schulen? In: Franz E. Weinert (Hg.): Leistungsmessungen in Schulen. Weinheim: Beltz, S. 337–352.

Rossi, P.H; Lipsey, M.W; Freeman, H.E (2004): Evaluation: A systematic approach. 7. Aufl. Thousand Oaks, CA: Sage.

Schmidt, Christiane [3](2010): Auswertungstechniken für Leitfadeninterviews. In: Barbara Friebertshäuser, Antje Langer, Annedore Prengel, Heike Boller und Sophia Richter (Hg.): Handbuch Qualitative Forschungsmethoden in der Erziehungswissenschaft. Weinheim: Juventa-Verlag, S. 473–486.

Schneewind, Julia (2006): Rückmeldungen als Motivator für die Teilnahme an Schulleistungsstudien? Die Rezeptionsstudie von BeLesen. In: Harm Kuper und Julia Schneewind (Hg.): Rückmeldung und Rezeption von Forschungsergebnissen. Zur Verwendung wissenschaftlichen Wissens im Bildungssystem. Münster: Waxmann, S. 107–126.

Schneewind, Julia (2007): Erfahrungen mit Ergebnisrückmeldungen im Projekt BeLesen -. Ergebnisse einer Interviewstudie. In: *Empirische Pädagogik* 21 (4), S. 368–382.

Schrader, Friedrich-Wilhelm; Helmke, Andreas (2004): Von der Evaluation zur Innovation? Die Rezeptionsstudie WALZER: Ergebnisse der Lehrerbefragung. In: *Empirische Pädagogik* 18 (1), S. 140–161.

Schratz, Michael (2001): Was soll denn das schon wieder heißen? Ein Glossar zur Qualitätsdiskussion. In: *Friedrich Jahresheft* XIX, S. 138.

Schubert, Klaus; Klein, Martina [4](2006): Das Politiklexikon. Bonn: Dietz.

Schwippert, Knut; Goy, Martin [2](2008): Leistungsvergleichs- und Schulqualitätsforschung. In: Werner Helsper und Jeanette Böhme (Hg.): Handbuch der Schulforschung. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 387–421.

Schymala, Martin (2010): Wie kann man das Konzept formativer Leistungsmessung unter den Bedingungen testbasierter Schulreformen in Deutschland umsetzen? Diplomarbeit. Pädagogische Hochschule, Schwäbisch Gmünd. Institut für Erziehungswissenschaft.

Stadt Schwäbisch Gmünd (Hg.) (2010): Bildung in Schwäbisch Gmünd 2010. Erster indikatorengestützter Bericht.

Stamm, Margrit (2003a): Evaluation im Spiegel ihrer Nutzung: Grande idée oder grande illusion des 21. Jahrhunderts. Online verfügbar unter http://www.zfev.de/fruehereAusgabe/ausgabe2003-2/artikel/ZfEv2-2003_2-Stamm.pdf, zuletzt aktualisiert am 20.03.2006, zuletzt geprüft am 23.06.2011.

Stamm, Margrit (2003b): Evaluation und ihre Folgen für die Bildung. Eine unterschätzte pädagogische Herausforderung. Münster: Waxmann.

Stamm, Margrit (2005): Unterschätzt oder überschätzt? Die Folgen von Evaluation im Bildungsbereich. Vortrag auf der SEVAL-Jahrestagung, 03.06.2005. Online verfügbar unter http://perso.unifr.ch/margrit.stamm/forschung/fo_downloads/fo_dl_kongr/050603_Seval.pdf, zuletzt geprüft am 29.06.2011.

Steffens, Ulrich (2007): Schulqualitätsdiskussion in Deutschland - Ihre Entwicklung im Überblick. In: Jürgen van Buer und Cornelia Wagner (Hg.): Qualität von Schule. Ein kritisches Handbuch. Frankfurt am Main: Lang, S. 21–51.

Steffens, Ulrich (2008): Nutzung und Nutzen von Schulrückmeldungen im Rahmen standardisierter Lernstandserhebungen / Vergleichsarbeiten. Zweites Positionspapier des EMSE-Netzwerkes– verabschiedet auf der 9. EMSE-Fachtagung am 16.-17. Dezember 2008 in Nürnberg. Unter Mitarbeit von Rainer Peek.

Steinke, Ines [2](2003): Gütekriterien qualitativer Forschung. In: Uwe Flick, Ernst von Kardoff und Ines Steinke (Hg.): Qualitative Forschung. Ein Handbuch. Reinbeck bei Hamburg: Rowohlt, S. 319–331.

Stryck, Tom (2000): Qualitätssicherung in der Geisterbahn. Was hat Schulaufsicht mit Schulqualität zu tun? In: Andreas Helmke, Walter Hornstein und Ewald Terhart (Hg.): Qualität und Qualitätssicherung im Bildungsbereich: Schule, Sozialpädagogik, Hochschule: Beltz, J, S. 111–128.

Terhart, Ewald (2002): Wie können die Ergebnisse von vergleichenden Leistungsstudien systematisch zur Qualitätsverbesserung in Schulen genutzt werden? In: *Zeitschrift für Pädagogik* 48, S. 91–110.

Thürmann, Eike (2002): Unbekanntes Land. Wie Bildungsstandards das Lernen verbessern sollen. In: *Forum Schule* 3, S. 22–25.

Tillmann, Klaus-Jürgen (2006): Systemsteuerung durch Leistungsvergleiche und Bildungsstandards? Oder: Kritische Anmerkungen zum gegenwärtigen Zeitgeist. In: Martin Heinrich und Ulrike Greiner (Hg.): Schauen, was 'rauskommt. Kompetenzförderung, Evaluation und Systemsteuerung im Bildungswesen. Wien: Lit (Österreichische Beiträge zur Bildungsforschung, 3), S. 13–35.

Tillmann, Klaus-Jürgen (2007): Qualitätssicherung durch Leistungsvergleiche und Bildungsstandards? In: Ulrich Herrmann (Hg.): In der Pädagogik etwas bewegen. Impulse für Bildungspolitik und Schulentwicklung. Weinheim: Beltz, S. 38–45.

Tresch, Sarah (2007): Potenzial Leistungstest. Wie Lehrerinnen und Lehrer Ergebnisrückmeldungen zur Sicherung und Steigerung ihrer Unterrichtsqualität nutzen. Bern: hep.

Universität Landau (2006): VERA - Vergleichsarbeiten in der Grundschule. Online verfügbar unter http://www.uni-landau.de/vera/, zuletzt aktualisiert am 02.08.2006, zuletzt geprüft am 10.05.2010.

van Ackeren, Isabell (2002): Von FIMS und FISS bis TIMSS und PISA. Schulleistungen in Deutschland im historischen und internationalen Vergleich. In: *Die Deutsche Schule* 94 (2), S. 157–175.

van Ackeren, Isabell (2003): Evaluation, Rückmeldung und Schulentwicklung. Erfahrungen mit zentralen Tests, Prüfungen und Inspektionen in England, Frankreich und den Niederlanden. Münster: Waxmann.

van Ackeren, Isabell; Bellenberg, Gabriele (2004): Parallelarbeiten, Vergleichsarbeiten und Zentrale Abschlussprüfungen. Bestandsaufnahme und Perspektiven. In: Heinz Günther Holtappels, Klaus Klemm, Hermann Pfeiffer, Hans-Günter Rolff und Renate Schulz-Zander (Hg.): Jahrbuch der Schulentwicklung. Weinheim & München: Juventa (13), S. 125–159.

van Ackeren, Isabell; Klemm, Klaus (2011): Entstehung, Struktur und Steuerung des deutschen Schulsystems. Eine Einführung. Wiesbaden.

Watermann, Reiner; Stanat, Petra (2004): Schulrückmeldungen in PISA 2000: Sozialnorm- und kriteriumsorienterite Rückmeldeverfahren. In: *Empirische Pädagogik* 18 (1), S. 40–61.

Weinert, Franz E. (Hg.) [2](2002): Leistungsmessungen in Schulen. Weinheim: Beltz.

Wermke, Matthias; Kunkel-Razum, Kathrin; Scholze-Stubenrecht, Werner [9](2007): Der Duden. Das Fremdwörterbuch.

Wissinger, Jochen (2007): Does School Governance matter? Herleitungen und Thesen aus dem Bereich "School Effectiveness and School Improvement". In: Herbert Altrichter, Thomas Brüsemeister und Jochen Wissinger (Hg.): Educational Governance. Handlungskoordination und Steuerung im Bildungssystem. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 105–129.

Zlatkin-Troitschanskaia, Olga (2007a): Effizienz und Effektivität der Steuerung des öffentlichen Schulwesens - Rechtsökonomische Reflexionen. In: Jürgen van Buer und Cornelia Wagner (Hg.): Qualität von Schule. Ein kritisches Handbuch. Frankfurt am Main: Lang, S. 183–199.

Zlatkin-Troitschanskaia, Olga (2007b): Steuerungsfähigkeit des öffentlichen Schulwesens versus Steuerbarkeit der Schule - ein Paradigmenwechsel? In: Jürgen van Buer und Cornelia Wagner (Hg.): Qualität von Schule. Ein kritisches Handbuch. Frankfurt am Main: Lang, S. 67–92.

Abbildungsverzeichnis

Abbildung 1: Modell der Qualität im Bildungswesen nach Fend (2000, S. 58) 15
Abbildung 2: Ein Angebots-Nutzungs-Modell des Unterrichts nach Helmke (Helmke 2006; Helmke & Schrader 2006) 16
Abbildung 3: Input-Prozess-Output-Modell nach Oelkers & Reussser (2008, S. 19) 19
Abbildung 4: Modell zur Qualität und Qualitätsentwicklung im Bildungsbereich (Ditton 2000, S. 79) 23
Abbildung 5: Rekontextualisierung im Bildungswesen nach (Fend 2008b) 27
Abbildung 6: School Performance Feedback Systems nach Visscher & Coe (2002) 98
Abbildung 7: „Von der Evaluation zur Innovation“ nach Helmke & Hosenfeld (2005) 101
Abbildung 8: Übersicht über die Fragebereiche in Anlehnung an Helmke & Hosenfeld (2005) 105
Abbildung 9: Allgemeines inhaltsanalytisches Ablaufmodell nach Mayring (2008, S. 54) 130
Abbildung 10: Ablaufmodell induktive Kategorienbildung 132
Abbildung 11: Beispiel für Kategorienbaum zu „Funktion von Vergleichsarbeiten“ 136
Abbildung 12: Ablaufmodell deduktive Kategorienbildung 138
Abbildung 13: Fragebogen zum schulischen Kommunikationsverhalten 142
Abbildung 14: Prozess der Typenbildung nach Kluge (1999) 143
Abbildung 15: VA als Diagnose- versus Kontrollinstrument 212

Tabellenverzeichnis:

Tabelle 2.4.1: Typisierung von Evaluationsformen nach van Ackeren & Klemm (2011, S. 160) 35
Tabelle 3.4.1: Überblick über zentrale nationale Vergleichsstudien zum„Schulmonitoring“ in Anlehnung an van Ackeren (2003, S. 52 80
Tabelle 3.4.2: Typologie der Nutzung von Evaluationswissen nach Stamm (2005, S.20) 94
Tabelle 4.2.1: Gesamtstichprobe der Interviewstudie 2007 113
Tabelle 4.3.1: Transkriptionsregeln nach Langer (2010) 124
Tabelle 4.4.1: Induktive Kategorienbildung am Beispiel "Funktionen der VA" 133
Tabelle 4.4.2: Beispiel für Präsentation der induktiv gewonnenen Ergebniskategorien 137
Tabelle 4.4.3: Beispiel für deduktive Kategorienbildung zu „Verständlichkeit der Rückmeldungen“ 139
Tabelle 4.4.4: Beispiel für Präsentation der deduktiv gewonnenen Ergebniskategorien 140
Tabelle 4.4.5: Übersicht über die Anwendung deduktiver Analyseverfahren bzgl. Fragestellungen 141
Tabelle 4.4.6: Übersicht über die Anwendung induktiver Analyseverfahren 141
Tabelle 5.1.1: Individuelle Bedingungen 148
Tabelle 5.2.1: Rezeption – Hauptkategorie I 152
Tabelle 5.2.2: Rezeption – Hauptkategorie II 158
Tabelle 5.3.1: Reflexion – Hauptkategorie I 163
Tabelle 5.3.2: Reflexion – Hauptkategorie II 171
Tabelle 5.3.3: Reflexion – Hauptkategorie III 177
Tabelle 5.3.4: Reflexion – Hauptkategorie IV 182
Tabelle 5.3.5: Reflexion – Hauptkategorie V 187
Tabelle 5.4.1: Aktion – Hauptkategorie I 196
Tabelle 5.5.1: Schulischer Kontext – Keine Gespräche 202
Tabelle 5.5.2: Schulischer Kontext – Informelle Gespräche 203
Tabelle 5.5.3: Schulischer Kontext – Systematische Diskussionen 204
Tabelle 5.5.4: Schulischer Kontext – Systematische Diskussionen aufgelistet nach Schulart 205
Tabelle 5.6.1: Gesamtübersicht aller Ergebnistabellen 207
Tabelle 5.7.1: Typenbildung zur Rezeption und Nutzung von Vergleichsarbeiten (VA) 214
Tabelle 5.7.2: Zuordnung der einzelnen Interviews zur Typologie 252