

Julia Kahnert

Das Zentralabitur im Fach Mathematik

**Eine empirische Analyse von
Abitur- und TIMSS-Daten im Vergleich**

Empirische Erziehungswissenschaft

herausgegeben von

Rolf Becker, Sigrid Blömeke, Wilfried Bos,
Hartmut Ditton, Cornelia Gräsel, Eckhard Klieme,
Rainer Lehmann, Thomas Rauschenbach,
Hans-Günther Roßbach, Knut Schwippert,
Ludwig Stecher, Christian Tarnai, Rudolf Tippelt,
Rainer Watermann, Horst Weishaupt

Band 47



Waxmann 2014
Münster • New York

Julia Kahnert

Das Zentralabitur im Fach Mathematik

Eine empirische Analyse von
Abitur- und TIMSS-Daten im Vergleich



Waxmann 2014
Münster • New York

Diese Arbeit wurde 2013 von der Technischen Universität Dortmund als Dissertation angenommen.

Bibliografische Informationen der Deutschen Nationalbibliothek

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

Empirische Erziehungswissenschaft, Band 47

ISSN 1862-2127

Print-ISBN 978-3-8309-3047-1

E-Book-ISBN 978-3-8309-8047-6

© Waxmann Verlag GmbH, 2014

Postfach 8603, 48046 Münster

www.waxmann.com

info@waxmann.com

Umschlaggestaltung: Pleßmann Design, Ascheberg

Druck: Systemdruck Köln

Gedruckt auf alterungsbeständigem Papier, säurefrei gemäß ISO 9706

Printed in Germany

Alle Rechte vorbehalten. Nachdruck, auch auszugsweise, verboten.

Kein Teil dieses Werkes darf ohne schriftliche Genehmigung des Verlages in irgendeiner Form reproduziert oder unter Verwendung elektronischer Systeme verarbeitet, vervielfältigt oder verbreitet werden.

Abstract

By passing the A-level examinations at the end of upper secondary high school, students gain the higher education entrance qualification which enables them to take up any subject offered at university. Therefore, it is necessary for the A-level examinations to meet a high degree of standardization (Blossfeld et al., 2011; Neumann, Nagy, Trautwein & Lüdtke, 2009; van Ackeren & Klemm, 2009). Apparently, such a high degree of standardization could not be achieved by a decentralized organisation of the A-level examinations. As a response to the poor results of German students in the school achievement comparison studies conducted between 1995 and 2001 such as *Third International Mathematics and Science Study* (TIMSS) and *Programme for International Student Assessment* (PISA), measures regarding the centralized and standardized testing and assessment of academic achievement were postulated regarding improvement of the quality assurance of the German school system (Fend, 2005; Klieme, 2004; Maag Merki, 2010; Tillmann, Dederich, Kneuper, Kuhlmann & Nessel, 2008). As a result, the centralized organization of the A-level examinations was implemented as a control mechanism over the course of the so-called New Governance structure in many federal states in Germany, as was the case in North Rhine-Westphalia in 2007.

The core of the central A-level examinations is made up of the examination questions which are to measure the competence of students at the end of upper secondary high school. The present state of research for the analysis of A-level examination questions shows the existence of research criteria concerning the issue whether central A-level examinations make a contribution to the quality assurance of the German school system by ensuring appropriate assessment of academic achievement. Using probabilistic scaling methods this study examines whether central A-level examinations hold valid significance for measuring mathematical competence. The pre-university mathematics test taken from the TIMSS/III study was chosen in order to validate the examination questions, since it features high curricular validity, is proven to measure mathematical competence and more than meets the requirements of the central A-level examinations (Klieme, 2000). Furthermore, the mathematics A-level examination questions undergo a detailed quality analysis. Up to now, for instance, comparison of the examination questions cannot be made according to the range of difficulty of the individual subtasks. Hence, all subtasks which are at the teachers' disposal to compile a written examination are analyzed in order to verify the assumption that they all come within the same level of difficulty.

The study at hand gives first information regarding the optimization of the compilation process of central written A-level examinations. The findings show the status quo of these examinations in order to improve of the examination question construction process. The comparison of both test formats is innovative and relevant, since written A-level examinations have not yet been analyzed according to test theory. Moreover, the expectation that written A-level examinations in mathematics are a valid instrument to measure mathematical competence has yet to be proven.

Inhalt

1	Einleitung und Überblick	11
1.1	Die Relevanz des Zentralabiturs im Fach Mathematik im 21. Jahrhundert.....	14
1.2	Forschungsfragen und Ziele dieser Arbeit	15
1.3	Aufbau dieser Arbeit	20
2	Zentralabitur: Theoretische Ausgangslage und empirische Ergebnisse	22
2.1	Theoretische Ausgangslage des Zentralabiturs: Neue Steuerung und Governance	22
2.2	Instrumente der Neuen Steuerung	33
2.3	Das Zentralabitur als Instrument der Neuen Steuerung.....	37
2.3.1	Entstehung und Entwicklung des Zentralabiturs in der BRD.....	40
2.3.2	Die Funktionen des Zentralabiturs	42
2.3.3	Die Organisation des Zentralabiturs in den Ländern der BRD.....	46
2.3.3.1	Die Organisation des Zentralabiturs in NRW.....	56
2.3.3.2	Zusammenfassung und Diskussion: Organisation des Zentralabiturs in der BRD	62
2.4	Internationale und nationale Forschungsbefunde zum Zentralabitur	66
2.4.1	Forschungsergebnisse international.....	66
2.4.2	Forschungsergebnisse national.....	75
2.5	Zusammenfassung des Kapitels	83
3	Zentrale Abiturprüfungen im Fach Mathematik in NRW	85
3.1	Vorgaben und Prüfungsanforderungen für das Fach Mathematik.....	86
3.2	Vorgaben für die Aufgabenkonstruktion für das Fach Mathematik (Leistungskurs).....	90
3.3	Erfassung fachlicher und methodischer mathematischer Kompetenzen durch die Abiturprüfung.....	91

3.4	Bewertung der Abituraufgaben und Einflussfaktoren auf die Bewertung	92
3.5	Messtheoretische Aspekte der Abituraufgaben	96
3.6	Zusammenfassung des Kapitels	102
4	Messung von mathematischen Kompetenzen am Beispiel der TIMSS/III-Studie im Vergleich zur Kompetenzmessung im Zentralabitur Mathematik	104
4.1	Erfassung von mathematischen Kompetenzen in der Oberstufe in empirischen Studien	104
4.2	Die TIMS-Studie in Deutschland – Untersuchungen der Sekundarstufe II	110
4.3	Voruniversitärer Mathematiktest TIMSS	112
4.3.1	Konzeption und Aufbau des Tests zur voruniversitären Mathematik ..	113
4.3.2	Erfassung mathematischer Kompetenzen durch den voruniversitären Mathematiktest	115
4.3.3	Bewertung der Testaufgaben	116
4.3.4	Messtheoretische Aspekte der Testaufgaben	118
4.3.5	Ergebnisse aus der TIMSS/III-Studie – voruniversitärer Mathematiktest	121
4.3.6	Zusammenfassung	123
4.4	Gegenüberstellung: Zentrale Abiturprüfung versus Schulleistungstest im Fach Mathematik	123
5	Ziele und Fragestellungen der eigenen Studie	128
5.1	Forschungslücken und Forschungsdesiderate	128
5.2	Zentrale Fragestellungen und Hypothesen	130
6	Anlage, Durchführung und Auswertungsmethoden der eigenen Studie	135
6.1	Untersuchungsgruppe und -zeitraum	135
6.2	Übersicht über die Instrumentierung	138
6.2.1	Abituraufgaben Mathematik	138
6.2.2	Expertenbefragung	158

6.2.3	Voruniversitärer TIMSS-Test im Fach Mathematik.....	160
6.3	Forschungsmethodische Grundlagen.....	165
6.3.1	Raschskalierung	167
6.3.2	Expertenbefragung	180
6.3.3	Rangkorrelationsanalyse	180
6.3.4	<i>Horizontal equating</i>	182
6.3.5	Umgang mit fehlenden Werten	183
6.3.6	Programme für die Datenanalyse	186
6.4	Zusammenfassung des Kapitels	186
7	Analyse der Daten	188
7.1	Deskriptive Analyse der Mathematikabiturdaten	188
7.2	Ergebnisse der Raschskalierung der Abiturdaten	193
7.2.1	Skalierung der Abiturdaten mit Fokus auf die Itemschwierigkeiten ...	195
7.2.2	Überprüfung der Mehrdimensionalität der Abiturdaten	229
7.2.3	Vergleich der Itemschwierigkeiten mit der Einschätzung von Experten	235
7.3	Analyse der Daten des voruniversitären TIMSS-Tests	245
7.3.1	Deskriptive Analysen der voruniversitären TIMSS Daten.....	246
7.3.2	Skalierung der voruniversitären TIMSS-Daten.....	248
7.3.3	Überprüfung der Mehrdimensionalität der voruniversitären TIMSS-Daten	252
7.4	Verbindung der voruniversitären TIMSS-Ergebnisse Mathematik mit den Abitureergebnissen im Fach Mathematik	253
7.4.1	Analyse der Eindimensionalität der Abitur- und TIMSS-Daten	254
7.4.2	Überprüfung der Mehrdimensionalität der Abitur- und TIMSS-Daten	257
7.4.3	Vergleich der Testergebnisse	260
7.5	Zusammenfassung des Kapitels	263
8	Diskussion und Schlussfolgerungen.....	266
8.1	Diskussion der empirischen Ergebnisse	268
8.2	Schlussfolgerungen	280

9	Literatur.....	287
10	Abbildungs- und Tabellenverzeichnis	308
10.1	Abbildungsverzeichnis	308
10.2	Tabellenverzeichnis.....	311

1 Einleitung und Überblick

Abiturientinnen und Abiturienten erwerben mit dem erfolgreichen Abschluss der gymnasialen Oberstufe die allgemeine Hochschulreife, die sie zum Studium aller Fächer an einer Hochschule befähigt (Trautwein, Köller, Lehmann & Lüdtke, 2007). Daher ist es notwendig, dass bei der Vergabe des Abiturs Standards gelten (Baumert & Watermann, 2000; Blossfeld et al., 2011; Neumann, Nagy, Trautwein & Lüdtke, 2009; Trautwein & Neumann, 2008; van Ackeren & Klemm, 2009).

Ein hohes Maß an Standardisierung konnte durch eine dezentrale Organisationsstruktur des Abiturs allem Anschein nach nicht erreicht werden. Denn als Folge der schlechten Leistungsergebnisse der deutschen Schülerinnen und Schüler, insbesondere in den zwischen 1995 und 2001 stattfindenden Schulleistungsvergleichsstudien, wie *Third International Mathematics and Science Study* (TIMSS), *Programme for International Student Assessment* (PISA) und *Internationale Grundschul-Lese-Untersuchung* (IGLU), wurden Maßnahmen gefordert, die Leistungen zentralisiert und standardisiert einfordern und überprüfen und so einen Beitrag zur Qualitätssicherung des Schulwesens leisten können (Fend, 2005; Klieme, 2004a; Klieme, 2004b; Maag Merki, 2010; Tillmann, Dederich, Kneuper, Kuhlmann & Nessel, 2008). Einen weiteren Grund für die Forderungen von standardisierten und zentralisierten Maßnahmen zur Qualitätssicherung des Schulwesens stellen Forschungsbefunde dar, die auf eine „nur geringe Standardsicherung, auf eine hohe Varianz zwischen einzelnen Bundesländern oder Schulen (Baumert & Watermann 2000; Köller, Watermann, Trautwein & Lüdtke, 2004; Köller, Baumert & Schnabel, 1999), auf die uneinheitliche Verwendung von Benotungsmaßstäben (Klieme, 2003a) und auf signifikante Interaktionseffekte zwischen Notengebung und Schulzugehörigkeit (Klieme, 2005)“ (Maag Merki, 2010, S. 145) verweisen. Weitere Forschungsbefunde weisen auf eine zu hohe soziale Selektivität hin (Böttcher, 2009; Maag Merki, 2010).

Während die Bildungspolitik sich zuvor vor allem auf die Steuerung durch Inputs wie Curricula, effiziente Ressourcen und die Lehrerprofessionalisierung konzentriert hat, rücken im neuen Steuerungsmodell Outputstandards und die empirische Messung dieser in den Vordergrund (Bellmann & Weiß, 2009; Diemer & Kuper, 2010; Helmke, Hornstein & Terhart, 2000; Kurz, 2006; Seiber, Krekkel & van Buer, 2000). Die Qualitätsarbeit an Schulen wird dadurch

erstmalig sichtbar und verhandelbar (Fend, 2011). Aufgrund dessen wird in Deutschland von einem Paradigmenwechsel von der Inputsteuerung zur Outputsteuerung, die auch als Ergebnissteuerung (Altrichter, 2011; Böttcher, 2004; Maag Merki, 2010), evidenzbasierte/evaluationsbasierte Steuerung (Fend, 2011; Heinrich, 2007, 2010), Neue Steuerung oder outputorientierte Inputsteuerung (Berkemeyer, 2010; Böttcher, 2009; Buchen & Rolff, 2009; Ditton, 2000) bezeichnet wird, gesprochen.

Infolgedessen wurde durch die bildungspolitische Reform der Outputsteuerung, neben Steuerungsinstrumenten wie zentrale Lernstandserhebungen, Schulinspektion, zentrale Abschlussprüfungen nach der Klasse 10, auch das Zentralabitur in acht der sechzehn Bundesländer implementiert. Sieben Bundesländer hatten das Zentralabitur bereits vor dem Paradigmenwechsel eingeführt. Lediglich Rheinland-Pfalz bleibt bei der dezentralen Organisationsstruktur des Abiturs. Zentrale Abiturprüfungen wurden in den deutschen Bundesländern eingeführt, „[...]da sie als Garant für die Qualitätssicherung und -entwicklung in der gymnasialen Oberstufe sowie für die Sicherung von Vergleichbarkeit der Anforderungen in den Abiturprüfungen gelten“ (Kühn, 2010, S. 25). Für alle Länder ist zur Sicherung gemeinsamer Standards im Abitur die Orientierung an den *Einheitlichen Prüfungsanforderungen für die Abiturprüfung (EPA)* verbindlich.

¹ Darüber hinaus gibt es länderspezifische Regelungen zur Qualitätssicherung wie Lehrpläne und die fach- und jahrgangsspezifischen Vorgaben zu den unterrichtlichen Voraussetzungen für die schriftlichen Prüfungen im Abitur in der gymnasialen Oberstufe. Jedoch unterliegt das Zentralabitur in Deutschland nicht einer komplett zentral organisierten Prüfungsstruktur. Die Aufgaben werden zentral gestellt, die Bewertung hingegen erfolgt dezentral und halbstandardisiert. Die Erstkorrektur der Abituraufgaben übernimmt dezentral die jeweilige Kurslehrkraft und die Zweitkorrektur in den meisten Fällen eine Lehrperson derselben Schule. Die halbstandardisierte Bewertung der Schülerleistungen findet unter Hinzunahme eines kriterienorientierten Bewertungsbogens statt, der jedoch gewisse Beurteilungsspielräume offen lässt, da nicht konkretisiert wird, für welche Leistung welche Punktzahl vergeben werden darf, wodurch eine mangelnde Auswertungsobjektivität entstehen kann.

Für die Abiturientinnen und Abiturienten ist das Zentralabitur von enormer Wichtigkeit, da es mit erheblichen Konsequenzen verbunden sein kann, wie

¹ Seit dem Jahr 2012 gibt es Bildungsstandards als Grundlage für die Abiturprüfungen für die Fächer Mathematik, Deutsch und Englisch und Französisch. Ab dem Schuljahr 2016/17 lösen sie die EPA in diesen Fächern vollständig ab.

z.B. das Nicht-Erlangen eines Studienplatzes aufgrund einer zu schlechten Abiturdurchschnittsnote. Es erfüllt daher die Funktion eines *high-stakes Tests* für die Schülerinnen und Schüler (Schirp, 2006). Die Qualifikationsfunktion stellt deshalb eine wichtige Funktion des Zentralabiturs dar (Fend, 1980; Kühn, 2010; Wacker, 2008). Zentrale Abschlussprüfungen sollen die Schülerinnen und Schüler für eine selbstständige, berufliche Lebensführung qualifizieren, indem sie ihnen Wissen und Kompetenzen bescheinigen (Kühn, 2010). Kompetenzen können dabei verstanden werden als:

[...] die bei Individuen verfügbaren oder durch sie erlernbaren kognitiven Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten, um die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können (Weinert, 2001b, S. 27f.).

Nach dieser Definition ist Kompetenz eine „[...] Disposition, die Personen dazu befähigt bestimmte Arten von Problemen zu lösen, also konkrete Anforderungssituationen eines bestimmten Typs zu bewältigen“ (ebd. S. 28).

Der Begriff der Kompetenz kommt jedoch erst im aktuellen bildungspolitischen Kontext zu Bildungsstandards für die Allgemeine Hochschulreife offensichtlich zur Geltung. Zwar sind auch in den EPAs kompetenzorientierte Elemente zu finden, aber erst in den Bildungsstandards werden diese für das Zentralabitur fachspezifisch ausformuliert (Kultusministerkonferenz [KMK], 2012b).

Den Kern der zentralen Abiturprüfungen bilden die Abituraufgaben, mit denen die Kompetenzen der Abiturientinnen und Abiturienten am Ende der gymnasialen Oberstufe erfasst werden sollen. Daher muss ihnen eine besondere Bedeutung beigemessen werden.

Der Forschungsstand zur Analyse von Abiturprüfungsaufgaben zeigt, dass Forschungsdesiderate hinsichtlich der Frage bestehen, ob zentrale Abiturprüfungen einen Beitrag zur Qualitätssicherung des Schulwesens leisten können, indem sie eine adäquate Leistungsmessung ermöglichen. Vor allem bleibt bisher unbeantwortet, ob Abituraufgaben ein valides Messinstrument zur Messung von Kompetenzen darstellen. Diese Forschungslücke wird in der vorliegenden Arbeit für das Fach Mathematik aufgegriffen.

1.1 Die Relevanz des Zentralabiturs im Fach Mathematik im 21. Jahrhundert

Die vorliegende Forschungsarbeit befasst sich mit dem Zentralabitur im Fach Mathematik. Das Fach Mathematik ist von besonderem Interesse, weil es bis zum Abitur von allen Schülerinnen und Schülern obligatorisch belegt werden muss. Ein Zitat von Baptist (2001) zeigt auf, warum auf die Mathematik als Pflichtfach in der Schule nicht verzichtet werden kann.

Die Grundfertigkeiten Lesen, Schreiben und Rechnen sind wesentliche Bestandteile der Schlüsselkompetenzen. [...]. Auch Rechenkompetenz, Mathematik – und Computerkenntnisse sowie naturwissenschaftliches Verständnis sind für die uneingeschränkte Teilhabe an der Wissensgesellschaft und für die Wettbewerbsfähigkeit der modernen Volkswirtschaften von größter Wichtigkeit (Kommission der Europäischen Gemeinschaften, 2008, S. 6).

Die mathematische Grundbildung („Mathematical Literacy“) kann als Grundkonstrukt der mathematischen Kompetenz angesehen werden und wird von der OECD wie folgt definiert:

Kein Mensch bestreitet ernsthaft die enorme und ständig wachsende Anwendbarkeit der Mathematik, und insofern gibt es auch keine Legitimationsprobleme für den obligatorischen Mathematikunterricht bis zum Abitur (Baptist & Winter, 2001, S. 67).

Weiter ist das Fach Mathematik wichtig, weil die mathematische Kompetenz zu den Schlüsselkompetenzen des 21. Jahrhunderts zählt.

Mathematische Grundbildung ist die Fähigkeit einer Person, die Rolle zu erkennen und zu verstehen, die die Mathematik in der Welt spielt, fundierte mathematische Urteile abzugeben und sich auf eine Weise mit der Mathematik zu befassen, die den Anforderungen des gegenwärtigen und künftigen Lebens dieser Person als konstruktivem, engagiertem und reflektierendem Bürger entspricht (OECD, 1999, S. 41).

Nicht zuletzt deshalb spielt Mathematik in Schulleistungsstudien wie TIMSS und PISA, in denen die Kompetenzen von Schülerinnen und Schülern erfasst werden, eine zentrale Rolle. Mathematik eignet sich gut für die Erfassung von Schülerleistungen, da das Fach in seiner wissenschaftlichen Konzeption als weitgehend objektiv angesehen wird (Prediger, 2001). Der Erwerb von mathematischen Kompetenzen gilt als eines der Hauptziele des allgemeinbildenden Schulsystems in Deutschland, weil sie eine wichtige Voraussetzung für die

gesellschaftliche Teilhabe und für eine selbstbestimmte, freie und unabhängige Lebensführung darstellen (Neumann & Nagy, 2010).

Als weiterer Grund für die Fachauswahl ist die Einführung der Bildungsstandards für die Mittelstufe (2003) und für die Primarstufe (2004) zu nennen. Klieme (2003) beschreibt Bildungsstandards folgendermaßen:

Bildungsstandards formulieren Anforderungen an das Lehren und Lernen in der Schule. Sie benennen Ziele für die pädagogische Arbeit, ausgedrückt als erwünschte Lernergebnisse der Schülerinnen und Schüler. Damit konkretisieren Standards den Bildungsauftrag, den allgemein bildende Schulen zu erfüllen haben. Bildungsstandards [...] greifen allgemeine Bildungsziele auf. Sie benennen die Kompetenzen, welche die Schule ihren Schülerinnen und Schülern vermitteln muss, damit bestimmte zentrale Bildungsziele erreicht werden. Die Bildungsstandards legen fest, welche Kompetenzen die Kinder oder Jugendlichen bis zu einer bestimmten Jahrgangsstufe erworben haben sollen. Die Kompetenzen werden so konkret beschrieben, dass sie in Aufgabenstellungen umgesetzt und prinzipiell mit Hilfe von Testverfahren erfasst werden können (Klieme, 2003b, S. 19).

Das Fach Mathematik wurde über alle Schulstufen hinweg in den Bildungsstandards aufgegriffen. Auch in den Bildungsstandards, die gerade für die Oberstufe vom Institut für Qualitätssicherung (IQB) für die allgemeine Hochschulreife entwickelt wurden, wurde die Umsetzung für das Fach Mathematik neben den Fächern Deutsch, Englisch und Französisch zuerst durchgeführt. Die Bildungsstandards für das Fach Mathematik wurden 2012 erlassen und sollen ab dem Schuljahr 2016/2017 als verbindliche Grundlage für die Abituraufgaben gelten (Blossfeld et al., 2011). Auch in Bezug auf den Vorschlag eines Kernabiturs und der Umsetzung eines Aufgabenpools für alle Bundesländer auf Grundlage der Bildungsstandards spielt Mathematik eine tragende Rolle (ebd.; Kultusministerkonferenz [KMK], 2012a). Zusammenfassend wird dem Fach Mathematik in der Bildungsforschung eine hohe Bedeutung zugesprochen.

1.2 Forschungsfragen und Ziele dieser Arbeit

Um die skizzierte Forschungslücke, ob Abituraufgaben ein valides Steuerungsinstrument zur Messung mathematischer Kompetenzen darstellen, zu schließen, werden die Abituraufgaben im Fach Mathematik aus dem Jahr 2011 mithilfe

des voruniversitären TIMSS-Tests im Fach Mathematik² einer Qualitätsanalyse unterzogen.³ Es werden in dieser Studie ein halbstandardisiertes Verfahren (Abitur) und ein standardisiertes Verfahren (TIMSS-Test) zur Messung mathematischer Kompetenzen einander gegenüber gestellt. Der TIMSS-Test wurde zur Validierung der Abituraufgaben ausgewählt, da er eine hohe curriculare Validität aufweist und den Anforderungen des Abiturs entspricht (Klieme, 2000).

Die Forschungsarbeit stellt ein Novum dar, da empirische Ergebnisse mit Erkenntnissen zum Vergleich von voruniversitären TIMSS-Daten mit detaillierten Abiturergebnissen im Fach Mathematik, die auch Rückschluss auf die Aufgaben aus den drei Sachgebieten Analysis, Lineare Algebra/Geometrie und Stochastik zulassen, bislang nicht vorliegen. Der Bezug auf den Zusammenhang zwischen Testleistung in Mathematik und Abiturnote wurde bisher in den Studien TIMSS/III und TOSCA (Transformation des Sekundarschulsystems und akademische Karrieren) und in der Studie von Holmeier (2012) untersucht, jedoch in TIMSS/III und TOSCA nur im Hinblick auf den Zusammenhang zwischen Testergebnis und der Gesamtabiturnote im Fach Mathematik (Baumert & Watermann, 2000; Köller et al., 2004).

Das Thema Qualitätsanalyse der zentralen Abituraufgaben stellt ein weiteres Forschungsdesiderat dar. Bisher können beispielsweise keine Aussagen über die Qualität der Abituraufgaben in Bezug auf die Vergleichbarkeit der Schwierigkeitsniveaus der einzelnen Auswahlaufgaben gemacht werden. Aufgrund dessen soll erstmals für das Fach Mathematik untersucht werden, ob alle Auswahlaufgaben, die die Lehrpersonen für die Erstellung einer Klausur zur Verfügung haben, gleich schwierig sind und ob sich drei Schwierigkeitsbereiche (leicht, mittel, schwierig) testtheoretisch nachweisen lassen. Die vorliegende Arbeit bearbeitet auf der Grundlage des Zentralabiturs in NRW im Fach Mathematik 2011 und des fast zeitgleich eingesetzten voruniversitären TIMSS-Test in der gleichen Schülerkohorte die folgenden forschungsleitenden Fragestellungen:

-
- 2 Der voruniversitäre TIMSS-Test stammt aus der TIMS/III-Studie zur Messung mathematischer Bildung am Ende der gymnasialen Oberstufe (Baumert, 2000b).
 - 3 Die vorliegende Studie wurde im Rahmen des Projekts „Qualitätssicherung von zentralen Abiturprüfungen in NRW“ am Institut für Schulentwicklungsforschung in Dortmund durchgeführt.

1. Inwiefern lassen sich die zentralen Abiturprüfungen im Fach Mathematik testtheoretisch beschreiben, unter besonderer Berücksichtigung der Aufgabenschwierigkeiten?
2. Erfüllen zentrale Abituraufgaben im Fach Mathematik ihre Steuerungsfunktion, indem sie eine valide Aussagekraft für die Erfassung mathematischer Kompetenzen von Schülerinnen und Schülern bieten? Messen die zentralen Abiturprüfungen im Fach Mathematik dieselbe mathematische Kompetenz wie der standardisierte voruniversitäre TIMSS-Test?

Diese Studie ist von besonderer Relevanz, weil sie erste Aufschlüsse zur Prozessoptimierung der Erstellung von zentralen Abituraufgaben liefern kann, indem die Ergebnisse aufzeigen, wie der Ist-Zustand bezüglich der Qualität der Abituraufgaben aussieht, um dann eventuelle Maßnahmen zur Verbesserung der Aufgabenerstellung treffen zu können. Weiter wird aufgezeigt, ob die Abituraufgaben im Fach Mathematik eine allgemeine mathematische Kompetenz messen. Dazu wird der voruniversitäre TIMSS-Test herangezogen, bei dem bereits gut erforscht ist, dass er eine allgemeine mathematische Kompetenz misst. Der Vergleich der beiden Testformate ist innovativ und forschungsweisend, da bisher die Abituraufgaben noch nicht testtheoretisch untersucht wurden und bisher noch keine Aussagen dahingehend getroffen werden konnten, ob die Abituraufgaben im Fach Mathematik valide mathematische Kompetenz messen.

Die detaillierten Forschungsfragen, denen in dieser Arbeit nachgegangen wird, und das zugrundeliegende Datenmaterial sind in der folgenden Tabelle überblicksweise zusammengefasst, getrennt nach den beiden Hauptfragestellungen mit den entsprechenden Unterfragen.

Tabelle 1: Überblick über die Forschungsfragen

Forschungsfragen	Datenquellen	Ergebnisse
<p>Inwiefern lassen sich die zentralen Abiturprüfungen im Fach Mathematik testtheoretisch beschreiben, unter besonderer Berücksichtigung der Aufgabenschwierigkeiten?</p> <ol style="list-style-type: none"> 1. Sind die Daten der Abituraufgaben im Fach Mathematik skalierbar und lassen sie sich mit einem Partial-Credit-Modell beschreiben? 2. Sind die Daten der Abituraufgaben im Fach Mathematik eindimensional und erfassen somit eine mathematische Fähigkeitsdimension? 3. Wie stellt sich die Schwierigkeit der Abituraufgaben dar? Lassen sich leichte, mittlere und schwierige Aufgaben für die Schülerinnen und Schüler identifizieren? 4. Lassen sich Abituraufgaben im Fach Mathematik besser durch mehrdimensionale Modelle beschreiben, welche die unterschiedlichen Inhaltsbereiche Analysis und Lineare Algebra/Geometrie berücksichtigen? 5. Wie schätzen Experten die Schwierigkeit der Abituraufgaben ein? Stimmen die Einschätzungen mit den statistisch bestimmten Werten überein? 	<ul style="list-style-type: none"> – halbstandardisierte Bewertungsbögen der Kurslehrkräfte/Erstkorrektur der Abiturklausuren (Dokumentation der erreichten Punktzahlen für Teilaufgaben) in der Stichprobe Leistungskurs Abitur Mathematik 2011 in NRW – Standardisierte Befragung von Expertinnen und Experten aus der Schulpraxis zur Einschätzung der Aufgabenschwierigkeiten 	<p>Abschnitt 7.2.1</p> <p>Abschnitt 7.2.1</p> <p>Abschnitt 7.2.1</p> <p>Abschnitt 7.2.1</p> <p>Abschnitt 7.2.3</p>

Tabelle 1 (Fortsetzung)

Forschungsfragen	Datenquellen	Ergebnisse
<p>Erfüllen zentrale Abituraufgaben im Fach Mathematik ihre Steuerungsfunktion, indem sie eine valide Aussagekraft für die Erfassung mathematischer Kompetenzen von Schülerinnen und Schülern bieten? Messen die zentralen Abiturprüfungen im Fach Mathematik dieselbe mathematische Kompetenz wie der standardisierte voruniversitäre TIMSS-Test?</p> <ol style="list-style-type: none"> 1. Weisen die TIMSS-Daten eine ein- oder mehrdimensionale Struktur auf? Messen sie eine mathematische Fähigkeitsdimension? 2. Inwieweit lassen sich die Abituraufgaben im Fach Mathematik, zusammen mit einem standardisierten Testformat (TIMSS), skalieren und mit dem eindimensionalen Partial-Credit-Modell beschreiben? Messen beide Testformen eine allgemeine mathematische Kompetenz? 3. Inwieweit lassen sich die Abiturprüfungsaufgaben zusammen mit dem standardisierten Testformat in einem zweidimensionalen Modell beschreiben, welches die unterschiedlichen Tests berücksichtigt? Messen beide Testformate unterschiedliche mathematische Kompetenzbereiche? 4. Wie verhalten sich die ermittelten Leistungsergebnisse der getesteten Schülerinnen und Schüler beider Testformen zueinander? Erbringen Abiturientinnen und Abiturienten, die gut in der Mathematikabiturklausur abschneiden, auch gute Leistungen im voruniversitären TIMSS-Test? 	<ul style="list-style-type: none"> – halbstandardisierte Bewertungsbögen der Kurslehrkräfte/Erstkorrektur der Abiturklausuren (Dokumentation der erreichten Punktzahlen für Teilaufgaben) in der Stichprobe Leistungskurs Abitur Mathematik 2011 in NRW – Testergebnisse der einzelnen Aufgaben von allen Schülerinnen und Schülern des voruniversitären TIMSS-Tests für das Fach Mathematik 	<p>Abschnitt 7.3.2/3</p> <p>Abschnitt 7.4.1</p> <p>Abschnitt 7.4.2</p> <p>Abschnitt 7.4.3</p>

1.3 Aufbau dieser Arbeit

Die nachfolgende Arbeit gliedert sich in einen theoretischen und einen empirischen Teil:

In Kapitel 2 werden die theoretische Ausgangslage dieser Arbeit und die empirischen Ergebnisse vorgestellt. Zunächst wird in Abschnitt 2.1 das Konzept der Neuen Steuerung und der Forschungsansatz *Educational Governance* im Hinblick auf das Zentralabitur dargestellt. Daran anschließend werden weitere Formen und Instrumente der Neuen Steuerung aufgezeigt. Daraus ableitend wird das Zentralabitur als ein Instrument der Neuen Steuerung detailliert abgebildet. Es wird die Entwicklung des Zentralabiturs in der BRD sowie dessen Organisation in den verschiedenen Bundesländern der BRD dargestellt. Abschließend wird der internationale und nationale Forschungsstand zum Zentralabitur dargelegt.

Kapitel 3 und 4 zeigen die theoretischen Darstellungen der in dieser Forschungsarbeit eingesetzten Instrumente wie zentrale Abiturprüfungen im Fach Mathematik und standardisierter voruniversitärer TIMSS-Test im Fach Mathematik, auf. Kapitel 3 beinhaltet das Zentralabitur in NRW mit dem Fokus auf das Zentralabitur im Fach Mathematik. Es werden zuerst die Vorgaben und Prüfungsanforderungen und die Vorgaben zur Aufgabenkonstruktion aufgeführt. Weiter wird auf die Erfassung mathematischer Kompetenzen durch Abiturprüfungsaufgaben eingegangen und auf die Bewertung der Abituraufgaben mit möglichen Einflussfaktoren verwiesen, um dann letztlich auf messtheoretische Aspekte der Abituraufgaben hinzuweisen. Kapitel 4 behandelt den Vergleich der Erfassung von Kompetenzen im Zentralabitur mit einem standardisierten Test am Beispiel des Faches Mathematik. Zuerst wird die Erfassung von mathematischen Kompetenzen in empirischen Studien vorgestellt. Daraus folgend geht es um die TIMSS/III-Studie, welche u.a. die mathematischen Kompetenzen von Schülerinnen und Schülern in der Sekundarstufe II untersucht und als Testinstrument den voruniversitären TIMSS-Test im Fach Mathematik eingesetzt hat, der für diese Forschungsarbeit von Relevanz ist. Es wird das Konzept und der Aufbau des Tests dargestellt und dann kongruent zum Zentralabitur im Fach Mathematik die Erfassung mathematischer Kompetenzen durch den voruniversitären Mathematiktest, die Bewertung der Testaufgaben, messtheoretische Aspekte der Testaufgaben und der Forschungsstand in Form von Ergebnissen zum voruniversitären Test aufgezeigt. Abschließend werden die Abiturprüfung im Fach Mathematik und der Schulleistungstest im Fach

Mathematik einander gegenübergestellt, um Unterschiede und Gemeinsamkeiten in ihrer Konstruktion aufzuzeigen.

Der theoretische Teil schließt mit dem 5. Kapitel, in dem die Ziele und Fragestellungen, die aus dem theoretischen Teil resultieren, formuliert werden.

Im 6. Kapitel folgt die Darstellung des empirischen Teils dieser Arbeit. Hier werden die Anlage, die Durchführung und die statistische Auswertung der Studie dargestellt. Weitergehend werden die Untersuchungsgruppe und der Untersuchungszeitraum aufgezeigt sowie eine detaillierte Beschreibung über die eingesetzten Testinstrumente gegeben. Es folgt die Beschreibung der eingesetzten Forschungsmethoden, Raschskalierung, Expertenbefragung, Rangkorrelation und *Horizontal Equating*. Abschließend wird der Umgang mit fehlenden Werten in der Forschungsarbeit dargestellt.

Die Dokumentation der Analyseergebnisse zu den jeweiligen Forschungsfragen (vgl. Tabelle 3) erfolgt in Kapitel 7. Mittels des Verfahrens der Raschskalierung wird der übergeordneten Frage nachgegangen, inwiefern sich die zentralen Abiturprüfungen im Fach Mathematik – unter besonderer Berücksichtigung der Aufgabenschwierigkeiten – testtheoretisch beschreiben lassen. Die testtheoretischen Ergebnisse werden unter Hinzunahme einer Expertenbefragung methodentriangulativ verglichen, um die Ergebnisse der Skalierung zu unterstützen oder zu hinterfragen. Ein weiterer Forschungsstrang bezieht sich auf die erwünschte Steuerungsfunktion von zentralen Abiturprüfungen. Es wird der Frage nachgegangen, ob die zentralen Abiturprüfungen im Fach Mathematik eine valide Aussagekraft für die Erfassung mathematischer Kompetenzen von Schülerinnen und Schülern bieten, in der Form, dass sie dieselbe mathematische Fähigkeit messen wie der standardisierte voruniversitäre TIMSS-Test im Fach Mathematik. Dabei wird wieder das statistische Verfahren der Raschskalierung genutzt. Des Weiteren werden Korrelationsanalysen genutzt, um den Zusammenhang zwischen der Mathematikleistung in der Abiturklausur und der Testleistung im voruniversitären TIMSS-Test Mathematik aufzuzeigen.

Eine abschließende Diskussion der Untersuchungsergebnisse sowie ein Ausblick auf weiterführende Untersuchungen erfolgt im 8. Kapitel dieser Forschungsarbeit.

2 Zentralabitur: Theoretische Ausgangslage und empirische Ergebnisse

In diesem Kapitel wird zuerst auf die theoretische Ausgangslage des Zentralabiturs im Rahmen der Neuen Steuerung und der Governanceperspektive eingegangen. Weiterführend wird das Zentralabitur in den Kontext des Forschungsansatzes der *Educational Governance* eingeordnet, um daraus die Relevanz dieser Forschungsarbeit abzuleiten. Darüber hinaus werden neben dem Zentralabitur andere Instrumente der Neuen Steuerung vorgestellt, um aufzuzeigen, wie vielfältig die Qualitätssicherung des Bildungswesens vorangeführt wird. Im weiteren Verlauf des Kapitels wird das Zentralabitur als Fokus dieser Forschungsarbeit näher in den Blick genommen. Es wird einführend auf die Entstehung und Entwicklung des Zentralabiturs in der BRD eingegangen. Daran anschließend werden die Funktionen des Zentralabiturs, die mit seiner Implementation verbunden sind, erläutert. Da das Abitur nicht in allen 16 Bundesländern gleich organisiert ist, wird auf die Organisation in den einzelnen Ländern verwiesen, mit Schwerpunkt auf Nordrhein-Westfalen. Das Kapitel schließt mit dem internationalen und nationalen Forschungsstand zu zentralen Abiturprüfungen. Es soll aufgezeigt werden, ob die Forschung die Begründung für die Einführung des Zentralabiturs empirisch stützen kann.

2.1 Theoretische Ausgangslage des Zentralabiturs: Neue Steuerung und Governance

Die elementare Funktion von Steuerung besteht nach Luhmann (2004) darin, Differenzen zwischen tatsächlichen und angestrebten Systemzuständen zu verringern (Luhmann, 1994). Übertragen auf die Schule bedeutet das, dass durch die Gestaltung von Lehr- und Lernprozessen die Kompetenzentwicklung der Schülerinnen und Schüler so gesteigert werden soll, dass sich die Differenz zu den Erwartungswerten, die sich primär auf Leistung beziehen, verringern (Diemer & Kuper, 2011). Die Steuerung kann somit als zentrale Dimension der Schulentwicklung betrachtet werden, da die Entwicklung der Einzelschulen nicht gänzlich losgelöst von Standards und deren Einhaltung erfolgen darf (Dedering, 2012). Die Aufsicht über das Schulwesen obliegt in Deutschland per Gesetz dem Staat und seinen Instanzen. Steuerung wird demnach als politische und administrative Rahmensetzung, die sich sowohl auf das gesamte Schulsys-

tem als auch auf die einzelnen Schulen bezieht, verstanden. In diesem Sinne handelt es sich um Fremdsteuerung, welche von der Selbststeuerung der Einzelschule abzugrenzen ist (Berkemeyer, Müller & Kummer, 2007). Die „Macht“ über Verfahren und Inhalte liegt bei einer Außeninstanz, in Bezug auf Schule bei der Bildungsadministration (Böttcher, Holtappels & Brohm, 2006, S. 12).

Der Steuerungsbegriff ist seit Beginn der Institutionalisierung des deutschen Schulwesens im erziehungswissenschaftlichen Diskurs im Gebrauch (Dedering, 2012; Wacker, 2008), jedoch wird ihm erst seit der Veröffentlichung der unzulänglichen deutschen Ergebnisse der internationalen Leistungsvergleichsstudien TIMSS, PISA und IGLU wieder eine hohe Bedeutung zugesprochen (Altrichter & Maag Merki, 2010a).

Infolgedessen wurden zahlreiche „Maßnahmen initiiert, um die Qualität des Bildungssystems und seiner Ergebnisse [...] weiterzuentwickeln und nachhaltig zu sichern“ (Altrichter, Brüsemeister & Wissinger, 2007, S. 9). Es wird gefordert Leistungen stärker einzufordern und zu überprüfen (Tillmann et al., 2008). Als weiterer Impuls für die Einführung neuer Maßnahmen zur Qualitätssicherung des Bildungssystems kann der gestiegene Bedarf nach Standardisierung aufgrund des Wunsches nach einer adäquaten Definition von Studierfähigkeit genannt werden (Blossfeld et al., 2011). Um beispielsweise Unterschiede zwischen Schulformen hinsichtlich fachlicher Anforderungsniveaus und Beurteilungsmaßstäbe auszugleichen, werden standardisierte Verfahren, wie z.B. das Zentralabitur eingeführt (Klieme, 2004a; Klieme, 2004b; Maag Merki, 2010). Der Fokus bei der Neuen Steuerung liegt auf der Qualitätssicherung durch Bildungsergebnisse. Ein System outputorientiert zu steuern, heißt demnach, auf der Grundlage von Informationen über den Output eines Systems sowie auf der Grundlage von Annahmen über den Zusammenhang von Input-, Prozess- und Outputvariablen, Eingriffe in Input- oder Prozessvariablen vorzunehmen (Diemer & Kuper, 2011). Der Begriff Output bezieht sich dabei auf den Erwerb formaler Abschlüsse und auf die Zertifizierung von Leistungen, aber auch auf den Aufbau von Qualifikationen und Kompetenzen (Avenarius et al., 2003; Ditton, 2000). Die Outputsteuerung organisiert die Lern- und Lehrprozesse des Bildungswesens von den gewünschten Ergebnissen und Erträgen her (Kurz, 2006). Das Konzept der Outputsteuerung impliziert nicht, dass auf die Inputsteuerung verzichtet wird, z.B. behalten Lehrpläne, Stundentafeln und gesetzliche Rahmenbedingungen weiterhin ihre Bedeutung (Dedering, 2012). Die Outputs dienen als Zwecke, und die Inputs und Prozesse als Mittel zur Erreichung dieser Zwecke (Luhmann, 2000).

Das neue Steuerungsmodell basiert auf zwei Kernbereichen (Bellmann & Weiß, 2009). Der erste Kernbereich kann als „*standards-based-reform*“ (Bellmann & Weiß, 2009, S. 287) oder Standardsicherung (Dubs, 1996; Schedler, 2007) bezeichnet werden, womit „eine Form von Outputsteuerung durch die Setzung von Bildungsstandards und die externe Evaluation von Schulleistungen gemeint ist“ (Bellmann & Weiß, 2009, S. 287). Bildungsstandards und Instrumente der externen Evaluationen wie Lernstanderhebungen und zentrale Abschlussprüfungen differenzieren sich insofern, als Bildungsstandards ein Inputmerkmal beschreiben, welches Ziele und Kompetenzen definiert, die von den Schülerinnen und Schülern zu einem bestimmten Zeitpunkt erreicht werden sollen (Klieme, 2003b). Lernstanderhebungen und zentrale Abschlussprüfungen hingegen stellen externe und administrative Verfahren dar, mit denen die in den Bildungsstandards definierten Kompetenzen am Ende einer bestimmten Lernzeit als Outputmerkmal überprüft werden sollen (ebd.). Die Steuerungsinstrumente werden „als Mittel der Qualitätssicherung und Qualitätsentwicklung im Schulsystem [eingesetzt]. Erwartet werden von ihnen insbesondere Fortschritte im Hinblick auf Effizienz und Gerechtigkeit (Bellmann & Weiß, 2009, S. 289). Mit diesen Erwartungen ist die Hypothese verbunden, dass sich die Neue Steuerung auf verbesserte Lernergebnisse der Schülerinnen und Schüler auswirkt (Fend, 2011), womit sie in das Feld der Schuleffektivitätsforschung eintritt, die sich mit den Faktoren für die Optimierung von Schulleistungen beschäftigt (vgl. u.a. Creemers, Scheerens & Reynolds, 2000; Fend, 2011; Luyten, 2003; Scheerens, Luyten, Steen & Luyten-de Thouars, 2006; Teddlie & Reynolds, 2000; van der Wolf, 1988).

Der zweite Kernbereich bezeichnet die Wettbewerbssteuerung, „*choice policies*“ (Bellmann & Weiß, 2009, S. 287) oder Qualitätssicherung (Dubs, 1996; Schedler, 2007), womit die „Etablierung von Quasi-Märkten im Bildungssystem durch Dezentralisierung, Schulautonomie und freie Schulwahl“ (Bellmann & Weiß, 2009, S. 287) gemeint ist.

Es gibt eine Wechselbeziehung zwischen der Outputsteuerung und der Wettbewerbssteuerung (ebd.).

Es zeigt sich,

dass sich beide Instrumentbereiche wechselseitig voraussetzen, ergänzen und verstärken. So macht Outputsteuerung Unterschiede sichtbar, die vorher in objektivierter Form nicht sichtbar waren. Damit werden nicht nur die Voraussetzungen für Wettbewerb geschaffen; es wird zugleich die Erwartung erzeugt von der Wahl der ‚besseren Angebote‘ auch tatsächlich Gebrauch machen zu kön-

nen. Umgekehrt gilt: Wettbewerbssteuerung setzt eine gewisse Vergleichbarkeit der Anbieter voraus. Die Etablierung verbindlicher Standards und die Überprüfung tatsächlich erbrachter Leistungen schafft sowas wie eine „gemeinsame Währung“, die die Voraussetzung für Qualitätswettbewerb darstellt (ebd., S. 288).

So können z.B. gute Outputergebnisse einer Schule, z.B. die Anzahl der bestandenen Abiturprüfungen, ein ausschlaggebendes Kriterium für die Schulwahl sein.

Modelle der Neuen Steuerung im Schulsystem

Die „Neue Steuerung steht im Kontext einer Auffassung, nach der Veränderungen und Entwicklungen im Bildungswesen auf der Grundlage ihrer nachgewiesenen Wirksamkeiten, also „evidence-based“ erfolgen sollen“ (Fend, 2011, S. 9). Es ist notwendig, die Wirksamkeit der Neuen Steuerung zu evaluieren, wobei das Erfolgskriterium die Steigerung der Lernergebnisse von Schülerinnen und Schülern ist (ebd.). Der Einfluss der Instrumente der Neuen Steuerung auf die Lernergebnisse kann in Modellen outputorientierter Steuerung festgehalten werden. Eine ganze Reihe von Modellen, die auf Steuerungssysteme im Schulwesen angewendet werden können, stehen bereits zur Verfügung (Bishop, 1999; Bos, Holtappels & Rösner, 2006; Maag Merki, 2010; Scheerens & Bosker, 1997; Stufflebeam, 1967) und ausgewählt werden im Folgenden vorgestellt, wobei immer das Steuerungsinstrument des Zentralabiturs in den Modellen verortet wird.

Das Input-Prozess-Output-Modell und Kontextmodell, welches sich auf Stufflebeam (1967) zurückführen lässt, hat sich als hilfreiche Ordnung der zentralen Faktorengruppen erwiesen (Stufflebeam, 1967). Das Modell ist einfach konstruiert. Der bestmögliche Input und die bestmöglichen Prozesse und Kontexte fördern den Output und die Lernergebnisse der Schülerinnen und Schüler. Das Zentralabitur ist am Output anzusiedeln, da es zum Ziel die Überprüfung der Kompetenzen von Schülerinnen und Schülern am Ende der Schullaufbahn im Rahmen eines Monitoringsystems hat. Es ist jedoch ebenfalls auf der Inputebene anzusiedeln, da die Vorgaben für das Zentralabitur als Input in das System hineingegeben werden. Die Outputsteuerung wird in diesem Modell als Zugabe zur Inputsteuerung verstanden und nicht als Option. Es kann von einer outputorientierten Inputsteuerung gesprochen werden (Berkemeyer, 2010). Die Schwachstelle des Modells von Stufflebeam (1967) liegt in der Ergebnisrückmeldung, die in einem Modell der Steuerung des Schulsystems nicht unberücksichtigt bleiben darf. Ohne die Rückkopplung der Ergebnisse, in Form von

Inputfaktoren, bleiben die Ergebnisse für die Schulentwicklung wirkungslos. Die Rückkopplung der Ergebnisse bleibt z.B. auch in dem Modell von Scheerens und Bosker (1997) unberücksichtigt. Es ist jedoch differenzierter als das Modell von Stufflebeam (1967), da die die Schule als Mehrebenensystem in das Modell einbezogen wird. Es wird zwischen der Systemebene, Schulebene, Klassen- und Unterrichtsebene unterschieden und es werden Aspekte der Schuleffektivitätsforschung (Scheerens & Bosker, 1997) in das Modell integriert. Neben der fehlenden Rückkopplung in dem Modell bleibt auch offen, wie die einzelnen Ebenen zusammenwirken und welche Rolle sie bei der Qualitätssicherung spielen. Wie in dem Modell von Stufflebeam (1967) finden sich Teilaspekte von zentralen Abiturprüfungen auf allen Ebenen wieder.

In Anlehnung an die Dimensionen Input, Prozess und Output haben Bos, Holtappels & Rösner (2006) ein Kreislaufmodell der Schulsteuerung entwickelt, welches eine Erweiterung des Modells von Schwippert und Bos (2005) darstellt (vgl. Abbildung 1).

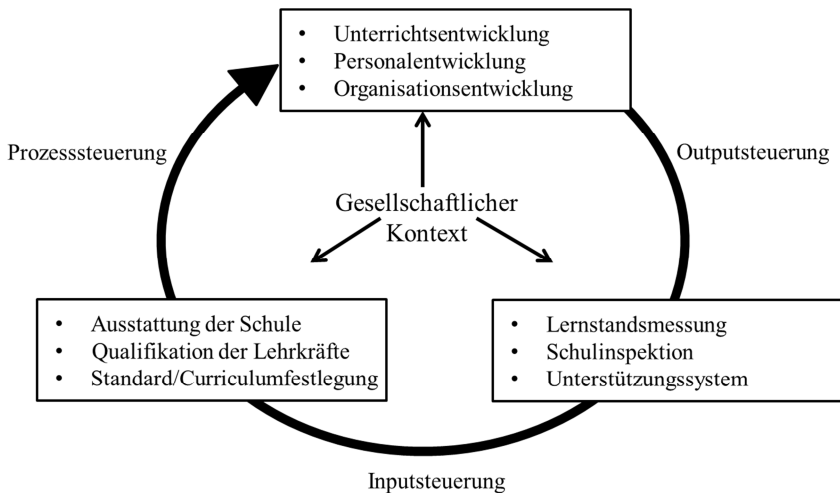


Abbildung 1: Kreislauf der Steuerung im Schulwesen (Bos, Holtappels & Rösner, 2006, S. 83)

Dieses Kreislaufmodell ergänzt, zu den bisher beschriebenen Modellen, die Wechselbeziehung zwischen den drei Steuerungstypen. Weitergehend impliziert es, dass die Ergebnisse aus dem Output auf die Inputebene wirken. Dadurch können Inputs stärker an erkannte Bedarfe ausgerichtet werden (Berkemeyer, 2010). Die Mehrebenenstruktur des Schulsystems wird in diesem

Modell nicht explizit berücksichtigt. Der Vorteil dieses Modells gegenüber den vorherigen ist jedoch die deutliche Betonung des Zusammenhangs der jeweiligen Steuerungsdimensionen. Unklar bleibt, wie die Wirkungsmechanismen zwischen den Ebenen funktionieren, wie z.B. aus den Outputs die Bedarfe abgeleitet werden, die als Input wieder in das System gelangen sollen. Das Modell bietet eine gute Grundlage zur Darstellung der Qualitätssicherung und Qualitätsverbesserung im Rahmen der Neuen Steuerung. Es müsste nur noch um die Makro-, Meso- und Mikroebene und die Wechselbeziehungen untereinander erweitert werden (Lorenz, 2013).

Genau diese Unklarheiten der Wirkungsmechanismen des Modells von Bos, Holtappels und Rösner (2006) greift Ditton (2000) mit seinem Modell auf. Als strukturelle Grundlage der Outputsteuerung kann das Modell von Ditton (2000) zur Qualitätskontrolle und -sicherung in Schule und Unterricht dienen (Tippelt, 2009), welches Input, Transformation/Prozesse und Output innerhalb des Systems Schule unterscheidet (Böttcher, 2009; Ditton, 2008b) (vgl. Abbildung 2).⁴

Die Inputebene umfasst auf der einen Seite alle Ressourcen, die eine Schule besitzt, und die Bedingungen an die eine Schule geknüpft ist (Böttcher, Bos, Döbert & Holtappels, 2008). Ressourcen können finanzieller, materieller, personeller und sozialer Art sein. Die Bedingungen beziehen sich auf die strukturellen Vorgaben des Schulsystems (Ditton, 2008b). Weiter gehören Bildungsziele (Lehrpläne), Bildungsstandards, Leistungen, Einstellungen und Haltungen zu den Inputfaktoren der Organisation Schule (ebd.) Die Inputfaktoren sind für die Schulen nicht beeinflussbar und umfassen die Eingangsbedingungen.

Die zweite Ebene wird als Prozessebene bezeichnet und beinhaltet Faktoren, die sich in Aspekte der Schulqualität und Unterrichtsqualität unterteilen lassen. Diese Ebene macht die pädagogische Ausgestaltung der Einzelschule aus, die Schule besitzt hier die Möglichkeit, die Prozessfaktoren zu beeinflussen (Eickelmann, 2010). Merkmale wie Schulkultur, Schulmanagement und Personalentwicklung betreffen die gesamte Einzelschule und sind auf der Seite der Schulqualität anzusiedeln. Prozesse, die sich konkret auf den Unterricht beziehen, wie Adäquatheit der Lerninhalte und -materialien und Qualität des Lehrens und Lernens werden der Unterrichtsqualität zugeordnet.

4 Das Modell von Ditton erhebt keinen Anspruch auf Vollständigkeit. Es stellt keine Theorie im eigentlichen Wortsinn dar, sondern bietet sich an als Planungsinstrument zur Evaluation schulischer Wirksamkeit (Tippelt, 2009).

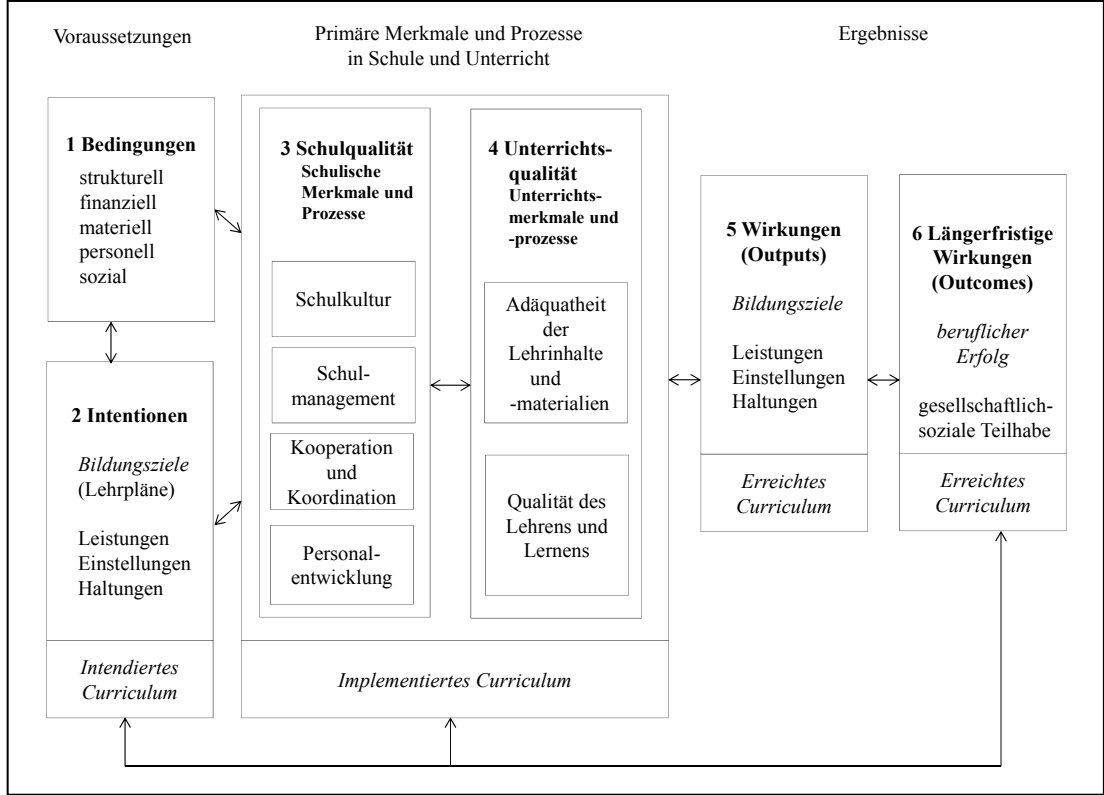


Abbildung 2: Strukturelle und dynamische Dimensionen der Schule (Ditton, 2000, S. 79)

Die erzielten Wirkungen auf der Schul- und Unterrichtsebene sind auf die Vorgabe von Intentionen (Inputebene) wie Bildungsziele, Leistungen, Haltungen und Einstellungen zurückzuführen, aber auch anhand der vorgegebenen Bedingungsstrukturen (finanziell, materiell, strukturell, personell) reflektierbar (Tippelt, 2009).

Die dritte Ebene ist die Outputebene, sie bezieht sich auf die Ergebnisse. Sie umfasst die Leistungsergebnisse der Schülerinnen und Schüler, aber auch ihre Einstellungen (Ditton, 2008b). Sie beinhaltet langfristige Wirkungen, die mit dem Erreichen einer Qualifikation verbunden sind, z.B. beruflicher Erfolg und soziale Teilhabe an der Gesellschaft (Tippelt, 2009). Der Blick auf der Outputebene ist darauf gerichtet, was die Organisation Schule erzeugt und bewirkt, aber nicht als unabhängig von der Input- und Prozessebene zu verstehen. „Outputsteuerung wäre in diesem Sinne eine Steuerung der Inputs und der Prozesse über die konsequenzhaltige Analyse von voran definierten erwünschten Qualifikationsergebnissen“ (Böttcher, 2009, S. 674).

Auf das Zentralabitur bezogen bedeutet das, dass das Zentralabitur mit seinen intendierten Vorgaben, als Input in das System Schule gegeben wird. Auf der Prozessebene müssen dann beispielsweise die Inhalte und geforderten Kompetenzen adäquat im Unterricht den Schülerinnen und Schülern vermittelt werden, damit der vorab definierte Output entstehen kann. Der Output kann sich z.B. darin definieren, möglichst viele Schülerinnen und Schüler zu einem vergleichbaren und qualifizierten Abschluss zu führen, der sie zur Aufnahme eines Studiums befähigt. Es werden Standards in das System Schule hineingegeben, um einen Beitrag zur Qualitätssicherung der Outputs und der Bildungsergebnisse zu leisten.

Die interessante Frage, die sich aus dem Modell ableitet, ist, welchen Einfluss die Bildungsadministration mit der Implementation von Standards und deren Überprüfungen mittels Testungen, auf die Lernergebnisse der Schülerinnen und Schüler haben (Fend, 2011). Die Systemsteuerung müsste nach dem Modell die Steigerung von Lernergebnissen beeinflussen (Fend, 2005). Die empirische Forschungslage zu diesem Forschungsbereich ist uneindeutig (Maag Merki, 2010). Zum Beispiel hat Wößmann (2002) auf Datenbasis der TIMS-Studien versucht zu überprüfen, welche Makrofaktoren einen Einfluss auf die Leistungsergebnisse haben. Als Ergebnis kam zwar heraus, dass die Länder bessere Leistungen erbringen, die eine zentrale Prüfungsorganisation aufweisen, aber das Instrument der zentralen Prüfungen kovarierte auch mit anderen Faktoren, sodass eine kausale Zuordnung schwierig ist (Wößmann, 2003). Auch ein Vergleich von Deutschland mit z.B. den USA kann nicht ohne weiteres vorge-

nommen werden, da *high-stakes Tests* dort mit erheblichen Konsequenzen für die Lehrpersonen und Schülerinnen und Schüler verbunden sind (Bellmann & Weiß, 2009; Kühn, 2010; Maag Merki, 2012e). Dies ist in Deutschland nicht der Fall, dort ist beispielsweise das Zentralabitur nur ein *high-stakes Test* für Schülerinnen und Schüler.

Zusammenfassend ist es schwierig zu modellieren, ob mit den Instrumenten der Neuen Steuerung eine Steigerung der Leistungsergebnisse erzielt werden kann. Es gestaltet sich als schwierig, direkte Wirkungen nachzuweisen; daher werden eher indirekte Wirkungen in die Forschungshypothesen integriert, wie die Erhöhung der Unterrichtsqualität (Fend, 2011). Es muss nach theoretischen Konzepten gesucht werden, die die Wirkungsketten von der Makrosteuerung (Bildungswesen) auf die Mikrosteuerung (Schülerinnen und Schüler) aufzuklären versuchen (ebd.). Hilfreich bei der Aufklärung kann das Konzept der *Educational Governance* sein (Altrichter & Heinrich, 2007; Altrichter & Maag Merki, 2010a; Berkemeyer, 2010; Maag Merki & Altrichter, 2010), wobei zunächst der theoretische Ansatz der *Governance* kurz erläutert wird.

Educational Governance

Mit dem Ansatz der *Governance* werden die „Koordination und Steuerung interdependenter Handlungen gesellschaftlicher Akteure“ bzw. allgemeiner „Regelungsaspekte in komplexen Strukturen“ (Benz, 2004a, S. 17) thematisiert. Wer den Governanceansatz als Forschungsperspektive einnimmt, will verstehen, was Steuern von sozialen Systemen heißen könnte, und will auf die Frage der Systemgestaltung eine differenzierte Antwort geben (Altrichter & Heinrich, 2005; Altrichter & Maag Merki, 2010b). Als Begriffskern des Governancekonzepts fasst Benz (2004a, S. 25) Folgendes zusammen:

1. Governance bedeutet Steuern und Koordinieren (oder auch Regieren) mit dem Ziel des Managements von Interdependenzen zwischen (in der Regel) kollektiven Akteuren.
2. Steuerung und Koordination beruhen auf institutionalisierten Regelsystemen, welche das Handeln der Akteure lenken sollen, wobei in der Regel Kombinationen aus unterschiedlichen Regelsystemen (Markt, Hierarchie, Mehrheitsregel, Verhandlungsregeln) vorliegen.
3. Governance umfasst auch Interaktionsmuster und Modi kollektiven Handelns, welche sich im Rahmen von Institutionen ergeben (Netzwerke, Koalitionen, Vertragsbeziehungen, wechselseitige Anpassung im Wettbewerb).