

Tests, Ergebnisrückmeldungen und dann?

Zur Nutzung von Ergebnisrückmeldungen aus einem
internetgestützten Diagnose- und Förderinstrument
durch Lehrkräfte

Dissertation

zur Erlangung des Doktorgrades

der Wirtschafts- und Sozialwissenschaftlichen Fakultät

der Eberhard-Karls-Universität Tübingen

vorgelegt von

Maike Sigel, M.A.

aus Kirchheim unter Teck

Tübingen

2015

Tag der mündlichen Prüfung: 11.06.2015

Dekan: Professor Dr. rer. soc. Josef Schmid

1. Gutachterin: Professorin Dr. phil. Kathrin Jonkmann

2. Gutachter: Professor Dr. phil. Ulrich Trautwein

Betreuer:

Professor Dr. paed. Stefan Jeuk

Betreuerin:

Dr. phil. Christiane Fiege

Danksagung

Die vorliegende Dissertation wurde im Rahmen des Kooperativen Promotionskollegs „Effektive Lehr-Lernarrangements: Empirische Evaluation und Intervention in der pädagogischen Praxis“ der Universität Tübingen und der Pädagogischen Hochschule Ludwigsburg verfasst und über ein damit verbundenes Promotionsstipendium finanziert. Mein herzlicher Dank gilt meinen Betreuern und Gutachtern Prof. Dr. Kathrin Jonkmann und Prof. Dr. Ulrich Trautwein. Ebenso und ganz besonders herzlich danke ich meiner Betreuerin Dr. Christiane Fiege, deren Unterstützung sehr wertvoll für mich war, sowie meinem Betreuer vonseiten der Pädagogischen Hochschule Ludwigsburg, Prof. Dr. Stefan Jeuk. Der Koordinatorin des Kooperativen Promotionskollegs Dr. Barbara Flunger gilt mein großer und herzlicher Dank für die Organisation des Kollegs und vor allem für die fachliche Begleitung meiner Arbeit, ebenso wie dem ehemaligen Koordinator Prof. Dr. Albrecht Wacker. Weiterhin möchte ich mich bei allen Doktorandinnen und Doktoranden für den Austausch und die Zusammenarbeit im Kooperativen Promotionskolleg bedanken. Außerdem danke ich den Mitarbeiterinnen und Mitarbeitern des Verlagsverbundes Bildungshaus. Ebenso danke ich dem Team des Hector-Instituts für Empirische Bildungsforschung (früher Abteilung Empirische Bildungsforschung und Pädagogische Psychologie) für die gute und freundliche Zusammenarbeit. Das Team am Hector-Institut habe ich als ganz besonders fortschrittlich und zukunftsweisend erlebt. Ich freue mich, alle Kolleginnen und Kollegen kennengelernt zu haben, und bin dankbar dafür, dass ich eine so schöne Zeit in Tübingen am Hector-Institut verbringen durfte. Ganz besonders möchte ich meinen lieben Kolleginnen Alena, Jessica und Hanna für die Unterstützung beim Verfassen dieser Dissertation danken. Den Zusammenhalt der Doktorandinnen und Doktoranden habe ich in meiner Zeit am Hector-Institut als sehr positiv erlebt. Schließlich möchte ich mich auch privat bei meinen Lieben für die Geduld und Unterstützung während der Promotion bedanken.

Zusammenfassung

Aktuelle Entwicklungen im Bildungssystem gaben den Anstoß zu den Untersuchungen in der vorliegenden Dissertation. Die vermehrte Durchführung von Schulleistungsstudien, Vergleichsarbeiten, Kompetenztests etc. wirft Fragen zur Nutzung der daraus gewonnenen objektivierten Informationen auf. Lehrkräfte etwa könnten durch objektivierte Informationen Unterstützung beispielsweise bei der individuellen Diagnostik und Förderung von Schülerinnen und Schülern erhalten. Das in den Schuljahren 2009/2010 bis 2012/2013 in ganz Baden-Württemberg eingesetzte Testverfahren *Online Diagnose* bot den Lehrkräften ein internetgestütztes Diagnose- und Förderinstrument. Das Testverfahren *Online Diagnose* umfasste verschiedene Tests (Haupt-, Detail- und Nachtest) sowie die Rückmeldung von Ergebnissen aus den Tests und die Bereitstellung individueller Fördermaterialien.

In Anlehnung an ein Nutzungsmodell von Ergebnisrückmeldungen, das Zyklusmodell der Rückmeldung (z.B. Helmke 2004; Helmke & Hosenfeld 2005, Hosenfeld & Groß Ophoff 2007), wurde die Nutzung der Ergebnisrückmeldungen und der Instrumente des Testverfahrens *Online Diagnose* durch die Lehrkräfte untersucht. Konkrete Teilziele der Dissertation bestanden darin, festzustellen, inwieweit die Lehrkräfte das Testverfahren *Online Diagnose* nutzten, welche Zusammenhänge sich zwischen bestimmten Aspekten der Nutzung und der Wirkung des Testverfahrens *Online Diagnose* ergaben und welche Einflussfaktoren auf die Aspekte der Nutzung wirkten. Die Bearbeitung dieser Fragestellungen erfolgte auf der Grundlage bisheriger Befunde zur Nutzung von Ergebnisrückmeldungen für die individuelle Förderung von Schülerinnen und Schülern.

Zur Untersuchung der Fragestellungen wurde im Schuljahr 2012/2013 eine Lehrkraftbefragung in ganz Baden-Württemberg durchgeführt. Insgesamt konnten zur Bearbeitung der Fragestellungen die Antworten von 150 Deutschlehrkräften der fünften Jahrgangsstufe aus Haupt-, Werkreal- und Gemeinschaftsschulen herangezogen werden.

Die Ergebnisse zeigten, dass im Rahmen des Testverfahrens *Online Diagnose* hauptsächlich der Haupttest und die Fördermaterialien genutzt wurden. Der Detail- und der Nachtest wurden eher nachrangig verwendet. Bei rund einem Viertel der Lehrkräfte erwies sich die Durchführung als technisch problematisch. Die aus den Tests resultierenden Ergebnisrückmeldungen wandte die Mehrheit der Lehrkräfte gezielt auf einzelne Schülerinnen und Schüler an. Außerdem tauschten sich die Lehrkräfte hauptsächlich im Kollegium über die Ergebnisrückmeldungen aus. Gründe für ein schwaches Testergebnis sahen die Lehrkräfte

hauptsächlich im tatsächlichen Förderbedarf der Schülerinnen und Schüler. Der Zeitaufwand für die Bearbeitung und Besprechung der Ergebnisrückmeldungen war im Vergleich zum Zeitaufwand für die Durchführung der Tests und Nutzung der Ergebnisrückmeldungen höher. Den Einfluss motivationaler Aspekte bei der Bearbeitung bzw. Anwendung des Haupttests und der Fördermaterialien schätzten die Lehrkräfte für sich selbst tendenziell höher ein als für die Schülerinnen und Schüler. Außerdem schätzten die Lehrkräfte die Qualität z.B. der Aufgaben des Testverfahrens *Online Diagnose* als hoch ein. Ein Vergleich mit objektiven Qualitätskriterien war allerdings nicht möglich.

Weiterhin konnte in einer zweiten Untersuchung gezeigt werden, dass die Lehrkräfte hinsichtlich der Wirkung auf Lernende eher Qualitätsaspekte des Haupttests und der Fördermaterialien, eine gezielte Nutzung von Ergebnisrückmeldungen und die Fördermaterialien selbst als relevant ansahen. Das bedeutet, dass die Wirkung auf Lernende vor allem mit den praktischen Maßnahmen und Instrumenten des Testverfahrens *Online Diagnose* in Zusammenhang stand. Hinsichtlich der Wirkung auf sich selbst sahen die Lehrenden hingegen eher den Bezug zum Bildungsplan der Grundschule, den Informationsgewinn aus den Ergebnisrückmeldungen und die Fördermaterialien als relevant an. Das heißt, die Wirkung auf Lehrende hing eher mit den unterstützenden Elementen des Testverfahrens *Online Diagnose* zusammen.

Im Rahmen einer dritten Untersuchung wurde deutlich, dass vor allem individuelle, schulische und externe Bedingungen, wie z.B. die Akzeptanz von externen Testverfahren durch die Lehrkraft und das schulische Umfeld, die Aspekte des Nutzungsprozesses beeinflussten. Auffallend war, dass vorrangig signifikante Zusammenhänge zwischen den Kriterien der individuellen Bedingungen und den Prozessschritten Rezeption und Reflexion, also dem selbstwirksamen Umgang mit den Ergebnisrückmeldungen bzw. dem Informationsgewinn aus den Ergebnisrückmeldungen auftraten. Außerdem zeigten sich eher signifikante Zusammenhänge im Verhältnis zwischen den schulischen und externen Bedingungen und dem Prozessschritt Evaluation, also der Wirkung des Testverfahrens *Online Diagnose* auf Lernende, das Kollegium und Lehrende.

Die Komplexität der Nutzung von Ergebnisrückmeldungen konnte durch die vorliegende Dissertation erneut bestätigt werden. Außerdem wurde die wichtige Rolle von Merkmalen und Einstellungen der Lehrkraft in diesem Prozess der Nutzung offensichtlich. Als Fazit kann aus den Ergebnissen die Erkenntnis gezogen werden, dass der Beschluss der Kultusministerkonferenz der Länder, die Lehrkräfte hinsichtlich ihrer Diagnosefähigkeit, in Bezug auf den Umgang mit Heterogenität und in der individuellen Förderung von Schülerinnen

und Schülern durch Aus- und Fortbildung zu unterstützen, von besonderer Relevanz ist. Die Lehrkräfte können in den genannten Punkten durch die Ergebnisrückmeldungen aus externen Evaluationen unterstützt werden. Hierfür ist die Förderung der Akzeptanz von externen Evaluationen sowie die Unterstützung der Lehrkräfte bei der Nutzung von (elektronischen) Ergebnisrückmeldungen z.B. durch Fort- und Weiterbildungsprogramme notwendig.

Inhaltsverzeichnis

1 Einleitung	3
2 Theoretischer Hintergrund.....	7
2.1 Bildungsmonitoring und Bildungsstandards.....	7
2.2 Testverfahren in Baden-Württemberg	10
2.3 Das Zyklenmodell der Rückmeldung	14
2.4 Die Nutzung von Ergebnisrückmeldungen.....	24
2.5 Zusammenfassung der Kapitel zum theoretischen Hintergrund.....	35
3 Das Testverfahren <i>Online Diagnose</i>	39
3.1 Ziele und Ablauf des Testverfahrens <i>Online Diagnose</i>	39
3.2 Technische Voraussetzungen zur Durchführung des Testverfahrens <i>Online Diagnose</i>	40
3.3 Aufgabenbereiche des Testverfahrens <i>Online Diagnose</i> im Vergleich mit den Bildungsstandards	42
3.4 Ergebnisrückmeldungen zum Testverfahren <i>Online Diagnose</i>	47
3.5 Das Testverfahren <i>Online Diagnose</i> in Bezug auf das Zyklenmodell der Rückmeldung	51
3.6 Zusammenfassung der Kapitel zum Testverfahren <i>Online Diagnose</i>	52
3.7 Fragestellungen.....	53
4 Datengrundlage.....	55
4.1 Beschreibung der Studie	55
4.2 Beschreibung der Stichprobe	58
5 Teilstudie 1: Deskription der Nutzung des Testverfahrens <i>Online Diagnose</i>	61
5.1 Statistische Analysen	61
5.2 Ergebnisse	62
5.2.1 Durchführung und Nutzung des Testverfahrens <i>Online Diagnose</i>	63
5.2.2 Motivation bei der Durchführung des Haupttests und dem Einsatz der Fördermaterialien	73
5.2.3 Subjektive Qualitätseinschätzung zum Testverfahren <i>Online Diagnose</i>	90
5.2.4 Prozessschritte der Nutzung von Ergebnisrückmeldungen.....	99
5.3 Diskussion und Implikationen	109
6 Teilstudie 2: Prozessschritte der Nutzung von Ergebnisrückmeldungen	123

6.1 Methode und Instrumente	124
6.2 Statistische Analysen	126
6.3 Ergebnisse.....	127
6.4 Diskussion und Implikationen	139
7 Teilstudie 3: Einflussfaktoren auf die Prozessschritte der Nutzung von Ergebnisrückmeldungen.....	149
7.1 Methode und Instrumente	150
7.2 Statistische Analysen	161
7.3 Ergebnisse.....	162
7.4 Diskussion und Implikationen	182
8 Gesamtdiskussion.....	195
8.1 Zentrale Befunde zu den Teilstudien 1 bis 3	195
8.2 Kritische Einordnung der Ergebnisse	199
8.3 Kritische Einordnung des Studiendesigns	203
8.4 Implikationen für die weitere Forschung und die Praxis.....	206
8.5 Fazit	209
9 Literatur	211
10 Verzeichnisse	233
10.1 Abbildungsverzeichnis	233
10.2 Tabellenverzeichnis	234

1 Einleitung

„Empirische Bildungsforschung hat drei zentrale Aufgaben: Sie hilft erstens bei der Diagnose von Kompetenzen der Lernenden bzw. der Funktionsfähigkeit des Systems. Neben der Diagnostik verfolgt sie zweitens das Ziel, die Ursachen für gelingende oder misslingende Lernprozesse und Bildungskarrieren zu identifizieren, sowohl auf der Ebene der Individuen und ihres unmittelbaren sozialen Umfeldes als auch auf der Ebene von Lerngruppen, Bildungseinrichtungen und ihren übergreifenden Strukturen. Eine dritte Aufgabe der empirischen Bildungsforschung ist die Begleitung und Evaluation von Veränderungen im Bildungssystem und die Entwicklung und Prüfung von Interventionsmaßnahmen.“

(Scheiter, Schrader, Trautwein & Hesse, 2014, S. 189)

Die unter anderem durch die PISA-Studie angestoßenen strukturellen Änderungen im Bildungssystem führen von der Input- zur evaluationsbasierten Output-Steuerung (Wacker, Maier & Wissinger, 2012). Lag der Schwerpunkt zuvor auf den Unterrichtsinhalten, so wird der Fokus nun im Sinne einer prozessgeleiteten Unterrichtsgestaltung auf die operativen Einheiten des Bildungssystems gelenkt. Vergleichende Schulleistungsstudien liefern die notwendigen Informationen zur Diagnostik leistungsbezogener Stärken und Schwächen von Schülerinnen und Schülern für eine prozessgeleitete Unterrichtsgestaltung (Kühle & van Ackeren, 2012). Die durch die externe Evaluation (z.B. Vergleichsarbeiten) überprüften kompetenzorientierten Standards (z.B. Bildungsstandards) sollen wiederum im Unterricht verwirklicht und in Lehrerhandeln umgesetzt werden (Kühle & van Ackeren, 2012). In diesem Rahmen muss der Diagnosefähigkeit von Lehrkräften eine stärkere Aufmerksamkeit zukommen. Das Erkennen von Förderbedarfen und das Umsetzen von Förderangeboten sollte Unterrichtsrealität sein. Forschungsbefunde zeigen allerdings deutliche Unterschiede in der Diagnosegenauigkeit bei den Lehrkräften (z.B. Karing, 2009; Zimmer-Müller, Hosenfeld & Koch, 2014). Werden die Ergebnisrückmeldungen aus externen Testergebnissen dennoch von den Lehrkräften genutzt? Empirische Studien weisen nach, dass das objektivierbar gewordene Wissen aus externen Evaluationen von den pädagogischen Akteuren der Schule überwiegend zur Kenntnis genommen wird und dass diese auch zur Reflexion bereit sind (z.B. Kühle & van Ackeren, 2012; Maier & Rauin, 2006; Schneewind & Kuper, 2009). Ob und wie diese Informationen tatsächlich Eingang in den Schulalltag finden, bedarf weiterer Untersuchungen. Aus Rezeptionsstudien, die die Nutzung zurückgemeldeter Individualdaten untersuchten, wird deutlich, dass die Ergebnisse zwar rezipiert und reflektiert wurden, dass dies allerdings selten Konsequenzen für den Unterricht hatte (z.B. Maier, 2009;

Schneewind & Kuper, 2009). Die Umsetzung wissenschaftlichen Wissens in Handlungswissen ist für die Nutzung von Ergebnisrückmeldungen aus externen Evaluationen notwendig (Rolff, 2002). Lehrkräfte nehmen Ergebnisrückmeldungen besonders dann zur Kenntnis und reflektieren sie, wenn deutlich wird, wo die Schülerinnen und Schüler stehen und in welcher Hinsicht sie Kompetenzen entwickeln müssen (Kühle & van Ackeren, 2012). Maier (2008b) zeigte in einer Gegenüberstellung von Vergleichsarbeiten aus Baden-Württemberg mit Kompetenztests aus Thüringen, dass eine Einbindung des Verfahrens in ein übergreifendes Fortbildungs- und Weiterentwicklungskonzept zu einer höheren Akzeptanz- und Nutzungseinschätzung führte (vgl. auch Zimmer-Müller, Hosenfeld & Koch, 2014). Die Übertragung von Ergebnisrückmeldungen in Handlungswissen bzw. die Unterstützung von Lehrkräften bei der Umsetzung von theoretischem Wissen in Handlungswissen ist für die Wirkung von Ergebnisrückmeldungen von Bedeutung. Das über einen Zeitraum von vier Jahren eingesetzte Testverfahren *Online Diagnose* soll Lehrkräfte bei der Diagnostik von Schülerleistungen unterstützen. Es bietet den Lehrkräften zwar keine direkte Übertragung von Ergebnisrückmeldungen in Handlungswissen an, unterbreitet ihnen jedoch konkrete Vorschläge zur Förderung in Form von Fördermaterialien. Der Ablauf des Testverfahrens *Online Diagnose* kann an das theoretische Zyklenmodell der Rückmeldung angelehnt werden. Dieses sieht die Nutzung von Ergebnisrückmeldungen über die Rezeption, Reflexion und Aktion bis hin zur Evaluation vor (z. B. Helmke & Hosenfeld, 2005; Hosenfeld & Groß Ophoff, 2007). Das Zyklenmodell berücksichtigt den Prozessweg von der Interpretation von Ergebnisrückmeldungen über die Suche nach Erklärungen bis zum Einsatz von konkreten Maßnahmen und der anschließenden Evaluation des Prozesses, es schließt damit Handlungen bzw. Handlungswissen mit ein.

Die vorliegende Dissertation hatte die Untersuchung der Nutzung von Ergebnisrückmeldungen aus einem internetgestützten Diagnose- und Förderinstrument zum Ziel. Der Aufbau der Arbeit gestaltet sich wie folgt: Zu Beginn wird die Entstehung und aktuelle Relevanz des Themas in Bezug auf die Umstrukturierung des Bildungssystems unter den Stichworten „Bildungsmonitoring“ und „Bildungsstandards“ dargestellt (Kapitel 2.1). Die durch das Bildungsmonitoring angestoßenen Veränderungen werden anhand beispielhafter Verfahren aus Baden-Württemberg berichtet (Kapitel 2.2) und schließlich in einen theoretischen Rahmen zur Nutzung von Ergebnisrückmeldungen eingefasst (2.3). Im dritten Abschnitt wird ausführlich das der Dissertation zugrunde liegende Testverfahren *Online Diagnose* vorgestellt (Kapitel 3). Das Testverfahren *Online Diagnose* deckt einige Anforderungen aus der

Theorie und den bisherigen empirischen Ergebnissen ab (Kapitel 3.6) und kann an das theoretische Zyklenmodell der Rückmeldung angelehnt werden (Kapitel 3.5). Daraus werden schließlich die drei für die Dissertation relevanten Fragestellungen abgeleitet (Kapitel 3.7). Im anschließenden Abschnitt wird die Datengrundlage der Dissertation erläutert (Kapitel 4). Die Kapitel 5 bis 7 bilden den Kern der Arbeit, indem sie die Problemstellung in drei aufeinander folgenden Teilstudien unter verschiedenen Gesichtspunkten behandeln. Die Ergebnisse dieser drei Teilstudien werden aufbereitet, strukturiert und anschließend diskutiert. Zunächst wird die Nutzung des Testverfahrens *Online Diagnose* beschrieben (Teilstudie 1; Kapitel 5). Die Schritte der Nutzung von Ergebnisrückmeldungen werden in Teilstudie 2 referiert und in Zusammenhang mit der Wirkung auf Lernende, der Wirkung auf das Kollegium und der Wirkung auf Lehrende untersucht (Kapitel 6). In Teilstudie 3 werden Einflussfaktoren auf die Schritte der Nutzung von Ergebnisrückmeldungen berücksichtigt. Die durch das Zyklenmodell der Rückmeldung strukturierten individuellen Bedingungen, Bedingungen der Klasse sowie schulische und externe Bedingungen werden in Zusammenhang mit den Schritten der Nutzung von Ergebnisrückmeldungen gebracht (Kapitel 7). Die Dissertation schließt mit der Gesamtdiskussion und den Implikationen aus den vorliegenden Ergebnissen der Teilstudien 1 bis 3.

2 Theoretischer Hintergrund

2.1 Bildungsmonitoring und Bildungsstandards

Gesamtstrategie Bildungsmonitoring zur Weiterentwicklung des Bildungssystems in Deutschland. Im Jahr 2006 wurde durch die Kultusministerkonferenz die Gesamtstrategie zum Bildungsmonitoring beschlossen (KMK & IQB, 2006). Neben dem Konstanzer Beschluss im Jahr 1997, in dem bereits die regelmäßige Durchführung länderübergreifender Vergleichsuntersuchungen auf den Weg gebracht wurde (KMK, 1997), ist die Gesamtstrategie zum Bildungsmonitoring ein Meilenstein in der Weiterentwicklung des Bildungssystems in Deutschland. Der Gesamtstrategie liegen sowohl politische Ziele, wie die gesellschaftliche Teilhabe, die individuelle Entwicklung und das berufliche Fortkommen der Bevölkerung, als auch der wirtschaftliche Erfolg des Landes zugrunde (KMK & IQB, 2006). Als Auslöser der Entwicklungen spielte das schlechte Abschneiden deutscher Schulen im internationalen Vergleich der PISA-Ergebnisse eine tragende Rolle. Durch die Orientierung an PISA verlagerte sich die Zielsetzung von einer bisher starken Akzentuierung der passenden Struktur und inhaltlichen Vorgaben (Input-Steuerung) auf die Prozesse und Ergebnisse des Lernens (KMK & IQB, 2006). Die Qualitätssicherung und Standardisierung auf allen Ebenen des Bildungssystems stellt für die Erreichung der neuen Ziele eine Notwendigkeit dar. In diesem Rahmen wurden vier konzeptionell verbundene Verfahren festgelegt: Internationale Schulleistungsstudien, Ländervergleiche zur Überprüfung der Erreichung der Bildungsstandards, Vergleichsarbeiten und schließlich die gemeinsame Bildungsberichterstattung von Bund und Ländern (KMK & IQB, 2006). Die Verfahren sollen Handlungsbedarfe in Bezug auf die frühzeitige Förderung z. B. von Schülerinnen und Schülern mit Migrationshintergrund oder sozialer Benachteiligung aufdecken. Sie sollen ebenso die kompetenz- und standardbasierte Unterrichtsentwicklung durch die Lehrkräfte erleichtern. Des Weiteren beinhalten die Beschlüsse die Aus- und Fortbildung von Lehrkräften hinsichtlich ihrer Diagnosefähigkeit, des Umgangs mit Heterogenität und der individuellen Förderung von Schülerinnen und Schülern. Die Erarbeitung von Konzepten, Materialien und Aufgabenpools, zunächst in den Unterrichtsfächern Deutsch und Mathematik, ist ebenfalls ein Teil der Gesamtstrategie (KMK & IQB, 2006).

Die Entwicklung von Bildungsstandards. Die Bildungsstandards für den in dieser Arbeit relevanten Primarbereich wurden im Jahr 2004 beschlossen (KMK, 2004a). Die Implementierung in den Fächern Deutsch und Mathematik begann im Schuljahr 2005/2006 (KMK, 2004a). Die Bildungsstandards haben generell eine Überprüfungs- und Entwicklungsfunktion. Die Stärken und Schwächen von Schülerinnen und Schülern in den zentralen Kompetenzbereichen sollen aufgedeckt und eine Qualitätssicherung ermöglicht werden. Dabei konzentrieren sich die Bildungsstandards auf erwünschte Lernergebnisse des jeweiligen Faches (KMK, 2004b). Die Bildungsstandards wurden als Regelstandards entwickelt. Diese beziehen sich auf eine mittlere Leistungsfähigkeit von Schülerinnen und Schülern. Damit soll eine Über- oder Unterforderung vermieden werden, die bei Mindest- oder Maximalstandards schneller auftreten kann (KMK, 2004b; KMK & IQB 2010). Im Unterschied zu den Lehrplänen, die ein strukturierendes Element darstellen, zeigen die Bildungsstandards eine Zielperspektive auf (KMK, 2004b).

Klieme (2004) bezieht die Bildungsstandards in Abhängigkeit vom jeweiligen Einsatzbereich auf folgende Funktionen: Die Bildungsstandards können in empirischem Sinne als gemessene Lernergebnisse bezeichnet werden. In Vergleichsuntersuchungen dienen diese Lernergebnisse z.B. als messbare Standards, die verglichen werden können. In diagnostischem Sinne können die Bildungsstandards als Benotungsmaßstab eingesetzt werden. In der individuellen Förderung von Schülerinnen und Schülern können die Bildungsstandards als Maßstab und Entscheidungshilfe in der Frage dienen, ob eine Förderung notwendig ist und worin die Förderung bestehen soll. Außerdem sind die Bildungsstandards in normativem Sinne als Erwartungen an die Leistung bzw. Lernergebnisse der Schülerinnen und Schüler sowie an die Unterrichtsinhalte (z.B. Lehrpläne) und an den Unterricht (z.B. Unterrichtsentwicklung) zu bezeichnen. Hierbei können die Bildungsstandards z.B. bei einer Differenz der erwarteten und tatsächlichen Leistung von Schülerinnen und Schülern, bei Abweichungen der vorgesehenen von den tatsächlich umgesetzten Unterrichtsinhalten sowie bei Abweichungen der erwarteten von der tatsächlichen Unterrichtsentwicklung zu strukturellen bildungspolitischen Entwicklungen beitragen.

Die durch die Bildungsstandards gewonnene Vergleichbarkeit wird ebenso durch das Institut zur Qualitätsentwicklung im Bildungswesen (IQB) zur Normierung von Testaufgaben sowie zur Überprüfung und Weiterentwicklung der Bildungsstandards genutzt (KMK, 2004b; IQB, n.d.).

Da in der vorliegenden Arbeit Deutschlehrkräfte befragt wurden, werden im Rahmen dieser Betrachtung die Bildungsstandards des Unterrichtsfaches Deutsch herangezogen (Abbildung 1). Die Sprache hat eine erhebliche Bedeutung für die kognitive, soziale und emotionale Entwicklung von Schülerinnen und Schülern. Deshalb ist eine grundlegende sprachliche Bildung für den Deutschunterricht in der Primarstufe vorgesehen. Die Bildungsstandards sind in die folgenden Bereiche gegliedert: Sprechen und Zuhören, Schreiben, Lesen – mit Texten und Medien umgehen sowie Sprache und Sprachgebrauch untersuchen (KMK, 2004a)¹.

<p>Sprechen und Zuhören</p> <ul style="list-style-type: none"> • zu anderen sprechen • verstehend zuhören • Gespräche führen • szenisch spielen • über Lernen sprechen 	<p>Schreiben</p> <ul style="list-style-type: none"> • über Schreibfertigkeiten verfügen • richtig schreiben • Texte planen • Texte schreiben • Texte überarbeiten 	<p>Lesen – mit Texten und Medien umgehen</p> <ul style="list-style-type: none"> • über Lesefähigkeiten verfügen • über Leseerfahrungen verfügen • Texte erschließen • Texte präsentieren
<p>Methoden und Arbeitstechniken</p> <p>Methoden und Arbeitstechniken werden jeweils in Zusammenhang mit den Inhalten jedes einzelnen Kompetenzbereichs erworben.</p>		
<p>Sprache und Sprachgebrauch untersuchen</p> <ul style="list-style-type: none"> • grundlegende sprachliche Strukturen und Begriffe kennen • sprachliche Verständigung untersuchen • an Wörtern, Sätzen, Texten arbeiten • Gemeinsamkeiten und Unterschiede von Sprachen entdecken 		

Abbildung 1. Bildungsstandards der Grundschule im Unterrichtsfach Deutsch (KMK, 2004a, S. 7)

Unterstützung von Lehrkräften in der Diagnostik. Die Diagnosefähigkeit von Lehrkräften ist von zentraler Bedeutung (Schrader, 1989) und dies in zweifacher Hinsicht: Zum einen für die möglichst objektive Vergabe von Noten und zum anderen für die Förderung des Lernerfolgs der Schülerinnen und Schüler (Brunner, Anders, Hachfeld & Kraus, 2011). Forschungsbefunde zeigten, dass die Einschätzung von z.B. Leistungsniveau, Leistungsheterogenität und Leistungsbereitschaft einer Klasse durch Lehrkräfte nur in geringem Maße adäquat ist (Brunner et al. 2011; vgl. auch McElvany, Schröder, Hachfeld, Baumert,

¹ In Abschnitt 3.3 werden die Bildungsstandards weiter ausgeführt.

Richter, Schnotz, Horz & Ullrich, 2009; Spinath, 2005). In weiteren Befunden wurde festgestellt, dass die Einschätzung der Leistungen von Schülerinnen und Schülern durch Lehrkräfte im Mittel zwar gut bzw. realistisch ist, die individuellen Unterschiede der Lehrerurteile allerdings groß sind (Karing, 2009; Kunter & Pohlmann, 2015). Die Qualität der Einschätzungen der Lehrkräfte ist z.B. abhängig von der Erfahrung der Lehrkräfte und von Merkmalen der Situation (z.B. Krolak-Schwerdt, Böhmer & Gräsel, 2009; Kunter & Pohlmann, 2015), ebenso konnte gezeigt werden, dass die Bandbreite der Variationen die Adäquatheit der Lehrkräfteeinschätzungen nicht systematisch erklären kann (Dicke, Lüdtker, Trautwein, Nagy & Nagy, 2012). Bis die im Rahmen der Gesamtstrategie Bildungsmonitoring geforderte Aus- und Fortbildung der Lehrkräfte in Bezug auf die Diagnosefähigkeit, den Umgang mit Heterogenität und die individuelle Förderung greift, ist die Unterstützung von Lehrkräften durch Ergebnisrückmeldungen von Individualdaten aus extern zur Verfügung gestellten diagnostischen Verfahren von besonderer Bedeutung. Ergebnisrückmeldungen der landesweiten Vergleichsarbeiten können hierzu ihren Beitrag leisten (Brunner et al. 2011). Ebenso kann das in Baden-Württemberg landesweit eingesetzte Testverfahren *Online Diagnose* die Lehrkräfte durch individuelle Rückmeldungen bei der Diagnostik von Schülerleistungen sowie beim Umgang mit Heterogenität in einer Klasse und durch die Fördermaterialien bei der individuellen bzw. gezielten Förderung unterstützen.

2.2 Testverfahren in Baden-Württemberg

Testverfahren in Baden-Württemberg im Rahmen der Gesamtstrategie Bildungsmonitoring. Beginnend mit der TIMSS²-Studie im Jahr 1995 wurde auch in Baden-Württemberg die Zeit der internationalen Vergleichsstudien eingeleitet (Kultusportal, Schulleistungsstudien). Im Jahr 1997 wurde im Rahmen der Konstanzer Beschlüsse der Kultusministerkonferenz der Länder die regelmäßige Teilnahme an internationalen Schulleistungsstudien beschlossen (KMK, 1997). Neben TIMSS nimmt das Land Baden-Württemberg an den folgenden Schulleistungsstudien teil: DESI³, PIRLS/IGLU⁴, PISA⁵ TOSCA⁶ und

² Trends in International Mathematics and Science Study (TIMSS)

³ Deutsch Englisch Schülerleistungen International (DESI)

⁴ Progress in International Reading Literacy Study (PIRLS) und Internationale Grundschul-Lese-Untersuchung (IGLU)

⁵ Programme for International Student Assessment (PISA)

⁶ Transformation des Sekundarschulsystems und akademische Karrieren (TOSCA)

TRAIN⁷ (Kultusportal, Schulleistungsstudien). Im Rahmen der Gesamtstrategie zum Bildungsmonitoring wurden die Ländervergleiche⁸ in der Sekundarstufe I im Schuljahr 2008/2009 und in der Primarstufe im Schuljahr 2010/2011 eingeführt. In ihrem Aufbau sind die Ländervergleichsstudien an internationale Studien angelehnt. Ziel der Ländervergleiche ist die Überprüfung der in Deutschland für alle Bundesländer verbindlichen nationalen Bildungsstandards (IQB, Ländervergleich; KMK, Ländervergleiche). Zur Selbstevaluation der allgemeinbildenden Schulen wurden im Schuljahr 2005/2006 die Instrumente VERA 3⁹ für die Primarstufe sowie im Schuljahr 2010/2011 die in Baden-Württemberg landesspezifischen DVA¹⁰ für die Sekundarstufe flächendeckend eingesetzt (Kultusportal, Vergleichsarbeiten; Landesbildungsserver, Vergleichsarbeiten DVA). Ab dem Schuljahr 2015/2016 wird sich das Land Baden-Württemberg auch an VERA 8¹¹ beteiligen (Kultusportal, Vergleichsarbeiten; Landesbildungsserver, Vergleichsarbeiten VERA8). VERA 8 wird die landesspezifischen DVA ablösen (Landesbildungsserver, Lernstandserhebungen). Außerdem wird im Schuljahr 2015/2016 zusätzlich ein landesspezifisches förderdiagnostisches Verfahren in allen fünften Klassen eingeführt: Lernstand 5. Dieses Instrument soll anhand von Basiskompetenzen fachdidaktische und differenzierte Informationen über den Lernstand von Schülerinnen und Schülern aufzeigen (Kultusportal, Vergleichsarbeiten; Kultus und Unterricht, 2014). Grund für die Einführung ist der Wegfall der verpflichtenden Grundschulempfehlung und die dadurch zu erwartende heterogenere Schülerschaft. Das Ziel ist – vergleichbar mit dem Ziel des Testverfahrens *Online Diagnose* –, eine Ausgangsbasis für das individuelle Weiterlernen in der Sekundarstufe zu erhalten. Aus diesem Grund werden die Instrumente im Primarbereich pilotiert (Landesbildungsserver, Lernstand 5).

Vergleichsarbeiten in Baden-Württemberg. Aufgrund von konzeptionellen Erfahrungen und bereits bestehenden wissenschaftlichen Erkenntnissen zu den Vergleichsarbeiten (VERA)¹² wird nachfolgend auf deren Rahmenbedingungen näher eingegangen (IQB, VERA; Landesbildungsserver, Lernstandserhebungen; KMK, Vergleichsarbeiten). Die Konzeptionierung der Vergleichsarbeiten unterscheidet sich von den internationalen (z.B.

⁷ Tradition und Innovation (TRAIN)

⁸ Ländervergleiche sind Teil eines umfassenden Systems von Lernstandserhebungen (IQB, Ländervergleich)

⁹ Vergleichsarbeiten an Grundschulen (VERA 3)

¹⁰ Vergleichsarbeiten an weiterführenden Schulen (DVA)

¹¹ Vergleichsarbeiten in der Sekundarstufe (VERA 8)

¹² Vergleichsarbeiten (VERA)

PISA) und nationalen (z.B. Ländervergleiche) Schulleistungsstudien. Ziel ist nicht das Systemmonitoring, sondern die Unterrichts- und Schulentwicklung (IQB, VERA). Vergleichsarbeiten untersuchen in allen allgemeinbildenden Schulen, Kompetenzen von Schülerinnen und Schülern, die sie zu einem bestimmten Zeitpunkt der dritten oder achten Klasse erreicht haben (IQB, VERA). Wird das Fach Deutsch in der dritten Jahrgangstufe als Beispiel herangezogen, sind Tests in den folgenden Kompetenzbereichen möglich: Lesen, Zuhören, Orthografie sowie Sprache und Sprachgebrauch. Die Unterschiede zwischen Vergleichsarbeiten und Klassenarbeiten liegen in der Leistungsbeurteilung und in der Form der gestellten Aufgaben. Diese greifen auf den Transfer von Kompetenzen zurück und fragen nicht spezielle Lernaufgaben ab (Wacker & Kramer, 2012). Die Ergebnisse im Rahmen von VERA sollen den Lehrkräften zeigen, wo ihre Schülerinnen und Schüler im Vergleich zu anderen Klassen oder Schulen stehen und an welcher Stelle sie besondere Stärken bzw. Schwächen aufweisen (IQB, FAQ). Die Vergleichswerte werden in Baden-Württemberg der zuvor erfolgten Pilotierung entnommen. Im Idealfall würden die Lehrkräfte auf der Grundlage der erhaltenen Informationen den Unterricht und die Lehr-/Lernprozesse weiterentwickeln können.

Für jede Durchführung der Vergleichsarbeiten werden unter Einbezug von Lehrkräften neue Aufgaben entwickelt, die durch das IQB pilotiert und in Testheften zusammengestellt werden (IQB, VERA). Das IQB ist hierbei in der zentralen Verantwortung der Aufgabenentwicklung, Aufgabenerprobung (Pilotierung), Ermittlung der Aufgabenschwierigkeiten (Skalierung), Testhefterstellung und Erarbeitung begleitender didaktischer Materialien (IQB, VERA). Alle weiteren Aufgaben liegen in der Verantwortung der einzelnen Länder und werden daher zum Teil unterschiedlich gehandhabt: Druck und Distribution der Testhefte, Testdurchführung, Korrektur und Dateneingabe, statistische Auswertung, Rückmeldungsgestaltung, Rückmeldung, Unterstützung der Schulen durch flankierende Maßnahmen nach der Ergebnisrückmeldung (IQB, VERA). In Baden-Württemberg werden die Vergleichsarbeiten von den Lehrkräften selbst durchgeführt und ausgewertet. In anderen Ländern bzw. bei Schulleistungsstudien werden in der Regel externe Testleiterinnen und Testleiter eingesetzt und die Tests zentral ausgewertet. Zur Auswertung der Vergleichsarbeiten in Baden-Württemberg erhalten die Lehrkräfte Unterstützung von den Landesinstituten für Schulentwicklung (LS, 2013). Ein Vorteil der Auswertung durch die Lehrkräfte ist, dass diese bereits während der Auswertung das Zustandekommen der Ergebnisse miterleben. Sie erhalten die Ergebnisrückmeldungen zu den Vergleichsarbeiten also sofort bei bzw. nach der

Auswertung (IQB, VERA). Der Umgang mit den zurückgemeldeten Ergebnissen liegt in Baden-Württemberg in der Verantwortung der jeweiligen Schule. Hierfür stellt das Landesinstitut für Schulentwicklung eine Handreichung zur Verfügung (LS, 2013). Diese soll mit konkreten Handlungsvorschlägen und -anregungen die Lehrkräfte bei der Nutzung der Ergebnisse unterstützen. Zu diesem Zweck wurden drei Schritte ausgearbeitet: 1. Analyse der Ergebnisse, 2. Interpretation der Ergebnisse und 3. Konsequenzen aus den Ergebnissen. Im ersten Schritt werden die Lehrkräfte angehalten, die Ergebnisse systematisch zu betrachten, um Auffälligkeiten festzustellen. Der zweite Schritt sieht die Interpretation der Ergebnisse im Kontext bzw. im Abgleich mit Vergleichswerten vor. Bereits hier wird den Lehrkräften empfohlen, erste Bewertungen und Erklärungsansätze vorzunehmen. Im dritten Schritt soll die Festlegung und Planung von Zielen und Maßnahmen für das weitere Vorgehen stattfinden. Für alle drei Schritte wird die Dokumentation der Auffälligkeiten, Ziele und Maßnahmen empfohlen. Ebenso werden die Lehrkräfte aufgefordert, die Auswertung und Festlegung der Maßnahmen im Team erfolgen zu lassen (LS, 2013).

Da die Anwendung der Vergleichsarbeiten in der Verantwortung der Länder liegt, hat dies unterschiedliche Vorgehensweisen der Nutzung zur Folge. Maier (2008b) verglich die Vorgehensweise der Länder Baden-Württemberg und Thüringen. Unterschiede in der Vorgehensweise waren z.B. in der Art der Implementierung der Tests, im Format und im zeitlichen sowie inhaltlichen Umfang der Leistungsrückmeldungen, in den Unterstützungssystemen und den projektinternen Evaluationskonzeptionen festzustellen (Maier, 2008b). In Baden-Württemberg werteten die Lehrkräfte die Vergleichsarbeiten mit Unterstützung der Landesinstitute für Schulentwicklung selbst aus. Die Thüringer Lehrkräfte erhielten zwei Wochen nach der Eingabe der Daten einen ersten Sofortbericht zu den Lösungshäufigkeiten der einzelnen Aufgaben sowie Klassenergebnisse zu einzelnen Kompetenz- und Anforderungsbereichen. Als Vergleichsmaßstab wurde in Baden-Württemberg die landesweite Pilotierungsstichprobe verwendet. In Thüringen erhielten die Lehrkräfte Landesmittelwerte, die in Bezug auf den sozioökonomischen Status der Familien korrigiert wurden. Zur Nutzung der Rückmeldungen erhielten die baden-württembergischen Lehrkräfte Unterstützung z.B. in Form einer Handreichung. In Thüringen wurden die Rückmeldungen mit weiterführenden Fehleranalysen verknüpft und in Fortbildungsangebote eingebettet.

In einer Lehrkraftbefragung ($n = 1136$) wurde untersucht, ob aufgrund der unterschiedlichen Vorgehensweise der beiden Länder die Akzeptanz- und Nutzungseinschätzung

der Lehrkräfte variierte und ob dadurch die schulinterne Diskussion über Testrückmeldungen beeinflusst wurde (Maier, 2008b). Die Ergebnisse der Untersuchung zeigten, dass die Tests bei Thüringer Lehrkräften eher akzeptiert und als weniger belastend empfunden wurden. Außerdem wurde der Nutzen der lerndiagnostischen Hinweise, die sich aus den Tests ergaben, von den Thüringer Lehrkräften als höher eingeschätzt. Dahingegen wurde durch die baden-württembergischen Lehrkräfte eher ein Hinweis für die Notengebung aus den Tests aufgenommen. Die Besprechung der zentralen Rückmeldungen wurde von den Thüringer Lehrkräften im Vergleich zu den baden-württembergischen Lehrkräften als systematischer eingeschätzt (Maier, 2008b). Die Ergebnisse zeigten, dass die Nutzung von Ergebnissrückmeldungen vielschichtig ist und eine (wissenschaftliche) Begleitung dieser Nutzung für die Nachhaltigkeit bedeutend. Das im nachfolgenden Kapitel beschriebene Modell greift die Komplexität der Nutzung von Ergebnissrückmeldungen auf und stellt diese in einem Prozess sowie in Bedingungen als Einflussfaktoren auf den Prozess dar. Das Zyklenmodell der Rückmeldung (z. B. Helmke & Hosenfeld, 2005; Hosenfeld & Groß Ophoff, 2007) wird anschließend näher ausgeführt.

2.3 Das Zyklenmodell der Rückmeldung

Die Verfahren der im Rahmen der bildungspolitischen Beschlüsse festgelegten Schulleistungsstudien, Ländervergleiche und Vergleichsarbeiten zur Qualitätssicherung und Standardisierung des Bildungssystems liefern den Ländern, Schulen, Lehrkräften, Schülerinnen und Schülern usw. Leistungsrückmeldungen. Diese Leistungsrückmeldungen können z. B. der frühzeitigen Förderung von Schülerinnen und Schülern sowie der kompetenz- und standardbasierten Unterrichtsentwicklung durch die Lehrkräfte dienen. Wie die im vorhergehenden Abschnitt dargestellten Beispiele zur Nutzung von Ergebnissrückmeldungen aus Vergleichsarbeiten aufzeigten, stellen Leistungsrückmeldungen notwendige, aber keine ausreichenden Voraussetzungen für die Verbesserung von Lehre und Lernen dar (Helmke & Hosenfeld, 2005). Das in Abbildung 2 dargestellte Zyklenmodell der Rückmeldung von Helmke und Hosenfeld (2005; siehe auch Helmke, 2004; Hosenfeld & Groß Ophoff, 2007) veranschaulicht den Weg von der Information zur Innovation als „(...) weit und beschwerlich“ (S. 147). Das bedeutet, dass Leistungsrückmeldungen zunächst verstanden, reflektiert, in konkrete Maßnahmen umgesetzt und wieder evaluiert werden müssen, damit sie überhaupt zu Ver-

besserungen und Entwicklungen führen können. Helmke und Hosenfeld (2005) nennen diesen Weg „(...) eine lange und in mehrfacher Hinsicht fragile Kette“ (S. 148). In jedem Abschnitt oder Übergang kann diese Kette, im Folgenden als Prozessschritte bezeichnet, unterbrochen oder abgebrochen werden. Weitere Einflüsse, im Zyklusmodell der Rückmeldung als Bedingungen dargestellt, können die Prozessschritte fördern oder behindern.

Im Folgenden werden die einzelnen Prozessschritte und Bedingungen des Zyklusmodells der Rückmeldung dargelegt. Zunächst wird die Art der Rückmeldung aufgezeigt, die im ersten Kasten der Abbildung 2 die Information durch die Leistungsrückmeldungen und die Vergleiche der Leistungsrückmeldungen mit Vergleichswerten bezeichnet. Im Anschluss werden die Prozessschritte Rezeption, Reflexion, Aktion und Evaluation im Einzelnen dargestellt und abschließend die individuellen, schulischen und externen Bedingungen aufgeführt.

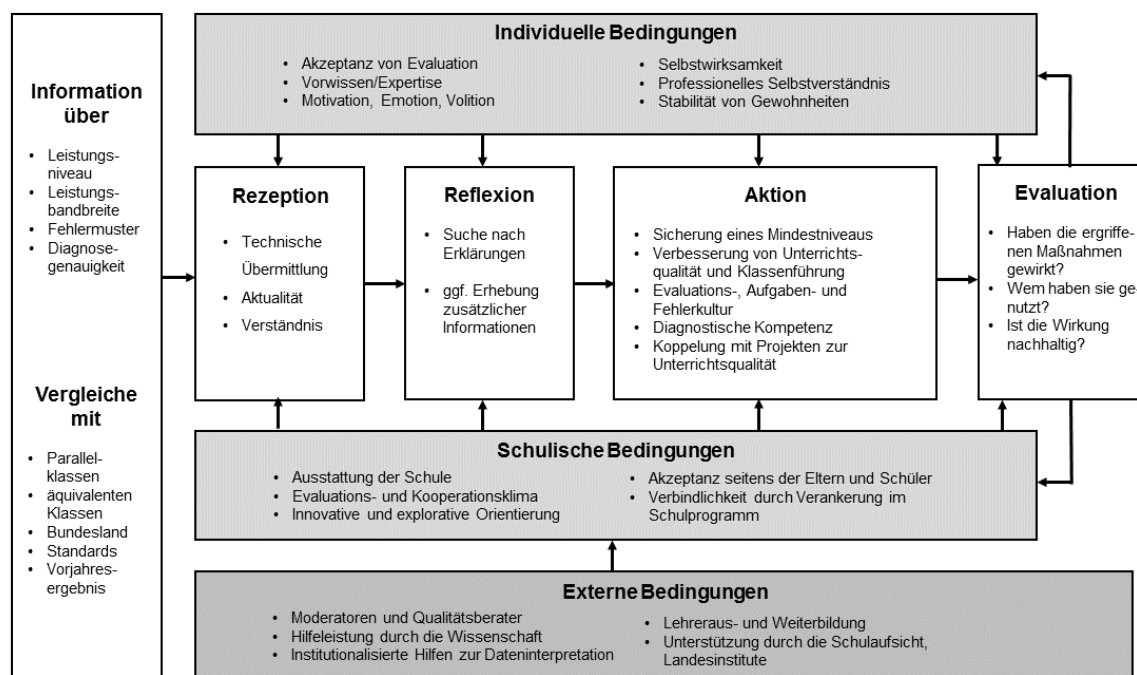


Abbildung 2. Zyklusmodell der Rückmeldung (vgl. Helmke, 2004; Helmke & Hosenfeld, 2005; Hosenfeld & Groß Ophoff, 2007)

Art der Ergebnissrückmeldungen. Helmke und Hosenfeld (2005) nennen zum Einstieg in das Zyklusmodell der Rückmeldung die zu berichtenden Informationen über Schülerinnen und Schüler, wie Leistungsniveau, Leistungsbandbreite, Fehlermuster und Diagnosegenauigkeit. Außerdem werden Vergleichswerte als Vergleiche mit äquivalenten Klassen

(faire Vergleiche), mit Vergleichswerten des Bundeslandes, mit Standards oder mit Vorjahresergebnissen genannt (Helmke & Hosenfeld, 2005). Schneewind (2006) ergänzt den Einstieg in das Zyklenmodell der Rückmeldung durch einige den Ergebnisrückmeldungen und dem Verwendungsprozess vorausgehende Schritte. Sie nennt Kriterien wie die Art der Evaluations- bzw. Leistungsstudie, die Ziele der Rückmeldung, die Zielgruppe der Rückmeldung, die Kommunikation der Ziele, die Stichprobe, das Aggregationsniveau der Daten sowie die Bezugsnormen. Die Kriterien fasst sie unter dem Begriff der „Gestaltung der Rückmeldung“ zusammen (Schneewind, 2006, S. 51).

Die Art der Ergebnisrückmeldung unterscheidet sich je nach Absicht, Form und Inhalt (Groß Ophoff, Koch, Hosenfeld & Helmke, 2006). Diese Kriterien sind bereits mit der Durchführung der Studie bzw. der Datenerhebung verknüpft (z.B. Schneewind, 2006). Hat eine Studie z.B. zum Ziel, Schülerinnen und Schüler individuell zu fördern, müssen Daten auf Schülerebene erhoben und zurückgemeldet werden. Eine Ergebnisrückmeldung, die sich auf die gesamte Klasse bezieht, würde das angegebene Ziel nicht erfüllen. Aus diesem Grund ist es nicht möglich, Ergebnisrückmeldungen unabhängig von den Studien, aus denen sie hervorgehen, zu betrachten (Schneewind, 2006). Neben der Abhängigkeit der Ergebnisrückmeldungen von den Zielen der Studie können hier ebenso Kriterien der Qualität interpretiert werden. Wenn als Beispiel die Diagnosegenauigkeit herangezogen wird, muss diese umso detaillierter sein, je individueller die Ergebnisrückmeldung sein soll. Außerdem kann die Diagnosegenauigkeit hinsichtlich der Messgenauigkeit eines Tests im Rahmen der psychologischen Psychometrie interpretiert werden.

Ein weiteres Qualitätskriterium von Ergebnisrückmeldungen stellen Vergleichswerte zu den Leistungsrückmeldungen dar. Zur Einordnung der Ergebnisrückmeldungen bzw. der Leistungen von Schülerinnen und Schülern werden Vergleichswerte benötigt. Im Rahmen von Schulleistungsstudien wurden die Möglichkeiten des Leistungsvergleiches definiert und erprobt (Rheinberg & Fries, 2010). Die von Rheinberg und Fries (2010) zusammengefasste Bezugsnormorientierung stellt folgende Möglichkeiten des Vergleichs dar: Die individuelle bzw. ipsative Bezugsnorm, die sachliche bzw. kriteriale Bezugsnorm und die soziale Bezugsnorm. Die individuelle bzw. ipsative Bezugsnorm besteht aus Vergleichen vorhergehender Leistungen derselben Schülerin bzw. desselben Schülers mit der aktuell erfassten Leistung. Die Leistung zu einem früheren Zeitpunkt dient also als Referenz für intraindividuelle Leistungsveränderungen (Fiege, 2013). Die sachliche bzw. kriteriale Bezugsnorm

wird von Rheinberg und Fries (2010) als der Vergleich der Leistung mit einer vorab festgelegten und inhaltlich begründeten Idealform beschrieben. In diesem Rahmen können z.B. die Bildungsstandards als Vergleichsmaßstab herangezogen werden. Die soziale Bezugsnorm stellt den Vergleich zwischen Schülerinnen und Schülern dar (Rheinberg & Fries, 2010). Dieser Vergleich basiert z.B. auf der Leistungsverteilung aller Schülerinnen und Schüler, deren Leistung erhoben wurde (Fiege, 2013). Im Rahmen des sozialen Vergleichs ist es jedoch nicht möglich, die reine Leistung als Vergleichsmaßstab heranzuziehen. Außerschulische Faktoren haben ebenfalls Einfluss auf das Lernen. So wurden zum Beispiel zu den Vergleichsarbeiten (VERA)¹³ im Rahmen der sogenannten fairen Vergleiche die Vergleichswerte unter anderem für das Vorwissen, den sozioökonomischen Status, das Geschlecht, die Muttersprache und die soziale Komposition der Klasse kontrolliert (Fiege, 2013). Diese Adjustierungsverfahren werden bereits in vielen Studien eingesetzt (z.B. Kühle & Peek, 2007; Nachtigall & Jantowski, 2007). Für die baden-württembergischen DVA¹⁴ wurden die Ergebnisse der zuvor durchgeführten Pilotierung als landesweite Vergleichswerte herangezogen (Maier, 2009). Erfahrungen aus der Studie BeLesen¹⁵ zeigten, dass die Lehrkräfte im Rahmen der sozialen Bezugsnorm der Einteilung von Schülerinnen und Schülern in Leistungsgruppen kritisch gegenüberstanden, vor allem deshalb, weil die Zuteilung der Schülerinnen und Schüler zu den Leistungsgruppen den Lehrkräften als nicht durchsichtig erschien (Schneewind, 2006). Tresch (2007) forderte nach den Auswertungen ihrer Untersuchung Check 5¹⁶, dass den Lehrkräften möglichst viele Vergleichsgrößen wie die soziale oder kriteriale Bezugsnorm oder der Vergleich mit vorhergehenden Ergebnissen zur Verfügung gestellt werden sollte.

Prozessschritt Rezeption. Der Prozessschritt Rezeption umfasst im Zyklusmodell der Rückmeldung die technische Übermittlung der Ergebnisrückmeldungen sowie die Wahrnehmung durch den Rezipienten und das Verständnis der Ergebnisrückmeldungen (Helmke & Hosenfeld, 2005; vgl. auch Schneewind, 2006). Mit anderen Worten: Die Aktualität, Nützlichkeit und Verständlichkeit von Ergebnisrückmeldungen erwies sich in bisher durchgeführten Untersuchungen als zentral (Müller, 2009). In Studien aus dem deutschsprachigen Raum, in deren Rahmen den Lehrkräften Individualdaten der einzelnen Schülerinnen und

¹³ Informationen zu den Vergleichsarbeiten (VERA) siehe in Abschnitt 2.2

¹⁴ Vergleichsarbeiten in der Sekundarstufe (DVA)

¹⁵ Berliner Längsschnittstudie zur Lesekompetenz (BeLesen)

¹⁶ Schulleistungsmessung im Kanton Aargau, Schweiz (Check 5)

Schüler zurückgemeldet wurden und an die eine Rezeptionsstudie anschloss (Dedering, 2011, S. 66-68), zeigten sich folgende Befunde:

Die Aktualität der Rückmeldung ist von Bedeutung. Dauert die Auswertung zu lange, können die Ergebnisse bereits veraltet sein oder den Lehrkräften als veraltet erscheinen und deshalb keine weitere Berücksichtigung finden. Außerdem muss ein möglicher Klassenwechsel bedacht werden. Unter Umständen steht zum Ende eines Schuljahres ein Klassenwechsel der durchführenden Lehrkraft an, sodass keine Folgeaktivitäten auf die Ergebnisrückmeldungen hin stattfinden könnten (Maier & Rauin, 2006). Bei einigen Studien wird daher auf ein System von mehreren Rückmeldezeitpunkten zurückgegriffen. Zur Studie VERA¹⁷ in Rheinland-Pfalz wurden in einer ersten Welle die Fähigkeitsniveaus der einzelnen Schülerinnen und Schüler zurückgemeldet. In einer zweiten Welle wurden die Fähigkeitsniveaus im Vergleich zum eigenen Bundesland, faire Vergleiche zu einer landesspezifischen Kontextgruppe und die Lösungshäufigkeiten in den Zentralaufgaben im Vergleich zur Gesamtstichprobe zurückgemeldet (Groß Ophoff, Koch, Hosenfeld & Helmke, 2006). Auch zu den Thüringer Kompetenztests gab es Sofortrückmeldungen nach Abschluss der Dateneingabe und acht Wochen später Klassen- und Schulergebnisberichte mit Vergleichswerten. Weitere vier Monate nach der Dateneingabe erhielten die Lehrkräfte die auf Landesebene analysierten Ergebnisrückmeldungen (Nachtigall & Jantowski, 2007). Die Wichtigkeit der Aktualität von Ergebnisrückmeldungen stellte auch Tresch (2007) im Rahmen der Untersuchungen zu Check 5¹⁸ fest. Zu den Ergebnisrückmeldungen erhielten die Lehrkräfte ein halbes Jahr nach der Durchführung der Leistungstests eine Handreichung mit Vergleichswerten und konkreten Beispielen.

Untersuchungen zur Nützlichkeit von Ergebnisrückmeldungen zeigten, dass die von den Lehrkräften angegebene Nützlichkeit mit der Häufigkeit der zurückgemeldeten Ergebnisse stieg (Kühle & Peek, 2007). Weiterhin wurde im Rahmen der Untersuchungen zu den Lernstandserhebungen in Nordrhein-Westfalen ein korrelativer Zusammenhang der Angaben zur Nützlichkeit und Verständlichkeit festgestellt (Kühle & Peek, 2007). Je nützlicher und verständlicher die Ergebnisrückmeldungen zur Studie BeLesen¹⁹ in der Wahrnehmung der Probanden waren, desto eher wurden sie für die Lehrplanung und den Förderunterricht genutzt (Schneewind, 2006).

¹⁷ Vergleichsarbeiten in der Grundschule (VERA)

¹⁸ Schulleistungsmessung Kanton Aargau, Schweiz (Check 5)

¹⁹ Berliner Längsschnittstudie zur Lesekompetenz (BeLesen)

Die Lehrkräfte bezeichneten die Ergebnisrückmeldungen in den unterschiedlichsten Studien mehrheitlich als verständlich (z. B. Schneewind, 2007). Hierbei wurde die durch die Lehrkräfte bewertete subjektive Verständlichkeit erfasst. Zur Studie BeLesen gaben die Lehrkräfte an, die Übertragung der zurückgemeldeten Testergebnisse auf die Schülerinnen und Schüler als leicht empfunden zu haben. Ebenso gaben die Lehrkräfte an, keine Schwierigkeiten gehabt zu haben, die Rückmeldungen mit der eigenen Einschätzung zu vergleichen (Schneewind, 2006). An den Lernstandserhebungen in Nordrhein-Westfalen, die mehrere Messzeitpunkte implizierten, konnte gezeigt werden, dass Verständnisschwierigkeiten bei wiederholten Rückmeldungen abnahmen (Kühle & Peek, 2007). Hinsichtlich Befragungen zur Darstellung von Ergebnisrückmeldungen zeigten sich die Antworten weniger positiv. Vor allem Fachbegriffe, Abkürzungen und statistische Bezeichnungen (wie z. B. *N*) gaben die Lehrkräfte im Rahmen der Studie BeLesen als schwer verständlich an. Der Eindruck der Lehrkräfte, die Ergebnisrückmeldungen nicht verstehen zu können, zeichnete sich als Hürde für die Nutzung ab (Schneewind, 2006). Für Angaben wie z. B. Durchschnittswerte der Klasse zeigten sich die Lehrkräfte wenig interessiert (Schneewind, 2006). Tresch (2007) resümierte aus den Ergebnissen ihrer Studie zu Check 5, dass zu den Ergebnisrückmeldungen eine Beschreibung der Testinhalte und Aufgabenbeispiele angefügt werden sollte (vgl. auch Kühle & Peek, 2007). Die Darstellung der Ergebnisrückmeldungen sollte einfach gehalten sein und sich auf Prozentangaben, Punkte oder Noten beziehen, die den Lehrkräften aus dem Unterrichtsalltag bekannt sind (Tresch, 2007). Abbildungen und Tabellen erwiesen sich für die Nutzung der Ergebnisrückmeldungen als hilfreich und übersichtlich (Tresch, 2007).

Prozessschritt Reflexion. Liefern die Ergebnisrückmeldungen den Rezipienten neue Informationen, beginnt im Prozessschritt Reflexion die Suche nach Erklärungen und gegebenenfalls die Beschaffung zusätzlicher Informationen (Helmke & Hosenfeld, 2005; vgl. auch Müller, 2009; Hosenfeld, 2010).

Die Lehrkräfte sind hier angehalten die Ergebnisrückmeldungen zu analysieren und mit ihren vorhandenen Kenntnissen zu vergleichen (Hosenfeld, 2010). Die Zuschreibung von Ursache und Wirkung kann auf unterschiedlichen Wegen geschehen (Müller, 2009): Zum einen beginnt die Suche beim Rezipienten bzw. der Lehrkraft selbst und zum anderen bei den Informationslieferanten bzw. bei den Schülerinnen und Schülern. Rezipient und Informationslieferant gehören häufig nicht derselben Personengruppe an. Geschieht die Zu-

schreibung von Ursache und Wirkung zu einseitig, kann dies zu einer verschobenen Attribuierung führen (Kohler, 2004; vgl. auch Schneewind, 2007; Kuper & Hartung, 2007), beispielsweise dann, wenn Lehrkräfte die Ergebnisrückmeldungen nicht auf das eigene Unterrichtshandeln beziehen, sondern nur auf die Einzelleistung ihrer Schülerinnen und Schüler (Schneewind, 2007).

Prozessschritt Aktion. Der an die Reflexion anschließende Prozessschritt Aktion beschreibt die konkrete Entwicklung und Umsetzung von Maßnahmen (Schneewind, 2007). Dazu nennen Helmke und Hosenfeld (2005) folgende beispielhafte Möglichkeiten: Sicherung eines Mindestniveaus an Kompetenz, Förderung leistungsschwacher Schülerinnen und Schüler, generelle Steigerung der Unterrichtsqualität und der Klassenführung, Änderung der Evaluations-, Aufgaben- und Fehlerkultur sowie Schulung der diagnostischen Kompetenz und Koppelung mit Projekten der Unterrichtsqualität. In diesem Sinne sollen Änderungen wie z.B. die Förderung individueller Leistungen durchgeführt werden (Müller, 2009). In Studien aus dem deutschsprachigen Raum, in deren Rahmen den Lehrkräften Individualdaten der einzelnen Schülerinnen und Schüler zurückgemeldet wurden und an die sich eine Rezeptionsstudie anschloss (Dedering, 2011, S. 66-68), zeigten sich folgende Befunde:

Neben der Verständlichkeit, Nützlichkeit und Akzeptanz von Ergebnisrückmeldungen und externen Leistungsstudien wurden daraus resultierende Konsequenzen erfasst. Die Lehrkräfte bezeichneten die Ergebnisrückmeldungen häufig als verständlich, zeigten sich jedoch hinsichtlich deren Umsetzung als eher hilflos. Die Übertragung der Ergebnisrückmeldungen auf den Unterricht kann nicht als automatisiert angenommen werden (Schneewind, 2006). Bei der Rückmeldung von Ergebnissen wird von Veränderungen oder Folgemaßnahmen ausgegangen. Einige Studien zeigten auf, dass die Ergebnisrückmeldungen durchaus zur Reflexion anregten, allerdings nicht zwingend zu Handlungen (z.B. Maier, 2009; Maier & Rauin, 2006; Schneewind & Kuper, 2009). Auch Tresch (2007) stellte in ihrer Untersuchung zur Studie Check 5²⁰ fest, dass viele Maßnahmen aus den Ergebnisrückmeldungen abgeleitet wurden, diese allerdings an der Umsetzung scheiterten. Lehrkräfte unterschätzten häufig die zeitlichen und materiellen Ressourcen, die für eine vollständige Um-

²⁰ Schulleistungsmessung Kanton Aargau, Schweiz (Check 5)

setzung der abgeleiteten Maßnahmen notwendig waren (Tresch, 2007). Hinsichtlich der Studie BeLesen²¹ wurden vor allem für solche Schülerinnen und Schüler Maßnahmen durchgeführt, die mit einer verzögerten Lernentwicklung auffielen (Schneewind, 2006). Lehrkräfte griffen dann auf Daten zurück, wenn es eine Verknüpfung zu bereits vorliegenden diagnostischen Informationen oder schulischen Leistungsdaten gab (Kühle & Peek, 2007). Ebenso wurden die Ergebnisse besonders dann rezipiert und auf unterrichtliche Leistungen bezogen, je stärker die Schule von den jeweiligen Inhalten der Ergebnisrückmeldungen betroffen war (Nachtigall & Jantowski, 2007). Einige Forscherinnen und Forscher berichteten von einer einseitigen Attribuierung im Rahmen der Nutzung von Ergebnisrückmeldungen (Schneewind & Kuper, 2009). Zum Teil beruht dies auf der Datenerhebung. Ergebnisrückmeldungen, die auf Leistungen der Schülerinnen und Schüler, also auf Schülerdaten beruhen, bezogen die Lehrkräfte nicht intuitiv auf ihre eigene Unterrichtsleistung (z.B. Schneewind, 2007). Somit wurden oftmals allein die Leistungen der Schülerinnen und Schüler für schwache Testergebnisse verantwortlich gemacht.

Prozessschritt Evaluation. Der Prozessschritt Evaluation stellt den Rückblick auf die gesamte Prozesskette dar und schließt damit den Kreislauf (Hosenfeld, 2010). Helmke und Hosenfeld (2005) schlagen vor zu überprüfen, ob die ergriffenen Maßnahmen wirkten, wem diese nutzten und ob die Wirkung nachhaltig war. Die Wirkung der Prozesskette wird in diesem letzten Schritt evaluiert und produziert automatisch neue Ergebnisrückmeldungen. Aus den zum Prozessschritt Aktion festgehaltenen Ergebnissen lässt sich schließen, dass keine Wirkungen evaluiert werden können, wenn keine Maßnahmen umgesetzt wurden. Auch bei einer oberflächlichen Auseinandersetzung mit den Ergebnisrückmeldungen wäre keine oder höchstens eine geringe Nachhaltigkeit zu erwarten (Groß Ophoff, Koch, Hosenfeld & Helmke, 2006).

Bedingungen des Zyklusmodells der Rückmeldung. Auf die Prozessschritte wirken unterschiedliche Bedingungsfaktoren, die im Zyklusmodell der Rückmeldung in die individuelle, schulische und externe Bedingung unterteilt wurden (Hosenfeld & Groß Ophoff, 2007; Helmke & Hosenfeld, 2005).

²¹ Berliner Längsschnittstudie zur Lesekompetenz (BeLesen)

Mit den individuellen Bedingungen werden Lehrermerkmale angesprochen. Helmke und Hosenfeld (2005) greifen hierzu eine Reihe von Kriterien auf: Ist-Soll-Vergleiche, Zufriedenheit, Vorwissen, Expertise, Motivation und Volition, Akzeptanz von Evaluationen, Medien- und IT-Kompetenz, Selbstwirksamkeit, professionelles Selbstverständnis und Stabilität von Gewohnheiten.

Die schulischen Bedingungen beeinflussen durch die Ausstattung der Schule, das Evaluations- und Kooperationsklima, die innovative und explorative Orientierung, die Akzeptanz durch die Eltern sowie Schülerinnen und Schüler und durch die Verbindlichkeit durch die Verankerung im Schulprogramm (Hosenfeld & Groß Ophoff, 2007; vgl. auch Müller, 2009). Außerdem kann die Akzeptanz der Testinstrumente und Ergebnisrückmeldungen durch das Kollegium und Schulleitung als wichtiges Kriterium genannt werden (Schneewind, 2006).

Die externen Bedingungen erfassen Hilfestellungen auf der Makroebene. Neben anderen gehören hierzu die im Zyklenmodell der Rückmeldung erfassten Punkte: Moderatoren und Qualitätsberater, Hilfestellungen durch die Wissenschaft, institutionalisierte Hilfen zur Dateninterpretation, Lehrer-Aus- und -Weiterbildungen sowie Unterstützung durch die Schulaufsicht (Hosenfeld & Groß Ophoff, 2007; vgl. auch Helmke & Hosenfeld, 2005; Müller, 2009).

In Studien aus dem deutschsprachigen Raum, in deren Rahmen den Lehrkräften Individualdaten der einzelnen Schülerinnen und Schüler zurückmeldet wurden und an die eine Rezeptionsstudie anschloss (Dedering, 2011, S. 66-68), zeigten sich folgende Befunde zur Akzeptanz und Kooperation: In einigen Studien hatte es sich gezeigt, dass die Akzeptanz von externen Leistungsstudien durch die Lehrkräfte einen starken Einfluss auf die weitere Verarbeitung von Ergebnisrückmeldungen hat. Maier (2008b) stellte den Zusammenhang zwischen der Akzeptanz und der Nutzung zentraler Tests fest. Ebenso erwiesen sich die Kriterien Akzeptanz und Nützlichkeit der Lernstandserhebungen in Nordrhein-Westfalen als prädiktiv für die Bereitschaft der nutzungsorientierten Rezeption von Ergebnisrückmeldungen (Kühle & Peek, 2007). Schneewind (2006) stellte zur Studie BeLesen²² fest, dass die Akzeptanz sowie die positive Einschätzung der Qualität für die Nutzung der Ergebnisrückmeldungen relevant sind. Die durch die Lehrkräfte subjektiv eingeschätzte Qualität kann allerdings durchaus von der objektiven Qualität eines Testverfahrens abweichen. Bei Unter-

²² Berliner Längsschnittstudie zur Lesekompetenz (BeLesen)

suchungen zu den Thüringer Kompetenztests wurde eine Zunahme der Akzeptanz, des innovativen Umgangs mit den Ergebnisrückmeldungen und der Veränderungsbereitschaft nach mehrmaliger Durchführung festgestellt (Nachtigall & Jantowski, 2007). Je häufiger die Kompetenztests bereits durchgeführt wurden, desto eher wurden diese von den Lehrkräften akzeptiert. Die Akzeptanz von Ergebnisrückmeldungen hängt unter anderem auch damit zusammen, wie die Ergebnisrückmeldungen kommuniziert wurden und ob diese Kommunikation ausreichend umfangreich war (Tresch, 2007). Fühlten sich die Lehrkräfte ausreichend und positiv informiert, wurden die Ergebnisrückmeldungen eher akzeptiert. Außerdem kann die positive Einstellung von Lehrkräften gegenüber externen Leistungserhebungen über den Einbezug und die Beteiligung der Lehrkräfte z.B. bei der Auswahl von Testaufgaben und der Datenerhebung gesteigert werden (Dedering, 2011). Somit haben die Lehrkräfte die Möglichkeit, für sie relevante Themen mitzubestimmen.

Die Kooperations- und Evaluationskultur ist maßgeblich für den Prozess der Auseinandersetzung mit Ergebnisrückmeldungen (Kühle & Peek, 2007). Ein fachgruppenbezogener Austausch fördert z.B. ein gemeinsames Beurteilungsschemata und einen einheitlichen Umgang mit den Rückmeldungen (Hartung-Beck, 2009). Von einer schulübergreifenden Kooperation hinsichtlich zurückgemeldeter Ergebnisse berichteten die Lehrkräfte z.B. im Rahmen der Thüringer Kompetenztests nur selten. Andererseits wurden die Ergebnisse verstärkt im Kollegium besprochen (Nachtigall & Jantowski, 2007). Auch im Rahmen der Studie BeLesen fanden Gespräche insbesondere zwischen Kolleginnen und Kollegen, die in derselben Klasse unterrichteten, oder mit Kolleginnen und Kollegen, die ebenfalls das Projekt durchführten, statt (Schneewind, 2006). Zu einigen Studien z.B. zu den baden-württembergischen DVA²³ fand der Austausch im Kollegium über die Ergebnisrückmeldungen eher auf informeller Ebene statt. Auch die Informationen an die Eltern fanden im Rahmen der DVA eher in reduzierter Form statt (Maier, 2009).

Das Zyklenmodell der Rückmeldung wurde bereits häufig zitiert und vielen Studien zugrunde gelegt (z.B. Hosenfeld, 2010; Koch, 2011; Müller, 2009; Schneewind, 2006). Die feinen Variationen der Studien erschweren eine Vereinheitlichung und den Vergleich der bereits vorliegenden Untersuchungen. Andererseits ermöglichen die individuellen Ausgestaltungen der Prozessschritte und Bedingungen eine flexible Passung zur jeweiligen Studie. In Anbetracht der Komplexität ist eine vollständige Überprüfung des Modells nicht erstrebenswert. Dennoch können für die Nutzung von Ergebnisrückmeldungen relevante Kriterien

²³ Vergleichsarbeiten in der Sekundarstufe (DVA)

herausgearbeitet werden: „Das Modell eignet sich, um Phasen der Nutzbarmachung und externe Einflussfaktoren zu lokalisieren. Für eine empirische Überprüfung fehlt jedoch die inhaltliche Spezifikation von Zusammenhängen.“ (Maier, 2008a, S. 97).

2.4 Die Nutzung von Ergebnisrückmeldungen

Zur Nutzung von Ergebnisrückmeldungen in Bezug auf die individuelle Förderung von Schülerinnen und Schülern können aus der spezifischen Literatur zwei Begriffe herangezogen werden: Rückmeldung und Feedback. Nachfolgend werden Unterschiede und Überschneidungen der beiden Begriffe dargestellt. Außerdem wird die Komplexität der Nutzung von Ergebnisrückmeldungen mit dem Fokus auf Lernende durch die multidimensionale Konzeptualisierung von Hattie und Wollenschläger (2014; vgl. auch Hattie & Timperley, 2007) aufgezeigt. Hierbei wird auf das direkte Feedback für eine leistende Person durch eine rückmeldende Person eingegangen. Anschließend werden weitere theoretische Aspekte zur Nutzung von Ergebnisrückmeldungen mit dem Fokus auf Lehrende berichtet. Die Erkenntnisse aus weiteren Forschungsbereichen ergänzen die Befunde der Rezeptionsforschung und zum Zyklenmodell der Rückmeldung (z.B. Helmke & Hosenfeld, 2005). Weiterhin werden Möglichkeiten von und Anforderungen an elektronische Ergebnisrückmeldungen zusammengefasst.

Rückmeldung vs. Feedback. Die Begriffe Rückmeldung und Feedback werden häufig synonym verwendet (Rolf, 2002). Feine Unterscheidungen hinsichtlich ihrer Definitionen sind jedoch festzustellen. Schneewind (2007) unterscheidet die Begriffe wie folgt: Das klassisch soziale Feedback besteht aus einer leistenden und einer prüfenden bzw. rückmeldenden Person (vgl. auch Hattie & Timperley, 2007; Müller & Ditton 2014). Im Gegensatz hierzu wird die Rückmeldung z.B. aus Schulleistungsstudien ausgeweitet auf an der Rückmeldesituation bedingt beteiligte Personen. Hierzu zählen Personen aus Auswertungsinstituten oder aus der Schulaufsicht, die den Lehrkräften die Ergebnisrückmeldungen zur Verfügung stellen. Die leistenden Personen bzw. Lernenden werden in den Rückmeldekreislauf häufig nicht direkt einbezogen (z.B. bei der Unterrichtsentwicklung) oder es tritt die Lehrkraft als übermittelnde Person auf (z.B. individuelle Förderung) (Schneewind, 2007). Müller und Ditton (2014) ziehen den Vergleich, indem sie feststellen, dass Feedback und Rückmeldung keine eindeutig getrennten Konzepte darstellen. Die Differenziertheit der übermittelten

Informationen kann zwar ein mögliches Kriterium der Unterscheidung der beiden Begriffe bieten, Feedback ist in der Regel detailgenauer, trotzdem ist der Übergang fließend. Als Fazit kann festgehalten werden: Die Überschneidungen sind insgesamt so umfassend, dass eine gemeinsame Verwendung der Erkenntnisse der Rückmeldungs- und Feedbackforschung sinnvoll erscheint (Müller & Ditton, 2014).

Nutzung von Ergebnisrückmeldungen mit dem Fokus auf Lernende. Zur Nutzung von Ergebnisrückmeldungen mit direktem Bezug auf Lernende werden die Metaanalyse von Hattie und Timperley (2007) und die ergänzende multidimensionale Konzeptualisierung von Feedback durch Hattie und Wollenschläger (2014) aufgegriffen. Das Anliegen von Feedback ist es, die Diskrepanz zwischen der aktuellen Leistung und der zu erreichenden Leistung zu verringern (Hattie & Wollenschläger, 2014). Die Autoren greifen in ihren Erläuterungen auf unterschiedliche theoretische Ansätze der Feedbackforschung zurück, die hier erwähnt, jedoch nicht weiter ausgeführt werden sollen (vgl. Narciss, 2014). Hierzu gehört das kybernetisch orientierte Response-Certainty-Modell (Kulhavy & Stock, 1989), das Mindful-Processing-Modell zur individuellen Feedbackverarbeitung (Bangert-Drowns, Kulik, Kulik & Morgan, 1991), das Feedback Modell für selbstreguliertes Lernen (Butler & Winne, 1995), die Feedback-Intervention-Theory zur Aufmerksamkeitsregulation (Kluger & DeNisi, 1996) sowie das multidimensionale Interactive-Two-Feedback-Loops-Model (Narciss, 2006; 2008).

Das komplexe Modell der multidimensionalen Konzeptualisierung von Feedback wird in Abbildung 3 dargestellt. Hattie und Wollenschläger (2014) nennen zum externen Feedback bzw. zu den externen Rückmeldung drei Einflusskriterien: Die Feedbackquelle, die Art der Feedbackdarstellung und der Feedbackinhalt. Diese Einflusskriterien des externen Feedbacks wirken auf die im Modell dargestellte aufgabenbezogene und prozessbezogene Ebene (Task Level und Process Level). Der metakognitiven Ebene wird eher eine mediatorische oder moderierende Wirkung zugeschrieben (Meta-Task Level; Hattie & Wollenschläger, 2014). Die drei Ebenen werden im Konzept der Feedbackverarbeitung als „Task Level“, „Process Level“ und „Meta-Task Level“ dargestellt (Hattie & Wollenschläger, 2014, S. 140). Um die Diskrepanz der aktuellen und anvisierten Lernziele zu verringern, sollten drei Hauptfragen verfolgt werden: „Where am I going? (What are the goals?), How am I going? (What process is being made toward the goal?), and Where to next? (What activities

need to be undertaken to make better progress?)“ (Hattie & Timperley, 2007, S. 86). Diese drei bzw. sechs Fragen strukturieren wiederum die drei Feedback-Ebenen.

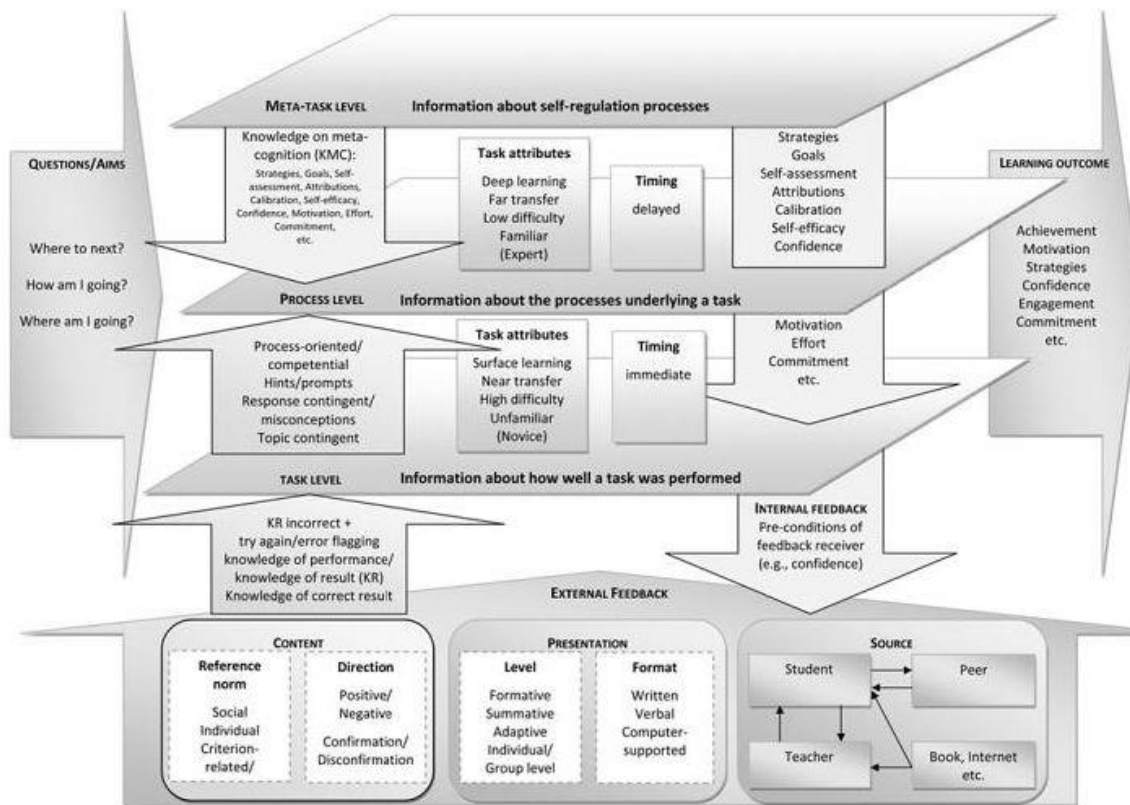


Abbildung 3. Multidimensionale Konzeptualisierung von Feedback (Hattie & Wollenschläger, 2014)

Die Nutzung von Ergebnissrückmeldungen ist neben den Bedingungen des Lernenden abhängig von weiteren drei Bedingungen: vom Feedback selbst, von der rückmeldenden Person und von weiteren Kontextbedingungen (Hattie & Timperley, 2007). Für die Nutzung von Ergebnissrückmeldungen ist neben den anderen genannten Bedingungen die Rolle der rückmeldenden Person zentral. Die rückmeldende Person muss die Ergebnissrückmeldungen interpretieren und Entscheidungen darüber treffen, wie eine Förderung umgesetzt werden kann. Ist zum Beispiel bei der leistenden Person kein oder wenig Wissen über das zu erreichende Lernziel vorhanden, hilft dieser eine einführende Instruktion möglicherweise mehr als ein Feedback (Hattie & Timperley, 2007).

Zudem sind Aushandlungsprozesse zwischen der rückmeldenden und der leistenden Person notwendig. Entscheidet die rückmeldende Person allein über die Fragen nach dem Ziel bzw. darüber, wie dieses erreicht werden kann und welches der nächste Schritt ist, so

werden motivationale Aspekte der leistenden Person, z.B. deren Selbstbestimmung, übergangen (Ryan & Deci, 2000, vgl. auch Gagné & Deci, 2014). Die Aufgabe der rückmeldenden Person kann somit als verantwortungsvoll und komplex angesehen werden. Erschwerend kann hinzukommen, dass die leistende Person nicht zwingend dasselbe unter dem gegebenen Feedback versteht wie die rückmeldende Person. Die konstruktivistische Theorie geht von einer eigenaktiven und konstruktiven Auseinandersetzung der leistenden Person mit dem gegebenen Feedback aus (Kopp & Mandl, 2014). Demnach ist Feedback kein einseitiger Vorgang, denn die leistende Person steuert den Feedbackprozess mit (Hattie & Timperley, 2007).

Die Geschwindigkeit und Präzision, mit der eine rückmeldende Person über ein Feedback z.B. im Unterrichtsgeschehen entscheiden muss, zeigt erneut die Komplexität der Aufgaben einer rückmeldenden Person. Vor allem auf der Prozessebene des Modells der multidimensionalen Konzeptualisierung muss die Entscheidung über ein Feedback innerhalb von Sekunden getroffen werden, und dabei sollten sämtliche in der Abbildung 3 dargestellten Kriterien eines Feedbacks berücksichtigt werden (Hattie & Timperley, 2007; Hattie & Wollenschläger, 2014). Nicht nur die leistende Person ist damit in der Rolle des Lernenden, auch die rückmeldende Person muss zum einen aus den Kriterien des Feedbacks lernen und erhält zum anderen durch das Feedback Informationen über die leistende Person (Hattie & Timperley, 2007).

Nutzung von Ergebnisrückmeldungen mit dem Fokus auf Lehrkräfte. Zusätzlich zu den Prozesselementen des im vorhergehenden Kapitel 2.3 beschriebenen theoretischen Zyklusmodells der Rückmeldung zur Nutzung von Ergebnisrückmeldungen werden in diesem Abschnitt Befunde aus weiteren Forschungsfeldern herangezogen. Zunächst wird der Begriff der Nutzung im Sinne der Verwendungsforschung definiert. Die Perspektive auf die Nutzung von Ergebnisrückmeldungen der Wissenschaftsrezeption und Verwendungsforschung ergänzen die Befunde der Rezeptionsforschung und des theoretischen Zyklusmodells der Rückmeldung. Diese werden im Anschluss an die Definition des Nutzungsbegriffs dargestellt. Abschließend werden Anforderungen an Lehrkräfte hinsichtlich der Nutzung von Ergebnisrückmeldungen herausgearbeitet.

Definition eines traditionellen Nutzungsbegriffes. In der traditionellen Verwendung des Begriffs Nutzung wird darunter die Anwendung oder Verwendung von Informationen

verstanden. Ein erweitertes Nutzungsverständnis bezieht sich auf ein breites Feld der Gestaltung von Informationen, der Rezeption, der Umsetzung, der Verarbeitung, der Integration und der Implementation (Stamm, 2003). Diesem Verständnis entsprechend definiert Stamm (2003):

[Die] Nutzung als bewusste oder unbewusste, lang- oder kurzfristige Rückmeldung, Verbreitung und Weiterverarbeitung von Evaluationsinformationen in Form von Produkten oder Ergebnissen zur Erreichung eines Ziels oder eines veränderten Zustands (...) [, die] von den Akteuren und Verantwortlichen in der Praxis als solche dargestellt und sichtbar gemacht wird (S. 39).

Perspektive auf die Nutzung von Ergebnisrückmeldungen der Wissenschaftsrezeption und Verwendungsforschung. Die Nutzung von Ergebnisrückmeldungen aus externen Testverfahren gleicht der Wissenschaftsrezeption, der das Ziel der begründeten Entscheidungsfindung zugrunde liegt (Bromme & Kienhues, 2012). Es geht nicht darum, aus Lehrkräften Experten z.B. von Forschungsmethoden zu machen, sondern um eine evidenzbasierte Orientierungs- und Entscheidungsgrundlage (Bromme & Kienhues, 2012). Die Übermittlung der nutzbaren Informationen ist abhängig von der individuellen Person bzw. dem Rezipienten. Bromme und Kienhues (2012) geben als beeinflussende Personenvariablen die Einstellung der Person zur jeweiligen Sache, die epistemischen Überzeugungen und weitere Personeneigenschaften wie z.B. Ambiguitätstoleranz an. In Anlehnung an die Sozialpsychologie kann angenommen werden, dass die Einstellung auf drei Komponenten beruht: Auf kognitiven, auf affektiven und auf verhaltensbezogenen Informationen (Haddock & Maio 2014; Zanna & Rempel, 1988). Kognitionen bezeichnen in diesem Sinne Gedanken und Überzeugungen, affektive Informationen bezeichnen Gefühle, und verhaltensbezogene Informationen beziehen sich auf Verhaltensweisen gegenüber einem Einstellungsgegenstand (Haddock & Maio, 2014). Diese drei Einstellungsmerkmale sind für zu treffende Entscheidungen relevant.

Auf der Grundlage der Verwendungsforschung bezieht sich Stamm (2003) auf drei Nutzungsvarianten von Informationen: Die direkte bzw. instrumentelle, die konzeptuelle sowie die selektive bzw. symbolische Nutzung. Die auf Rossi und Freeman (z. B. Rossi, Lipsey & Freeman, 2004) zurückgehenden Varianten der Nutzung bezeichnen mit dem Begriff der direkten bzw. instrumentellen Nutzung die konkreten Handlungsentscheidungen (Maier, 2008a). In der konzeptuellen Nutzung wird die Rückmeldung bewertet, sie beeinflusst somit

das Denken des Nutzers und darauffolgende Entscheidungen. Die selektive bzw. symbolische Nutzung ist eher eine Legitimation, um den eigenen, bereits festgelegten Standpunkt zu stützen (Maier, 2008a).

Die durch die Wissenschaftsrezeption und Verwendungsforschung dargelegten Hintergründe bestätigen, was zum Zyklenmodell der Rückmeldung bereits gesagt wurde: „(...) der Weg von der Information (...) bis zur Innovation (...) [ist] weit und beschwerlich“ (Helmke & Hosenfeld, 2005, S. 147). Die Ergebnisrückmeldungen müssen z.B. nicht zu konkreten Handlungsentscheidungen führen, sondern können lediglich als Legitimation der bisherigen Handlungen verwendet werden (Verwendungsforschung). Dabei würden die Ergebnisrückmeldungen keine Veränderungen bewirken. Ebenso können z.B. affektive Informationen aus Ergebnisrückmeldungen für die Lehrkräfte negativ belegt sein und dadurch die Nutzung der Ergebnisrückmeldungen beeinträchtigen (Wissenschaftsrezeption). Die Befunde zeigen auf, dass die Nutzung von Ergebnisrückmeldungen von unterschiedlichen Faktoren abhängig ist und je nach Rezipient individuell verläuft.

Anforderungen an Lehrkräfte hinsichtlich der Nutzung von Ergebnisrückmeldungen.

Die Anforderungen an Lehrkräfte beziehen sich auf das Verstehen der zum Teil elektronisch zurückgemeldeten Ergebnisse. Hierbei ist zwischen technischen, statistischen und inhaltlichen Rückmeldungen zu unterscheiden (Stauke, 2008). Die Lehrkräfte müssen erstens mit den technischen Mitteln umgehen können, d. h. die Hard- und Software nutzen können. Zweitens müssen sie statistische Kennwerte entschlüsseln und Grafiken interpretieren können. Drittens müssen sie sich inhaltlich und didaktisch in den zum Teil spezifischen Defiziten oder Fortschritten von Schülerinnen und Schülern auskennen (Stauke, 2008). Die Anforderungen an den Nutzer sind komplex. Aus diesem Grund sollten Lehrkräfte Unterstützung z.B. in Form von Handlungswissen erhalten (Rolff, 2002) bzw. sollten die Ergebnisrückmeldungen an die Kompetenzen der Empfänger angepasst sein. Nicht nur das Verstehen von (elektronischen) Ergebnisrückmeldungen, auch die Nutzung der Ergebnisrückmeldungen hinsichtlich zu treffender Entscheidungen ist eine Anforderung an die Lehrkräfte. Vor allem ungeübte Lehrkräfte benötigen Fort- und Weiterbildungen hinsichtlich der evidenzbasierten Entscheidungsfindung (z.B. Breiter & Light, 2006, Zimmer-Müller, Hosenfeld & Koch, 2014). In diesen Schulungen sollten Lehrkräfte zur Nutzung von Ergebnisrückmeldungen in Form von Materialbeispielen und neuen Strategien unterstützt werden (Breiter & Light,

2006). Im Speziellen bei elektronischen Ergebnisrückmeldungen ist ein entdeckender Umgang mit computer- bzw. internetbasierten Rückmeldungen hilfreich (Breiter & Stauke, 2007). Das Analyse- und Auswertungspotenzial der Informationstechnologie ist hierfür längst nicht ausgeschöpft (Breiter & Stauke, 2007) und kann bei wachsenden Kompetenzen der Nutzer ausgeweitet werden.²⁴

Zu den Kompetenzen der Nutzer von Ergebnisrückmeldungen hinsichtlich graphischer Darstellungen soll hier ein kurzer Exkurs in das Review „Graph Comprehension Research“ (S. 47) bei Schülerinnen und Schülern von Shah und Hoeffner (2002) unternommen werden. Die Befunde können auf weitere Nutzer von Grafiken übertragen werden. Ob und wie eine Grafik verstanden wird, ist abhängig von der Grafik selbst und den Kenntnissen des Nutzers. Die Kenntnisse des Nutzers können in statistische und inhaltliche Kenntnisse unterschieden werden. Shah und Hoeffner (2002) fassen in ihrem Review drei Hauptkomponenten zusammen, die für das Verständnis von Grafiken notwendig sind: 1. Die Nutzer müssen die Anordnungen und wichtigen Eigenschaften der Grafiken identifizieren können, 2. Die visuellen Bestandteile der Grafik müssen mit den inhaltlichen Aussagen in Beziehung gesetzt werden, 3. Die Nutzer müssen die Werte der Grafik nachvollziehen und mit den Ausgangswerten in Beziehung setzen. Die Charakteristik der Grafik sowie das Wissen des Nutzers über Grafiken und deren Inhalt spielen beim Verstehen einer Grafik eine wichtige Rolle. Außerdem wird die Interpretation von Grafiken beeinflusst von den Erwartungen, die ein Nutzer an deren Inhalte mitbringt. Shah und Hoeffner (2002) geben im Rahmen ihres Reviews relevante Empfehlungen, die bei der Erstellung einer Grafik berücksichtigt werden sollten. Die Darstellungsformen einer Grafik sollten an die Ausgangswerte angepasst werden, z. B. werden Liniendiagramme üblicherweise für Verläufe oder Zeitreihen genutzt, Balkendiagramme für diskontinuierliche Werte und Kreisdiagramme für Prozentwerte. Sollen multiple Ausgangswerte dargestellt werden, ist es nützlich, dieselben Daten in unterschiedlichen Formaten abzubilden, z. B. können die quantifizierten Werte in einer Tabelle und die dazugehörenden Zeitreihen in einer Grafik aufgeführt werden. Animationen sollten mit Vorsicht und nur dann verwendet werden, wenn sie für die entsprechende Fragestellung einen zusätzlichen Beitrag leisten können. Die Anstrengung des Arbeitsgedächtnisses sollte für unnötige Kriterien gering gehalten werden, z. B. sollte die Farb- und Symbolauswahl sowie die Legende bzw. Beschriftung der Grafik mit Bedacht ausgewählt werden. Die Farbauswahl wird von den Autoren als wichtiger Punkt erwähnt, so sollten z. B. kontinuierliche Werte

²⁴ Siehe auch im nachfolgenden Abschnitt: Elektronische Ergebnisrückmeldungen.

nicht farblich kontrastiert werden. Dreidimensionale Darstellungen können ansprechend wirken, sollten jedoch überlegt gewählt werden. Ebenso sollten die Relationen und Größenverhältnisse der Grafiken bewusst genutzt werden, da hierdurch sehr leicht die Interpretation der Nutzer beeinflusst werden kann. Abschließend empfehlen die beiden Autoren, dass die Grafik und der Text konsistent gestaltet werden sollten. Vor allem für Nutzer, die im Umgang mit Grafiken ungeübt sind, ist dies unterstützend.

Der Abschnitt zeigt auf, dass die Anforderungen an Lehrkräfte unterschiedlich gelagert sind. Nicht nur Fachwissen ist für die Nutzung von Ergebnisrückmeldungen relevant, die Lehrkräfte müssen auch Leistungen wie z. B. den Umgang mit Computern, Grafiken und statistischen Begriffen erbringen. In diesem Zusammenhang zeigt sich wiederum die Komplexität der Nutzung von Ergebnisrückmeldungen.

Elektronische Ergebnisrückmeldungen. In den nachfolgenden Abschnitten wird auf die Entwicklung technischer Möglichkeiten der elektronischen Datenerhebung und im Speziellen der Ergebnisrückmeldung eingegangen. Weiterhin werden Anforderungen an elektronische Ergebnisrückmeldungen aus der Perspektive der empirischen Forschung festgehalten.

Entwicklungen technischer Möglichkeiten hinsichtlich elektronischer Ergebnisrückmeldungen. Auswirkungen der im Laufe des 20. und 21. Jahrhunderts zunehmend Raum greifenden digitalen Revolution treffen auch die Schule und das schulische Umfeld (z. B. Schulz-Zander, Eickelmann, Moser, Niesyto & Grell 2012). Nicht nur die Schulverwaltung und der Unterricht sind in die technischen Veränderungen eingeschlossen, auch eine evidenzbasierte Schul- und Unterrichtsentwicklung sowie eine evidenzbasierte individuelle Förderung von Schülerinnen und Schülern, die durch die Technisierung erst in diesem Umfang möglich wurde, wird zunehmend zum Status Quo (z. B. Altrichter & Maag Merki 2010; Kühle & van Ackeren, 2012; Wacker, Maier & Wiesinger, 2012). Der Begriff Evidenz meint hier die Überprüfung bzw. Begründung von Annahmen durch empirisch-wissenschaftliche Methoden (vgl. van Ackeren, Zlatkin-Troitschanskaia, Binnewies, Clausen, Dormann, Preisendörfer, Rosenbusch & Schmidt, 2011; Bromme, Prenzel & Jäger, 2014; Schrader, 2014). Im nachfolgenden Abschnitt wird der Einfluss der zunehmenden Einbindung technischer Möglichkeiten auf Ergebnisrückmeldungen erörtert.

Im Sinne der Betriebswirtschaftslehre und Wirtschaftsinformatik unterscheiden Breiter und Stauke (2007) bei der Informationssammlung, -verarbeitung und -kommunikation zwischen externen und internen Informations- und Kommunikationssystemen. Externe Informations- und Kommunikationssysteme betreffen im Rahmen der Schule alle Bereiche der Öffentlichkeit wie z. B. Internetauftritte. Interne Informations- und Kommunikationssysteme betreffen die Organisation innerhalb der Schule, wie z. B. die Schulverwaltung. Auch die im Rahmen der vorliegenden Studie interessierenden elektronischen Rückmeldungen werden den internen Informations- und Kommunikationssystemen zugeordnet (Breiter & Stauke, 2007). Die Informations- und Kommunikationssysteme sind ein Zusammenspiel aus Datenverarbeitungstechniken, Anwendungsbereichen und Nutzerverhalten (Gluchowski, Gabriel & Dittmar, 2008). Entwicklungen und Forschungsergebnisse aus dem Bereich der Wirtschaftsinformatik können für die evidenzbasierte Schul- und Unterrichtsentwicklung bzw. individuelle Förderung richtungsweisend sein. Als kurze Anmerkung soll auf die Erkenntnis der vergangenen Jahre hingewiesen werden, dass formatives Assessment zu einer besseren Leistungsentwicklung bei Schülerinnen und Schülern führt (Black & Wiliam, 1998; Sainsbury & Benton, 2011). Black und Wiliam (2009) definieren formatives Assessment wie folgt:

„Practice in a classroom is formative to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about the next steps in instruction that are likely to better, or better founded, than the decisions they would have taken in the absence of the evidence that was elicited.“ (S. 9)

Die elektronischen Ergebnisrückmeldungen können für die Optimierung von Lehr-Lernprozessen im Sinne eines formativen Assessments (Maier, 2010) besonders nützlich sein. Der Einsatz sich entwickelnder Technologien ermöglicht die Vernetzung von Datensätzen (data warehouse), einen internetgestützten Zugriff auf Daten und Ergebnisrückmeldungen (web-based interactive data report) oder die computergestützte Erhebung von Daten (electronic diagnostic assessment tool), durch die eine Dateneingabe überflüssig wird (Brunner & Light, 2008). Die Bereitstellung technischer Verfahren zur individuellen Ergebnisrückmeldung sowie Daten zu Vergleichszwecken auf internationaler, nationaler und lokaler Ebene bieten für die schulische Bildung neue Möglichkeiten (Breiter & Stauke, 2007). Die Integration der Informationssammlung, -verarbeitung und -kommunikation in einen laufenden Prozess der

Schul- und Unterrichtsentwicklung sowie der individuellen Förderung ist vereinzelt bereits Realität. Dabei ist das Zusammenwirken der drei Komponenten Datenverarbeitungstechniken, Anwendungsbereiche und Nutzer, wie die Wirtschaftsinformatik aufgedeckt hat, unabdingbar (Gluchowski, Gabriel & Dittmar, 2008). Fachkräfte entwickeln Datenverarbeitungstechniken, um Aufgaben in den Anwendungsbereichen, wie z.B. den Umgang mit Leistungsunterschieden in einer Klasse, zu lösen (Brunner & Light, 2006). Die Endnutzer, die das System anwenden und bedienen, müssen in den Entwicklungsprozess einbezogen werden (Gluchowski, Gabriel & Dittmar, 2008). Die Definition der Anforderungen der Datenverarbeitungstechniken sowie die Konzeption einer Benutzeroberfläche sollten vom Endnutzer mitgestaltet werden. Dies erhöht letztendlich auch die Akzeptanz einer Nutzung von Ergebnisrückmeldungen (Gluchowski, Gabriel & Dittmar, 2008; Breiter & Stauke, 2007).

Anforderungen an elektronische Ergebnisrückmeldungen. Die Anforderungen an elektronische Ergebnisrückmeldungen der operativen Ebene werden in diesem Abschnitt zusammengefasst. Bei elektronischen Ergebnisrückmeldungen kann zunächst zwischen drei Möglichkeiten unterschieden werden: statische Rückmeldungen wie z.B. unveränderbare PDF-Dateien²⁵, computerbasierte, offline übermittelte Rückmeldungen (z.B. durch eine Daten-CD) und internetbasierte Rückmeldungen (Stauke, 2008). Die computer- bzw. internetbasierten Rückmeldungen leisten einen signifikanten Beitrag zur Organisation von Ergebnisrückmeldungen und tragen damit zur Erfüllung von Anforderungen auf der operativen Ebene bei (Visscher, 2008).

Ein entscheidendes Kriterium von Ergebnisrückmeldungen ist ihre Aktualität. Elektronische Ergebnisrückmeldungen auf operativer Ebene sollten hochaktuell sein und deshalb zeitnah angeboten werden (Breiter & Light, 2006). Die Informationen aus den elektronischen Ergebnisrückmeldungen sollten möglichst differenziert, handlungsleitend und klar verständlich sein, um Entscheidungen in spezifischen Bereichen unterstützen zu können (Breiter & Stauke, 2007; Rolff, 2002). Das bedeutet auch, dass trotz der technischen Möglichkeit, viele Informationen komprimiert zu übermitteln, die Ergebnisrückmeldungen ausgewählte Informationen enthalten sollten, um eine Überforderung der Lehrkräfte durch einen Überfluss an Informationen zu vermeiden (Brunner & Light, 2006; vgl. auch Helmke, 2004; Rolff, 2002). Um die Navigation innerhalb der Ergebnisrückmeldungen zu erleichtern, bietet sich eine Strukturierung an, die sich am Aufbau der Tests orientiert (Stauke, 2008).

²⁵ Diese können auch den papierbasierten Rückmeldungen zugeordnet werden.

Ein Wiedererkennungseffekt ist demnach nützlich. Generell sollten die Ergebnismeldungen benutzerfreundlich gestaltet sein. Unter anderem wirkt eine Visualisierung der Ergebnisse beim Erfassen der Informationen unterstützend (Breiter & Stauke, 2007). Bedenken von Lehrkräften hinsichtlich des Datenschutzes kann dadurch begegnet werden, dass die Dateneingabe ohne direkten Personenbezug erfolgt, dass also z.B. die Schülerinnen und Schüler die Tests unter einem Pseudonym ausfüllen und die Zuordnung zur einzelnen Schülerin bzw. zum einzelnen Schüler in der Schule verbleibt (Stauke, 2008).

Die Testentwicklung und die sich darauf beziehende Ergebnismeldung können mit den nationalen Bildungsstandards abgestimmt werden. In diesem Zusammenhang müssen weitere Anforderungen an elektronische Ergebnismeldungen im Sinne von Vergleichswerten zur Einordnung der Ergebnisse berücksichtigt werden. Elektronische Ergebnismeldungen können aufgrund der Interaktionsmöglichkeiten der Informations- und Kommunikationssysteme hinsichtlich der Bezugsnormorientierung (Rheinberg & Fries, 2010) und der fairen Vergleiche (Fiege, 2013) bisher ungenutzte Möglichkeiten bieten (Breiter & Stauke, 2007). Zum Beispiel könnten durch die Schule erhobene Daten zum sozioökonomischen Status mit Ergebnissen aus Leistungstest verknüpft werden. Ebenso wäre es möglich, dass Lehrkräfte internetbasiert je nach Erkenntnisinteresse unterschiedliche Vergleichswerte in Beziehung zu den Leistungsergebnissen einer Schülerin oder eines Schülers setzen.

Rolff (2002) bringt in die Diskussion zur Nutzung von Ergebnismeldungen die Anforderung ein, dass die wissenschaftlichen Erkenntnisse in Handlungswissen übersetzt werden müssen. Die Wirksamkeit von Ergebnismeldungen ist zwar stark abhängig von der Qualität der Ergebnisse und der Qualität der Ergebnismeldung, aber erst die Übersetzung von Ergebnismeldungen in einen Handlungskontext, um neues Handlungswissen zu generieren, vervollständigt den Prozess. Das setzt hohe Anforderungen an die elektronischen Ergebnismeldungen voraus. Diese müssen so gestaltet sein, dass die Notwendigkeit von Verbesserungsmaßnahmen erkannt und zielführende Wege zur Verbesserung aufgezeigt werden. Handlungswissen über Verbesserungen auf operativer Ebene bezieht sich auf didaktisches Wissen (Rolff, 2002). Wie die Erkenntnisse aus der Wirtschaftsinformatik aufzeigen, wird auch hier die Notwendigkeit der Zusammenarbeit von Fachkräften der Datenverarbeitungstechnik, Fachkräften der empirischen Bildungsforschung und Fachkräften der Methodik, Didaktik und des Fachwissens deutlich (Gluchowski, Gabriel & Dittmar, 2008).

2.5 Zusammenfassung der Kapitel zum theoretischen Hintergrund

Im Rahmen von Beschlüssen der Kultusministerkonferenz der Länder wurde die regelmäßige Teilnahme an internationalen Vergleichsstudien, die Durchführung von Vergleichsarbeiten und die gemeinsame Bildungsberichterstattung von Bund und Ländern beschlossen. Außerdem wurden Kriterien entwickelt, die eine Vergleichbarkeit zentraler Kompetenzbereiche ermöglichen sollen: die Bildungsstandards. Neben den Verfahren, die sich hauptsächlich auf die Leistungserfassung von Schülerinnen und Schülern beziehen, soll auch die Qualifizierung von Lehrkräften mittels Aus- und Fortbildung vorangebracht werden.

Lehrkräfte können durch objektivierte Informationen, z.B. aus Vergleichsarbeiten, bei der Diagnostik von Schülerleistungen unterstützt werden. Vor allem erscheint die Unterstützung dann hilfreich, wenn die Aus- und Fortbildung der Lehrkräfte hinsichtlich diagnostischer Kompetenzen noch nicht greift. Empirische Studien zeigten allerdings auf, dass der Weg einer Nutzung von Ergebnisrückmeldungen von der Information bis zur Innovation als weit und beschwerlich angesehen werden kann (Helmke & Hosenfeld, 2005). Das Zyklusmodell der Rückmeldung strukturiert als theoretische Grundlage den Prozess der Nutzung von Ergebnisrückmeldungen durch die Prozessschritte Rezeption, Reflexion, Aktion und Evaluation (z.B. Helmke & Hosenfeld, 2005). Zunächst wird im Rahmen des Prozessschrittes Rezeption die Verständlichkeit der Ergebnisdarstellungen erfasst. Im Prozessschritt Reflexion reflektieren die Lehrkräfte den Informationsgewinn aus den Ergebnisrückmeldungen. Anschließend werden im Prozessschritt Aktion Maßnahmen als Konsequenzen aus den Ergebnisrückmeldungen aufgegriffen und umgesetzt. Der Prozessschritt Evaluation dient dem reflektierten Rückblick auf den Nutzungsprozess. Hier soll die Wirkung der eingesetzten Maßnahmen bzw. sollen die Konsequenzen aus dem Prozessweg überprüft werden. Außerdem greift das Zyklusmodell der Rückmeldung Faktoren auf, die den Prozess der Nutzung beeinflussen, wie etwa individuelle, schulische und externe Bedingungen. Vor allem die Einstellung und die Akzeptanz von Ergebnisrückmeldungen durch die Lehrkräfte und das schulische Umfeld sind in diesem Zusammenhang relevant.

Die Nutzung von Ergebnisrückmeldungen ist abhängig von den zugrundeliegenden Daten und Zielen. Soll eine individuelle und gezielte Förderung von Schülerinnen und Schülern stattfinden, müssen entsprechend detaillierte und den interessierenden Kompetenzbereich betreffende Informationen zurückgemeldet werden. Die der Nutzung von Ergebnisrückmeldungen vorangehenden Rückmeldungen bzw. Feedbacks gehen auf unterschiedliche Definitionen zurück. Im Vergleich der Konzepte von Rückmeldung und Feedback werden

feine Unterschiede deutlich, dennoch wird für eine gemeinsame Verwendung der Erkenntnisse aus der jeweiligen Forschung plädiert.

Das direkte Feedback einer rückmeldenden Person an eine leistende Person wird anhand des auf der Metaanalyse von Hattie und Timperley (2007) basierenden Modells der multidimensionalen Konzeptualisierung von Feedback (Hattie & Wollenschläger, 2014) erfasst. Das Modell ist als theoretischer Hintergrund für die Nutzung von Ergebnisrückmeldungen relevant, da es unter anderem die Komplexität von Rückmeldungen bzw. Feedback aufzeigt. Die häufigste Form der Rückmeldung aus z.B. externen Evaluationen geschieht über die Lehrkraft. Deshalb erscheint neben der Zusammensetzung der Ergebnisse, den Zielen der Datennutzung und den Rückmeldeprozessen der Fokus auf die Lehrkräfte besonders bedeutsam. Lehrkräfte erhalten Ergebnisse zu Schülerleistungen oder werten die Leistungstests selbst aus. Sie stellen bildlich gesprochen den „Trichter“ dar, durch den die Ergebnisrückmeldungen zu den Schülerinnen und Schülern gelangen. Sie strukturieren die jeweiligen Rückmeldeprozesse. Aus diesem Grund ist die Fokussierung der Erhebung auf die Lehrkräfte für die vorliegende Arbeit von besonderer Relevanz.

Zur Nutzung von Ergebnisrückmeldungen werden in der Forschung der Wissenschaftsrezeption kognitive, affektive und verhaltensbezogene Komponenten hinsichtlich der Einstellung der nutzenden Person als bedeutend für eine begründete Entscheidungsfindung benannt. Ebenso werden in der Verwendungsforschung für die Nutzung von Ergebnisrückmeldungen drei bedeutsame Komponenten beschrieben: a) die direkte bzw. instrumentelle Nutzung, die sich auf konkrete Handlungsentscheidungen bezieht, b) die konzeptuelle Nutzung, die das Denken des Nutzers beeinflusst, und c) die selektive bzw. symbolische Nutzung, die eher als Legitimation des bestehenden Standpunkts dient. Aus der Wissenschaftsrezeption und Verwendungsforschung lässt sich ableiten, dass die Nutzung von Ergebnisrückmeldungen auf folgende Weise ablaufen kann: Die Rückmeldungen werden in konkrete Handlungen bzw. Aktivitäten umgesetzt, die Rückmeldungen treten in Form einer Verhaltens- oder Einstellungsänderung der nutzenden Person auf oder die Rückmeldungen rufen keine Veränderungen hervor, sondern werden allenfalls als Bestätigung des eigenen Standpunktes aufgefasst. Neben der Art und Weise, wie Ergebnisrückmeldungen genutzt werden können, werden bei der Nutzung von Ergebnisrückmeldungen Anforderungen an die Lehrkräfte gestellt. Elektronisch zurückgemeldete Ergebnisse erfordern ein technisches, statistisches und inhaltliches Verständnis der Rückmeldungen. Die technische Rückmeldung meint

im Rahmen von elektronisch übermittelten Ergebnissen den Umgang mit der Hard- und Software und die Navigation innerhalb der Ergebnisrückmeldungen. Die statistischen Rückmeldungen beziehen sich auf die Darstellungsweise und die Begriffe der Ergebnisrückmeldungen. Die inhaltlichen Rückmeldungen beziehen sich auf das jeweilige Fachwissen und die spezifischen Defizite der Schülerinnen und Schüler.

3 Das Testverfahren *Online Diagnose*

Im vorliegenden Abschnitt wird das Testverfahren *Online Diagnose* vorgestellt. Zunächst werden die Ziele und der Ablauf des Testverfahrens berichtet. Im Anschluss daran wird auf die technischen Voraussetzungen eingegangen, die zur Durchführung des Testverfahrens erfüllt sein sollten. Weiterhin wird ein Vergleich der Aufgabenbereiche aus dem Testverfahren *Online Diagnose* mit den Bildungsstandards vorgenommen. Abschließend werden die Ergebnismeldungen aus dem Testverfahren *Online Diagnose* dargestellt und in Bezug zu den Anforderungen elektronischer Ergebnismeldungen gesetzt.

3.1 Ziele und Ablauf des Testverfahrens *Online Diagnose*

Das durch die Bildungshaus Schulbuchverlage Westermann, Schroedel, Diesterweg, Schöningh, Winklers GmbH²⁶ bereitgestellte Testverfahren *Online Diagnose* hat zum Ziel, den Lehrkräften das Erkennen heterogener Lernstände von Schülerinnen und Schülern zu ermöglichen bzw. zu erleichtern. Des Weiteren hat das Testverfahren zum Ziel, Maßnahmen zur Förderung von Stärken und zur Beseitigung von Schwächen einzuführen (Bildungshaus, 2008; 2013). Im Unterrichtsfach Deutsch²⁷ werden Stärken und Schwächen jeder Schülerin und jedes Schülers hinsichtlich der Fertigkeiten im Bereich Leseverständnis, Lesegeschwindigkeit, Hörverständnis, Grammatik und Rechtschreibung identifiziert, um auf der Basis dieser Diagnose individuelle Fördermaterialien ausgeben zu können.

In Baden-Württemberg wurden ab dem Schuljahr 2009/2010 in allen Eingangsklassen der Haupt- und Werkrealschulen die Deutsch- und Mathematikkenntnisse der Schülerinnen und Schüler anhand des Testverfahrens *Online Diagnose* internetgestützt diagnostiziert (Kultusportal, Onlinebasierte Lernstandserhebung). Bereits zu Beginn des Schuljahres 2008/2009 setzten mehr als zwei Drittel der baden-württembergischen Hauptschulen das Testverfahren *Online Diagnose* ein. Die landesweite Lizenz endete mit dem Schuljahr 2012/2013. Die Durchführung des Haupttests war für die ersten Monate der fünften Klasse angesetzt. Für die Bearbeitung des Haupttests wurden ca. 30 Minuten veranschlagt (Bil-

²⁶ Im Folgenden „Bildungshaus Schulbuchverlage“ genannt.

²⁷ Die vorliegende Studie verarbeitet Informationen von Deutschlehrkräften. Aus diesem Grund wird als Beispiel das Unterrichtsfach Deutsch herangezogen.

derungshaus, 2008; 2013). Zusätzlich wurde ein Detailtest eingesetzt, der gezielt weitere Aufgaben zu Bereichen stellte, für die im Haupttest ein individueller Förderbedarf ermittelt worden war (Bildungshaus, 2008; 2013). Am Ende des Schuljahres bzw. nach einer mehrmonatigen Förderphase stand ein Nachtest zur Verfügung, der die Veränderungen der Kompetenzen erfasste. Der Nachtest dauerte maximal 40 Minuten und enthielt Aufgaben des Haupt- und Detailtests (Bildungshaus, 2012). Die Auswertungen der Haupt-, Detail- und Nachtests konnten direkt nach der Testdurchführung des jeweiligen Tests online abgerufen werden. Passend zur Ergebnisrückmeldung der jeweiligen Haupt-, Detail- oder Nachtests konnten individuelle Fördermaterialien für jede Schülerin und jeden Schüler heruntergeladen und ausgedruckt werden. Die Fördermaterialien waren auf die in den Bildungsstandards und Lehrplänen definierten Kompetenzen zugeschnitten (Bildungshaus, 2008; 2013). Der Ablauf der Durchführung der Tests wird in Abbildung 4 veranschaulicht.



Abbildung 4. Ablauf und Testdurchführung des Testverfahrens *Online Diagnose*

3.2 Technische Voraussetzungen zur Durchführung des Testverfahrens *Online Diagnose*

Von den Bildungshaus Schulbuchverlagen wurde für jede Schule ein Zugang zum Testverfahren *Online Diagnose* angelegt (siehe Abbildung 5). Für diesen Zugang bekam die Schule

eine Benutzerkennung zugesandt (Schulbuchzentrum, n.d.). Durch den Login mit dieser Benutzerkennung konnten die entsprechenden Klassenkonten für die Lehrkräfte angelegt werden. Die Klassenkonten wurden durch ein persönliches Kennwort der jeweiligen Lehrkräfte gesichert (Bildungshaus, n.d.). Im letzten Schritt konnten die Lehrkräfte die Schülerinnen und Schüler ihrer Klasse innerhalb dieses Klassenkontos anlegen und die entsprechenden Tests auswählen. Außerdem mussten die Lehrkräfte ein Leistungsniveau festlegen, wovon die Zuweisung der Vergleichsgruppe und der Fördermaterialien abhing (Bildungshaus, n.d.). Zum Abschluss dieses Vorgangs erzeugten die Lehrkräfte für jede Schülerin und jeden Schüler einen Benutzernamen und ein Kennwort. Die Schülerinnen und Schüler waren damit individuell berechtigt und konnten die vorausgewählten Tests durchführen.

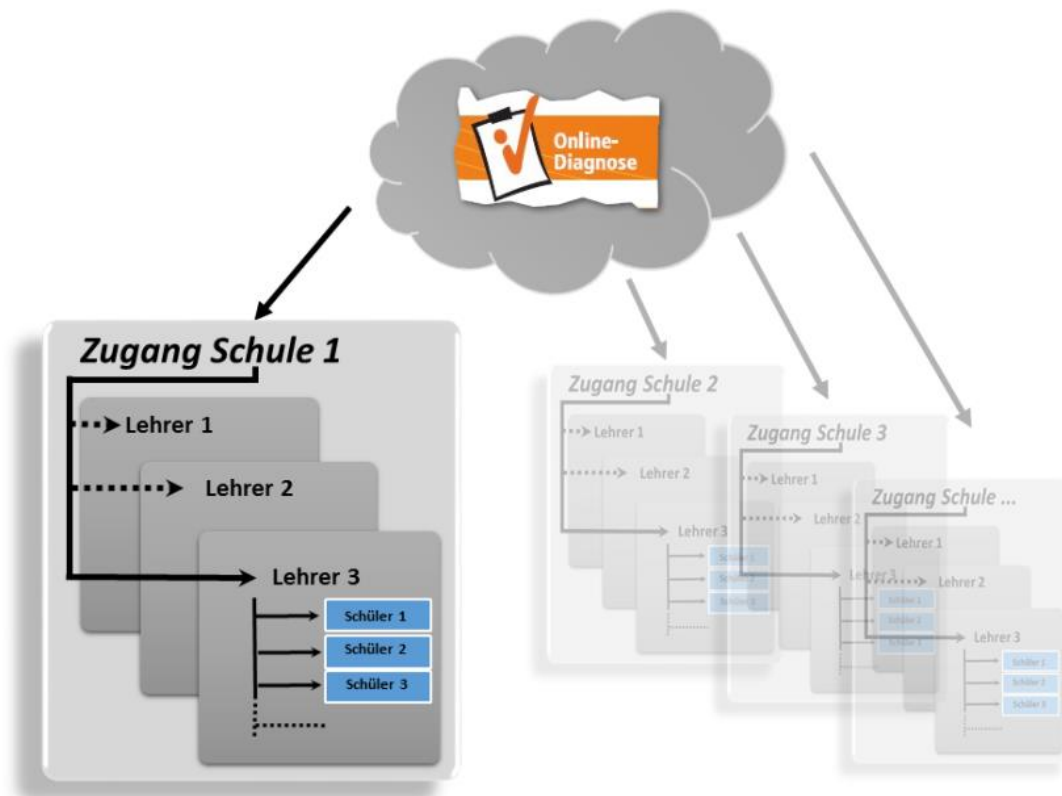


Abbildung 5. Zugang und Benutzerkennung zum Testverfahren *Online Diagnose*

Die jeweiligen Tests wurden von den Schülerinnen und Schülern am Computer ausgefüllt, somit konnten die Ergebnisse direkt auf einem Server abgespeichert werden (Bildungshaus, 2008; 2013). Pro Schülerin bzw. Schüler standen täglich 40 Minuten zur Testbearbeitung zur Verfügung. Überschritt die Bearbeitungsdauer diesen Zeitrahmen, wurde der Test der

einzelnen Schülerin bzw. des Schülers automatisch abgebrochen (Bildungshaus, 2012). Die Lehrkräfte erhielten nach Abschluss der Tests über ihr Klassenkonto die Ergebnisse für jede Schülerin und jeden Schüler. Die passenden Fördermaterialien bzw. Arbeitsblätter konnten für jede Schülerin und jeden Schüler heruntergeladen und ausgedruckt werden. Außerdem konnten die Arbeitsblätter auf bestimmte Kapitel aus Lehrbüchern des Verlagsverbundes Bildungshaus abgestimmt werden (Bildungshaus, 2012).

Zur Durchführung des internetgestützten Tests mussten die Computer bestimmte technische Systemvoraussetzungen erfüllen. Notwendig war ein installierter Adobe Flash Player Version 9 oder höher zum Abspielen der Tondateien. Außerdem musste sichergestellt sein, dass der Download von MP3-Dateien nicht z.B. durch die Firewall oder durch einen Router des Schulnetzwerks unterbunden werden konnte. Zusätzlich musste Java Script aktiviert sein und Session-Cookies mussten zugelassen werden. Eine Bildschirmauflösung von 980 Pixel Breite wurde zwar empfohlen, war aber nicht zwingend notwendig (Bildungshaus, Bedienungshinweise).

Die Bildungshaus Schulbuchverlage empfahlen die folgende Möglichkeit zur Wahrung des Datenschutzes bei der Online-Registrierung (Bildungshaus, 2010): Schülerinnen und Schüler, die für ein Internetangebot angemeldet wurden, mussten nicht mit vollständigem Namen angemeldet werden. Für die Anmeldung der Schülerinnen und Schüler genügte z.B. der Vorname, Spitzname oder ein Pseudonym. Für eine Anmeldung unter vollständigem Namen der jeweiligen Schülerinnen und Schüler wurde von den Bildungshaus Schulbuchverlagen eine Einwilligung der Eltern empfohlen. Die Daten zum Zwecke der Leistungserbringung durch die Bildungshaus Schulbuchverlage wurden nach Einwilligung der Schulen erhoben, gespeichert, verarbeitet und konnten jederzeit widerrufen werden. Zu Zwecken der Optimierung des Online-Angebots wurden laut den Datenschutzhinweisen die Antworten der Schülerinnen und Schüler nicht personenbezogen ausgewertet oder genutzt (Bildungshaus, 2010).

3.3 Aufgabenbereiche des Testverfahrens *Online Diagnose* im Vergleich mit den Bildungsstandards

In diesem Abschnitt werden die Aufgabenbereiche des Testverfahrens *Online Diagnose* den Bildungsstandards gegenübergestellt. Wie bereits in Abschnitt 2.1 aufgeführt, wurden die

Bildungsstandards zur Erfassung von Stärken und Schwächen in zentralen Kompetenzbereichen für den deutschlandweiten Einsatz entwickelt. Damit erscheinen sie als inhaltliche Orientierung für einen Vergleich mit dem Testverfahren *Online Diagnose* geeignet.²⁸ Die Bildungsstandards der Grundschule sind in die folgenden Kompetenzbereiche des Faches Deutsch gegliedert (Abbildung 6): Sprechen und Zuhören, Schreiben (Texte verfassen), Sprache und Sprachgebrauch untersuchen sowie Lesen – mit Texten und Medien umgehen (KMK, 2004a).

Das Testverfahren *Online Diagnose* ist in die folgenden Aufgabenbereiche unterteilt (Abbildung 7): Lesen und Hören sowie Rechtschreibung, Schreiben und Sprache. Nachfolgend werden die Aufgabenbereiche des Testverfahrens *Online Diagnose* mit den Bildungsstandards verglichen und mögliche Grenzen aufgezeigt. Die nun folgenden Abschnitte sind anhand der Bildungsstandards strukturiert.

²⁸ Alle Informationen sind den Bildungsstandards der Primarstufe des Faches Deutsch entnommen (KMK, 2004a). Die Aufgabenbereiche zum Testverfahren *Online Diagnose* sind den Auswertungen entnommen.

Sprechen und Zuhören

- Gespräche führen z. B. sich an Gesprächen beteiligen
- Zu anderen sprechen z. B. an der gesprochenen Standardsprache orientiert und artikuliert sprechen
- Verstehend zuhören z. B. Inhalte zuhörend verstehen
- Szenen spielen z. B. Perspektiven einnehmen
- Über Lernen sprechen z. B. Beobachtungen wiedergeben

Schreiben

- Über Schreibfertigkeiten verfügen z. B. eine gut lesbare Handschrift flüssig schreiben
- Richtig schreiben z. B. geübte, rechtschreibwichtige Wörter normgerecht schreiben
- Texte planen z. B. Schreibabsicht, Schreibsituation, Adressaten und Verwendungszusammenhang klären
- Texte schreiben z. B. nach Anregungen (Texte, Bilder, Musik) eigene Texte schreiben
- Texte überarbeiten z. B. Texte an der Schreibaufgabe überprüfen

Lesen – mit Texten und Medien umgehen

- Über Lesefähigkeit verfügen z. B. altersgemäße Texte sinnverstehend lesen
- Über Leseerfahrung verfügen z. B. verschiedene Sorten von Sach- und Gebrauchstexten kennen, sich in einer Bücherei orientieren
- Texte erschließen z. B. Verfahren zur ersten Orientierung über einen Text nutzen
- Texte präsentieren z. B. selbstgewählte Texte zum Vorlesen vorbereiten und sinngestaltend vorlesen

Sprache und Sprachgebrauch untersuchen

- Sprachliche Verständigung untersuchen z. B. Beziehung zwischen Absicht – sprachlichen Merkmalen – Wirkungen untersuchen
- An Wörtern, Sätzen, Texten arbeiten z. B. Wörter strukturierten und Möglichkeiten der Wortbildung kennen
- Gemeinsamkeiten und Unterschiede von Sprachen entdecken z. B. gebräuchliche Fremdwörter untersuchen
- Grundlegende sprachliche Strukturen und Begriffe kennen und verwenden z. B. Satzzeichen

Abbildung 6. Bildungsstandards mit Beispielen



Abbildung 7. Testinhalte Online Diagnose²⁹

²⁹ Die Vollständigkeit der Testinhalte kann nicht garantiert werden.

Bildungsstandards: Sprechen und Zuhören. Der Bereich der Bildungsstandards „Sprechen und Zuhören“ bezieht sich z.B. auf folgende Punkte „Gespräche führen“, „zu anderen sprechen“, „szenisch spielen“ und „über Lernen sprechen“. Unter Zuhilfenahme von Computern und Programmen sind diese Punkte nur begrenzt umsetzbar, deshalb kann das Testverfahren *Online Diagnose* diesen Bereich der Bildungsstandards nur bedingt erfüllen. Daraus kann geschlossen werden, dass nicht alle Bildungsstandards durch den Einsatz von Computern und Programmen abgedeckt werden können. Kreative Lösungen könnten hier allerdings die Umsetzung weiterer Bildungsstandards möglich machen wie z.B. in einem Klassenchat über Lernen sprechen. Den Bereich der Bildungsstandards „verstehend hören“ greift das Testverfahren *Online Diagnose* zum Teil durch die Aufgabenbereiche im Abschnitt „Hören“ auf (Abbildung 7). Nach einem systematischen Vergleich durch Jeuk (2013) entspricht das Testniveau im Aufgabenbereich „Hören“ des Testverfahrens *Online Diagnose* allerdings eher den Standards der zweiten Jahrgangsstufe der Grundschule, anstatt wie vorgesehen der vierten Jahrgangsstufe. Einen Nutzen könnte die Testung dadurch für mehrsprachige Schülerinnen und Schüler bieten (Jeuk, 2013). Hierfür ist das Testverfahren jedoch nicht explizit entwickelt und bedürfte einer spezifischen Überprüfung.

Bildungsstandards: Schreiben (Texte erfassen) und Sprache und Sprachgebrauch untersuchen. Die Bereiche der beiden Bildungsstandards „Schreiben (Texte verfassen)“ sowie „Sprache und Sprachgebrauch untersuchen“ (Abbildung 6) greift das Testverfahren *Online Diagnose* größtenteils durch den gesamten Aufgabenbereich „Rechtschreiben, Schreiben und Sprache“ auf (Abbildung 7). Die Punkte der Bildungsstandards „über Schreibfertigkeiten verfügen“, „richtig schreiben“, „Texte überarbeiten“, „an Wörtern, Sätzen, Texten arbeiten“ und „grundlegende sprachliche Strukturen und Begriffe kennen und verwenden“ werden durch das Testverfahren *Online Diagnose* mehr oder weniger ausführlich abgedeckt. Dies verdeutlicht vor allem die in Abbildung 7 dargestellte Auflistung zur Rechtschreibung, aber auch die Auflistungen zu Schreiben und Sprache bzw. Grammatik. Dennoch werden nicht alle Bereiche der Bildungsstandards abgedeckt. Die folgenden Bereiche der Bildungsstandards sind nicht enthalten: „Texte planen“, „Texte schreiben“, „sprachliche Verständigung untersuchen“ sowie „Gemeinsamkeiten und Unterschiede von Sprachen entdecken“ (Abbildung 6). Jeuk (2013) stellte im Vergleich der Bildungsstandards mit dem Aufgabenbereich Rechtschreiben des Testverfahrens *Online Diagnose* fest, dass die Testaufgaben ausführlich sind und nahezu alle Regelungsbereiche der Orthographie betreffen.

Für Schülerinnen und Schüler der 5. Klasse der Haupt- bzw. Werkrealschule sind die Anforderungen durch die Aufgabenbereiche jedoch als eher schwer einzuordnen (Jeuk, 2013). Die beiden Bereiche Schreiben und Grammatik sind in ihrer Anforderung wiederum eher dünn besetzt (Jeuk, 2013).

Bildungsstandards: Lesen – mit Texten und Medien umgehen. Im Rahmen der Bildungsstandards „Lesen – mit Texten und Medien umgehen“, können die beiden Punkte „über Leseerfahrung verfügen“ und „Texte präsentieren“ durch Computer und Programme nur begrenzt abgedeckt werden. Hierzu geforderte Kriterien wie z.B. Bibliotheken benutzen sind durch die Zuhilfenahme von Computern und Programmen nicht im implizierten Sinne möglich. Zu den beiden Punkten aus den Bildungsstandards „über Lesefähigkeit verfügen“ und „Texte erschließen“ greift das Testverfahren *Online Diagnose* Elemente zum Aufgabenbereich „Lesen“ auf (Abbildung 7). Im Vergleich der Bildungsstandards mit dem Testverfahren *Online Diagnose* stellte Jeuk (2013) fest, dass dieses allerdings nur einen Bruchteil der den Bildungsstandards zugeschriebenen Kriterien bzw. Umsetzungsmöglichkeiten erfasst. Auch scheint das Niveau wiederum eher der zweiten Jahrgangsstufe der Grundschule zuzuordnen zu sein (Jeuk, 2013).

3.4 Ergebnismrückmeldungen zum Testverfahren *Online Diagnose*

In diesem Abschnitt werden zunächst die Ergebnismrückmeldungen des Testverfahrens *Online Diagnose* dargestellt und beschrieben. Im Anschluss werden die Ergebnismrückmeldungen in Bezug zu Anforderungen an elektronische Ergebnismrückmeldungen gesetzt.

Darstellung der Ergebnismrückmeldungen zum Testverfahren *Online Diagnose*.

Die Ergebnisse werden in Grafiken z.B. Kreis- und Balkendiagramme und als Text dargestellt und können auf unterschiedlichen Aggregationsebenen gelesen werden (Bildungshaus, 2012). Zum Beispiel kann eine Zusammenfassung zum Leistungsprofil Lesen für eine Schülerin oder einen Schüler durch Auswahlfelder auf die jeweiligen Aufgabenblöcke selektiert werden (Bildungshaus, 2012). Die Leistungen der Schülerinnen und Schüler z.B. zu den individuellen Stärken- und Schwächen-Analysen werden anhand von Ampelfarben dargestellt. Die Farbe Rot bedeutet, es liegt ausdrücklich ein Förderbedarf vor, die Farbe Orange zeigt ebenfalls einen Förderbedarf an, und die Farbe Grün stellt dar, dass die Schülerinnen

und Schüler gefordert werden sollen. Für jede Schülerin und jeden Schüler können die Ergebnisse im Vergleich zu einer vorab festgelegten Vergleichsgruppe der jeweiligen Schulform oder im Vergleich zu vorhergehenden Testleistungen, z.B. Vergleiche zwischen Haupt- und Nachtest, eingesehen werden. Die schwarzen Markierungen (z.B. in Abbildung 8) zeigen die Vergleichswerte mit einer vorab festgelegten Vergleichsgruppe an. Über eine Navigationsleiste (z.B. Abbildung 8, links) kann zwischen einer Vergleichsgruppe aus Baden-Württemberg und einer bundesweiten Vergleichsgruppe gewechselt werden.

Der Vergleich zwischen Haupt- und Nachtest (Abbildung 9), also der Vergleich mit vorhergehenden Testleistungen, wird über einen Pfeil dargestellt, welcher in unterschiedlichen Winkelgraden entweder nach oben, nach rechts oder nach unten zeigt (Bildungshaus, 2012). Außerdem können die Ergebnisse aller Schülerinnen und Schüler einer Klasse anhand einer Ergebnis-Geschwindigkeits-Matrix betrachtet werden. Hier wird der Testerfolg mit der Arbeitsgeschwindigkeit einer Schülerin bzw. eines Schülers verglichen (Abbildung 10).

Wie in den Abbildungen 8 bis 10 deutlich wird, können über eine Navigationsleiste (links) die einzelnen Ergebnismeldungen angewählt werden. Im mittleren Teil der Benutzeroberfläche werden die Ergebnisse aufgeführt. Rechterhand ist der Quick-Tipp zu finden. Dieser gibt den Nutzern eine kurze Erläuterung, welche Ergebnismeldungen angezeigt werden, wie die schwarzen Markierungen zu verstehen sind und welche weiteren Möglichkeiten die Seite bietet bzw. auf der Seite selektiert werden können.

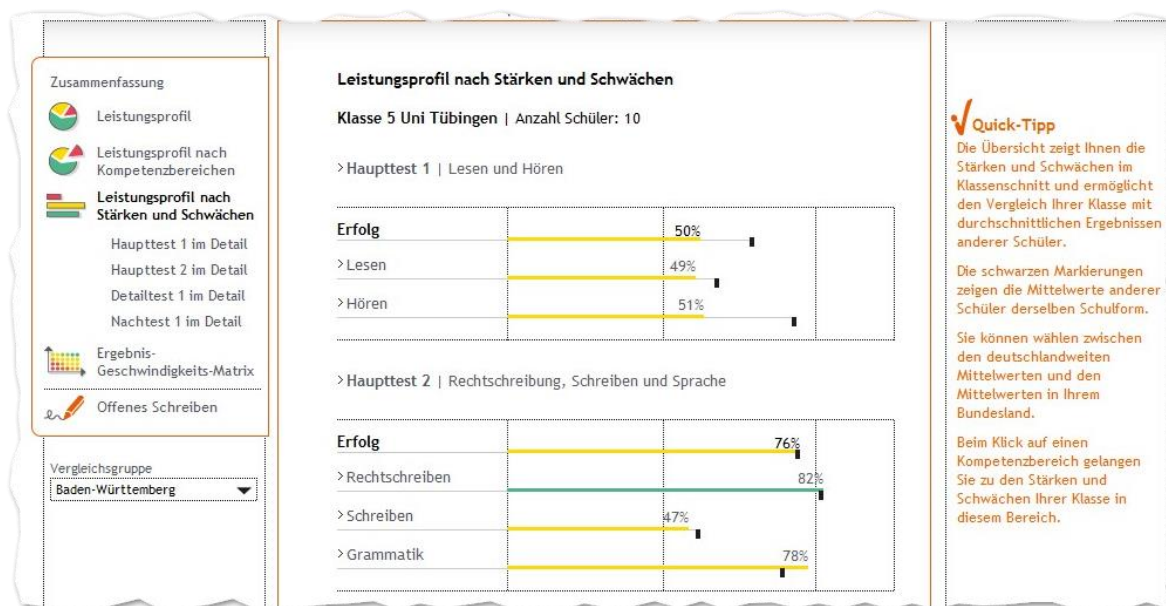


Abbildung 8. Leistungsprofil nach Stärken und Schwächen der gesamten Klasse

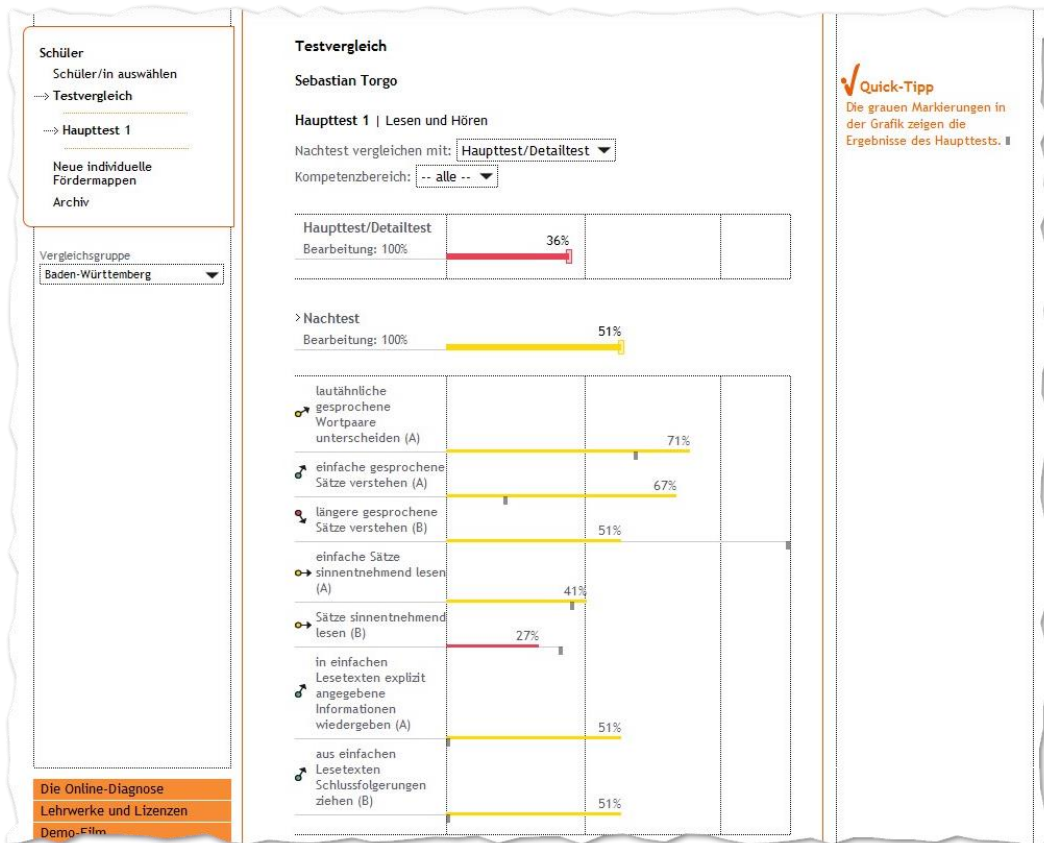


Abbildung 9. Testvergleich des Haupttests/Detailtests mit dem Nachtest

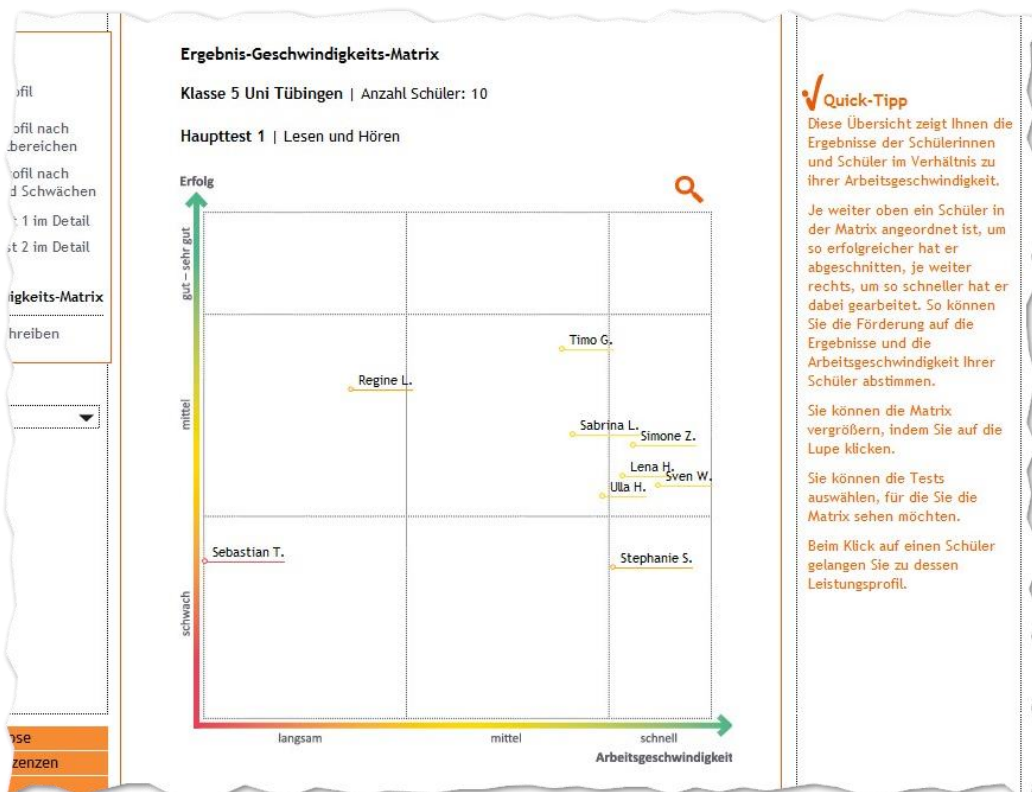


Abbildung 10. Ergebnis-Geschwindigkeits-Matrix für jede Schülerin und jeden Schüler

Betrachtung der Ergebnisrückmeldungen des Testverfahrens *Online Diagnose* in Bezug auf die Anforderungen an elektronische Ergebnisrückmeldungen. Das Testverfahren *Online Diagnose* erfüllt die Anforderungen an elektronische Ergebnisrückmeldungen größtenteils. Die Ergebnisrückmeldungen werden über das jeweilige Klassenkonto der Lehrkraft internetbasiert zurückgemeldet. Durch die automatische Auswertung der Tests geschieht die Rückmeldung direkt nach der Bearbeitung der jeweiligen Tests. Die Ergebnisrückmeldungen sind also zeitnah abzurufen. Die Ergebnisse werden in Grafiken, z.B. als Kreis- und Balkendiagramme, und als Text dargestellt. Sie können daher in unterschiedlichen Formaten gelesen werden und erleichtern somit das Verstehen der Informationen. Lehrkräfte, die Schwierigkeiten mit der Interpretation einer Grafik haben, können die Texte als Informationsquelle nutzen. Die Datenmenge der zurückgemeldeten Ergebnisse ist enorm. Diese können allerdings in unterschiedliche Aggregationsebenen differenziert werden. Zum Beispiel kann eine Zusammenfassung zum Bereich Lesen für eine Schülerin oder einen Schüler auf die jeweiligen Aufgabenblöcke selektiert werden. Die Lehrkraft kann bzw. muss somit selbst entscheiden, wie viele Informationen sie zu welcher Schülerin bzw. welchem Schüler benötigt. Die Leistung der Schülerinnen und Schüler werden anhand von Ampelfarben dargestellt, so kann zwischen einer guten und einer schlechten Leistung unterschieden werden. Die Ergebnisrückmeldungen bieten dadurch einen schnellen Überblick über die Leistungen der Klasse. Die Navigation durch die Ergebnisrückmeldungen wird über Erklärungen am Seitenrand erleichtert. Die Lehrkräfte können also eine direkte Hilfeleistung im Umgang mit dem Onlineinstrument nutzen. In Bezug auf die Tests besteht allerdings kein Wiedererkennungseffekt zu den Ergebnisrückmeldungen, der eine Interpretation der Ergebnisse erleichtern würde. Im Rahmen des Testverfahrens *Online Diagnose* werden die Anforderungen an Ergebnisrückmeldungen nach Vergleichswerten erfüllt. Allerdings entsprechen diese nicht den Maßstäben sogenannter fairer Vergleiche. Der Anforderung an Ergebnisrückmeldungen, dass die wissenschaftlichen Erkenntnisse in Handlungswissen übersetzt werden sollten, kann indirekt in Form der je nach Testleistung ausgegebenen Fördermaterialien entsprochen werden.

3.5 Das Testverfahren *Online Diagnose* in Bezug auf das Zyklusmodell der Rückmeldung

Das Testverfahren *Online Diagnose* lässt sich in das Zyklusmodell der Rückmeldung einbetten (Abbildung 11).

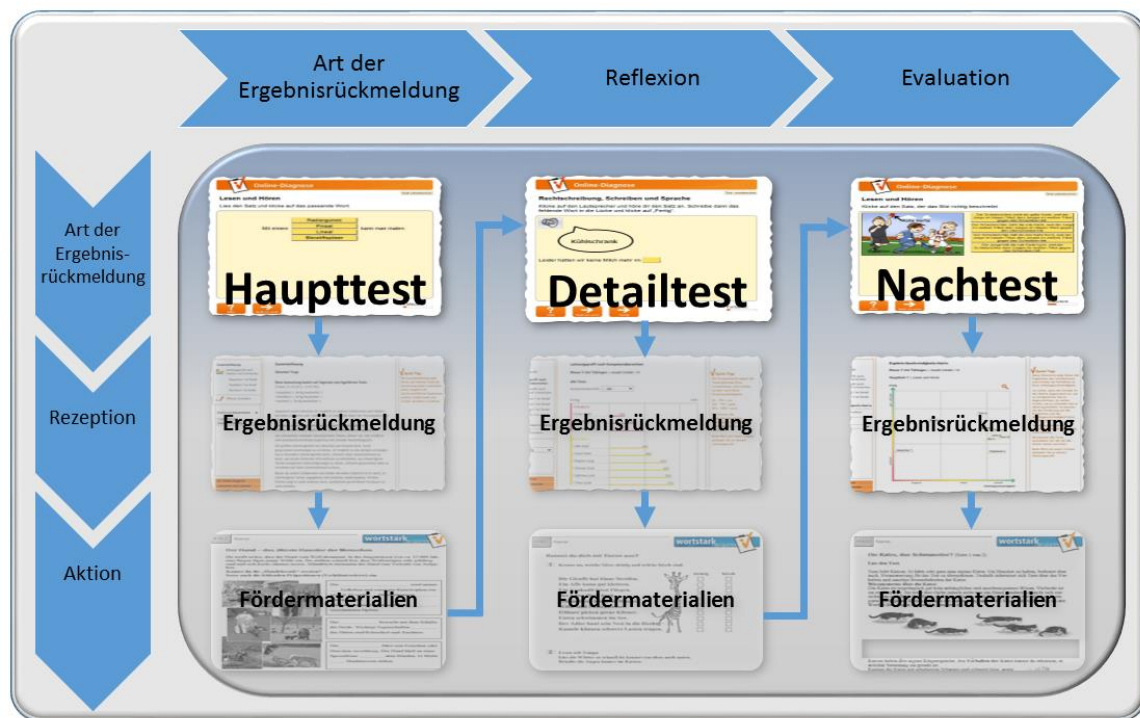


Abbildung 11. Ablauf Testverfahren *Online Diagnose* in Anlehnung an das Zyklusmodell der Rückmeldung

Haupttest, Detailtest, Nachtest, die Ergebnismeldungen und die Fördermaterialien lassen sich den Prozessschritten des Zyklusmodells der Rückmeldung zuordnen. Der Haupttest stellt im Ablauf des Testverfahrens *Online Diagnose* die erste Erfassung der Leistung von Schülerinnen und Schülern dar. Die Ergebnismeldungen aus dem Haupttest dienen damit als objektivierbare Informationen und müssen von den Lehrkräften interpretiert und verstanden werden. Somit ist der Haupttest der Art der Ergebnismeldungen zuzuordnen und die Ergebnismeldungen dem Prozessschritt Rezeption. Im Anschluss an den Prozessschritt Rezeption folgt im Zyklusmodell der Rückmeldung der Prozessschritt Reflexion. Die Lehrkräfte haben durch das Testverfahren *Online Diagnose* die Möglichkeit, ihre Interpretationen aus den Ergebnismeldungen des Haupttests und die daran angeschlossenen Maßnahmen empirisch zu überprüfen bzw. zu reflektieren. Damit sind der Detailtest und die sich

daran anschließenden Ergebnisrückmeldungen dem Prozessschritt Reflexion zuzuordnen. Der Einsatz von Maßnahmen, wie z.B. die Ausgabe der durch das Testverfahren *Online Diagnose* vorgegebenen Fördermaterialien, kann dem Prozessschritt Aktion des Zyklusmodells der Rückmeldung zugeordnet werden. Der Nachtest und die daraus entstandenen Ergebnisrückmeldungen können als empirisch basierte Evaluation angesehen und damit dem Prozessschritt Evaluation zugeordnet werden.

3.6 Zusammenfassung der Kapitel zum Testverfahren *Online Diagnose*

Die Ziele des Testverfahrens *Online Diagnose* beziehen sich auf die Erkennung heterogener Lernstände von Schülerinnen und Schülern und auf Maßnahmen zur Förderung von Stärken und zur Beseitigung von Schwächen. Insgesamt können drei Tests, die Haupt-, Detail- und Nachtests, eingesetzt werden. Außerdem können die Lehrkräfte Ergebnisrückmeldungen zu jeder Schülerin und jedem Schüler online abrufen sowie entsprechende individuelle Fördermaterialien.

In Baden-Württemberg konnten die Lehrkräfte vom Schuljahr 2009/2010 bis 2012/2013 eine landesweite Lizenz des Testverfahrens *Online Diagnose* nutzen. Das Testverfahren *Online Diagnose* stellt Anforderungen an die Lehrkräfte hinsichtlich ihrer Computerkenntnisse. Die Lehrkräfte müssen sich z.B. auf der Benutzeroberfläche des Testverfahrens zurechtfinden, innerhalb des Klassenkontos die Accounts für Schülerinnen und Schüler anlegen und für die Tests freischalten. Im Programm enthaltene Hilfestellung bei der Nutzung des Onlinetestverfahrens erhalten die Lehrkräfte direkt am Seitenrand der Benutzeroberfläche. Zusammenfassend kann aus der Gegenüberstellung des Testverfahrens *Online Diagnose* mit den Bildungsstandards der Grundschule festgehalten werden, dass die Aufgabenbereiche des Testverfahrens *Online Diagnose* nicht die gesamten Bereiche der Bildungsstandards abdecken. Außerdem enthält das Testverfahren *Online Diagnose* teilweise Aufgaben, die nicht in die Bildungsstandards der Grundschule aufgenommen wurden. Das Testverfahren *Online Diagnose* eignet sich daher eher als ein die Bildungsstandards der Grundschule ergänzendes Instrument. Eine Erfassung zentraler Kompetenzen, die für die Sekundarstufe notwendig sind, kann das Testverfahren *Online Diagnose* nicht vollständig erbringen. Die Ergebnisrückmeldungen zum Testverfahren *Online Diagnose* erfüllen größtenteils die Befunde zu den Anforderungen elektronischer Ergebnisrückmeldungen. Sie bieten den Lehrkräften als Nutzer der Ergebnisrückmeldungen eine Auswahl unterschiedlicher

Formate, z.B. Grafiken und Texte, die die Interpretation der Informationen erleichtern. Die Ergebnismrückmeldungen können auf unterschiedlichen Aggregationsebenen abgerufen werden und markante Ergebnisse sind visuell gekennzeichnet. Die Auswahl der Ergebnisse und die Detailgenauigkeit kann die Lehrkraft, je nach Bedarf flexibel treffen bzw. wählen. Ebenso bietet das Testverfahren *Online Diagnose* für die Ergebnisse Vergleichswerte an, die allerdings nicht den Standards fairer Vergleiche entsprechen. In seiner Funktion kann das Testverfahren *Online Diagnose* als empirisches Verfahren auf der Matrix des Zyklusmodells der Rückmeldung abgebildet werden. Wurde das Testverfahren wie vorgesehen eingesetzt, bot es zu mehreren Zeitpunkten im Verlauf des Schuljahres objektivierte Informationen aus empirischen Daten und somit eine Unterstützung der Lehrkräfte bei diagnostischen Aufgaben. Das Testverfahren bietet durch die individuellen Rückmeldungen eine Lösung für den Umgang mit Heterogenität in der Klasse und durch die Ausgabe von Fördermaterialien eine Lösung für die individuelle bzw. gezielte Förderung. Das Testverfahren kann somit als Unterstützung und Übergangslösung angesehen werden, bis die im Rahmen der Kultusministerkonferenz der Länder geforderte Qualifizierung der Lehrkräfte durch Aus- und Fortbildung greift.

3.7 Fragestellungen

Aus den theoretischen Erkenntnissen lassen sich folgende Fragestellungen ableiten:

In der ersten Fragestellung soll die Durchführung des Testverfahrens *Online Diagnose* und die Nutzung der daraus resultierenden Ergebnisse untersucht werden. Neben anderen Kriterien interessieren z.B. die Art der Durchführung und technische Probleme bei der Durchführung. Außerdem soll überprüft werden, ob sich die Prozessschritte Rezeption, Reflexion, Aktion und Evaluation des Zyklusmodells der Rückmeldung aus den erhobenen Daten zum Testverfahren *Online Diagnose* ableiten lassen.

1. Wie setzen die Lehrkräfte das Testverfahren *Online Diagnose* ein und inwiefern werden die Ergebnismrückmeldungen genutzt? Lassen sich aus der Erhebung zum internetgestützten Diagnose- und Förderinstrument *Online Diagnose* Schritte der Nutzung ableiten?

Weiterhin soll überprüft werden, ob das zur Diagnostik und Förderung bereitgestellte Testverfahren als Diagnose- und Förderinstrument genutzt werden kann. Die Schritte der Nutzung werden in Zusammenhang mit der Wirkung des Testverfahrens *Online Diagnose* auf Lernende, auf das Kollegium und auf Lehrende überprüft.

2. Sind die Prozessschritte der Nutzung: Rezeption, Reflexion und Aktion mit der Wirkung des Testverfahrens *Online Diagnose* auf Lernende, das Kollegium und Lehrende assoziiert?

Zudem werden Einflussfaktoren auf die Schritte der Nutzung untersucht. In empirischen Studien, die Individualdaten von Schülerinnen und Schülern erhoben hatten, zeigte sich z.B. die Akzeptanz der Testverfahren als wichtiger Prädiktor. Die Einflussfaktoren werden anhand der Bedingungen des Zyklusmodells der Rückmeldung strukturiert.

3. Welche Bedingungen sind mit den Prozessschritten der Nutzung des Testverfahrens *Online Diagnose* assoziiert?

Die Teilstudien und Forschungsfragen beziehen sich auf die nachfolgend dargelegte Befragung von Lehrkräften (Kapitel 4) und beschränken sich dadurch auf die Perspektive der Lehrkräfte.

4 Datengrundlage

4.1 Beschreibung der Studie

Die Grundgesamtheit der im Schuljahr 2012/2013 durchgeführten Untersuchung zum internetbasierten Diagnose- und Förderinstrument *Online Diagnose* umfasste in ganz Baden-Württemberg rund 1012 Deutschlehrkräfte³⁰ der Jahrgangsstufe fünf. Für den Versand der Fragebögen wurde die Grundgesamtheit der Lehrkräfte anhand des Verzeichnisses für Grund-, Werkreal- und Hauptschulen des statistischen Landesamtes Baden-Württemberg aus dem Schuljahr 2010/2011 ermittelt (Statistisches Landesamt, 2011). Das Schulverzeichnis wurde für den Versand der Fragebögen wie folgt bearbeitet: Zunächst wurden alle Grundschulen und private Haupt- und Werkrealschulen aus der Liste entfernt, da sich die verpflichtende Durchführung und Finanzierung des Testverfahrens auf öffentliche Haupt-, Werkreal- und Gemeinschaftsschulen beschränkte. Die Schulen, die im Schuljahr 2010/2011 eine fünfte, sechste oder siebte Klasse führten, wurden aus der Übersicht ausgewählt. Schulen, die laut Schulverzeichnis im Jahr 2010/2011 weder eine fünfte noch eine sechste Klasse hatten, wurden im Internet recherchiert. Anhand der Informationen aus der Internetrecherche wurden die Schulen, die auch im aktuellen Schuljahr 2012/2013 keine fünfte Klasse hatten, gelöscht – dies betraf 25 Schulen. Nach dieser Prozedur konnte auf rund 870 Haupt- Werkreal- und Gemeinschaftsschulen zurückgegriffen werden.

Als Außenstellen aufgeführte Schulen wurden der Hauptstelle zugeordnet, da eine Verteilung der Fragebögen über die Schulleitung und dadurch eine entsprechende Genehmigung der Schulleitung erwünscht war. Die Anzahl der pro Schule zu versendenden Fragebögen wurde zunächst anhand des Schulverzeichnisses des statistischen Landesamtes festgelegt und im Anschluss durch eine Recherche auf den jeweiligen Schulhomepages aktualisiert. Bei den Schulen, deren aktuelle Klassenanzahl nicht über die jeweilige Schulhomepage nachvollzogen werden konnte, wurde die Anzahl der zu versendenden Fragebögen wie folgt generalisiert: An Schulen, die im Schuljahr 2010/2011 vier oder fünf 5. Klassen (inkl. Außenstellen) führten und über die im Internet keine Angaben zu finden waren, wurde der Unterlagensatz in dreifacher Ausfertigung verschickt. Schulen, die im Schuljahr 2010/2011 drei fünfte Klassen führten, erhielten die Unterlagen in zweifacher Ausfertigung. An Schulen,

³⁰ Die Gesamtanzahl der unterrichtenden Deutschlehrkräfte in den fünften Klassen konnte nur geschätzt werden, da von 8,2 % der Haupt-, Werkrealschulen und Gemeinschaftsschulen keine Angaben über die Anzahl der Deutschlehrkräfte vorlagen.

die im Schulverzeichnis im Schuljahr 2010/2011 mit einer fünften Klasse aufgeführt waren, wurde der Unterlagensatz einmal versandt. Durch ein Fax der Schulleitung an das Datenunternehmen Daten.Werk sollten die recherchierten Angaben korrigiert werden, damit gegebenenfalls Fragebögen nachgeliefert werden konnten.

Aus Gründen des Datenschutzes wurde für den Versand und die Erfassung des Rücklaufs das Datenunternehmen Daten.Werk³¹ engagiert. Über das Unternehmen wurden die Fragebögen verschickt und gegebenenfalls nachträglich in der benötigten Anzahl versandt (Abbildung 12). Ebenfalls nahm das Datenunternehmen den Rücklauf entgegen und erstellte einen anonymisierten Datensatz. Die Fragebögen, die Anschreiben an die Schulleitungen und Lehrkräfte sowie ein sogenanntes Zuordnungsblatt und eine Gewinnspielkarte wurden, wie in Abbildung 12 dargestellt, am 8. und 9. April 2013 per Post an die Schulleitungen zur Kenntnisnahme und mit der Bitte zur Weiterreichung an die entsprechenden Deutschlehrkräfte versandt. Der Versandtermin nach den Osterferien wurde in der Annahme gewählt, dass alle Lehrkräfte das internetbasierte Diagnose- und Förderinstrument bereits eingesetzt hatten und auch die Fördermaterialien sowie der Detailtest zur Anwendung gekommen waren. Zudem konnte davon ausgegangen werden, dass die Zeit nach Ausgabe der Halbjahreszeugnisse eine für die Lehrkräfte weniger beanspruchende Phase im Schuljahr ist.

Aufgrund des geringen Rücklaufs wurde noch vor den Pfingstferien am 14. Mai 2013 eine Nachfassaktion gestartet. Alle Schulen, die sich bis dahin nicht zurückgemeldet hatten, wurden telefonisch kontaktiert. Bis zum 3. Juli 2013 hatten alle Schulen mindestens eine telefonische Erinnerung erhalten. Die Telefonate waren nach folgender Systematik organisiert: Zuerst wurden die Schulen angerufen, die zwei oder mehr fünfte Klassen führten. Anschließend wurden alle Schulen, die eine fünfte Klasse führten, durch ein Zufallsverfahren in eine Reihenfolge gebracht und nach und nach telefonisch kontaktiert. Weiterhin wurden die Schulen, die dem Daten.Werk ein Fax zugesandt, jedoch keinen Fragebogen zurückgesandt hatten, ab dem 8. Juli 2013 telefonisch kontaktiert. Insgesamt wurden 664 Schulen angerufen. Im Zuge dieser Nachfassaktion wurden 291 Fragebögen erneut versandt. Schulen, die telefonisch nicht erreicht werden konnten (10. Juli 2013), die erneut Unterlagen zugesandt bekommen hatten, und Schulen, die bereits ein Fax, jedoch keinen Fragebogen zurückgeschickt hatten (17. Juli 2013), sowie Schulen, die beim Telefonat angegeben hatten, die Unterlagen erhalten zu haben, jedoch ebenfalls keinen Fragebogen zurückgesandt hatten (18. Juli 2013), wurden per E-Mail nochmals an die Befragung erinnert. Einschränkungen

³¹ <http://www.daten-werk.de/>

des E-Mail-Versands bestanden, wenn die entsprechende E-Mail-Adresse der jeweiligen Schule nicht auf der Schulhomepage verfügbar war. Vereinzelt Fragebögen kamen während und nach den Sommerferien an. Die Erhebung endete schließlich am 25. September 2013. Der Rücklauf betrug letztendlich 150 Fragebögen aus 137 Haupt-, Werkreal- und Gemeinschaftsschulen. Dies entspricht einer Rücklaufquote von 14,82 % der insgesamt rund 1012 Deutschlehrkräfte.

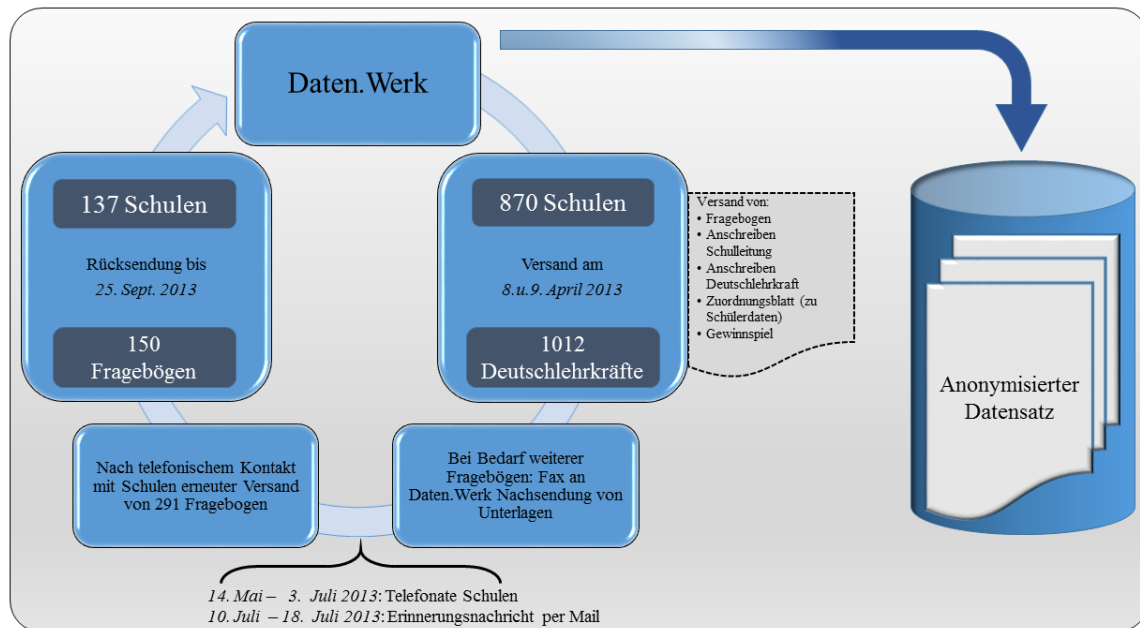


Abbildung 12. Ablauf der Erhebung der Lehrkraftbefragung zum Testverfahren *Online Diagnose*

Durch die Freiwilligkeit der Beantwortung des Fragebogens sowie die geringe Rücklaufquote ist die Repräsentativität der Stichprobe eingeschränkt. Selbst durch die Ausschreibung eines Gewinnspiels, Telefonate mit den Akteuren der Schule sowie die versendeten E-Mails konnte die Teilnahmequote nicht erhöht werden. Ein Grund hierfür ist sicherlich ein generell geringer Rücklauf zusätzlicher Erhebungen nach entsprechenden Hauptuntersuchungen (z.B. Ditton & Arnold, 2004; Schwippert, 2004). In den Telefonaten wurden die Lehrkräfte nach Gründen für die Nichtteilnahme gefragt; hier wurden folgende Gründe in absteigender Häufigkeit genannt: Die Lehrkräfte waren z.B. durch die Umstellung auf eine andere Schulform überlastet; die durchführende Lehrkraft war aufgrund von Mutterschutz oder Rente etc. nicht mehr an der Schule tätig; technische Probleme hinderten die Durchführung des Testverfahrens *Online Diagnose*; aufgrund anderer Evaluationen wurde nicht am Testverfahren teilgenommen.

4.2 Beschreibung der Stichprobe

Zu den Lehrkräften. In der Untersuchung zum Diagnose- und Förderinstrument *Online Diagnose* wurden Lehrkräfte zur Anwendung des Testverfahrens und zum Umgang mit den Ergebnisrückmeldungen befragt. Im Zuge der Erhebung sendeten 150 der ca. 1012 im Schuljahr 2012/13 angeschriebenen Deutschlehrkräfte den ausgefüllten Fragebogen zurück. Dies entspricht etwa 15 % der unterrichtenden Deutschlehrkräfte in den fünften Klassen der Haupt-, Werkreal- und Gemeinschaftsschulen in Baden-Württemberg. An der Befragung nahmen mehr weibliche Lehrkräfte 81,6 % (Alter $M = 39$ Jahre, $SD = 12$ Jahre), als männlich Lehrkräfte 18,4 % (Alter $M = 45$ Jahre, $SD = 13$ Jahre) teil. Das durchschnittliche Alter aller teilnehmenden Lehrkräfte betrug 40 Jahre ($SD = 12$ Jahre). Die Lehrkräfte waren einschließlich der Referendariats- und der Vorbereitungszeit im Mittel 13 Jahre berufstätig ($SD = 11$ Jahre). Sie hatten durchschnittlich an $M = 2.6$ ($SD = 1.19$) Schulen länger als drei Monate unterrichtet. Die Lehrkräfte gaben an, ein nominelles Unterrichtsdeputat von $M = 24$ ($SD = 4.37$) Wochenstunden zu haben. Zudem hatten 20 % der antwortenden Lehrkräfte eine Funktionsstelle wie z.B. eine Beratungslehrertätigkeit oder die Stellvertretung der Schulleitung inne. Insgesamt 96 % der Lehrkräfte dieser Stichprobe hatten sich für ein Studium an einer Pädagogischen Hochschule in Baden-Württemberg (94 %), mit dem Abschluss im Bereich der Grund-, Haupt- und Werkrealschule (96 %) entschieden. In Tabelle 1 ist die erste und zweite Studienfachwahl für das Unterrichtsfach Deutsch sowie die entsprechende Note festgehalten. Ebenso wird die Note der praktischen Ausbildungsphase berichtet. Aus den Angaben wird deutlich, dass 26 % der Lehrkräfte das Fach Deutsch nicht als Hauptfach, sondern im zweiten Studienfach studiert hatten. Die zweithäufigste Wahl für das erste Studienfach war Mathematik. Für das zweite Studienfach wurde an zweiter Stelle das Fach Geschichte gewählt. Die Zeitspanne der Abschlüsse bewegte sich zwischen 1973 und 2013.

Tabelle 1
Studienfach Deutsch und Abschlussnoten

	Anzahl	Abschlussnote	
	%	M	SD
Studienfach 1: Deutsch	44	1.9	.57
Studienfach 2: Deutsch	26	1.9	.55
Zweite praktische Ausbildungsphase	44	1.8	.67

Anmerkung. $N = 150$.

Zu den Klassen. Die teilnehmenden Lehrkräfte unterrichteten durchschnittlich eine Klasse im Unterrichtsfach Deutsch und diese ca. fünf Schulstunden wöchentlich ($SD = 1.40$). In dieser Klasse waren 77 % als Klassenlehrer tätig. Die Klassengröße betrug im Mittel $M = 18$ ($SD = 6$) Schülerinnen und Schüler. Den Haupttest ($M = 2.6$, $SD = 2.66$) und die Fördermaterialien ($M = 2.4$, $SD = 2.58$) wendeten die Lehrkräfte in der Klasse insgesamt häufiger an als den Detailtest ($M = 1.6$, $SD = 1.92$) oder den Nachtest ($M = 1.4$, $SD = 2.47$). In der erwähnten Klasse setzte die Lehrkraft die Instrumente im Durchschnitt bei folgender Anzahl an Schülerinnen und Schülern ein: Haupttest $M = 17$ ($SD = 7$), Ergebnismeldungen $M = 15$ ($SD = 8$), Detailtest $M = 11$ ($SD = 9$), Nachtest $M = 8$ ($SD = 9$) und Fördermaterialien $M = 15$ ($SD = 7$).

5 Teilstudie 1: Deskription der Nutzung des Testverfahrens *Online Diagnose*

Ziel des Diagnose- und Förderinstruments *Online Diagnose*³² ist die Unterstützung von Lehrkräften bei der individuellen Diagnostik von Förderbedarfen bei Schülerinnen und Schülern sowie der Ausgleich möglicher Defizite bzw. die Förderung der im Bildungsplan der Grundschule geforderten Kompetenzen. Nach der Bearbeitung des Haupttests erhalten die Lehrkräfte die Ergebnismeldungen zur Leistung ihrer Schülerinnen und Schüler sowie Vorschläge zur Förderung der Schülerinnen und Schüler in Form von Förderaufgaben. Ebenso haben die Lehrkräfte die Möglichkeit, einen Detail- und Nachtest durchzuführen. In Teilstudie 1 sollen Fragen zur Nutzung und den Materialien des Testverfahrens *Online Diagnose* beantwortet werden. Zum einen wird der Frage nachgegangen, wie die Lehrkräfte das Testverfahren *Online Diagnose* einsetzen und inwiefern die Ergebnismeldungen genutzt wurden. Zum anderen soll geprüft werden, ob sich aus der Erhebung zum internetgestützten Diagnose- und Förderinstrument *Online Diagnose* Prozessschritte der Nutzung ableiten lassen. Nachfolgend werden der Ablauf und die Nutzung des Testverfahrens *Online Diagnose* deskriptiv berichtet. Außerdem werden Skalen zur Erfassung der Prozessschritte des Zyklusmodells der Rückmeldung aus der Befragung abgeleitet.³³

5.1 Statistische Analysen

Die Ergebnisse werden deskriptiv berichtet: Mittelwert (M), Modus und Standardabweichungen (SD). Die jeweiligen positiven und die jeweiligen negativen Antwortkategorien einer Likert-Skala wie z. B. 1 (trifft nicht zu) und 2 (trifft eher nicht zu) werden zusammengefasst und prozentual dargestellt. Außerdem werden in den Tabellen jeweils die Prozentangaben der Zustimmung zu einem Item durch die beiden zusammengefassten Antwortmöglichkeiten 3 (trifft eher zu) und 4 (trifft zu) berichtet.

Die Skalenbildung anhand von exploratorischen Faktorenanalysen und Reliabilitätsanalysen wurde in SPSS 22 durchgeführt. Es wurde das Extraktionsverfahren der Hauptkomponentenanalysen mit der Promax-Rotation und Kaiser-Normalisierung verwendet. Die

³² Das Testverfahren *Online Diagnose* wird in Kapitel 3 ausführlich beschrieben.

³³ Die Prozessschritte des theoretischen Zyklusmodells der Rückmeldung (z. B. Helmke & Hosenfeld, 2005) können in Kapitel 2.3 nachgelesen werden.

Voreinstellung in SPSS von 25 Iterationen wurde beibehalten (z.B. Tabellen 9, 10, 12, 13, 17 und 23). Neben den Faktorladungen werden der Eigenwert und die Kommunalitäten berichtet. Der Eigenwert eines Faktors sollte > 1 sein, damit klärt der Faktor mehr Varianz auf als ein einzelnes Item. Die Kommunalitäten berichten die durch die extrahierten Faktoren aufgeklärten Varianzanteile (Bühner, 2011). Das exploratorische Vorgehen wurde als Methode gewählt, weil die entwickelten Items keinem etablierten Instrument entnommen sind und deshalb die Zuordnung der Items zu den speziellen Eigenschaften bisher nicht empirisch belegt ist (Bühner, 2011). Zur internen Konsistenz im Rahmen der Reliabilitätsanalysen wird Cronbachs α berichtet. Außerdem wurden die Trennschärfekoeffizienten (r_{it}) der jeweiligen Items überprüft.

Sämtliche bivariaten Korrelationen wurden in dem Programm Mplus Version 7.1 (Muthén & Muthén, 1998-2013) berechnet. Mplus bietet die Möglichkeit, das Full Information Maximum Likelihood Verfahren (FIML) zum Umgang mit fehlenden Werten zu verwenden. Dieses stellt eines der momentan am besten geeigneten Verfahren zum Umgang mit fehlenden Werten dar (z.B. Lüdtke, Robitzsch, Trautwein & Köller, 2007; Enders, 2010). Insgesamt lagen die fehlenden Werte zwischen 0.7 % und 70.7 %. Die fehlenden Werte werden zu jedem Item in den jeweiligen Tabellen berichtet.

5.2 Ergebnisse

Die deskriptiven Befunde zum Testverfahren *Online Diagnose* werden in die folgenden Abschnitte untergliedert: Durchführung und Nutzung des Diagnose- und Förderinstruments *Online Diagnose*, Motivation bei der Durchführung des Haupttests und dem Einsatz der Fördermaterialien, subjektive Qualitätseinschätzung zum Testverfahren *Online Diagnose* und Prozessschritte der Nutzung von Ergebnisrückmeldungen.

Der erste Abschnitt geht auf den Ablauf, die Durchführung und die Nutzung der Testinstrumente und Fördermaterialien ein. Ebenso werden die Rahmenbedingungen hinsichtlich der technischen Ausstattung an Schulen und die möglichen Einflüsse für ein schwaches Testergebnis berichtet. Zudem geht der Abschnitt schließlich noch auf die Unterstützung der Lehrkräfte u. a. in Form schriftlicher Hilfen im Rahmen des Testverfahrens *Online Diagnose* ein. Im darauffolgenden Abschnitt werden Wertaspekte zur Motivation aus Schülerperspektive (Einschätzung durch die Lehrkraft) und aus Lehrerperspektive (eigene Einschätzung) in Bezug auf den Haupttest sowie die Fördermaterialien berichtet. Im Rahmen des nächsten Abschnitts wird auf die subjektive Qualitätseinschätzung hinsichtlich der Verständlichkeit

und Eignung der Testaufgaben des Haupttests eingegangen. Ebenso wird als Qualitätsaspekt ein Bezug zum Bildungsplan der Grundschule hergestellt und die subjektive Qualitätseinschätzung des Haupttests, der Ergebnisrückmeldungen und der Fördermaterialien durch die Lehrkräfte festgehalten. Abschließend werden Aspekte der Nutzung durch die Prozessschritte Rezeption, Reflexion, Aktion und Evaluation des theoretischen Zyklenmodells der Rückmeldung operationalisiert.

5.2.1 Durchführung und Nutzung des Testverfahrens *Online Diagnose*

Angaben zum Ablauf des Testverfahrens *Online Diagnose*. Als Information über den Einsatz des Haupttests interessierte der konkrete Ablauf des Testverfahrens *Online Diagnose*. Die in Tabelle 2 aufgeführten Items konnten in der Antwortkategorie 0 (trifft nicht zu) und 1 (trifft zu) beantwortet werden. Beinahe alle an der Befragung teilnehmenden Lehrkräfte (97 %) informierten ihre Klasse vorab über den Zweck des Haupttests. Etwa drei Viertel der an der Befragung teilnehmenden Lehrkräfte gab an, die Schülerinnen und Schüler einer Klasse würden den Test zeitgleich bearbeiten. Ein Viertel der Lehrkräfte kreuzte an, dass die Schülerinnen und Schüler den Haupttest in kleinen Gruppen nacheinander bearbeiteten. Laut den Lehrkräften bearbeitete nur ein Prozent der Schülerinnen und Schüler den Haupttest als Hausaufgabe.

Tabelle 2

*Deskriptive Statistik der Items zum Ablauf des Testverfahrens *Online Diagnose**

Item	Itemtext	<i>M</i>	<i>SD</i>	Fehlende Angaben (Prozent)
1.	Ich informiere die Klasse vorab über den Zweck des Haupttests.	.97	0.16	2.0
2.	Die gesamte Klasse bearbeitet den Haupttest gleichzeitig.	.76	0.43	8.7
3.	Die Klasse wird zur Bearbeitung des Haupttests in kleine Gruppen geteilt, die nacheinander den Test bearbeiten.	.24	0.43	12.7
4.	Die Schülerinnen und Schüler bearbeiten den Haupttest als Hausaufgabe.	.01	0.12	8.0

Anmerkung. *N* = 150; Skalierung: 0 (trifft nicht zu) bis 1 (trifft zu).

Nutzungsumfang und Zeitaufwand für die Anwendung des Testverfahrens *Online Diagnose*. Um den Nutzungsumfang des Testverfahrens *Online Diagnose* zu erfassen,

wurde die Anzahl der genutzten Tests pro Klasse erhoben (Abbildung 13). Die durchschnittliche Klassengröße betrug $M = 18$ Schülerinnen und Schüler pro Klasse ($SD = 5.7$; $N = 147$). Die Lehrkräfte nutzten den Haupttest für durchschnittlich $M = 17$ Schülerinnen und Schüler ihrer Klasse ($SD = 6.7$; $N = 137$). Deutlich weniger Lehrkräfte wandten den Detailtest an, nämlich bei durchschnittlich $M = 11$ Schülerinnen und Schülern ihrer Klasse ($SD = 9.3$; $N = 102$). Der Nachtest wurde durchschnittlich bei $M = 8$ Schülerinnen und Schülern der jeweiligen Klasse angewandt ($SD = 8.7$; $N = 90$). Zum Detailtest gaben 35 % der antwortenden Lehrkräfte an, diesen bei keiner Schülerin bzw. keinem Schüler eingesetzt zu haben. Noch weniger Lehrkräfte setzten den Nachtest ein, 46 % der antwortenden Lehrkräfte gaben an, den Nachtest nicht verwendet zu haben.

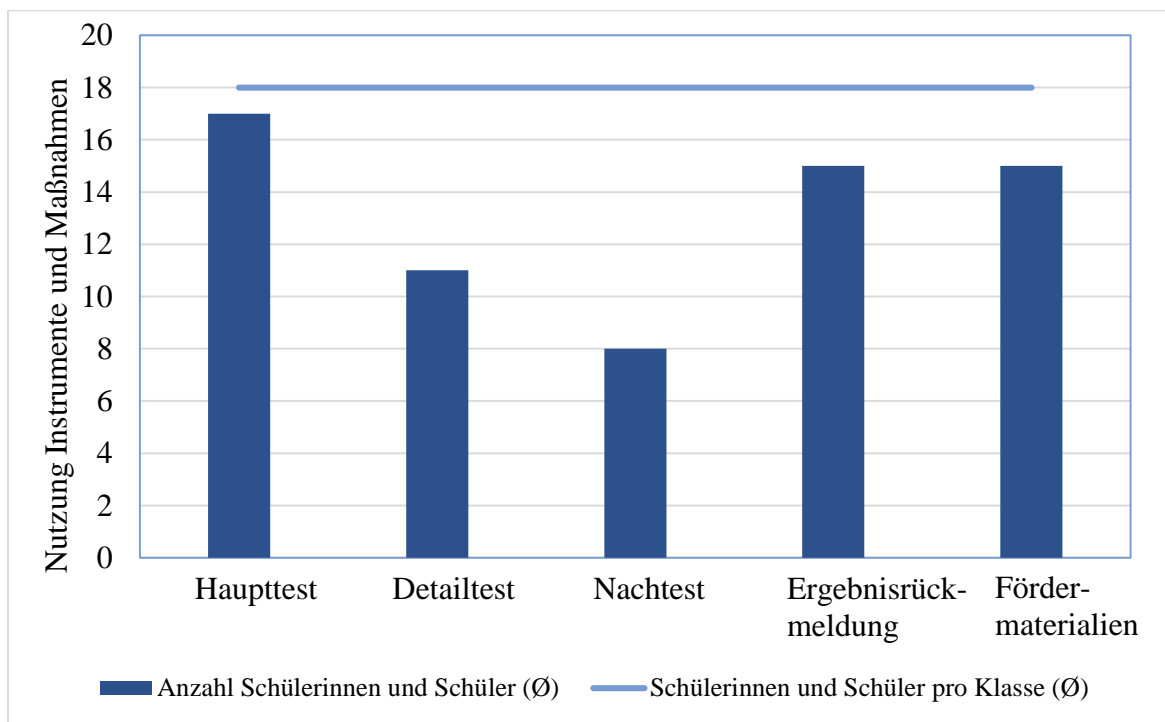


Abbildung 13. Nutzung Instrumente und Maßnahmen

Neben dem Nutzungsumfang wurden die Lehrkräfte befragt, wie viel Zeit sie für die Vorbereitung, Durchführung und Nachbereitung des Haupttests, des Detailtests und des Nachtests während und außerhalb des Unterrichts aufwandten (Abbildung 14 und Tabelle 3). Während des Unterrichts waren es durchschnittlich vier Schulstunden, die für die Vorbereitung, Durchführung und Nachbereitung des Haupttests benötigt wurden. Auf den Detailtest und Nachtest wurden jeweils zwei Schulstunden verwendet. Außerhalb des Unterrichts wandten die Lehrkräfte durchschnittlich drei Schulstunden für den Haupttest auf. Für den Detailtest

war es eine und für den Nachtest waren es ebenfalls knappe eineinhalb Schulstunden außerhalb des Unterrichts.

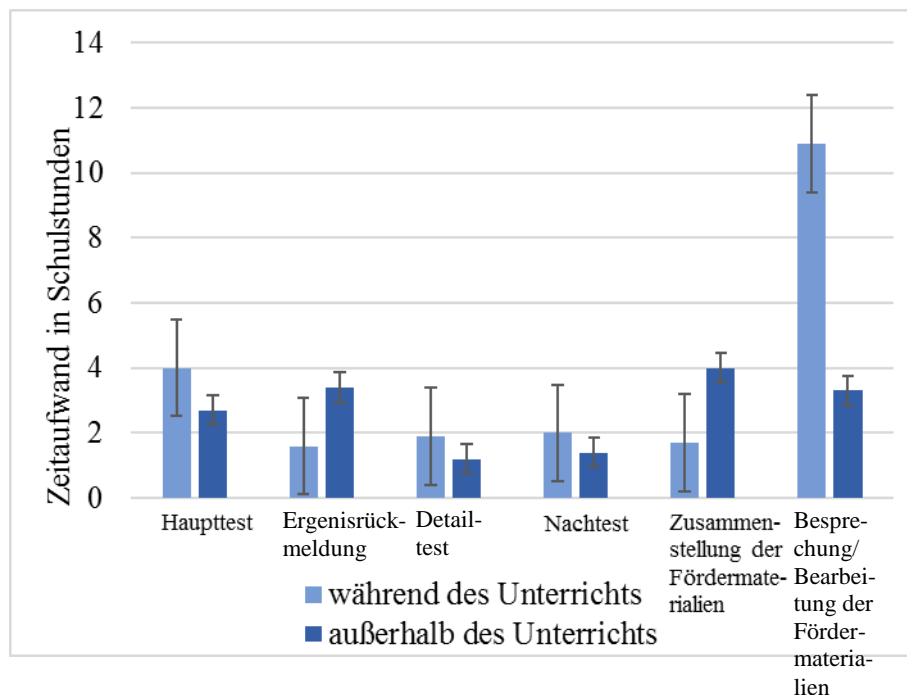


Abbildung 14. Zeitaufwand während und außerhalb des Unterrichts

Die Ergebnisrückmeldungen nutzten die Lehrkräfte für durchschnittlich $M = 15$ Schülerinnen und Schüler ihrer Klasse ($SD = 7.74$, $N = 130$). Hierzu gaben 11 % der antwortenden Lehrkräfte an, die Ergebnisrückmeldungen für keine Schülerin bzw. keinen Schüler genutzt zu haben. Für das Durcharbeiten der individuellen Ergebnisrückmeldungen benötigten die Lehrkräfte knappe zwei Schulstunden während des Unterrichts und durchschnittlich drei Schulstunden außerhalb des Unterrichts (Abbildung 14).

Die Fördermaterialien wurden im Durchschnitt bei $M = 15$ Schülerinnen und Schülern ($SD = 7.29$, $N = 136$) eingesetzt. Insgesamt gaben 7 % der antwortenden Lehrkräfte an, die Fördermaterialien nicht eingesetzt zu haben. Der Zeitaufwand ist bei den Fördermaterialien ebenso ein wichtiges Maß für den Nutzungsumfang. Die Lehrkräfte wurden zum einen gefragt, wie viel Zeit sie für die Zusammenstellung der Fördermaterialien benötigten, und zum anderen, wie viel Zeit sie für die Bearbeitung sowie Besprechung der Materialien aufwendeten (Tabelle 3). Beide Items beantworteten die Lehrkräfte für den Zeitaufwand während und außerhalb des Unterrichts. Für die Zusammenstellung der Fördermaterialien benö-

tigten die Lehrkräfte durchschnittlich zwei Schulstunden während der Unterrichtszeit. Außerhalb des Unterrichts benötigten sie für die Zusammenstellung der Fördermaterialien mehr als die doppelte Zeit: durchschnittlich vier Schulstunden. Für die Besprechung und Bearbeitung der Fördermaterialien wurden während des Unterrichts im Durchschnitt insgesamt elf Schulstunden benötigt – im Vergleich hierzu unterrichteten die befragten Lehrkräfte in den jeweiligen Klassen durchschnittlich ein bis sieben Schulstunden wöchentlich im Unterrichtsfach Deutsch. Außerhalb des Unterrichts wurden drei Schulstunden für die Besprechung und Bearbeitung der Fördermaterialien verwendet. Insgesamt wurden im Durchschnitt während des Unterrichts $M = 13$ Schulstunden für das Testverfahren *Online Diagnose* aufgewendet ($SD = 14.60$; $N = 67$; inkl. Bearbeitung und Besprechung der Fördermaterialien). Dies entsprach einer Unterrichtszeit von durchschnittlich $M = 3$ Wochen ($SD = 3.65$; $N = 61$) für den Einsatz des Testverfahrens *Online Diagnose*, denn die Lehrkräfte unterrichteten wöchentlich durchschnittlich $M = 5$ Deutschstunden in der jeweiligen Klasse ($SD = 1.53$, $N = 135$). Außerhalb des Unterrichts betrug der Stundenaufwand in etwa $M = 7$ Schulstunden ($SD = 8.82$; $N = 61$; inkl. der Bearbeitung und Besprechung der Fördermaterialien).

Die Tabelle 3 fasst die beschriebenen Durchschnittswerte für den Zeitaufwand während und außerhalb des Unterrichts zusammen. Auffällig ist hierbei der hohe Anteil an fehlenden Angaben, dieser liegt zwischen 17 % und 58 %.

Tabelle 3

Deskriptive Statistik der Items zum Zeitaufwand während und außerhalb des Unterrichts

Item	Itemtext	Zeitaufwand während des Unterrichts			Zeitaufwand außerhalb des Unterrichts		
		<i>M</i>	<i>SD</i>	Fehlende Angaben (Prozent)	<i>M</i>	<i>SD</i>	Fehlende Angaben (Prozent)
1.	Haupttest ^a	4.04	2.48	10.7	2.72	2.64	30.7
2.	Detailtest ^a	1.85	1.76	40.7	1.24	1.94	54.7
3.	Nachtest ^a	2.03	2.14	46.7	1.41	2.00	54.0
4.	Ergebnisrückmeldungen	1.64	1.99	44.7	3.00	2.61	27.3
5.	Fördermaterialien Zusammenstellung	1.73	3.20	50.7	4.03	3.15	17.3
6.	Fördermaterialien Bearbeitung/ Besprechung	10.90	13.58	23.3	3.27	8.22	58.0

Anmerkung. $N = 150$; ^aErfasst wurde der Zeitaufwand zur Vorbereitung, Durchführung und Nachbereitung.

Austausch über die Ergebnisrückmeldungen mit pädagogischen Akteuren der Schule. Die weiterführende Nutzung des Diagnose- und Förderinstrumentes *Online Diagnose* nach dessen Durchführung bestand in der Auseinandersetzung der Lehrkraft mit den zurückgemeldeten Ergebnissen zur Leistung ihrer Schülerinnen und Schüler. Die Lehrkräfte wurden befragt, mit wem sie die Ergebnisse der Schülerinnen und Schüler besprachen (Tabelle 4). Die Items konnten mit 0 (trifft nicht zu) und 1 (trifft zu) beantwortet werden. Ein Teil der antwortenden Lehrkräfte (7 %) gab an, die Ergebnisrückmeldungen des Haupttests mit keiner anderen Person besprochen zu haben, wohingegen 44 % der Lehrkräfte die Ergebnisse im persönlichen Gespräch mit allen teilnehmenden Schülerinnen und Schülern ihrer Klasse besprachen. 69 % besprachen die Ergebnisse mit ausgewählten Schülerinnen und Schülern ihrer Klasse im persönlichen Gespräch. 74 % der durchführenden Lehrkräfte pflegten den Austausch über die Ergebnisse des Haupttests mit anderen Lehrerinnen und Lehrern dieser Klasse. Bei 18 % der antwortenden Lehrkräfte galt dies auch für die Schulleitung. Für das persönliche Gespräch mit allen Eltern der am Testverfahren *Online Diagnose* teilnehmenden Schülerinnen und Schüler der Klasse nutzten 18 % der Lehrkräfte die Informationen aus den Ergebnisrückmeldungen. 70 % der Lehrkräfte nutzten die Informationen für das persönliche Gespräch mit einzelnen Eltern.

Tabelle 4

Deskriptive Statistik der Items zum Austausch der Lehrkräfte mit pädagogischen Akteuren der Schule

Item	Itemtext	<i>M</i>	<i>SD</i>	Fehlende Angaben (Prozent)
1.	Mit niemandem	.07	0.25	10.7
2.	Mit allen Schülerinnen und Schülern im persönlichen Gespräch	.44	0.50	7.3
3.	Mit einzelnen Schülerinnen und Schülern im persönlichen Gespräch	.69	0.47	8.7
4.	Mit anderen Lehrerinnen und Lehrern dieser Klasse	.74	0.44	4.7
5.	Mit der Schulleitung	.18	0.39	10.7
6.	Mit den Eltern aller Schülerinnen und Schülern	.18	0.39	6.0
7.	Mit den Eltern einzelner Schülerinnen und Schülern	.70	0.46	6.7

Anmerkung. *N* = 150; Skalierung: 0 (trifft nicht zu) bis 1 (trifft zu).

Anwendung der Fördermaterialien. Die Lehrkräfte wurden befragt, wie sie die Fördermaterialien in ihrer Klasse einsetzten. Einerseits wurde hierbei auf die Verfahrensweise bei der Anwendung eingegangen und andererseits wurde nach den Rahmenbedingungen der

Anwendung gefragt. Außerdem beantworteten die Lehrkräfte die Frage, ob die Besprechung der Lösungen in der Schule eingebunden wurde. Die einzelnen Aussagen konnten mit 0 (trifft nicht zu) oder 1 (trifft zu) angekreuzt werden. Die jeweiligen Items sind in Tabelle 5 aufgeführt.

Insgesamt vergaben 4 % der Lehrkräfte die gleichen Förderaufgaben an alle Schülerinnen und Schüler einer Klasse. 71 % der Lehrkräfte gaben zudem an, dass die nicht-förderbedürftigen Schülerinnen und Schüler die Förderaufgaben ebenfalls bearbeiteten. Der Einsatz von identischen Aufgaben bei allen förderbedürftigen Schülerinnen und Schülern fand bei 10 % statt. Im Gegensatz hierzu kreuzten 41 % der Lehrkräfte an, die Förderaufgaben für die förderbedürftigen Schülerinnen und Schüler selbst zusammengestellt zu haben. Gleichzeitig gaben 91 % der Lehrkräfte an, dass alle förderbedürftigen Schülerinnen und Schüler die von der *Online Diagnose* individuell zusammengestellten Förderaufgaben bearbeiteten. Die Lehrkräfte setzten die Fördermaterialien auf unterschiedlichem Wege ein: 66 % der Schülerinnen und Schüler bearbeiteten die Förderaufgaben im Deutschunterricht, 60 % bearbeiteten die Förderaufgaben in einem speziellen Förderunterricht, 32 % bearbeiteten die Förderaufgaben als eine verpflichtende zusätzliche Hausaufgabe und 20 % bearbeiteten die Förderaufgaben als eine freiwillige zusätzliche Hausaufgabe. Auch eine Rückkopplung der Materialien in den Unterricht zur Besprechung der Lösungen der bearbeiteten Aufgaben führte mehr als die Hälfte der Lehrkräfte (62 %) durch.

Tabelle 5
Deskriptive Statistik der Items zur Anwendung der Fördermaterialien

Item	Itemtext	<i>M</i>	<i>SD</i>	Fehlende Angaben (Prozent)
Verfahrensweise bei der Anwendung Fördermaterialien:				
1.	Die gesamte Klasse bearbeitet die gleichen Förderaufgaben.	.04	0.20	6.7
2.	Die nicht-förderbedürftigen Schülerinnen und Schüler bearbeiten ebenfalls Förderaufgaben.	.71	0.45	6.7
3.	Alle förderbedürftigen Schülerinnen und Schüler bearbeiten die gleichen Förderaufgaben.	.10	0.30	8.0
4.	Die förderbedürftigen Schülerinnen und Schüler bearbeiten von mir individuell zusammengestellte Förderaufgaben.	.41	0.49	7.3
5.	Die förderbedürftigen Schülerinnen und Schüler bearbeiten die von der <i>Online Diagnose</i> individuell zusammengestellten Förderaufgaben.	.91	0.29	2.0

(Fortsetzung)

Item	Itemtext	<i>M</i>	<i>SD</i>	Fehlende Angaben (Prozent)
Rahmenbedingungen der Anwendung der Fördermaterialien:				
6.	Die Förderaufgaben werden im Deutschunterricht bearbeitet.	.66	0.47	6.7
7.	Die Förderaufgaben werden in einem speziellen Förderunterricht bearbeitet.	.60	0.49	4.0
8.	Die Bearbeitung der Förderaufgaben ist eine freiwillige zusätzliche Hausaufgabe.	.20	0.40	8.0
9.	Die Bearbeitung der Förderaufgaben ist eine verpflichtende zusätzliche Hausaufgabe.	.32	0.47	9.3
Einbindung der Anwendung der Fördermaterialien:				
10.	Die Lösungen der Förderaufgaben werden in der Schule besprochen.	.62	0.49	6.7

Anmerkung. *N* = 150; Skalierung: 0 (trifft nicht zu) bis 1 (trifft zu).

Technische Ausstattung der Schulen. Zur Durchführung des Testverfahrens *Online Diagnose* war eine angemessene technische Ausstattung der Schulen eine notwendige Voraussetzung. Eine ausreichende Anzahl an Computern mit einer entsprechenden Qualität der Internetverbindung sollte vorhanden sein. Aus diesem Grund wurden die Lehrkräfte um eine Auskunft zur Situation der technischen Ausstattung in der Schule gebeten. Die technische Ausstattung der Schule konnte von den Lehrkräften von 1 (sehr schlecht) bis 4 (sehr gut) bewertet werden. Die Ergebnisse dieser Befragung können Tabelle 6 entnommen werden. Die Mehrheit der befragten Lehrkräfte gab an, eine ausreichende Anzahl an Computern zur Verfügung zu haben (86 %). Ebenso wurde die Qualität der Computer von nahezu drei Viertel aller antwortenden Lehrkräfte positiv bewertet (74 %). Anders als bei den Computern war die Anzahl der verfügbaren Kopfhörer oder Lautsprecher (62 %) etwas weniger zufriedenstellend. In den meisten Schulen war laut Angaben der Lehrkräfte ein Großteil der Internetverbindungen qualitativ in Ordnung (70 %), auch die Druckmöglichkeiten an der Schule bewerteten die Lehrkräfte als eher gut bis sehr gut (80 %). Zusätzlich zu den genannten Prozentangaben zeigte der Modus eine insgesamt positive Beantwortung der Items. Die häufigsten Angaben fielen auf die Antwortkategorie 3 (eher gut).

Tabelle 6

Deskriptive Statistik der Items zur technischen Ausstattung von Schulen

Item	Itemtext	<i>M</i>	<i>SD</i>	Modus	Sehr / eher gut in %	Fehlende Angaben (Prozent)
1.	Anzahl der verfügbaren Computer	3.26	0.75	3	85.9	0.7
2.	Qualität der verfügbaren Computer	2.95	0.76	3	73.8	0.7
3.	Anzahl der verfügbaren Kopfhörer/ Lautsprecher	2.71	1.00	3	61.9	2.0
4.	Qualität der Internetverbindung	2.84	0.79	3	70.3	1.3
5.	Druckmöglichkeiten (z. B. für die Fördermaterialien)	3.14	0.85	3	80.4	1.3

Anmerkung. *N* = 150; Skalierung: 1 (sehr schlecht) bis 4 (sehr gut).

Mögliche Gründe für ein schwaches Testergebnis. Mögliche Gründe für ein schwaches Testergebnis konnten erstens auf externe Störungen während der Testdurchführung zurückgeführt werden, zweitens auf den Test selbst, oder drittens auf die Leistungen der Schülerinnen und Schülern (siehe Tabelle 7). Die externen Störungen wie Zeitdruck, Lärm oder technische Probleme, die Aufgabenstellungen des Tests sowie die mangelnde Motivation von Schülerinnen und Schülern sollten das Testergebnis möglichst wenig bzw. nicht beeinflussen. Der tatsächliche Förderbedarf von Schülerinnen und Schülern als Grund für ein schwaches Testergebnis sollte durch den Test aufgedeckt werden. Deshalb sollte dieses Item von den Lehrkräften als zutreffend bewertet werden (Abbildung 15).

Wie in Abbildung 15 aufgezeigt wird, bewertete ein Viertel der befragten Lehrkräfte den Lärm und den Zeitdruck als mögliche externe Störungen während der Testdurchführung (24 % und 29 %). Technische Probleme sahen rund 42 % der Lehrkräfte als hinderlich an. Unverständlichkeit der Aufgabenstellungen wurde von 44 % der Lehrkräfte als verantwortlich für ein schwaches Testergebnis angesehen. Weitere Gründe für ein schwaches Testergebnis lagen für 47 % der Lehrkräfte in der mangelnden Motivation von Schülerinnen und Schülern. Vor allem Sprachschwierigkeiten von Schülerinnen und Schülern mit Migrationshintergrund (81 %) gaben Lehrkräfte als Begründung für ein schwaches Testergebnis an. Außerdem gaben die Lehrkräfte den tatsächlichen Förderbedarf bei Schülerinnen und Schülern als häufigsten Grund für ein schlechtes Testergebnis an (97 %).

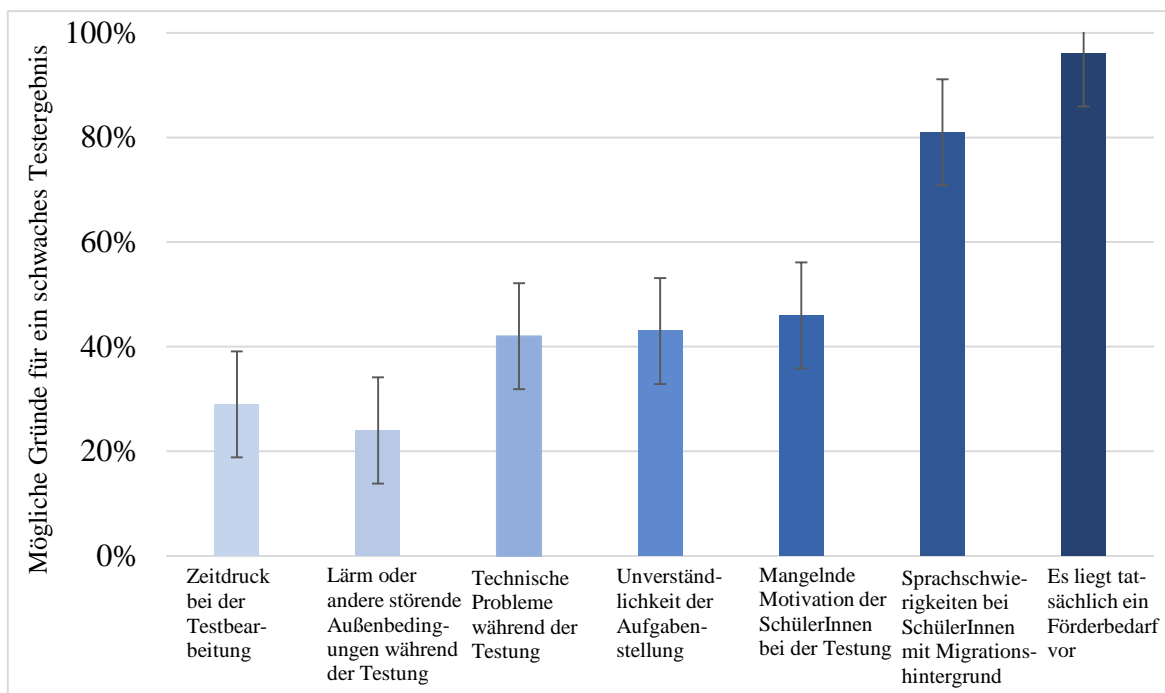


Abbildung 15. Mögliche Gründe für ein schwaches Testergebnis

Anhand des Modus in Tabelle 7 ist abzulesen, dass die Mehrheit der Items mit der Antwortkategorie 2 (trifft eher nicht zu) beantwortet wurden. Die beiden nachfolgenden Items bildeten im Antwortverhalten der Lehrkräfte eine Ausnahme: Der Modus des Items „Sprachschwierigkeiten bei Schülerinnen und Schülern mit Migrationshintergrund“ lag bei 3 (trifft eher zu) und der Modus des Items „Es liegt tatsächlich ein Förderbedarf vor“ lag bei 4 (trifft zu). Ein tatsächlich vorliegender Förderbedarf der Schülerinnen und Schülern war laut den Angaben der Lehrkräfte der Hauptgrund für ein schwaches Testergebnis. Dennoch waren laut den Angaben der Lehrkräfte unerwünschte Einflussfaktoren auf ein schwaches Testergebnis, wie technische Probleme oder die Aufgabenstellung, recht hoch.

Tabelle 7

Deskriptive Statistik der Items zu möglichen Gründen für ein schwaches Testergebnis

Item	Itemtext	<i>M</i>	<i>SD</i>	Modus	Trifft (eher) zu in %	Fehlende Angaben (Prozent)
Durchführungsmodalität:						
1.	Zeitdruck bei der Testbearbeitung.	2.08	0.94	2	29.1	4.0

(Fortsetzung)

Item	Itemtext	<i>M</i>	<i>SD</i>	Modus	Trifft (eher) zu in %	Fehlende Angaben (Prozent)
2.	Lärm oder andere störende Außenbedingungen während der Testung.	1.94	0.88	2	24.1	3.3
3.	Technische Probleme während der Testung.	2.41	1.08	2	42.1	3.3
Test:						
4.	Unverständlichkeit der Aufgabenstellung.	2.46	0.79	2	43.7	4.0
Schülerinnen und Schüler:						
5.	Mangelnde Motivation der Schülerinnen und Schüler bei der Testung	2.44	0.95	2	46.6	4.0
6.	Sprachschwierigkeiten bei Schülerinnen und Schülern mit Migrationshintergrund.	3.14	0.90	3	81.3	4.0
7.	Es liegt tatsächlich ein Förderbedarf vor.	3.44	0.59	4	96.5	4.0

Anmerkung. *N* = 150; Antwortkategorien: 1 (trifft nicht zu) bis 4 (trifft zu).

Unterstützung für die Lehrkräfte bei der Nutzung des Testverfahrens *Online Diagnose*. Die Nutzung des Diagnose- und Förderinstrumentes kann durch entsprechende Unterstützungsangebote erleichtert werden. Im Rahmen des Testverfahrens *Online Diagnose* wurden unterschiedliche Angebote bereitgestellt. Die Möglichkeit, eine Fortbildung zu besuchen, nutzten zehn der befragten Lehrkräfte. Häufiger genutzte Unterstützungsangebote werden hier unter dem Begriff „schriftliche Hilfen“ zusammengefasst. In Tabelle 8 sind die entsprechenden Angebote aufgeführt. Vorausgesetzt, dass die Lehrkräfte die schriftlichen Hilfen benutzt hatten, konnten sie diese von 1 (nicht hilfreich) bis 4 (hilfreich) bewerten. Die Kurzanleitung „erste Schritte mit der *Online Diagnose*“ wurde von 90 % der Lehrkräfte als überwiegend hilfreich bis hilfreich eingeschätzt. Darauf folgte der „Quick Tipp“, der von 89 % der Lehrkräfte als überwiegend hilfreich bzw. hilfreich eingeschätzt wurde. Dieser begleitete die Lehrkräfte bei der Bearbeitung des Online-Instrumentes. Er beschrieb den Inhalt und die Nutzungsmöglichkeit der geöffneten Internetseite.³⁴ Auch die Sammlung häufig gestellter Fragen (FAQ, 83 %) und das Nutzerhandbuch der zentralen Klassenverwaltung (75 %) sowie die Konzeption zum Testverfahren *Online Diagnose* (74 %) erschienen den

³⁴ Für weitere Informationen zum „Quick Tipp“ des Testverfahrens *Online Diagnose* siehe Kapitel 3.

Lehrkräften als (überwiegend) hilfreich. Wie die Tabelle 8 aufzeigt, lagen die fehlenden Angaben zwischen 30 % und 70.7 % und waren damit sehr hoch. Die gegebenen Antworten waren also nur auf einen kleineren Anteil der teilnehmenden Lehrkräfte zurückzuführen. Der Modus zeigte nochmals deutlich, dass der Anteil der antwortenden Lehrkräfte die Items hauptsächlich mit der Antwortkategorie 3 (eher hilfreich) bewertet hatten.

Tabelle 8

Deskriptive Statistik der Items zur Unterstützung für die Lehrkräfte bei der Nutzung der Online Diagnose

Item	Itemtext	<i>M</i>	<i>SD</i>	Modus	Genutzt & (überwiegend) hilfreich in %	Nicht genutzt in %	Nicht bekannt in %	Fehlende Angaben (Prozent)
1.	Kurzanleitung: Erste Schritte mit der <i>Online Diagnose</i>	3.20	0.70	3	89.5	14.1	12.0	5.3
2.	Quick Tipp (am rechten Rand der Onlineseiten)	3.13	0.80	3	89.1	44.1	15.4	9.3
3.	Sammlung häufig gestellter Fragen (FAQ)	3.06	0.83	3	83.3	51.8	8.8	8.7
4.	Nutzerhandbuch Zentrale Klassenverwaltung	2.89	0.75	3	75.0	47.5	20.9	7.3
5.	Konzeption <i>Online Diagnose</i>	2.90	0.78	3	74.2	33.6	21.2	8.7

Anmerkung. *N* = 150; Skalierung: 1 (genutzt & nicht hilfreich) bis 4 (genutzt & hilfreich).

5.2.2 Motivation bei der Durchführung des Haupttests und dem Einsatz der Fördermaterialien

Das Forschungsfeld zur menschlichen Motivation ist komplex. In aller Kürze formuliert, stellen Motivationstheorien Erklärungsversuche für die Ursachen und Gründe menschlichen Handelns dar (Krapp & Harsch, 2014). Eine Untersuchung des Einflusses von Motivation auf Leistungssituationen kann über das Erwartungs-Wert-Modell vorgenommen werden (vgl. Atkinson, 1957; Eccles, 1983, 2005). Dieses geht zum einen von der Erwartung aus, dass eine Handlung erfolgreich durchgeführt werden kann, und zum anderen davon, dass durch die Handlung ein erstrebenswertes Ziel erreicht werden kann. Motivation ist nicht direkt erfassbar bzw. kann nicht direkt gemessen werden, deshalb sind Forscherinnen und Forscher auf die Selbst- und Fremdeinschätzung angewiesen (Krapp & Harsch, 2014).

Die dem Erwartungs-Wert-Modell entnommenen „Subjektive Task Value“ (Eccles, 2005, S. 106) wurden für die vorliegende Studie anhand der Untersuchungen und Ausführungen von Eccles (1983, 2005) operationalisiert. Jacquelynne Eccles ging davon aus, dass der Wertaspekt von der Sache und von der Person abhängig ist. Hieraus schloss sie auf vier Bestandteile des Wertaspekts: die Wichtigkeit z.B. des Themas, der intrinsische Wert bzw. das Interesse, die Nützlichkeit für zukünftige Ziele und die Bereitschaft, einen bestimmten Aufwand bzw. relative Kosten aufzubringen. Im Interesse besserer Lesbarkeit wurden die Wertaspekte auf die folgenden vier Begriffe beschränkt: Wichtigkeit, intrinsischer Wert, Nützlichkeit und relative Kosten. Im Rahmen dieser Studie wurden die Lehrkräfte darum gebeten, die vier Bestandteile des Wertaspekts aus der Schülerperspektive und der eigenen Perspektive, also der Lehrerperspektive, einzuschätzen. Die Skalierung reichte von 1 (trifft nicht zu) bis 4 (trifft zu). Die vier Wertaspekte wurden in Bezug auf den Haupttest sowie in Bezug auf die Fördermaterialien erfasst. Pro Wertaspekt wurden zwei bis fünf Items entwickelt.

Für die Auswertung wurden die Skalen in zwei Abschnitten dargestellt: Der erste Abschnitt umfasst alle Bestandteile der Wertaspekte, die in Bezug auf den Haupttest erfasst wurden, und der zweite Abschnitt umfasst die Bestandteile der Wertaspekte in Bezug auf die Fördermaterialien. Innerhalb der beiden Abschnitte wurde zwischen den Skalen, die die Einschätzung durch die Lehrkräfte für die Schülerinnen und Schüler abbilden (Schülerperspektive in Einschätzung durch die Lehrkraft), und den Skalen zu den Lehrkräften (eigene Einschätzung) unterschieden. Insgesamt wurden vier exploratorische Faktorenanalysen mit den Items zu den folgenden Bereichen durchgeführt: Wertaspekte in Bezug auf den Haupttest aus Schülerperspektive (Einschätzung durch die Lehrkraft), Wertaspekte in Bezug auf den Haupttest aus Lehrerperspektive (eigene Einschätzung), Wertaspekte in Bezug auf die Fördermaterialien aus Schülerperspektive (Einschätzung durch die Lehrkraft), Wertaspekte in Bezug auf die Fördermaterialien aus Lehrerperspektive (eigene Einschätzung). Außerdem wurde der Mittelwert, die Standardabweichung, der Modus und die fehlenden Werte zu den Items und Skalen dargestellt sowie die interne Konsistenz und Trennschärfekoeffizienten.

Wertaspekte in Bezug auf den Haupttest aus Schülerperspektive (Einschätzung durch die Lehrkraft). In diesem Abschnitt werden die Items und Skalen in Bezug auf den Haupttest berichtet. Zunächst wurde eine exploratorische Faktorenanalyse mit den Items zu

den Wertaspekten aus Schülerperspektive durchgeführt. Hierbei zeigte sich nicht die erwartete Faktorenstruktur: Wichtigkeit, intrinsischer Wert, Nützlichkeit und relative Kosten. Die in Tabelle 9 dargestellten Faktorenladungen zeigten, dass die durch die Lehrkräfte eingeschätzte Wichtigkeit und der intrinsische Wert des Haupttests auf einen Faktor laden. Ein möglicher Grund für das Ergebnis könnte in der Einschätzung der Schülerperspektive durch die Lehrkräfte liegen. Für die Lehrkräfte als Außenstehende waren die Wichtigkeit und der intrinsische Wert möglicherweise nicht als unterschiedliche Wertaspekte wahrnehmbar. Ebenso kann Krapp (z.B. 2002) zitiert werden, der in seiner Theorie zur Interessenentwicklung keinen Unterschied zwischen der persönlichen Wichtigkeit und der intrinsischen Motivation von Aktivitäten annahm. Die entstandene Skala aus dem intrinsischen Wert und der Wichtigkeit wurde dennoch als intrinsischer Wert/Wichtigkeit bezeichnet. Die Items zur Nützlichkeit und den relativen Kosten des Haupttests luden wie erwartet auf zwei weiteren Faktoren. Alle drei Faktoren entsprachen dem Eigenwertkriterium > 1 . Die dreifaktorielle Lösung erklärte 66.38 % der Gesamtvarianz der Items.

Bei der Einschätzungen der Wertaspekte aus der Schülerperspektive (durch die Lehrkraft) gab die knappe Mehrheit der Lehrkräfte an, dass die Schülerinnen und Schüler den Haupttest zwar als wichtig ansahen (57 %), dass ihnen jedoch die Teilnahme am Haupttest nicht sehr viel bedeutete (74 %). Auch waren die Lehrkräfte der Ansicht, die Schülerinnen und Schüler würden keinen großen Wert auf den Haupttest legen (67 %). Laut den Lehrkräften freuten sich die Schülerinnen und Schüler auf den Haupttest (59 %), hatten Spaß bei der Bearbeitung (78 %) und führten den Haupttest mit Freude durch (62 %). Die Mehrheit der Lehrkräfte schätzte es für die Schülerinnen und Schüler als nützlich ein über ihre eigene Leistung informiert zu werden (62 %). Wiederum sagte die knappe Mehrheit aus, dass die Ergebnismrückmeldungen nicht hilfreich waren, um die eigene Leistung einschätzen zu können (57 %). Etwas mehr als die Hälfte der Lehrkräfte vermutete (56 %), dass die Schülerinnen und Schüler die Ergebnismrückmeldung des Haupttests als nützlich empfanden, da sie dadurch ihre Stärken und Schwächen erkennen konnten. Die Lehrkräfte gaben an, dass sich die Schülerinnen und Schüler durch den Haupttest nicht in Stress versetzt fühlten (82 %). Außerdem wurde von den Lehrkräften bestätigt, dass die Schülerinnen und Schüler weder befürchteten schlecht abzuschneiden (67 %) noch bei der Bearbeitung des Haupttests angespannt waren (61 %). Der Modus lag ausschließlich auf den mittleren Antwortkategorien 2 (trifft eher nicht zu) und 3 (trifft eher zu).

Tabelle 9

Deskriptive Statistik, Trennschärfenoeffizienten und Faktorenladungen zu den Wertaspekten in Bezug auf den Haupttest aus Schülerperspektive

Item	Itemtext	M	SD	Modus	Trifft (eher) zu in %	Fehlende Werte (Prozent)	r_{it}	Intrin. Wert/ Wicht.	Nütz- lich- keit	Rela- tive Kosten
1.	Die Schülerinnen und Schüler finden es wichtig am Haupttest teilzunehmen.	2.63	0.95	3	56.7	4.7	.59	.75	.07	.18
2.	Für die Schülerinnen und Schüler bedeutet die Teilnahme am Haupttest sehr viel.	2.06	0.80	2	26.3	6.0	.66	.81	.01	.35
3.	Die Schülerinnen und Schüler legen großen Wert auf den Haupttest.	2.16	0.80	2	32.7	4.0	.68	.80	.07	.17
4.	Die Schülerinnen und Schüler freuen sich auf den Haupttest.	2.59	0.92	3	59.3	3.3	.72	.81	-.09	-.20
5.	Die Schülerinnen und Schüler haben Spaß bei der Bearbeitung des Haupttests.	2.93	0.78	3	77.6	4.7	.64	.68	-.03	-.30
6.	Für die Schülerinnen und Schüler ist die Durchführung des Haupttests mit Freude verbunden.	2.69	0.81	3	62.4	6.0	.66	.68	.01	-.36
7.	Die Schülerinnen und Schüler finden es nützlich, über ihre eigene Leistung informiert zu werden.	2.71	0.73	3	62.3	4.7	.56	.20	.72	-.08
8.	Die Schülerinnen und Schüler finden die Ergebnisrückmeldung des Haupttests hilfreich für die Einschätzung ihrer Leistung.	2.40	0.69	2	42.9	6.7	.66	-.10	.90	.06
9.	Die Schülerinnen und Schüler finden die Ergebnisrückmeldung nützlich, um ihre Stärken und Schwächen zu erkennen.	2.39	0.69	2	45.8	5.3	.71	.02	.88	-.05
10.	Die Durchführung des Haupttests bedeutet für die Schülerinnen und Schüler großen Stress.	1.92	0.80	2	17.9	3.3	.50	.13	-.20	.86
11.	Die Schülerinnen und Schüler befürchten, im Haupttest schlecht abzuschneiden.	2.25	0.78	2	33.1	5.3	.57	-.03	.15	.67
12.	Bei der Bearbeitung des Haupttests ist den Schülerinnen und Schülern die Anspannung anzumerken.	2.30	0.88	2	38.5	4.7	.61	.03	.03	.76
	Eigenwert des Faktors							4.14	1.26	2.56
	Erklärte Varianz des Faktors in %							34.51	10.52	21.35

Anmerkung. $N = 150$; Skalierung: 1 (trifft nicht zu) bis 4 (trifft zu); Faktorladungen $> .40$ sind fettgedruckt.

Wertaspekte in Bezug auf den Haupttest aus Lehrerperspektive (eigene Einschätzung). Des Weiteren gaben die Lehrkräfte in einer eigenen Einschätzung Auskunft über die Wertaspekte in Bezug auf den Haupttest. Die jeweiligen Items sind in Tabelle 10 aufgeführt. Auch hier ergab die exploratorische Faktorenanalyse nicht die erwartete Faktorenstruktur: Wichtigkeit, intrinsischer Wert, Nützlichkeit und relative Kosten. Die Items der Wichtigkeit und der Nützlichkeit des Haupttests luden auf denselben Faktor. Möglicherweise waren die Formulierungen der Items nicht optimal, um zwischen den beiden Bestandteilen der Wertaspekte zu differenzieren. Ein weiterer Grund könnte sein, dass die Wichtigkeit und Nützlichkeit für die Lehrkräfte im Rahmen des Testverfahrens *Online Diagnose* gleichbedeutend war. Die Verknüpfung der empfundenen Nützlichkeit des Haupttests und dessen Wichtigkeit wurde als Nützlichkeit/Wichtigkeit bezeichnet. Der intrinsische Wert und die relativen Kosten ließen sich in zwei weitere Faktoren unterteilen. Das Item „Die Auseinandersetzung mit den Ergebnissen des Haupttests bedeutet für mich einen hohen Zeitaufwand“ (Item 22) lud auf alle drei Faktoren $> .40$. Aufgrund der theoretischen Zuordnung und der höchsten Ladung wurde das Item dennoch dem Faktor relative Kosten zugeordnet. Der Trennschärfekoeffizient des Items und die interne Konsistenz des Faktors waren angemessen. Bei den Faktoren zur eigenen Einschätzung der Wertaspekte in Bezug auf den Haupttest entsprach das Eigenwertkriterium > 1 . Die drei Faktoren erklärten 62.78 % der Gesamtvarianz der Items.

Die Lehrkräfte kreuzten im Fragebogen an, sich auf die Arbeit mit den Ergebnissen des Haupttests zu freuen (58 %). Sie notierten, die individuellen Ergebnisse einer jeden Schülerin und eines jeden Schülers spannend zu finden (80 %). Allerdings gab auch eine knappe Mehrheit der Lehrkräfte an (52 %), es hätte ihnen keinen Spaß bereitet, sich mit den Ergebnissen des Haupttests zu beschäftigen. Die Lehrkräfte bewerteten die Ergebnisse des Haupttest als wichtig, um die Schülerinnen und Schüler angemessen fördern zu können (80 %). Die individuellen Ergebnisse gaben die Lehrkräfte als bedeutsam für ihre Arbeit mit den Schülerinnen und Schülern an (63 %). Ebenso sagte die Mehrheit der Lehrkräfte aus, großen Wert auf die Ergebnisse des Haupttests zu legen (61 %). Sie waren der Ansicht, dass die Ergebnisse es ihnen ermöglichten, die Schwächen der Schülerinnen und Schüler zu identifizieren (82 %), und dass sie durch das Instrument in der Förderdiagnostik unterstützt wurden (90 %). Zudem waren die Ergebnisse für ihre Arbeit mit den Schülerinnen und Schülern nützlich (81 %). Der Stress bei der Durchführung des Haupttests (64 %) und der Aufwand an Unterrichtszeit (59 %) für die Durchführung des Haupttests schätzten die Lehrkräfte mehrheitlich als begrenzt ein. Eine knappe Mehrheit schätzte den logistischen Aufwand für

den Einsatz des Testinstrumentes als eher hoch bis hoch ein (55 %). Die Lehrkräfte gaben an, dass die Auseinandersetzung mit den Ergebnissen des Haupttests einen hohen Zeitaufwand bedeutete (73 %). Der Modus lag ausschließlich auf den mittleren Antwortkategorien 2 (trifft eher nicht zu) und 3 (trifft eher zu).

Tabelle 10

Deskriptive Statistik, Trennschärfenkoeffizienten und Faktorenladungen zu den Wertaspekten in Bezug auf den Haupttest aus Lehrerperspektive

Item	Itemtext	M	SD	Modus	Trifft (eher) zu in %	Fehlende Werte (Prozent)	r_{it}	Intrin- sischer Wert	Nütz. / Wicht	Rela- tive Kosten
1.	Ich freue mich immer darauf, mit den Ergebnissen des Haupttests zu arbeiten.	2.59	0.72	3	58.0	4.7	.53	.86	.00	.14
2.	Es macht mir Spaß, mich mit den Ergebnissen zu beschäftigen.	2.52	0.71	2	48.3	3.3	.59	.69	.08	.01
3.	Ich finde die individuellen Ergebnisse der Schülerinnen und Schüler spannend.	3.06	0.76	3	79.7	4.7	.46	.60	.23	-.06
4.	Für mich ist das Ergebnis des Haupttests wichtig, um die Schülerinnen und Schüler angemessen fördern zu können.	3.01	0.74	3	80.4	4.7	.75	.07	.82	-.07
5.	Die individuellen Ergebnisse des Haupttests sind in meiner Arbeit mit den Schülerinnen und Schülern bedeutsam.	2.71	0.75	3	63.2	4.0	.65	.37	.50	-.00
6.	Ich lege großen Wert auf die Ergebnisse des Haupttests.	2.72	0.80	3	61.3	5.3	.65	.39	.52	.09
7.	Die Ergebnisse ermöglichen es mir, die Schwächen der Schülerinnen und Schüler besser zu identifizieren.	3.06	0.69	3	82.0	4.0	.69	-.20	.97	-.01
8.	Die Ergebnisse des Haupttests unterstützen mich in der Förderdiagnostik.	3.15	0.65	3	89.7	3.3	.64	.20	.60	.00
9.	Die Ergebnisse des Haupttests sind für meine Arbeit mit den Schülerinnen und Schülern nützlich.	2.94	0.68	3	80.7	3.3	.71	.27	.61	-.07
10.	Die Durchführung des Haupttests bedeutet für mich großen Stress.	2.30	0.91	2	35.8	3.3	.60	.13	-.02	.84
11.	Der Einsatz des Haupttests bedeutet für mich einen hohen logistischen Aufwand.	2.60	0.89	3	55.3	4.7	.55	.14	-.07	.80
12.	Für die Durchführung des Haupttests geht zu viel Unterrichtszeit verloren.	2.42	0.94	2	40.7	3.3	.48	.07	-.29	.68
13.	Die Auseinandersetzung mit den Ergebnissen des Haupttests bedeutet für mich einen hohen Zeitaufwand.	2.95	0.78	3	72.9	4.0	.45	-.44	.40	.63
	Eigenwert des Faktors							2.11	5.04	1.02
	Erklärte Varianz des Faktors in %							16.19	38.75	7.83

Anmerkung. $N = 150$; Skalierung: 1 (trifft nicht zu) bis 4 (trifft zu); Faktorladungen $> .40$ sind fettgedruckt.

Ein Überblick über alle Skalen zu den Wertaspekten in Bezug auf den Haupttest zeigt, dass die Mittelwerte der Lehrerperspektive (eigene Einschätzungen) des intrinsischen Werts und der Nützlichkeit positiver ausfallen als die Mittelwerte zur Schülerperspektive (Einschätzung durch die Lehrkraft). Die Mittelwerte der relativen Kosten hingegen fallen in der Schülerperspektive (Einschätzung durch die Lehrkraft) positiver aus. Wie in Tabelle 11 dargestellt, weisen alle Skalen eine gute bis sehr gute interne Konsistenz (α) auf. Die in den Tabellen 9 und 10 dargestellten Trennschärfekoeffizienten (r_{it}) weisen ebenfalls gute Werte auf.

Tabelle 11

Deskriptive Statistik und interne Konsistenz der Skalen zu den Wertaspekten in Bezug auf den Haupttest aus Schülerperspektive und Lehrerperspektive

	<i>M</i>	<i>SD</i>	Fehlende Werte (Prozent)	α
Schülerperspektive (Einschätzung durch die Lehrkräfte)				
Intrinsischer Wert/Wichtigkeit	2.51	0.64	4.0	.86
Nützlichkeit	2.51	0.60	4.7	.80
Relative Kosten	2.16	0.66	4.7	.73
Lehrerperspektive (eigene Einschätzung)				
Intrinsischer Wert	2.72	0.58	3.3	.71
Nützlichkeit/Wichtigkeit	2.93	0.56	3.3	.88
Relative Kosten	2.57	0.65	4.0	.73

Anmerkung. $N = 150$; Skalierung: 1 (trifft nicht zu) bis 4 (trifft zu).

Wertaspekte in Bezug auf die Fördermaterialien aus Schülerperspektive (Einschätzung durch die Lehrkraft). Im Sinne des Erwartungs-Wert-Modells von Eccles (1983, 2005) wurden die Lehrkräfte auch in Bezug auf die Fördermaterialien um ihre Einschätzung der Wichtigkeit, des intrinsischen Werts, der Nützlichkeit und der relative Kosten gebeten. Wie anfangs bereits erläutert, wurden die Wertaspekte zu den Fördermaterialien aus der Schülerperspektive (Einschätzung durch die Lehrkraft) und aus der Lehrerperspektive (eigene Einschätzung) erfasst. Für die Schülerperspektive und die Lehrerperspektive wurden getrennte exploratorische Faktorenanalysen durchgeführt.

Zuerst wird die exploratorische Faktorenanalyse der Wertaspekte in Bezug auf die Fördermaterialien aus Schülerperspektive aufgeführt (Tabelle 12). Die Items in Bezug auf

die Fördermaterialien luden ebenso wie die Items des intrinsischen Werts und der Wichtigkeit in Bezug auf den Haupttest auf einen Faktor. Hier konnte, wie im Abschnitt zum Haupttest bereits erwähnt, vermutet werden, dass die Lehrkräfte im Rahmen der Einschätzung der Schülerperspektive nicht zwischen dem intrinsischen Wert und der Wichtigkeit unterscheiden konnten. Auch hierzu soll noch einmal die Interessentheorie von Krapp (2002) genannt werden, der annahm, dass es im Rahmen einer Aktivität keinen Unterschied zwischen der persönlichen Wichtigkeit und der intrinsischen Motivation gibt. Somit konnten die Ladungen auf einen Faktor begründet werden. Dieser Faktor wurde als intrinsischer Wert/Wichtigkeit bezeichnet. Die Items der beiden anderen Wertespekte ließen sich, wie in Tabelle 12 dargestellt, eindeutig in den Faktor Nützlichkeit und den Faktor relative Kosten trennen. Die Eigenwertkriterien der drei Faktoren entsprachen wie erwünscht einem Wert > 1 . Die drei Faktoren klärten 66.01 % der Gesamtvarianz der Items auf.

Die Lehrkräfte schätzten die Bedeutung der Fördermaterialien für die Schülerinnen und Schüler als gering ein (88 %). Rund 76 % der Lehrkräfte gaben an, dass die Schülerinnen und Schüler keinen großen Wert auf die Fördermaterialien legten. Laut den Lehrkräften freute sich ein geringer Anteil der Schülerinnen und Schüler (34 %) auf die Bearbeitung der Fördermaterialien. Auch der Aussage, die Bearbeitung der Fördermaterialien sei bei den Schülerinnen und Schülern mit Freude verbunden, stimmten nur 33 % der Lehrkräfte zu. Weniger als die Hälfte der Schülerinnen und Schüler hatte, nach Angabe der Lehrkräfte (41 %), Spaß bei der Bearbeitung der Fördermaterialien. Die knappe Mehrheit der Lehrkräfte (51 %) war der Ansicht, die Schülerinnen und Schüler würden die Fördermaterialien für die Verbesserung ihrer Leistung als nützlich empfinden. Die Bearbeitung der Fördermaterialien schätzte die knappe Mehrheit der Lehrkräfte aus Sicht der Schülerinnen und Schüler als nützlich ein (52 %). Sie waren außerdem der Meinung, dass die Schülerinnen und Schüler die Fördermaterialien als hilfreich empfanden (71 %). Die deutliche Mehrheit der Lehrkräfte gab an, dass die Bearbeitung der Fördermaterialien bei den Schülerinnen und Schülern keinen Stress auslöste (72 %). Auch die Befürchtung, die Aufgaben nicht lösen zu können, gab mehr als die Hälfte der Lehrkräfte als nicht zutreffend an (65 %). Die Lehrkräfte gaben außerdem an, dass den Schülerinnen und Schülern keine Anspannung anzumerken war (88 %). Der Modus lag ausschließlich auf den mittleren Antwortkategorien 2 (trifft eher nicht zu) und 3 (trifft eher zu).

Tabelle 12

Deskriptive Statistik, Trennschärfenoeffizienten und Faktorenladungen zu den Wertaspekten in Bezug auf die Fördermaterialien aus Schülerperspektive

Item	Itemtext	M	SD	Modus	Trifft (eher) zu in %	Fehlende Werte (Prozent)	r_{it}	Intrin. Wert/ Wicht.	Nütz- lich- keit	Rela- tive Kosten
1.	Den Schülerinnen und Schülern bedeutet die Bearbeitung der Fördermaterialien sehr viel.	1.87	0.65	2	12.0	6.0	.65	.86	-.06	.22
2.	Die Schülerinnen und Schüler legen großen Wert auf die Fördermaterialien.	2.06	0.69	2	23.9	5.3	.67	.79	.10	.25
3.	Die Schülerinnen und Schüler freuen sich auf die Bearbeitung der Fördermaterialien.	2.21	0.78	2	34.3	6.7	.77	.87	-.06	-.06
4.	Die Schülerinnen und Schüler haben Spaß bei der Bearbeitung der Fördermaterialien.	2.34	0.70	2	40.6	4.7	.78	.77	.06	-.28
5.	Die Bearbeitung der Fördermaterialien ist bei den Schülerinnen und Schülern mit Freude verbunden.	2.25	0.67	2	32.6	8.0	.76	.74	.06	-.23
6.	Die Schülerinnen und Schüler finden die Fördermaterialien nützlich, um ihre Leistung zu verbessern.	2.45	0.72	3	51.5	6.7	.53	.19	.62	-.05
7.	Die Schülerinnen und Schüler finden, dass ihnen die Bearbeitung der Fördermaterialien etwas bringt.	2.46	0.69	3	52.2	9.3	.63	.26	.72	.16
8.	Die Fördermaterialien sind für die Schülerinnen und Schüler hilfreich.	2.74	0.63	3	71.2	7.3	.58	-.18	.97	-.04
9.	Die Bearbeitung der Fördermaterialien bedeutet für die Schülerinnen und Schüler großen Stress.	2.13	0.84	2	28.4	6.0	.34	-.32	.10	.60
10.	Die Schülerinnen und Schüler befürchten die Aufgaben der Fördermaterialien nicht lösen zu können.	2.25	0.84	2	34.8	6.0	.33	.18	-.32	.72
11.	Bei der Bearbeitung der Fördermaterialien ist den Schülerinnen und Schülern die Anspannung anzumerken.	1.91	0.61	2	11.8	4.0	.25	.08	.24	.65
	Eigenwert des Faktors							4.55	1.52	1.19
	Erklärte Varianz des Faktors in %							41.38	13.84	10.80

Anmerkung. $N = 150$; Skalierung: 1 (trifft nicht zu) bis 4 (trifft zu); Faktorladungen $> .40$ sind fettgedruckt.

Wertaspekte in Bezug auf die Fördermaterialien aus Lehrerperspektive (eigene Einschätzung). Die in der Lehrerperspektive (eigene Einschätzung) erhobenen Items zu den Wertaspekten in Bezug auf die Fördermaterialien luden wie die Items in Bezug auf den Haupttest nicht auf vier Faktoren (Tabelle 13). Die Items zur Wichtigkeit und zur Nützlichkeit der Fördermaterialien luden auf einen Faktor. Auch hier konnte vermutet werden, dass die Formulierungen der Items nicht genau genug differenzierten bzw. die Lehrkräfte die Wichtigkeit und Nützlichkeit im Rahmen des Testverfahrens *Online Diagnose* als einen einzigen Aspekt wahrnahmen. Der Faktor wurde als Nützlichkeit/Wichtigkeit bezeichnet. Die weiteren Items luden wie erwartet auf die Faktoren intrinsischer Wert und relative Kosten. Mit Ausnahme des Items „Meiner Meinung nach geht für die Bearbeitung der Fördermaterialien die ohnehin schon knappe Unterrichtszeit verloren“, dieses lud negativ auf den Faktor Nützlichkeit. Da das Item inhaltlich nicht zu dem Faktor Nützlichkeit passte und die interne Konsistenz des Faktors nicht verbesserte, wurde es nicht in die Skala aufgenommen. Die drei Faktoren zu den Wertaspekten in der Lehrerperspektive entsprachen dem Eigenwertkriterium > 1 . Die Faktoren erklärten 68.02 % der Gesamtvarianz der Items.

Laut den Antworten der Lehrkräfte bereitete der Einsatz der Fördermaterialien der knappen Mehrheit Freude (56 %). Allerdings hatten die Lehrkräfte an der Überprüfung der Fördermaterialien keinen Spaß (75 %) und ebenso die Mehrheit der Lehrkräfte bewertete die Arbeit mit den Fördermaterialien als nicht spannend (60 %). Die Lehrkräfte sahen die Fördermaterialien für sich selbst als wichtig an, um die Schülerinnen und Schüler angemessen fördern zu können (73 %). Im Besonderen für die individuelle Förderung bewertete die knappe Mehrheit der Lehrkräfte die Fördermaterialien ebenfalls als bedeutsam (56 %). Zudem kreuzten 64 % der Lehrkräfte an, dass die Auseinandersetzung mit den Fördermaterialien ihre Arbeit bereicherte und sie großen Wert auf den Einsatz der Fördermaterialien legten (61 %). Die Lehrkräfte teilten im Fragebogen mit, dass die Fördermaterialien ihnen die individuelle Förderung von Schülerinnen und Schüler ermöglichte (83 %). Ebenso fühlten sich mehr als drei Viertel der Lehrkräfte bei der individuellen Förderung von Schülerinnen und Schülern unterstützt (80 %). Sie empfanden die Fördermaterialien für ihre Arbeit mit den Schülerinnen und Schülern als nützlich (77 %). Die Anwendung der Fördermaterialien bewertete die Mehrheit der Lehrkräfte weder als Stress (72 %) noch als hohen logistischen Aufwand (56 %). Auch war die Mehrheit der Lehrkräfte der Meinung, die für die Bearbeitung der Fördermaterialien aufgewendete Unterrichtszeit, sei keine verlorene Zeit (69 %).

Der Modus lag ausschließlich auf den mittleren Antwortkategorien 2 (trifft eher nicht zu) und 3 (trifft eher zu).

Tabelle 13

Deskriptive Statistik, Trennschärfenkoeffizienten und Faktorenladungen zu den Wertaspekten in Bezug auf die Fördermaterialien aus Lehrerperspektive

Item	Itemtext	M	SD	Modus	Trifft (eher) zu in %	Fehlende Werte (Prozent)	r_{it}	Intrin- sischer Wert	Nütz. / Wicht	Rela- tive Kosten
1.	Ich freue mich immer darauf, die Fördermaterialien mit den Schülerinnen und Schülern zu bearbeiten.	2.60	0.76	3	56.1	6.0	.61	.80	-.00	-.13
2.	Es macht mir Spaß, die bearbeiteten Fördermaterialien zu überprüfen.	2.05	0.78	2	24.5	4.7	.62	.92	-.09	.01
3.	Ich finde die Arbeit mit den Fördermaterialien spannend.	2.31	0.79	2	40.3	4.0	.66	.62	.25	-.06
4.	Für mich sind die Fördermaterialien wichtig, um die Schülerinnen und Schüler angemessen fördern zu können.	2.92	0.83	3	72.9	4.0	.74	.08	.80	.17
5.	Die individuellen Fördermaterialien sind für meine Arbeit mit den Schülerinnen und Schülern sehr bedeutsam.	2.60	0.81	3	56.4	5.3	.78	.22	.70	.02
6.	Die Auseinandersetzung mit den Fördermaterialien bereichert meine Arbeit.	2.67	0.82	3	64.1	5.3	.73	.33	.52	-.15
7.	Ich lege großen Wert auf den Einsatz der Fördermaterialien.	2.63	0.83	3	61.4	3.3	.75	.12	.78	.15
8.	Die Fördermaterialien ermöglichen es mir die Schülerinnen und Schüler individuell zu fördern.	3.07	0.77	3	83.4	4.0	.72	.27	.60	.04
9.	Die Fördermaterialien unterstützen mich in der individuellen Förderung der Schülerinnen und Schüler.	2.99	0.81	3	79.7	4.7	.78	.10	.81	.13
10.	Die Fördermaterialien sind für meine Arbeit mit den Schülerinnen und Schülern nützlich.	2.91	0.76	3	77.2	3.3	.71	-.19	.89	-.13
11.	Die Anwendung der Fördermaterialien bedeutet für mich großen Stress.	2.13	0.82	2	27.8	4.0	.46	.01	-.00	.84
12.	Der Einsatz der Fördermaterialien bedeutet für mich einen hohen logistischen Aufwand.	2.45	0.95	2	44.1	4.7	.46	-.21	.12	.79
13.	Meiner Meinung nach geht für die Bearbeitung der Fördermaterialien die ohnehin schon knappe Unterrichtszeit verloren.	2.21	0.82	2	31.0	3.3	.36	.31	-.75	.34
	Eigenwert des Faktors							1.42	6.40	1.02
	Erklärte Varianz des Faktors in %							10.95	49.21	7.85

Anmerkung. $N = 150$; Skalierung: 1 (trifft nicht zu) bis 4 (trifft zu); Faktorladungen $> .40$ sind fettgedruckt.

Ein Überblick über alle Skalen zu den Wertaspekten in Bezug auf die Fördermaterialien wird in Tabelle 14 dargestellt. Ein Vergleich der Skalen zeigt, dass die Mittelwerte der Nützlichkeit bzw. Nützlichkeit/Wichtigkeit in der Schüler- sowie Lehrerperspektive höher ausfallen als die Mittelwerte zum intrinsischen Wert/Wichtigkeit bzw. intrinsischen Wert. Die Mittelwerte zu den relativen Kosten deuten darauf hin, dass die Lehrkräfte diese für die Schülerinnen und Schüler geringer einschätzten als für sich selbst. Allerdings streuen die Angaben zu den relativen Kosten im Rahmen der Schülerperspektive (Einschätzung durch die Lehrkräfte) stärker. Fünf der extrahierten Faktoren zeigen jeweils, wie in Tabelle 14 dargestellt, eine akzeptable bis sehr gute interne Konsistenz (α). Die Skala relative Kosten in der Schülerperspektive (Einschätzung durch die Lehrkräfte) ergab einen nicht akzeptablen α -Wert. Wurden die entsprechenden Trennschärfekoeffizienten dieser Skala überprüft, fiel das Item „Bei der Bearbeitung der Fördermaterialien ist den Schülerinnen und Schülern die Anspannung anzumerken“ mit einem geringen Trennschärfekoeffizienten auf ($r_{it} = .25$). Würde das Item aus der Skala entfernt werden, sänke die interne Konsistenz weiter. Die relativen Kosten, die durch die Lehrkräfte für die Schülerinnen und Schüler eingeschätzt wurden, konnten mit den verwendeten Items nicht zufriedenstellend gemessen werden. Die Tabellen 12 und 13 stellen alle Trennschärfekoeffizienten (r_{it}) der einzelnen Items dar.

Tabelle 14

Deskriptive Statistik und interne Konsistenz der Skalen zu Wertaspekten in Bezug auf die Fördermaterialien aus Schülerperspektive und Lehrerperspektive

	<i>M</i>	<i>SD</i>	Fehlende Werte (Prozent)	α
Schülerperspektive (Einschätzung durch die Lehrkräfte)				
Intrinsischer Wert/Wichtigkeit	2.16	0.57	4.0	.89
Nützlichkeit	2.55	0.64	6.7	.75
Relativen Kosten	2.10	0.52	4.0	.48
Lehrerperspektive (eigene Einschätzung)				
Intrinsischer Wert	2.32	0.67	4.7	.79
Nützlichkeit/Wichtigkeit	2.82	0.64	4.0	.92
Relativen Kosten	2.28	0.76	4.0	.62

Anmerkung. $N = 150$; Skalierung: 1 (trifft nicht zu) bis 4 (trifft zu).

Vergleich zwischen den Wertaspekten des Haupttests und der Fördermaterialien.

Im Vergleich zwischen den Wertaspekten in Bezug auf den Haupttest und den Wertaspekten in Bezug auf die Fördermaterialien fiel Folgendes auf: Die Standardabweichungen zu den Skalen in Bezug auf den Haupttest reichten in der Schülerperspektive (Einschätzung durch die Lehrkraft) von 0.60 bis 0.66 und waren damit höher als die Standardabweichungen der Lehrerperspektive (eigene Einschätzung) 0.56 bis 0.65. In umgekehrter Form waren die Standardabweichungen der Skalen in Bezug auf die Fördermaterialien in der Schülerperspektive (Einschätzung durch die Lehrkraft) geringer 0.52 bis 0.64, als die Standardabweichungen in der Lehrerperspektive (eigene Einschätzung) 0.64 bis 0.76. Die Standardabweichungen werden in den Tabellen 11 und 14 berichtet. Hieraus könnte der Schluss gezogen werden, dass es den Lehrkräften leichter fiel, die Angaben zum Haupttest für sich selbst einzuschätzen. Umgekehrt könnte vermutet werden, dass es den Lehrkräften leichter fiel, die Angaben zu den Fördermaterialien für die Schülerinnen und Schüler einzuschätzen.

Bivariate Korrelationen der Wertaspekte in Bezug auf den Haupttest sowie der Wertaspekte in Bezug auf die Fördermaterialien. Die in Tabelle 15 dargestellten signifikanten Korrelationen der jeweiligen Wertaspekte reichen von $r = -.29$ ($p < .01$) bis $r = .68$ ($p < .001$). Die Zusammenhänge sind als mittel- bis stark einzuordnen (vgl. Spinath, 2011; Trautwein & Köller, 2003). Die beiden höchsten signifikanten Zusammenhänge ergaben sich im Rahmen der Lehrerperspektive zwischen den Wertaspekten intrinsischer Wert und Nützlichkeit/Wichtigkeit in Bezug auf den Haupttest und ebenso in Bezug auf die Fördermaterialien. Alle Korrelationen zwischen den Wertaspekten zum intrinsischen Wert bzw. intrinsischer Wert/Wichtigkeit und der Nützlichkeit bzw. Nützlichkeit/Wichtigkeit waren signifikant positiv. Auffallend war, dass sich ausschließlich Korrelationen mit den Wertaspekten der relativen Kosten als nicht signifikant zeigten. Möglicherweise waren die Items für die Befragung ungeeignet oder für die Lehrkräfte schwer zu beantworten. Die interne Konsistenz der Angaben zu den relativen Kosten war teilweise auch gering (z. B. relative Kosten der Fördermaterialien aus Schülerperspektive). Ebenfalls auffallend war, dass der Wertaspekt relative Kosten in Bezug auf den Haupttest aus Schülerperspektive signifikant positiv mit dem Wertaspekt Nützlichkeit in Bezug auf die Fördermaterialien aus Schülerperspektive korrelierte $r = .18$ ($p < .05$). Ebenso korrelierte der Wertaspekt relative Kosten bezüglich der Fördermaterialien in der Schülerperspektive signifikant positiv mit dem Wertaspekt Nützlichkeit in Bezug auf den Haupttest in der Schülerperspektive $r = .21$ ($p < .05$).

Eine höher eingeschätzte Nützlichkeit bei den Schülerinnen und Schülern im Haupttest hing also mit verstärkten relativen Kosten bei den Fördermaterialien zusammen. Ebenso hingen höhere relative Kosten beim Haupttest mit einer höher eingeschätzten Nützlichkeit für die Schülerinnen und Schüler bei den Fördermaterialien zusammen. Möglicherweise war hier die durch die Lehrkräfte eingeschätzte Bereitschaft, Kosten in Kauf zu nehmen, mit einer verstärkt empfundenen Nützlichkeit verknüpft. Diese Annahme müsste durch weitere Untersuchungen belegt werden. Alle weiteren signifikanten Korrelationen zwischen den relativen Kosten und dem intrinsischen Wert oder der Nützlichkeit waren negativ.

Tabelle 15

Bivariate Korrelationen der Wertaspekte in Bezug auf den Haupttest und der Wertaspekte in Bezug auf die Fördermaterialien

	1	2	3	4	5	6	7	8	9	10	11	12
Wertaspekte in Bezug auf den Haupttest aus Schülerperspektive (Einschätzung durch die Lehrkraft)												
1 Intrinsischer Wert/Wichtigkeit												
2 Nützlichkeit	.41**											
3 Relative Kosten	-.20*	.14										
Wertaspekte in Bezug auf den Haupttest aus Lehrerperspektive (eigene Einschätzung)												
4 Intrinsischer Wert	.22**	.41**	.16									
5 Nützlichkeit/Wichtigkeit	.18*	.42**	.15	.68**								
6 Relative Kosten	-.19*	-.25**	.16	-.19*	-.18*							
Wertaspekte in Bezug auf die Fördermaterialien aus Schülerperspektive (Einschätzung durch die Lehrkraft)												
7 Intrinsischer Wert/Wichtigkeit	.41**	.32**	.00	.39**	.32**	-.20*						
8 Nützlichkeit	.21*	.55**	.18*	.43**	.52**	-.22*	.53**					
9 Relative Kosten	-.09	.21*	.39**	.04	-.00	.18	-.20*	-.15				
Wertaspekte in Bezug auf die Fördermaterialien aus Lehrerperspektive (eigene Einschätzung)												
10 Intrinsischer Wert	.21*	.28**	.15	.60**	.43**	-.20*	.50**	.49**	-.11			
11 Nützlichkeit/Wichtigkeit	.20*	.27**	.12	.55**	.65**	-.24**	.40**	.54**	-.14	.67**		
12 Relative Kosten	-.06	-.01	.02	-.06	-.01	.42**	-.20*	-.18	.43**	-.27**	-.29*	

Anmerkung. $N = 146$; * $p < .05$; ** $p < .01$.

5.2.3 Subjektive Qualitätseinschätzung zum Testverfahren *Online Diagnose*

Im Rahmen der Befragung der Deutschlehrkräfte wurden diese nach ihrer Einschätzung der Qualität des Testverfahrens *Online Diagnose* gefragt. Es wird angenommen, dass eine positive Einschätzung der Qualität eines Testverfahrens zu einer Verstärkung der Nutzung führt. Umgekehrt könnte die Annahme bedeuten, dass durch eine negative Einschätzung der Qualität eines Testverfahrens durch die Lehrkräfte dasselbe weniger genutzt wird.³⁵ Aus diesem Grund ist die Untersuchung der Einschätzung der Lehrkräfte hinsichtlich der Qualität eines Testinstrumentes von Bedeutung. Die Einschätzung der Qualität wird als subjektive Qualitätseinschätzung bezeichnet. Für die subjektive Qualitätseinschätzung des Diagnose- und Förderinstrumentes *Online Diagnose* wurden sechs qualitätsrelevante Aspekte gewählt: Die Qualitätseinschätzung der Testaufgaben hinsichtlich ihrer Verständlichkeit (1) und Eignung (2), der Bezug der Testaufgaben zum Bildungsplan der Grundschule (3) und schließlich eine allgemeine Qualitätseinschätzung zu den Instrumenten Haupttest (4), Ergebnismeldung (5) und Fördermaterial (6).

Subjektive Qualitätseinschätzung der Testaufgaben. Die subjektive Qualitätseinschätzung der Testaufgaben wurde über zwei Merkmale erfasst: Zum einen über die Verständlichkeit der Testaufgaben für die Schülerinnen und Schüler und zum anderen über die Eignung der Testaufgaben, einen Förderbedarf feststellen zu können. Hierfür wurden Beispielaufgaben ausgewählt (z.B. Abbildung 16 bis 18; Bildungshaus, 2012), die jeden Aufgabentyp des Haupttests der *Online Diagnose* im Unterrichtsfach Deutsch repräsentieren.³⁶ Die Lehrkräfte wurden darum gebeten, die Verständlichkeit und Eignung dieser Beispielaufgaben für Schülerinnen und Schüler anhand einer Skala von 1 (überhaupt nicht verständlich bzw. überhaupt nicht geeignet) bis 4 (völlig verständlich bzw. völlig geeignet) einzuschätzen. Die Fragestellungen zur Einschätzung der subjektiven Qualität der Testaufgaben nahmen ausschließlich auf die leistungsmäßig schwächeren Schülerinnen und Schüler Bezug: „*Verständlichkeit*: Wird die Aufgabenstellung von leistungsmäßig schwächeren Schü-

³⁵ Der Zusammenhang der subjektiven Qualitätseinschätzung mit der Wirkung des Testverfahrens *Online Diagnose* wird in Teilstudie 2 untersucht.

³⁶ Die jeweils zuerst angezeigte Aufgabe eines Aufgabenblocks aus dem Testverfahren *Online Diagnose* wurde als Beispielaufgabe für den Fragebogen ausgewählt. Der im Testverfahren im Bereich „Hören“ gesprochene Text, wurde im Fragebogen symbolisch als Sprechblase eingefügt. Farbige Darstellungen wurden im Fragebogen in graustufen abgedruckt, deshalb wurden entscheidende farbige Merkmale dieser Darstellungen, wie z. B. eine rote Karte beim Fußballspiel, textlich erfasst.

lerinnen und Schülern verstanden?“ und „Eignung: Halten Sie diesen Aufgabentyp für geeignet, um einen Förderbedarf festzustellen?“. Die deskriptive Statistik ist in Tabelle 16 nachzulesen.

Online-Diagnose
Test unterbrechen

Lesen und Hören

Lies den Satz und klicke auf das passende Wort.

Mit einem

Radiergummi
Pinsel
Lineal
Bleistiftspitzer

 kann man malen.

?
Hilfe

→
Weiß ich nicht

Fortschritt:

Abbildung 16. Beispielaufgabe Radiergummi

Online-Diagnose
Test unterbrechen

Lesen und Hören

Klicke auf den Satz, der das Bild richtig beschreibt.

Der Schiedsrichter zieht die gelbe Karte, weil der Junge im blauen Trikot dem Jungen im weißen Trikot gegen das Schienbein tritt.

Der Schiedsrichter zieht die rote Karte, weil der Junge im weißen Trikot dem Jungen im blauen Trikot gegen den Oberschenkel tritt.

Der Schiedsrichter hält die rote Karte hoch, weil der Junge im blauen Trikot dem Jungen im weißen Trikot gegen das Schienbein tritt.

Der Junge hält die rote Karte hoch, weil der Schiedsrichter dem Jungen im weißen Trikot gegen das Schienbein tritt.

?
Hilfe

→
Weiß ich nicht

Fortschritt:

Abbildung 17. Beispielaufgabe Schiedsrichter

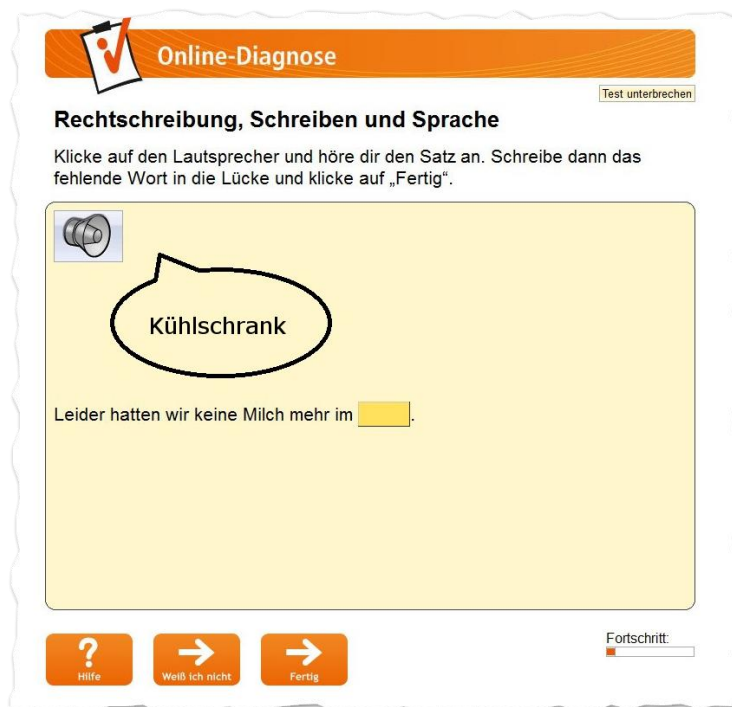


Abbildung 18. Beispielaufgabe Kühlschrank

Im Rahmen des Testverfahrens *Online Diagnose* wurden die Aufgaben, ähnlich dem Bildungsplan der Grundschule (2004) zum Unterrichtsfach Deutsch, in die Bereiche „Lesen und Hören“ und „Rechtschreibung, Schreiben und Sprache“ unterteilt. In der Datenauswertung wurde die Trennung der beiden Bereiche „Lesen und Hören“ sowie „Rechtschreibung, Schreiben und Sprache“ im Hinblick auf Verständlichkeit und Eignung berücksichtigt. Alle vier Skalen wiesen eine hohe interne Konsistenz auf: Verständlichkeit zum Bereich Lesen und Hören ($\alpha = .88$, $M = 3.33$, $SD = .49$, $N = 140$), Verständlichkeit zum Bereich Rechtschreibung, Schreiben und Sprache ($\alpha = .83$, $M = 3.32$, $SD = .47$, $N = 137$), Eignung zum Bereich Lesen und Hören ($\alpha = .84$, $M = 3.30$, $SD = .47$, $N = 132$) und Eignung des Bereichs Rechtschreibung, Schreiben und Sprache ($\alpha = .83$, $M = 3.44$, $SD = .47$, $N = 132$). Die Trennschärfekoeffizienten sind in Tabelle 16 abzulesen. Der Modus zeigt an, dass nur einmal die Antwortkategorie 2 (eher nicht verständlich bzw. eher nicht geeignet) häufiger als die anderen Antwortkategorien gewählt wurde. Weiterhin wurden ausschließlich die Antwortkategorien 3 (eher verständlich bzw. eher geeignet) und 4 (völlig verständlich bzw. völlig geeignet) häufiger als die anderen Antwortkategorien angekreuzt.

Tabelle 16

Deskriptive Statistik und Trennschärfeffizienten zu den Bereichen Lesen und Hören sowie Rechtschreibung, Schreiben und Sprache

Item		Verständlichkeit						Eignung					
		<i>M</i>	<i>SD</i>	Modus	Völlig/ e-her ver-ständlich/ ge-eignet	Fehlende Angaben (Prozent)	<i>r_{it}</i>	<i>M</i>	<i>SD</i>	Modus	Völlig/ e-her ver-ständlich/ ge-eignet	Fehlende Angaben (Prozent)	<i>r_{it}</i>
Lesen und Hören													
1.	Radiergummi	3.64	0.55	4	98.0	2.0	.64	3.27	0.69	3	90.4	2.7	.45
2.	Lesewettbewerb	3.10	0.76	3	78.8	2.7	.63	3.17	0.81	3	81.1	1.3	.58
3.	München	3.03	0.81	3	73.3	2.7	.60	3.08	0.81	3	76.7	2.7	.59
4.	Johanna	3.33	0.72	4	87.8	2.0	.47	3.38	0.69	4	91.0	4.0	.52
5.	Schiedsrichter	3.27	0.80	4	82.3	2.0	.56	3.33	0.73	4	86.1	4.0	.51
6.	Säugetiere	3.08	0.81	3	76.8	2.7	.61	3.20	0.79	3 ^a	81.6	5.3	.58
7.	Nagel	3.60	0.67	4	91.1	3.3	.57	3.42	0.70	4	89.5	4.7	.47
8.	Giraffen	3.52	0.68	4	92.5	2.7	.53	3.45	0.70	4	92.3	4.7	.41
9.	Sonntag	3.41	0.76	4	87.6	3.3	.63	3.38	0.73	4	86.7	4.7	.44
10.	Papagei	3.47	0.66	4	93.8	3.3	.65	3.41	0.68	4	90.3	3.3	.58
11.	Hund	3.08	0.84	3	78.1	2.7	.64	3.19	0.68	4	81.9	4.7	.53
Rechtschreibung, Schreiben, Sprache													
1.	Kühlschrank	3.70	0.52	4	98.6	3.3	.53	3.59	0.60	4	95.8	4.7	.46
2.	Zug	3.49	0.67	4	91.7	3.3	.49	3.51	0.65	4	93.0	4.7	.44
3.	Schule	3.37	0.72	4	88.4	2.0	.56	3.54	0.60	4	95.8	4.0	.56
4.	Museum	2.67	0.89	2	54.1	4.0	.52	2.85	0.95	3	66.7	4.0	.51
5.	Satzglied	2.83	0.87	3	64.4	2.7	.53	3.30	0.72	3	88.8	5.3	.66
6.	Rose	3.68	0.56	4	96.5	3.3	.66	3.64	0.67	4	93.1	3.3	.52
7.	Adjektiv	3.40	0.72	4	90.3	4.0	.55	3.61	0.57	4	97.2	4.0	.59
8.	Geburtstag	3.37	0.74	4	88.4	2.7	.54	3.51	0.62	4	95.0	5.3	.61
9.	Finale	3.41	0.75	4	86.7	4.7	.55	3.45	0.69	4	93.0	4.7	.56

Anmerkung. *N* = 150; Skalierung: 1 (überhaupt nicht verständlich bzw. überhaupt nicht geeignet) bis 4 (völlig verständlich bzw. völlig geeignet); ^a Es sind mehrere Modi vorhanden, der kleinste Wert wurde angegeben.

Bezug zum Bildungsplan der Grundschule. Das Testverfahren *Online Diagnose* sollte den Leistungsstand von Schülerinnen und Schülern beim Übergang in die Haupt-, Werkreal- oder Gemeinschaftsschule erfassen. Die Aufgabenstellungen orientierten sich deshalb am Bildungsplan der Grundschule (2004). Im Rahmen der Befragung wurden die Deutschlehrkräfte um ihre Einschätzung gebeten, inwieweit der Haupttest des Testverfahrens *Online Diagnose* die Inhalte des Bildungsplans der Grundschule abdeckt. Die drei Items werden in Tabelle 18 aufgeführt. Die interne Konsistenz der aus den Items gebildeten Skala war angemessen ($\alpha = .77$, $M = 3.15$, $SD = .49$, $N = 142$). Die in Tabelle 18 aufgeführten Trennschärfekoeffizienten (r_{it}) zeigten ebenfalls eine gute Passung der Items mit der Skala. Die Mehrheit der Lehrkräfte war gegenüber dem Bezug des Testverfahrens *Online Diagnose* zum Bildungsplan der Grundschule sehr positiv eingestellt. Insgesamt waren 94 % der Lehrkräfte der Meinung, der Haupttest würde einen klaren Bezug zum Bildungsplan der Grundschule aufweisen. Von den antwortenden Lehrkräften gaben 89 % an, der Haupttest würde alle wichtigen Bereich des Bildungsplans abdecken. Die Aufgaben seien geeignet, die geforderten Kompetenzen des Bildungsplans zu überprüfen, dieser Ansicht waren 91 % der Lehrkräfte. Laut den Lehrkräften ist dem Instrument *Online Diagnose* der Bezug zum Bildungsplan der Grundschule gut gelungen.

Subjektive Qualitätseinschätzung des Haupttests. Zur Qualität des Haupttests wurden den Lehrkräften zwei Aussagen vorgelegt. Diese werden in Tabelle 18 aufgeführt. Die daraus gebildete Skala wies eine geringe interne Konsistenz auf ($\alpha = .53$, $M = 2.89$, $SD = .58$, $N = 145$). Die geringe interne Konsistenz der Skala Qualität des Haupttests ist aufgrund der wenigen Items akzeptabel (Lienert, 1969). Die in Tabelle 18 abgebildeten Trennschärfekoeffizienten (r_{it}) waren gering, jedoch akzeptabel. Die Mehrheit der Lehrkräfte (61 %) war der Ansicht, dass der Haupttest die zentralen Fähigkeiten für eine erfolgreiche Mitarbeit im Deutschunterricht abprüft. Ebenfalls sahen sie den Haupttest der *Online Diagnose* als ein gutes Instrument an, um Förderbedarfe feststellen zu können (90 %). Die Qualität des Haupttests aus dem Testverfahren *Online Diagnose* nahmen die Lehrkräfte positiv wahr.

Subjektive Qualitätseinschätzung der Ergebnisrückmeldungen. Die Zusammenhörigkeit der Items zur Qualität der Ergebnisrückmeldung wurde anhand einer exploratorischen Faktorenanalyse überprüft (Tabelle 17). Da zum einen eine allgemeine Einschätzung

zu den Ergebnisrückmeldungen und zum anderen eine Einschätzung zur Darstellung der Ergebnisrückmeldungen erfasst wurde. Die sechs Items luden wie erwartet auf zwei Faktoren: Qualität der Ergebnisrückmeldungen und Qualität der Darstellung der Ergebnisrückmeldungen. Ein Item lud auf beide Faktoren $> .40$. Dieses wird aufgrund der inhaltlichen Zugehörigkeit sowie der höheren Ladung dem zweiten Faktor Qualität der Darstellung der Ergebnisrückmeldungen zugeordnet. Das Eigenwertkriterium der beiden Faktoren beträgt einen Wert > 1 . Die zwei Faktoren klären insgesamt 71.85 % der Gesamtvarianz der Items auf. Die interne Konsistenz (α) beider Skalen ist angemessen: Qualität der Ergebnisrückmeldungen ($\alpha = .75$, $M = 3.12$, $SD = .62$, $N = 148$) und Qualität der Darstellung der Ergebnisrückmeldungen ($\alpha = .80$, $M = 3.04$, $SD = .65$, $N = 148$). Auch die in Tabelle 18 aufgeführt Trennschärfekoeffizienten (r_{it}) sind sehr gut.

Die Lehrkräfte nahmen die Ergebnisrückmeldungen als verständlich wahr (89 %). Auch gaben die Lehrkräfte an, dass die Ergebnisrückmeldungen hinsichtlich ihres Umfangs angemessen waren (76 %). Die Informationsmenge zu den jeweiligen Schülerinnen und Schülern empfanden die Lehrkräfte als ausreichend (84 %). Die Items zur Skala Qualität der Darstellung der Ergebnisrückmeldungen wurden ähnlich positiv bewertet. Knapp drei Viertel der Lehrkräfte gaben an, dass die Grafiken selbsterklärend waren (74 %). Mehr als drei Viertel der Lehrkräfte sagten aus, dass die Ergebnisrückmeldungen übersichtlich waren (78 %). Insgesamt zeigten sich die Lehrkräfte mit den Darstellungen der Testergebnisse zufrieden (81 %).

Tabelle 17
Faktorenladungen einer exploratorischen Faktorenanalyse mit Promax Rotation zur Qualität der Ergebnisrückmeldungen

Item	Itemtext	Qualität der Ergebnisrückmeldungen	Qualität der Darstellung der Ergebnisrückm.
1	Die Informationen zu den jeweiligen Schülerinnen und Schülern sind ausreichend.	.89	-.14
2	Den Umfang der Ergebnisrückmeldungen finde ich angemessen.	.84	-.11
3	Die Ergebnisrückmeldungen sind verständlich.	.71	.23
1	Die Grafiken sind selbsterklärend.	-.19	.96
2	Die Ergebnisrückmeldungen sind sehr übersichtlich.	.02	.92
3	Mit der Darstellung der Testergebnisse bin ich insgesamt zufrieden.	.43	.51

(Fortsetzung)

Item	Itemtext	Qualität der Ergebnisrück- meldungen	Qualität der Darstellung der Ergebnisrückm.
	Eigenwert des Faktors	3.23	1.08
	Erklärte Varianz des Faktors in %	53.87	17.99

Anmerkung. Skalierung: 1 (trifft nicht zu) bis 4 (trifft zu). Extraktionsmethode: Hauptkomponentenanalyse; Rotationsmethode: Promax mit Kaiser-Normalisierung; Faktorladungen > .40 sind fettgedruckt.

Subjektive Qualitätseinschätzung der Fördermaterialien. Die Lehrkräfte wurden hinsichtlich der Fördermaterialien um ihre allgemeine Einschätzung der Qualität gebeten. Die beiden Items sind in Tabelle 18 aufgeführt. Die interne Konsistenz der Skala aus den beiden Items war entsprechend gut ($\alpha = .74$, $M = 2.83$, $SD = .64$, $N = 145$). Ebenso waren die in Tabelle 18 aufgezeigten Trennschärfekoeffizienten (r_{it}) der Items angemessen. Die Mehrheit der Lehrkräfte bezeichnete die Aufgaben der Fördermappe als gut (80 %). Ebenso waren die Lehrkräfte der Meinung, die Fördermappe wäre ein gutes Instrument, um Defizite zu reduzieren (68 %).

Tabelle 18

Deskriptive Statistik und Trennschärfeffizienten der subjektiven Qualitätseinschätzung der Instrumente und Fördermaterialien

Skalen	Item	Itemtext	<i>M</i>	<i>SD</i>	Modus	Trifft (eher) zu in %	Fehlende Angaben (Prozent)	<i>r_{it}</i>
Bezug zum Bildungsplan der Grundschule	1.	Der Haupttest weist einen klaren Bezug zum Bildungsplan der Primarstufe auf.	3.21	0.58	3	94.3	6.7	.50
	2.	Der Haupttest deckt alle wichtigen Bereiche des Bildungsplans ab.	3.12	0.62	3	89.3	6.7	.64
	3.	Die Aufgaben sind geeignet, die im Bildungsplan geforderten Kompetenzen zu überprüfen	3.12	0.59	3	90.8	6.0	.67
Qualität des Haupttests	1.	Der Haupttest ist ein gutes Instrument, um Förderbedarf festzustellen.	3.14	0.62	3	89.6	4.0	.37
	2.	Der Haupttest prüft die zentralen Fähigkeiten für eine erfolgreiche Mitarbeit im Deutschunterricht ab.	2.65	0.77	3	61.1	4.0	.37
Qualität der Ergebnismeldungen	1.	Die Informationen zu den jeweiligen Schülerinnen und Schülern sind ausreichend.	3.21	0.77	3	84.4	2.0	.65
	2.	Den Umfang der Ergebnismeldungen finde ich angemessen.	2.95	0.83	3	75.6	2.0	.54
	3.	Die Ergebnismeldungen sind verständlich.	3.22	0.67	3	89.1	2.0	.65
Qualität der Darstellung der Ergebnismeldungen	1.	Die Grafiken sind selbsterklärend.	3.08	0.85	3 ^a	74.4	1.3	.59
	2.	Die Ergebnismeldungen sind sehr übersichtlich.	3.00	0.74	3	78.4	1.3	.80
	3.	Mit der Darstellung der Testergebnisse bin ich insgesamt zufrieden.	3.04	0.71	3	81.1	1.3	.57
Qualität der Fördermaterialien	1.	Die Fördermappe enthält gute Aufgaben.	2.95	0.70	3	79.9	4.0	.59
	2.	Die Fördermappe ist ein gutes Instrument, um Defizite zu reduzieren.	2.70	0.75	3	68.1	4.6	.59

Anmerkung. *N* = 150; Skalierung: 1 (trifft nicht zu) bis 4 (trifft zu); ^a Es sind mehrere Modi vorhanden, der kleinste Wert wurde angegeben.

Bivariate Korrelationen der Skalen zur subjektiven Qualitätseinschätzung. Die vier Skalen zur subjektiven Qualitätseinschätzung der Testaufgaben zeigten erwartungsgemäß signifikante, starke Zusammenhänge auf (Tabelle 19). Besonders stark korrelierte die Einschätzung der Verständlichkeit und Eignung des Bereichs Lesen und Hören ($r = .81$; $p < .001$). Auch die Einschätzung der Verständlichkeit und Eignung des Bereichs Rechtschreibung, Schreiben und Sprache ($r = .80$; $p < .001$) korrelierten besonders stark. Die weiteren Skalen der subjektiven Qualitätseinschätzung des Bezugs zum Bildungsplan der Grundschule, des Haupttests, der Ergebnisrückmeldungen, der Darstellung der Ergebnisrückmeldungen und der Fördermaterialien wiesen schwache bis mittlere signifikante Zusammenhänge auf ($r = .20$; $p < .01$ bis $r = .51$; $p < .001$). Die folgenden Skalen wiesen die höchsten Zusammenhänge auf: Bezug zum Bildungsplan der Grundschule und Qualität des Haupttests ($r = .46$; $p < .001$), Qualität des Haupttests und Qualität der Fördermaterialien ($r = .46$; $p < .001$) sowie Qualität der Ergebnisrückmeldungen und Qualität der Darstellung der Ergebnisrückmeldungen ($r = .51$; $p < .001$). Auffallend war, dass zwischen den beiden Skalen Qualität der Ergebnisrückmeldungen und der Qualität der Fördermaterialien kein signifikanter Zusammenhang bestand. Hier könnte die vage Annahme aufgestellt werden, dass die Lehrkräfte die Ergebnisrückmeldungen nicht in Verbindung mit den Fördermaterialien brachten. Dies wäre für eine gezielte individuelle Förderung jedoch erforderlich. Deshalb würden weitere Untersuchungen zu dieser Annahme wichtige Informationen liefern.

Tabelle 19

Bivariate Korrelationen der Skalen zur subjektiven Qualitätseinschätzung

	1	2	3	4	5	6	7	8	9
1. Verständlichkeit zum Bereich Lesen und Hören									
2. Verständlichkeit zum Bereich Rechtschreibung, Schreiben und Sprache	.73**								
3. Eignung zum Bereich Lesen und Hören	.81**	.59**							
4. Eignung zum Bereich Rechtschreibung, Schreiben und Sprache	.66**	.80**	.72**						
5. Bezug zum Bildungsplan der Grundschule	.29**	.33**	.32**	.32**					
6. Qualität des Haupttests	.22*	.20*	.23**	.17*	.46**				
7. Qualität der Ergebnisrückmeldungen	.06	.16*	.11	.13	.21*	.20*			
8. Qualität der Darstellung der Ergebnisrückmeldungen	.27**	.21**	.19*	.21**	.20**	.22**	.51**		
9. Qualität der Fördermaterialien	.15	.12	.17	.17*	.33**	.46**	.13	.18	

Anmerkung. $N = 148$; * = $p < .05$; ** $p < .01$.

5.2.4 Prozessschritte der Nutzung von Ergebnisrückmeldungen

Ein für die Nutzung von Ergebnisrückmeldungen umfassendes und in der deutschsprachigen Forschung häufig verwendetes Modell ist das prozedurale Zyklusmodell der Rückmeldung (Helmke, 2004; Helmke & Hosenfeld, 2005; Hosenfeld & Groß Ophoff, 2007)³⁷. Dieses Modell bildet die theoretische Grundlage in den Teilstudien 2 und 3. Eine Modellüberprüfung ist im Rahmen der vorliegenden Arbeit nicht vorgesehen (vgl. Maier, 2008a). In den nachfolgenden Abschnitten werden Skalen zu den vier Prozessschritten des theoretischen Zyklusmodells der Rückmeldung Rezeption, Reflexion, Aktion und Evaluation operationalisiert.

Prozessschritt Rezeption. Helmke und Hosenfeld (2005) definierten den Prozessschritt *Rezeption* als die technische Übermittlung der Ergebnisse, die Wahrnehmung und das Verständnis der Ergebnisrückmeldung. In Bezug auf das Diagnose- und Förderinstrument *Online Diagnose* können die drei Kriterien wie folgt erläutert werden: Zunächst geschah die technische Übermittlung der Ergebnisse internetbasiert über den Zugriff der Lehrkraft auf das entsprechende Klassenkonto. Die Lehrkräfte hatten nach der Durchführung der Tests jederzeit die Möglichkeit, die Ergebnisse abzurufen. Von Seiten des Diagnose- und Förderinstruments *Online Diagnose* dürfte es zu keinen zeitlichen Verzögerungen der Ergebnisübermittlungen gekommen sein, deshalb wurde hier auf eine weitere Untersuchung verzichtet.³⁸

Die von Helmke und Hosenfeld (2005) aufgeschlüsselten Kriterien Wahrnehmung und Verständnis wurden für die vorliegende Studie differenzierter definiert: Durch die computerbasierte Auswertung, der von den Schülerinnen und Schülern ausgefüllten Tests und die problemlose Übermittlung der Ergebnisse wurden viele Informationen zur Leistung einzelner Schülerinnen und Schülern produziert. Deshalb interessierte in dieser Studie, wie positiv bzw. negativ die Lehrkräfte den Zeitaufwand empfanden, der für die Wahrnehmung der Ergebnisrückmeldungen benötigt wurde. Die Befragung sollte aufdecken, ob der Zeitaufwand für die Nutzung der Ergebnisrückmeldungen als angemessen empfunden wurde. Das von Helmke und Hosenfeld (2005) aufgeführte Kriterium Verständnis kann ebenso mit den Begriffen Interpretation und Verstehen beschrieben werden. Da sich die vorliegende Studie

³⁷ Nähere Ausführungen zum Zyklusmodell der Rückmeldung sind in Kapitel 2.3 aufgeführt.

³⁸ Weitere Angaben zur technischen Ausstattung der Schulen und zu möglichen technischen Problemen bei der Übermittlung der Ergebnisse können in diesem Kapitel unter Abschnitt 5.2.1 nachgelesen werden.

auf eine Selbsteinschätzung der Lehrkräfte bezog, war eine objektive Erfassung der Interpretation und des Verstehens von Ergebnisrückmeldungen nicht möglich. Aus diesem Grund erschien die Erfassung der Selbstwirksamkeit der Lehrkräfte im Umgang mit den Ergebnisrückmeldungen des Testverfahrens *Online Diagnose* als geeignet, eine zuverlässige Aussage hierüber zu gewinnen. Von selbstwirksamem Umgang mit den Ergebnisrückmeldungen kann zwar nicht auf die Kenntnisse der Lehrkräfte über die Ergebnisrückmeldungen geschlossen werden, ein solcher spricht jedoch für einen souveränen Umgang der Lehrkräfte mit den Ergebnisrückmeldungen (z.B. Edelstein, 2002; Krapp & Ryan, 2002; Schwarzer & Jerusalem, 2002). Die Lehrkräfte wurden deshalb gefragt, ob sie persönlich einen Erfolg im Umgang mit den Ergebnisrückmeldungen erwarteten. Es wurde davon ausgegangen, dass in diesem selbstwirksamen Umgang mit den Ergebnisrückmeldungen das subjektive Verständnis impliziert sei. Von Lehrkräften, die selbstwirksam mit den Ergebnisrückmeldungen umgingen, wurde ein höheres Engagement erwartet (Schmitz & Schwarzer, 2000; Edelstein, 2002).

Zeitaufwand für die Nutzung der Ergebnisrückmeldungen. Wie einleitend bereits berichtet, wurden die Lehrkräfte um ihre Auskunft zum Zeitaufwand für die Nutzung der Ergebnisrückmeldungen gebeten (Tabelle 20). Die interne Konsistenz der Skala war zufriedenstellend ($\alpha = .69$, $M = 2.45$, $SD = .74$, $N = 148$). Die in Tabelle 20 aufgezeigten Trennschärfekoeffizienten (r_{it}) stellen angebrachte Werte dar. Die Hälfte der Lehrkräfte (51 %) kreuzte an, es wäre ihnen zeitlich nicht möglich gewesen, die Ergebnisrückmeldungen von allen Schülerinnen und Schülern zu lesen. 65 % der Lehrkräfte empfanden das Durcharbeiten aller Ergebnisrückmeldungen als zu zeitaufwendig. Rund 45 % der Lehrkräfte gaben an, die Ergebnisrückmeldungen dann besonders genau zu lesen, wenn sie selbst einen Förderbedarf bei den Schülerinnen und Schülern sahen. Dennoch war die Mehrheit der Lehrkräfte der Meinung, die Ergebnisrückmeldungen seien nicht zu detailliert (59 %).

Selbstwirksamkeit im Umgang mit den Ergebnisrückmeldungen. Die in Tabelle 20 aufgelisteten Items zur Skala Selbstwirksamkeit im Umgang mit den Ergebnisrückmeldungen zeigen gute Trennschärfekoeffizienten (r_{it}). Auch die interne Konsistenz ist zufriedenstellend ($\alpha = .74$, $M = 2.99$, $SD = .53$, $N = 144$). Der Großteil der Lehrkräfte notierte, dass ihnen die Interpretation der Ergebnisse des Haupttests leicht fiel (82 %). Sie gaben ebenso an, dass sie gut mit den Ergebnissen zurechtkamen (84 %). Die Mehrheit der antwortenden

Lehrkräfte (67 %) war überzeugt davon, dass ihnen die Arbeit mit den Ergebnisrückmeldungen lag. Ebenfalls war die Mehrheit der Lehrkräfte (83 %) zuversichtlich, dass ihnen eine angemessene Interpretation gelungen war.

Für die Skalenbildung der beiden Skalen wurden die negativ formulierten Items rekodiert. Dies betraf insgesamt zwei Items. Der Modus bezog sich beinahe ausschließlich auf die Antwortkategorie 3 (trifft eher zu), mit einer Ausnahme die auf der Antwortkategorie 2 (trifft eher nicht zu) liegt.

Tabelle 20
Deskriptive Statistik und Trennschärfeeffizienten der Items zum Prozessschritt Rezeption

Skala	Item	Itemtext	<i>M</i>	<i>SD</i>	Modus	Trifft (e-her) zu in %	Fehlende Angaben (Prozent)	<i>r_{it}</i>
Zeitaufwand für die Nutzung der Ergebnisrückmeldungen	1.	Es ist mir zeitlich nicht möglich, die Ergebnisrückmeldungen von allen Schülerinnen und Schülern ausführlich zu lesen.	2.49	1.06	3	50.6	1.3	.65
	2.	Das Durcharbeiten aller Ergebnisrückmeldungen ist zu zeitaufwendig.	2.84	0.99	3	65.3	2.0	.49
	3.	Ich lese die Ergebnisrückmeldungen der Schülerinnen und Schüler nur dann besonders genau, wenn ich selbst einen Förderbedarf sehe.	2.41	1.05	2	44.9	2.0	.40
	4.	Die Informationen in den Rückmeldungen sind zu detailliert.	3.21	0.77	3	41.1	2.0	.36
Selbstwirksamkeit im Umgang mit den Ergebnisrückmeldungen	1.	Es fällt mir leicht, die Ergebnisse des Haupttests zu interpretieren.	3.03	0.65	3	82.1	3.3	.57
	2.	Ich komme gut mit den Ergebnisrückmeldungen zurecht.	3.00	0.71	3	83.5	3.3	.59
	3.	Es liegt mir, mit den Ergebnisrückmeldungen des Haupttests zu arbeiten.	2.75	0.71	3	66.9	5.3	.43
	4.	Eine angemessene Interpretation der Ergebnisse gelingt mir nicht.	3.17	0.75	3	16.8	4.7	.53

Anmerkung. *N* = 150; Skalierung: 1 (trifft nicht zu) bis 4 (trifft zu).

Prozessschritt Reflexion. Die beiden Autoren Helmke und Hosenfeld (2005) beschrieben den Prozessschritt *Reflexion* als die Suche nach Erklärungen. Weil im Rahmen der vorliegenden Untersuchung die Nutzung der Ergebnisrückmeldungen stärker im Fokus des Interesses lag als die Bewertung der Schülerleistungen, wurden die Lehrkräfte befragt, ob sie mittels der Ergebnisrückmeldungen neue Informationen erhielten. Die Befragung erfasste also nicht die Reflexion der Lehrkräfte über die einzelnen Schülerleistungen, wie im Zyklenmodell der Rückmeldung vorgesehen, sondern fokussierte auf die Reflexion über den Informationsgewinn aus den Ergebnisrückmeldungen. Aus den Items wurde die Skala Informationsgewinn aus den Ergebnisrückmeldungen gebildet.

Bei der Skalenbildung (Tabelle 21) erwies sich das Item „Meine Einschätzungen der Schülerinnen und Schüler decken sich mit den Ergebnisrückmeldungen aus dem Haupttest“ als unpassend: Der Trennschärfenkoeffizienten r_{it} war $< .3$ und der α -Wert verschlechterte sich durch die Hinzunahme des Items deutlich ($\alpha = .52$). Aus diesem Grund wurde die Skala Informationsgewinn aus Ergebnisrückmeldungen aus den drei verbleibenden Items gebildet. Die Skala wies schließlich eine akzeptable interne Konsistenz auf ($\alpha = .75$, $M = 2.26$, $SD = .65$, $N = 143$). Die Trennschärfenkoeffizienten (r_{it}) der Items sind ebenfalls zufriedenstellend (Tabelle 21). Für die Skalenbildung wurden die negativ formulierten Items rekodiert. Dies betraf zwei Items. Der Modus bezog sich auf die beiden mittleren Antwortkategorien 2 (trifft eher nicht zu) und 3 (trifft eher zu).

Die Lehrkräfte wurden gefragt, ob sie aus den Ergebnisrückmeldungen des Haupttests neue Informationen über die Leistung der Schülerinnen und Schüler ableiten konnten (Tabelle 21). Hierzu kreuzte ein Großteil der Lehrkräfte (72 %) an, dass ihnen die Informationen aus den Ergebnisrückmeldungen des Haupttests bereits bekannt waren. Die Hälfte der antwortenden Lehrkräfte gab an, aus den Ergebnisrückmeldungen des Haupttests neue Informationen erhalten zu haben (50 %). Zudem gaben 67 % der Lehrkräfte an, dass sie auch ohne die Ergebnisrückmeldungen auf die Probleme bzw. Schwächen von Schülerinnen und Schülern gestoßen wären. Weiterhin gaben 88 % der Lehrkräfte an, dass sich ihre Einschätzung mit den Ergebnisrückmeldungen aus dem Haupttest des Testverfahrens *Online Diagnose* deckte.

Tabelle 21

Deskriptive Statistik und Trennschärfekoeffizienten der Items zum Prozessschritt Reflexion

Item	Itemtext	<i>M</i>	<i>SD</i>	Modus	Trifft (e-her) zu in %	Fehlende Angaben (Prozent)	<i>r_{it}</i>
1.	Die Ergebnisrückmeldungen des Haupttests bestätigen zu- meist nur, was ich sowieso schon wusste.	2.91	0.76	3	71.8	5.3	.58
2.	Durch die Ergebnisrückmel- dungen des Haupttests erfahre ich nichts Neues.	2.52	0.85	2	49.7	4.7	.72
3.	Ohne die Ergebnisrückmel- dungen des Haupttests würde ich auf manche Probleme bzw. Schwächen von Schülerinnen und Schülern nicht stoßen.	2.19	0.80	2	32.9	4.7	.43
4.	Meine Einschätzungen der Schülerinnen und Schüler de- cken sich mit den Ergebnis- rückmeldungen aus dem Haupttest.	3.10	0.65	3	87.5	4.0	.15

Anmerkung. *N* = 150; Skalierung: 1 (trifft nicht zu) bis 4 (trifft zu).

Prozessschritt Aktion. Der Prozessschritt *Aktion* bezieht sich auf Maßnahmen, die Schülerleistungen verbessern sollen. Helmke und Hosenfeld (2005) führten hierzu im Zy-
klenmodell der Rückmeldung eine Reihe von Handlungsmöglichkeiten auf: Die Sicherung
eines Mindestniveaus an fachlicher Kompetenz, die Steigerung der Unterrichtsqualität und
die Kopplung bzw. Einführung weiterer Projekte zur Unterrichtsqualität, die Veränderung
der Evaluations-, Aufgaben und Fehlerkultur sowie die Schulung der diagnostischen Kom-
petenz von Lehrkräften. Das Diagnose- und Förderinstrument *Online Diagnose* fokussierte
hier vor allem auf die gezielte Förderung von Schülerinnen und Schülern. Hierfür stellte das
Instrument individualisierte Fördermaterialien bereit. Für die vorliegende Untersuchung in-
teressierte zum Prozessschritt *Aktion*, wie die Ergebnisrückmeldungen genutzt und der Ein-
satz der Fördermaterialien umgesetzt wurde. Hierzu wurden zwei Facetten erfasst. Zum ei-
nen die in Abschnitt 5.2.1 aufgeführte Anwendung der Fördermaterialien und zum anderen
wurden die Lehrkräfte befragt, ob sich ihr Unterricht und ihr Verhalten vor und nach der
Einführung des Testverfahrens *Online Diagnose* unterschieden hatten. Die Items konnten
von 1 (viel weniger als vorher) bis 5 (viel mehr als vorher) beantwortet werden. Aus den
Items wurde die Skala gezielte Förderung gebildet.

Die gezielte individuelle Förderung einzelner Schülerinnen und Schüler wurde nach
der Durchführung des Testverfahrens *Online Diagnose* häufiger eingesetzt als zuvor – 64 %

der Lehrkräfte bestätigten dies. Die Förderung in Form von bestimmten Schülergruppen wurde von 45 % der Lehrkräfte häufiger als zuvor durchgeführt. Die Möglichkeit der Vertiefung oder Wiederholung von Unterrichtsinhalten wurde von 31 % der Lehrkräfte verstärkt eingesetzt. Bei zwei Items der Skala gezielte Förderung wurde durch den Modus deutlich, dass die Mehrheit der Lehrkräfte die Antwortkategorie 3 (genauso viel wie vorher) angekreuzt hatte. Zum Item „Gezielte individuelle Förderung einzelner Schülerinnen und Schüler“ wurde von den Lehrkräften die Antwortkategorie 4 (mehr als vorher) häufiger als die anderen Antwortkategorien angekreuzt. Die Mittelwerte zeigten die Tendenz hin zur Antwortkategorie 4 (mehr als vorher). Die interne Konsistenz der Skala war akzeptabel ($\alpha = .63$, $M = 3.52$, $SD = .45$, $N = 139$) und auch die Trennschärfekoeffizienten (r_{it}) waren annehmbar.

Tabelle 22

Deskriptive Statistik und Trennschärfekoeffizienten der Items zum Prozessschritt Aktion

Item	Itemtext	<i>M</i>	<i>SD</i>	Modus	(Viel) mehr als vorher in %	Fehlende Angaben (Prozent)	<i>r_{it}</i>
1.	Gezielte individuelle Förderung einzelner Schülerinnen und Schüler	3.74	0.64	4	64.3	6.7	.59
2.	Gezielte Förderung bestimmter Schülergruppen	3.50	0.61	3	45.4	6.0	.45
3.	Vertiefung bzw. Wiederholung von bestimmten Unterrichtsinhalten	3.32	0.54	3	31.2	6.0	.31

Anmerkung. $N = 150$; Skalierung: 1 (viel weniger als vorher) bis 5 (viel mehr als vorher).

Prozessschritt Evaluation. In diesem Abschnitt wird der Frage nachgegangen, wie sich das Diagnose- und Förderinstrument *Online Diagnose* auf die Schülerinnen und Schüler, das Kollegium und die eigene Person auswirkte. Im Sinne des Zyklusmodells der Rückmeldung (Helmke & Hosenfeld, 2005) könnte die Fragestellung zum Prozessschritt Evaluation wie folgt formuliert werden: Wem hat das Diagnose- und Förderinstrument genutzt und wie nachhaltig war die Wirkung? Da für die Studie nur ein Erhebungszeitpunkt geplant war und dieser im Schuljahr der Durchführung des Testverfahrens *Online Diagnose* lag, musste von einer Erfassung der Nachhaltigkeit des Testverfahrens abgesehen werden. Im Rahmen dieser Studie wurden zum Prozessschritt Evaluation drei Skalen abgeleitet: Wirkung auf Lernende, Wirkung auf das Kollegium und Wirkung auf Lehrende. Der Begriff „Wirkung“

war hierbei nicht als kausal zu verstehen. Als Selbstauskunft sollten die Lehrkräfte die Veränderungen für die Schülerinnen und Schüler, das Kollegium und die eigene Person als Lehrkraft einschätzen. Die exploratorische Faktorenanalyse in Tabelle 23 zeigt eine klare Trennung der Items in drei Faktoren. Alle drei Faktoren trafen das Eigenwertkriterium > 1 . Die drei Faktoren erklärten 60.03 % der Gesamtvarianz der Items.

Wirkung auf Lernende. Die Skala zur Wirkung auf Lernende zeigte eine zufriedenstellende interne Konsistenz auf ($\alpha = .86$, $M = 2.13$, $SD = .48$, $N = 141$). Auch die in Tabelle 23 dargestellten Trennschärfekoeffizienten (r_{it}) waren entsprechend hoch. Die Mehrheit der Lehrkräfte (57 %) war der Ansicht, dass eine sorgfältige Bearbeitung der Fördermaterialien durch die Schülerinnen und Schüler zu einer Steigerung der Kompetenzen führte. Dieser positive Aspekt konnte durch die weiteren Aussagen nicht bestätigt werden: Die Lehrkräfte waren der Ansicht, dass sich die Noten der förderbedürftigen Schülerinnen und Schüler durch das Diagnose- und Förderinstrument nicht verbessert hatten (86 %). Die Mehrheit der Lehrkräfte (68 %) war der Meinung, die Motivation der Schülerinnen und Schüler wurde nicht gesteigert. Auch die Beteiligung am Unterricht (80 %) verbesserte sich laut den Aussagen der Lehrkräfte nicht. Die Lehrkräfte gaben an, dass das Instrument den Anschluss einzelner förderbedürftiger Schülerinnen und Schüler an den Unterricht nicht herstellte (71 %). Zudem empfanden die Lehrkräfte keine Annäherungen der Leistungsunterschiede zwischen den Schülerinnen und Schülern (85 %).

Wirkung auf das Kollegium. Die interne Konsistenz der Skala Wirkung auf das Kollegium ($\alpha = .76$, $M = 1.83$, $SD = .61$, $N = 144$) war angemessen hoch. Auch die in Tabelle 23 dargestellten Trennschärfekoeffizienten (r_{it}) zu den Items, die von der jeweiligen Lehrkraft zum Kollegium beantwortet wurden, waren entsprechend angemessene Werte. In Bezug auf das Kollegium gaben die Lehrkräfte weder eine stärkere Kooperation (79 %) noch häufigere Absprachen über einzelne Schülerinnen und Schüler (81 %) als zutreffend an. Die Ergebnisrückmeldungen schienen den Lehrkräften bei Besprechungen nicht als gemeinsamer Referenzrahmen nützlich gewesen zu sein (85 %).

Wirkung auf Lehrende. Die Skala Wirkung auf Lehrende wies eine geringe, aber dennoch akzeptable interne Konsistenz auf ($\alpha = .69$, $M = 2.33$, $SD = .56$, $N = 144$). Die Höhe der in Tabelle 23 dargestellten Trennschärfekoeffizienten (r_{it}) war passabel. Die Lehrkräfte

gaben an, keine nachhaltigen Veränderungen an der Unterrichtsplanung durchgeführt zu haben (76 %). Auch fühlte sich die Mehrheit der Lehrkräfte (72 %) nicht weniger beansprucht. Doch gab die Mehrheit der Lehrkräfte als zutreffend an, dass ihnen die individuelle Förderung im Unterricht durch das Diagnose- und Förderinstrument *Online Diagnose* leichter fiel (67 %). Ebenfalls empfand die knappe Mehrheit der Lehrkräfte eine Steigerung ihrer persönlichen Diagnosekompetenz (56 %).

Der Modus zeigte auf, dass sich die Antworten hauptsächlich auf die mittleren Antwortkategorien 2 (trifft eher nicht zu) und 3 (trifft eher zu) bezogen. Hierbei überwogen die Rückmeldungen der Lehrkräfte zur Antwortkategorie 2 (trifft eher nicht zu).

Tabelle 23

Deskriptive Statistik, Trennschärfenoeffizienten und Faktorenladungen einer exploratorischen Faktorenanalyse zum Prozessschritt Evaluation

Item	Itemtext	<i>M</i>	<i>SD</i>	Modus	Trifft (e-her) zu in %	Fehlende Angaben (Prozent)	<i>r_{it}</i>	Wirkung auf Lernende	Wirkung auf das Kollegium	Wirkung auf Lehrende
1.	Die sorgfältige Bearbeitung der Fördermaterialien führt zu einer Steigerung der Kompetenzen.	2.52	0.70	3	57.3	4.7	.59	.68	-.04	.10
2.	Die förderbedürftigen Schülerinnen und Schüler erzielen nun bessere Noten.	1.97	0.55	2	13.7	7.3	.68	.70	.00	.17
3.	Die förderbedürftigen Schülerinnen und Schüler sind motivierter.	2.15	0.70	2	31.5	4.7	.64	.82	-.08	-.06
4.	Die förderbedürftigen Schülerinnen und Schüler beteiligen sich mehr am Unterricht.	2.01	0.62	2	20.0	6.7	.67	.83	.02	-.14
5.	Die förderbedürftigen Schülerinnen und Schüler finden im Unterricht wieder Anschluss.	2.17	0.65	2	29.3	6.7	.60	.66	.19	-.01
6.	Die Leistungsunterschiede zwischen den Schülerinnen und Schülern sind kleiner geworden.	1.99	0.75	2	15.1	7.3	.70	.84	-.06	-.00
7.	Im Kollegium besprechen wir nun häufiger die Leistungen einzelner Schülerinnen und Schüler.	1.91	0.75	2	20.8	4.0	.48	.12	.74	-.12
8.	Durch die <i>Online Diagnose</i> haben wir bei den Besprechungen im Kollegium einen gemeinsamen Referenzrahmen.	1.81	0.73	2	18.6	3.3	.64	-.09	.91	-.02
9.	Es findet jetzt eine stärkere Kooperation im Kollegium statt.	1.77	0.75	2	14.7	4.7	.66	-.04	.86	.07
10.	Ich habe meine Unterrichtsplanung nachhaltig verändert.	2.08	0.75	2	24.0	2.7	.50	.01	.05	.74
11.	Ich fühle mich weniger beansprucht.	2.10	0.86	2	28.1	5.3	.33	-.07	-.21	.66
12.	Individuelle Förderung im Unterricht fällt mir nun leichter.	2.65	0.73	3	67.4	4.0	.62	.03	.04	.78
13.	Ich konnte meine eigene Diagnosekompetenz steigern.	2.51	0.76	3	55.9	4.7	.49	.03	.12	.68
	Eigenwert des Faktors							4.68	1.89	1.25
	Erklärte Varianz des Faktors in %							35.99	14.45	9.60

Anmerkungen. N = 150; Antwortformat: 1 (trifft nicht zu) bis 4 (trifft zu); Extraktionsmethode: Hauptkomponentenanalyse; Rotationsmethode: Promax mit Kaiser-Normalisierung; Faktorladungen > .40 sind fettgedruckt.

5.3 Diskussion und Implikationen

In dieser Teilstudie wurden deskriptive Befunde zum Testverfahren *Online Diagnose* berichtet. Die Befunde werden in den nachfolgenden Abschnitten wie folgt zusammengefasst und diskutiert: Durchführung und Nutzung des Diagnose- und Förderinstruments *Online Diagnose*, die Motivation bei der Durchführung des Haupttests und dem Einsatz der Fördermaterialien, die subjektive Qualitätseinschätzung zum Testverfahren *Online Diagnose* und die Prozessschritte der Nutzung von Ergebnisrückmeldungen.

Durchführung und Nutzung des Diagnose- und Förderinstruments *Online Diagnose*.

Angaben zum Ablauf des Testverfahrens Online Diagnose. Die Ergebnisse zeigten, dass die Schülerinnen und Schüler von der Lehrkraft über den Zweck des Haupttests informiert wurden. Im Anschluss ließen drei Viertel der Lehrkräfte den Haupttest zur *Online Diagnose* alle Schülerinnen und Schüler gleichzeitig bearbeiten. Lediglich ein Viertel der Lehrkräfte ließ die Schülerinnen und Schüler den Haupttest in mehreren Gruppen bearbeiten. Dies deutet zum einen auf eine ausreichende technische Ausstattung der Schulen hin und zum anderen darauf, dass die Bearbeitung des Haupttests im Klassenkontext stattgefunden hat. Somit konnte die Mehrheit der teilnehmenden Schülerinnen und Schüler sich nicht vorab über die Aufgabenstellungen unterhalten.

Nutzungsumfang und Zeitaufwand für die Anwendung des Testverfahrens Online Diagnose. Bei den Häufigkeitsangaben zum Nutzungsumfang und dem Zeitaufwand bei der Anwendung des Testverfahrens *Online Diagnose* wurde deutlich, dass der Haupttest häufiger genutzt wurde als der Detailtest oder der Nachtest und dass mehr Zeit für ihn aufgewendet wurde. Außerdem wurde mehr Zeit während des Unterrichts für den Haupttest investiert als außerhalb des Unterrichts. Insgesamt wurden für das Testverfahren *Online Diagnose* durchschnittlich $M = 13$ Schulstunden während des Unterrichts aufgewendet (inkl. der Bearbeitung und Besprechung der Fördermaterialien). Außerhalb des Unterrichts betrug der Stundenaufwand in etwa sieben Schulstunden (inkl. der Bearbeitung und Besprechung der Fördermaterialien). Zu den Häufigkeitsangaben zum Nutzungsumfang und dem Zeitaufwand bei der Anwendung des Testverfahrens *Online Diagnose* fiel ein hoher Anteil an fehlenden Werten auf. Besonders hoch waren die fehlenden Angaben zum Detailtest und zum Nachtest. Gleichzeitig nahm die Häufigkeit der Anwendung und der Aufwand an Zeit beim

Detailtest und Nachtest deutlich ab. Ebenso waren die Standardabweichungen auffallend hoch. Das bedeutet, die Angaben waren schief verteilt. Hieraus konnte geschlossen werden, dass der Detailtest und der Nachtest von wenigen Lehrkräften genutzt wurde. Gaben die Lehrkräfte an, den Detail- und Nachtest genutzt zu haben, so war damit die Auskunft verbunden, dass dies in einem geringen Zeitumfang geschah. Für die Nutzung des Testverfahrens *Online Diagnose* bedeutet dies, dass das Verfahren selten wie vorgesehen genutzt wurde. Gründe für die eingeschränkte Nutzung konnten durch die Untersuchung nicht ausgemacht werden. Möglicherweise gab es organisatorische Probleme, wie z.B. die Belegung des Computerraumes, oder das Testverfahren beanspruchte zu viel Unterrichtszeit, die bereits für andere Themen fixiert war. Gründe für die eingeschränkte Nutzung des Testverfahrens *Online Diagnose* sollten vor einer weiteren Nutzung untersucht werden.

Für das Durcharbeiten der individuellen Ergebnisrückmeldungen brachten die Lehrkräfte mehr Zeit außerhalb des Unterrichts auf als während des Unterrichts. Insgesamt benötigten die Lehrkräfte, die während und außerhalb des Unterrichts Zeit für das Durcharbeiten der individuellen Ergebnisrückmeldungen aufwandten, durchschnittlich $M = 4.3$ Schulstunden ($SD = 3.6$, $N = 60$). Im Vergleich hierzu gab Peek (2004) in seiner Studie QuaSUM³⁹ eine Dauer von durchschnittlich 2.6 Stunden an, in der sich die Mathematiklehrkräfte mit den klassen- bzw. kursbezogenen Rückmeldungen auseinandersetzten ($N = 250$). Bezogen auf die Gesamtzahl der Schülerinnen und Schüler, für die die Lehrkräfte die Ergebnisrückmeldungen durcharbeiteten, ist davon auszugehen, dass nicht viel Zeit für die gesamte Ergebnisrückmeldung aufgewendet wurde. Durch die Fokussierung auf Teilbereiche von Ergebnisrückmeldungen könnten in zukünftigen Untersuchungen genauere Aussagen über die Nutzung der Ergebnisrückmeldungen getroffen werden. Unter anderem könnte überprüft werden, für welche Bereiche der Ergebnisrückmeldung die Lehrkräfte besonders viel Zeit aufwandten und welche Bereiche sie zudem besonders hilfreich empfanden.

Für die Zusammenstellung der Fördermaterialien wendeten die Lehrkräfte mehr als doppelt so viel Zeit außerhalb des Unterrichts auf wie während des Unterrichts. Die Bearbeitung und Besprechung der Fördermaterialien nahm den größten Zeitaufwand während des Unterrichts in Anspruch, insgesamt $M = 10.9$ Schulstunden. Die Bearbeitung und Besprechung der Fördermaterialien außerhalb des Unterrichts betrug ca. ein Drittel der für die Bearbeitung und Besprechung der Fördermaterialien während des Unterrichts verwendeten Zeit. Bei den Zeitangaben zu den Fördermaterialien waren die Standardabweichungen hoch

³⁹ Qualitätsuntersuchung an Schulen zum Unterricht in Mathematik (QuaSUM)

und deuteten auf eine schiefe Verteilung der Antworten hin. Vor allem die fehlenden Angaben zur Bearbeitung und Besprechung der Fördermaterialien außerhalb des Unterrichts waren mit 58 % enorm hoch. Die Ergebnisse lassen erkennen, dass es beim Zeitaufwand für die Nutzung der Fördermaterialien starke Abweichungen gab, was wiederum darauf hinweist, dass die Nutzung der Fördermaterialien unterschiedlich gehandhabt wurde. Als Folge dieser Ergebnisse wären Interviews der Lehrkräfte hilfreich, um herauszufinden worin mögliche Probleme bei der Nutzung der Fördermaterialien lagen. Außerdem könnte durch die Erfahrungen der Lehrkräfte eine Optimierung des Einsatzes der Fördermaterialien vorgenommen werden und den Lehrkräften für die zukünftige Nutzung eine Handlungsanleitung zur Verfügung gestellt werden.

Austausch über die Ergebnismeldungen mit pädagogischen Akteuren der Schule. Zum Fragenkomplex Austausch der Lehrkräfte mit pädagogischen Akteuren der Schule berichteten Erstere hauptsächlich die Nutzung der Ergebnismeldungen in Gesprächen mit anderen Lehrkräften der jeweiligen Klasse, mit Eltern einzelner Schülerinnen und Schülern der Klasse, sowie mit einzelnen Schülerinnen und Schülern im persönlichen Gespräch. Im Rahmen der Untersuchung von Peek (2004; QuaSUM) gaben 58 % der Lehrkräfte ($N = 250$) an, sich mit Kolleginnen und Kollegen auszutauschen, die in der Testklasse unterrichteten. Ähnlich umfangreich war der Austausch der Lehrkräfte der vorliegenden Untersuchung zum Testverfahren *Online Diagnose* (74 %, $N = 143$).

Anwendung der Fördermaterialien. Hinsichtlich der Anwendung der Fördermaterialien gaben 91 % der Lehrkräfte an, die im Rahmen des Testverfahrens *Online Diagnose* individuell zusammengestellten Fördermaterialien eingesetzt zu haben. Gleichzeitig gaben 41 % der Lehrkräfte an, die Fördermaterialien selbst zusammengestellt zu haben. Hier ist zu beachten, dass sich die Angaben überschneiden. Das bedeutet, dass die beiden Items nicht genau differenzierten und anhand der Antworten somit nicht konkret unterschieden werden konnte, wie die Fördermaterialien eingesetzt wurden. Eine genauere Erfassung, ob den Schülerinnen und Schülern im Rahmen der Fördermaterialien die gesamten Aufgaben vorgelegt oder ob einzelne Aufgaben aus dem Förderpaket ausgewählt wurden, wäre hinsichtlich des Erkenntnisinteresses nützlicher.

Die Ergebnisse verdeutlichen, dass der Einsatz der Fördermaterialien im Sinne des Testverfahrens *Online Diagnose* differenziert wurde. Es kam selten vor, dass die gesamte

Klasse oder alle förderbedürftigen Schülerinnen und Schüler dieselben Aufgaben bearbeiten. Über die Ergebnismeldungen aus dem Testverfahren *Online Diagnose* wurden die Schülerinnen und Schüler bzw. deren Leistungen in die drei Kategorien Rot, Orange und Grün eingestuft, wobei die beiden Farben Rot und Orange für den Bedarf an Förderung der Schülerinnen und Schüler standen und die Farbe Grün für die Förderung der Schülerinnen und Schüler. Somit erhielten auch die nicht-förderbedürftigen Schülerinnen und Schüler Aufgaben.

Die Schülerinnen und Schüler bearbeiteten die Fördermaterialien mehrheitlich während des Deutschunterrichts oder während eines speziellen Förderunterrichts. Ein Teil der Lehrkräfte gab die Fördermaterialien als verpflichtende oder freiwillige Hausaufgabe. Außerdem überprüfte mehr als die Hälfte der Lehrkräfte die Lösungen der Förderaufgaben im Unterricht.

Technische Ausstattung der Schulen. Im Rahmen dieser Stichprobe gab die Mehrheit der Lehrkräfte an, dass die Anzahl und Funktion der Computer, die Internetverbindung und die Druckmöglichkeiten ausreichend waren. An der zusätzlichen Ausstattung wie Kopfhörer oder Lautsprecher mangelte es eher. Dem Einsatz des Testverfahrens *Online Diagnose* dürfte jedoch von Seiten der technischen Ausstattung im Sinne der Hard- und Software mehrheitlich nichts im Wege gestanden haben. Hosenfeld, Koch, Groß Ophoff, und Scherthan bescheinigten den Schulen bereits in ihrem im Jahr 2008 im Rahmen von Untersuchungen zu VERA veröffentlichten Artikel eine entsprechend gute Hard- und Software-Ausstattung hinsichtlich internetfähiger Rechner und Druckmöglichkeiten sowie unter anderem einer nützlichen Bildschirmauflösung und eines aktuellen Browsers (Hosenfeld, Koch, Groß Ophoff & Scherthan, 2008).

Mögliche Gründe für ein schwaches Testergebnis. Die Lehrkräfte bewerteten die äußeren Einflüsse bzw. Störungen auf die Durchführung des Haupttests insgesamt als wenig problematisch. Die Einflussfaktoren Lärm und Zeitdruck stellten im Rahmen dieser Stichprobe bei einem Viertel der Schulen ein Problem dar. Einen stärkeren Einfluss hatten die technischen Probleme während der Testung, diese Antwort gaben rund 42 % der Lehrkräfte. In diesem Bereich herrscht durchaus Optimierungsbedarf, der gedeckt sein müsste, um eine störungsfreie Testung zu ermöglichen. Genauere Kenntnis darüber, was zu den technischen

Problemen führte und worin diese bestanden, wäre zur Verbesserung der Situation notwendig. Möglicherweise könnten einige Probleme mit Hilfe eines telefonischen Supports gelöst werden. Andere müssten eventuell über eine programm-basierte Lösung verbessert werden. Hosenfeld et al. (2008) berichten in ihrer Untersuchung im Rahmen von VERA, dass rund 33 % der Lehrkräfte sich mit Anfragen an den technischen Support gewendet hatten ($N = 6.400$). Die Autoren konnten zudem den Rückgang an Support-Anfragen bereits im Folgejahr aufzeigen, obwohl die Vergleichsarbeiten dennoch vollständig durchlaufen wurden.

Auf die Frage nach weiteren Gründen für ein schwaches Testergebnis gaben beinahe alle Lehrkräfte (97 %) den tatsächlichen Förderbedarf der Schülerinnen und Schüler an. Dies würde mit Bezugnahme auf die subjektive Einschätzung der Lehrkräfte bedeuten, dass der Haupttest geeignet war, den tatsächlichen Förderbedarf festzustellen. Diese subjektive Einschätzung sollte hinsichtlich ihrer psychometrischen Güte durch die Untersuchung belegt oder widerlegt werden. Auch Sprachschwierigkeiten von Schülerinnen und Schülern mit Migrationshintergrund trugen laut den Lehrkräften zu einem schwachen Testergebnis bei. Hinsichtlich der Schülerinnen und Schüler mit Migrationshintergrund müsste das Testverfahren *Online Diagnose* speziell auf seine Eignung überprüft werden. Jeuk (2009) brachte hierzu neben anderen Kriterien das Argument vor, dass nicht nur das Alter hinsichtlich der Prüfung vorliegender Sprachkenntnisse berücksichtigt werden müsse, sondern auch die Dauer, die Qualität und die Quantität des Kontakts mit der deutschen Sprache. Die Unverständlichkeit der Aufgabenstellungen und die mangelnde Motivation der Schülerinnen und Schüler sahen nahezu 45 % der Lehrkräfte als weitere mögliche Gründe für ein schwaches Testergebnis an. Auch hier sollte die psychometrische Güte der Tests überprüft werden, um eine genauere Aussage zur Verortung möglicher Gründe für ein schwaches Testergebnis treffen zu können.

Unterstützung für Lehrkräfte bei der Nutzung des Testverfahrens Online Diagnose. Lediglich zehn der an der Befragung teilnehmenden Lehrkräfte nutzten eine Fortbildung als Unterstützungsangebot für die Durchführung des Testverfahrens *Online Diagnose*. Der Grund für die geringe Nutzerquote könnte unter anderem bei den Lehrkräften liegen oder am mangelnden Angebot an Fortbildungsmaßnahmen, dies konnte im Rahmen der vor-

liegenden Untersuchung nicht überprüft werden. Ditton und Arnoldt (2004) berichten in ihrer Untersuchung zu QuaSSU⁴⁰ ebenfalls eine geringe Ambition der Lehrkräfte hinsichtlich der Wahrnehmung von Unterstützungsangeboten: nur ca. 8 % der Lehrkräfte gaben an, eine persönliche Beratung für erforderlich zu halten (N = 98).

Die sogenannten „schriftlichen Hilfen“ nutzten die Lehrkräfte umfassend. Vor allem die Kurzanleitung zur *Online Diagnose* und der „Quick Tipp“ am Seitenrand einer geöffneten Internetseite erschien den Lehrkräften als hilfreich. Auch die Sammlung häufig gestellter Fragen (FAQ), das Nutzerhandbuch und die Konzeption beurteilten die antwortenden Lehrkräfte als hilfreich. Vor allem durch den „Quick Tipp“ konnte von einer direkten Unterstützung der Lehrkräfte bei der Online-Nutzung z. B. der Ergebnisrückmeldungen ausgegangen werden.

Motivation bei der Durchführung des Haupttests und dem Einsatz der Fördermaterialien. Die aus Schülerperspektive (Einschätzung durch die Lehrkraft) und Lehrerperspektive (eigene Einschätzung) gewonnenen Items zu den Wertaspekten in Bezug auf den Haupttest und die Fördermaterialien, ließen sich zu zwölf Skalen zusammenfassen. Die Reliabilitätsanalysen ergaben bei elf Skalen eine gute interne Konsistenz und auch die Trennschärfekoeffizienten waren entsprechend gut. Die Skala relative Kosten in Bezug auf die Fördermaterialien aus Schülerperspektive (Einschätzung durch die Lehrkraft) ergab eine geringe interne Konsistenz und auch dementsprechend geringe Trennschärfekoeffizienten.

Die im Folgenden genannte erwartete Faktorenstruktur konnte weder in Bezug auf den Haupttest noch in Bezug auf die Fördermaterialien gefunden werden: Wichtigkeit, intrinsischer Wert, Nützlichkeit und relative Kosten. Im Rahmen der Schülerperspektive (Einschätzung durch die Lehrkraft) luden die Wichtigkeit und der intrinsische Wert auf einen Faktor. Die Faktorenladungen verhielten sich in Bezug auf den Haupttest und in Bezug auf die Fördermaterialien gleich. Die Ladungen können möglicherweise mithilfe der Interesstheorie begründet werden (z. B. Krapp, 2002). In diesem Zusammenhang wurde angenommen, dass die persönliche Wichtigkeit und die intrinsische Motivation für eine Aktivität nicht voneinander zu unterscheiden sind. Ein weiterer möglicher Grund für das Ergebnis könnte in der Einschätzung der Schülerperspektive durch die Lehrkräfte liegen. Für die Lehrkräfte als Außenstehende sind die Wichtigkeit und der intrinsische Wert möglicherweise

⁴⁰ Qualitätssicherung in Schule und Unterricht (QuaSSU)

nicht als unterschiedliche Wertaspekte wahrnehmbar. Grundsätzlich muss eine Einschätzung der Motivation und im Speziellen der intrinsischen Motivation durch Außenstehende kritisch diskutiert werden. Die Kombination der Einschätzungen der Lehrkraft mit Beurteilungen der Schülerinnen und Schüler selbst (Schülerdaten) könnte hierzu eine neue Perspektive eröffnen.

Außerdem luden im Rahmen der Lehrerperspektive (eigene Einschätzung) die Wichtigkeit und die Nützlichkeit auf einen Faktor. Die Faktorenladungen verhielten sich in Bezug auf den Haupttest und in Bezug auf die Fördermaterialien gleich. Möglicherweise waren die Formulierungen der Items in diesem Zusammenhang nicht optimal gewählt, um zwischen den beiden Bestandteilen der Wertaspekte zu differenzieren (z.B. Tabelle 10 und 13). Ein weiterer Grund könnte sein, dass die Wichtigkeit und Nützlichkeit für die Lehrkräfte im Rahmen des Testverfahrens *Online Diagnose* gleichbedeutend war.

Wertaspekte in Bezug auf den Haupttest. Zusammenfassend ist festzustellen, dass laut den Angaben der Lehrkräfte die Schülerinnen und Schüler den Haupttest als eher wichtig ansahen bzw. der intrinsische Wert eher hoch war. Die Ergebnisrückmeldungen aus dem Haupttest schätzten die Lehrkräfte für die Schülerinnen und Schüler ebenfalls als eher nützlich ein. Insgesamt bildete sich jedoch kein eindeutiges Meinungsbild ab. Die relativen Kosten bei der Bearbeitung des Haupttests schätzten die Lehrkräfte für die Schülerinnen und Schüler als nicht besonders hoch ein. Für sich selbst bewerteten die Lehrkräfte den intrinsischen Wert der Ergebnisse des Haupttests als überwiegend hoch. Die Nützlichkeit und Wichtigkeit des Haupttests bezeichneten die Lehrkräfte als hoch. Die relativen Kosten gaben die Lehrkräfte als eher hoch an. Jedoch auch hier ließ sich kein klares Meinungsbild der Lehrkräfte darstellen: Den Stress bei der Durchführung des Tests und den Aufwand an Unterrichtszeit schätzte die Mehrheit der Lehrkräfte als nicht sehr hoch ein. Den logistischen Aufwand bei der Testdurchführung und den Zeitaufwand bei der Auseinandersetzung mit den Ergebnisrückmeldungen empfand die Mehrheit der Lehrkräfte allerdings als unangemessen.

Wertaspekte in Bezug auf die Fördermaterialien. Eine Zusammenfassung der Einschätzungen der Lehrkräfte in Bezug auf die Fördermaterialien macht deutlich, dass der intrinsische Wert und die Wichtigkeit für die Schülerinnen und Schüler in den Einschätzungen eher gering waren. In der projektiv eingenommenen Schülerperspektive schätzten die

Lehrkräfte die Fördermaterialien als eher nützlich ein. Die relativen Kosten für die Schülerinnen und Schüler bewerteten die Lehrkräfte als gering. Eine Gesamtschau der Wertaspekte aus der Lehrerperspektive ergab, dass der Einsatz der Fördermaterialien eher nicht intrinsisch motiviert war. Die Nützlichkeit bzw. Wichtigkeit der Fördermaterialien bewertete die Mehrheit der Lehrkräfte insgesamt als hoch. Die relativen Kosten wurden von der Mehrheit der Lehrkräfte als eher gering bewertet.

Die Ergebnisse zu den Wertaspekten zeigen, dass sich vor allem die Lehrkräfte hinsichtlich der Anwendung der Instrumente und Maßnahmen des Testverfahrens *Online Diagnose* als motivierter betrachteten als die Schülerinnen und Schüler und zugleich die Instrumente und Maßnahmen des Testverfahrens als nützlicher bzw. wichtiger für sich selbst einschätzten als für die Schülerinnen und Schüler. Auch die relativen Kosten schätzten die Lehrkräfte für sich selbst im Vergleich zu den Schülerinnen und Schülern als höher ein. Möglicherweise wirkten die Schülerinnen und Schüler auf die Lehrkräfte tatsächlich nicht besonders motiviert und interessiert. Allerdings ist eine Fremdeinschätzung der Schülerperspektive hinsichtlich der Motivation durchaus als kritisch zu bewerten und sollte in jedem Fall mit direkten Aussagen von Schülerinnen und Schüler verglichen werden.

Subjektive Qualitätseinschätzung des Testverfahrens *Online Diagnose*. Um zu einer subjektiven Qualitätseinschätzung zu gelangen, wurden die Lehrkräfte darum gebeten, die Qualität der Testaufgaben hinsichtlich ihrer Eignung und Verständlichkeit zu bewerten. Außerdem wurden die Lehrkräfte dazu befragt, ob sie einen Zusammenhang zwischen dem Haupttest und dem Bildungsplan der Grundschule sähen. Weiterhin wurden die Lehrkräfte um eine allgemeine Qualitätseinschätzung zum Haupttests, den Ergebnisrückmeldungen und den Fördermaterialien gebeten. Bis auf die Skala Qualität des Haupttests wiesen alle berichteten Skalen eine ausreichende interne Konsistenz auf und auch die Trennschärfekoeffizienten waren entsprechend hoch.

Wie die Ergebnisse zeigen, wurden die Testaufgaben der *Online Diagnose* hinsichtlich ihrer Verständlichkeit und Eignung von den Lehrkräften sehr positiv bewertet. Den Bezug zum Bildungsplan der Grundschule schätzten die Lehrkräfte als passend ein. Auch mit der Qualität des Haupttests war die Mehrheit der Lehrkräfte zufrieden. Die Qualität der Ergebnisrückmeldungen wurde von den Lehrkräften als positiv wahrgenommen. Die Qualität

der Darstellung der Ergebnismeldungen beschrieben die Lehrkräfte ebenfalls als zufriedenstellend. Die Skala zur Qualität der Fördermaterialien wurde wie die anderen Skalen von der Mehrheit der Lehrkräfte als gut eingeschätzt. Für die Überprüfung der Testaufgaben würden sich psychometrische Verfahren anbieten. In einigen Studien (z.B. Koch 2011; Kühle & Peek, 2007; Schneewind, 2006) wird deutlich, dass die Verständlichkeit und Qualität von Testverfahren durch die Lehrkräfte als sehr gut eingeschätzt wurde. Dies wirft die Frage auf, wie zuverlässig die Einschätzungen der Lehrkräfte sein können, zumal die in Kapitel 3.3 zum theoretischen Hintergrund dargestellte Gegenüberstellung der Testinhalte aus dem Testverfahren *Online Diagnose* mit den Bildungsstandards keine umfassenden Überschneidungen aufzeigen. Teilweise konnte z.B. angenommen werden, dass die Testaufgaben zum Testverfahren *Online Diagnose* Inhalte der zweiten Jahrgangsstufe der Grundschule abfragten, anstatt Inhalte der vierten Jahrgangsstufe. Objektivere Methoden zur Untersuchung der Testqualität wären eine angemessene Ergänzung.

Der bivariate Zusammenhang der Skala Qualität der Ergebnismeldungen mit der Skala Qualität der Fördermaterialien war nicht signifikant. Dem sollte insofern Aufmerksamkeit geschenkt werden, als die Lehrkräfte möglicherweise der Verbindung zwischen den zurückgemeldeten Leistungen der Schülerinnen und Schüler und dem Einsatz der Fördermaterialien keine Beachtung schenken. Einer gezielten individuellen Förderung dient jedoch nicht nur der Einsatz von Fördermaterialien, sondern vielmehr der anhand von Ergebnismeldungen geplante Einsatz der Fördermaterialien. Die Ergebnisse vermitteln den Eindruck, dass die Ergebnismeldungen nicht unmittelbar als relevant für die Förderung der Schülerinnen und Schüler angesehen wurden. Hierfür lieferten die Ergebnisse allerdings keine fundierten Beweise, sondern lediglich Vermutungen, die in weiteren Untersuchungen aufgegriffen werden sollten.

Prozessschritte der Nutzung von Ergebnismeldungen. Abschließend wurden die zu den Prozessschritten der Nutzung von Ergebnismeldungen abgeleiteten Skalen in Bezug auf die Rezeption, Reflexion, Aktion und Evaluation des theoretischen Zyklusmodells der Rückmeldung operationalisiert:

Prozessschritt Rezeption. Aus dem Zyklusmodell der Rückmeldung wurden zum Prozessschritt Rezeption für die vorliegende Untersuchung zwei Skalen abgeleitet: Der Zeit-

aufwand für die Nutzung der Ergebnisrückmeldungen und die Selbstwirksamkeit im Umgang mit den Ergebnisrückmeldungen. Die Einstellung der Lehrkräfte zum Zeitaufwand für die Nutzung der Ergebnisrückmeldungen tendierte im Rahmen dieser Stichprobe in keine eindeutige Richtung. Ein Teil der Lehrkräfte bewertete den Zeitaufwand als hoch, ein anderer Teil als gering. Die Standardabweichungen zu den Items des Zeitaufwands für die Nutzung der Ergebnisrückmeldungen waren auffallend hoch.

Den eigenen Angaben zufolge nutzte der Großteil der Lehrkräfte die Ergebnisrückmeldungen aus dem Testverfahren *Online Diagnose* selbstwirksam. Auch Schneewind (2006) kommt im Rahmen der Studie BeLesen zu dem Ergebnis, dass die Lehrkräfte die Ergebnisrückmeldungen als eher nützlich sowie verständlich wahrnahmen und zudem akzeptierten. Ähnliches zur positiven Beurteilung der Nützlichkeit und Verständlichkeit von Ergebnisrückmeldungen berichten z.B. Nachtigall und Jantowski (2007) zu den Thüringer Kompetenztests, Schwippert (2004) zu IGLU sowie Ditton und Arnold (2004) zur Studie QuaSSU. Kühle und Peek (2007) verweisen sogar auf eine weitere Abnahme von Verständnisschwierigkeiten bei der Nutzung von Ergebnisrückmeldungen im zweiten Erhebungsjahr. Koch (2011) konnte zudem nachweisen, dass die Auseinandersetzung mit den Ergebnisrückmeldungen umso intensiver war und als umso nützlicher erlebt wurde, je verständlicher diese in der Wahrnehmung der Lehrkräfte waren. Im Gegensatz hierzu schreiben Kühle und Peek (2007), dass sich die Verständlichkeit der Ergebnisrückmeldungen nicht als Prädiktor für die Bereitschaft der Nutzung eigne. Allerdings korreliert die Verständlichkeit im Rahmen ihrer Untersuchung bivariat hoch mit dem Prädiktor Nützlichkeit. Die Operationalisierung des Prozessschrittes Rezeption als Selbstwirksamkeit im Umgang mit Ergebnisrückmeldungen kann durchaus kritisch gesehen werden, denn die Verständlichkeit der Ergebnisrückmeldungen wurde hierbei zwar impliziert, es war jedoch kein Beweis möglich, dass Lehrkräfte, die sich im Umgang mit den Ergebnisrückmeldungen als selbstwirksam betrachteten, dieselben auch als subjektiv verständlich einschätzten. Dennoch wurde im Rahmen dieser Untersuchungen dem selbstwirksamen Umgang mit Ergebnisrückmeldungen eine heuristisch zuverlässigere Aussagekraft zugeschrieben als der subjektiv eingeschätzten Verständlichkeit der Ergebnisrückmeldungen. Begründen lässt sich diese Zuschreibung damit, dass dem selbstwirksamen Umgang ein Bezug zu Handlungen mit den Ergebnisrückmeldungen möglich war.

Prozessschritt Reflexion. Zum Prozessschritt Reflexion wurde die Skala Informationsgewinn aus den Ergebnisrückmeldungen abgeleitet. Die Lehrkräfte tendierten zu einer negativen Beantwortung der vorgelegten Items. Das bedeutet, die Mehrheit der Lehrkräfte war der Meinung, dass sie auch ohne die Ergebnisrückmeldungen entsprechend informiert gewesen wären. Die Ergebnisrückmeldungen lieferten ihnen nicht unbedingt neue Informationen. Schneewind (2006) erhielt in ihrer Untersuchung ein ähnliches Ergebnis. Die Mehrheit der Lehrkräfte gab an, lediglich für ein Viertel ihrer Schülerinnen und Schüler neue Informationen erhalten zu haben. Ebenso gab die Mehrheit der Lehrkräfte an, dass die Ergebnisrückmeldungen bei drei Vierteln der Schülerinnen und Schüler mit der eigenen Einschätzung übereinstimmten. Diese Aussage deckt sich mit der vorliegenden Untersuchung. Kohler (2005) kam in ihrer Untersuchung zur Problematik der Studie TIMS zu dem Ergebnis, dass die Lehrkräfte sich bei Aussagen dazu, ob das Ergebnis des Abschneidens der Schülerinnen und Schüler erwartet worden war, unentschlossen zeigten.

In Bezug auf die positive subjektive Qualitätseinschätzung der Ergebnisrückmeldungen zeigte sich hier überraschend, dass der Informationsgewinn als eher gering befunden wurde. Die Lehrkräfte sahen die Qualität der Ergebnisrückmeldungen als gut an, gaben jedoch gleichzeitig an, keine neuen Erkenntnisse daraus zu gewinnen. Weder mit der subjektiven Qualitätseinschätzung der Ergebnisrückmeldungen noch mit der Selbstwirksamkeit im Umgang mit den Ergebnisrückmeldungen korrelierte der Informationsgewinn aus Ergebnisrückmeldungen bivariat. Im Gegensatz hierzu korrelierte der Informationsgewinn aus den Ergebnisrückmeldungen bivariat signifikant mit der gezielten Förderung (bivariaten Korrelationen, siehe Tabelle 25). Die Lehrkräfte schienen hier also keine Verbindung zwischen der Qualität der Ergebnisrückmeldungen und dem Informationsgewinn zu ziehen. Die Aufmerksamkeit der Lehrkräfte auf die Verknüpfung zwischen Ergebnisrückmeldungen und Informationsgewinn zu lenken und weitere Untersuchungen in diesem Bereich zu forcieren, erscheint sinnvoll, zumal der Informationsgewinn wiederum einen möglichen Einfluss auf den Einsatz von Fördermaterialien aufweist.

Prozessschritt Aktion. Aus der Vielfalt der Handlungsvorschläge des Zyklenmodells der Rückmeldung wurde im Rahmen der Untersuchung zum Testverfahren *Online Diagnose* zum Prozessschritt Aktion die Skala gezielte Förderung abgeleitet. Da sich das Testverfahren *Online Diagnose* durch die Ausgabe der Fördermaterialien nahezu ausschließlich auf die

gezielte Förderung von Schülerinnen und Schülern bezog. Gemäß dem Prinzip des Testverfahrens *Online Diagnose* wurde die Einzelförderung häufiger angewandt als die beiden anderen Fördermöglichkeiten der gezielten Förderung von Schülergruppen und der gezielten Vertiefung oder Wiederholung von Unterrichtsinhalten.

Anhand eines offenen Antwortformats in einem Verfahren der Kategorienbildung konnte Schneewind (2006) im Rahmen der Studie BeLesen aufzeigen, dass die gezielte Förderung leistungsschwacher und -starker Schülerinnen und Schüler ein Hauptpunkt hinsichtlich der Veränderung der Arbeit von Lehrkräften war. Die Änderungen umfassten unter anderem zusätzliche Förderstunden oder ein verstärktes Üben im Unterricht. Groß Ophoff (2013) bezeichnet die im Rahmen von VERA erfassten Unterrichtsveränderungen (Wiederholung bzw. Vertiefung von Unterrichtsinhalten) als ein Merkmal mit einer starken Anwendungsausprägung. Die gezielte Förderung einzelner Schülerinnen und Schüler wurde im Zusammenhang dieses Fragenblocks allerdings nicht erfasst. Hinsichtlich einer weiteren Befragung zu VERA, in der ein offenes Antwortformat verwendet wurde, kristallisierte sich ebenfalls die Aussage heraus, dass Lehrkräfte Ergebnisrückmeldungen zur Ableitung von Maßnahmen für den Unterricht nutzten (Groß Ophoff, Koch, Hosenfeld & Helmke, 2006).

Prozessschritt Evaluation. Zum Prozessschritt Evaluation wurden die drei Skalen Wirkung auf Lernende, Wirkung auf das Kollegium und Wirkung auf Lehrende abgeleitet. Den Angaben der Lehrkräfte war zu entnehmen, dass sie insgesamt eine geringe Wirkung auf Lernende durch das Testverfahren *Online Diagnose* wahrnahmen. Lediglich die Aussage, die sorgfältige Bearbeitung der Fördermaterialien führe zu einer Steigerung der Kompetenzen aufseiten der Lernenden, wurde durch mehr als die Hälfte der Lehrkräfte bestätigt.

Noch geringer fiel die Wirkung auf das Kollegium aus. Das Testverfahren *Online Diagnose* schien laut den Aussagen der Lehrkräfte nicht zu den genannten Wirkungen auf das Kollegium geführt zu haben. Der Mehrheit der Lehrkräfte erschienen die Kooperationen und die Absprachen im Kollegium eher unverändert. Die geringen Mittelwerte deuten darauf hin, dass die Items zur Variablen Wirkung auf das Kollegium möglicherweise ungünstig gewählt waren. Inhaltlich zielten die Items auf eine durch das Testverfahren *Online Diagnose* verstärkte bzw. häufigere Kooperation im Kollegium ab. Eine bereits bestehende umfangreiche Kooperation müsste jedoch nicht verstärkt werden.

Für die Lehrkräfte selbst veränderte sich laut deren Angaben ebenfalls eher wenig. Vor allem nicht hinsichtlich der Unterrichtsplanung oder des Gefühls, beansprucht zu sein.

Stärkere Veränderungen fanden sich hinsichtlich der Erleichterung der individuellen Förderung im Unterricht und der Steigerung der eigenen Diagnosekompetenz. Im Rahmen der Studie BeLesen hält Schneewind (2006) ein ähnliches Ergebnis fest. Ein knappes Viertel der Lehrkräfte gab an, aufgrund der Rückmeldung die Lehrmethode verändert zu haben ($N = 52$). Auch Kühle und Peek (2007) berichten von einer Veränderung der Unterrichtsgestaltung und -methode von lediglich 25 % der Lehrkräfte ($N = 232$). Ebenso berichten Schrader und Helmke (2004) von ca. 25 bis 30 % vorgenommenen Veränderungen durch die Lehrkräfte im Rahmen der Fachdidaktik bzw. -methodik ($N = 81$).

6 Teilstudie 2: Prozessschritte der Nutzung von Ergebnisrückmeldungen

Im Rahmen empirischer Forschung wurde häufig argumentiert, Ergebnisrückmeldungen aus Testverfahren seien notwendig, aber nicht ausschlaggebend für Verbesserungen von Lehre und Lernen (Helmke & Hosenfeld, 2005; vgl. auch Arnold, 2002; Terhart, 2002; Watermann, Stanat, Kunter, Klieme & Baumert, 2003; Müller, 2009). Die Bereitstellung möglichst objektiver Informationen über Schülerleistungen ist für Verbesserungen von Lehre und Lernen notwendig, doch ebenso ist zu untersuchen, ob diese Informationen zur Verbesserung von Lehre und Lernen tatsächlich genutzt werden.

In Konsequenz daraus postulieren Helmke und Hosenfeld (2005) im Zyklenmodell der Rückmeldung vier Prozessschritte zur Nutzung von Ergebnisrückmeldungen und Fördermaßnahmen: Rezeption, Reflexion, Aktion und Evaluation.⁴¹ Die beiden Autoren sehen in diesem Modell nicht nur die Interpretation der Ergebnisrückmeldungen und die Anwendung daraus abgeleiteter Maßnahmen vor, sondern auch die Reflexion der Ergebnisrückmeldungen und die Evaluation des gesamten Prozesses. Dem Modell liegt die Annahme zugrunde, dass das Durchlaufen der vier Prozessschritte durch die gezielte Auseinandersetzung mit den Ergebnisrückmeldungen und den Fördermaterialien Verbesserungen in Lehre und Lernen bewirken kann (Helmke & Hosenfeld, 2005). In die vorliegende Untersuchung wurde zusätzlich zu den vier Prozessschritten des Zyklenmodells der Rückmeldung die subjektive Qualitätseinschätzung des Testverfahrens als eine Voraussetzung für gute Instrumente, Ergebnisrückmeldungen und Maßnahmen mit aufgenommen.

Dem Modell folgend wurde angenommen, dass nach der Rückmeldung der Ergebnisse das Durchlaufen eines jeweiligen Prozessschrittes einen Zusammenhang mit der Wirkung des Testverfahrens *Online Diagnose* aufzeigt. Deshalb wurde untersucht, ob die Wirkung auf Lernende, die Wirkung auf das Kollegium und die Wirkung auf Lehrende (Prozessschritt Evaluation) mit der subjektiven Qualitätseinschätzung des Testverfahrens durch die Lehrkräfte sowie den Prozessschritten Rezeption, Reflexion und Aktion assoziiert war. Der Erhebungszeitpunkt der Befragung zum Testverfahren *Online Diagnose* war so gewählt, dass die Lehrkräfte die Prozessschritte Rezeption, Reflexion und Aktion bereits durchlaufen

⁴¹ Das der Studie zugrunde liegende theoretische Zyklenmodell der Rückmeldung (z. B. Helmke & Hosenfeld) wird in Kapitel 2.3 näher ausgeführt.

haben sollten. Zudem sollten die Lehrkräfte mittels der schriftlichen Befragung zu evaluativen Überlegungen in Bezug auf den gesamten Prozess der Durchführung des Testverfahrens *Online Diagnose* angeregt werden. Hinsichtlich des Prozessschrittes Evaluation wurde davon ausgegangen, dass den Lehrkräften eine Aussage zur Wirkung des Testverfahrens *Online Diagnose* nicht nur auf Schülerinnen und Schüler (im Folgenden als Lernende bezeichnet), sondern auch auf das Kollegium und auf die Lehrenden selbst möglich war.

Zur Untersuchung der Nutzung von Ergebnisrückmeldungen anhand der genannten Prozessschritte wurden drei Hypothesen aufgestellt. Es wurde erwartet, dass die durch die Lehrkraft wahrgenommene subjektive Qualitätseinschätzung des Testverfahrens sowie die Prozessschritte Rezeption, Reflexion und Aktion in Zusammenhang mit dem Prozessschritt Evaluation stehen. Der Prozessschritt Evaluation wurde hierbei als abhängige Variable behandelt. Die drei Hypothesen wurden wie folgt formuliert:

(a) Die subjektive Qualitätseinschätzung und die Prozessschritte Rezeption, Reflexion und Aktion sind mit der durch die Lehrkräfte eingeschätzten Wirkung auf Lernende (Prozessschritt Evaluation) assoziiert.

(b) Die subjektive Qualitätseinschätzung und die Prozessschritte Rezeption, Reflexion und Aktion sind mit der durch die Lehrkräfte eingeschätzten Wirkung auf das Kollegium (Prozessschritt Evaluation) assoziiert.

(c) Die subjektive Qualitätseinschätzung sowie die Prozessschritte Rezeption, Reflexion und Aktion sind mit der durch die Lehrkräfte eingeschätzten Wirkung auf Lehrende (Prozessschritt Evaluation) assoziiert.

In den folgenden Abschnitten werden die eingesetzten Instrumente vorgestellt sowie die statistischen Analysen zur Überprüfung der Hypothesen dargelegt. Es folgt die Darstellung der Ergebnisse, die abschließend diskutiert werden.

6.1 Methode und Instrumente

Im Folgenden werden die in Teilstudie 2 verwendeten Instrumente dargestellt. Die Items wurden im Rahmen einer Lehrkraftbefragung zum Testverfahren *Online Diagnose* entwickelt.⁴² Itembeispiele zur jeweiligen Skala sind in einem kurzen Überblick in Tabelle 24

⁴² Die hierzu erfasste Stichprobe umfasst Lehrkräfte der fünften Jahrgangsstufe an Haupt-, Werkreal- und Gemeinschaftsschulen. Weitere Informationen zur verwendeten Stichprobe sind in Kapitel 4 nachzulesen. Ebenso sind Faktoren- und Reliabilitätsanalysen zur Skalenbildung in Kapitel 5.2.4 dargestellt.

aufgeführt. Die Variablen zur subjektiven Qualitätseinschätzung und zu den Prozessschritten Rezeption, Reflexion und Aktion werden als unabhängige Variablen eingesetzt. Die drei Variablen zum Prozessschritt Evaluation werden in den nachfolgenden Analysen als abhängige Variablen behandelt. Die Variablen zum Prozessschritt *Aktion* wurden mit einer fünfstufigen Likert-Skala gemessen (1 = *viel weniger als vorher*, 2 = *weniger als vorher*, 3 = *genauso viel wie vorher*, 4 = *mehr als vorher*, 5 = *viel mehr als vorher*). Alle weiteren Variablen wurden mit einer vierstufigen Likert-Skala gemessen (1 = *trifft nicht zu*, 2 = *trifft eher nicht zu*, 3 = *trifft eher zu*, 4 = *trifft zu*).

Tabelle 24

Itembeispiele und Cronbachs α zu den unabhängigen und abhängigen Variablen

Variablen	α	Itemanzahl	Itembeispiele
<i>Prozessschritt Evaluation:</i>			
Wirkung auf Lernende	.86	6	„Die förderbedürftigen Schülerinnen und Schüler erzielen nun bessere Noten“
Wirkung auf das Kollegium	.76	3	„Im Kollegium besprechen wir nun häufiger die Leistungen einzelner Schülerinnen und Schüler“
Wirkung auf Lehrende	.69	4	„Ich konnte meine Diagnosekompetenz steigern“
<i>Subjektive Qualitätseinschätzung:</i>			
Eignung Testaufgaben	.89	20	„Halten Sie diesen Aufgabentyp für geeignet, um einen Förderbedarf festzustellen?“
Bezug zum Bildungsplan der Grundschule	.77	3	„Der Haupttest weist einen klaren Bezug zum Bildungsplan der Primarstufe auf“
Qualität des Haupttests	.53	2	„Der Haupttest ist ein gutes Instrument, um Förderbedarfe festzustellen“
Qualität der Ergebnisrückmeldung	.75	3	„Die Informationen zu den jeweiligen Schülerinnen und Schülern sind ausreichend“
Qualität der Darstellung der Ergebnisrückmeldungen	.80	3	„Die Grafiken sind selbsterklärend“
Qualität der Fördermaterialien	.74	2	„Die Fördermappe enthält gute Aufgaben“
<i>Prozessschritt Rezeption:</i>			
Selbstwirksamkeit im Umgang mit Ergebnisrückmeldungen	.74	4	„Es fällt mir leicht, die Ergebnisse des Haupttests zu interpretieren“
Zeitaufwand für die Nutzung der Ergebnisse	.69	4	„Es ist mir zeitlich nicht möglich, die Ergebnisrückmeldungen von allen Schülerinnen und Schülern ausführlich zu lesen“
<i>Prozessschritt Reflexion:</i>			
Informationsgewinn aus Ergebnisrückmeldungen	.75	3	„Ohne die Ergebnisrückmeldungen des Haupttests würde ich auf manche Probleme bzw. Schwächen von Schülerinnen und Schülern nicht stoßen“

(Fortsetzung)

Variablen	α	Item- anzahl	Itembeispiele
<i>Prozessschritt Aktion:</i>			
Gezielte Förderung	.63	3	„Welche Auswirkungen hatte die <i>Online Diagnose?</i> – Gezielte individuelle Förderung einzelner Schülerinnen und Schüler“
Dummy: Bearbeitung im Deutschunterricht			„Die Förderaufgaben werden im Deutschunterricht bearbeitet“
Dummy: Bearbeitung im Förderunterricht			„Die Förderaufgaben werden in einem speziellen Förderunterricht bearbeitet.“
Dummy: Einbindung der Fördermaterialien			„Die Lösungen der Förderaufgaben werden in der Schule besprochen.“

Anmerkung. Cronbachs α .

6.2 Statistische Analysen

Zur Prüfung der Hypothesen wurden lineare multiple Regressionsanalysen schrittweise durchgeführt. Die Ergebnisse werden anhand des standardisierten Regressionskoeffizienten β in den Tabellen 26 bis 28 berichtet (vgl. z.B. Cohen, Cohen, West & Aiken, 2003). Außerdem werden die Standardfehler aufgeführt. Als Kennwert für die Stärke des Zusammenhangs wird der Determinationskoeffizient R^2 dargestellt (vgl. z.B. Steyer & Eid, 2001). Als abhängige Variablen wurde die Wirkung auf Lernende, die Wirkung auf das Kollegium und die Wirkung auf Lehrende (Prozessschritt Evaluation) eingesetzt. Als unabhängige Variablen wurden die in Tabelle 24 dargestellten Domänen subjektive Qualitätseinschätzung sowie die Prozessschritte Rezeption, Reflexion und Aktion eingesetzt. Für jede Hypothese wurden insgesamt vier Regressionsanalysen berechnet. Zur Hypothese a wurden die Modelle 1 bis 4 berechnet, wobei die abhängige Variable Wirkung auf Lernende geprüft wurde (Tabelle 26). Zur Prüfung von Hypothese b wurden die Modelle 5 bis 8 berechnet, wobei die abhängige Variable Wirkung auf das Kollegium geprüft wurde (Tabelle 27). Zur Prüfung von Hypothese c wurden die Modelle 9 bis 12 berechnet, wobei die abhängige Variable Wirkung auf Lehrende geprüft wurde (Tabelle 28).

Vorbereitende Analysen wurden mit SPSS 22 durchgeführt. Die Normalverteilung der Residuen – als eine Voraussetzung der linearen Regression (z. B. Eid, Gollwitzer & Schmitt, 2011) – konnte durch den Kolmogorov-Smirnoff-Test nicht bestätigt werden. Aus diesem Grund wurden alle linearen multiplen Regressionsanalysen mit dem robusteren MLR-Schätzer (robust maximum likelihood estimator) in Mplus Version 7.1 (Muthén &

Muthén, 1998-2013) durchgeführt. Für alle Regressionskoeffizienten wurden die Standardized Model Results (STDYX) angegeben. Die bivariaten Zusammenhänge wurden ebenfalls mit dem Programm Mplus berechnet. Zum Umgang mit fehlenden Werten wurde das Full Information Maximum Likelihood Verfahren (FIML) verwendet. Das Verfahren kann lediglich bei Fällen, die auf allen Variablen fehlende Werte aufweisen, nicht angewandt werden (z.B. Geiser, 2011). Es stellt damit eine der momentan am besten geeigneten Methoden zum Umgang mit fehlenden Werten dar (z.B. Lüdtke, Robitzsch, Trautwein & Köller, 2007; Enders, 2010). Insgesamt lagen die fehlenden Werte bei den für diese Analysen verwendeten Variablen zwischen 1.3 % und 7.3 %.

6.3 Ergebnisse

Interkorrelationen der Variablen. Zunächst werden bivariate Zusammenhänge der einzelnen Variablen in Tabelle 25 berichtet. Die höchsten negativen und positiven Korrelationen waren im Bereich der Interkorrelationen der unabhängigen Variablen zu finden. Diese reichten von $r = -.37$ ($p < .001$) bis $r = .51$ ($p < .001$). Die stärkste negative Interkorrelation bestand zwischen den beiden Dummy-Variablen zu den Fördermaterialien: Bearbeitung im Deutschunterricht und Bearbeitung im Förderunterricht. Die stärksten positiven Interkorrelationen der unabhängigen Variablen bestanden zwischen den Variablen Qualität der Ergebnisrückmeldungen und Qualität der Darstellung der Ergebnisrückmeldungen sowie der Selbstwirksamkeit im Umgang mit Ergebnisrückmeldungen und Qualität der Darstellung der Ergebnisrückmeldungen. Da der stärkste Zusammenhang lediglich eine mittlere Korrelation von $r = .51$ ($p < .001$) aufwies, kann eine Multikollinearität ausgeschlossen werden (Marsh, Downson, Pietsch & Walker, 2004).

Tabelle 25
Interkorrelationen der abhängigen und unabhängigen Variablen

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Abhängige Variablen:																
<i>Prozessschritt Evaluation:</i>																
1. Wirkung auf Lernende																
2. Wirkung auf das Kollegium	.28**															
3. Wirkung auf Lehrende	.44**	.33**														
Unabhängige Variablen:																
<i>Subjektive Qualitätseinschätzung:</i>																
4. Eignung Testaufgaben	.09	-.03	.23*													
5. Bezug zum Bildungsplan der Grundschule	.11	.14	.35**	.34**												
6. Qualität des Haupttests	.41**	.14	.39**	.23*	.46**											
7. Qualität der Ergebnisrückmeldung	.16	.03	.00	.14	.22*	.21*										
8. Qualität der Darstellung der Ergebnisrückmeldungen	.14	-.08	.10	.21**	.20**	.22*	.51**									
9. Qualität der Fördermaterialien	.44**	.15*	.37**	.18*	.33**	.46**	.13	.18								
<i>Prozessschritt Rezeption:</i>																
10. Selbstwirksamkeit im Umgang mit Ergebnissen	.24*	.02	.33**	.28**	.37**	.48**	.38**	.51**	.43**							
11. Zeitaufwand für die Nutzung der Ergebnisse	.16	.11	-.10	-.14	-.10	.03	-.09	-.05	-.03	-.17*						

(Fortsetzung)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
<i>Prozessschritt Reflexion:</i>																
12. Informationsgewinn aus Ergebnisrückmeldungen	.27**	.16*	.38**	.19*	.11	.35**	-.07	-.01	.23*	.09	-.05					
<i>Prozessschritt Aktion:</i>																
13. Gezielte Förderung	.36**	.16*	.39**	.21**	.19*	.31**	.03	.19*	.22*	.30**	.06	.37**				
14. Dummy: Bearbeitung im Deutschunterricht	.03	-.13	.13	-.02	-.02	.10	-.14	-.06	-.05	-.07	-.12	.08	-.18*			
15. Dummy: Bearbeitung im Förderunterricht	.11	.13	.21*	-.03	.14	.03	-.06	.03	.13	.10	.00	-.07	.14	-.37**		
16. Dummy: Einbindung der Fördermaterialien	.17*	.11	.22*	-.13	-.07	.05	-.03	.00	.01	.12	-.02	-.03	.05	.13	-.06	

Anmerkung. $N = 148$; * $p < .05$; ** $p < .01$.

Wirkung auf Lernende in Abhängigkeit von der subjektiven Qualitätseinschätzung und den Prozessschritten Rezeption, Reflexion und Aktion. Im Rahmen der Hypothese a wurde erwartet, dass die Wirkung auf Lernende mit den unabhängigen Variablen zur subjektiven Qualitätseinschätzung sowie zu den Prozessschritten Rezeption, Reflexion und Aktion assoziiert ist. Die Ergebnisse werden in Tabelle 26 berichtet.

In Modell 1 wurden die folgenden unabhängigen Variablen der subjektiven Qualitätseinschätzung in die Regressionsanalyse aufgenommen: Eignung Testaufgaben, Bezug zum Bildungsplan der Grundschule, Qualität des Haupttests, Qualität der Ergebnisrückmeldungen, Qualität der Darstellung der Ergebnisrückmeldungen und Qualität der Fördermaterialien. Hierbei standen lediglich die unabhängigen Variablen Qualität des Haupttests ($\beta = .32; p < .001$) und Qualität der Fördermaterialien ($\beta = .33; p < .001$) in einem statistisch signifikanten Zusammenhang mit der abhängigen Variable Wirkung auf Lernende. Insgesamt klärten die Prädiktoren der subjektiven Qualitätseinschätzung 26 % der Varianz in der abhängigen Variable auf.

In Modell 2 wurden zusätzlich zu den unabhängigen Variablen der subjektiven Qualitätseinschätzung die folgenden unabhängigen Variablen zum Prozessschritt Rezeption aufgenommen: Selbstwirksamkeit im Umgang mit den Ergebnisrückmeldungen und Zeitaufwand für die Nutzung der Ergebnisrückmeldungen. Zu den beiden Variablen Qualität des Haupttests und Qualität der Fördermaterialien war in Modell 2 die Variable Zeitaufwand für die Nutzung der Ergebnisrückmeldungen signifikant mit der abhängigen Variable Wirkung auf Lernende assoziiert ($\beta = .16; p < .05$). Insgesamt wurden 29 % der Varianz in der abhängigen Variable aufgeklärt.

In Modell 3 wurde zusätzlich zu den Variablen der subjektiven Qualitätseinschätzung und denen des Prozessschritts Rezeption die unabhängige Variable zum Prozessschritt Reflexion mit aufgenommen: Informationsgewinn aus Ergebnisrückmeldungen. Die folgenden vier unabhängigen Variablen erwiesen sich in Modell 3 als signifikante Prädiktoren: Qualität des Haupttests, Qualität der Fördermaterialien, Zeitaufwand für die Nutzung der Ergebnisrückmeldungen und Informationsgewinn aus Ergebnisrückmeldungen ($\beta = .14; p < .05$). In Modell 3 wurden insgesamt 31 % der Varianz in der abhängigen Variable aufgeklärt.

In Modell 4 wurden schließlich zusätzlich zu den Variablen der subjektiven Qualitätseinschätzung und denen der Prozessschritte Rezeption und Reflexion die unabhängigen Variablen zum Prozessschritt Aktion in das Modell aufgenommen: Gezielte Förderung und

die Dummies Bearbeitung im Deutschunterricht, Bearbeitung im Förderunterricht, Einbindung der Fördermaterialien. Die in Modell 4 hinzugenommenen Prädiktoren gezielte Förderung ($\beta = .22$; $p < .05$) und der Dummy Einbindung der Fördermaterialien ($\beta = .16$; $p < .05$) standen in einem statistisch signifikanten Zusammenhang mit der abhängigen Variable Wirkung auf Lernende. Neben der Qualität des Haupttests und der Qualität der Fördermaterialien zeigte sich in Modell 4 auch die Variable Qualität der Ergebnisrückmeldungen signifikant mit der abhängigen Variable Wirkung auf Lernende assoziiert. Die unabhängige Variable Informationsgewinn aus Ergebnisrückmeldungen zeigte in Modell 4 keinen signifikanten Zusammenhang mit der abhängigen Variable Wirkung auf Lernende. Die Regressionskoeffizienten der Prädiktoren Bezug zum Bildungsplan der Grundschule, Dummy Bearbeitung im Deutschunterricht und Dummy Bearbeitung im Förderunterricht lagen auf einem Signifikanzniveau $p < .10$. In Modell 4 wurden insgesamt 40 % der Varianz in der abhängigen Variable Wirkung auf Lernende aufgeklärt.

Zusammenfassung. Die Regressionskoeffizienten der unabhängigen Variablen Qualität des Haupttests und Qualität der Fördermaterialien waren in allen vier Modellen signifikant (subjektive Qualitätseinschätzung). Die unabhängige Variable Qualität der Ergebnisrückmeldungen war lediglich in Modell 4 unter Aufnahme aller unabhängigen Variablen signifikant mit der abhängigen Variable Wirkung auf Lernende assoziiert. Eine signifikante bivariate Korrelation der unabhängigen Variable Qualität der Ergebnisrückmeldungen mit der abhängigen Variable Wirkung auf Lernende bestand nicht. Weiterhin war der Prädiktor Zeitaufwand für die Nutzung der Ergebnisrückmeldungen (Prozessschritt Rezeption) in den Modellen 2 bis 4 signifikant mit der Wirkung auf Lernende assoziiert. Im Rahmen der bivariaten Korrelationen zeigte die Variable Zeitaufwand für die Nutzung der Ergebnisrückmeldungen lediglich einen signifikant negativen Zusammenhang mit der Selbstwirksamkeit im Umgang mit Ergebnisrückmeldungen. Ein signifikant bivariater Zusammenhang mit der abhängigen Variable Wirkung auf Lernende bestand nicht. Der Zusammenhang wurde also durch weitere Prädiktoren beeinflusst. Weiterhin war die unabhängige Variable Informationsgewinn aus Ergebnisrückmeldungen (Prozessschritt Reflexion) lediglich in Modell 3 gering signifikant mit der abhängigen Variable Wirkung auf Lernende assoziiert. Durch die Aufnahme der unabhängigen Variable Informationsgewinn aus Ergebnisrückmeldungen nahm die Varianzaufklärung nur geringfügig zu. Unter Aufnahme aller unabhängigen Vari-

ablen waren die Variablen gezielte Förderung und die Dummy-Variable Einbindung der Fördermaterialien (Prozessschritt Aktion) signifikant mit der Wirkung auf Lernende assoziiert. Die anfangs aufgestellte Hypothese konnte also größtenteils bestätigt werden. Variablen zu nahezu jedem Prozessschritt bzw. der subjektiven Qualitätseinschätzung waren mit der Wirkung auf Lernende assoziiert. Lediglich der Prozessschritt Reflexion zeigte in Modell 4 keinen Zusammenhang und eine geringe Varianzaufklärung.

Wirkung auf das Kollegium in Abhängigkeit von der subjektiven Qualitätseinschätzung und den Prozessschritten Rezeption, Reflexion und Aktion. Die in Hypothese b erwarteten Assoziationen der abhängigen Variable Wirkung auf das Kollegium mit den unabhängigen Variablen zur subjektiven Qualitätseinschätzung und zu den Prozessschritten Rezeption, Reflexion und Aktion werden in den Modellen 5 bis 8 in Tabelle 27 dargestellt. Wie bereits die bivariaten Korrelationen andeuteten, zeigte sich bei den Regressionsanalysen zur abhängigen Variable Wirkung auf das Kollegium ein schwaches Ergebnis. Keines der in Tabelle 27 dargestellten Modelle 5 bis 8 konnte signifikante Zusammenhänge bestätigen. Auch unter Berücksichtigung aller unabhängigen Variablen konnten keine signifikanten Assoziationen aufgedeckt werden (Modell 8). Gründe hierfür werden nachfolgend diskutiert.

Wirkung auf Lehrende in Abhängigkeit von der subjektiven Qualitätseinschätzung und den Prozessschritten Rezeption, Reflexion und Aktion. Die Ergebnisse der im Rahmen der Hypothese c erwarteten Assoziationen der abhängigen Variable Wirkung auf Lehrende mit den Variablen zur subjektiven Qualitätseinschätzung und zu den Prozessschritten Rezeption, Reflexion und Aktion werden in den Modellen 9 bis 12 in Tabelle 28 dargestellt.

In Modell 9 wurden die folgenden unabhängigen Variablen zur subjektiven Qualitätseinschätzung in die Regressionsanalyse aufgenommen: Eignung Testaufgaben, Bezug zum Bildungsplan der Grundschule, Qualität des Haupttests, Qualität der Ergebnisrückmeldungen, Qualität der Darstellung der Ergebnisrückmeldungen und Qualität der Fördermaterialien. Drei der in Modell 9 dargestellten unabhängigen Variablen standen in statistisch signifikantem Zusammenhang mit der abhängigen Variable Wirkung auf Lehrende: Bezug zum Bildungsplan der Grundschule ($\beta = .17$; $p < .05$), Qualität des Haupttests ($\beta = .22$; $p < .05$) und Qualität der Fördermaterialien ($\beta = .21$; $p < .05$). Der Prädiktor Qualität der Ergebnisrückmeldungen war zwar gering, lag jedoch noch auf einem Signifikanzniveau $p < .10$. Die

genannten unabhängigen Variablen erklärten einen Varianzanteil von 25 % in der abhängigen Variable.

In Modell 10 wurden zusätzlich zu den unabhängigen Variablen der subjektiven Qualitätseinschätzung die folgenden unabhängigen Variablen zum Prozessschritt Rezeption aufgenommen: Selbstwirksamkeit im Umgang mit den Ergebnisrückmeldungen und Zeitaufwand für die Nutzung der Ergebnisrückmeldungen. Die Zusammenhänge der in Modell 10 zusätzlich aufgenommenen Prädiktoren Selbstwirksamkeit im Umgang mit Ergebnisrückmeldungen und Zeitaufwand für die Nutzung der Ergebnisrückmeldungen waren nicht signifikant. Ebenfalls lag der Zusammenhang des Prädiktors Qualität der Fördermaterialien nun auf einem Signifikanzniveau $p < .10$. Allerdings zeigte sich in Modell 10 ein schwacher negativer Zusammenhang der abhängigen Variable Wirkung auf Lehrende mit der unabhängigen Variable Qualität der Ergebnisrückmeldungen. Durch die Aufnahmen der zusätzlichen Prädiktoren verringerten sich die signifikanten Regressionskoeffizienten der unabhängigen Variablen Bezug zum Bildungsplan der Grundschule und Qualität des Haupttests minimal. In Modell 10 steigerte sich die aufgeklärte Varianz in der abhängigen Variable nur gering auf 27 %.

In Modell 11 wurde zusätzlich zu den genannten Variablen der subjektiven Qualitätseinschätzung und des Prozessschritts Rezeption die folgende unabhängige Variable zum Prozessschritt Reflexion aufgenommen: Informationsgewinn aus den Ergebnisrückmeldungen. Die unabhängige Variable Informationsgewinn aus den Ergebnisrückmeldungen stand in einem statistisch signifikanten Zusammenhang mit der abhängigen Variable Wirkung auf Lehrende ($\beta = .26$; $p < .01$). Außerdem war die unabhängige Variable Bezug zum Bildungsplan der Grundschule signifikant mit der abhängigen Variable assoziiert. Der Zusammenhang der abhängigen Variable Wirkung auf Lehrende mit der unabhängigen Variable Selbstwirksamkeit im Umgang mit den Ergebnisrückmeldungen lag nun auf einem Signifikanzniveau $p < .10$. In Modell 11 wurde insgesamt ein Varianzanteil von 32 % in der abhängigen Variable Wirkung auf Lehrende aufgeklärt.

In Modell 12 wurden schließlich zusätzlich zu den Variablen der subjektiven Qualitätseinschätzung und denen der Prozessschritte Rezeption und Reflexion die folgenden unabhängigen Variablen zum Prozessschritt Aktion simultan in das Modell einbezogen: Die gezielte Förderung sowie die Dummies Bearbeitung im Deutschunterricht, Bearbeitung im Förderunterricht und Einbindung der Fördermaterialien. Neben den beiden unabhängigen

Variablen Bezug zum Bildungsplan der Grundschule und Informationsgewinn aus Ergebnismeldungen standen die Prädiktoren gezielte Förderung ($\beta = .20$; $p < .05$) und die Dummies Bearbeitung im Deutschunterricht ($\beta = .23$; $p < .01$), Bearbeitung im Förderunterricht ($\beta = .25$; $p < .001$) und Einbindung der Fördermaterialien ($\beta = .20$; $p < .01$) in statistisch signifikantem Zusammenhang mit der abhängigen Variable Wirkung auf Lehrende. Die unabhängige Variable Qualität der Fördermaterialien lag in Modell 12 auf einem Signifikanzniveau $p < .10$. Die aufgeklärte Varianz in der abhängigen Variablen betrug insgesamt 45 %.

Zusammenfassung. Die Regressionskoeffizienten der unabhängigen Variable Bezug zum Bildungsplan der Grundschule waren zwar schwach, wurden aber in allen Modellen (9 bis 12) signifikant. Weitere Variablen der subjektiven Qualitätseinschätzung, die Qualität des Haupttests, die Qualität der Ergebnismeldungen und die Qualität der Fördermaterialien waren lediglich in den Modellen 9 bzw. 10 mit der abhängigen Variable Wirkung auf Lehrende assoziiert. Die unabhängige Variable Selbstwirksamkeit im Umgang mit den Ergebnismeldungen (Prozessschritt Rezeption) lag in Modell 11 auf einem Signifikanzniveau $p < .10$. In Modell 12 unter Aufnahme der Variablen zum Prozessschritt Aktion war die unabhängige Variable Selbstwirksamkeit im Umgang mit den Ergebnismeldungen nicht signifikant. Weiterhin war der Prädiktor Informationsgewinn aus Ergebnismeldungen des Prozessschrittes Reflexion in den Modellen 11 und 12 signifikant mit der Wirkung auf Lehrende assoziiert. Die Prädiktoren zum Prozessschritt Aktion (Modell 12) waren unter Aufnahme aller unabhängigen Variablen signifikant mit der Wirkung auf Lehrende assoziiert. Die zur Wirkung auf Lehrende aufgestellte Hypothese konnte größtenteils bestätigt werden. Variablen der Prozessschritte Reflexion und Aktion sowie der subjektiven Qualitätseinschätzung waren mit den abhängigen Variablen assoziiert. Lediglich die Variablen zum Prozessschritt Rezeption zeigten ein schwaches bzw. nicht-signifikantes Ergebnis.

Zusammenfassung insgesamt. Zusammenfassend kann festgehalten werden, dass die Hypothesen nicht uneingeschränkt bestätigt werden konnten. Assoziationen der Prädiktoren mit der Wirkung auf das Kollegium konnten nicht bestätigt werden (Hypothese b). Zu den Assoziationen der Prädiktoren mit der Wirkung auf Lernende und der Wirkung auf Lehrende wurde folgendes deutlich (Hypothesen a und c): Auffallend war, dass hinsichtlich der Wirkung auf Lernende die Lehrkräfte eher Qualitätsaspekte, den Zeitaufwand und die

Förderung in den Fokus rückten, wohingegen sie bei der Wirkung auf Lehrende den Bezug zum Bildungsplan, den Informationsgewinn aus den Ergebnisrückmeldungen und die Förderung für relevant hielten. Die Ergebnisse werden nachfolgend diskutiert.

Tabelle 26

Lineare multiple Regression zur abhängigen Variable Wirkung auf Lernende

Prädiktor	Wirkung auf Lernende											
	Modell 1			Modell 2			Modell 3			Modell 4		
	β	<i>SE</i>	<i>p</i>	β	<i>SE</i>	<i>p</i>	β	<i>SE</i>	<i>p</i>	β	<i>SE</i>	<i>p</i>
Subjektive Qualitätseinschätzung:												
Eignung Testaufgaben	.01	0.07	.92	.02	0.06	.66	.01	0.07	.92	.03	0.06	.66
Bezug zum Bildungsplan Grundschule	-.15	0.09	.11	-.13	0.09	.13	-.12	0.09	.16	-.14	0.08	.08
Qualität des Haupttests	.31	0.09	.00	.29	0.10	.00	.23	0.10	.02	.19	0.10	.05
Qualität der Ergebnismrückmeldung	.10	0.12	.40	.11	0.11	.33	.13	0.11	.24	.21	0.10	.05
Qualität der Darstellung der Ergebnismrückmeldungen	-.02	0.09	.85	-.01	0.10	.93	-.00	0.10	.97	-.03	0.10	.74
Qualität der Fördermaterialien	.33	0.10	.00	.33	0.10	.00	.32	0.10	.00	.32	0.09	.00
Rezeption:												
Selbstwirksamkeit im Umgang mit Ergebnismrückmeldungen				-.01	0.10	.89	.01	0.10	.96	-.08	0.10	.43
Zeitaufwand für die Nutzung der Ergebnisse				.16	0.08	.05	.17	0.08	.03	.17	0.08	.03
Reflexion:												
Informationsgewinn aus Ergebnismrückmeldungen							.14	0.07	.04	.10	0.07	.19
Aktion:												
Gezielte Förderung										.22	0.09	.02
Dummy: Bearbeitung im Deutschunterricht										.13	0.07	.06
Dummy: Bearbeitung im Förderunterricht										.14	0.08	.07
Dummy: Einbindung der Fördermaterialien										.16	0.07	.02
<i>R</i> ²	.26			.29			.31			.40		

Anmerkung. β = standardisierte Regressionskoeffizient, *p* = Signifikanzniveau, *SE* = Standardfehler, *R*² = aufgeklärte Varianz.

Tabelle 27

Lineare multiple Regression zur abhängigen Variable Wirkung auf das Kollegium

Prädiktor	Wirkung auf das Kollegium											
	Modell 5			Modell 6			Modell 7			Modell 8		
	β	<i>SE</i>	<i>p</i>	β	<i>SE</i>	<i>p</i>	β	<i>SE</i>	<i>p</i>	β	<i>SE</i>	<i>p</i>
Subjektive Qualitätseinschätzung:												
Eignung Testaufgaben	-.08	0.09	.37	-.06	0.09	.47	-.09	0.09	.33	-.07	0.08	.41
Bezug zum Bildungsplan Grundschule	.12	0.10	.24	.13	0.10	.20	.14	0.10	.16	.13	0.10	.19
Qualität des Haupttests	.07	0.10	.48	.07	0.11	.52	.01	0.11	.91	.02	0.11	.84
Qualität der Ergebnismrückmeldung	.08	0.10	.44	.09	0.10	.37	.11	0.10	.27	.12	0.10	.20
Qualität der Darstellung der Ergebnismrückmeldungen	-.17	0.10	.09	-.15	0.12	.20	-.15	0.12	.20	-.15	0.11	.17
Qualität der Fördermaterialien	.11	0.09	.19	.13	0.09	.16	.11	0.09	.22	.10	0.09	.26
Rezeption:												
Selbstwirksamkeit im Umgang mit Ergebnissen				-.04	0.12	.76	-.02	0.12	.85	-.10	0.12	.42
Zeitaufwand für die Nutzung der Ergebnisse				.10	0.08	.22	.11	0.08	.18	.09	0.08	.29
Reflexion:												
Informationsgewinn aus Ergebnismrückmeldungen							.14	0.09	.11	.13	0.10	.18
Aktion:												
Gezielte Förderung										.10	0.09	.29
Dummy: Bearbeitung im Deutschunterricht										-.10	0.08	.24
Dummy: Bearbeitung im Förderunterricht										.09	0.08	.30
Dummy: Einbindung der Fördermaterialien										.14	0.08	.09
<i>R</i> ²	.06			.07			.09			.14		

Anmerkung. β = standardisierte Regressionskoeffizient, p = Signifikanzniveau, *SE* = Standardfehler, R^2 = aufgeklärte Varianz.

Tabelle 28

Lineare multiple Regression zur abhängigen Variable Wirkung auf Lehrende

Prädiktor	Wirkung auf Lehrende											
	Modell 9			Modell 10			Modell 11			Modell 12		
	β	<i>SE</i>	<i>p</i>	β	<i>SE</i>	<i>p</i>	β	<i>SE</i>	<i>p</i>	β	<i>SE</i>	<i>p</i>
Subjektive Qualitätseinschätzung:												
Eignung Testaufgaben	.09	0.10	.38	.07	0.10	.48	.03	0.10	.75	.07	0.08	.34
Bezug zum Bildungsplan Grundschule	.17	0.08	.04	.15	0.08	.05	.18	0.07	.02	.16	0.07	.03
Qualität des Haupttests	.22	0.08	.01	.20	0.08	.02	.10	0.09	.26	.02	0.08	.78
Qualität der Ergebnisrückmeldung	-.14	0.08	.08	-.16	0.08	.05	-.12	0.08	.12	-.01	0.07	.89
Qualität der Darstellung der Ergebnisrückmeldungen	.03	0.09	.76	-.03	0.09	.77	-.02	0.09	.80	-.06	0.08	.44
Qualität der Fördermaterialien	.21	0.09	.02	.17	0.10	.08	.14	0.10	.15	.14	0.08	.08
Rezeption:												
Selbstwirksamkeit im Umgang mit Ergebnissen				.15	0.10	.12	.18	0.09	.06	.10	0.09	.27
Zeitaufwand für die Nutzung der Ergebnisse				-.07	0.07	.36	-.05	0.07	.49	-.03	0.07	.67
Reflexion:												
Informationsgewinn aus Ergebnisrückmeldungen							.26	0.08	.00	.23	0.07	.00
Aktion:												
Gezielte Förderung										.20	0.08	.01
Dummy: Bearbeitung im Deutschunterricht										.23	0.08	.00
Dummy: Bearbeitung im Förderunterricht										.25	0.07	.00
Dummy: Einbindung der Fördermaterialien										.20	0.07	.00
<i>R</i> ²	.25			.27			.32			.45		

Anmerkung. β = standardisierte Regressionskoeffizient, *p* = Signifikanzniveau, *SE* = Standardfehler, *R*² = aufgeklärte Varianz.

6.4 Diskussion und Implikationen

In dieser Teilstudie wurden zur Beantwortung der oben bereits genannten Fragestellung subjektive Qualitätseinschätzungen und Prozessschritte der Nutzung von Ergebnisrückmeldungen als Prädiktoren für die Wirkung auf Lernende, die Wirkung auf das Kollegium und die Wirkung auf Lehrende untersucht: Sind die Prozessschritte der Nutzung mit der Wirkung des Testverfahrens *Online Diagnose* auf Lernende, das Kollegium und Lehrende assoziiert? Im Folgenden werden die Ergebnisse aus der vorliegenden Untersuchung wie folgt zusammengefasst und diskutiert: Wirkung auf Lernende (Hypothese a), Wirkung auf das Kollegium (Hypothese b) und Wirkung auf Lehrende (Hypothese c).

Die Annahmen zur Hypothese a konnten zum Teil bestätigt werden: Prädiktoren der subjektiven Qualitätseinschätzung sowie der Prozessschritte Rezeption und Aktion waren unter Einbezug aller Variablen signifikant mit der Wirkung auf Lernende assoziiert.

Die Annahmen im Rahmen der Hypothese b, dass die subjektive Qualitätseinschätzung sowie die Prozessschritte Rezeption, Reflexion und Aktion mit der Wirkung auf das Kollegium assoziiert sind, konnten nicht bestätigt werden.

Die Annahmen zur Hypothese c konnten teilweise bestätigt werden. Prädiktoren der subjektiven Qualitätseinschätzung sowie der Prozessschritte Reflexion und Aktion waren unter Einbezug aller Variablen signifikant mit der Wirkung auf Lehrende assoziiert.

Wirkung auf Lernende. Deutlich wurde, dass vor allem Lehrkräfte, die die Qualität des Haupttests und der Fördermaterialien der *Online Diagnose* als gut einschätzten (subjektive Qualitätseinschätzung), gleichzeitig eine Wirkung auf förderbedürftige Schülerinnen und Schüler angaben. Das bedeutet auch, dass Lehrkräfte, die die Wirkung auf Lernende höher einschätzten, die Qualität des Haupttests und der Fördermaterialien besser bewerteten. Nahmen die Lehrkräfte z.B. Verbesserungen der Kompetenzen, der Noten oder der Motivation ihrer Schülerinnen und Schüler wahr, so fiel ihre Bewertung der Qualität der Testinstrumente besser aus, auch wenn eingeräumt werden muss, dass das Signifikanzniveau niedrig ist. Es kann insofern angenommen werden, dass Lehrkräfte, die die Qualität der Testinstrumente aus der *Online Diagnose* als gut einschätzten, diese häufiger einsetzten und dadurch eine Verbesserung der Schülerleistungen erzielen konnten. Es wäre allerdings notwendig, diese Annahmen in weiteren Untersuchungen zu bestätigen, z.B. unter Hinzunahme von tatsächlichen Leistungstest, die von Schülerinnen und Schülern ausgefüllt werden. Fazit aus der vorliegenden Untersuchung ist, dass die wahrgenommene Qualität der Testinstrumente

in Zusammenhang mit der wahrgenommenen Wirkung auf Schülerleistungen steht. Das bedeutet, dass sich Schülerleistungen verbessern können, wenn die Lehrkräfte die Qualitätsmerkmale der Testinstrumente erkennen und die entsprechenden qualitativ hochwertigen Ergebnisrückmeldungen nutzen. Zusätzlich sollte überprüft werden, wie objektiv die Qualitätseinschätzungen der Lehrkräfte bewertet werden können. Würde die subjektive und objektive Qualität übereinstimmen, so könnte davon ausgegangen werden, dass sich die Verbesserung der Schülerleistungen auf die Nutzung des Testverfahrens bezieht. Stimmt die subjektive und objektive Qualität nicht überein, bliebe es bei einer rein subjektiven Wahrnehmung der Lehrkräfte. Dies konnte jedoch mit den vorliegenden Daten nicht überprüft werden.

Weiterhin war der Prädiktor Zeitaufwand für die Nutzung der Ergebnisrückmeldungen (Prozessschritt Rezeption) in den Modellen 2 bis 4 signifikant mit der Wirkung auf Lernende assoziiert. Lehrkräfte, die die Nutzung der Ergebnisrückmeldung als zeitaufwendig angaben und sich demnach gezielter mit den Ergebnisrückmeldungen derjenigen Schülerinnen und Schüler beschäftigten, bei denen sie selbst einen Förderbedarf annahmen, gaben gleichzeitig eine stärkere Wirkung auf Lernende an. Dieses Ergebnis kann wie folgt interpretiert werden: Lehrkräfte, die berichteten, Verbesserungen der Kompetenzen, der Noten oder der Motivation ihrer Schülerinnen und Schüler wahrzunehmen, und die den Zeitaufwand für die Nutzung der Ergebnisrückmeldungen als hoch bezeichneten, die sich also intensiver mit den Ergebnisrückmeldungen bzw. mit speziell ausgewählten Ergebnissen beschäftigten, sahen eine Wirkung des Testverfahrens *Online Diagnose* auf Lernende. Das würde bedeuten, dass Lehrkräfte bei der selektiven Auswahl der Nutzung von Ergebnisrückmeldungen unterstützt werden sollten. In weiteren Untersuchungen, die Schülerdaten einbeziehen, sollte untersucht werden, ob der Zeitaufwand für die Nutzung der Ergebnisrückmeldungen durch die Diagnosekompetenz der Lehrkräfte moderiert bzw. modifiziert werden kann.

Unter Aufnahme aller unabhängigen Variablen waren die Variable gezielte Förderung und die Dummy-Variable Einbindung der Fördermaterialien (Prozessschritt Aktion) signifikant mit der Wirkung auf Lernende assoziiert. Die Lehrkräfte schätzten die Wirkung auf Lernende dann besonders hoch ein, wenn sie gleichzeitig angaben, Maßnahmen der gezielten Förderung häufiger als vor der Einführung des Testverfahrens *Online Diagnose* eingesetzt zu haben. Anders formuliert: Lehrkräfte, die die Maßnahmen der gezielten Förderung häufiger einsetzten, gaben an, eine stärkere Wirkung des Testverfahrens *Online Diagnose*

auf Lernende feststellen zu können. Außerdem berichteten Lehrkräfte, die angaben, die Besprechung der Lösungen in den Unterricht einzubinden, gleichzeitig von einer stärkeren Wirkung auf Lernende. Ebenso kann formuliert werden: Lehrkräfte, die von einer Wirkung des Testverfahrens auf Lernende berichteten, gaben gleichzeitig an, die Lösungen der Förderaufgaben im Unterricht zu besprechen. Lehrkräfte, die die Fördermaterialien häufiger und umfassend anwandten, waren gleichzeitig zuversichtlich, dass diese eine Wirkung auf Lernende haben. Lehrkräfte, die durch das Testverfahren *Online Diagnose* die Förderung von Schülerinnen und Schülern häufiger einsetzten und auch die Besprechung der Lösungen in den Unterricht einbezogen, nahmen z.B. Verbesserungen der Kompetenzen, der Noten und der Motivation ihrer Schülerinnen und Schüler wahr. Der Einsatz von Fördermaterialien aus dem Testverfahren *Online Diagnose* bei gleichzeitiger Einbindung derselben in den Unterrichtsalltag konnte hinsichtlich der Einschätzung der Lehrkräfte zu einer Wirkung auf Schülerleistungen führen. Neben der Bereitstellung von Fördermaterialien könnten den Lehrkräften Vorschläge für deren Anwendung zur Verfügung gestellt werden. Ein ausgearbeitetes Konzept würde den Lehrkräften die Vorbereitung für die Anwendung der Ergebnisrückmeldungen erleichtern und dadurch die Wahrscheinlichkeit der Nutzung erhöhen.

Zusammenfassung. Zusammenfassend kann festgehalten werden, dass in Zusammenhang mit der Wirkung auf Lernende unter Einbezug aller unabhängigen Variablen vor allem die Qualität des Haupttests, die Qualität der Fördermaterialien, der Zeitaufwand für die Nutzung der Ergebnisrückmeldungen, die gezielte Förderung und die Einbindung der Fördermaterialien prädiktiv waren. Die unabhängigen Variablen bezogen sich vorrangig auf die Instrumente und Maßnahmen der *Online Diagnose* und weniger auf die Qualität der Ergebnisrückmeldungen, die Selbstwirksamkeit der Lehrkräfte im Umgang mit den Ergebnisrückmeldungen oder den Informationsgewinn aus den Ergebnisrückmeldungen. Daraus konnte abgeleitet werden, dass die Lehrkräfte die Ergebnisse und Informationen aus dem Testverfahren *Online Diagnose* als weniger relevant für die Schülerinnen und Schüler ansahen als die Maßnahmen wie z.B. die Tests und Fördermaterialien selbst. Wie in Teilstudie 1 (Kapitel 5) dargelegt ist, gab die Mehrheit der Lehrkräfte zwar an, die Ergebnisse aus dem Testverfahren *Online Diagnose* mit den Schülerinnen und Schülern besprochen zu haben, die Auswertung legt jedoch den Schluss nahe, dass sie konkrete Handlungen für die Schülerinnen und Schüler wie z.B. die Anwendung der Fördermaterialien als eher relevant ansahen.

Rolff (2002) führt in seinen Überlegungen aus, dass für Lehrkräfte vor allem Handlungsanweisungen zur konkreten Förderung von Schülerinnen und Schülern bedeutsam sind. Auch die vorliegende Studie bestätigt dies, die Lehrkräfte beurteilten in Verbindung mit der Wirkung des Testverfahrens *Online Diagnose* auf Lernende vor allem die Instrumente und Maßnahmen der *Online Diagnose* als wichtig. Um die Ergebnisse abzusichern, sollten in weiteren Untersuchungen Schülerdaten einbezogen werden. Die Wirkung auf Lernende sollte anhand von Schülerdaten in einem Prä-/Post-Design überprüft werden. Außerdem könnte in weiteren Untersuchungen der in den Regressionsanalysen signifikante Zusammenhang der unabhängigen Variable Zeitaufwand für die Nutzung der Ergebnisrückmeldungen mit der abhängigen Variable Wirkung auf Lernende überprüft werden. Bei der Interpretation dieses Zusammenhangs kann ein indirekter Effekt z. B. der Diagnosekompetenz der Lehrkräfte vorliegen. Möglicherweise beschränkten sich die Lehrkräfte mit einer guten Diagnosekompetenz – angesichts des hohen Zeitaufwands für die Nutzung aller Ergebnisrückmeldungen – auf Ergebnisse von Schülerinnen und Schülern, bei denen sie selbst einen Förderbedarf sahen. Genauere Aussagen können jedoch anhand dieser Ergebnisse nicht getroffen werden, sie wären zu vage. Die Ergebnisse könnten allerdings als Grundlage für weitere Untersuchungen dienen. Ebenfalls könnten weitere Studien zum Vergleich der subjektiven Qualitätseinschätzung durch die Lehrkräfte mit der tatsächlich erfassten Testqualität neue Erkenntnisse liefern. Die vorliegende Teilstudie 2 zeigte auf, dass die subjektive Qualitätseinschätzung durch die Lehrkräfte einen Zusammenhang mit der Wirkung auf Lernende aufweist. Unklar bleibt jedoch, wie zuverlässig die subjektive Qualitätseinschätzung der Lehrkräfte in Verbindung mit der tatsächlichen Testqualität ist. An dieses Ergebnis schließen sich weitere Fragestellungen an: Woran machen Lehrkräfte eine gute Qualität von Test- und Förderinstrumenten fest? Und schließt ein solches Qualitätskriterium an objektive Kriterien der Qualität an?

Abschließend soll hier der negative Zusammenhang der unabhängigen Variable Bezug zum Bildungsplan der Grundschule mit der abhängigen Variable Wirkung auf Lernende aufgegriffen werden. Der Zusammenhang war zwar gering und lag knapp über dem Signifikanzniveau $p < .05$, jedoch ist die tendenziell negative Ausprägung des Regressionskoeffizienten zu beachten. Ein stärkerer Bezug zum Bildungsplan der Grundschule würde, laut den Angaben der Lehrkräfte, eine geringere Wirkung auf Lernende bedeuten. Anders formuliert wäre z. B. eine durch die Lehrkräfte wahrgenommene Verbesserung der Kompetenzen, der Noten oder der Motivation ihrer Schülerinnen und Schüler mit einem geringeren Bezug zum

Bildungsplan der Grundschule verknüpft. Wird die Feststellung aus dem Theorieteil berücksichtigt, dass die Aufgaben des Testverfahrens *Online Diagnose* größtenteils mit den Bildungsstandards der zweiten Jahrgangsstufe der Grundschule verglichen werden können, könnte der negative Zusammenhang wie folgt interpretiert werden: Je höher die Lehrkräfte den Bezug zum Bildungsplan der Grundschule einschätzen, das heißt größtenteils den Bezug zur zweiten Jahrgangsstufe des Bildungsplans der Grundschule, desto geringer wäre die Wirkung auf Lernende, gemessen an den Anforderungen für die fünfte Jahrgangsstufe der Haupt- und Werkrealschule.

Wirkung auf das Kollegium. Die Annahmen im Rahmen der Hypothese b konnten nicht bestätigt werden. Mögliche Interpretationen, z. B. dahingehend, dass die Ergebnisrückmeldungen und die Fördermaterialien zum Testverfahren *Online Diagnose* nicht im Rahmen der Kooperation innerhalb des Kollegiums genutzt wurden, bleiben anhand der Datenlage spekulativ. In einer Befragung von 56 Grundschullehrkräften im Rahmen der Studie BeLesen⁴³ stellte Schneewind (2007; vgl. auch Koch, 2011; Peek, 2004) einen Zusammenhang der Nutzung von Ergebnisrückmeldungen für die Kooperation im Kollegium mit einem bereits bestehenden kooperativen Klima im Kollegium fest. Die Items der vorliegenden Studie zur Variable Wirkung auf das Kollegium waren möglicherweise ungünstig gewählt. Dies bestätigten auch die in Teilstudie 1 (Kapitel 5) dargestellten geringen Mittelwerte. Die Items zielten auf eine durch das Testverfahren *Online Diagnose* verstärkte bzw. häufigere Kooperation im Kollegium ab. Eine bereits etablierte oder umfassende Kooperation innerhalb des Kollegiums muss jedoch nicht verstärkt werden. Hier würden die Inhalte der Kooperation mehr Aufschluss über die Nutzung von Informationen aus dem Testverfahren *Online Diagnose* liefern. Die Ausweitung der Lehrkraftbefragung zum Testverfahren *Online Diagnose* auf Fragen zum bestehenden kooperativen Klima innerhalb des Kollegiums wäre eine nützliche Ergänzung. Speziell müsste hier ein Bezug zur Kooperation bei externen Evaluationen bzw. Vergleichsstudien gezogen werden. Die Variablen aus der vorliegenden Studie sind hierfür zu unspezifisch. Ebenfalls deckte z. B. Schneewind (2006) auf, dass Informationen aus z. B. externen Evaluationen hauptsächlich innerhalb der Schule und innerhalb des Fachbereichs ausgetauscht wurden. Im Rahmen der Untersuchung zum Testverfahren *Online Diagnose* gaben die Lehrkräfte als häufigste Antwort an, die Ergebnisse mit anderen Lehrkräften der entsprechenden Klasse zu besprechen, gefolgt von der Nutzung der Ergebnisse für

⁴³ Berliner Längsschnittstudie zur Lesekompetenz (BeLesen)

Gespräche mit Eltern einzelner Schülerinnen und Schülern und mit einzelnen Schülerinnen und Schülern im persönlichen Gespräch. Die Ergebnisse zu Teilstudie 1 sind in Kapitel 5 ausgeführt. Damit kann die eingangs notierte mögliche Interpretation zu Hypothese b eher ausgeschlossen werden.

Wirkung auf Lehrende. Der Bezug zum Bildungsplan der Grundschule war zwar schwach, allerdings in allen Modellen (9 bis 12) signifikant mit der Wirkung auf Lehrende assoziiert. Schätzten die Lehrkräfte den Bezug des Testverfahrens *Online Diagnose* zum Bildungsplan der Grundschule als passend ein, schätzten sie gleichzeitig die Wirkung auf Lehrende als höher ein. Umgekehrt formuliert bedeutet dies, dass Lehrkräfte, die die Wirkung auf sich als Lehrende höher einschätzten, gleichzeitig den Bezug des Testverfahrens *Online Diagnose* zum Bildungsplan der Grundschule als eher gegeben ansahen. Das Ergebnis kann wie folgt interpretiert werden: War ein Bezug zum Bildungsplan der Grundschule nach der Meinung der Lehrkräfte gegeben, fühlten sich diese in ihrer Diagnose- und Förderkompetenz unterstützt. Eine nicht nur empfundene, sondern auch tatsächliche Unterstützung der Lehrkräfte in ihrer Diagnose- und Förderkompetenz kann nur stattfinden, wenn die objektive Qualität des Testverfahrens korrekt ist. Fragen zur objektiven Qualität des Testverfahrens *Online Diagnose* konnten im Rahmen dieser Studie nicht beantwortet werden und müssen in weiteren Untersuchungen nachgeholt werden. Die Überprüfung der Haupt-, Detail- und Nachttests anhand psychometrischer Verfahren wäre hierzu notwendig.

Hinsichtlich des Informationsgewinns aus Ergebnisrückmeldungen des Prozessschrittes Reflexion wurde aus den oben dargelegten Ergebnissen deutlich, dass die Lehrkräfte die Ergebnisrückmeldungen dann als besonders hilfreiche Unterstützung ihrer Diagnose- und Förderkompetenz empfanden, wenn sie gleichzeitig angaben, dass sie aus den Ergebnisrückmeldungen neue Informationen erhielten. Ebenso kann formuliert werden, dass Lehrkräfte, die einen Informationsgewinn aus den Ergebnisrückmeldungen zogen, gleichzeitig die Wirkung auf Lehrende hinsichtlich der Unterstützung der Diagnose- und Förderkompetenz als höher einschätzten. Lehrkräfte, die die Ergebnisrückmeldungen als Informationsgewinn wahrnahmen, fühlten sich durch die Wirkung des Testverfahrens *Online Diagnose* unterstützt. Das Testverfahren *Online Diagnose* bot laut den Angaben einiger Lehrkräfte durchaus neue Informationen, die ihnen ohne die Ergebnisrückmeldungen nicht zur Verfügung gestanden hätten.

Die Variable gezielte Förderung sowie die Dummy-Variablen Bearbeitung im Deutschunterricht, Bearbeitung im Förderunterricht und Einbindung der Fördermaterialien des Prozessschritts Aktion waren prädiktiv für die Wirkung auf Lehrende. Setzten die Lehrkräfte Maßnahmen der gezielten Förderung häufiger ein als vor der Durchführung des Testverfahrens *Online Diagnose*, gaben sie gleichzeitig an, eine stärkere Wirkung auf Lehrende zu erleben. Ebenso kann formuliert werden, dass Lehrkräfte, die die Wirkung auf Lehrende stärker einschätzten, die Maßnahmen zur gezielten Förderung häufiger eingesetzt hatten. Die Maßnahmen der gezielten Förderung wurden von den Lehrkräften als Unterstützung ihrer Diagnose- und Förderkompetenz angesehen. Unklar bleibt im Rahmen dieser Ergebnisinterpretation, ob die Unterstützung durch die Quantität der Maßnahmen bzw. dadurch zustande kam, dass Lehrkräfte überhaupt Maßnahmen zur Förderung der Schülerinnen und Schüler erhielten, oder ob die Unterstützung in der Qualität der Fördermaterialien bestand. Wurden die Fördermaterialien im Deutschunterricht oder im Förderunterricht bearbeitet und wurden die Lösungen der Förderaufgaben im Unterricht besprochen, gaben die Lehrkräfte gleichzeitig eine stärkere Wirkung auf Lehrende an. Umgekehrt formuliert kann die Wirkung auf Lehrende hinsichtlich der Unterstützung ihrer Diagnose- und Förderkompetenz als stärker bezeichnet werden, wenn die Lehrkräfte angaben, die Fördermaterialien im Deutschunterricht oder im Förderunterricht eingesetzt und die Lösung der Fördermaterialien in den Unterricht eingebunden zu haben. Wendeten die Lehrkräfte die Fördermaterialien tatsächlich an und besprachen mit den Schülerinnen und Schülern die Lösungen, so kann angenommen werden, dass die Lehrkräfte durch das Testverfahren *Online Diagnose* unterstützt wurden und sich durch diese Unterstützung weniger beansprucht fühlten. Hierdurch bietet das Testverfahren *Online Diagnose* einen Vorteil gegenüber anderen Rückmeldungen aus Schulleistungsstudien. Die Übermittlung von entsprechenden Fördermaterialien zu den Ergebnisrückmeldungen ist bei den wenigstens Rückmeldeformaten vorgesehen.

Zusammenfassung. Aus den Ergebnisrückmeldungen wurde deutlich, dass in Zusammenhang mit der Wirkung auf Lehrende unter Einbezug aller unabhängiger Variablen vor allem der Bezug zum Bildungsplan der Grundschule, der Informationsgewinn aus Ergebnisrückmeldungen, die gezielte Förderung und die Dummy-Variablen Bearbeitung im Deutschunterricht, Bearbeitung im Förderunterricht und Einbindung der Fördermaterialien prädiktiv waren. Ähnlich wie bei den Ergebnissen zur Wirkung auf Lernende schienen in

Zusammenhang mit der abhängigen Variable Wirkung auf Lehrende vor allem die Fördermaßnahmen einen wichtigen Einfluss gehabt zu haben. Ebenfalls kann hier Rolff (2002) zitiert werden, der es als notwendig ansah, den Lehrkräften zu den Ergebnisrückmeldungen Handlungswissen an die Hand zu geben. Allerdings konnten die Fördermaterialien nur im weiteren Sinne unter dem Begriff Handlungswissen kategorisiert werden, da die Fördermaterialien kein direktes Handlungswissen bzw. keine Handlungsanleitungen für Lehrkräfte anboten, die umgesetzt werden konnten, ohne dass sich die Lehrkräfte inhaltlich genauer damit befassten. Ob die Lehrkräfte die Ergebnisrückmeldungen inhaltlich in Bezug zu den Fördermaterialien setzten, konnte im Rahmen dieser Untersuchung nicht überprüft werden. Die Lehrkräfte fühlten sich allerdings durch die Fördermaterialien in der Diagnostik von Schülerleistungen unterstützt und entlastet.

Überraschend war, dass keine der beiden Variablen zum Prozessschritt Rezeption signifikant mit der abhängigen Variable Wirkung auf Lehrende assoziiert war. Von den beiden Variablen wurde ein Zusammenhang mit der abhängigen Variable Wirkung auf Lehrende mit Bezug auf die Unterstützung der Diagnose- und Förderkompetenz erwartet. Auch Maier (2008a) konnte in seiner Untersuchung keinen Einfluss der Lehrerselbstwirksamkeit auf das Zyklenmodell der Rückmeldung nachweisen. Bonsen, Büchter und Peek (2006) zeigen in ihrer Untersuchung ebenfalls keinen direkten Effekt der Intensität der Auseinandersetzung mit Rückmeldungen und der Bewertung der Verständlichkeit von Rückmeldungen auf die Bedeutung der Rückmeldungen für die Unterrichtsentwicklung. Die beiden Variablen beeinflussen hier allerdings die Bewertung der Nützlichkeit der Rückmeldung, welche einen Effekt auf die Bedeutung der Rückmeldung für die Unterrichtsentwicklung aufweist (Bonsen, Büchter & Peek, 2006). In weiteren Untersuchungen sollten für den Prozessschritt Rezeption und auch für den Prozessschritt Reflexion zusätzliche Aspekte überprüft werden, um weitere Einflussfaktoren abzudecken. Außerdem sollte weiterhin geprüft werden, ob die Prozessschritte Rezeption und Reflexion die Art der Anwendung von Fördermaterialien moderieren. Eine Moderation könnte der Abfolge der Prozessschritte eher gerecht werden, wie Koch (2011) in ihrer Untersuchung zu indirekten Effekten auf den Prozessschritt Aktion beschreibt.

Insgesamt konnten alle unabhängigen Variablen, die einen signifikanten Zusammenhang mit der abhängigen Variable Wirkung auf Lehrende aufwiesen, als für die Lehrkraft unterstützend bezeichnet werden. Der Bezug zum Bildungsplan enthielt eine unterstützende Funktion in dem Sinne, dass den Lehrkräften die Auseinandersetzung mit dem Bildungsplan

der Grundschule erleichtert wurde. Der Erhalt neuer Informationen über die Leistung der Schülerinnen und Schüler konnte ebenso als unterstützend angesehen werden. Und auch der Einsatz der durch das Testverfahren *Online Diagnose* bereitgestellten Fördermaterialien erschien als unterstützend. Daraus konnte folgende Interpretation abgeleitet werden: Lehrkräfte, die Ergebnisrückmeldungen als Unterstützung wahrnahmen bzw. durch diese unterstützt wurden, schätzten die Wirkung auf Lehrende als höher ein. Dieser Zusammenhang kann im Rahmen der vorliegenden Ergebnisse jedoch nicht kausal interpretiert werden. Vielmehr decken die Erkenntnisse aus der vorliegenden Untersuchung weitere Fragestellungen auf. In Folgestudien zur Nutzung von Ergebnisrückmeldungen sollte die empfundene Unterstützung hinsichtlich der Materialien eines Testverfahrens z.B. im Rahmen einer Intervention kontrolliert untersucht werden.

7 Teilstudie 3: Einflussfaktoren auf die Prozessschritte der Nutzung von Ergebnisrückmeldungen

Die vorliegende Untersuchung liefert einen Forschungsbeitrag zu den Einflussfaktoren auf die Nutzung von Ergebnisrückmeldungen. Unterschiedlichste Einflussfaktoren müssen bei der Nutzung von Ergebnisrückmeldungen zur Erreichung des übergeordneten Ziels der Verbesserung von Lehre und Lernen berücksichtigt werden. Neben anderen können personelle und kontextuelle Einflussfaktoren bei der Nutzung von Ergebnisrückmeldungen hinderlich oder förderlich wirken (Helmke & Hosenfeld, 2005). Die personellen und kontextuellen Einflussfaktoren auf die Nutzung von Ergebnisrückmeldungen orientierten sich im Rahmen dieser Studie an den im Zyklenmodell der Rückmeldung vorgesehenen individuellen, schulischen und externen Bedingungen (z.B. Helmke & Hosenfeld, 2005; Hosenfeld & Groß Ophoff, 2007). Für die vorliegende Untersuchung wurden Bedingungen der Klasse ergänzt. Die Nutzung der Ergebnisrückmeldungen wurde anhand der vier Prozessschritte Rezeption, Reflexion, Aktion und Evaluation strukturiert. Diese erfassten z.B. die Selbstwirksamkeit der Lehrkräfte im Umgang mit den Ergebnisrückmeldungen, den Informationsgewinn aus den Ergebnisrückmeldungen, den Einsatz von Maßnahmen und die Wirkung des Testverfahrens *Online Diagnose* auf Lernende, das Kollegium und Lehrende.⁴⁴

Die folgende Fragestellung soll mit dieser Teilstudie beantwortet werden: Welche Bedingungen sind mit den Prozessschritten der Nutzung aus dem Testverfahren *Online Diagnose* assoziiert? Zur Beantwortung der Fragestellung wurden zwei Schritte unternommen:

1. Schritt 1 bestand in einer Überprüfung der Annahme, dass die jeweiligen Variablen der individuellen Bedingungen oder der Bedingungen der Klasse oder der schulischen und externen Bedingungen mit den vier Prozessschritten Rezeption, Reflexion, Aktion und Evaluation zusammenhängen.
2. Schritt 2 beinhaltet weiterhin die Überprüfung der Annahme, dass die in Schritt 1 abgebildeten Zusammenhänge bestehen bleiben, wenn alle Bedingungen gleichzeitig in das jeweilige Modell zu den Prozessschritten Rezeption, Reflexion, Aktion und Evaluation aufgenommen werden.

⁴⁴ Zusammenhänge der Prozessschritte der Nutzung von Ergebnisrückmeldungen mit der Wirkung auf Lernende, auf das Kollegium und Lehrende können in Teilstudie 2 in Kapitel 6 nachgelesen werden.

In den nachfolgenden Abschnitten werden zunächst die Methode und die eingesetzten Instrumente berichtet. Im Anschluss daran werden die Ergebnisse anhand von bivariaten Korrelationen und linearen multiplen Regressionsanalysen berechnet, dargestellt und abschließend diskutiert.

7.1 Methode und Instrumente

Die Daten zur vorliegenden Teilstudie beziehen sich auf die Befragung zum internetgestützten Diagnose- und Förderinstrument *Online Diagnose*. Befragt wurden Deutschlehrkräfte an Haupt-, Werkreal- und Gemeinschaftsschulen der fünften Jahrgangstufe. Nähere Informationen zum Testverfahren *Online Diagnose* können in Kapitel 3 und zur Stichprobe in Kapitel 4 nachgelesen werden.

Prozessschritte Rezeption, Reflexion, Aktion und Evaluation. Die Variablen zu den Prozessschritten Rezeption, Reflexion, Aktion und Evaluation wurden in Teilstudie 3 als abhängige Variablen eingesetzt. Die jeweiligen Items wurden im Rahmen einer Lehrkraftbefragung zum Testverfahren *Online Diagnose* entwickelt. Faktoren- und Reliabilitätsanalysen zur Skalenbildung sind in Kapitel 5 dargestellt. In der nachfolgenden Tabelle 29 werden im Sinne eines kurzen Überblicks Itembeispiele angeführt. Die Prozessschritte Rezeption, Reflexion und Evaluation wurden mit einer vierstufigen Likert-Skala gemessen (1 = *trifft nicht zu*, 2 = *trifft eher nicht zu*, 3 = *trifft eher zu*, 4 = *trifft zu*). Der Prozessschritt Aktion, wurde mit einer fünfstufigen Likert-Skala gemessen (1 = *viel weniger als vorher*, 2 = *weniger als vorher*, 3 = *genauso viel wie vorher*, 4 = *mehr als vorher*, 5 = *viel mehr als vorher*).

Tabelle 29

Itembeispiele und Cronbachs α zu den Prozessschritten Rezeption, Reflexion, Aktion und Evaluation

Variablen	α	Itemanzahl	Itembeispiele
<i>Prozessschritt Rezeption:</i>			
Selbstwirksamkeit im Umgang mit Ergebnismeldungen	.74	4	„Es fällt mir leicht, die Ergebnisse des Haupttests zu interpretieren“
Zeitaufwand für die Nutzung der Ergebnisse	.69	4	„Es ist mir zeitlich nicht möglich, die Ergebnismeldungen von allen Schülerinnen und Schülern ausführlich zu lesen“

(Fortsetzung)

Variablen	α	Item- anzahl	Itembeispiele
<i>Prozessschritt Reflexion:</i>			
Informationsgewinn aus Ergebnisrückmeldungen	.75	3	„Ohne die Ergebnisrückmeldungen des Haupttests würde ich auf manche Probleme bzw. Schwächen von Schülerinnen und Schülern nicht stoßen“
<i>Prozessschritt Aktion:</i>			
Gezielte Förderung	.63	3	„Welche Auswirkungen hatte die <i>Online Diagnose?</i> – Gezielte individuelle Förderung einzelner Schülerinnen und Schüler“
<i>Prozessschritt Evaluation:</i>			
Wirkung auf Lernende	.86	6	„Die förderbedürftigen Schülerinnen und Schüler erzielen nun bessere Noten“
Wirkung auf das Kollegium	.76	3	„Im Kollegium besprechen wir nun häufiger die Leistungen einzelner Schülerinnen und Schüler“
Wirkung auf Lehrende	.69	4	„Ich konnte meine Diagnosekompetenz steigern“

Anmerkung. Cronbachs α .

Individuelle Bedingungen. Die individuellen Bedingungen beziehen sich auf persönliche Einstellungen und Kompetenzen von Lehrkräften. Dazu wurden im Rahmen dieser Studie Variablen zur Einstellung zu Testuntersuchungen, computerbezogene Einstellung, Computerängstlichkeit, Diagnosekompetenz, Enthusiasmus für das Unterrichten und emotionale Belastung untersucht. Im Folgenden wird die Bedeutung der Variablen für die vorliegende Studie und ihre Herkunft genauer erläutert. Die Itembeispiele sind in Tabelle 30 aufgeführt.

Einstellung zu Testuntersuchungen. Die von den Lehrkräften bewertete Einstellung zu externen Testuntersuchungen hat einen Einfluss auf den Umgang mit Informationen aus Schulleistungsstudien (z.B. Kohler, 2005; Schrader & Helmke, 2004). So kann sich die allgemeine Einstellung gegenüber Untersuchungsergebnissen auf die Individual- und Förderdiagnostik auswirken, weswegen sie in der vorliegenden Untersuchung erhoben wurde. Die verwendeten Items sind dem DFG-Projekt QuaSSU⁴⁵ (Ditton & Merz, 2000) entnommen. Auch in der Studie QuaSUM⁴⁶ (Peek, 2004) wurden die Items bereits erfolgreich eingesetzt.

⁴⁵ Qualitätssicherung in Schule und Unterricht (QuaSSU)

⁴⁶ Qualitätsuntersuchungen zum Unterricht in Mathematik (QuaSUM)

Computerbezogene Einstellung. Die kognitionsbasierte Einstellung im Umgang mit Computern wurde in dieser Untersuchung als computerbezogene Einstellung bezeichnet. Die Items erfassten die Einstellung der Lehrkräfte in Bezug auf die Nutzung von Computern in der Schule. Lehrkräfte mit einer positiven computerbezogenen Einstellung könnten auch dem internetgestützten Diagnose- und Förderinstrument *Online Diagnose* positiv gegenüberstehen. Die Items sind der revidierten Fassung des Inventars zur Computerbildung entnommen (INCOBI-R; Richter, Naumann & Horz, 2010).

Computerängstlichkeit. Die affektbasierte Einstellung von Lehrkräften im Umgang mit Computern wurde über die Computerängstlichkeit erfasst. Stehen die Lehrkräfte Computern ängstlich gegenüber, wird ein vermeidendes Verhalten gegenüber Computern angenommen. Die Computerängstlichkeit kann dadurch ein Einflussfaktor in Bezug auf die Nutzung der internetbasiert zurückgemeldeten Ergebnisrückmeldungen aus dem Testverfahren *Online Diagnose* sein. Die fünf Items sind dem INCOBI-R⁴⁷ entnommen (Richter, Naumann & Horz, 2010).

Diagnosekompetenz. Der Begriff Diagnosekompetenz wurde im Rahmen der Untersuchung auf die Beurteilung allgemeiner Verständnisfragen beschränkt. Das heißt, von einer diagnostischen Einschätzung der Unterrichtsleistung von Schülerinnen und Schülern hinsichtlich des Fachwissens der Lehrkraft wurde abgesehen (Baumert & Kunter, 2011). Im Rahmen der vorliegenden Studie war die Variable relevant, um einen möglichen Zusammenhang der persönlich eingeschätzten Diagnosekompetenz mit der Nutzung von Ergebnisrückmeldungen aus einem Diagnose- und Förderinstrument feststellen zu können. Die Items sind der Studie COACTIV⁴⁸ entnommen (Baumert et al., 2009).

Enthusiasmus für das Unterrichten. Der Enthusiasmus für das Unterrichten erwies sich in der Studie COACTIV als eine wichtige Eigenschaft von Lehrkräften (Baumert & Kunter, 2011). Positive Zusammenhänge zwischen dem Enthusiasmus für das Unterrichten mit Merkmalen der Unterrichtsqualität sowie der Leistungs- und Motivationsentwicklung von Schülerinnen und Schülern wurden aufgedeckt (Baumert & Kunter, 2011). Der Enthu-

⁴⁷ Revidierte Fassung des Inventars zur Computerbildung (INCOBI-R)

⁴⁸ Professionswissen von Lehrkräften, kognitiv aktivierende Mathematikunterricht und die Entwicklung von mathematischer Kompetenz (COAKTIV)

siasmus für das Unterrichten wurde im Rahmen dieser Studie als ein Indikator für das Engagement der Lehrkräfte hinsichtlich der Nutzung von Ergebnisrückmeldungen angenommen. Die Items sind der COACTIV-Studie entnommen (Baumert et al., 2009).

Emotionale Belastung. Beanspruchungen der Lehrkräfte können z. B. über Erschöpfungszustände, Lustlosigkeit oder Überlastung in der Schule beschrieben werden. Andauernde Belastungen im Lehrerberuf können, insbesondere bei Abwesenheit von Ressourcen, zu Beanspruchungen führen (z. B. Rothland, 2012; Cramer & Friedrich, under revision). Zu den Indikatoren einer geringen emotionalen Belastung gehört z. B., dass sich Lehrkräfte von Arbeitsbelangen distanzieren können, wenn es für den Erhalt ihrer Ressourcen notwendig ist, und dass sie auch mit Misserfolg umgehen können (Klusmann, Kunter, Trautwein & Baumert, 2011). Hinsichtlich der vorliegenden Studie wurde davon ausgegangen, dass Lehrkräfte, die sich lediglich in geringem Maße emotional belastet fühlen, mehr Ressourcen für eine Nutzung der Ergebnisrückmeldungen frei haben. Die Items sind dem COACTIV Fragebogen entnommen (Baumert et al., 2009).

Die nachfolgende Tabelle 30 stellt die deskriptiven Statistiken der jeweiligen Variablen dar. Für die Skalenbildung wurden die negativen Items rekodiert. Dies betraf insgesamt drei Items. Eine linkssteile Verteilung wies die Variable Computerängstlichkeit auf. Das bedeutet, dass sich die Mehrheit der Lehrkräfte als nicht computerängstlich einschätzte. Bei den weiteren Variablen tendierten die Lehrkräfte zu einer eher zustimmenden Antwort. Die Mittelwerte lagen zwischen $M = 1.65$ und $M = 3.64$. Dabei hatte die Variable Computerängstlichkeit den niedrigsten und die Variable Enthusiasmus für das Unterrichten den höchsten Mittelwert. Die Standardabweichungen reichten von $SD = 0.42$ bis $SD = 0.62$. Die interne Konsistenz (α) wies insgesamt gute bis akzeptable Werte auf. Den geringsten α -Wert besaß die Variable Einstellung zu Testuntersuchungen ($\alpha = .69$), den höchsten Wert die Variable Enthusiasmus für das Unterrichten ($\alpha = .90$).

Tabelle 30

Deskriptive Statistik sowie interne Konsistenz der Variablen zu den individuellen Bedingungen

Variablen	Itembeispiele	Item- anzahl	<i>M</i>	<i>SD</i>	Modus	Fehlende Angaben (Prozent)	α
Einstellung zu Testuntersuchungen	Zentrale (landesweite) Tests sind für die Arbeit der Schulen sehr wichtig.	3	2.67	0.58	3	5.3	.69
Computerbezogene Einstellung	Lernen mit dem Computer ermöglicht in hohem Maße selbstbestimmtes und entdeckendes Lernen.	5	3.07	0.55	3	4.7	.81
Computerängstlichkeit	Wenn möglich, vermeide ich das Arbeiten am Computer.	5	1.65	0.62	1	4.0	.85
Diagnosekompetenz	Es fällt mir leicht, festzustellen, ob eine Schülerin oder ein Schüler eine Aufgabe verstanden hat.	4	3.15	0.42	3	3.3	.75
Enthusiasmus für das Unterrichten	Mir macht das Unterrichten großen Spaß.	4	3.64	0.46	4	5.3	.90
Emotionale Belastung	Ich fühle mich in der Schule oft erschöpft.	4	2.01	0.59	2	7.3	.79

Anmerkung. $N = 150$; Skalierung: 1 (trifft nicht zu) bis 4 (trifft zu) oder 1 (stimme nicht zu) bis 4 (stimme zu); Für die Skalenbildung wurden negative Items rekodiert, dies betraf drei Items.

Bedingungen der Klasse. Für die vorliegende Studie wurden die individuellen, die schulischen und die externen Bedingungen des Zyklenmodells der Rückmeldung um die Bedingungen der Klasse ergänzt bzw. ihnen neu zugeordnet. Grund für die Ergänzung war die Annahme, dass Kriterien der Klasse wie z.B. die Zusammensetzung der Klassen einen Einfluss auf die Nutzung der Ergebnisrückmeldungen haben könnten. Als Bedingungen der Klasse wurden Kriterien erfasst, die einen Überblick über die Zusammensetzung der Klasse ermöglichten. Hierzu wurden die folgenden Variablen verwendet: Unterrichtsstörungen, Klassengröße, Klassenanteil der Kinder aus Familien mit Migrationshintergrund und Klassenanteil der Kinder aus engagierten, sich kümmernden Familien. Die Verteilungen der jeweiligen Variablen werden in Tabelle 31 dargestellt. Im Folgenden werden die vier Variablen, ihre Bedeutung für die vorliegende Studie und ihre Herkunft genauer erläutert.

Unterrichtsstörungen. Inhaltlich erfassten die Items Kriterien, die Störungen im Unterricht oder den Zeitaufwand für Themen, die nicht unterrichtsrelevant sind, charakterisieren. Ein Ziel für den Unterricht ist es, die Unterrichtszeit als Lernzeit optimal zu nutzen und Zeitverluste durch von den Lerninhalten ablenkende Handlungen zu vermeiden (z.B. Voss & Kunter, 2011). Die Variable Unterrichtsstörungen war für die vorliegende Studie relevant, da je nach Zusammensetzung der Klasse eine gezielte individuelle Förderung der Schülerinnen und Schüler für die Lehrkraft als gewinnbringend oder, aufgrund der Störungen, als nicht nützlich erscheinen konnte. Für die Variable Unterrichtsstörungen wurden Items aus COACTIV verwendet (Baumert et al., 2009). Die Verteilung zeigte sich eher linkssteil, das bedeutet, die Unterrichtsstörungen traten laut den Angaben der Lehrkräfte nicht sehr häufig auf. Die interne Konsistenz der Skala Unterrichtsstörungen war hoch ($\alpha = .94$).

Klassengröße. Die Klassengröße wurde durch ein offenes Antwortformat erfasst. Die Spannweite reichte von vier bis 30 Schülerinnen und Schüler pro Klasse. Die Klassengröße war relevant für die vorliegende Untersuchung, da die individuelle Umsetzung der Ergebnisrückmeldungen je nach Größe der Klasse eine subjektive Beanspruchung darstellen konnte.

Klassenanteil der Kinder aus Familien mit Migrationshintergrund. Der Anteil der Familien mit Migrationshintergrund wurde durch ein offenes Antwortformat erfasst. Die Spannweite reichte von 0 bis 25 Schülerinnen und Schüler pro Klasse. 2.5 % der Lehrkräfte

berichteten, keine Kinder aus Familien mit Migrationshintergrund in ihrer Klasse zu haben. Der Anteil an Familien mit Migrationshintergrund war für die Studie deshalb von Bedeutung, weil das Testverfahren *Online Diagnose* für die Leistungsmessung von Schülerinnen und Schülern mit Migrationshintergrund aufgrund möglicher Sprach- oder Verständnisschwierigkeiten nicht geeignet ist.

Klassenanteil der Kinder aus engagierten, sich kümmernden Familien. Der Anteil engagierter und sich kümmernder Familien wurde durch ein offenes Antwortformat erfasst. Die Spannweite reichte von 0 bis 21 Familien pro Klasse. 1 % der Lehrkräfte berichtete, keine Kinder aus engagierten, sich kümmernden Familien in ihrer Klasse zu haben. Die Angaben zu den sich engagierenden und sich kümmernden Familien waren für die Studie relevant, weil die Lehrkräfte möglicherweise durch diese Familien bei der Förderung der Schülerinnen und Schüler und im Speziellen z.B. beim Einsatz der Fördermaterialien aus dem Testverfahren *Online Diagnose* unterstützt wurden.

Tabelle 31

Deskriptive Statistik (sowie interne Konsistenz) der Variablen zu den Bedingungen der Klasse

Variablen	Itembeispiele/Fragestellungen	Item- anzahl	<i>M</i>	<i>SD</i>	Modus	Fehlende Angaben (Prozent)	α
Unterrichtsstörungen ^a	Ich muss in dieser Klasse viel ermahnen, um für Ruhe zu sorgen.	4	2.33	0.84	2	3.3	.94
Klassengröße	Wie groß ist dies Klasse?		18.07	5.70	20	2.0	
Klassenanteil an Familien mit Migrationshintergrund	Wie hoch ist in etwa die Anzahl der Schülerinnen und Schüler in dieser Klasse aus der folgenden Gruppe?		7.78	5.86	2	19.3	
Klassenanteil engagierter, sich kümmernder Familien	Wie hoch ist in etwa die Anzahl der Schülerinnen und Schüler in dieser Klasse aus der folgenden Gruppe?		7.72	4.67	5	23.3	

Anmerkung. $N = 150$; ^aSkalierung: 1 (trifft nicht zu) bis 4 (trifft zu).

Schulische und externe Bedingungen. Die schulischen und externen Bedingungen bezeichnen die Einbettung von Testverfahren in das schulische Umfeld und die darin er-fahrene Unterstützung für die ausführenden Lehrkräfte. Zu den schulischen Bedingungen wurden gezählt: Allgemeine Unterstützung durch das Kollegium und die Schulleitung sowie die Akzeptanz der *Online Diagnose* durch das Kollegium und die Schulleitung. Auch die Verpflichtung der Lehrkraft zur Teilnahme am Testverfahren *Online Diagnose* durch die Schulleitung fiel unter die schulischen Bedingungen. Zu den externen Bedingungen zählten Merkmale und Einstellung der Elternschaft: Von der allgemeinen Unterstützung durch die Eltern und der Akzeptanz der *Online Diagnose* durch die Eltern wurde ein positiver Einfluss auf die Nutzung und Wirkung des Testverfahrens *Online Diagnose* erwartet. Im Folgenden wird die Bedeutung der Variablen für die vorliegende Studie und ihre Herkunft genauer er-läutert. Die schulischen und externen Bedingungen werden in den nächsten Abschnitten ge-meinsam berichtet. Die jeweiligen Itembeispiele sind in Tabelle 32 zusammengefasst.

Allgemeine Unterstützung durch das Kollegium, die Schulleitung und die Eltern. Die Lehrkräfte gaben zur allgemeinen Unterstützung durch das Kollegium, die Schulleitung und die Eltern unter anderem an, ob Kooperationen stattfanden und ob sie sich durch diese unterstützt fühlten. Für die vorliegende Untersuchung waren die Variablen wichtig, da ein unterstützendes Umfeld möglicherweise bei der Interpretation der Ergebnisrückmeldungen weiterhelfen konnte und damit die Nutzung von Ergebnisrückmeldungen förderte. Die Items zur Unterstützung durch die Schulleitung wurden der Studie PISA entnommen (Ramm et al., 2003). Alle Items wurden in der Studie TRAIN eingesetzt (Jonkmann, Rose & Trautwein, 2013).

Akzeptanz der Online Diagnose durch das Kollegium, die Schulleitung und die Eltern. Neben der persönlichen Einstellung von Lehrkräften zu externen Testverfahren (in-dividuelle Bedingungen), wurde ebenfalls von einer entsprechenden Einstellung gegenüber Testverfahren des Kollegiums, der Schulleitung und der Eltern ein Einfluss erwartet (schu-lische und externe Bedingungen). Koch (2011) konnte in ihren Untersuchungen einen Zu-sammenhang der Akzeptanz eines Testverfahrens durch das Kollegium mit der Nutzung von Ergebnisrückmeldungen aufdecken. Im Rahmen dieser Studie wurden die Lehrkräfte um eine Einschätzung der Akzeptanz des Testverfahrens *Online Diagnose* ihres Kollegiums, der

Schulleitung und der Eltern gebeten. Es wurde angenommen, dass die Akzeptanz des Testverfahrens *Online Diagnose* im schulischen Umfeld die Nutzung der daraus resultierenden Ergebnisrückmeldungen beeinflusst. Die Items wurden für die vorliegende Studie entwickelt.

Verpflichtung durch die Schulleitung. Die Variable Verpflichtung durch die Schulleitung beinhaltete Informationen darüber, ob die Lehrkräfte die Durchführung des Testverfahrens *Online Diagnose* als Verpflichtung wahrnahmen. Da das Testverfahren *Online Diagnose* flächendeckend in Baden-Württemberg und zudem kostenlos angeboten wurde, wurden die Lehrkräfte gezielt hierzu befragt. Eine mögliche Verpflichtung der Lehrkräfte zur Durchführung des Testverfahrens *Online Diagnose* könnte zu einem geringen Engagement der Lehrkräfte und damit zu einem minimalistischen Aufwand für den Einsatz des Testverfahrens geführt haben. Die Items wurden für diese Befragung neu entwickelt.

Die nachfolgende Tabelle 32 zeigt die deskriptive Statistik der Variablen zu den schulischen und externen Bedingungen. Für die Skalenbildung wurden die negativen Items rekodiert. Dies betraf insgesamt zwei Items. Die Werte der Variablen Akzeptanz der *Online Diagnose* durch das Kollegium und Akzeptanz der *Online Diagnose* durch die Eltern waren eher linkssteil verteilt. Das bedeutet, das Testverfahren *Online Diagnose* wurde eher nicht akzeptiert. Die Mittelwerte streuten von einer eher geringen Akzeptanz der *Online Diagnose* durch die Eltern ($M = 2.04$) bis zu einer hohen Akzeptanz der *Online Diagnose* durch die Schulleitung ($M = 3.36$). Die Standardabweichungen lagen zwischen $SD = 0.64$ und $SD = 0.93$. Die interne Konsistenz war ausreichend mit Ausnahme der Variable Akzeptanz der *Online Diagnose* durch die Schulleitung, die mit $\alpha = .52$ in einem eher niedrigen Bereich lag. Die geringe interne Konsistenz war aufgrund der wenigen Items dennoch akzeptabel (Lienert, 1969). Den höchsten α -Wert besaß die Variable allgemeine Unterstützung durch das Kollegium ($\alpha = .87$).

Tabelle 32

Deskriptive Statistik sowie interne Konsistenz der Variablen zu den schulischen und externen Bedingungen

Variablen	Itembeispiele	Item- anzahl	<i>M</i>	<i>SD</i>	Modus	Fehlende Angaben (Prozent)	α
Allgemeine Unterstützung Kollegium	Die anderen Lehrkräfte sind mir eine große Hilfe, wenn es um Unterrichtsmethoden und Unterrichtsinhalte geht.	4	3.09	0.71	4	2.7	.87
Allgemeine Unterstützung Schulleitung	Die Schulleitung ist zu Fragen des Unterrichts ansprechbar und kompetent.	3	2.80	0.71	3	2.7	.81
Akzeptanz der OD Kollegium	Die <i>Online Diagnose</i> ist im Kollegium ein akzeptiertes Hilfsmittel, um den Unterricht zu gestalten.	3	2.29	0.65	2	4.0	.76
Akzeptanz der OD Schulleitung	Die Schulleitung findet die <i>Online Diagnose</i> wichtig.	2	3.36	0.64	4	3.3	.52
Verpflichtung durch die Schulleitung	Die Schulleitung verpflichtet die Lehrkräfte zur Durchführung der Fördermaterialien.	2	2.20	0.93	2	5.3	.67
Allgemeine Unterstützung Eltern	Die Eltern unterstützen meine Arbeit.	4	2.85	0.65	3	2.7	.83
Akzeptanz der OD Eltern	Die Eltern finden die <i>Online Diagnose</i> zur individuellen Förderung der Schülerinnen und Schüler wichtig.	3	2.04	0.72	1 ^a	4.7	.77

Anmerkung. $N = 150$; OD = *Online Diagnose*; Skalierung: 1 (trifft nicht zu) bis 4 (trifft zu); ^a Es sind mehrere Modi vorhanden, der kleinste Wert wurde angegeben; Für die Skalenbildung wurden negative Items rekodiert, dies betraf zwei Items.

Kovariaten. Die Variablen Geschlecht und Alter der Lehrkräfte wurden als Kovariate in die Analysen einbezogen, um weitere Merkmale der jeweiligen Lehrkraft zu berücksichtigen. In der Stichprobe antworteten 82 % Lehrerinnen und 18 % Lehrer. Die Lehrkräfte waren im Durchschnitt 40 Jahre alt. Dabei zeigte sich, dass die männlichen Lehrkräfte älter waren als die weiblichen Lehrkräfte.⁴⁹

7.2 Statistische Analysen

Zur Prüfung der Fragestellungen wurden zunächst bivariate Zusammenhänge der abhängigen und unabhängigen Variablen betrachtet. Im Anschluss daran wurden multiple lineare Regressionsanalysen durchgeführt. Alle Analysen wurden mit dem Programm Mplus Version 7.1 (Muthén & Muthén, 1998-2013) berechnet. Das Programm bot den Vorteil eines robusteren MLR-Schätzers (robust maximum likelihood estimator), der für lineare Regressionsanalysen, deren Residuen nicht normalverteilt sind, geeignet ist. Die Normalverteilung der Residuen (z.B. Eid, Gollwitzer & Schmitt, 2011) wurde in Vorabanalysen mit SPSS 22 durch den Kolmogorov-Smirnoff-Test überprüft. Dabei wurde festgestellt, dass die Residuen nicht normalverteilt waren. Weiterhin bestand durch das Programm Mplus die Möglichkeit, das Full Information Maximum Likelihood Verfahren (FIML) zu verwenden. Es stellt eine der momentan am besten geeigneten Methoden zum Umgang mit fehlenden Werten dar (z.B. Lüdtke, Robitzsch, Trautwein & Köller, 2007; Enders, 2010). Das Verfahren wurde in allen Analysen angewandt. Insgesamt lagen die fehlenden Werte bei den verwendeten Variablen zwischen 1.3 % und 7.3 %. Zu den nachfolgenden Regressionen wurden die standardisierten Regressionskoeffizienten β (z.B. Cohen et al., 2003), die Standardfehler SE und der Determinationskoeffizient R^2 berichtet (z.B. Steyer & Eid, 2001). Bei allen Regressionskoeffizienten wurden die Standardized Model Results (STDYX Standardization) berichtet.

In allen Modellen wurden die Prozessschritte Rezeption, Reflexion, Aktion und Evaluation als abhängige Variablen eingesetzt. Die unabhängigen Variablen wurden durch die individuellen Bedingungen, die Bedingungen der Klasse sowie die schulischen und externen Bedingungen abgebildet. In allen multiplen linearen Regressionsanalysen wurde für das Geschlecht und das Alter der Lehrkräfte kontrolliert. Nachfolgend werden die Ergebnisse wie

⁴⁹ Weitere Informationen zur Stichprobe sind in Kapitel 4 nachzulesen.

folgt berichtet: Zunächst werden die bivariaten Zusammenhänge aufgeführt. Im Anschluss werden die Ergebnisse der jeweils getrennt für jeden Prozessschritt und jede Bedingung berechneten Regressionsanalysen aufgeführt (Schritt 1). Weiterhin werden die Ergebnisse zu den Regressionsanalysen berichtet, in die alle unabhängigen Variablen gleichzeitig aufgenommen wurden (Schritt 2). Zusätzlich werden die Ergebnisse für die Regressionsanalysen aufgezeigt, in die lediglich die signifikanten unabhängigen Variablen der Bedingungen aufgenommen wurden (Schritt 2).

7.3 Ergebnisse

Bivariate Zusammenhänge. Zunächst werden die bivariaten Zusammenhänge der abhängigen und unabhängigen Variablen dargestellt. Die signifikanten Interkorrelationen der abhängigen Variablen reichten von $r = -.18$ ($p < .05$) bis $r = .44$ ($p < .001$).⁵⁰ Die signifikanten Interkorrelationen der unabhängigen Variablen lagen zwischen $r = -.57$ ($p < .001$) und $r = .60$ ($p < .001$). Das Problem der Multikollinearität liegt laut Marsh, Downson, Pietsch und Walker (2004) erst bei einer Interkorrelation ab $r = .93$ vor. Aufgrund der vorliegenden Werte konnte deshalb eine Multikollinearität der Variablen ausgeschlossen werden.

Die in Tabelle 33 dargestellten signifikanten, bivariaten Zusammenhänge der individuellen Bedingungen mit den Prozessschritten Rezeption, Reflexion, Aktion und Evaluation reichten von $r = -.30$ ($p < .001$) bis $r = .35$ ($p < .001$). Die stärkste signifikant negative Korrelation bestand zwischen der abhängigen Variable Informationsgewinn aus Ergebnismeldungen und der unabhängigen Variable Diagnosekompetenz. Die stärkste signifikant positive Korrelation bestand zwischen der abhängigen Variable Wirkung auf Lehrende und der unabhängigen Variable Einstellung zu Testuntersuchungen. Zusammenfassend lässt sich feststellen, dass die abhängige Variable Selbstwirksamkeit im Umgang mit den Ergebnismeldungen die häufigsten signifikanten Korrelationen mit den unabhängigen Variablen aufwies. Außerdem kann festgehalten werden, dass die unabhängige Variable Einstellung zu Testuntersuchungen beinahe mit allen abhängigen Variablen signifikant korrelierte.

Die signifikanten, bivariaten Zusammenhänge der Bedingungen der Klasse mit den Prozessschritten Rezeption, Reflexion, Aktion und Evaluation lagen zwischen $r = -.17$

⁵⁰ Weitere Informationen zu den bivariaten Korrelationen der Prozessschritte Rezeption, Reflexion, Aktion und Evaluation können in Teilstudie 2 in Kapitel 6 nachgelesen werden.

($p < .05$) und $r = .20$ ($p < .05$). Die stärkste signifikante negative Korrelation bestand zwischen der abhängigen Variable gezielte Förderung und der unabhängigen Variable Klassenanteil an Kindern aus Familien mit Migrationshintergrund. Die stärkste signifikante positive Korrelation bestand zwischen der abhängigen Variable gezielte Förderung und der unabhängigen Variable Klassenanteil an Kindern mit engagierten, sich kümmernden Familien. Zusammenfassend lässt sich feststellen, dass die beiden Korrelationen die einzigen signifikanten bivariaten Zusammenhänge mit den abhängigen Variablen darstellten.

Die signifikanten, bivariaten Zusammenhänge der schulischen und externen Bedingungen mit den Prozessschritten Rezeption, Reflexion, Aktion und Evaluation lagen zwischen $r = .16$ ($p < .05$) und $r = .52$ ($p < .001$). Die schwächste signifikante positive Korrelation bestand zwischen der abhängigen Variable gezielte Förderung und der unabhängigen Variable Akzeptanz der *Online Diagnose* durch die Eltern. Die stärkste signifikant positive Korrelation bestand zwischen der abhängigen Variable Wirkung auf das Kollegium und der unabhängigen Variable Akzeptanz der *Online Diagnose* durch das Kollegium. Zusammenfassend lässt sich feststellen, dass vor allem die drei abhängigen Variablen Wirkung auf Lernende, Wirkung auf das Kollegium und Wirkung auf Lehrende häufiger als die anderen abhängigen Variablen mit den unabhängigen Variablen zu den schulischen und externen Bedingungen signifikant korrelierten. Die abhängige Variable Zeitaufwand für die Nutzung der Ergebnisse hing mit keiner der unabhängigen Variablen signifikant zusammen. Außerdem hing die unabhängige Variable allgemeine Unterstützung durch die Eltern mit keiner abhängigen Variable signifikant zusammen.

Tabelle 33

Bivariate Korrelationen der Variablen der Prozessschritte Rezeption, Reflexion, Aktion und Evaluation mit den individuellen Bedingungen, Bedingungen der Klasse sowie schulischen und externen Bedingungen

	1	2	3	4	5	6	7
Rezeption:							
1 Selbstwirksamkeit im Umgang mit Ergebnismeldungen							
2 Zeitaufwand für die Nutzung der Ergebnisse	-.18*						
Reflexion:							
3 Informationsgewinn aus Ergebnismeldungen	.09	-.05					
Aktion:							
4 Gezielte Förderung	.29**	.07	.36**				
Evaluation:							
5 Wirkung auf Lernende	.24*	.16	.27**	.36**			
6 Wirkung auf das Kollegium	.02	.09	.15	.15	.27**		
7 Wirkung auf Lehrende	.33**	-.10	.38**	.37**	.44**	.33**	
Kovariaten:							
8 Geschlecht (0 = männlich)	.10	.03	-.15	.11	-.12	-.02	.02
9 Alter	-.05	-.06	-.14	-.17*	-.06	.15	-.01
Individuelle Bedingungen:							
10 Einstellung zu Testuntersuchungen	.28**	.07	.29**	.16*	.15	.18*	.35**
11 Computerbezogene Einstellungen	.23*	.00	.20*	.15	.17	.03	.19
12 Computerängstlichkeit	-.26**	.01	-.09	-.05	.04	.18*	-.09
13 Diagnosekompetenz	.09	.02	-.30**	.02	-.02	-.11	-.07
14 Enthusiasmus für das Unterrichten	.18*	.18*	.08	.14	.15*	.07	.14
15 Emotionale Belastung	-.20*	-.03	-.09	-.00	-.14	-.04	-.04
Bedingungen der Klasse:							
16 Unterrichtstörungen	-.08	.05	-.03	.02	-.09	.13	.07
17 Klassengröße	-.06	.12	-.12	-.04	-.06	-.01	-.14
18 Klassenanteil an Kindern aus Familien mit Migrationshintergrund	.02	-.01	-.13	-.17*	-.14	-.07	-.05
19 Klassenanteil an Kindern mit engagierten, sich kümmernden Familien	.08	.07	.08	.20*	.12	.06	-.06
Schulische und externe Bedingungen:							
20 Allgemeine Unterstützung Kollegium	.06	-.04	.03	.07	.08	.28**	.22**
21 Allgemeine Unterstützung Schulleitung	.26**	-.01	-.00	.04	.15	.09	.25**
22 Akzeptanz der OD Kollegium	.33**	-.07	.10	.14	.36**	.52**	.40**
23 Akzeptanz der OD Schulleitung	.15*	.06	.14	.11	.31**	.02	.27**
24 Verpflichtung durch die Schulleitung	-.01	.05	-.10	-.13	-.05	.28**	.07
25 Allgemeine Unterstützung Eltern	.13	-.11	.04	.13	.17	.09	.04
26 Akzeptanz der OD Eltern	.10	-.13	.20*	.16*	.40**	.29**	.33**

Anmerkung. $N = 150$; * $p < .05$; ** $p < .01$; OD = Online Diagnose.

Assoziationen der jeweiligen Prozessschritte mit den jeweiligen individuellen Bedingungen, Bedingungen der Klasse oder schulischen und externen Bedingungen (Schritt 1). In den Tabellen 34 bis 36 sind die multiplen linearen Regressionsanalysen zum ersten Schritt der Beantwortung der folgenden Fragestellung aufgeführt: Welche Bedingungen sind mit den Prozessschritten der Nutzung aus dem Testverfahren *Online Diagnose* assoziiert? In diesem ersten Schritt wurden die Assoziationen der jeweiligen Bedingungen mit den jeweiligen Prozessschritten in getrennten Modellen analysiert. Die Ergebnisse der Regressionsanalysen werden wie folgt berichtet: Zunächst die Modelle zu den individuellen Bedingungen (Tabelle 34), dann die Modelle zu den Bedingungen der Klasse (Tabelle 35) und als Drittes die Modelle zu den schulischen und externen Bedingungen (Tabelle 36). Für das Geschlecht und Alter der Lehrkräfte wurde jeweils kontrolliert.

Individuelle Bedingungen. Zu den individuellen Bedingungen (Tabelle 34), unter Kontrolle des Geschlechts und des Alters der Lehrkraft, wurden Assoziationen der unabhängigen Variable Einstellung zu Testuntersuchungen mit den folgenden abhängigen Variablen aufgedeckt: Die Lehrkräfte, die gegenüber Testuntersuchungen positiv eingestellt waren, berichteten gleichzeitig eine stärkere Wirkung des Testverfahrens *Online Diagnose* auf Lehrende ($\beta = .33, p < .001$), eine höhere Selbstwirksamkeit im Umgang mit den Ergebnissen ($\beta = .28; p < .001$), einen höheren Informationsgewinn aus den Ergebnissen ($\beta = .21; p < .01$) und einen verstärkten Einsatz von Maßnahmen der gezielten Förderung ($\beta = .21; p < .01$). Weiterhin berichteten Lehrkräfte, die sich eine etwas stärker ausgeprägte Computerängstlichkeit zuschrieben von einem geringeren selbstwirksamen Umgang mit Ergebnisrückmeldungen ($\beta = -.23; p < .01$). Umgekehrt berichteten die Lehrkräfte, die sich eine etwas stärker ausgeprägte Computerängstlichkeit zuschrieben, gleichzeitig von einer stärkeren Wirkung auf Lernende ($\beta = .23; p < .01$) und einer stärkeren Wirkung auf das Kollegium ($\beta = .22; p < .01$). Lehrkräfte, die sich selbst eine bessere Diagnosekompetenz zuschrieben, gaben an, weniger neue Informationen aus den Ergebnisrückmeldungen zu gewinnen ($\beta = -.21; p < .01$). Ferner berichteten Lehrkräfte, die sich als enthusiastisch hinsichtlich des Unterrichts beschrieben, von einem höheren Zeitaufwand für die Nutzung der Ergebnisse ($\beta = .23; p < .05$). Die Varianzaufklärung der Variablen zu den individuellen Bedingungen reichte von 4 % bis 19 %. Die häufigste und stärkste Vorhersagekraft der individuellen Bedingungen hatten die unabhängigen Variablen Einstellung zu Testuntersuchungen und Computerängstlichkeit.

Bedingungen der Klasse. Bezugnehmend auf die Bedingungen der Klasse (Tabelle 35) wurden, unter Kontrolle des Geschlechts und des Alters der Lehrkraft, die folgende signifikante Assoziation festgestellt: Lehrkräfte die einen hohen Klassenanteil an Kindern aus engagierten und sich kümmernden Familien hatten, gaben auch an, Maßnahmen zur gezielten Förderung häufiger einzusetzen ($\beta = .23; p < .05$). Die Varianzaufklärung im Modell der gezielten Förderung lag bei 10 %.

Schulische und externe Bedingungen. Zu den schulischen und externen Bedingungen (Tabelle 36) wurden, unter Kontrolle des Geschlechts und des Alters der Lehrkräfte, signifikante Assoziationen der unabhängigen Variable allgemeine Unterstützung durch das Kollegium mit den folgenden abhängigen Variablen erfasst: Lehrkräfte, die berichteten, durch das Kollegium unterstützt zu werden, gaben gleichzeitig an, weniger selbstwirksam mit den Ergebnisrückmeldungen umzugehen ($\beta = -.22; p < .01$) sowie eine stärkere Wirkung auf das Kollegium zu erleben ($\beta = .24; p < .001$). Außerdem berichteten Lehrkräfte, die sich von der Schulleitung im Allgemeinen unterstützt fühlten, gleichzeitig von einem selbstwirksameren Umgang mit den Ergebnisrückmeldungen ($\beta = .28; p < .001$) und einer stärkeren Wirkung auf Lehrende ($\beta = .18; p < .01$). Die unabhängige Variable Akzeptanz der *Online Diagnose* durch das Kollegium war mit mehreren abhängigen Variablen assoziiert: Lehrkräfte, die von der Akzeptanz der *Online Diagnose* durch das Kollegium berichteten, gaben gleichzeitig einen selbstwirksameren Umgang mit den Ergebnisrückmeldungen an ($\beta = .34; p < .001$) sowie eine stärkere Wirkung auf Lernende ($\beta = .31; p < .001$), das Kollegium ($\beta = .43; p < .001$) und Lehrende ($\beta = .30; p < .001$). Ebenso berichteten die Lehrkräfte, die eine stärkere Akzeptanz der *Online Diagnose* durch die Schulleitung erlebten, gleichzeitig von einer stärkeren Wirkung auf Lernende ($\beta = .19; p < .05$) und umgekehrt von einer geringeren Wirkung auf das Kollegium ($\beta = -.15; p < .05$). Wenn die Lehrkräfte außerdem von der Verpflichtung des Einsatzes z.B. der Fördermaterialien durch die Schulleitung berichteten, gaben sie gleichzeitig eine geringere Wirkung auf Lernende ($\beta = -.14; p < .05$) und umgekehrt eine stärkere Wirkung auf das Kollegium an ($\beta = .22; p < .001$). Berichteten die Lehrkräfte positiv von der allgemeinen Unterstützung der Eltern, gaben sie gleichzeitig eine geringere Wirkung auf Lehrende an ($\beta = -.21; p < .001$). Außerdem gaben Lehrkräfte, die von einer stärkeren Akzeptanz der *Online Diagnose* durch die Eltern berichteten, gleichzeitig einen höheren Informationsgewinn aus Ergebnisrückmeldungen ($\beta = .18; p < .05$), eine stärkere Wirkung auf Lernende ($\beta = .34; p < .001$), eine stärkere Wirkung auf das Kollegium

($\beta = .20$; $p < .05$) und eine stärkere Wirkung auf Lehrende an ($\beta = .32$; $p < .001$). Die unabhängigen Variablen zu den schulischen und externen Bedingungen klärten, im Vergleich zu den individuellen Bedingungen und den Bedingungen der Klasse, den größten Anteil an Varianz in den abhängigen Variablen auf. Die Varianzaufklärung lag zwischen 12 % und 41 %. Die häufigste und stärkste Vorhersagekraft der schulischen und externen Bedingungen hatten die unabhängigen Variablen Akzeptanz der *Online Diagnose* durch das Kollegium und Akzeptanz der *Online Diagnose* durch die Eltern.

Tabelle 34
 Regressionsanalysen zu den individuellen Bedingungen

	Rezeption			Reflexion			Aktion			Evaluation											
	Selbstwirksamkeit im Umgang mit Ergebnisrückmel- dungen			Zeitaufwand für die Nutzung der Ergebnisse			Informationsge- winn aus Ergebnis- rückmeldungen			Gezielte Förderung			Wirkung auf Lernende			Wirkung auf das Kollegium			Wirkung auf Lehrende		
	β	SE	p	β	SE	p	β	SE	p	β	SE	p	β	SE	p	β	SE	p	β	SE	p
Kovariaten:																					
Geschlecht (0 = männlich)	.09	0.08	.23	-.05	0.08	.57	-.19	0.08	.02	.05	0.09	.58	-.18	0.07	.01	-.05	0.09	.58	.02	0.08	.78
Alter	.07	0.09	.47	-.05	0.10	.61	-.11	0.10	.26	-.13	0.09	.15	-.10	0.09	.25	.12	0.09	.20	.06	0.09	.49
Individuelle Bedingungen:																					
Einstellung zu Testuntersuchungen	.28	0.09	.00	.07	0.09	.44	.21	0.08	.01	.21	0.07	.01	.13	0.09	.15	.12	0.09	.19	.33	0.09	.00
Computerbezogene Einstellung	.05	0.10	.60	-.05	0.11	.65	.06	0.10	.54	.09	0.09	.31	.16	0.10	.12	.09	0.11	.41	.07	0.12	.54
Computerängstlichkeit	-.23	0.09	.01	.04	0.12	.72	.04	0.10	.67	.06	0.10	.56	.23	0.09	.01	.22	0.09	.01	-.03	0.10	.73
Diagnosekompetenz	.16	0.10	.10	.03	0.09	.71	-.21	0.08	.01	.10	0.09	.26	.03	0.09	.71	-.12	0.09	.21	.01	0.09	.89
Enthusiasmus für das Unterrichten	.04	0.11	.73	.23	0.11	.03	.08	0.09	.40	.15	0.10	.13	.11	0.09	.25	.13	0.09	.14	.14	0.10	.14
Emotionale Belastung	-.07	0.08	.36	.08	0.11	.47	-.07	0.09	.44	.15	0.09	.10	-.10	0.11	.33	-.05	0.09	.60	.09	0.10	.39
R^2	.19		.01	.04		.21	.19		.00	.10		.02	.11		.05	.11		.03	.15		.03

Anmerkung. β = standardisierte Regressionskoeffizienten; p = Signifikanzniveau; SE = Standardfehler; R^2 = aufgeklärte Varianz.

Tabelle 35
 Regressionsanalysen zu den Bedingungen der Klasse

	Rezeption			Reflexion			Aktion			Evaluation											
	Selbstwirksamkeit im Umgang mit Ergebnisrückmeldungen			Zeitaufwand für die Nutzung der Ergebnisse			Informationsgewinn aus Ergebnisrückmeldungen			Gezielte Förderung			Wirkung auf Lernende			Wirkung auf das Kollegium			Wirkung auf Lehrende		
	β	SE	p	β	SE	p	β	SE	p	β	SE	p	β	SE	p	β	SE	p	β	SE	p
Kovariaten:																					
Geschlecht (0 = männlich)	.11	0.08	.18	.01	0.08	.86	-.22	0.08	.00	.05	0.08	.53	-.15	0.07	.03	-.01	0.09	.92	.03	0.08	.67
Alter	-.03	0.09	.77	-.08	0.09	.40	-.18	0.09	.04	-.16	0.08	.05	-.09	0.09	.33	.12	0.09	.19	-.01	0.09	.90
Bedingungen der Klasse:																					
Unterrichtsstörungen	-.03	0.10	.77	.04	0.10	.72	.07	0.09	.44	.14	0.09	.14	-.02	0.10	.84	.16	0.09	.09	.11	0.10	.23
Klassengröße	-.16	0.13	.22	.14	0.12	.27	-.18	0.12	.15	-.09	0.12	.45	-.07	0.11	.56	-.06	0.13	.67	-.21	0.12	.09
Klassenanteil an Familien mit Migrationshintergrund	.14	0.13	.27	-.07	0.10	.49	-.05	0.11	.64	-.13	0.12	.27	-.05	0.12	.67	-.02	0.12	.86	.06	0.12	.63
Klassenanteil mit engagierten, sich kümmernden Familien	.15	0.12	.24	-.00	0.12	.99	.19	0.11	.10	.23	0.10	.02	.15	0.11	.19	.12	0.11	.26	.05	0.13	.71
R^2	.04		.26	.02		.39	.11		.05	.10		.02	.05		.19	.05		.20	.03		.22

Anmerkung. β = standardisierte Regressionskoeffizienten; p = Signifikanzniveau; SE = Standardfehler; R^2 = aufgeklärte Varianz.

Tabelle 36
Regressionsanalysen zu den schulischen und externen Bedingungen

	Rezeption			Reflexion			Aktion			Evaluation											
	Selbstwirksamkeit im Umgang mit Er- gebnisrück-meldun- gen			Zeitaufwand für die Nutzung der Ergebnisse			Informations-ge- winn aus Ergebnis- rückmeldungen			Gezielte Förderung			Wirkung auf Lernende			Wirkung auf das Kollegium			Wirkung auf Lehrende		
	β	<i>SE</i>	<i>p</i>	β	<i>SE</i>	<i>p</i>	β	<i>SE</i>	<i>p</i>	β	<i>SE</i>	<i>p</i>	β	<i>SE</i>	<i>p</i>	β	<i>SE</i>	<i>p</i>	β	<i>SE</i>	<i>p</i>
Kovariaten:																					
Geschlecht (0 = männlich)	.13	0.08	.09	.01	0.08	.87	-.18	0.08	.03	.10	0.09	.25	-.08	0.06	.20	-.06	0.08	.48	.04	0.08	.59
Alter	-.03	0.08	.70	-.08	0.09	.37	-.18	0.08	.03	-.16	0.08	.04	-.12	0.08	.11	.09	0.07	.22	-.03	0.07	.66
Schulische Bedingungen:																					
Allgemeine Unterstützung Kollegium	-.22	0.09	.01	-.03	0.10	.80	.02	0.10	.87	-.02	0.11	.86	-.10	0.08	.21	.24	0.08	.00	-.01	0.09	.95
Allgemeine Unterstützung Schulleitung	.28	0.09	.00	.00	0.09	.96	-.07	0.10	.52	-.02	0.10	.88	.05	0.08	.53	-.10	0.08	.23	.18	0.07	.01
Akzeptanz der OD Kollegium	.34	0.08	.00	-.07	0.09	.46	.08	0.10	.43	.12	0.09	.18	.31	0.08	.00	.43	0.07	.00	.30	0.08	.00
Akzeptanz der OD Schulleitung	.06	0.09	.48	.12	0.10	.25	.12	0.09	.15	.07	0.09	.41	.19	0.08	.02	-.15	0.07	.02	.11	0.08	.16
Verpflichtung durch die Schulleitung	-.13	0.08	.10	.07	0.09	.45	-.09	0.08	.26	-.15	0.09	.10	-.14	0.07	.05	.22	0.07	.00	-.04	0.08	.60
Externe Bedingungen:																					
Allgemeine Unterstützung Eltern	.01	0.08	.88	-.05	0.11	.64	.01	0.10	.95	.05	0.10	.62	-.01	0.10	.91	-.04	0.09	.62	-.21	0.07	.00
Akzeptanz der OD Eltern	.02	0.09	.81	-.11	0.11	.29	.18	0.09	.05	.16	0.09	.07	.34	0.08	.00	.20	0.08	.02	.32	0.08	.00
<i>R</i> ²	.20		.00	.04		.26	.12		.02	.11		.03	.33		.00	.41		.00	.30		.00

Anmerkung. Online Diagnose (OD); β = standardisierte Regressionskoeffizienten; *p* = Signifikanzniveau; *SE* = Standardfehler; *R*² = aufgeklärte Varianz.

Assoziationen der jeweiligen Prozessschritte mit allen individuellen Bedingungen, Bedingungen der Klasse sowie schulischen und externen Bedingungen gemeinsam (Schritt 2). In den Tabellen 37 und 38 sind die multiplen linearen Regressionsanalysen zum zweiten Schritt der Beantwortung der folgenden Fragestellung aufgeführt: Welche Bedingungen sind mit den Prozessschritten der Nutzung aus dem Testverfahren *Online Diagnose* assoziiert? In diesem zweiten Schritt wurden zunächst die Assoziationen aller individuellen Bedingungen, Bedingungen der Klasse sowie die schulischen und externen Bedingungen mit den jeweiligen Prozessschritten in einem Modell analysiert. In Schritt 2 wurde überprüft, ob die Abhängigkeiten der in Schritt 1 signifikanten Zusammenhänge unter Aufnahme weiterer unabhängiger Variablen bestehen bleibt. Im Anschluss daran wurden weitere Regressionsanalysen mit lediglich den Variablen aus den individuellen Bedingungen, Bedingungen der Klasse sowie schulischen und externen Bedingungen durchgeführt, die in Tabelle 37 signifikant mit dem jeweiligen Prozessschritt assoziiert waren. Mit diesen zusätzlichen Regressionsanalysen wurde überprüft, wie viel Varianz die signifikanten unabhängigen Variablen allein in der abhängigen Variable aufklärten. Für das Geschlecht und Alter der Lehrkräfte wurde jeweils kontrolliert.

Die Darstellung der Ergebnisse wurden anhand der abhängigen Variablen strukturiert: Selbstwirksamkeit im Umgang mit den Ergebnismeldungen und der Zeitaufwand für die Nutzung der Ergebnismeldungen (Prozessschritt Rezeption), der Informationsgewinn aus den Ergebnismeldungen (Prozessschritt Reflexion), die gezielte Förderung (Prozessschritt Aktion), die Wirkung auf Lernende, auf das Kollegium und die Wirkung auf Lehrende (Prozessschritt Evaluation). Zunächst wurde berichtet, welche unabhängigen Variablen in allen Regressionsanalysen zu Schritt 1 (Tabellen 34 bis 36) und zu Schritt 2 (Tabellen 37 und 38) signifikant mit der jeweiligen abhängigen Variable assoziiert waren. Anschließend wurden die signifikanten Assoziationen, die lediglich in einzelnen Regressionsanalysen bestanden, berichtet. Abschließend wurde die Varianzaufklärung der im zweiten Schritt in Tabelle 37 und 38 berechneten Regressionsanalysen verglichen.

Selbstwirksamkeit im Umgang mit den Ergebnismeldungen (Prozessschritt Rezeption). Zur abhängigen Variable Selbstwirksamkeit im Umgang mit den Ergebnismeldungen waren in allen Regressionsanalysen (Schritt 1 und 2) die folgenden unabhängigen Variablen signifikant assoziiert: Einstellung zu Testuntersuchung, Computerängstlich-

keit, Diagnosekompetenz, allgemeine Unterstützung durch das Kollegium, allgemeine Unterstützung durch die Schulleitung und Akzeptanz der *Online Diagnose* durch das Kollegium. Unter Aufnahme aller Bedingungen in die Regressionsanalyse zeigte auch die unabhängige Variable Diagnosekompetenz einen signifikanten Zusammenhang mit der abhängigen Variable (Tabelle 37).

Die unabhängigen Variablen waren in geringer bis mittlerer Stärke mit der abhängigen Variable Selbstwirksamkeit im Umgang auf Ergebnismrückmeldungen assoziiert (Tabelle 38). Der geringste Zusammenhang auf dem Signifikanzniveau $< .05$ bestand zwischen der Diagnosekompetenz und der abhängigen Variable. Lehrkräfte, die ihre Diagnosekompetenz selbst als ausgeprägter einschätzten, gaben einen selbstwirksamen Umgang mit den Ergebnismrückmeldungen an ($\beta = .18; p < .05$). Im Rahmen der individuellen Bedingungen war außerdem die unabhängige Variable Einstellung zu Testuntersuchungen mit der abhängigen Variable assoziiert. Lehrkräfte, die positiv gegenüber Testuntersuchungen eingestellt waren, gaben gleichzeitig einen selbstwirksameren Umgang mit den Ergebnismrückmeldungen an ($\beta = .21; p < .05$). Negativ war die unabhängige Variable Computerängstlichkeit mit der abhängigen Variable assoziiert. Lehrkräfte, die sich als eher computerängstlich einordneten, berichteten gleichzeitig von einem weniger selbstwirksamen Umgang mit Ergebnismrückmeldungen ($\beta = -.24; p < .001$). Im Rahmen der schulischen und externen Bedingungen war die unabhängige Variable allgemeine Unterstützung durch das Kollegium negativ mit der Selbstwirksamkeit im Umgang mit Ergebnismrückmeldungen assoziiert. Lehrkräfte, die eine stärkere allgemeine Unterstützung durch das Kollegium wahrnahmen, gaben gleichzeitig an, weniger selbstwirksam mit den Ergebnismrückmeldungen zu sein ($\beta = -.18; p < .05$). Die unabhängige Variable allgemeine Unterstützung durch die Schulleitung war positiv mit der abhängigen Variable assoziiert. Fühlten sich die Lehrkräfte allgemein von der Schulleitung unterstützt, stieg der selbstwirksame Umgang mit den Ergebnismrückmeldungen ($\beta = .20; p < .05$). Der stärkste Zusammenhang im Rahmen der abhängigen Variable Selbstwirksamkeit im Umgang mit Ergebnismrückmeldungen bestand mit der unabhängigen Variable Akzeptanz der *Online Diagnose* durch das Kollegium. Schätzten die Lehrkräfte die *Online Diagnose* als ein im Kollegium akzeptiertes Instrument ein, stieg ihre Annahme, selbstwirksam mit den Ergebnismrückmeldungen zu sein ($\beta = .30; p < .001$).

Im Vergleich zum Modell in Tabelle 37, in welches alle unabhängigen Variablen in die Analyse aufgenommen wurden ($R^2 = .32$), klärte das Modell in Tabelle 38 dennoch 29 %

an Varianz in der abhängigen Variable auf. Das bedeutet, die signifikanten Variablen in Tabelle 38 trugen den Hauptanteil zur Varianzaufklärung in der abhängigen Variable bei.

Zeitaufwand für die Nutzung der Ergebnisrückmeldungen (Prozessschritt Rezeption). Mit der abhängigen Variable Zeitaufwand für die Nutzung der Ergebnisrückmeldungen war lediglich die unabhängige Variable Enthusiasmus für das Unterrichten in allen Regressionsanalysen (Schritt 1 und 2) signifikant assoziiert. Lehrkräfte, die sich selbst einen hohen Enthusiasmus für das Unterrichten zuschrieben, waren außerdem gegenüber dem Zeitaufwand für die Nutzung der Ergebnisrückmeldungen positiv eingestellt ($\beta = .18$; $p < .05$). Die aufgeklärte Varianz sank im Vergleich zum Modell in Tabelle 37 ($R^2 = .11$) auf 4 % ab und war nicht signifikant (Tabelle 38).

Informationsgewinn aus Ergebnisrückmeldungen (Prozessschritt Reflexion). Zur abhängigen Variable Informationsgewinn aus Ergebnisrückmeldungen waren die folgenden unabhängigen Variablen in allen Regressionsanalysen (Schritt 1 und 2) signifikant assoziiert: Einstellung zu Testuntersuchungen und Diagnosekompetenz. Die unabhängige Variable Akzeptanz der *Online Diagnose* durch die Eltern war lediglich im getrennt berechneten Modell zu Schritt 1 (Tabelle 36) mit der abhängigen Variable signifikant assoziiert.

Lehrkräfte, die positiv gegenüber Testuntersuchungen eingestellt waren, gaben einen höheren Informationsgewinn aus Ergebnisrückmeldungen an ($\beta = .24$; $p < .001$). Andererseits gaben die Lehrkräfte, die ihre Diagnosekompetenz als ausgeprägter einschätzten, einen geringeren Informationsgewinn aus Ergebnisrückmeldungen an ($\beta = -.19$; $p < .05$).

Im Vergleich zum Modell in Tabelle 37, in welches alle unabhängigen Variablen in die Analyse aufgenommen wurden ($R^2 = .26$), klärten die unabhängigen Variablen in Tabelle 38 immerhin noch 17 % an Varianz in der abhängigen Variable auf. Die signifikanten, unabhängigen Variablen trugen also einen Großteil zur Varianzaufklärung in der abhängigen Variable bei.

Gezielte Förderung (Prozessschritt Aktion). Zur abhängigen Variable gezielte Förderung waren die folgenden unabhängigen Variablen lediglich in den getrennt berechneten Modellen zu Schritt 1 signifikant assoziiert (Tabelle 35 und 36): Klassenanteil an Kindern mit engagierten, sich kümmernden Familien und Akzeptanz der *Online Diagnose* durch die Eltern. Wurden alle Bedingungen in das Modell aufgenommen (Tabelle 37), zeigte lediglich

die unabhängige Variable emotionale Belastung einen signifikanten Zusammenhang mit der abhängigen Variable. Die unabhängige Variable Klassenanteil an Kindern mit engagierten, sich kümmernden Familien lag in diesem Modell (Tabelle 37) nur knapp über dem Signifikanzniveau $< .05$ ($\beta = .20$; $p < .10$). In Tabelle 38 wurde das Modell zur gezielten Förderung vollständigshalber berechnet, ein signifikanter Zusammenhang der einzelnen Variable emotionale Belastung mit der abhängigen Variable bestand allerdings nicht. Die Varianzaufklärung des Modells, in das alle unabhängigen Variablen gleichzeitig aufgenommen wurden, lag bei 21 % (Tabelle 37).

Wirkung auf Lernende (Prozessschritt Evaluation). Zur abhängigen Variable Wirkung auf Lernende waren die folgenden unabhängigen Variablen in allen Regressionsanalysen (Schritt 1 und 2) signifikant assoziiert: Akzeptanz der *Online Diagnose* im Kollegium, Akzeptanz der *Online Diagnose* durch die Schulleitung und Akzeptanz der *Online Diagnose* durch die Eltern. Die unabhängigen Variablen Computerängstlichkeit und Verpflichtung durch die Schulleitung waren lediglich in den getrennt berechneten Modellen zu Schritt 1 (Tabellen 34 und 36) signifikant mit der abhängigen Variable assoziiert.

Die abhängige Variable Wirkung auf Lernende hing in den Analysen in Tabelle 37 ausschließlich mit unabhängigen Variablen der schulischen und externen Bedingungen zusammen. Der geringste Zusammenhang auf dem Signifikanzniveau $p < .05$ bestand zwischen der Akzeptanz der *Online Diagnose* durch die Schulleitung und der abhängigen Variable. Lehrkräfte, die die Akzeptanz der *Online Diagnose* durch die Schulleitung positiv einschätzten, gaben gleichzeitig eine stärkere Wirkung des Testverfahrens auf Lernende an ($\beta = .18$; $p < .05$). Einen stärkeren signifikanten Zusammenhang zeigte die unabhängige Variable Akzeptanz der *Online Diagnose* durch das Kollegium mit der abhängigen Variable. Lehrkräfte, die die Akzeptanz der *Online Diagnose* durch das Kollegium positiv einschätzten, gaben gleichzeitig eine stärkere Wirkung des Testverfahrens auf Lernende an ($\beta = .28$; $p < .001$). Der stärkste Zusammenhang im Rahmen der abhängigen Variable Wirkung auf Lernende bestand mit der Variable Akzeptanz der *Online Diagnose* durch die Eltern. Je positiver die Lehrkräfte die Akzeptanz der *Online Diagnose* durch die Eltern einschätzten, desto eher gaben sie gleichzeitig bei der Beantwortung des Fragebogens eine höhere Wirkung auf Lernende an ($\beta = .32$; $p < .001$).

Im Vergleich zum Modell in Tabelle 37, in welches alle unabhängigen Variablen in die Analyse aufgenommen wurden ($R^2 = .35$), klärten die verbleibenden unabhängigen Variablen in Tabelle 38 immer noch 31 % der Varianz in der abhängigen Variable auf. Die signifikanten unabhängigen Variablen trugen also einen Großteil zur Varianzaufklärung in der abhängigen Variable bei.

Wirkung auf das Kollegium (Prozessschritt Evaluation). Zur abhängigen Variable Wirkung auf das Kollegium waren die folgenden unabhängigen Variablen in allen Regressionsanalysen (Schritt 1 und 2) signifikant assoziiert: Allgemeine Unterstützung durch das Kollegium, Akzeptanz der *Online Diagnose* im Kollegium, Akzeptanz der *Online Diagnose* durch die Schulleitung, Verpflichtung durch die Schulleitung und Akzeptanz der *Online Diagnose* durch die Eltern. Die unabhängige Variable Computerängstlichkeit war lediglich in dem getrennt berechneten Modell in Schritt 1 (Tabelle 34) signifikant mit der abhängigen Variable assoziiert.

Die abhängige Variable Wirkung auf das Kollegium war in der Analyse in Tabelle 37 ausschließlich mit unabhängigen Variablen der schulischen und externen Bedingungen assoziiert. Die geringsten Zusammenhänge bestanden zwischen der Akzeptanz der *Online Diagnose* durch die Schulleitung sowie der Akzeptanz der *Online Diagnose* durch die Eltern mit der abhängigen Variable. Schätzten die Lehrkräfte die Akzeptanz der *Online Diagnose* durch die Schulleitung als gering ein, schätzten sie gleichzeitig die Wirkung der Ergebnisrückmeldungen auf das Kollegium als höher ein ($\beta = -.17$; $p < .01$). Umgekehrt hing die Akzeptanz der *Online Diagnose* durch die Eltern positiv mit der Wirkung auf das Kollegium zusammen ($\beta = .17$; $p < .05$). Des Weiteren waren die allgemeine Unterstützung durch das Kollegium, die Verpflichtung durch die Schulleitung und die Akzeptanz der *Online Diagnose* durch das Kollegium positiv mit der abhängigen Variable verknüpft. Wenn sich die Lehrkräfte durch ihr Kollegium allgemein unterstützt fühlten, gaben sie gleichzeitig eine stärkere Wirkung der Ergebnisrückmeldungen auf das Kollegium an ($\beta = .21$; $p < .01$). Fühlten sich die Lehrkräfte wiederum durch die Schulleitung zur Durchführung des Testverfahrens *Online Diagnose* verpflichtet, gaben sie gleichzeitig eine stärkere Wirkung der Ergebnisrückmeldungen auf das Kollegium an ($\beta = .21$; $p < .01$). Der stärkste Zusammenhang bestand zwischen der Einschätzung der Lehrkräfte, die *Online Diagnose* würde durch das Kollegium akzeptiert werden, und der Wirkung der Ergebnisrückmeldungen auf das Kollegium ($\beta = .43$; $p < .001$).

Im Vergleich zum Modell in Tabelle 37, in welches alle unabhängigen Variablen in die Analyse aufgenommen wurden ($R^2 = .48$), klärten die verbleibenden unabhängigen Variablen in Tabelle 38 immer noch 40 % der Varianz in der abhängigen Variable auf. Die unabhängigen Variablen in Tabelle 38 klärten einen Großteil der Varianz in der abhängigen Variable auf.

Wirkung auf Lehrende (Prozessschritt Evaluation). Zur abhängigen Variable Wirkung auf Lehrende waren die folgenden unabhängigen Variablen in allen Regressionsanalysen (Schritt 1 und 2) signifikant assoziiert: Einstellung zu Testuntersuchungen, Akzeptanz der *Online Diagnose* durch das Kollegium und Akzeptanz der *Online Diagnose* durch die Eltern. Die unabhängige Variable Akzeptanz der *Online Diagnose* durch die Schulleitung war lediglich in den Modellen zu Schritt 2 signifikant mit der abhängigen Variable assoziiert (Tabelle 37 und 38). Die unabhängigen Variablen allgemeine Unterstützung durch die Schulleitung und allgemeine Unterstützung durch die Eltern waren lediglich im getrennt gerechneten Modell in Schritt 1 (Tabelle 36) signifikant mit der abhängigen Variable assoziiert.

Die abhängige Variable Wirkung auf Lehrende hing mit unabhängigen Variablen der individuellen Bedingungen sowie der schulischen und externen Bedingungen zusammen. Die unabhängige Variable Einstellung zu Testuntersuchungen der individuellen Bedingungen war signifikant positiv mit der abhängigen Variable assoziiert. Gaben die Lehrkräfte eine positive Einstellung gegenüber Testuntersuchungen an, schätzten sie gleichzeitig die Wirkung auf Lehrende als stärker ein ($\beta = .20$; $p < .05$). Im Rahmen der schulischen und externen Bedingungen war die Akzeptanz der *Online Diagnose* durch die Eltern und die Akzeptanz der *Online Diagnose* durch das Kollegium positiv mit der abhängigen Variable assoziiert. Schätzten die Lehrkräfte die Akzeptanz der *Online Diagnose* durch die Eltern positiv ein, gaben sie gleichzeitig eine stärkere Wirkung auf Lehrende an ($\beta = .21$; $p < .001$). Schätzten die Lehrkräfte die Akzeptanz der *Online Diagnose* durch das Kollegium als positiv ein, gaben sie gleichzeitig eine stärkere Wirkung auf Lehrende an ($\beta = .27$; $p < .001$). Bei einer positiven Einschätzung der Akzeptanz der *Online Diagnose* durch die Schulleitung gaben sie ebenfalls eine stärkere Wirkung auf Lehrende an ($\beta = .16$; $p < .05$).

Im Vergleich zum Modell in Tabelle 37, in welches alle unabhängigen Variablen in die Analyse aufgenommen wurden ($R^2 = .37$), klärten die verbleibenden unabhängigen Variablen in Tabelle 38 immer noch 29 % der Varianz in der abhängigen Variable auf.

Kovariate. Im Rahmen der Analysen in den Tabellen 37 und 38 zeigte die Kontrollvariable Geschlecht einen signifikanten Zusammenhang mit der abhängigen Variable Informationsgewinn aus Ergebnisrückmeldungen auf ($\beta = -.17$; $p < .05$). Jüngere Lehrkräfte gaben eher an, einen Informationsgewinn aus den Ergebnisrückmeldungen zu erhalten. Weitere Zusammenhänge der Kovariaten auf dem Signifikanzniveau $p < .05$ bestanden nicht.

Zusammenfassung. Zusammenfassend kann festgehalten werden, dass zu den beiden unabhängigen Variablen Akzeptanz der *Online Diagnose* durch das Kollegium und Akzeptanz der *Online Diagnose* durch die Eltern ein Regressionskoeffizient in mittlerer Höhe ($\beta > .30$) berichtet werden konnte. Regressionskoeffizienten größer als .25 waren ebenfalls bei den Assoziationen zur unabhängigen Variable Akzeptanz der *Online Diagnose* durch das Kollegium zu finden. Regressionskoeffizienten größer als .20 zeigten die unabhängigen Variablen Einstellung zu Testuntersuchungen, Computerängstlichkeit, allgemeine Unterstützung durch das Kollegium, Verpflichtung durch die Schulleitung und Akzeptanz der *Online Diagnose* durch die Eltern auf.

Tabelle 37

Regressionsanalysen der abhängigen Variablen mit allen unabhängigen Variablen der Bedingungen

	Rezeption			Reflexion			Aktion			Evaluation											
	Selbstwirksamkeit im Umgang mit Ergebnisrück- meldungen			Zeitaufwand für die Nutzung der Ergebnisse			Informations- gewinn aus Ergebnisrück- meldungen			Gezielte Förderung			Wirkung auf Lernende			Wirkung auf das Kollegium			Wirkung auf Lehrende		
	β	<i>SE</i>	<i>p</i>	β	<i>SE</i>	<i>p</i>	β	<i>SE</i>	<i>p</i>	β	<i>SE</i>	<i>p</i>	β	<i>SE</i>	<i>p</i>	β	<i>SE</i>	<i>p</i>	β	<i>SE</i>	<i>p</i>
Kovariaten:																					
Geschlecht (0 = männlich)	.13	0.07	.06	-.06	0.09	.51	-.18	0.08	.02	.07	0.10	.49	-.10	0.07	.15	-.07	0.09	.40	.07	0.08	.34
Alter	.09	0.08	.31	-.02	0.10	.82	-.12	0.09	.17	-.14	0.09	.11	-.13	0.08	.08	.05	0.08	.51	-.00	0.08	.96
Individuelle Bedingungen:																					
Einstellung zu Testuntersuchungen	.21	0.09	.02	.10	0.09	.28	.17	0.09	.05	.11	0.08	.18	-.04	0.08	.60	.01	0.06	.89	.19	0.09	.04
Computerbezogene Einstellung	-.07	0.10	.47	-.01	0.13	.93	.08	0.09	.38	.03	0.09	.75	.03	0.10	.77	-.01	0.09	.89	.04	0.11	.74
Computerängstlichkeit	-.28	0.09	.00	.06	0.12	.60	.00	0.11	.98	-.02	0.09	.84	.08	0.09	.34	.12	0.08	.13	-.10	0.09	.29
Diagnosekompetenz	.17	0.08	.04	.03	0.08	.72	-.26	0.09	.00	.07	0.08	.39	-.00	0.08	.97	-.07	0.07	.27	.01	0.07	.88
Enthusiasmus für das Unterrichten	.06	0.11	.60	.24	0.11	.03	.06	0.10	.57	.16	0.10	.11	.11	0.10	.27	.06	0.09	.51	.04	0.10	.66
Emotionale Belastung	.00	0.09	.97	.03	0.14	.81	-.06	0.09	.53	.21	0.10	.03	.04	0.12	.75	.08	0.09	.37	.16	0.10	.13
Bedingungen der Klasse:																					
Unterrichtsstörungen	.02	0.09	.82	-.01	0.10	.96	.03	0.09	.74	.12	0.09	.17	-.01	0.09	.92	.11	0.08	.15	.08	0.08	.29
Klassengröße	-.04	0.11	.73	.07	0.13	.58	-.05	0.12	.67	-.07	0.12	.58	.04	0.11	.70	.12	0.11	.27	-.02	0.10	.82
Klassenanteil an Familien mit Migrationshintergrund	.11	0.11	.30	-.10	0.12	.40	-.14	0.12	.24	-.10	0.13	.41	-.11	0.12	.35	-.13	0.09	.15	-.05	0.10	.66
Klassenanteil mit engagierten, sich kümmernden Familien	.05	0.12	.67	.06	0.13	.63	.02	0.12	.85	.20	0.11	.06	-.02	0.10	.88	.02	0.10	.85	-.11	0.10	.30

(Fortsetzung)

	Selbstwirksamkeit im Umgang mit Ergebnisrück- meldungen			Zeitaufwand für die Nutzung der Ergebnisse			Informations- gewinn aus Ergebnisrück- meldungen			Gezielte Förderung			Wirkung auf Lernende			Wirkung auf das Kollegium			Wirkung auf Lehrende		
	β	<i>SE</i>	<i>p</i>	β	<i>SE</i>	<i>p</i>	β	<i>SE</i>	<i>p</i>	β	<i>SE</i>	<i>p</i>	β	<i>SE</i>	<i>p</i>	β	<i>SE</i>	<i>p</i>	β	<i>SE</i>	<i>p</i>
Schulische und externe Bedingungen:																					
Allgemeine Unterstützung Kollegium	-.23	0.08	.00	-.06	0.10	.57	.07	0.09	.43	-.01	0.10	.95	-.12	0.09	.16	.25	0.08	.00	.02	0.09	.78
Allgemeine Unterstützung Schulleitung	.18	0.09	.05	-.01	0.10	.92	-.09	0.10	.38	.03	0.11	.77	.07	0.08	.41	-.03	0.09	.73	.15	0.08	.06
Akzeptanz der OD Kollegium	.33	0.09	.00	-.07	0.09	.43	-.06	0.10	.58	.09	0.10	.37	.32	0.09	.00	.44	0.07	.00	.25	0.10	.02
Akzeptanz der OD Schulleitung	.06	0.09	.52	.05	0.10	.61	.14	0.09	.10	.03	0.09	.77	.17	0.09	.05	-.19	0.06	.00	.15	0.08	.05
Verpflichtung durch die Schulleitung	-.14	0.07	.06	.05	0.09	.58	-.09	0.08	.26	-.16	0.08	.06	-.14	0.08	.07	.22	0.07	.00	-.02	0.08	.84
Allgemeine Unterstützung Eltern	.02	0.10	.81	-.13	0.13	.31	.01	0.10	.95	-.01	0.10	.92	-.06	0.11	.61	-.02	0.10	.84	-.11	0.09	.20
Akzeptanz der OD Eltern	.03	0.10	.76	-.13	0.11	.26	.10	0.10	.30	.16	0.09	.06	.35	0.09	.00	.19	0.08	.02	.30	0.08	.00
<i>R</i> ²	.32		.00	.11		.06	.26		.00	.21		.00	.35		.00	.48		.00	.37		.00

Anmerkung. Online Diagnose (OD); β = standardisierte Regressionskoeffizienten; *p* = Signifikanzniveau; *SE* = Standardfehler; *R*² = aufgeklärte Varianz.

Tabelle 38

Regressionsanalysen der abhängigen Variablen mit allen signifikanten unabhängigen Variablen der Bedingungen

	Rezeption			Reflexion			Aktion			Evaluation											
	Selbstwirksamkeit im Umgang mit Ergebnisrückmel- dungen			Zeitaufwand für die Nutzung der Ergebnisse			Informationsgewinn aus Ergebnisrück- meldungen			Gezielte Förderung			Wirkung auf Lernende			Wirkung auf das Kollegium			Wirkung auf Lehrende		
	β	SE	p	β	SE	p	β	SE	p	β	SE	p	β	SE	p	β	SE	p	β	SE	p
<i>Kovariaten:</i>																					
Geschlecht (0 = männlich)	.12	0.07	.09	-.03	0.08	.73	-.17	0.08	.03	.09	0.08	.26	-.11	0.06	.08	-.07	0.08	.37	.03	0.07	.67
Alter	.05	0.08	.54	-.05	0.09	.62	-.13	0.09	.13	-.14	0.09	.09	-.14	0.08	.09	.09	0.07	.22	-.03	0.07	.66
<i>Individuelle Bedingungen:</i>																					
Einstellung zu Testuntersuchungen	.21	0.09	.02				.24	0.08	.00										.20	0.09	.03
Computerbezogene Einstellung																					
Computerängstlichkeit	-.24	0.07	.00																		
Diagnosekompetenz	.18	0.09	.04				-.19	0.08	.02												
Enthusiasmus für das Unterrichten				.18	0.08	.03															
Emotionale Belastung										.04	0.09	.67									
<i>Bedingungen der Klasse:</i>																					
Unterrichtsstörungen																					
Klassengröße																					
Klassenanteil an Familien mit Migrationshintergrund																					
Klassenanteil mit engagierten, sich kümmernden Familien																					

(Fortsetzung)

	Selbstwirksamkeit im Umgang mit Ergebnisrück- meldungen			Zeitaufwand für die Nutzung der Ergebnisse			Informations- gewinn aus Ergebnisrück- meldungen			Gezielte Förderung			Wirkung auf Lernende			Wirkung auf das Kollegium			Wirkung auf Leh- rende			
	β	<i>SE</i>	<i>p</i>	β	<i>SE</i>	<i>p</i>	β	<i>SE</i>	<i>p</i>	β	<i>SE</i>	<i>p</i>	β	<i>SE</i>	<i>p</i>	β	<i>SE</i>	<i>p</i>	β	<i>SE</i>	<i>p</i>	
Schulische und externe Be- dingungen:																						
Allgemeine Unterstützung Kollegium	-.18	0.08	.03													.21	0.08	.01				
Allgemeine Unterstützung Schulleitung	.20	0.08	.02																			
Akzeptanz der OD Kollegium	.30	0.08	.00									.28	0.08	.00	.43	0.07	.00	.27	0.08	.00		
Akzeptanz der OD Schulleitung												.18	0.08	.02	-.17	0.06	.01	.16	0.08	.04		
Verpflichtung durch die Schulleitung															.21	0.07	.01					
Allgemeine Unterstützung Eltern																						
Akzeptanz der OD Eltern												.32	0.07	.00	.17	0.08	.02	.21	0.07	.00		
<i>R</i> ²	.29		.00	.04	.25	.17	.00	.03	.28	.31	.00	.40		.00	.29		.00					

Anmerkung. Online Diagnose (OD); β = standardisierte Regressionskoeffizienten; *p* = Signifikanzniveau; *SE* = Standardfehler; *R*² = aufgeklärte Varianz.

7.4 Diskussion und Implikationen

In Teilstudie 3 wurde die folgende Fragestellung bearbeitet: Welche Bedingungen sind mit den Prozessschritten der Nutzung aus dem Testverfahren *Online Diagnose* assoziiert? Zur Untersuchung der Fragestellung wurden in einem ersten Schritt – getrennt für jede Bedingung und jeden Prozessschritt – Regressionsanalysen durchgeführt. Weiterhin wurden in einem zweiten Schritt alle unabhängigen Variablen der individuellen Bedingungen, der Bedingungen der Klasse sowie der schulischen und externen Bedingungen zu der jeweiligen abhängigen Variable gleichzeitig in die Regressionsanalyse aufgenommen. Außerdem wurde die Höhe der Varianzaufklärung durch die signifikanten unabhängigen Variablen in einer weiteren Regressionsanalyse überprüft. Die Varianzaufklärung in den Gesamtmodellen lag zwischen 11 % und 48 % (Tabelle 37) und in den vereinfachten Modellen immer noch zwischen 3 % und 40 % (Tabelle 38). Bereits hier fiel auf, dass die individuellen Bedingungen und die Bedingungen der Klasse eher in Bezug zu den Variablen der Prozessschritte Rezeption, Reflexion und Aktion standen. Ebenso standen die schulischen und externen Bedingungen eher in Bezug zu den Variablen des Prozessschritts Evaluation.

Im Folgenden werden die Ergebnisse nach den Bedingungen sortiert, zusammengefasst und diskutiert. Die Sortierung nach Bedingungen wurde gewählt, da Letztere im Fokus der Beantwortung der Fragestellung stehen. Dabei wird zunächst auf die individuellen Bedingungen eingegangen, im Anschluss auf die Bedingungen der Klasse und abschließend auf die schulischen und externen Bedingungen.

Einstellung zu Testuntersuchungen (Individuelle Bedingungen). Die unabhängige Variable Einstellung zu Testuntersuchungen war mit mehreren abhängigen Variablen assoziiert: Selbstwirksamkeit im Umgang mit den Ergebnisrückmeldungen, Informationsgewinn aus den Ergebnisrückmeldungen, gezielte Förderung und Wirkung auf Lehrende. Mit Ausnahme der gezielten Förderung bestanden die Zusammenhänge hauptsächlich im Abgleich mit Variablen, die die Lehrkraft als Person betrafen. Daraus konnte abgeleitet werden, dass die Einstellung der Lehrkräfte damit zusammenhing, wie sie mit Ergebnisrückmeldungen umgingen und wie sie diese bewerteten: selbstwirksamer Umgang, Neues erfahren und einen Nutzen davon ableiten. Auch die Literatur weist die Einstellung der Lehrkräfte zu externen Testuntersuchungen als einen wichtigen Prädiktor aus (Bromme & Kienhues, 2012). Groß Ophoff (2013) zeigt einen Effekt der Akzeptanz flächendeckender Schulleistungstests durch die Lehrkräfte auf die Fortbildungsaktivität und Unterrichtsveränderungen

auf. Lehrkräfte die externe Leistungsuntersuchungen akzeptierten, waren also auch eher bereit, an Fortbildungen teilzunehmen und Veränderungen im Unterricht durchzuführen. Fortbildungen können z.B. bei der Interpretation der Ergebnisrückmeldungen unterstützen oder Konzepte für Folgemaßnahmen bieten. Damit kann die Wahrscheinlichkeit der Nutzung von Ergebnisrückmeldungen erhöht werden. Kühle und Peek (2007) berichten ebenfalls einen Effekt der allgemeinen Akzeptanz von Lernstandserhebungen zusammen mit den Prädiktoren Nützlichkeit und Verständlichkeit auf die Bereitschaft der nutzenorientierten Rezeption von Ergebnisrückmeldungen. Auch Kohler (2005) weist einen bivariaten Zusammenhang der Bereitschaft zur Teilnahme an Evaluationsstudien mit der generellen Haltung zu vergleichenden Leistungsstudien nach.

Hinsichtlich der vorliegenden Untersuchung und der bereits veröffentlichten Ergebnisse ist anzunehmen, dass die Einstellung der Lehrkräfte einen Einfluss auf die Nutzung von Ergebnissen aus der Individual- und Förderdiagnostik haben kann. Dieses Ergebnis war zu erwarten, verdeutlicht jedoch auch in Bezug auf die Erfordernisse einer externen Individual- und Förderdiagnostik, dass die entsprechenden Instrumente nicht nur aufschlussreiche Informationen über die Leistung von Schülerinnen und Schüler liefern sollten, sondern dass auch die positive Einstellung der Lehrkräfte ein solches Instrument zu fördern vermag. Dies kann z.B. dadurch geschehen, dass Lehrkräfte bereits vor der Durchführung über Ziel und Zweck der Individual- und Förderdiagnostik informiert werden (z.B. Scriven, 1972). Ebenso kann die Einbindung der Nutzung der Ergebnisrückmeldungen in ein Fortbildungskonzept zu einer Einstellungsänderung bzw. zu einer positiven Einstellung führen.

Die Einstellung zu externen Testuntersuchungen sollte in Untersuchungen, die die Nutzung von Ergebnisrückmeldungen erfassen, grundsätzlich berücksichtigt werden. Da die Bewertung der Lehrkraft, wie z.B. die (subjektive) Einschätzung der Verständlichkeit oder Nützlichkeit von Ergebnisrückmeldungen, durch die allgemeine Einstellung zu externen Testuntersuchungen beeinflusst werden könnte. Die Einstellung zu externen Testuntersuchungen sollte als möglicher Moderator berücksichtigt werden. Im Rahmen dieser Untersuchung sind trotz der positiven Einstellung der Lehrkräfte zu externen Testuntersuchungen ($M = 2.67$) Unterschiede in den abhängigen Variablen festzustellen.

Idealerweise sollten die Effekte der Einstellung zu Testuntersuchungen im Rahmen einer Intervention überprüft werden. Die Testpersonen könnten vorab anhand einer Fortbildung oder Informationsveranstaltung über den Nutzen und die Möglichkeiten externer Eva-

lationen geschult werden. Im Vergleich mit einer Kontrollgruppe könnten mögliche Unterschiede bezüglich der Effekte der Einstellung von Lehrkräften auf die Nutzung von Ergebnisrückmeldungen festgestellt werden.

Computerängstlichkeit (Individuelle Bedingungen). Die Computerängstlichkeit wies Zusammenhänge mit drei abhängigen Variablen auf: Selbstwirksamkeit im Umgang mit Ergebnisrückmeldungen, Wirkung auf Lernende und Wirkung auf das Kollegium. Die Zusammenhänge schienen zunächst ambivalent: Zum einen war die Computerängstlichkeit mit der Selbstwirksamkeit im Umgang mit Ergebnisrückmeldungen negativ assoziiert, zum anderen war jedoch die Assoziation mit der Wirkung auf Lernende und der Wirkung auf das Kollegium positiv. Dass eine ausgeprägte Computerängstlichkeit mit einem geringeren selbstwirksamen Umgang mit den Ergebnisrückmeldungen zusammenhing, oder umgekehrt eine starke Selbstwirksamkeit im Umgang mit den Ergebnisrückmeldungen mit einer geringen Computerängstlichkeit zusammenhing, war plausibel. Ebenso könnte die Anwendung eines internetbasierten Diagnose- und Förderinstrumentes bei Lehrkräften, deren Computerängstlichkeit ausgeprägt ist, zu einer verstärkten Kooperation mit dem Kollegium und damit zu einer verstärkten Wirkung auf das Kollegium führen. Oder umgekehrt hinge eine geringere Kooperation mit dem Kollegium mit einer geringeren Computerängstlichkeit zusammen. Der Zusammenhang einer ausgeprägten Computerängstlichkeit mit einer verstärkten Wirkung auf Lernende war zunächst nicht offensichtlich. Umgekehrt bedeutete eine geringe Wirkung auf Lernende eine geringere Computerängstlichkeit der Lehrkräfte. Möglicherweise zeigten sich Lehrkräfte, die sich im Umgang mit Computern unsicher fühlten, beeindruckt von der Möglichkeit des internetbasierten Diagnose- und Förderinstrument und schrieben dem Instrument deshalb eine stärkere Wirkung zu. Gleichzeitig bewerteten Lehrkräfte, deren Computerängstlichkeit weniger ausgeprägt war, die Wirkung auf Lernende als weniger überzeugend. Ein signifikanter Zusammenhang über alle Regressionsanalysen hinweg blieb allerdings nur zwischen der Computerängstlichkeit und der Selbstwirksamkeit im Umgang mit Ergebnisrückmeldungen bestehen. Der Zusammenhang von Computerängstlichkeit und Testleistung wurde in der Forschung bisher hauptsächlich bei Schülerinnen und Schülern untersucht, allerdings uneinheitlich beantwortet (z.B. Gnambs, Batinic & Hertel, 2011; Shermis & Lombard, 1998; Williams & McCord, 2004). Richter, Naumann und Horz (2010) berichten von einem negativen Zusammenhang der Computerängstlichkeit z.B. mit den Jahren der Computernutzung oder der wöchentlichen Dauer der Computernutzung von

Studierenden. Ebenfalls wurde aufgezeigt, dass sich die Computerängstlichkeit bei Altersvergleichen der Probanden mit zunehmendem Alter ausgeprägter zeigt (z.B. Spannagel & Bescherer, 2009). Peschke (2011) geht auf die Professionalisierung der Lehrkräfte ein und bezeichnet diese im Bereich Medienbildung als noch nicht ausreichend. Deutlich wurde, dass vor allem ältere Lehrkräfte möglicherweise eine stärkere Unterstützung bei der Anwendung und Interpretation von computerbasierten Testverfahren benötigen. In den vorliegenden Analysen zur Lehrkraftbefragung im Rahmen des Testverfahrens *Online Diagnose* ließ sich ebenfalls ein bivariater Zusammenhang der Computerängstlichkeit mit dem Alter der Lehrkräfte feststellen ($r = .36$; $p < .001$), allerdings kein signifikanter Zusammenhang der Computerängstlichkeit bei älteren Lehrkräften mit der Nutzung von Ergebnisrückmeldungen. Ebenso hängen Geschlechtsunterschiede mit dem Umgang mit Computern zusammen. Bei Männern wurde allgemein eine höhere Computerselbstwirksamkeit gefunden (Whitley, 1997). Auch Richter, Naumann und Groeben (2001) fanden Geschlechterunterschiede zu Gunsten der Männer bei den Skalen zum Computer Literacy. Ein signifikanter Zusammenhang der Computerängstlichkeit speziell bei Lehrerinnen mit der Nutzung von Ergebnisrückmeldungen ließ sich in den vorliegenden Analysen im Rahmen des Testverfahrens *Online Diagnose* allerdings nur tendenziell finden. Möglicherweise ist dies durch die eingeschränkte Stichprobe zu erklären.

Diagnosekompetenz (Individuelle Bedingungen). Die Diagnosekompetenz hing in allen Regressionsanalysen signifikant negativ mit der abhängigen Variable Informationsgewinn aus den Ergebnisrückmeldungen zusammen. Folglich ging eine bessere Einschätzung der eigenen Diagnosekompetenz mit einem geringeren Informationsgewinn aus den Ergebnisrückmeldungen einher. Oder umgekehrt formuliert, Lehrkräfte, die den Informationsgewinn aus den Ergebnisrückmeldungen als größer angaben, schätzten gleichzeitig ihre eigene Diagnosekompetenz als weniger gut ein.

Unter Aufnahme aller Bedingungen in das Modell zur Selbstwirksamkeit im Umgang mit Ergebnisrückmeldungen war der Zusammenhang mit der Diagnosekompetenz ebenfalls signifikant (Tabelle 37). Der positive Zusammenhang zeigte an, dass Lehrkräfte, die ihre Diagnosekompetenz als gut einschätzten, auch ihren Umgang mit den Ergebnisrückmeldungen als selbstwirksam einschätzten. Das bedeutet, dass Lehrkräfte, die ihre Kenntnisse über Schülerleistungen als gut einschätzten, ebenso davon ausgingen, dass sie gut mit den Ergebnisrückmeldungen umgehen konnten. Eine geringe Selbstwirksamkeit im Umgang mit den

Ergebnisrückmeldungen hing also mit einer geringen Diagnosekompetenz zusammen. Das Ergebnis deutet an, dass die Lehrkräfte ihre Kompetenzen hinsichtlich der Nutzung von Ergebnisrückmeldungen aus einem internetgestützten Testinstrument ähnlich wie ihre allgemeine Diagnosekompetenz einschätzten.

Schlussfolgernd kann hier angemerkt werden, dass Lehrkräfte den Ergebnisrückmeldungen aus dem Testverfahren *Online Diagnose* nur dann vermehrt neue Informationen entnehmen konnten, wenn ihre Diagnosekompetenz nicht sehr ausgeprägt war. Dies könnte dafür sprechen, dass das Testverfahren *Online Diagnose* Lehrkräfte bei der Diagnostik von Schülerleistungen unterstützen kann. Andererseits könnte dies auch für eine geringe Testqualität sprechen. Das würde bedeuten, dass nur Lehrkräfte mit einer geringen Diagnosekompetenz Informationen aus den Ergebnisrückmeldungen entnehmen konnten, da diese nicht sehr gehaltvoll bzw. informativ waren. Weiterhin müsste dennoch überprüft werden, ob fehlende objektive Qualitätskriterien für das vorliegende Ergebnis verantwortlich waren. Weitere Untersuchungen zum Zusammenhang zwischen der Diagnosekompetenz und dem Informationsgewinn aus den Ergebnisrückmeldungen könnten überprüfen, ob die objektive Qualität der Tests hinsichtlich des Vorzeichens des Zusammenhangs eine moderierende Funktion hat.

Enthusiasmus für das Unterrichten (Individuelle Bedingungen). Die unabhängige Variable Enthusiasmus für das Unterrichten war in allen Regressionsanalysen signifikant mit der abhängigen Variable Zeitaufwand für die Nutzung der Ergebnisrückmeldungen assoziiert. Eine Lehrkraft, die sich hinsichtlich des Unterrichtens als enthusiastisch bezeichnete, gab gleichzeitig an, dass der Zeitaufwand für die Nutzung der Ergebnisrückmeldungen nicht zu hoch wäre. Umgekehrt formuliert würde dies bedeuten, dass eine Lehrkraft, die den Zeitaufwand für die Nutzung der Ergebnisrückmeldungen als zu hoch empfindet, gleichzeitig ihren Enthusiasmus für das Unterrichten als geringer einschätzt. In der Erwartung, dass Lehrkräfte Ergebnisrückmeldungen nutzen, ist ein gewisses Maß an Enthusiasmus unterstützend. Die Ergebnisse erhärten die Forschungsergebnisse in der bestehenden Literatur zur Bedeutung von Lehrerenthusiasmus und zur intrinsischen Motivation. Intrinsisch motivierte Lehrkräfte zeigen z.B. eine größere Bereitschaft zur Fortbildung oder Nutzung neuer Methoden (z.B. Baumert & Kunter, 2011; Ryan & Deci, 2000). Dieses Engagement ist vergleichbar mit dem Zeitaufwand für die Nutzung der Ergebnisrückmeldungen aus dem Testverfahren *Online Diagnose*. Der Enthusiasmus für das Unterrichten, also die Freude und

Begeisterung, Schülerinnen und Schülern etwas beizubringen, scheint dieses Engagement zu begünstigen.

Klassenanteil an Kindern mit engagierten, sich kümmernden Familien (Bedingungen der Klasse). Zu den Bedingungen der Klasse war die Variable Klassenanteil an Kindern mit engagierten, sich kümmernden Familien mit der abhängigen Variable gezielte Förderung signifikant assoziiert. Diese signifikante Assoziation bestand lediglich in der separat zu den Bedingungen der Klasse berechneten Regressionsanalyse. Lehrkräfte, deren Klassenanteil an Kindern aus engagierten, sich kümmernden Familien größer war, gaben an, häufiger Maßnahmen der gezielten Förderung eingesetzt zu haben. Die Angabe der Lehrkräfte, dass ihre Klasse einen geringeren Anteil an Kindern aus engagierten, sich kümmernden Familien hatte, würde in Verbindung mit einem geringeren Einsatz an Maßnahmen der gezielten Förderung stehen. Jeynes (2007) zeigte in seiner Meta-Analyse auf, dass Elternengagement einen Einfluss auf die Schülerleistungen hat. Ebenso berichtet Sacher (2012) von einem Einfluss von Familien auf den Schulerfolg von Schülerinnen und Schülern. Die Ergebnisse der vorliegenden Studie könnten in dieser Hinsicht dahingehend interpretiert werden, dass der Einfluss des Elternengagements auf Schülerleistungen auch das Engagement der Lehrkräfte beeinflusst. Möglicherweise wirkte das Engagement der Eltern auf die Lehrkräfte motivierend. Umgekehrt könnte das Ergebnis der vorliegenden Studie so interpretiert werden, dass ein geringes Elternengagement für die Lehrkräfte einen größeren Aufwand z. B. bei der Hausaufgabenbesprechung bedeutet. Dadurch könnte weniger Zeit für weitere Fördermaßnahmen verbleiben. Ebenso könnte das Interesse bzw. Engagement der Eltern an der Schulleistung ihrer Kinder die Lehrkräfte zu einem verstärkten Einsatz der Fördermaterialien aus dem Testverfahren *Online Diagnose* motiviert haben. Engagierte Eltern könnten aufgrund der Ergebnisrückmeldungen eine Erwartungshaltung in Bezug auf entsprechende Folgen hinsichtlich der Förderung der Schülerinnen und Schüler entwickeln. Diese Interpretationen müssten in Folgeuntersuchungen mit einer konkreteren Fragestellung zu den Erwartungen der Eltern in Bezug auf das Testverfahren *Online Diagnose* untersucht werden. Kohler (2009) weist in diesem Zusammenhang darauf hin, dass in ihrer Untersuchung die jeweilige Gruppe der Eltern oder der Lehrkräfte die Ursachen für schlechte Testergebnisse am wenigsten bei sich selbst suchten. Eltern suchten die Ursache schlechter Testergebnisse am häufigsten bei den Schülerinnen und Schülern selbst.

Allgemeine Unterstützung durch das Kollegium (Schulische und externe Bedingungen). In allen Regressionsanalysen war die allgemeine Unterstützung durch das Kollegium signifikant mit der Selbstwirksamkeit im Umgang mit den Ergebnisrückmeldungen sowie der Wirkung auf das Kollegium assoziiert. Lehrkräfte, die angaben, weniger selbstwirksam im Umgang mit den Ergebnisrückmeldungen zu sein, erhielten mehr Unterstützung durch das Kollegium. Oder umgekehrt formuliert: Lehrkräfte, die eine stärkere Unterstützung durch das Kollegium erhielten, gaben an, weniger selbstwirksam im Umgang mit den Ergebnisrückmeldungen zu sein. Außerdem berichteten Lehrkräfte, die sich allgemein durch das Kollegium unterstützt fühlten, ebenfalls von einer Wirkung des Testverfahrens *Online Diagnose* auf das Kollegium. Ebenso kann formuliert werden, dass bereits ein unterstützender Umgang im Kollegium bestanden haben dürfte, wenn die Lehrkräfte eine Wirkung des Testverfahrens auf das Kollegium wahrnahmen. Schneewinds Resultate (2006) im Rahmen von Interviews zur Studie BeLesen bestätigen das Ergebnis dieser Studie. Sie berichtet, dass vor allem dann eine Kooperation zu den Ergebnissen der Studie im Kollegium stattfand, wenn Teamarbeit bereits etabliert war. Mit Bezug auf diese Ergebnisse kann angenommen werden, dass der selbstwirksame Umgang mit den Ergebnisrückmeldungen für Lehrkräfte, die sich durch ihr Kollegium unterstützt fühlten, weniger relevant war. Für die Nutzung von Ergebnisrückmeldungen bedeutet dies, dass vor allem ein unterstützendes Kollegium und eine etablierte Kooperation im Kollegium hilfreich sein können. Auch Rothland (2007) zeigt auf, dass eine wahrgenommene soziale Unterstützung durch das Kollegium oder die Schulleitung zu einer Entlastung bei beruflichen Belastungen führen kann. Wird angenommen, dass Lehrkräfte Schwierigkeiten mit der Interpretation von Ergebnisrückmeldungen haben, kann diese als Belastung erlebt werden. Die Unterstützung und Hilfe im Kollegium wirkt hierbei wiederum entlastend.

Allgemeine Unterstützung durch die Schulleitung (Schulische und externe Bedingungen). In allen Regressionsanalysen war die allgemeine Unterstützung durch die Schulleitung signifikant mit der Selbstwirksamkeit im Umgang mit den Ergebnisrückmeldungen assoziiert. Außerdem war die allgemeine Unterstützung durch die Schulleitung in der getrennt zu den schulischen und externen Bedingungen berechneten Regressionsanalyse mit der Wirkung auf Lehrkräfte assoziiert. Fühlten sich die Lehrkräfte von der Schulleitung unterstützt, gaben sie an, selbstwirksamer mit den Ergebnisrückmeldungen umzugehen.

Gleichzeitig bedeutet dies, dass Lehrkräfte, die eine geringere Ausprägung des selbstwirksamen Umgangs mit den Ergebnisrückmeldungen angaben, auch die Unterstützung der Schulleitung geringer einschätzten. Eine als unterstützend empfundene Schulleitung ging damit einher, dass die Lehrkräfte gleichzeitig eine stärkere Wirkung des Testverfahrens *Online Diagnose* auf die Lehrkräfte angaben. Anders formuliert würde das Ergebnis bedeuten, dass eine stärkere Wirkung des Testverfahrens *Online Diagnose* auf die Lehrkräfte gleichzeitig mit der Angabe einer stärkeren allgemeinen Unterstützung durch die Schulleitung zusammenhing. Hier kann ähnlich wie bei der allgemeinen Unterstützung durch das Kollegium angenommen werden, dass eine unterstützende Schulleitung die Lehrkräfte entlastet (z.B. Rothland, 2007). Ein unterstützendes Klima in der Schule hilft der Lehrkraft im Umgang mit Ergebnisrückmeldungen, z.B. hinsichtlich deren Interpretation. Dadurch kann die Lehrkraft auch eine Unterstützung für ihre Diagnose- und Förderkompetenz, also eine stärkere Wirkung auf sich als Lehrkraft erleben.

Akzeptanz der *Online Diagnose* durch das Kollegium (Schulische und externe Bedingungen). Nicht nur die im Rahmen der individuellen Bedingungen erfasste Einstellung der Lehrperson gegenüber externen Testuntersuchungen war mit mehreren abhängigen Variablen assoziiert. Auch die Akzeptanz der *Online Diagnose* im Speziellen durch das Kollegium wies einen Zusammenhang mit vier abhängigen Variablen auf: Selbstwirksamkeit im Umgang mit den Ergebnisrückmeldungen, Wirkung auf Lernende, Wirkung auf das Kollegium und Wirkung auf Lehrende. Diese Zusammenhänge waren in allen Regressionsanalysen der vorliegenden Untersuchung signifikant. Lehrkräfte, die die Akzeptanz der *Online Diagnose* durch das Kollegium als höher beurteilten, gaben gleichzeitig eine positivere Selbstwirksamkeit und eine stärkere Wirkung auf Lernende, auf das Kollegium sowie auf Lehrende an. Ebenso bedeutete eine stärkere Wirkung auf Lehrende, dass die Lehrkräfte die Akzeptanz der *Online Diagnose* durch das Kollegium als positiver einschätzten. Auch die Aussagen der Lehrkräfte, eine stärkere Wirkung auf das Kollegium wahrzunehmen, hing mit der Aussage zusammen, das Kollegium würde das Testverfahren *Online Diagnose* akzeptieren. Ebenso verhielt es sich mit der Wirkung auf Lernende: wurde diese als höher wahrgenommen, waren auch die Angaben zur Akzeptanz des Testverfahrens *Online Diagnose* positiver. Genauso kann ein stärkerer selbstwirksamer Umgang mit den Ergebnisrückmeldungen mit der Zunahme an positiven Aussagen zur Akzeptanz des Testverfahrens *Online Diagnose* durch das Kollegium vermerkt werden.

Aus den Ergebnissen kann abgeleitet werden, dass die Akzeptanz des Testverfahrens *Online Diagnose* durch das Kollegium vor allem die wahrgenommene Wirkung des Testverfahrens stark prägen kann. Wie diese beiden Faktoren zusammenhängen, konnte durch die vorliegende Studie nicht geklärt werden. Denkbar wäre, dass eine starke Akzeptanz des Testverfahrens die Kooperation und die Gespräche im Kollegium beeinflusst und dadurch die Nutzung der Ergebnisrückmeldungen und Instrumente erleichtert bzw. erhöht und somit zu einer stärkeren Wirkung führt. Andererseits ist ebenso denkbar, dass die Akzeptanz des Testverfahrens *Online Diagnose* durch das Kollegium lediglich die Wahrnehmung der antwortenden Lehrkräfte beeinflusst und diese dadurch die Wirkung auf Lernende, das Kollegium und Lehrende als positiver wahrnehmen. Koch (2011) wies im Rahmen der Studie VERA darauf hin, dass die Akzeptanz von Lernstandserhebungen der durchführenden Lehrkräfte die Verständlichkeit, die Intensität der Nutzung und vor allem die Nützlichkeit beeinflusste. Bereits Schwippert (2005) wies auf eine Einstellungsänderung von Lehrkräften gegenüber Schulleistungsstudien hin. Um die Veränderungen nachhaltig zu fixieren, vermerkte er die notwendige Aus- und Weiterbildung von Lehrkräften hinsichtlich empirischer Methodenkenntnisse. Daraus lässt sich schließen, dass eine Aus- und Weiterbildung der Lehrkräfte, angepasst an die sich verändernde Schullandschaft, zur Akzeptanz evidenzbasierter Entscheidungsfindungen beiträgt. Die Akzeptanz bedingt wiederum die Nutzung und Wirkung der evidenzbasierten Entscheidungsfindung.

Akzeptanz der *Online Diagnose* durch die Schulleitung (Schulische und externe Bedingungen). Die Akzeptanz der *Online Diagnose* durch die Schulleitung war in allen Regressionsanalysen mit den Variablen Wirkung auf Lernende und Wirkung auf das Kollegium assoziiert. Wurde die Akzeptanz der *Online Diagnose* durch die Schulleitung höher eingeschätzt, wurde gleichzeitig die Wirkung auf Lernende höher eingeschätzt. Dies bedeutet außerdem, dass die Lehrkräfte die Akzeptanz der *Online Diagnose* durch die Schulleitung als gering einschätzten, wenn gleichzeitig die Wirkung auf Lernende als gering eingeschätzt wurde. Diese Ergebnisse konnten in dieser Form erwartet werden. Die Angaben der Lehrkräfte zur Akzeptanz der *Online Diagnose* durch die Schulleitung und die Wirkung auf das Kollegium waren negativ assoziiert. Eine höhere Akzeptanz der *Online Diagnose* durch die Schulleitung, ging mit Angaben der Lehrkräfte zu einer geringeren Wirkung auf das Kollegium einher. Umgekehrt formuliert würde eine geringere Wirkung auf das Kollegium eine

höhere Akzeptanz der *Online Diagnose* durch die Schulleitung bedeuten. Eine bivariate Korrelation zwischen den beiden Variablen bestand nicht.

Das Ergebnis bedeutet möglicherweise, dass eine höhere Akzeptanz der Schulleitung gleichzeitig ein größeres Interesse und dadurch eine stärkere Unterstützung der Lehrkraft durch die Schulleitung hervorrief, so dass eine Kooperation mit dem Kollegium weniger wichtig erschien. Im Rahmen der oben dargestellten Unterstützung durch die Schulleitung wurde ein Zusammenhang mit dem selbstwirksamen Umgang mit den Ergebnisrückmeldungen durch die Lehrkraft festgestellt. Denkbar wäre, dass die Unterstützung der Schulleitung den negativen Zusammenhang zwischen der Akzeptanz der *Online Diagnose* durch die Schulleitung und der Wirkung auf das Kollegium beeinflusste.

Verpflichtung durch die Schulleitung (Schulische und externe Bedingungen).

Die unabhängige Variable Verpflichtung zur Durchführung der *Online Diagnose* durch die Schulleitung hing negativ mit der Variable Wirkung auf Lernende und positiv mit der Variable Wirkung auf das Kollegium zusammen, wobei lediglich der Zusammenhang zwischen den Variablen Verpflichtung durch die Schulleitung und Wirkung auf das Kollegium in allen Regressionsanalysen bestand. Bei einer empfundenen Verpflichtung zur Durchführung der *Online Diagnose* durch die Schulleitung gaben die Lehrkräfte gleichzeitig eine geringere Wirkung auf Lernende an. Ebenso kann das Ergebnis wie folgt beschrieben werden: Wenn Lehrkräfte die Wirkung auf Lernende weniger stark wahrnahmen, gaben sie gleichzeitig an, eine stärkere Verpflichtung zu empfinden. Schätzten die Lehrkräfte die Verpflichtung zur Teilnahme an der *Online Diagnose* durch die Schulleitung höher ein, schätzten sie gleichzeitig die Wirkung auf das Kollegium höher ein. Dies bedeutet ebenso, dass eine höher eingeschätzte Wirkung auf das Kollegium mit einer höher empfundenen Verpflichtung durch die Schulleitung einherging.

Die Ergebnisse können wie folgt interpretiert werden. Führen die Lehrkräfte das Testverfahren *Online Diagnose* aufgrund der Verpflichtung durch die Schulleitung durch, muss dies nicht bedeuten, dass sie selbst von dem Konzept der *Online Diagnose* überzeugt sind. Ist das Testverfahren von Lehrkräften nicht akzeptiert, bewerten sie unter Umständen die Wirkung auf Lernende negativ. Ebenso ist durch die Verpflichtung zur Durchführung des Testverfahrens *Online Diagnose* eine verstärkte Kooperation und dadurch Wirkung auf das Kollegium denkbar. Möglicherweise suchen die Lehrkräfte besonders dann Unterstützung im Kollegium, wenn die Schulleitung den Einsatz des Testverfahrens verfolgt und evtl.

bewertet. Ob Effekte auf die Wirkung des Testverfahrens durch eine Verpflichtung zur Durchführung verursacht werden, wäre für weitere Untersuchungen von Interesse.

Allgemeine Unterstützung durch die Eltern (Schulische und externe Bedingungen). Der negative Zusammenhang der Variable allgemeine Unterstützung durch die Eltern mit der abhängigen Variable Wirkung auf Lehrende trat nur in der getrennt berechneten Regressionsanalyse zu den schulischen und externen Bedingungen auf (Schritt 1). Ein Zusammenhang zwischen den Variablen in den Regressionsanalysen zu Schritt 2 bestand nicht. Ebenso korrelierten die Variablen auch bivariat nicht miteinander. Aus diesem Grund wird das Ergebnis nicht näher interpretiert.

Akzeptanz der *Online Diagnose* durch die Eltern (Schulische und externe Bedingungen). Die Akzeptanz der *Online Diagnose* durch die Eltern war mit vier abhängigen Variablen assoziiert: Informationsgewinn aus den Ergebnisrückmeldungen, Wirkung auf Lernende, Wirkung auf das Kollegium und Wirkung auf die Lehrkräfte. Allerdings war der Zusammenhang zwischen den Variablen Informationsgewinn aus den Ergebnisrückmeldungen und der Akzeptanz der *Online Diagnose* durch die Eltern lediglich im Rahmen der getrennt zu den schulischen und externen Bedingungen gerechneten Regressionsanalyse signifikant. Der signifikante Zusammenhang zwischen Akzeptanz der *Online Diagnose* durch die Eltern und den weiteren abhängigen Variablen bestand in allen Regressionsanalysen. Gingen die Lehrkräfte von einer höheren Akzeptanz der *Online Diagnose* durch die Eltern aus, schätzten sie auch die Wirkung auf Lernende, die Wirkung auf das Kollegium und die Wirkung auf Lehrende als höher ein. Umgekehrt bedeutete eine höhere Wirkung auf Lernende gleichzeitig eine höhere Einschätzung der Akzeptanz der *Online Diagnose* durch die Eltern. Dies galt ebenso für die Wirkung auf das Kollegium und die Wirkung auf Lehrende: wurden diese als höher eingeschätzt, wurde gleichzeitig die Akzeptanz der *Online Diagnose* durch die Eltern als höher eingeschätzt. Ein verstärkter Informationsgewinn aus den Ergebnisrückmeldungen hing mit einer durch die Lehrkräfte höher eingeschätzten Akzeptanz der *Online Diagnose* durch die Eltern zusammen. Anders formuliert war die Akzeptanz des Testverfahrens *Online Diagnose* durch die Eltern prädiktiv für den Informationsgewinn aus den Ergebnisrückmeldungen.

Ähnlich wie im Abschnitt Klassenanteil an Kindern engagierter, sich kümmernder Eltern kann auch hier ein positiver Effekt des Elternengagements auf die Wirkung des Testverfahrens *Online Diagnose* vermutet werden (z. B. Jeynes, 2007). Akzeptierten die Eltern das Testverfahren *Online Diagnose*, zeigten sie gegenüber der Durchführung möglicherweise größeres Interesse. Dadurch könnte ein höheres Engagement der Lehrkräfte angenommen werden und dadurch wiederum ein stärkerer Informationsgewinn und eine stärkere Wirkung auf Lernende, das Kollegium und Lehrende. Die aufgeführten Interpretationen bleiben jedoch Annahmen und müssten durch weitere Untersuchungen zunächst bestätigt werden. Ebenso wäre in weiteren Untersuchungen zu überprüfen, ob offizielle Informationen oder Informationsveranstaltungen für Eltern von Seiten einer Forschungseinrichtung oder Einrichtungen der Schulinspektion zu einem insgesamt größeren Interesse an externen Erhebungsverfahren führen kann.

Zusammenfassung. Zusammenfassend kann aus den vorliegenden Ergebnissen abgeleitet werden, dass vor allem Merkmale, die die Lehrperson und die pädagogischen Akteure der Schule betrafen, relevant für die Nutzung von Ergebnisrückmeldungen waren. Merkmale der Klasse schienen für die Nutzung von Ergebnisrückmeldungen weniger bedeutsam zu sein. Die Merkmale der Lehrperson bezogen sich beispielsweise auf die Einstellung der Lehrkraft gegenüber externen Testuntersuchungen. Merkmale der pädagogischen Akteure der Schule bezogen sich unter anderem ebenfalls auf Kriterien der Akzeptanz eines Testverfahrens.

Auffallend war, dass vorrangig signifikante Zusammenhänge zwischen Kriterien der individuellen Bedingungen und den Prozessschritten Rezeption und Reflexion auszumachen waren. Ebenso waren signifikante Zusammenhänge eher zwischen Kriterien der schulischen und externen Bedingungen und den Variablen zum Prozessschritt Evaluation auszumachen. In den getrennt berechneten Regressionsanalysen zu Schritt 1 zeigte sich außerdem ein signifikanter Zusammenhang zwischen den Bedingungen der Klasse und dem Prozessschritt Aktion.

Als Fazit kann abgeleitet werden, dass für die direkte Nutzung von Ergebnisrückmeldungen die Merkmale einer Lehrkraft besonders relevant sind. Für die Wirkung eines Testverfahrens im Allgemeinen sind eher die Einflussfaktoren des schulischen Umfelds relevant.

Diese Ergebnisse sind nicht unbedingt überraschend, liefern jedoch einen empirischen Nachweis dafür, wie Lehrkräfte die Nutzung von Ergebnisrückmeldungen bewerten. Auf diese Ergebnisse können weitere Studien aufbauen.

8 Gesamtdiskussion

Das Erkenntnisinteresse dieser Arbeit galt der Nutzung von Ergebnisrückmeldungen aus einem Diagnose- und Förderinstrument. Durch das Testverfahren *Online Diagnose* wurden Individualdiagnostiken an die Lehrkräfte zurückgemeldet. Zur Erhebung der Schülerleistungen wurden im Rahmen des Testverfahrens *Online Diagnose* Haupt-, Detail- und Nachtests eingesetzt. Die daraus resultierenden Ergebnisse wurden auf unterschiedlichen Aggregatensebenen online an die Lehrkräfte zurückgemeldet. Außerdem erhielten die Lehrkräfte entsprechende Fördermaterialien für jede Schülerin und jeden Schüler. Die Besonderheit des Testverfahrens *Online Diagnose* lag in der elektronischen Datenerhebung und -rückmeldung.

Die Nutzung der Ergebnisrückmeldungen wurde im Rahmen von drei Teilstudien untersucht. Zunächst wurden deskriptive Befunde berichtet (Teilstudie 1). Außerdem wurden Assoziationen der Prozessschritte der Nutzung von Ergebnisrückmeldungen mit der Wirkung des Testverfahrens *Online Diagnose* überprüft (Teilstudie 2). Abschließend wurden Einflussfaktoren (individuelle Bedingungen, Bedingungen der Klasse sowie schulische und externe Bedingungen) auf die Prozessschritte der Nutzung überprüft (Teilstudie 3).

Die einzelnen Ergebnisse wurden bereits im Anschluss an die jeweiligen Teilstudien zusammengefasst und diskutiert. Im Rahmen der Gesamtdiskussion werden die zentralen Befunde der Teilstudien 1 bis 3 dargelegt und diskutiert. Anschließend erfolgt eine kritische Einordnung der Ergebnisse. Ergänzend wird das Studiendesign kritisch diskutiert. Weiterhin werden Implikationen für die weitere Forschung und Praxis aufgeführt. Die Gesamtdiskussion wird mit einem Fazit abgeschlossen.

8.1 Zentrale Befunde zu den Teilstudien 1 bis 3

Die Gesamtstrategie zum Bildungsmonitoring in Deutschland beinhaltet unter anderem die Durchführung von Schulleistungsstudien, die individuelle Förderung von Schülerinnen und Schülern und die Förderung der Diagnosekompetenz von Lehrkräften (KMK & IQB, 2006). Von einer automatisierten Nutzung von Ergebnisrückmeldungen nach der Rückmeldung von Schulleistungsstudien oder Individualdiagnostiken kann allerdings nicht ausgegangen werden (z.B. Helmke & Hosenfeld, 2005). Vielmehr sind einige Anforderungen zu erfüllen, um

das Gelingen der Nutzung von Ergebnisrückmeldungen zu ermöglichen. Zunächst stehen Anforderungen an Lehrkräfte, die eine Übermittlungsfunktion im Rückmeldeprozess innehaben, im Vordergrund (z.B. Hattie & Timperley, 2007). Neben anderen Kriterien ist das technische, statistische und inhaltliche Verständnis der Rückmeldungen zu nennen (z.B. Stauke, 2008). Außerdem sind Anforderungen an die Ergebnisrückmeldungen selbst zu stellen (z.B. Stauke, 2008). Kriterien wie z.B. die zeitnahe Übermittlung der Informationen oder die Bereitstellung von Vergleichswerten sind relevant (z.B. Breiter & Light, 2006). Einzelne Rezeptionsstudien, die Individualdaten erfassten, konnten aufzeigen, dass Lehrkräfte die Ergebnisrückmeldungen generell als verständlich bewerteten, die Akzeptanz der Rückmeldestudien einen wichtigen Einflussfaktor darstellte und die Kooperation eher auf informellem Wege innerhalb der Schule stattfand (z.B. Dederling, 2011; Groß Ophoff, 2013; Hosenfeld, 2010; Koch, 2011; Müller, 2009; Schneewind, 2006). Es konnte auch aufgezeigt werden, dass aus den Ergebnisrückmeldungen eher selten Konsequenzen folgten (Tresch, 2007). Wurden allerdings Konsequenzen umgesetzt, verwendeten die Lehrkräfte eher bekannte als innovative Verfahren. Das zur Nutzung von Ergebnisrückmeldungen entwickelte Zyklusmodell der Rückmeldung strukturiert die Nutzung in die Prozessschritte Rezeption, Reflexion, Aktion und Evaluation (z.B. Helmke, 2004; Helmke & Hosenfeld, 2005; Hosenfeld & Groß Ophoff, 2007). Außerdem sieht es die Berücksichtigung von Einflussfaktoren wie die individuellen Bedingungen, die Bedingungen der Klasse sowie die schulischen und externen Bedingungen vor. Die in dieser Arbeit durchgeführten Teilstudien 2 und 3 orientierten sich an dieser Strukturierung.

Zentrale Befunde der deskriptiven Ergebnisse (Teilstudie 1). Die Lehrkräfte setzten vor allem den Haupttest ein und nutzten die zurückgemeldeten Fördermaterialien. Die Detail- und Nachtests wurden hinsichtlich der Durchführung eher nachrangig behandelt. Außerdem nutzte die Mehrheit der Lehrkräfte die Ergebnisrückmeldungen gezielt für einzelne Schülerinnen und Schüler und weniger für die gesamte Klasse. Für die Bearbeitung und Besprechung der Fördermaterialien wurde im Vergleich zu den Tests und Ergebnisrückmeldungen mehr Zeit aufgewandt. Die Lehrkräfte tauschten sich hauptsächlich im Kollegium über die Ergebnisrückmeldungen aus. Die Schülerinnen und Schüler sowie die Eltern wurden eher über die Ergebnisrückmeldungen informiert. Die technische Ausstattung erwies sich laut Angaben der Lehrkräfte als gut. Allerdings gab dennoch mehr als ein Viertel der Lehrkräfte an, bei der Durchführung der Tests technische Probleme gehabt zu haben. Gründe

für ein schwaches Testergebnis sahen die Lehrkräfte hauptsächlich im tatsächlichen Förderbedarf, dem Migrationshintergrund oder der Motivation der Schülerinnen und Schüler. Einflüsse wie Lärm und Zeitdruck schienen nicht störend gewirkt zu haben. Fortbildungsangebote wurden sehr selten genutzt. Gründe hierfür liegen allerdings nicht vor. Sogenannte schriftliche Hilfen wie z.B. die Kurzanleitung und der Quick-Tipp zum Testverfahren *Online Diagnose* wurden als hilfreich eingeschätzt.

Hinsichtlich motivationaler Aspekte schätzten die Lehrkräfte den Haupttest für die Schülerinnen und Schüler als eher wichtig bzw. den intrinsischen Wert als eher hoch ein. Die relativen Kosten für Lernende wurden als gering bewertet. Etwas höher schätzten die Lehrkräfte die Nützlichkeit des Haupttests für die Schülerinnen und Schüler ein. Für sich selbst bewerteten die Lehrkräfte die Nützlichkeit/Wichtigkeit und den intrinsischen Wert als hoch. Die relativen Kosten schätzte die Hälfte der Lehrkräfte für sich selbst ebenfalls als hoch, die andere Hälfte als gering ein.

Hinsichtlich der Fördermaterialien schätzten die Lehrkräfte den intrinsischen Wert bzw. die Wichtigkeit für die Schülerinnen und Schüler als gering ein. Ebenso wurden die relativen Kosten für die Schülerinnen und Schüler durch die Lehrkräfte als gering bewertet. Die Nützlichkeit wurde allerdings als höher eingeschätzt. Für sich selbst schätzten die Lehrkräfte den intrinsischen Wert hinsichtlich der Fördermaterialien als eher gering ein. Die Nützlichkeit/Wichtigkeit wurde als hoch und die relativen Kosten als angemessen bewertet.

Die Verständlichkeit und Eignung der Testaufgaben zur *Online Diagnose* bewerteten die Lehrkräfte als sehr gut. Ebenso sahen sie den Bezug des Testverfahrens *Online Diagnose* zum Bildungsplan der Grundschule als sehr gut an. Weitere subjektive Qualitätseinschätzungen der Lehrkräfte zum Haupttest, den Ergebnisrückmeldungen und den Fördermaterialien fielen ebenfalls sehr gut aus.

Zu den Schritten der Nutzung konnten folgende deskriptive Befunde erfasst werden: Die Angaben, die die Lehrkräfte zum Prozessschritt Rezeption machten, waren hinsichtlich des Zeitaufwands für die Ergebnisrückmeldungen uneinheitlich. Ein Teil gab den Zeitaufwand als hoch, ein anderer Teil als gering an. Zudem schätzte die Mehrheit der Lehrkräfte ihren Umgang mit den Ergebnisrückmeldungen als selbstwirksam ein. Zum Prozessschritt Reflexion erschien, laut den Angaben der Lehrkräfte, der Informationsgewinn aus den Ergebnisrückmeldungen als nicht unbedingt groß. Die Lehrkräfte gaben im Rahmen des Prozessschrittes Aktion an, die gezielte Förderung häufiger als vor der Durchführung des Test-

verfahrens *Online Diagnose* eingesetzt zu haben. Die Wirkung auf Lernende, auf das Kollegium und auf Lehrende des Prozessschrittes Evaluation schätzten die Lehrkräfte eher als gering ein.

Zentrale Befunde zu den Prozessschritten der Nutzung: Rezeption, Reflexion, Aktion und Evaluation (Teilstudie 2). Die zweite Teilstudie zum Testverfahren *Online Diagnose* erfasste Assoziationen der Prozessschritte Rezeption, Reflexion und Aktion mit der Wirkung auf Lernende, auf das Kollegium und auf Lehrende.

An den Ergebnissen zur Wirkung auf Lernende wurde deutlich, dass die Lehrkräfte hierbei vor allem einen Zusammenhang mit Kriterien sahen, die sich auf die Instrumente und Maßnahmen des Testverfahrens *Online Diagnose* bezogen. Die Hypothesen zur subjektiven Qualitätseinschätzung des Haupttests und der Fördermaterialien, zum Prozessschritt Rezeption und zum Prozessschritt Aktion konnten vorrangig bestätigt werden. Der Prozessschritt Rezeption bezog sich hauptsächlich auf den Zeitaufwand für die Nutzung der Ergebnisrückmeldungen. Der Prozessschritt Aktion bezog sich hierbei auf die gezielte Förderung und die Einbindung der Fördermaterialien in den Unterricht.

Die Annahme eines Zusammenhangs zwischen der Wirkung auf das Kollegium, der subjektiven Qualitätseinschätzung und den Prozessschritten Rezeption, Reflexion und Aktion konnte nicht bestätigt werden. Die Ergebnisse werden im Rahmen der Teilstudie 2 (Kapitel 6) diskutiert.

Hinsichtlich der Wirkung auf Lehrende wurden Zusammenhänge vor allem mit Kriterien festgestellt, durch die sich die Lehrkräfte unterstützt fühlten. Die Hypothesen konnten für den Bezug zum Bildungsplan der Grundschule und die Prozessschritte Reflexion und Aktion bestätigt werden. Der Prozessschritt Reflexion bezog sich hierbei auf den Informationsgewinn aus den Ergebnisrückmeldungen und der Prozessschritt Aktion auf die gezielte Förderung, den Einsatz der Fördermaterialien im Deutsch- oder Förderunterricht und die Einbindung von Fördermaterialien in den Unterricht.

Zentrale Befunde mit Bezug auf individuelle Bedingungen, Bedingungen der Klasse sowie schulische und externe Bedingungen (Teilstudie 3). Die dritte Teilstudie zum Testverfahren *Online Diagnose* erfasste Assoziationen von Einflussfaktoren mit den

Prozessschritten der Nutzung. Die Häufigkeit der Zusammenhänge, die Höhe der Regressionskoeffizienten sowie die Höhe der Varianzaufklärung wurden unter anderem mitberücksichtigt.

Die häufigsten und höchsten Assoziationen in insgesamt mittlerer Höhe hatte die Akzeptanz der *Online Diagnose* durch das Kollegium (schulische und externe Bedingungen) mit den Schritten der Nutzung von Ergebnisrückmeldungen, gefolgt von der Akzeptanz der *Online Diagnose* durch die Eltern (schulische und externe Bedingungen) und der Einstellung zu externen Testuntersuchungen der Lehrkraft (individuelle Bedingungen). An diesen Ergebnissen wurde für die Nutzung von Ergebnisrückmeldungen deutlich, dass die Einstellung und Akzeptanz der Lehrkraft und der pädagogischen Akteure der Schule äußerst wichtig ist. Weiterhin waren, wenn auch in eher geringer Höhe, die Assoziationen der Prozessschritte der Nutzung von Ergebnisrückmeldungen mit den folgenden Variablen zu nennen: allgemeine Unterstützung durch das Kollegium, allgemeine Unterstützung durch die Schulleitung und Akzeptanz der *Online Diagnose* durch die Schulleitung (schulische und externe Bedingungen). Außerdem waren folgende Variablen mindestens einmal mit den Prozessschritten der Nutzung von Ergebnisrückmeldungen assoziiert: Computerängstlichkeit, Diagnosekompetenz (individuelle Bedingungen) und Verpflichtung durch die Schulleitung (schulische und externe Bedingungen). Eine detaillierte Darstellung der Assoziationen enthält Kapitel 7 dieser Arbeit.

8.2 Kritische Einordnung der Ergebnisse

Die Befunde aus den drei Teilstudien müssen kritisch diskutiert werden. Die geringe Rücklaufquote (15 %; N = 150)⁵¹ legt die Vermutung nahe, dass besonders engagierte Lehrkräfte den Fragebogen beantwortet haben. Diese Annahme wurde durch die Angaben der Lehrkräfte zum Enthusiasmus für das Unterrichten bestätigt. Die dem Enthusiasmus für das Unterrichten zugehörigen Items wurden durchweg positiv beantwortet. Außerdem konnte in Teilstudie 3 aufgezeigt werden, dass der Enthusiasmus für das Unterrichten mit dem Zeitaufwand für die Nutzung der Ergebnisrückmeldungen zusammenhängt.

⁵¹ Die Repräsentativität der Stichprobe war nur bedingt gegeben und wurde hauptsächlich durch die geringe Rücklaufquote negativ beeinflusst. Die Repräsentativität wird im nachfolgenden Abschnitt 8.3 zur kritischen Einordnung des Studiendesigns nochmals aufgegriffen.

Die Stichprobe zur Befragung der Deutschlehrkräfte war wie folgt zu beurteilen: Der weibliche Anteil der antwortenden Lehrkräfte im Rahmen des Testverfahrens *Online Diagnose* betrug 82 %. Das statistische Landesamt Baden-Württemberg wies für das Schuljahr 2012/2013 einen ähnlichen Wert aus. Insgesamt betrug der Anteil weiblicher Lehrkräfte an Grund-, Haupt- und Werkrealschulen 79 % (Statistisches Landesamt, 2014). Die Lehrkräfte waren bei der Beantwortung des Fragebogens im Durchschnitt 40 Jahre alt. An den baden-württembergischen Grund-, Haupt- und Werkrealschulen lag im Schuljahr 2012/2013 der Prozentsatz der 35- bis 44-Jährigen bei einem Anteil von 24 % der Lehrkräfte. Die Tendenz an den Grund-, Haupt- und Werkrealschulen ging eher hin zu einem höheren Altersdurchschnitt. Der Anteil an 55-jährigen und älteren Lehrkräften betrug im Schuljahr 2012/2013 sogar 36 % (Statistisches Landesamt, 2014). Damit lag die vorliegende Stichprobe eher im Durchschnitt. Ebenso konnte im Verlauf der Jahre eine Abnahme der Anzahl an Schülerinnen und Schülern pro Lehrkraft verzeichnet werden. Der Durchschnittswert im Schuljahr 2012/2013 mit ca. 20 Schülerinnen bzw. Schülern pro Lehrkraft (Statistisches Landesamt, 2014) lag allerdings über dem Wert, den die Lehrkräfte im Rahmen der Befragung zum Testverfahren *Online Diagnose* nannten. Der durchschnittliche Anteil betrug laut den Angaben der teilnehmenden Lehrkräfte 18 Schülerinnen und Schüler pro Klasse.

Das Testverfahren *Online Diagnose* war in seiner Konzeption und Anlage geeignet für die individuelle Förderung von Schülerinnen und Schülern. Nach einem Eingangstest, dem sogenannten Haupttest, konnten Fördererfolge in einem Detail- und Nachtest überprüft werden. Zu den Ergebnisrückmeldungen auf unterschiedlicher Aggregationsebene erhielten die Lehrkräfte individuelle Fördermaterialien für jede Schülerin und jeden Schüler. Das Testverfahren *Online Diagnose* endete also nicht mit der Rückmeldung der Ergebnisse, sondern bot den Lehrkräften Fördermaterialien an, die in weiterem Sinne als Handlungsvorschläge verstanden werden konnten, und begleitete dadurch den Einsatz der Förderung. Ein Vergleich mit Ergebnisrückmeldungen aus unterschiedlichen anderen Studien lässt erkennen, dass solche Handlungsvorschläge in Form von Fördermaterialien selten sind. Das Testverfahren unterstützte also die Lehrkräfte bei der Übersetzung theoretischen Wissens in Handlungen (vgl. Rolff, 2002). Dennoch ist der „blinde“ Einsatz der Fördermaterialien ungeeignet. Die vorliegende Teilstudie 2 zeigte, dass einige Lehrkräfte den Zeitaufwand für den Einsatz der Ergebnisrückmeldungen zwar als hoch empfanden, sich dabei jedoch gezielt auf Ergebnisrückmeldungen förderbedürftiger Schülerinnen und Schüler bezogen haben und

dadurch auch die Wirkung auf dieselben als ausgeprägter ansahen. Eine reine Handlungsanleitung bzw. Rückmeldung von Fördermaterialien ist so gesehen ungeeignet, sie muss mit diagnostischem Wissen der Lehrkräfte kombiniert werden, damit die Ergebnisrückmeldungen gezielt selektiert werden können (vgl. hierzu auch Tresch, 2007).

Neben diesen Aspekten muss das Testverfahren *Online Diagnose* weiterhin kritisch beurteilt werden. Eine Gegenüberstellung der Aufgabeninhalte des Haupttests mit den Bildungsstandards der Primarstufe zeigte auf, dass das Testverfahren *Online Diagnose* zum einen nicht die gesamten Bildungsstandards abdeckte und zum anderen die Bildungsstandards ergänzende Aufgabenbereiche beinhaltete (Kapitel 3.3). Außerdem lag die Testschwierigkeit zum Teil auf dem Niveau der Anforderungen der Bildungsstandards an die zweite Jahrgangsstufe der Grundschule anstatt auf dem Level der vorausgesetzten Anforderungen der vierten Jahrgangsstufe der Grundschule. Die starke Zustimmung durch die Lehrkräfte hinsichtlich der Verständlichkeit und Eignung der Testaufgaben im Rahmen der subjektiven Qualitätseinschätzung (Teilstudie 1) sollte durch Kriterien der psychometrischen Qualität überprüft werden. Ein Vergleich der subjektiven Qualitätseinschätzung durch die Lehrkräfte mit objektiven Qualitätskriterien konnte in den vorliegenden Teilstudien nicht gezogen werden und steht daher noch aus. Trotz der positiven Beurteilung der Qualität des Testverfahrens bewerteten die Lehrkräfte den Informationsgewinn aus den Ergebnisrückmeldungen als eher gering (Teilstudie 1). Außerdem stellen die Ergebnisse heraus, dass der Zeitaufwand für die Nutzung der Ergebnisrückmeldungen zu hoch war (Teilstudie 1). Trotz der unterschiedlichen Aggregationsebenen, in denen die Ergebnisrückmeldungen genutzt werden konnten, erschien einem Teil der Lehrkräfte der Zeitaufwand als zu hoch. Eine Beratung der Lehrkräfte zum Umgang bzw. selektiven Umgang mit den Ergebnisrückmeldungen wäre an dieser Stelle nützlich (vgl. Rolff, 2002; Maier 2008b). Die Untersuchung zeigte allerdings auch auf, dass Fortbildungsangebote zum Testverfahren *Online Diagnose* sehr selten genutzt wurden oder selten zur Verfügung standen (Teilstudie 1).

Im Rahmen der elektronisch zurückgemeldeten Ergebnisse standen den Lehrkräften optional verschiedene Vergleichswerte zur Verfügung. Hierbei bot die elektronische Ergebnisrückmeldung einen großen Vorteil. Die Lehrkräfte hatten die Möglichkeit, die Vergleichswerte auf eine deutschlandweite oder baden-württembergische Vergleichsgruppe umzustellen. Ebenso konnten die Schülerleistungen innerhalb der Klasse verglichen werden. Zudem stand nach der Durchführung des Nachttests ein Vergleich mit vorhergehenden Test-

leistungen zur Verfügung. Die elektronischen Ergebnismeldungen kommen der Forderung von Tresch (2007) nach, den Lehrkräften möglichst verschiedene Vergleichswerte zur Verfügung zu stellen. Kritisch muss allerdings angemerkt werden, dass im Rahmen des Testverfahrens *Online Diagnose* keine Adjustierungsverfahren hinsichtlich fairer Vergleiche vorgenommen wurden (z. B. Rheinberg & Fries, 2010; Fiege, 2013).

Die für die vorliegenden Teilstudien verwendeten Operationalisierungen wurden theoretisch anhand des im deutschsprachigen Raum etablierten Zyklusmodells der Rückmeldung erstellt (Helmke & Hosenfeld, 2005; vgl. auch Müller, 2009). Kritisch ist zu vermerken, dass aus einer Vielzahl möglicher Kriterien, welche die Prozessschritte Rezeption, Reflexion, Aktion und Evaluation sowie die individuellen Bedingungen, Bedingungen der Klasse, schulischen und externen Bedingungen repräsentieren, lediglich ein Bruchteil umgesetzt werden konnte. Die Teilstudien umfassten also lediglich einen Ausschnitt des theoretischen Modells. Außerdem muss angemerkt werden, dass zu den Prozessschritten lediglich ein bis drei Skalen eingesetzt werden konnten (Teilstudie 1). Weitere Studien könnten die bisherigen Befunde fundieren und damit eine Fokussierung auf wichtige Kriterien vornehmen. Außerdem könnte zu gegebener Zeit eine Meta-Analyse einen Überblick über die bestehenden empirischen Einzelbefunde verschaffen.

Das Testverfahren *Online Diagnose* spiegelt, wie im Theorieteil beschrieben, den Nutzungsprozess aus dem Zyklusmodell der Rückmeldung wider. Bei einem vollständigen Einsatz der jeweiligen Tests kann der Haupttest als Eingangstest angesehen werden. Die Ergebnismeldungen könnten dem Prozessschritt Rezeption zugeordnet werden, die Fördermaterialien aus dem Testverfahren *Online Diagnose* könnten dem Prozessschritt Aktion zugeordnet werden, der Detailtest dem Prozessschritt Reflexion, und der Nachtest könnte dem Prozessschritt Evaluation zugeordnet werden. Das Testverfahren *Online Diagnose* kann anhand des Zyklusmodells der Rückmeldung als empirisches Modell strukturiert werden. Die deskriptiven Befunde zeigten allerdings auf, dass die Detail- und Nachtests selten genutzt wurden (Teilstudie 1). Hauptsächlich wurde der Haupttest und die Fördermaterialien angewandt. Somit bleiben Aussagen zum Nutzungsprozess vage und können sich lediglich auf die einzelnen Schritte der Nutzung beziehen (Teilstudie 2).

Aus den Ergebnissen zu Teilstudie 3 ließen sich hauptsächlich Kriterien der Akzeptanz und Unterstützung als prädiktiv für die Nutzung von Ergebnismeldungen herausfiltern. Die Ergebnisse lieferten statistisch abgesicherte Bestätigungen bisheriger Befunde.

Zur weiteren Absicherung der Ergebnisse sollten im Studiendesign zukünftiger Studien die im nachfolgenden kritisch diskutierten Punkte berücksichtigt werden.

8.3 Kritische Einordnung des Studiendesigns

Der vorliegende Abschnitt bezieht sich auf das Studiendesign der durchgeführten Teilstudien und schlägt mögliche Verbesserungen des Designs vor. Da im Rahmen dieser Untersuchungen nur die Einschätzung der Lehrkräfte zur Verfügung stand, ist z.B. die Aussagekraft zu den abhängigen Variablen Wirkung auf Lernende und Wirkung auf das Kollegium begrenzt. Die Lehrkräfte schätzten zwar die Sichtweise der Schülerinnen und Schüler sowie ihres Kollegiums ein, dennoch bleiben die Aussagen auf die Wahrnehmung der jeweiligen Lehrkraft beschränkt. Die Erhebung weiterer Daten durch die Befragung von Schülerinnen und Schülern oder des Kollegiums könnte ergänzende Informationen zur Nutzung und Wirkung des Testverfahrens *Online Diagnose* liefern. Ebenso muss an dieser Stelle erwähnt werden, dass auch die Daten über die Lehrkräfte aus Selbstberichten stammen. Selbstberichte waren hierbei geeignet, da es um die subjektive Wahrnehmung der Lehrkräfte ging, dennoch birgt diese Methode die Gefahr einer starken Einseitigkeit. Ein Problem des Selbstberichts ist, dass Merkmale der sozialen Erwünschtheit in die Beantwortung von Fragen und Items hineinspielen (Eid, Gollwitzer & Schmitt, 2011). Abgesehen von den bereits genannten Einschränkungen des Datensatzes, der allein aus Einschätzungen der Lehrkräfte besteht, würde die Kombination aus unterschiedlichen Datenquellen, z.B. Schüler- und Lehrerdaten, weitere methodische Möglichkeiten bieten.

In Bezug auf unterschiedliche Datenquellen wurde im Laufe der theoretischen Darstellung ersichtlich, dass die elektronische Datenerhebung bzw. -nutzung besonders viele Möglichkeiten und Vorteile bietet. Im Rahmen des Testverfahrens *Online Diagnose* wurden z.B. die Antworten über ein Onlineformular direkt erfasst und zentral auf einem Server gespeichert. Ebenso ist z.B. eine interaktive Gestaltung der Datenerhebung und -rückmeldung denkbar (z.B. Brunner & Light, 2008). Praktisch müssen allerdings noch einige Hindernisse beim Einsatz von Informations- und Kommunikationssystemen in der Zusammenarbeit von Datenverarbeitungstechniken, Anwendungsbereichen und Nutzern ausgeräumt werden. Die Evaluation des Testverfahrens *Online Diagnose* war zunächst nicht vorgesehen, somit wurden die Antworten der Schülerinnen und Schüler in einem für diese Zwecke üblichen Aufbau

einer Datenbank erfasst. Solch ein Aufbau einer Datenbank sieht allerdings die Wiederverwendung und für eine Evaluation notwendige Verkettung von Daten nicht generell vor. Dies müsste bereits direkt bei der Speicherung der Daten berücksichtigt werden. Die Nutzung der Schülerdaten für die vorliegende Untersuchung war aufgrund des zu betreibenden Aufwands, die Daten aus der Datenbank zu ziehen und zu verknüpfen, unverhältnismäßig hoch und deshalb nicht möglich. Generell ist die Verknüpfung unterschiedlicher Perspektiven, d. h. die Verknüpfung von Schüler- und Lehrerdaten, anzustreben. In Zukunft bieten die sich weiterentwickelnden technischen Systeme die besten Voraussetzungen dafür.

Trotz der Informationen, die aus den vorliegenden Teilstudien abgeleitet werden konnten, muss die Größe der Stichprobe zu bedenken gegeben werden. Der Rücklauf lag bei 15 % ($N = 150$). Schwippert (2004) berichtet in seiner Rezeptionsstudie im Anschluss an IGLU ebenfalls von einer geringen Rücklaufquote von 20 %. Ebenso berichten Ditton und Arnold (2004) in ihrer Feedbackbefragung im Anschluss an QuaSSU von einer Rücklaufquote von 26 % und in den Folgejahren von 15 %. Grund für die geringen Rücklaufquoten könnte die mangelnde Motivation der Lehrkräfte sein bzw. deren geringe Bereitschaft, nach der Durchführung von Testuntersuchungen nochmals Zeit und Engagement in eine zusätzliche Befragung zu investieren. Im Rahmen des Testverfahrens *Online Diagnose* kann die geringe Rücklaufquote zudem auf Strukturänderungen im Schulsystem zurückgeführt werden. Die Telefonate, die mit den Lehrkräften zur Erinnerung an die Befragung geführt wurden, ließen den folgenden Eindruck entstehen: Die häufigste Begründung für die Nichtteilnahme war die Umstellung auf eine andere Schulform und die dadurch empfundene Überlastung. Außerdem wurde angegeben, dass die durchführende Lehrkraft aufgrund von Mutterschutz, Rente oder Schulwechsel nicht mehr zur Befragung zur Verfügung stehe. Weitere Gründe waren technische Probleme oder die Teilnahme an der Durchführung anderer Evaluationen. Daher muss eingeräumt werden, dass die Repräsentativität der Studie nur beschränkt gegeben ist. Die Ergebnisse liefern jedoch interessante Befunde, an die Folgestudien anschließen können. Durch die Sicherung einer größeren Stichprobe würden sich zudem weitere methodische Möglichkeiten bieten. Unter anderem könnten Moderations- und Mediationseffekte überprüft werden. Diese könnten der Abfolge der Prozessschritte eher gerecht werden, wie z. B. Koch (2011) in ihrer Untersuchung zu indirekten Effekten auf den Prozessschritt Aktion beschreibt. In zukünftigen Studien könnte somit nicht nur die Assozi-

ation der drei Prozessschritte Rezeption, Reflexion und Aktion mit dem Prozessschritt Evaluation überprüft werden (Teilstudie 2), sondern auch der Ablauf bzw. die Reihenfolge der Prozessschritte.

Die Berücksichtigung der Mehrebenenstruktur (Raudenbush & Bryk, 2002), die in der aktuellen Bildungsforschung zum State of the Art gehört, würde in weiteren Untersuchungen gesicherte Erkenntnisse liefern. Zur Berücksichtigung der Mehrebenenstruktur sind Schüler- und Lehrerdaten notwendig sowie eine größere Stichprobe. Durch die elektronische Erfassung von Schülerdaten sollten – unter Berücksichtigung der notwendigen technischen Voraussetzungen (wie oben beschrieben) – große Stichproben kein Problem darstellen.

Die vorliegenden Teilstudien beruhen auf einer Querschnittsuntersuchung. Ein längsschnittlicher Vergleich würde der Konzeption des Testverfahrens *Online Diagnose* allerdings eher gerecht werden. Das Testverfahren sieht einen Haupttest zu Beginn des Schuljahres, einen Detailtest nach einer gewissen Zeit der Förderung und einen Nachtest zur Überprüfung der Leistungsentwicklung vor. Durch ein Prä-/Post-Design könnten Effekte der Nutzung von Ergebnisrückmeldungen und der Fördermaterialien festgehalten werden. Vor allem für die in Teilstudie 2 und 3 eingesetzten Variablen zur Erfassung von Veränderungen bzw. der Wirkung auf Lernende, das Kollegium und Lehrende würde ein Prä-/Post-Design die Aussagekraft deutlich stärken.

Eine weitere Möglichkeit zur Untersuchung der Nutzung von Ergebnisrückmeldungen besteht in der Anwendung von Interventionen. Lehrkräfte könnten im Rahmen einer Fortbildung in der Nutzung von Ergebnisrückmeldungen unterstützt werden. Diese könnte begleitend z.B. zu externen Vergleichsuntersuchungen angeboten werden. Zur Testung des Vorwissens könnte allen Lehrkräften in der ersten Fortbildungssitzung dieselbe Ergebnisrückmeldung zur Leistung einer Schülerin oder eines Schülers zur Bewertung vorgelegt werden. Dies kann in Form eines offenen Antwortformats geschehen und zusätzlich könnten Kriterien der im Theorieteil beschriebenen Anforderungen an Ergebnisrückmeldungen und Lehrkräfte abgefragt werden (z.B. Shah & Hoeffner, 2002). Damit könnte eine Vergleichbarkeit und Einstufung der Kenntnisse zur Nutzung von Ergebnisrückmeldungen von Lehrkräften erreicht werden. Außerdem könnten in solch einem Untersuchungsdesign kausale Zusammenhänge überprüft werden (vgl. Koch, 2011). Zudem zeigte sich in bisherigen Studien, dass die Einbindung der Nutzung von Ergebnisrückmeldungen in ein Weiterbildungs- und Fortbildungskonzept den Umfang der Nutzung erhöhen kann (z.B. Maier, 2008b; Zimmer-Müller, Hosenfeld & Koch, 2014)

8.4 Implikationen für die weitere Forschung und die Praxis

Implikationen für die Rezeptionsforschung. Die Befunde der vorliegenden Arbeit geben Hinweise zur Nutzung von Ergebnisrückmeldungen. Im Speziellen beziehen sich die Ergebnisrückmeldungen auf die Individual- und Förderdiagnostik eines internetgestützten Testverfahrens. Befunde aus bisherigen empirischen Untersuchungen zeigen, dass die Akzeptanz von Rückmeldestudien ein wichtiger Einflussfaktor im Hinblick auf die Nutzung von Ergebnisrückmeldungen ist (z.B. Helmke, 2004; Müller, 2009). Die vorliegende Untersuchung weist ebenfalls darauf hin, dass die Einstellung der Lehrkräfte zu externen Testuntersuchungen ebenso wie die Akzeptanz des Testverfahrens *Online Diagnose* in einem Zusammenhang mit den Prozessschritten der Nutzung steht (Teilstudie 3). Um die Bereitschaft der Lehrkräfte, an Untersuchungen teilzunehmen, zu erhöhen, sollte vor der Durchführung bzw. Rückmeldung der Ergebnisse die Akzeptanz der beteiligten Akteure gewonnen werden (vgl. Helmke, 2004). Denkbar wäre dies z.B. durch Aufklärungsarbeit über den Forschungszweck und den Zweck der geplanten Ergebnisrückmeldung. Außerdem wären beispielsweise Unterstützungsprogramme wie eine Fortbildung oder wissenschaftliche Begleitung des Nutzungsprozesses hilfreich.

Es konnte zudem aufgezeigt werden, dass die Kooperation eher auf informellem Wege innerhalb der Schule stattfand und seltener im Rahmen von z.B. Fachkonferenzen (z.B. Dederling, 2011). Die Untersuchung zum Testverfahren *Online Diagnose* wies allerdings auf den Zusammenhang zwischen der durch die Lehrkraft wahrgenommenen Unterstützung (beispielsweise durch Kooperationen) und der Wirkung des Testverfahrens hin (Teilstudie 3). Ein einheitliches Testinstrument über die Jahrgangsstufen hinweg, das lediglich die Anforderungen an die Testleistung variiert, kann die gegenseitige Unterstützung bei der Interpretation der Ergebnisse im Lehrerkollegium (auch in einzügigen Schulen) erhöhen. Ein solches Instrument, das für mehrere Jahrgangsstufen gleich aufgebaut ist und in dem sich die Lehrkräfte auskennen, erhöht auch die Kapazität der Aufmerksamkeit für die eigentlichen Ergebnisrückmeldungen. Außerdem könnten im Schulalltag fest vereinbarte Besprechungszeiten über Ergebnisrückmeldungen aus Testverfahren den Austausch erhöhen.

Weitere Befunde aus der Forschung zeigen auf, dass Anschlussaktivitäten an die zurückgemeldeten Ergebnisse bislang eher selten stattfinden (z.B. Dederling, 2011). Hierfür bietet das Testverfahren *Online Diagnose* Fördermaterialien an. Die vorliegende Untersuchung stellte den Zusammenhang zwischen einer verstärkten Anwendung von Fördermaßnahmen und der Wirkung des Testverfahrens heraus (Teilstudie 2). Die Bereitstellung von

Förderaufgaben bzw. von Beispielaufgaben kann für Lehrkräfte bei der Einführung von Anschlussaktivitäten hilfreich sein. Allerdings kann hierbei nicht im Sinne von Rolff (2002) von einer Übersetzung wissenschaftlichen Wissens in Handlungswissen ausgegangen werden. Inwiefern seitens der Lehrkräfte die Informationen aus den Ergebnisrückmeldungen in einen Bezug zu den Fördermaterialien gebracht wurden, konnte anhand der vorliegenden drei Teilstudien nicht untersucht werden. Dennoch führte die Durchführung des Testverfahrens *Online Diagnose* häufiger zu Konsequenzen bzw. zum Einsatz von Fördermaterialien, als dies bei Rückmeldungen aus vergleichbaren Untersuchungen der Fall war. Die Bereitstellung von Fördermaterialien erweist sich so als sinnvoll für zukünftige Ergebnisrückmeldungen. Optimal wäre eine zusätzliche Einbindung der Rückmeldungen in ein Fortbildungskonzept, das Lehrkräfte bei der Nutzung der Ergebnisse und dem Einsatz der Fördermaterialien unterstützt (z. B. Maier 2008b).

Anhand der Ergebnisse wurde deutlich, dass ein externes Testverfahren die Diagnosekompetenz von Lehrkräfte unterstützen oder ergänzen, allerdings nicht ersetzen kann. In Teilstudie 3 konnte ein Zusammenhang zwischen der allgemeinen Einschätzung der Diagnosekompetenz durch die Lehrkräfte und dem selbstwirksamen Umgang mit den Ergebnisrückmeldungen aufgezeigt werden. Ebenso wurde deutlich, dass besonders die Lehrkräfte, die ihre Diagnosekompetenz als geringer einschätzten, einen Informationsgewinn aus den Ergebnisrückmeldungen erhielten. Allerdings konnte Teilstudie 2 aufzeigen, dass eine selektive Nutzung von Ergebnisrückmeldungen, also ein gezielter Zeitaufwand für die Nutzung der Ergebnisrückmeldungen, in der Ansicht der Lehrkräfte eine stärkere Wirkung auf Lernende hat. Lehrkräfte benötigen diagnostische Kompetenzen, um die Ergebnisrückmeldungen selektiv bzw. gezielt verwenden zu können. Das bedeutet für die Rezeptionsforschung, dass bei zukünftigen Untersuchungen die Diagnosekompetenz der Lehrkräfte berücksichtigt und als mögliche Kovariate einbezogen werden sollte.

In Teilstudie 1 wurde deutlich, dass hauptsächlich der Haupttest und die Fördermaterialien aus dem Testverfahren *Online Diagnose* genutzt wurden. Auch wurde insgesamt mehr Zeit für die Durchführung des Haupttests sowie für den Einsatz der Fördermaterialien aufgewendet als für den Detail- und Nachttest. Deutlich wurde, dass der gesamte Zeitaufwand für die Durchführung und den Einsatz der Materialien im Durchschnitt 13 Schulstunden entsprach. Hinsichtlich des Durchschnittswerts an Zeitaufwand kann vermutet werden, dass keine zusätzliche Unterrichtszeit für den Einsatz der Detail- und Nachttests zur Verfügung stand, weshalb diese Möglichkeiten eher selten genutzt wurden. Zukünftige Planungen

und Beschlüsse zur Implementierung von Erhebungsverfahren sollten den gesamten Testablauf und die dafür aufzuwendende Zeit fest (in den Lehrplan) integrieren.

Implikationen für die Erhebung und Nutzung elektronischer Ergebnisrückmeldungen. Die Ergebnisse zeigen, dass der Einsatz elektronischer Instrumente und Ergebnisrückmeldungen für die Individual- und Förderdiagnostik genutzt werden kann. Das Testverfahren *Online Diagnose* fungiert als Werkzeug für die Erfassung, Auswertung und Rückmeldung von Schülerleistungen. Es hilft damit bei der datengestützten Entscheidungsfindung in Bezug auf die individuelle bzw. gezielte Förderung (Breiter, 2008). Die Technologie bietet der zukünftigen Konzeption individueller bzw. gezielter Förderung im Sinne eines formativen Assessments (z.B. Black & Wiliam, 1998) nützliche Möglichkeiten. Auch die Bereitstellung von unterschiedlichen Vergleichswerten (z.B. faire Vergleiche), die optional genutzt werden können, wird durch die Technik erleichtert. Allerdings erfordert die Zunahme hinsichtlich der Quantität von Daten und deren Komplexität entsprechende Kompetenzen der Lehrkräfte, um die Daten lesen und interpretieren zu können (Breiter, 2008). Vor allem der selektive Umgang mit der Datenmenge erfordert wiederum diagnostische Kompetenzen. Lehrkräfte müssen in der gezielten Suche und Beantwortung eines Erkenntnisinteresses hinsichtlich der Leistung einer Schülerin oder eines Schülers kompetent sein. Dies zeigte sich im Rahmen der Teilstudie 2: Lehrkräfte die den Zeitaufwand für die Nutzung der Ergebnisrückmeldungen als hoch einschätzten, dadurch diese allerdings gezielt nutzten, gaben gleichzeitig eine stärkere Wirkung des Testverfahrens *Online Diagnose* auf Lernende an. Für die Nutzung elektronischer Ergebnisrückmeldungen bedeutet dies, dass Lehrkräfte nicht nur in der Interpretation von Ergebnissen geschult werden müssen, sondern auch im selektiven Umgang mit einer großen Datenmenge.

Die sich weiter entwickelnden Informations- und Kommunikationssysteme und die damit verbundene Erfassung und Speicherung schülerbezogener Daten erfordern ein umfassendes und landesweit möglichst einheitliches Konzept des Datenschutzes. In Zukunft wird der Datenschutz hinsichtlich der datengestützten Diagnostik und Förderung ein unumgängliches Thema darstellen (z.B. Stauke, 2008).

8.5 Fazit

Als Fazit aus der vorliegenden Untersuchung wird deutlich, dass die Nutzung von Ergebnisrückmeldungen eine komplexe Aufgabe darstellt. Lehrkräfte nehmen darin eine entscheidende Rolle als Übermittler der Ergebnisse ein. Diese Vermittlungsprozesse werden wiederum von Faktoren wie z.B. die Unterstützung der Lehrkräfte bei der Nutzung von Ergebnisrückmeldungen beeinflusst. Die Rezeptionsforschung hat hierbei weiterhin den Auftrag, die Vielzahl der Einflussfaktoren auf die Nutzung von Ergebnisrückmeldungen zu benennen und zu konkretisieren. Der Beschluss der Kultusministerkonferenz der Länder, die Diagnosekompetenz von Lehrkräften durch die Aus- und Weiterbildung zu verbessern, scheint auch auf dem Hintergrund der vorliegenden Untersuchung eine wichtige Unternehmung zu sein. Zur Nutzung von Ergebnisrückmeldungen z.B. aus externen Evaluationen benötigen die Lehrkräfte diagnostische Kompetenzen. Diese können durch objektivierte Informationen ergänzt, jedoch nicht ersetzt werden.

Weiterhin kann die Erhebung und Nutzung elektronischer Ergebnisrückmeldungen in vielerlei Hinsicht eine Erleichterung bedeuten. Die direkte elektronische Erfassung und Speicherung von Antworten (z. B. bei Leistungstests von Schülerinnen und Schülern) erspart die aufwendige Datenerfassung bzw. -eingabe. Elektronische Ergebnisrückmeldung kann allgemeine Anforderungen an Ergebnisrückmeldungen erfüllen, wie beispielsweise die zeitnahe Rückmeldung der Ergebnisse oder die Bereitstellung unterschiedlicher Vergleichswerte. Als Fazit kann aus der vorliegenden Untersuchung geschlossen werden, dass die Weiterentwicklung von Informations- und Kommunikationssystemen in Bezug auf die Kombination von Datenverarbeitungstechniken, Anwendungsbereichen und Nutzerverhalten bedeutend ist. Eine verbesserte Datenverarbeitung und steigende Kompetenzen der Lehrkräfte im Umgang mit den Anwendungsbereichen und allgemein im Umgang mit Computern kann die generelle Anwendungseffizienz elektronischer Systeme erhöhen. Dadurch könnten beispielsweise mit geringerem Aufwand mehr Schülerinnen und Schüler im Rahmen einer Studie befragt werden. Ein größerer Datensatz hätte wiederum eine beträchtlichere Auswahl an passenden statistischen Auswertungsmethoden zur Folge.

9 Literatur

- Ahnert, L. (Hrsg.). (2014). *Theorien in der Entwicklungspsychologie*. Berlin, Heidelberg: Springer.
- Altrichter, H. & Maag Merki, K. (Hrsg.). (2010). *Educational governance: Vol. 7. Handbuch neue Steuerung im Schulsystem*. Wiesbaden: VS, Verl. für Sozialwiss.
- Arnold, K.-H. (2002). Schulentwicklung durch Rückmeldung der Lernwirksamkeit an die Einzelschule: Möglichkeiten und Grenzen der Schuleffizienzforschung. *Zeitschrift für Pädagogik*, 48(5), 741–764.
- Atkinson, J. W. (1957). Motivational Determinants of Risk-Taking Behavior. *Psychological Review*, 64(6), 359–372.
- Bangert-Drowns, R. L., Kulik, C.-L. C., Kulik, J. A., & Morgan Mary Teresa. (1991). The instructional effect of feedback in test-like events. *Review of educational research*, 61(2), 213–238.
- Bar-Tal, D. & Kruglanski, A. W. (Hrsg.). (1988). *The Social psychology of knowledge*. Cambridge, New York, Paris: Cambridge University; Editions de la maison des sciences de l'homme.
- Baumert, J., Blum, W., Brunner, M., Dubberke, T., Jordan, A., Klusmann, U., . . . Tsai Yi-Miau. (2009). *Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung von mathematischer Kompetenz (COACTIV): Dokumentation der Erhebungsinstrumente. Materialien aus der Bildungsforschung: Vol. 83*. Berlin: Max-Planck-Inst. für Bildungsforschung.
- Baumert, J. & Kunter, M. (2011). Das Kompetenzmodell von COAKTIV. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (pp. 29–53). Münster, München [u.a.]: Waxmann.
- Bildungshaus Schulbuchverlage Westermann Schroedel Diesterweg Schöningh Winklers GmbH. (n.d.). *Bedienungshinweise*. Zugriff am 02.09.2014, von <http://www.onlinediagnose.de/index.php?page=hilfe>
- Bildungshaus Schulbuchverlage Westermann Schroedel Diesterweg Schöningh Winklers GmbH. (n.d.). *FAQ: Fragen zur Benutzung*. Zugriff am 09.01.2014, von <https://www.onlinediagnose.de/index.php?page=faq>

- Bildungshaus Schulbuchverlage Westermann Schroedel Diesterweg Schöningh Winklers GmbH. (2008). *Konzeption Online-Diagnose*.
- Bildungshaus Schulbuchverlage Westermann Schroedel Diesterweg Schöningh Winklers GmbH. (2010). *Datenschutzhinweis*.
- Bildungshaus Schulbuchverlage Westermann Schroedel Diesterweg Schöningh Winklers GmbH. (2012). *Erste Schritte mit der Online-Diagnose*. von http://www.onlinediagnose.de/downloads/Kurzanleitung_Online-Diagnose.pdf
- Bildungshaus Schulbuchverlage Westermann Schroedel Diesterweg Schöningh Winklers GmbH. (2013). *Konzeption Online-Diagnose*. von http://www.onlinediagnose.de/downloads/Konzeption_Online-Diagnose_2013.pdf
- Bildungsplan der Grundschule. (2004). Zugriff am 27.02.2015, von http://www.bildungstaerkt-menschen.de/service/downloads/Bildungsplaene/Grundschule/Grundschule_Bildungsplan_Gesamt.pdf
- Black, P. & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education*, 5(1), 7–74.
- Black, P. & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31. doi:10.1007/s11092-008-9068-5
- Bohl, T. & Kiper, H. (Hrsg.). (2009). *Lernen aus Evaluationsergebnissen: Verbesserungen planen und implementieren*. Bad Heilbrunn: Klinkhardt.
- Bonsen, M., Büchter, A. & Peek, R. (2006). Datengestützte Schul- und Unterrichtsentwicklung: Bewertungen der Lernstandserhebungen in NRW durch Lehrerinnen und Lehrer. In W. Bos, H. G. Holtappels, H. Pfeiffer, H.-G. Rolff & R. Schulz-Zander (Hrsg.), *Eine Veröffentlichung der Arbeitsstelle für Schulentwicklungsforschung der Universität Dortmund. Jahrbuch der Schulentwicklung. Daten, Beispiele und Perspektiven* (pp. 125–148). Weinheim [u.a.]: Juventa Verl.
- Bos, W., Holtappels, H. G., Pfeiffer, H., Rolff, H.-G. & Schulz-Zander, R. (Hrsg.). (2006). *Eine Veröffentlichung der Arbeitsstelle für Schulentwicklungsforschung der Universität Dortmund. Jahrbuch der Schulentwicklung: Daten, Beispiele und Perspektiven*. Weinheim [u.a.]: Juventa Verl.
- Brägger, G., Bucher, B. & Landwehr, N. (Hrsg.). (2005). *Pädagogik. Schlüsselfragen zur externen Schulevaluation* (1. Aufl). Bern: Hep.

- Breiter, A. (2008). Integrating School Management Information System. In A. Breiter, A. Lange, & E. Stauke (Hrsg.), *School information systems and data-based decision-making* (1st ed., pp. 11–20). Frankfurt am Main, New York: P. Lang.
- Breiter, A., Lange, A., & Stauke, E. (Hrsg.). (2008). *School information systems and data-based decision-making* (1. Aufl). Frankfurt am Main, New York: P. Lang.
- Breiter, A. & Light, D. (2006). Data for School Improvement: Factors for designing effective information systems to support decision-making in schools. *Educational Technology & Society*, 9(3), 206–217.
- Breiter, A. & Stauke, E. (2007). Anforderungen an elektronische Rückmeldesysteme aus Nutzersicht. *Empirische Pädagogik*, 21(4), 383–400.
- Breiter, A. & Wind, M. (Hrsg.). (2011). *Kritische Informatik: Bd. 6. Informationstechnik und ihre Organisationslücken: Soziale, politische und rechtliche Dimensionen aus der Sicht von Wissenschaft und Praxis*. Berlin, Münster: Lit.
- Bromme, R. & Kienhues, D. (2012). Rezeption von Wissenschaft. Mit besonderem Fokus auf Bio- und Gentechnologie und konfligierende Evidenz. In : *SpringerLink : Bücher, Biotechnologie-Kommunikation. Kontroversen, Analysen, Aktivitäten*, pp. 303–348. Berlin, Heidelberg: Springer Vieweg.
- Bromme, R., Prenzel, M. & Jäger, M. (2014). Empirische Bildungsforschung und evidenzbasierte Bildungspolitik. *Zeitschrift für Erziehungswissenschaft*, 17(S4), 3–54. doi:10.1007/s11618-014-0514-5
- Brunner, M., Anders, Y., Hachfeld, A. & Krauss, S. (2011). Diagnostische Fähigkeiten von Mathematiklehrkräften. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (pp. 215–234). Münster, München [u.a.]: Waxmann.
- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion* (3., aktualisierte und erw. Aufl). *Psychologie*. München, Boston [u.a.]: Pearson Studium.
- Bühner, M., Schmidt-Atzert, L., Grieshaber, E. & Lux, A. (2001). Faktorenstruktur verschiedener neuropsychologischer Tests. *Zeitschrift für Neuropsychologie*, 12(3), 181–187. doi:10.1024//1016-264X.12.3.181
- Butler, D. L. & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of educational research*, 65(3), 245–281.

- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed). Mahwah, N.J.: L. Erlbaum Associates.
- Cramer, C. & Friedrich, A. (in Vorbereitung). Beanspruchung im Lehrerberuf: Ein neues, integratives Rahmenmodell.
- Dederling, K. (2011). Hat Feedback eine positive Wirkung?: Zur Verarbeitung extern erhobener Leistungsdaten in Schulen. *Unterrichtswissenschaft*, 39(1), 63–82.
- Dicke, A.-L., Lüdtke, O., Trautwein, U., Nagy, G., & Nagy, N. (2012). Judging students' achievement goal orientations: Are teacher ratings accurate? *Learning and Individual Differences*, 22(6), 844–849. doi:10.1016/j.lindif.2012.04.004
- Ditton, H. & Arnoldt, B. (2004). Schülerbefragungen zum Fachunterricht: Feedback an Lehrkräfte. *Empirische Pädagogik*, 18(1), 115–139.
- Ditton, H. & Merz, D. (2000). *Qualität von Schule und Unterricht: Kurzbericht über erste Ergebnisse einer Untersuchung an bayerischen Schulen*. von <http://www.quassu.net/Bericht1.pdf>
- Ditton, H. & Müller, A. (Hrsg.). (2014). *Feedback und Rückmeldungen: Theoretische Grundlagen, empirische Befunde, praktische Anwendungsfelder*. Münster, New York, NY: Waxmann.
- Döbert, H. & Fuchs, H.-W. (Hrsg.). (2005). *Studien zur international und interkulturell vergleichenden Erziehungswissenschaft: Bd. 6. Leistungsmessungen und Innovationsstrategien in Schulsystemen: Ein internationaler Vergleich*. Münster, New York, München, Berlin: Waxmann.
- Driscoll, M. P., Jonassen, D. H., & Harris, P. (Hrsg.). (2001). *AECT Series, 2. Handbook of Research for Educational Communications and Technology: A Project of the Association for Educational Communications and Technology* (3rd ed). Hoboken: Taylor & Francis.
- Eccles, J. (2005). Subjektive Task Value and the Eccles et al. Model of Achievement-Related Choices. In A. J. Elliot & C. S. Dweck (Hrsg.), *Handbook of competence and motivation* (pp. 105–121). New York: Guilford Press.
- Eccles, J., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). Expectancies, Values, and Academic Behaviors. In J. T. Spence (Hrsg.),

- A Series of books in psychology. Achievement and achievement motives. Psychological and sociological approaches* (pp. 75–146). San Francisco: W.H. Freeman.
- Edelstein, W. (2002). Selbstwirksamkeit, Innovation und Schulreform: Zur Diagnose der Situation. In M. Jerusalem & D. Hopf (Hrsg.), *Zeitschrift für Pädagogik : Beiheft: Vol. 44. Selbstwirksamkeit und Motivationsprozesse in Bildungsinstitutionen* (pp. 13–27). Weinheim [u.a.]: Beltz.
- Eid, M., Gollwitzer, M. & Schmitt, M. (2010). *Statistik und Forschungsmethoden: Lehrbuch ; mit Online-Materialien* (1. Aufl). Weinheim [u.a.]: Beltz.
- Elliot, A. J. & Dweck, C. S. (Hrsg.). (2005). *Handbook of competence and motivation*. New York: Guilford Press.
- Enders, C. K. (2010). *Applied missing data analysis. Methodology in the social sciences*. New York: Guilford Press.
- Fiege, C. (2013). *Faire Vergleiche in der Schulleistungsforschung. Methodologische Grundlagen und Anwendung auf Vergleichsarbeiten* (Dissertation). Friedrich-Schiller-Universität Jena, Jena.
- Fiege, C., Reuther, F. & Nachtigall, C. (2011). Faire Vergleiche?: Berücksichtigung von Kontextbedingungen des Lernens beim Vergleich von Testergebnissen aus deutschen Vergleichsarbeiten. *Zeitschrift für Bildungsforschung*, 1(2), 133–149. doi:10.1007/s35834-011-0009-x
- Gagné, M. (Hrsg.). (2014). *Oxford library of psychology. The Oxford handbook of work engagement, motivation, and self-determination theory*. New York: Oxford University Press.
- Gagné, M. & Deci, E. L. (2014). The History of Self-Determination Theory in Psychology and. The Oxford Handbook of Work Engagement, Motivation, and Self-Determination Theory. In M. Gagné (Hrsg.), *Oxford library of psychology. The Oxford handbook of work engagement, motivation, and self-determination theory* (pp. 1–9). New York: Oxford University Press.
- Geiser, C. (2011). *Datenanalyse mit Mplus: Eine anwendungsorientierte Einführung* (2nd ed.). *Lehrbuch*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Gluchowski, P., Gabriel, R. & Dittmar, C. (2008). *Management-Support-Systeme und Business intelligence: Computergestützte Informationssysteme für Fach- und Führungskräfte* (2nd ed.). Berlin, Heidelberg: Springer.

- Gnambs, T., Batinic, B. & Hertel, G. (2011). 10. Kapitel: Internetbasierte psychologische Diagnostik. In L. F. Hornke, M. Amelang & M. Kersting (Hrsg.), *Enzyklopädie der Psychologie : Themenbereich B, Methodologie und Methoden : Serie 2, Psychologische Diagnostik: Vol. 3. Leistungs-, Intelligenz- und Verhaltensdiagnostik* (pp. 447–498). Göttingen [u.a.]: Hogrefe.
- Groß Ophoff, J. (2013). *Lernstandserhebungen: Reflexion und Nutzung* (1. Aufl). Münster: Waxmann Verlag GmbH.
- Groß Ophoff, J., Koch, U., Hosenfeld, I. & Helmke, A. (2006). Ergebnissrückmeldungen und ihre Rezeption im Projekt VERA. In H. Kuper (Hrsg.), *Rückmeldung und Rezeption von Forschungsergebnissen. Zur Verwendung wissenschaftlichen Wissens im Bildungssystem* (pp. 19–40). Münster, New York, München, Berlin: Waxmann.
- Haddock, G. & Maio, G. R. (2014). Einstellungen. In K. Jonas, W. Stroebe & M. Hewstone (Hrsg.), *Springer-Lehrbuch. Sozialpsychologie* (6th ed., pp. 198–228). Berlin: Springer Gabler.
- Hartung-Beck, V. (2009). *Schulische Organisationsentwicklung und Professionalisierung: Folgen von Lernstandserhebung an Gesamtschulen* (1. Aufl). Wiesbaden: VS, Verl. für Sozialwiss.
- Hattie, J. & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77(1), 81–112. doi:10.3102/003465430298487
- Hattie, J. & Wollenschläger, M. (2014). A conceptualization of feedback. In H. Ditton & A. Müller (Hrsg.), *Feedback und Rückmeldungen. Theoretische Grundlagen, empirische Befunde, praktische Anwendungsfelder* (pp. 135–149). Münster, New York, NY: Waxmann.
- Helmke, A. (2004). Von der Evaluation zur Innovation: Pädagogische Nutzbarmachung von Vergleichsarbeiten in der Grundschule. *Das Seminar*, 10(2), 90–112.
- Helmke, A. & Hosenfeld, I. Vergleichsarbeiten - Standards - Kompetenzstufen: Begriffliche Klärung und Perspektiven für VERA.
- Helmke, A. & Hosenfeld, I. (2005). Standardbezogene Unterrichtsevaluation. In G. Brägger, B. Bucher & N. Landwehr (Hrsg.), *Pädagogik. Schlüsselfragen zur externen Schullevaluation* (1st ed.). Bern: Hep.

- Hornke, L. F., Amelang, M. & Kersting, M. (Hrsg.). (2011). *Enzyklopädie der Psychologie : Themenbereich B, Methodologie und Methoden : Serie 2, Psychologische Diagnostik: Vol. 3. Leistungs-, Intelligenz- und Verhaltensdiagnostik* ([Vollst. Neuausg.]). Göttingen [u.a.]: Hogrefe.
- Hosenfeld, A. (2010). *Führt Unterrichtsrückmeldung zu Unterrichtsentwicklung?: Die Wirkung von videographischer und schriftlicher Rückmeldung bei Lehrkräften der vierten Jahrgangsstufe. Empirische Erziehungswissenschaft: Vol. 22*. Münster, München [u.a.]: Waxmann.
- Hosenfeld, I. & Groß Ophoff, J. (2007). Nutzung und Nutzen von Evaluationsstudien in Schule und Unterricht. *Empirische Pädagogik*, 21(4).
- Hosenfeld, I., Koch, U., Groß Ophoff, J. & Scherthan, F. (2008). Projekt VERA: Ergebnisorientierte Unterrichtsentwicklung durch internetgestützte externe Evaluation? In A. Breiter, A. Lange & E. Stauke (Hrsg.), *School information systems and data-based decision-making* (1st ed., pp. 49–59). Frankfurt am Main, New York: P. Lang.
- Hu, L. & Bentler, P. M. (1999). Cutoff criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. doi:10.1080/10705519909540118
- Institut zur Qualitätsentwicklung im Bildungswesen. (n.d.). *VERA: Ein Überblick*. Zugriff am 11.11.2014, von <https://www.iqb.hu-berlin.de/vera>
- Institut zur Qualitätsentwicklung im Bildungswesen (IQB). (n.d.). *Bildungsstandards*. Zugriff am 11.11.2014, von <https://www.iqb.hu-berlin.de/bista>
- Institut zur Qualitätsentwicklung im Bildungswesen (IQB). (n.d.). *FAQ: Häufig gestellte Fragen*. Zugriff am 11.11.2014, von <https://www.iqb.hu-berlin.de/vera/faq>
- Institut zur Qualitätsentwicklung im Bildungswesen (IQB). (n.d.). *Ländervergleich*. Zugriff am 03.11.2014, von <https://www.iqb.hu-berlin.de/laendervergleich>
- Institut zur Qualitätsentwicklung im Bildungswesen (IQB). (n.d.). *Überprüfen von Bildungsstandards: Ländervergleich*. Zugriff am 11.11.2014, von <https://www.iqb.hu-berlin.de/bista/control>
- Jerusalem, M. & Hopf, D. (Hrsg.). (2002). *Zeitschrift für Pädagogik : Beiheft: Vol. 44. Selbstwirksamkeit und Motivationsprozesse in Bildungsinstitutionen*. Weinheim [u.a.]: Beltz.

- Jeuk, S. (2009). Aktuelle Verfahren zur Einschätzung des Stands der Sprachaneignung bei mehrsprachigen Kindern im Grundschulalter. In S. Jeuk & I. Schmid-Barkow (Hrsg.), *Differenzen diagnostizieren und Kompetenzen fördern im Deutschunterricht* (pp. 61–81). Freiburg im Breisgau: Fillibach.
- Jeuk, S. (2013). *Überlegungen zur Online Diagnose Baden-Württemberg: Unveröffentlichtes Manuskript*. Institut für Sprachen, Pädagogische Hochschule Ludwigsburg, Deutschland.
- Jeuk, S. & Schmid-Barkow, I. (Hrsg.). (2009). *Differenzen diagnostizieren und Kompetenzen fördern im Deutschunterricht*. Freiburg im Breisgau: Fillibach.
- Jeynes, W. H. (2007). The Relationship Between Parental Involvement and Urban Secondary School Student Academic Achievement: A Meta-Analysis. *Urban Education*, 42(1), 82–110. doi:10.1177/0042085906293818
- Jonas, K., Stroebe, W. & Hewstone, M. (Hrsg.). (2014). *Springer-Lehrbuch. Sozialpsychologie* (6., vollst. überarb. Aufl). Berlin: Springer Gabler.
- Jonkmann, K., Rose, N. & Trautwein, U. (2013). *Tradition und Innovation: Entwicklungsverläufe an Haupt- und Realschulen in Baden-Württemberg und Mittelschulen in Sachsen*. Abschlussbericht für die Länder Baden-Württemberg und Sachsen.
- Karing, C. (2009). Diagnostische Kompetenz von Grundschul- und Gymnasiallehrkräften im Leistungsbereich und im Bereich Interessen. *Zeitschrift für Pädagogische Psychologie*, 23(34), 197–209. doi:10.1024/1010-0652.23.34.197
- Klieme, E. (2004). Begründung, Implementation und Wirkung von Bildungsstandards: Aktuelle Diskussionslinien und empirische Befunde. *Zeitschrift für Pädagogik*, 50(5), 625–634.
- Klieme, E. & Warwas, J. (2011). Konzepte der Individuellen Förderung. *Zeitschrift für Pädagogik*, 57.(6), 805–818.
- Kluger, A. N. & DeNisi, A. (1996). The Effects of Feedback Interventions on Performance: A Historical Review, a Meta-Analysis, and a Preliminary Feedback Intervention Theory. *Psychological Bulletin*, 119(2), 254–284.
- Klusmann, U., Kunter, M., Trautwein, U. & Baumert, J. (2006). Lehrerbelastung und Unterrichtsqualität aus der Perspektive von Lehrenden und Lernenden. *Zeitschrift für Pädagogische Psychologie*, 20(3), 161–173. doi:10.1024/1010-0652.20.3.161

- Koch, U. (2011). *Verstehen Lehrkräfte Rückmeldungen aus Vergleichsarbeiten?: Datenkompetenz von Lehrkräften und die Nutzung von Ergebnismrückmeldungen aus Vergleichsarbeiten 2011. Empirische Erziehungswissenschaft: Vol. 31.* Münster: Waxmann.
- Kohler, B. (2004). Zur Rezeption externer Evaluation durch Lehrkräfte, Eltern sowie Beamte der Schulaufsicht. *Empirische Pädagogik*, 18(1), 18–39.
- Kohler, B. (2005). *Rezeption internationaler Schulleistungsstudien: Wie gehen Lehrkräfte, Eltern und die Schulaufsicht mit Ergebnissen schulischer Evaluationsstudien um? Pädagogische Psychologie und Entwicklungspsychologie: Vol. 46.* Münster, New York, München, Berlin: Waxmann.
- Kohler, B. (2009). Umgang von Lehrer/innen, Eltern und Schulaufsicht mit Ergebnissen internationaler Schulleistungsstudien. In T. Bohl & H. Kiper (Hrsg.), *Lernen aus Evaluationsergebnissen. Verbesserungen planen und implementieren* (pp. 81–96). Bad Heilbrunn: Klinkhardt.
- Kopp, B. & Mandl, H. (2014). Aspekte der Feedbacknachricht. In H. Ditton & A. Müller (Hrsg.), *Feedback und Rückmeldungen. Theoretische Grundlagen, empirische Befunde, praktische Anwendungsfelder* (pp. 151–162). Münster, New York, NY: Waxmann.
- Krapp, A. (2002). Structural and dynamic aspects of interest development: Theoretical considerations from an ontogenetic perspective. *Learning and Instruction*, 12(4), 383–409.
- Krapp, A. & Hascher, T. (2014). Die Erforschung menschlicher Motivation. In L. Ahnert (Hrsg.), *Theorien in der Entwicklungspsychologie* (pp. 234–251). Berlin, Heidelberg: Springer.
- Krapp, A. & Ryan, R. M. (2002). Selbstwirksamkeit und Lernmotivation. Eine kritische Betrachtung der Theorie von Bandura aus der Sicht der Selbstbestimmungstheorie und der pädagogisch-psychologischen Interessentheorie. In M. Jerusalem & D. Hopf (Hrsg.), *Zeitschrift für Pädagogik : Beiheft: Vol. 44. Selbstwirksamkeit und Motivationsprozesse in Bildungsinstitutionen* (pp. 54–82). Weinheim [u.a.]: Beltz.
- Krolak-Schwerdt, S., Böhmer, M. & Gräsel, C. (2009). Verarbeitung von schülerbezogener Information als zielgeleiteter Prozess. *Zeitschrift für Pädagogische Psychologie*, 23(34), 175–186. doi:10.1024/1010-0652.23.34.175

- Kühle, B. (2010). *Zentrale Lernstandserhebungen - ergebnisorientierte Unterrichtsentwicklung?: Schulische Strategien beim Umgang mit Ergebnissen aus den Schulrückmeldungen im Kontext der ersten Lernstandserhebungen 2004/2005 in Nordrhein-Westfalen* (1st ed.). D 465. Berlin: Köster.
- Kühle, B. & Peek, R. (2007). Lernstandserhebungen in Nordrhein-Westfalen: Evaluationsbefunde zur Rezeption und zum Umgang mit Ergebnisrückmeldungen in Schulen. *Empirische Pädagogik*, 21(4), 428–447.
- Kühle, B. & van Ackeren, I. (2012). Wirkungen externer Evaluationsformen für eine evidenzbasierte Schul- und Unterrichtsentwicklung. In M. Ratermann & S. Stöbe-Blossey (Hrsg.), *Governance von Schulund Elementarbildung* (pp. 45–62). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Kulhavy, R. W. & Stock, W. A. (1989). Feedback in written instruction: The place of response certitude. *Educational Psychology Review*, 1(4), 279–308.
- Kultus und Unterricht. (2014). *Amtlicher Teil: Verwaltungsvorschrift des Kultusministeriums zu den zentral angefertigten Lernstandserhebungen* (No. 17).
- Kultusministerkonferenz (KMK). (2004b). *Bildungsstandards der Kultusministerkonferenz: Erläuterungen zur Konzeption und Entwicklung*. Zugriff am 11.11.2014, von http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Bildungsstandards-Konzeption-Entwicklung.pdf
- Kultusministerkonferenz (KMK). (2004a). *Bildungsstandards im Fach Deutsch Bildungsstandards im Fach Deutsch für den Primarbereich (Jahrgangsstufe 4)*. Zugriff am 11.11.2014, von http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_10_15-Bildungsstandards-Deutsch-Primar.pdf
- Kultusministerkonferenz (KMK). (n.d.). *Ländervergleiche: Zentrale Überprüfung des Erreichens der Bildungsstandards im Ländervergleich*. Zugriff am 03.11.2014, von <http://www.kmk.org/bildung-schule/qualitaetssicherung-in-schulen/bildungsmonitoring/laendervergleiche.html>
- Kultusministerkonferenz (KMK). (n.d.). *Vergleichsarbeiten: Vergleichsarbeiten zur landesweiten Überprüfung der Leistungsfähigkeit aller Schulen*. Zugriff am 18.08.2014, von <http://www.kmk.org/bildung-schule/qualitaetssicherung-in-schulen/bildungsmonitoring/vergleichsarbeiten.html>

- Kultusministerkonferenz (KMK). (1997). *Grundsätzliche Überlegungen zu Leistungsvergleichen innerhalb der Bundesrepublik Deutschland: Konstanzer Beschluss*. Zugriff am 11.11.2014, von http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/1997/1997_10_24-Konstanzer-Beschluss.pdf
- Kultusministerkonferenz (KMK) & Institut zur Qualitätsentwicklung im Bildungswesen (IQB). (2006). *Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring*. Zugriff am 11.11.2014, von http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2006/2006_01_01-Gesamtstrategie-Endf.pdf
- Kultusministerkonferenz (KMK) & Institut zur Qualitätsentwicklung im Bildungswesen (IQB). (2010). *Konzeption der Kultusministerkonferenz zur Nutzung der Bildungsstandards für die Unterrichtsentwicklung*. Zugriff am 12.12. 2014, von http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2010/2010_00_00-Konzeption-Bildungsstandards.pdf
- Kultusportal Baden-Württemberg. (n.d.). *Onlinebasierte Lernstandserhebung*. Zugriff am 01.09.2014, von <http://www.kultusportal-bw.de/Lde/770225>
- Kultusportal Baden-Württemberg. (n.d.). *Schulleistungsstudien: Internationale Vergleichsstudien*. Zugriff am 18.08.2014, von <http://www.kultusportal-bw.de/Lde/Startseite/schulebw/Schulleistungsstudien>
- Kultusportal Baden-Württemberg. (n.d.). *Vergleichsarbeiten*. Zugriff am 03.11.2014, von <http://www.kultusportal-bw.de/Lde/771669>
- Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S. & Neubrand, M. (Hrsg.). (2011). *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV*. Münster, München [u.a.]: Waxmann.
- Kunter, M. & Pohlmann, B. (2014). Lehrer. In E. Wild & J. Möller (Hrsg.), *Springer-Lehrbuch. Pädagogische Psychologie* (2nd ed., pp. 262–282). Berlin: Springer.
- Kuper, H. (Hrsg.). (2006). *Rückmeldung und Rezeption von Forschungsergebnissen: Zur Verwendung wissenschaftlichen Wissens im Bildungssystem*. Münster, New York, München, Berlin: Waxmann.
- Kuper, H. & Hartung, V. (2007). Überzeugungen zur Verwendung des Wissens aus Lernstandserhebungen: Eine professionstheoretische Analyse. *Zeitschrift für Erziehungswissenschaft*, 10.(2), 214–229. doi:10.1007/s11618-007-0028-5

- Landesbildungsserver Baden-Württemberg. (n.d.). *Lernstand 5*. Zugriff am 11.11.2014, von <http://www.schule-bw.de/entwicklung/lernstandserhebungen/lernstand5/konzeption/>
- Landesbildungsserver Baden-Württemberg. (n.d.). *Lernstandserhebungen*. Zugriff am 11.11.2014, von <http://www.schule-bw.de/entwicklung/lernstandserhebungen/>
- Landesbildungsserver Baden-Württemberg. (n.d.). *Vergleichsarbeiten DVA*. Zugriff am 11.11.2014, von <http://www.schule-bw.de/entwicklung/lernstandserhebungen/dva/>
- Landesbildungsserver Baden-Württemberg. (n.d.). *Vergleichsarbeiten VERA 8*. Zugriff am 11.11.2014, von <http://www.schule-bw.de/entwicklung/lernstandserhebungen/vera8/>
- Landesinstitut für Schulentwicklung (LS). (2013). *Vergleichsarbeiten VERA: Umgang mit den Ergebnissen im Rahmen der Selbstevaluation der Schulen*. Stuttgart. von <http://www.schule-bw.de/entwicklung/lernstandserhebungen/dokumente/vera3docs/v3-13-lk-handreichung-umgang-ergebnisse.pdf>
- Lienert, G. A. (1969). *Testaufbau und Testanalyse* (3., durch einen Anhang über Faktorenanalyse ergänzte Auflage). Weinheim, Berlin, Basel: Julius Beltz.
- Lütke, O., Robitzsch, A., Trautwein, U. & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung. *Psychologische Rundschau*, 58(2), 103–117. doi:10.1026/0033-3042.58.2.103
- Maier, U. (2008a). Rezeption und Nutzung von Vergleichsarbeiten aus der Perspektive von Lehrkräften. *Zeitschrift für Pädagogik*, 54(1), 95–117.
- Maier, U. (2008b). Vergleichsarbeiten im Vergleich: Akzeptanz und wahrgenommener Nutzen standardbasierter Leistungsmessungen in Baden-Württemberg und Thüringen. *Zeitschrift für Erziehungswissenschaft*, 11(3), 453–474. doi:10.1007/s11618-008-0036-0
- Maier, U. (2009). *Wie gehen Lehrerinnen und Lehrer mit Vergleichsarbeiten um?: Eine Studie zu testbasierten Schulreformen in Baden-Württemberg und Thüringen*. Schul- und Unterrichtsforschung: Bd. 7. Baltmannsweiler: Schneider-Verl. Hohengehren.
- Maier, U. (2010). Formative Assessment: Ein erfolgversprechendes Konzept zur Reform von Unterricht und Leistungsmessung? *Zeitschrift für Erziehungswissenschaft*, 13(2), 293–308. doi:10.1007/s11618-010-0124-9

- Maier, U. & Rauin, U. (2006). Vergleichsarbeiten: Hilfe zur Unterrichtsentwicklung. *Die Deutsche Schule*, (4), 403–421.
- Marsh, H. W., Dowson, M., Pietsch, J., & Walker, R. (2004). Why Multicollinearity Matters: A Reexamination of Relations Between Self-Efficacy, Self-Concept, and Achievement. *Journal of Educational Psychology*, 96(3), 518–522. doi:10.1037/0022-0663.96.3.518
- McElvany, N., Schroeder, S., Hachfeld, A., Baumert, J., Richter, T., Schnotz, W., . . . Ullrich, M. (2009). Diagnostische Fähigkeiten von Lehrkräften. *Zeitschrift für Pädagogische Psychologie*, 23(34), 223–235. doi:10.1024/1010-0652.23.34.223
- Ministerium für Kultus, J. u. S. B.-W. *Schulleistungsstudien*. Zugriff am 11.11.2014, von <http://www.kultusportal-bw.de/Lde/Startseite/schulebw/Schulleistungsstudien>
- Ministerium für Kultus, J. u. S. B.-W. (2014). *Onlinebasierte Lernstandserhebung*. von <http://www.kultusportal-bw.de/Lde/770225>
- Ministerium für Kultus, J. u. S. B.-W. (2014). *Vergleichsarbeiten*. von <http://www.kultusportal-bw.de/Lde/771669>
- Müller, A. (2009). *Rückmeldungen nach Vergleichsarbeiten im Kontext des schulischen Qualitätsmanagements. Drei explorative Studien zu Gestaltung und Rezeption im Anschluss an KOALA-S* (Dissertation). Ludwig-Maximilians-Universität München, München.
- Müller, A. & Ditton, H. (2014). Feedback: Begriff, Formen und Funktionen. In H. Ditton & A. Müller (Hrsg.), *Feedback und Rückmeldungen. Theoretische Grundlagen, empirische Befunde, praktische Anwendungsfelder* (pp. 11–28). Münster, New York, NY: Waxmann.
- Muthén, B. & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, 38, 171–189.
- Muthén, B. & Kaplan, D. (1992). A comparison of some methodologies for the factor analysis of non-normal Likert variables: A note on the size of the model. *British Journal of Mathematical and Statistical Psychology*, 45, 19–30.
- Muthén, L. K. & Muthén, B. O. (1998-2013). *Mplus User's Guide*. Los Angeles, CA: Muthén & Muthén.

- Nachtigall, C. & Jantowski, A. (2007). Die Thüringer Kompetenztests unter besonderer Berücksichtigung der Evaluationsergebnisse zum Rezeptionsverhalten. *Empirische Pädagogik*, 21(4), 401–410.
- Narciss, S. (2001). Feedback Strategies for Interactive Learning Tasks. In M. P. Driscoll, D. H. Jonassen, & P. Harris (Hrsg.), *AECT Series, 2. Handbook of Research for Educational Communications and Technology. A Project of the Association for Educational Communications and Technology* (3rd ed., pp. 125–140). Hoboken: Taylor & Francis.
- Narciss, S. (2006). *Informatives tutorielles Feedback: Entwicklungs- und Evaluationsprinzipien auf der Basis instruktionspsychologischer Erkenntnisse. Pädagogische Psychologie und Entwicklungspsychologie: Vol. 56*. Münster, München [u.a.]: Waxmann.
- Narciss, S. (2014). Modelle zu den Bedingungen und Wirkungen von Feedback in Lehr-Lernsituationen. In H. Ditton & A. Müller (Hrsg.), *Feedback und Rückmeldungen. Theoretische Grundlagen, empirische Befunde, praktische Anwendungsfelder* (pp. 43–82). Münster, New York, NY: Waxmann.
- Naumann, J., Richter, T. & Groeben, N. (2001). Validierung des INCOBI anhand eines Vergleichs von Anwendungsexperten und Anwendungsnovizen. *Zeitschrift für Pädagogische Psychologie*, 15(3), 219–232. doi:10.1024//1010-0652.15.34.219
- Peek, R. (2004). Qualitätsuntersuchung an Schulen zum Unterricht in Mathematik (QuaSUM): Klassenbezogene Ergebnisrückmeldungen und ihre Rezeption in Brandenburger Schulen. *Empirische Pädagogik*, 18(1), 82–114.
- Peek, R. & Nilshon, I. (2004). *Schulrückmeldungen im Rahmen von Schulleistungsstudien am Beispiel des QuaSUM-Projektes (Qualitätsuntersuchungen an Schulen zum Unterricht in Mathematik): Zwei Untersuchungen zur Wirksamkeit*. Brandenburg.
- Peschke, R. (2011). Medien- und IT-Management in Schulen am Scheideweg. In A. Breiter & M. Wind (Hrsg.), *Kritische Informatik: Bd. 6. Informationstechnik und ihre Organisationslücken. Soziale, politische und rechtliche Dimensionen aus der Sicht von Wissenschaft und Praxis* (pp. 111–132). Berlin, Münster: Lit.
- Podsakow, P. M. & Farh, J.-L. (1989). Effects of Feedback Sign and Credibility on Goal Setting and Task Performance. *Organizational Behavior And Human Decision Processes*, 44, 45–67.

- Ramm, G., Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., . . . Schiefele, U. (2006). *PISA 2003: Dokumentation der Erhebungsinstrumente*. Münster, München [u.a.]: Waxmann.
- Ratermann, M. & Stöbe-Blossey, S. (Hrsg.). (2012). *Governance von Schulund Elementarbildung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed). *Advanced quantitative techniques in the social sciences: Vol. 1*. Thousand Oaks: Sage Publications.
- Reinders, H., Ditton, H., Gräsel, C. & Gniewosz, B. (Hrsg.). (2011). *Empirische Bildungsforschung: Gegenstandsbereiche* (1. Aufl). Wiesbaden: VS, Verl. für Sozialwiss.
- Rheinberg, F. & Fries, S. (2010). Bezugsnormorientierung. In D. H. Rost (Hrsg.), *Programm PVU, Psychologie-Verlags-Union. Handwörterbuch pädagogische Psychologie* (4th ed., pp. 61–68). Weinheim [u.a.]: Beltz.
- Richter, T., Naumann, J. & Holz, H. (2010). Eine revidierte Fassung des Inventars zur Computerbildung (INCOBI-R). *Zeitschrift für Pädagogische Psychologie*, 24(1), 23–37.
- Rolff, H.-G. (1998-2002). Rückmeldung und Nutzung der Ergebnisse von großflächigen Leistungsuntersuchungen. In H.-G. Rolff, H. G. Holtappels, K. Klemm, H. Pfeiffer & R. (. Schulz-Zander (Hrsg.), *Eine Veröffentlichung des Instituts für Schulentwicklungsforschung der Universität Dortmund. Jahrbuch der Schulentwicklung. Daten, Beispiele und Perspektiven* . Weinheim [etc.]: Juventa Verlag.
- Rolff, H.-G., Holtappels, H. G., Klemm, K., Pfeiffer, H. & Schulz-Zander, R. (Hrsg.). (1998-2002). *Eine Veröffentlichung des Instituts für Schulentwicklungsforschung der Universität Dortmund. Jahrbuch der Schulentwicklung: Daten, Beispiele und Perspektiven*. Weinheim [etc.]: Juventa Verlag.
- Rossi, P. H., Lipsey, M. W. & Freeman, H. E. (2004). *Evaluation: A systematic approach* (7th ed). Thousand Oaks, CA: Sage Publications.
- Rost, D. H. (Hrsg.). (2010). *Programm PVU, Psychologie-Verlags-Union. Handwörterbuch pädagogische Psychologie* (4., überarb. und erw. Aufl). Weinheim [u.a.]: Beltz.
- Rothland, M. (Hrsg.). (2013). *Belastung und Beanspruchung im Lehrerberuf: Modelle, Befunde, Interventionen* (2., überarb. Aufl. 2013). Wiesbaden: Springer Fachmedien Wiesbaden; Imprint: Springer VS.

- Ryan, R. M. & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist*, 55(1), 68–78.
- Sacher, W. (2012). Erziehungs- und Bildungspartnerschaften in der Schule: zum Forschungsstand. In W. Stange, R. Krüger, A. Henschel & C. Schmitt (Hrsg.), *Erziehungs- und Bildungspartnerschaften. Grundlagen und Strukturen von Elternarbeit* (pp. 232–243). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Sadler, R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119–144.
- Sainsbury, M. & Benton, T. (2011). Designing a formative e-assessment: Latent class analysis of early reading skills. *British Journal of Educational Technology*, 42(3), 500–514. doi:10.1111/j.1467-8535.2009.01044.x
- Scheiter, K., Schrader, J., Trautwein, U. & Hesse, F. W. (2014). Analyse und Förderung effektiver Lehr-Lernprozesse im Kontext evidenzbasierter Bildungsreform – Beiträge der Tübinger Forschergruppe für Empirische Bildungsforschung. *Zeitschrift für Erziehungswissenschaft*, 17(2), 189–192. doi:10.1007/s11618-014-0535-0
- Schmitz, G. S. & Schwarzer, R. (2000). Selbstwirksamkeitserwartung von Lehrern: Längsschnittbefunde mit einem neuen Instrument. *Pädagogische Psychologie*, 14 (14), 12–25.
- Schneewind, J. (2006). *Wie Lehrkräfte mit Ergebnisrückmeldungen aus Schulleistungsstudien umgehen. Ergebnisse aus Befragungen von Berliner Grundschullehrerinnen* (Dissertation). Freie Universität Berlin, Berlin.
- Schneewind, J. (2007). Erfahrungen mit Ergebnisrückmeldungen im Projekt BeLesen: Ergebnisse der Interviewstudie. *Empirische Pädagogik*, 21(4), 368–382.
- Schneewind, J. & Kuper, H. (2009). Rückmeldeformate und Verwendungsmöglichkeiten der Ergebnisse aus zentralen Lernstandserhebungen. In T. Bohl & H. Kiper (Hrsg.), *Lernen aus Evaluationsergebnissen. Verbesserungen planen und implementieren*. Bad Heilbrunn: Klinkhardt.
- Schrader, F.-W. (1989). *Diagnostische Kompetenzen von Lehrern und ihre Bedeutung für die Gestaltung und Effektivität des Unterrichts. Europäische Hochschulschriften*.

- Reihe VI, Psychologie Publications universitaires européennes. Série VI, Psychologie European university studies. Series VI, Psychology: vol. 289.* Frankfurt am Main, New York: P. Lang.
- Schrader, F.-W. & Helmke, A. (2004). Von der Evaluation zur Innovation? Die Rezeptionsstudie WALZER: Ergebnisse der Lehrerbefragung. *Empirische Pädagogik*, 18(1), 140–161.
- Schrader, J. (2014). Analyse und Förderung effektiver Lehr-Lernprozesse unter dem Anspruch evidenzbasierter Bildungsreform. *Zeitschrift für Erziehungswissenschaft*, 17(2), 193–223. doi:10.1007/s11618-014-0540-3
- Schulbuchzentrum. (n.d.). *Nutzerhandbuch Zentrale Klassenverwaltung*. Zugriff am 09.01.2015, von <http://zkv.schulbuchzentrum-online.de/admintool/downloads/Nutzerhandbuch%20ZKV%20Schuladmin.pdf>
- Schulz-Zander, R., Eickelmann, B., Moser, H., Niesyto, H. & Grell, P. (2012). Editorial: Qualitätsentwicklung in der Schule und medienpädagogische Professionalisierung. In R. Schulz-Zander, B. Eickelmann, H. Moser, H. Niesyto & P. Grell (Hrsg.), *Jahrbuch Medienpädagogik: Vol. 9. Jahrbuch Medienpädagogik 9* (pp. 9–14). Wiesbaden: Springer VS.
- Schulz-Zander, R., Eickelmann, B., Moser, H., Niesyto, H. & Grell, P. (Hrsg.). (2012). *Jahrbuch Medienpädagogik: Vol. 9. Jahrbuch Medienpädagogik 9*. Wiesbaden: Springer VS.
- Schwarzer, R. & Jerusalem, M. (2002). Das Konzept der Selbstwirksamkeit. In M. Jerusalem & D. Hopf (Hrsg.), *Zeitschrift für Pädagogik : Beiheft: Vol. 44. Selbstwirksamkeit und Motivationsprozesse in Bildungsinstitutionen* (pp. 28–53). Weinheim [u.a.]: Beltz.
- Schwippert, K. (2004). Leistungsrückmeldungen an Grundschulen im Rahmen der Internationalen Grundschul-Lese-Untersuchung (IGLU). *Empirische Pädagogik*, 18(1), 62–81.
- Schwippert, K. (2005). Zur gewandelten Akzeptanz von Schulrückmeldungen. In H. Döbert & H.-W. Fuchs (Hrsg.), *Studien zur international und interkulturell vergleichenden Erziehungswissenschaft: Bd. 6. Leistungsmessungen und Innovationsstrategien in Schulsystemen. Ein internationaler Vergleich* (pp. 63–78). Münster, New York, München, Berlin: Waxmann.

- Scriven, M. (1972). Die Methodologie der Evaluation. In C. Wulf (Hrsg.), *Erziehung in Wissenschaft und Praxis: Vol. 18. Evaluation. Beschreibung und Bewertung von Unterricht, Curricula und Schulversuchen* (pp. 60–91). München: R. Piper.
- Shah, P. & Hoeffner, J. (2002). Review of Graph Comprehension Research: Implications for Instruction. *Educational Psychology Review*, 14(1), 47–69.
- Shermis, M. D. & Lombard, D. (1998). Effects of Computer-Based Test Administrations on Test Anxiety and Performance. *Computers in Human Behavior*, 14(1), 111–123.
- Spannagel, C. & Bescherer, C. (2009). Computerbezogene Selbstwirksamkeitserwartung in Lehrveranstaltungen mit Computernutzung. *Notes on Educational Informatics (NEI) — Section A: Concepts and Techniques*, 5(1).
- Spence, J. T. (Hrsg.). (1983). *A Series of books in psychology. Achievement and achievement motives: Psychological and sociological approaches*. San Francisco: W.H. Freeman.
- Spinath, B. (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie*, 19(1), 85–95. doi:10.1024/1010-0652.19.1.85
- Spinath, B. (2011). Lernmotivation. In H. Reinders, H. Ditton, C. Gräsel & B. Gniewosz (Hrsg.), *Empirische Bildungsforschung. Gegenstandsbereiche* (1st ed., pp. 45–56). Wiesbaden: VS, Verl. für Sozialwiss.
- Stamm, M. (2003). *Evaluation und ihre Folgen für die Bildung: Eine unterschätzte pädagogische Herausforderung. Internationale Hochschulschriften: Vol. 419*. Münster, New York, München, Berlin: Waxmann.
- Stange, W., Krüger, R., Henschel, A. & Schmitt, C. (Hrsg.). (2012). *Erziehungs- und Bildungspartnerschaften: Grundlagen und Strukturen von Elternarbeit*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Statistisches Landesamt Baden-Württemberg. (2011). *Grund-, Werkreal- und Hauptschulen in Baden-Württemberg: Adressverzeichnis*. Schuljahr 2010/2011. Stuttgart.
- Statistisches Landesamt Baden-Württemberg. (2014). *Statistik Aktuell: Lehrkräfte in Baden-Württemberg*. Zugriff am 21.02.2015, von http://www.statistik.baden-wuerttemberg.de/Veroeffentl/Statistik_AKTUELL/803414004.pdf

- Stauke, E. (2008). Rückmeldesystem zur Kompetenzmessung in Ungarn. In A. Breiter, A. Lange & E. Stauke (Hrsg.), *School information systems and data-based decision-making* (1st ed., pp. 83–94). Frankfurt am Main, New York: P. Lang.
- Steyer, R. & Eid, M. (2001). *Messen und Testen: Mit Übungen und Lösungen*. Springer-Lehrbuch. Berlin: Springer.
- Terhart, E. (2002). Wie können die Ergebnisse von vergleichenden Leistungsstudien systematisch zur Qualitätsverbesserung in Schulen genutzt werden? *Zeitschrift für Pädagogik*, 48(1), 91–110.
- Trautwein, U. & Köller, O. (2003). Was lange währt, wird nicht immer gut: Zur Rolle selbstregulativer Strategien bei der Hausaufgabenerledigung. *Zeitschrift für Pädagogische Psychologie*, 17(3/4), 199–209. doi:10.1024//1010-0652.17.3.199
- Tresch, S. (2007). *Potenzial Leistungstest: Wie Lehrerinnen und Lehrer Ergebnisrückmeldungen zur Sicherung und Steigerung ihrer Unterrichtsqualität nutzen*. Bern: Hep.
- Tschannen-Moran, M. & Woolfolk Hoy, A. (2001). Teacher efficacy: capturing an elusive construct. *Teaching and Teacher Education*, 17, 783–805.
- van Ackeren, I., Zlatkin-Troitschanskaia, O., Binnewies, C., Clausen, M., Dormann, C., Preisendörfer, P., . . . Schmidt, U. (2011). Evidenzbasierte Schulentwicklung. Ein Forschungsüberblick aus interdisziplinärer Perspektive. *DDS - Die Deutsche Schule*, 103(2), 170–184.
- Visscher, A. J. (2008). The utilization of school performance feedback system for school improvement. In A. Breiter, A. Lange, & E. Stauke (Hrsg.), *School information systems and data-based decision-making* (1st ed., pp. 23–36). Frankfurt am Main, New York: P. Lang.
- Voss, T. & Kunter, M. (2011). Pädagogisch-psychologisches Wissen von Lehrkräften. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (pp. 193–214). Münster, München [u.a.]: Waxmann.
- Wacker, A. & Kramer, J. (2012). Vergleichsarbeiten in Baden-Württemberg. *Zeitschrift für Erziehungswissenschaft*, 15(4), 683–706. doi:10.1007/s11618-012-0326-4
- Wacker, A., Maier, U. & Wissinger, J. (2012). Ergebnisorientierte Steuerung – Bildungspolitische Strategie und Verfahren zur Initiierung von Schul- und Unterrichtsreformen. In A. Wacker, U. Maier & J. Wissinger (Hrsg.), *Schul- und Unterrichtsreform*

- durch ergebnisorientierte Steuerung (pp. 9–33). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Wacker, A., Maier, U. & Wissinger, J. (Hrsg.). (2012). *Schul- und Unterrichtsreform durch ergebnisorientierte Steuerung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Wacker, A., Maier, U. & Wissinger, J. (2012). *Schul- und Unterrichtsreform durch ergebnisorientierte Steuerung: Empirische Befunde und forschungsmethodische Implikationen*. SpringerLink : Bücher. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Waldersee, R. & Luthans, F. (1994). The Impact of Positive and Corrective Feedback on Customer Service Performance. *Journal of Organizational Behavior*, 15(1), 83–95.
- Watermann, R., Stanat, P., Kunter, M., Klieme, E. & Baumert, J. (2003). Schulrückmeldungen im Rahmen von Schulleistungsuntersuchungen: Das Disseminationskonzept von PISA-2000. *Zeitschrift für Pädagogik*, 49(1), 92–111.
- Weitze, M.-D., Pühler, A., Heckl, W. M., Müller-Röber, B., Renn, O., Weingart, P. & Wess, G. (2012). *Biotechnologie-Kommunikation: Kontroversen, Analysen, Aktivitäten*. Berlin, Heidelberg: Springer Vieweg.
- Whitley, B. E. (1997). Gender Differences in Computer-Related Attitudes and Behavior: A Meta-Analysis. *Computers in Human Behavior*, 13(1), 1–22.
- Wild, E. & Möller, J. (Hrsg.). (2014). *Springer-Lehrbuch. Pädagogische Psychologie* (2. vollst. überarb. und aktualisierte Aufl). Berlin: Springer.
- Williams, J. E. & McCord, D. M. (2006). Equivalence of standard and computerized versions of the Raven Progressive Matrices Test. *Computers in Human Behavior*, 22(5), 791–800. doi:10.1016/j.chb.2004.03.005
- Wulf, C. (Hrsg.). (1972). *Erziehung in Wissenschaft und Praxis: Vol. 18. Evaluation: Beschreibung und Bewertung von Unterricht, Curricula und Schulversuchen*. München: R. Piper.
- Zanna, M. P. & Rempel, J. K. (1988). Attitudes: A new look at an old concept. In D. Bar-Tal & A. W. Kruglanski (Hrsg.), *The Social psychology of knowledge* (pp. 315–334). Cambridge, New York, Paris: Cambridge University; Editions de la maison des sciences de l'homme.
- Zimmer-Müller, M., Hosenfeld, I. & Koch, U. (2014). Rückmeldungen nach Vergleichsarbeiten in Grund- und Sekundarschulen. In H. Ditton & A. Müller (Hrsg.), *Feedback*

und Rückmeldungen. Theoretische Grundlagen, empirische Befunde, praktische Anwendungsfelder (pp. 195–212). Münster, New York, NY: Waxmann.

10 Verzeichnisse

10.1 Abbildungsverzeichnis

<i>Abbildung 1.</i> Bildungsstandards der Grundschule im Unterrichtsfach Deutsch (KMK, 2004a, S. 7).....	9
<i>Abbildung 2.</i> Zyklenmodell der Rückmeldung (vgl. Helmke, 2004; Helmke & Hosenfeld, 2005; Hosenfeld & Groß Ophoff, 2007).....	15
<i>Abbildung 3.</i> Multidimensionale Konzeptualisierung von Feedback (Hattie & Wollenschläger, 2014).....	26
<i>Abbildung 4.</i> Ablauf und Testdurchführung des Testverfahrens <i>Online Diagnose</i>	40
<i>Abbildung 5.</i> Zugang und Benutzerkennung zum Testverfahren <i>Online Diagnose</i>	41
<i>Abbildung 6.</i> Bildungsstandards mit Beispielen.....	44
<i>Abbildung 7.</i> Testinhalte <i>Online Diagnose</i>	45
<i>Abbildung 8.</i> Leistungsprofil nach Stärken und Schwächen der gesamten Klasse	48
<i>Abbildung 9.</i> Testvergleich des Haupttests/Detailtests mit dem Nachtest	49
<i>Abbildung 10.</i> Ergebnis-Geschwindigkeits-Matrix für jede Schülerin und jeden Schüler..	49
<i>Abbildung 11.</i> Ablauf Testverfahren <i>Online Diagnose</i> in Anlehnung an das Zyklenmodell der Rückmeldung	51
<i>Abbildung 12.</i> Ablauf der Erhebung der Lehrkraftbefragung zum Testverfahren <i>Online Diagnose</i>	57
<i>Abbildung 13.</i> Nutzung Instrumente und Maßnahmen.....	64
<i>Abbildung 14.</i> Zeitaufwand während und außerhalb des Unterrichts	65
<i>Abbildung 15.</i> Mögliche Gründe für ein schwaches Testergebnis	71
<i>Abbildung 16.</i> Beispielaufgabe Radiergummi.....	91
<i>Abbildung 17.</i> Beispielaufgabe Schiedsrichter.....	91
<i>Abbildung 18.</i> Beispielaufgabe Kühlschranks	92

10.2 Tabellenverzeichnis

Tabelle 1	<i>Studienfach Deutsch und Abschlussnoten</i>	58
Tabelle 2	<i>Deskriptive Statistik der Items zum Ablauf des Testverfahrens Online</i> <i>Diagnose</i>	63
Tabelle 3	<i>Deskriptive Statistik der Items zum Zeitaufwand während und außerhalb</i> <i>des Unterrichts</i>	66
Tabelle 4	<i>Deskriptive Statistik der Items zum Austausch der Lehrkräfte mit</i> <i>pädagogischen Akteuren der Schule</i>	67
Tabelle 5	<i>Deskriptive Statistik der Items zur Anwendung der Fördermaterialien</i>	68
Tabelle 6	<i>Deskriptive Statistik der Items zur technischen Ausstattung von Schulen</i>	70
Tabelle 7	<i>Deskriptive Statistik der Items zu möglichen Gründen für ein schwaches</i> <i>Testergebnis</i>	71
Tabelle 8	<i>Deskriptive Statistik der Items zur Unterstützung für die Lehrkräfte bei der</i> <i>Nutzung der Online Diagnose</i>	73
Tabelle 9	<i>Deskriptive Statistik, Trennschärfenkoeffizienten und Faktorenladungen zu</i> <i>den Wertaspekten in Bezug auf den Haupttest aus Schülerperspektive</i>	76
Tabelle 10	<i>Deskriptive Statistik, Trennschärfenkoeffizienten und Faktorenladungen zu</i> <i>den Wertaspekten in Bezug auf den Haupttest aus Lehrerperspektive</i>	79
Tabelle 11	<i>Deskriptive Statistik und interne Konsistenz der Skalen zu den Wertaspekten</i> <i>in Bezug auf den Haupttest aus Schülerperspektive und Lehrerperspektive</i>	80
Tabelle 12	<i>Deskriptive Statistik, Trennschärfenkoeffizienten und Faktorenladungen zu</i> <i>den Wertaspekten in Bezug auf die Fördermaterialien aus Schülerperspektive</i>	82
Tabelle 13	<i>Deskriptive Statistik, Trennschärfenkoeffizienten und Faktorenladungen zu</i> <i>den Wertaspekten in Bezug auf die Fördermaterialien aus Lehrerperspektive</i>	85
Tabelle 14	<i>Deskriptive Statistik und interne Konsistenz der Skalen zu Wertaspekten in</i> <i>Bezug auf die Fördermaterialien aus Schülerperspektive und</i> <i>Lehrerperspektive</i>	86
Tabelle 15	<i>Bivariate Korrelationen der Wertaspekte in Bezug auf den Haupttest und</i> <i>der Wertaspekte in Bezug auf die Fördermaterialien</i>	89
Tabelle 16	<i>Deskriptive Statistik und Trennschärfenkoeffizienten zu den Bereichen</i> <i>Lesen und Hören sowie Rechtschreibung, Schreiben und Sprache</i>	93
Tabelle 17	<i>Faktorenladungen einer exploratorischen Faktorenanalyse mit Promax</i> <i>Rotation zur Qualität der Ergebnisrückmeldungen</i>	95
Tabelle 18	<i>Deskriptive Statistik und Trennschärfenkoeffizienten der subjektiven</i> <i>Qualitätseinschätzung der Instrumente und Fördermaterialien</i>	97
Tabelle 19	<i>Bivariate Korrelationen der Skalen zur subjektiven Qualitätseinschätzung</i>	98
Tabelle 20	<i>Deskriptive Statistik und Trennschärfenkoeffizienten der Items zum</i> <i>Prozessschritt Rezeption</i>	102
Tabelle 21	<i>Deskriptive Statistik und Trennschärfenkoeffizienten der Items zum</i> <i>Prozessschritt Reflexion</i>	104
Tabelle 22	<i>Deskriptive Statistik und Trennschärfenkoeffizienten der Items zum</i> <i>Prozessschritt Aktion</i>	105

Tabelle 23	<i>Deskriptive Statistik, Trennschärfenkoeffizienten und Faktorenladungen einer exploratorischen Faktorenanalyse zum Prozessschritt Evaluation</i>	108
Tabelle 24	<i>Itembeispiele und Cronbachs α zu den unabhängigen und abhängigen Variablen</i>	125
Tabelle 25	<i>Interkorrelationen der abhängigen und unabhängigen Variablen</i>	128
Tabelle 26	<i>Lineare multiple Regression zur abhängigen Variable Wirkung auf Lernende</i>	136
Tabelle 27	<i>Lineare multiple Regression zur abhängigen Variable Wirkung auf das Kollegium</i>	137
Tabelle 28	<i>Lineare multiple Regression zur abhängigen Variable Wirkung auf Lehrende</i>	138
Tabelle 29	<i>Itembeispiele und Cronbachs α zu den Prozessschritten Rezeption, Reflexion, Aktion und Evaluation</i>	150
Tabelle 30	<i>Deskriptive Statistik sowie interne Konsistenz der Variablen zu den individuellen Bedingungen</i>	154
Tabelle 31	<i>Deskriptive Statistik (sowie interne Konsistenz) der Variablen zu den Bedingungen der Klasse</i>	157
Tabelle 32	<i>Deskriptive Statistik sowie interne Konsistenz der Variablen zu den schulischen und externen Bedingungen</i>	160
Tabelle 33	<i>Bivariate Korrelationen der Variablen der Prozessschritte Rezeption, Reflexion, Aktion und Evaluation mit den individuellen Bedingungen, Bedingungen der Klasse sowie schulischen und externen Bedingungen</i>	164
Tabelle 34	<i>Regressionsanalysen zu den individuellen Bedingungen</i>	168
Tabelle 35	<i>Regressionsanalysen zu den Bedingungen der Klasse</i>	169
Tabelle 36	<i>Regressionsanalysen zu den schulischen und externen Bedingungen</i>	170
Tabelle 37	<i>Regressionsanalysen der abhängigen Variablen mit allen unabhängigen Variablen der Bedingungen</i>	178
Tabelle 38	<i>Regressionsanalysen der abhängigen Variablen mit allen signifikanten unabhängigen Variablen der Bedingungen</i>	180