

Jan Hochweber · Christian Brühwiler · Dimitra Kolovou · Giang Pham ·
Nevil Knöpfli · Ann Christin Hochweber

Aktive Lernzeit und Lernerfolg für ALLE

Schlussbericht zur Evaluation des Pilotprojekts

November 2020

Auftraggeberin

Bildungsdirektion des Kantons Zürich

Auftragnehmerin

Pädagogische Hochschule St.Gallen
Institut für Kompetenzdiagnostik
Notkerstrasse 27
CH-9000 St.Gallen

evaluation-alle@phsg.ch
www.phsg.ch/forschung

Bezugsquelle

Bildungsdirektion des Kantons Zürich
Bildungsplanung
Walcheturm, Walcheplatz 2
CH-8090 Zürich

www.zh.ch/bildungsplanung
www.zh.ch/studien-bildung

Der Bericht liegt nur in elektronischer Form vor.

Zitationsvorschlag

Hochweber, J., Brühwiler, C., Kolovou, D., Pham, G., Knöpfli, N. & Hochweber, A. C. (2020). Aktive Lernzeit und Lernerfolg für ALLE. Schlussbericht zur Evaluation des Pilotprojekts. Zürich: Bildungsdirektion Kanton Zürich.

Danksagung

Das Evaluationsteam bedankt sich für die vielfältige Unterstützung, die den vorliegenden Bericht erst ermöglicht hat. Wir bedanken uns bei den ehemaligen Mitarbeiterinnen im Evaluationsteam Dr. Nadja Abt Gürber, Dr. Benita Affolter und Bengü Kavadarli sowie bei den Praktikantinnen und studentischen Hilfskräften Thea Böni, Barbara Ehrbar, Taina Giezendanner, Nina Mitterer, Cassandra Vicini und Corina Waldispühl für ihr engagiertes Mitwirken. Prof. Dr. Titus Guldemann danken wir für seinen Beitrag zur Konzeption und Planung der Evaluation, Astrid Sperger, Patrizia Egli und Olivia Fuchs wir für ihre Mitarbeit bei der Interviewtranskription, Prof. Dr. Michael Beck für seine Unterstützung beim Datenmanagement. Weiterhin bedanken wir uns bei der Leiterin des Pilotprojekts Madeleine Wolf sowie bei Dr. Sybille Bayard, Hans-Martin Binder und Peter Nussbaum von der Bildungsplanung des Kantons Zürich für die vertrauensvolle Zusammenarbeit und konstruktive Unterstützung. Prof. em. Dr. Andreas Helmke danken wir für die fundierte Beratung in allen Phasen des Projekts. Dem Lehrmittelverlag St.Gallen und der Firma Arcadix danken wir für die Möglichkeit, Aufgabenpools und Infrastruktur der Lernfördersysteme für die Durchführung von Tests nutzen zu können. Und schliesslich gilt unserer besonderer Dank den Schulen, die an der Evaluation des Pilotprojekts beteiligt waren, den Schulleitungen und Lehrpersonen für ihr grosses zeitliches und persönliches Engagement sowie den Schülerinnen und Schülern für ihre hohe und anhaltende Teilnahmebereitschaft.

Inhaltsverzeichnis

Executive Summary	7
1 Einleitung	12
2 Fragestellungen	16
3 Methodische Grundlagen	22
3.1 Untersuchungsdesign	22
3.2 Durchführung der Datenerhebungen	24
3.2.1 Vorbereitungen an den Schulen	24
3.2.2 Durchführung der Kompetenztests und Befragungen	25
3.2.3 Durchführung der Interviews	28
3.3 Datengrundlage	29
3.3.1 Gewinnung der Stichproben für die Evaluation ALLE	29
3.3.2 Beschreibung der Stichproben	33
3.3.2.1 <i>Stichprobe der Schülerinnen und Schüler</i>	33
3.3.2.2 <i>Stichprobe der Lehrpersonen</i>	39
3.3.2.3 <i>Stichproben in den Befragungen zum Pilotprojekt und in den Interviews</i>	42
3.4 Beschreibung der Erhebungsinstrumente	44
3.4.1 Kompetenztests	44
3.4.2 Fragebogeninstrumente	47
3.4.3 Interviews	49
3.5 Datenaufbereitung und Datenanalyse	52
3.5.1 Aufbereitung der soziodemografischen Merkmale	53
3.5.2 IRT-Skalierung der Kompetenztests und Linking	54
3.5.3 Datenimputation	55
3.5.4 Schätzung der fachlichen Lernergebnisse: Plausible Values	57
3.5.5 Datenanalysen mit PVs und multipel imputierten Datensätzen	58
3.5.6 Transkription der Interviews	58
3.5.7 Kodierprozess und Kategoriensystem	59
3.6 Hinweise zur Interpretation der Ergebnisse	61
4 Ergebnisse	65
4.1 Rahmenbedingungen des Pilotprojekts ALLE	65
4.1.1 Finanzielle Voraussetzungen zur Umsetzung der Massnahmen	65
4.1.2 Qualität der innerschulischen Kooperation	67

4.2	Einsatz der Schulassistenzen	69
4.2.1	Rekrutierung, personale Konstanz und Unterstützung durch die Schulleitung	70
4.2.2	Qualifikationen der Schulassistenzen	71
4.2.3	Umfang und Art des Einsatzes	74
4.2.4	Qualität der Zusammenarbeit	86
4.2.5	Wahrgenommene Unterstützung durch die Schulassistenzen	94
4.2.6	Veränderungen im Unterrichtsgeschehen	97
4.2.7	Allgemeine Zufriedenheit mit der Umsetzung der Massnahme	100
4.3	Aufgabenhilfe in Deutsch und Mathematik	102
4.3.1	Integration der Aufgabenhilfe und Unterstützung durch die Schulleitung	103
4.3.2	Verfügbarkeit, Einsatz und Durchführung der Aufgabenhilfe	104
4.3.3	Wahrgenommene Unterstützung durch die Aufgabenhilfe	110
4.3.4	Allgemeine Zufriedenheit mit der Umsetzung der Massnahme	115
4.4	Weiterbildung der Lehrpersonen	117
4.4.1	Besuchte Weiterbildungen	118
4.4.2	Wahrgenommene Nützlichkeit und Verbesserungen im Fachunterricht	119
4.4.3	Allgemeine Zufriedenheit mit der Umsetzung der Massnahme	123
4.4.4	Überlegungen der Weiterbildungspersonen zu Optimierungen des Angebots	125
4.5	Beratungen für Lehrpersonen	128
4.5.1	Umfang und Art der in Anspruch genommenen Beratungen	128
4.5.2	Allgemeine Zufriedenheit mit der Umsetzung der Massnahme	130
4.5.3	Überlegungen der Beratungspersonen zum Beratungserfolg und zu Optimierungen	132
4.6	Einsatz von EMU	135
4.6.1	Umsetzung von EMU	136
4.6.2	Wahrgenommene Wirkungen von EMU	141
4.6.3	Allgemeine Zufriedenheit mit der Umsetzung der Massnahme	145
4.7	Individualisierung im Fachunterricht	146
4.8	Professionelle Kompetenzen der Lehrpersonen	152
4.8.1	Überzeugungen zum Lehren und Lernen	153
4.8.2	Urteilsakkuratheit und Urteilssicherheit der Lehrpersonen	157
4.8.2.1	<i>Indikatoren der Urteilsakkuratheit und Urteilssicherheit</i>	158
4.8.2.2	<i>Ergebnisse zu Urteilsakkuratheit und Urteilssicherheit</i>	159
4.9	Qualität des Unterrichts	167
4.9.1	Indikatoren der wahrgenommenen Unterrichtsqualität	167
4.9.2	Unterrichtsqualität aus Sicht der Schülerinnen und Schüler	171
4.9.3	Unterrichtsqualität aus Sicht der Lehrpersonen	179

4.10	Lernergebnisse der Schülerinnen und Schüler	184
4.10.1	Motivational-affektive Lernergebnisse	185
4.10.1.1	<i>Lernergebnisse in der gesamten Evaluationsstichprobe</i>	186
4.10.1.2	<i>Lernergebnisse bei ungünstiger Lernausgangslage</i>	194
4.10.2	Fachliche Lernergebnisse	196
4.10.2.1	<i>Lernergebnisse in der gesamten Evaluationsstichprobe</i>	197
4.10.2.2	<i>Lernergebnisse bei ungünstiger Lernausgangslage</i>	207
4.10.2.3	<i>Lernergebnisse im mittleren und unteren Leistungsniveau</i>	208
4.10.2.4	<i>Ergebnisse der Lernstandserhebung «Texte schreiben»</i>	210
4.11	Explorative Analyse: Gelingensbedingungen von ALLE	211
4.11.1	Bestimmung der Analyseebene und Datenbasis	212
4.11.1.1	<i>Klärung der Analyseebene</i>	212
4.11.1.2	<i>Stichprobe</i>	213
4.11.1.3	<i>Variablenauswahl</i>	214
4.11.2	Analysestrategie	217
4.11.3	Schritte der Datenanalyse und Ergebnisse	219
5	Bewertung der Ergebnisse und Entwicklungsperspektiven	226
5.1	Rahmenbedingungen des Pilotprojekts ALLE	226
5.2	Das Massnahmenbündel im Pilotprojekt ALLE	227
5.2.1	Schulassistenzen	228
5.2.2	Aufgabenhilfe	232
5.2.3	Weiterbildungen der Lehrpersonen	235
5.2.4	Beratungsangebote für die Lehrpersonen	236
5.2.5	EMU	238
5.2.6	Individualisierung im Fachunterricht	240
5.3	Wirkungen des Pilotprojekts ALLE	241
5.3.1	Professionelle Kompetenzen der Lehrpersonen	242
5.3.2	Qualitätsmerkmale des Unterrichts	244
5.3.3	Motivational-affektive Lernergebnisse der Schülerinnen und Schüler	245
5.3.4	Fachliche Lernergebnisse der Schülerinnen und Schüler	246
5.3.5	Fazit	248
5.4	Grenzen der Evaluation	250
5.4.1	Erhebungsdesign	250
5.4.2	Datengrundlage	251
5.4.3	Erhebungsinstrumente	252
5.4.4	Sonstige Einschränkungen	254
5.5	Handlungsfelder und Empfehlungen	254

6	Literatur	261
7	Anhänge	271
7.1	Anhang A: Kennwerte der Skalen für die Kapitel 4.8 bis 4.9	271
7.2	Anhang B: Deskriptive Statistiken zu den Abbildungen in den Kapiteln 4.2 bis 4.7	277
7.3	Anhang C: Messzeitpunktspezifische Ergebnisse zu den Kapiteln 4.8 und 4.9	298
8	Abbildungsverzeichnis	306
9	Tabellenverzeichnis	311
10	Abkürzungsverzeichnis	315

Executive Summary

Das Pilotprojekt ALLE

Im Pilotprojekt «Aktive Lernzeit und Lernerfolg für ALLE» erprobten neun öffentliche Sekundarschulen im Kanton Zürich verschiedene pädagogische Massnahmen mit dem Ziel, die Kompetenzen der Schülerinnen und Schüler in den Fächern Deutsch und Mathematik zu verbessern. Das Pilotprojekt startete im Schuljahr 2016/17 und endete mit dem Schuljahr 2018/19. Das Massnahmenbündel umfasste:

- Weiterbildungen für Lehrpersonen und Schulleitungen,
- Beratungen für die Lehrpersonen,
- kriterienorientierte, datenbasierte Unterrichtsreflexion (Kollegial- und Schülerfeedback) mittels EMU («Evidenzbasierte Methoden der Unterrichtsdiagnostik und -entwicklung»),
- Integration einer Aufgabenhilfe für die Fächer Deutsch und Mathematik und
- Einsatz von Schulassistenten im Unterricht zur Unterstützung der Lehrpersonen.

Die Evaluation des Pilotprojekts ALLE

Zur Evaluation des Pilotprojekts wurde eine Längsschnittstudie mit vier Messzeitpunkten durchgeführt, in der die Ergebnisse der neun am Pilotprojekt teilnehmenden Schulen (ALLE-Schulen) den Ergebnissen von neun weiteren Schulen der Sekundarschule aus dem Kanton Zürich (Vergleichsschulen) gegenübergestellt wurden. Die Erhebungen erfolgten zu Beginn der 7. Klassenstufe (Anfang des Schuljahres 2016/17) sowie gegen Ende der 7., 8. und 9. Klassenstufe (Ende des Schuljahres 2018/19). Sie umfassten Kompetenztests mit den Schülerinnen und Schülern sowie Befragungen von Lehrpersonen, Schulassistenten, Schulleitungen und Schülerinnen und Schülern. Um die Erkenntnisse zur Umsetzung der Massnahmen zu vertiefen, wurden zusätzliche Befragungen und Interviews mit Lehrpersonen, schulischen Heilpädagoginnen und Heilpädagogen, Schulassistenten, Schulleitungen sowie Beratungs- und Weiterbildungspersonen durchgeführt.

An den ALLE- und Vergleichsschulen nahmen insgesamt 1'687 Schülerinnen und Schüler an mindestens einer der vier Erhebungen der Evaluation teil. Sämtliche ALLE-Schulen und sechs der neun Vergleichsschulen mit insgesamt 1'480 Schülerinnen und Schülern (87.7% der Gesamtstichprobe) waren an allen vier Messzeitpunkten beteiligt.

Befunde zu den Wirkungen von ALLE

Die Grundidee des Pilotprojekts ALLE war es, die Qualität des Unterrichts weiterzuentwickeln, um bei den teilnehmenden Schülerinnen und Schülern eine Verbesserung der fachlichen Lernergebnisse – ferner auch der motivational-affektiven Lernergebnisse – zu erreichen. Die Evaluation ergab, dass mit dem Einsatz des ALLE-Massnahmenbündels zumindest teilweise die erhofften Wirkungen erzielt werden konnten. Gegenüber den Vergleichsschulen ergaben sich für die ALLE-Schulen Hinweise auf eine günstigere Leistungsentwicklung für das Fach Deutsch (im Gesamttest) sowie für den Kompetenzbereich «Form und Raum» im Fach Mathematik. Für die untersuchten motivational-affektiven Merkmale ergaben sich Unterschiede zugunsten der ALLE-Schulen bei der anerkennungs- und leistungsbezogenen Motivation der Schülerinnen und Schüler sowie (der Tendenz nach) bei der intrinsischen Motivation, die allerdings auf das Fach Mathematik beschränkt blieben. Die erzielten Effektgrößen sind eher gering, die zwischen ALLE- und Vergleichsschulen festgestellten Unterschiede demnach nicht gross. Berücksichtigt man jedoch, dass sich die Unterschiede auf sämtliche an ALLE teilnehmenden Schülerinnen und Schüler beziehen, sind sie als bedeutsam einzustufen.

Nur punktuell zeigte sich, dass ALLE den leistungsschwächeren Schülerinnen und Schülern zugute kam, deren Förderung im Pilotprojekt einen besonderen Stellenwert eingenommen hatte. Für die Schülerinnen und Schüler in den ALLE-Schulen, die zu Beginn der 7. Klassenstufe in den Kompetenztests zu den Schwächeren zählten, ergaben sich Anzeichen einer günstigeren Entwicklung für die bereits angesprochenen motivationalen Merkmale: die anerkennungs- und leistungsbezogene sowie intrinsische Motivation im Fach Mathematik. Dies ist von Bedeutung, da eine hinreichende Lernmotivation auch als wichtige Voraussetzung für die Förderung und Entwicklung der fachlichen Lernergebnisse anzusehen ist. Die Befunde der Evaluation zeigen zudem, dass die verstärkten Bemühungen in den ALLE-Schulen um die Förderung der leistungsschwächeren Schülerinnen und Schüler nicht zulasten der leistungsstärkeren Schülerinnen und Schüler gingen.

Für die Förderung des Lernens spielt qualitativ hochwertiger – das heisst lernförderlicher, effizienter, die Schülerinnen und Schüler unterstützender Unterricht – eine zentrale Rolle. Zwar nahmen die Schülerinnen und Schüler den Unterricht in beiden Fächern – bezüglich aller untersuchten Qualitätsmerkmale und zu allen Messzeitpunkten – überwiegend positiv wahr. Eine günstige Wirkung von ALLE auf die erfassten fachübergreifenden Qualitätsmerkmale des Unterrichts konnte jedoch weder aus der Sicht der Schülerinnen und Schüler noch aus der Sicht der Lehrpersonen schlüssig nachgewiesen werden. Am deutlichsten zeigte sich die Tendenz, dass die Lehrpersonen an den ALLE-Schulen von einer verbesserten individuellen Förderung und höheren Leistungserwartungen berichteten, während dies an den Vergleichsschulen nicht oder in geringerem Umfang erkennbar war.

Auch bei den untersuchten professionellen Kompetenzen der Lehrpersonen, die als wichtige Grundlage für einen guten, lernförderlichen Unterricht anzusehen sind, fanden sich nur vereinzelt Hinweise auf Wirkungen der ALLE-Massnahmen. Ein günstiges Ergebnis zeigte sich für das Ausmass, in dem die Lehrpersonen die Leistungen ihrer Schülerinnen und Schüler auf deren Begabung beziehungsweise auf angeborene Fähigkeitsunterschiede zurückführten: Die Überzeugung, dass für die Lernergebnisse vor allem die Begabung entscheidend ist, nahm bei den Deutsch-Lehrpersonen in den ALLE-Schulen während des Pilotprojekts eher ab, in den Vergleichsschulen dagegen noch zu. Da eine starke Zuschreibung von Leistungen auf die Begabung durch die Lehrperson pädagogisch wenig wünschenswert ist, lässt sich dies als Hinweis auf eine positive Wirkung von ALLE interpretieren.

Einordnung der Befunde: Umsetzung des Massnahmenbündels

Die oben zusammengefassten Befunde der Evaluation enthalten einige ermutigende Ergebnisse zur Wirkung der im Pilotprojekt ALLE eingesetzten Massnahmen. Zugleich dürften sie nicht in jeder Hinsicht den hohen Erwartungen an die Erträge des Pilotprojekts entsprechen. Wie sind die Befunde der Evaluation einzuordnen, wenn man die Herausforderungen durch das Pilotprojekt für die Schulen und die Erkenntnisse über die Umsetzung des Pilotprojekts in den Blick nimmt?

Für die Bewertung der Evaluationsergebnisse ist zum einen zu berücksichtigen, dass die Umsetzung der ALLE-Massnahmen an den beteiligten Schulen mit erheblichen Herausforderungen verknüpft war, beispielsweise mit der Rekrutierung und Einbindung geeigneter Schulassistenten, der Integration der Aufgabenhilfe in die bestehenden Stundenpläne oder der Bereitstellung zeitlicher Ressourcen für den Besuch von Weiterbildungen und Beratungsangeboten. Zudem resultierte für die Schulen ein erheblicher Koordinationsaufwand und die Notwendigkeit zur Einrichtung neuer Kommunikationskanäle und Kooperationsanlässe. Gerade bei der Implementation umfassender, anspruchsvoller Entwicklungsprogramme besteht die Gefahr, dass das Aufbrechen existierender Praktiken zunächst auch gut funktionierende Prozesse in Mitleidenschaft zieht und damit kurzfristig sogar negative Wirkungen mit sich bringen kann. Demgegenüber konnten an den ALLE-Schulen die fachlichen Lernergebnisse und die Motivation der Schülerinnen und Schüler zumindest in einzelnen Bereichen gesteigert werden. Bei einer noch länger angelegten Umsetzung der ALLE-Massnahmen liesse sich zudem erwarten, dass die genannten «disruptiven» – in der Eingangsphase potenziell ungünstigen – Effekte des Pilotprojekts weniger ins Gewicht fallen würden.

Wichtig für die Bewertung der Befunde ist zum anderen, was über die Umsetzung der Massnahmen an den ALLE-Schulen bekannt ist. Das Ausmass, in dem die Evaluationsergebnisse die möglichen Wirkungen der ALLE-Massnahmen widerspiegeln, ist davon abhängig, ob diese von den Schulen wie vorgesehen und in ihrer Gesamtheit umgesetzt wurden. Trotz gewisser Einschränkungen lässt

sich festhalten, dass sich aus der Evaluation keine Hinweise auf grundlegende Schwierigkeiten in den finanziellen und organisationalen Rahmenbedingungen ergaben. Demgegenüber fällt das Bild bezogen auf die konkrete Umsetzung der Massnahmen an den ALLE-Schulen heterogener aus. Betrachtet man die Aussagen der Beteiligten zu den einzelnen Massnahmen, wird deutlich, dass die Umsetzung von ALLE nicht immer mit der vollen erhofften Wirkung realisiert werden konnte (z.B. eingeschränkter Einsatz von Schulassistenten an einigen Schulen; eher geringe Inanspruchnahme der Beratungsangebote; Besuch der Aufgabenhilfe zum Teil nur durch wenige Schülerinnen und Schüler). Als Ursachen deuteten sich unter anderem begrenzte zeitliche und finanzielle Ressourcen, die teilweise gewährte Freiwilligkeit beziehungsweise geringe Verbindlichkeit in der Umsetzung der Massnahmen und vereinzelt auch Schwierigkeiten in der innerschulischen Zusammenarbeit an. Auch ergaben sich Anzeichen, dass Massnahmen nicht immer so verstanden und umgesetzt wurden, wie dies inhaltlich-konzeptionell seitens der ALLE-Projektleitung intendiert war oder aus Sicht von Expertinnen und Experten wünschenswert wäre (z.B. bei der Auswahl der Einsatzgebiete für die Schulassistenten).

Schliesslich wird an einigen Stellen Optimierungsbedarf bei den angebotenen Massnahmen selbst sichtbar. Dies betrifft grob unterteilt einerseits die wahrgenommene Kontinuität beziehungsweise Verschränktheit der Massnahmen über die Projektlaufzeit. So konnten beispielsweise die Beratungen nicht durchweg so gestaltet oder offeriert werden, dass diese an die vorher besuchten Weiterbildungen anknüpften und von den Lehrpersonen als hilfreich empfunden wurden. Andererseits lässt sich hinterfragen, ob die Verankerung von Verantwortlichkeiten an den Schulen immer in der notwendigen Breite und Verbindlichkeit erfolgte, oder beispielsweise ein weiter gefasster Einbezug von relevanten schulischen Akteuren hätte sichergestellt werden müssen.

Gegenüber möglichen Kritikpunkten ist hervorzuheben, dass ALLE als *Pilotprojekt* konzipiert war, und somit Erfahrungen – positive wie gegebenenfalls negative – bei der Umsetzung eines umfangreichen Schul- und Unterrichtsentwicklungsprogramms gesammelt werden sollten. Dass sich selbst bei nicht immer idealer Umsetzung Hinweise auf gewünschte Wirkungen ergaben, legt nahe, dass bei einer noch kohärenteren Implementierung der Massnahmen noch grössere Effekte erzielt werden könnten. Darüber hinaus muss betont werden, dass die Lehrpersonen, die an den jeweiligen Massnahmen Anteil hatten – und mehr noch die Schülerinnen und Schüler, die Schulleitungen und Schulassistenten – diese überwiegend positiv bis sehr positiv bewerteten.

Handlungsempfehlungen

Auf Basis der berichteten Ergebnisse werden folgende Handlungsempfehlungen formuliert:

- Erarbeitung von Schulentwicklungskonzepten, die mit den ALLE-Zielsetzungen übereinstimmen, und deren Umsetzung von den Schulleitungen gefördert und eingefordert wird;

- Sicherstellung der zeitlichen Ressourcen des pädagogischen Personals, um diesem ausreichend Gelegenheiten für die eigene professionelle Entwicklung zu geben, wie sie von den Massnahmen im Pilotprojekt (z.B. Weiterbildungen) intendiert ist;
- stärker verpflichtender Einsatz von Schulassistenzen, mit einer den fachlichen Voraussetzungen angemessenen Aufgabenteilung, ausreichenden Gelegenheiten zur Kommunikation und obligatorischer Weiterbildung zur Zusammenarbeit von Lehrpersonen und Schulassistenzen;
- Überprüfung der Konzeption der Aufgabenhilfe (u.a. der Zielsetzungen, der Einbettung in schulhausspezifische Angebote) und Erhöhung der Verbindlichkeit der Teilnahme, damit das Potenzial für eine verstärkte Individualisierung ausgeschöpft werden kann;
- Förderung eines Schulleitungshandelns nach dem Konzept der unterrichtswirksamen Führung, in dem hohe Erwartungen an die Lehrpersonen und Schülerinnen und Schüler mit der Etablierung klarer Lernziele und der Unterstützung eines lernförderlichen Unterrichtsklimas im Schulhaus verknüpft sind;
- gezieltere Nutzung der Expertise des Schulhausteams für die Unterrichtsentwicklung (z.B. für kollegiale Unterrichtsreflexionen) und Förderung eines dafür erforderlichen offenen, vertrauensvollen Schulhausklimas sowie von Überzeugungen im Kollegium, gemeinschaftlich erfolgreich handeln zu können («Collective Teacher Efficacy»);
- kontinuierliche, verbindlich eingeforderte und mit geeigneten Beratungsangeboten eng verknüpfte Weiterbildung von Lehrpersonen im Rahmen einer langfristig angelegten Personalentwicklungsstrategie der Schulen;
- Weiterführung und moderate Ausweitung von ALLE (z.B. auf andere Fächer; die Primarstufe) im Rahmen eines neuen, idealerweise über einen noch längeren Zeitraum angelegten Projekts, welches die in der Evaluation festgestellten Herausforderungen und Entwicklungspotenziale des Pilotprojekts systematisch berücksichtigt.

1 Einleitung

Bildungsstudien der zurückliegenden Jahre, wie die PISA-Erhebungen oder die Zürcher Lernstandserhebung aus dem Jahr 2012, weisen darauf hin, dass im Kanton Zürich ein nicht zu vernachlässigender Anteil der Schülerinnen und Schüler am Ende der Sekundarstufe I über unzureichende Kompetenzen in den Fächern Deutsch und Mathematik verfügt. Als Folge dieser Befunde erarbeitete das Volksschulamt (VSA) ab Juni 2014 ein Rahmenkonzept für das Pilotprojekt «Aktive Lernzeit und Lernerfolg für ALLE» (im Folgenden kurz ALLE), mit dem Ziel, die Leistungsentwicklung der Schülerinnen und Schüler zu fördern. Dieses Rahmenkonzept wurde im November 2014 vom Bildungsrat des Kantons Zürich genehmigt.

Die Ausschreibung der wissenschaftlichen Evaluation für das Pilotprojekt erfolgte im Februar 2015. Mit der Umsetzung der Evaluation wurde im Juni 2016 das Institut Professionsforschung & Kompetenzentwicklung der Pädagogischen Hochschule St.Gallen (PHSG) beauftragt. Ab September 2017 wurden die Arbeiten am Institut für Kompetenzdiagnostik der PHSG fortgeführt. Die Durchführung der Evaluation erfolgte in enger Kooperation mit der Projektleitung des Pilotprojekts beziehungsweise der Bildungsdirektion des Kantons Zürich. Der vorliegende Bericht fasst die Ergebnisse der wissenschaftlichen Evaluation des Pilotprojekts ALLE zusammen.

Das Pilotprojekt ALLE

Der Kern des Pilotprojekts ALLE war es, durch eine Weiterentwicklung des Unterrichts gezielt Einfluss auf die Qualität des Lehrens und Lernens zu nehmen, um so die Lernleistungen – insbesondere der leistungsschwächeren Schülerinnen und Schüler – zu verbessern. Wie dem Rahmenkonzept des Pilotprojekts entnommen werden kann, «sollen Erfahrungen und Erkenntnisse gewonnen werden, ob und inwiefern verschiedene Massnahmen und Interventionen die individuellen Lernleistungen zu steigern vermögen. Dabei wird das Augenmerk auf ein solides, kleinschrittiges und empirisch fundiertes Projekt- und Evaluationsdesign gelegt, mit klaren Rahmenvorgaben für die Umsetzung der Fördermassnahmen in den Pilotschulen.» (Bildungsdirektion Kanton Zürich/Volksschulamt, 2014, S. 4).

Bei den im Projektpilot implementierten Massnahmen handelte es sich im Einzelnen um:

- den Einsatz von Schulassistenten im Unterricht zur Unterstützung der Lehrpersonen, um die Lernwirksamkeit des Unterrichts zu erhöhen,
- die Integration einer Aufgabenhilfe für die Fächer Deutsch und Mathematik, um die Lernzeit der Schülerinnen und Schüler auszuweiten,
- die Durchführung von Weiterbildungen für Lehrpersonen und Schulleitungen,

- die Bereitstellung von Beratungen für die Lehrpersonen und
- eine kriterienorientierte, datenbasierte Unterrichtsreflexion der Lehrpersonen mittels EMU («Evidenzbasierte Methoden der Unterrichtsdiagnostik und -entwicklung»)¹

Eine genauere Darstellung der Umsetzung der Massnahmen kann dem Rahmenkonzept des Pilotprojekts ALLE entnommen werden (Bildungsdirektion Kanton Zürich/Volksschulamt, 2014).

Mit den genannten Massnahmen wurden insbesondere folgende Ziele verfolgt (vgl. ebd., S. 5):

- die individuelle Aufarbeitung grundlegender Defizite im Lernstand leistungsschwächerer Schülerinnen und Schüler im Unterricht und das Erzielen besserer Lernleistungen;
- die Steigerung der Lernmotivation und des Selbstvertrauens der leistungsschwächeren Schülerinnen und Schüler;
- die Orientierung der Lehrpersonen an Merkmalen lernwirksamen Unterrichts bei der Förderung der leistungsschwächeren Schülerinnen und Schüler;
- die Durchführung von Weiterbildungs- und Beratungsangeboten, die seitens der Lehrpersonen als wirksam und gewinnbringend für die Umsetzung im Unterricht erlebt werden;
- die Erfahrung der Zusammenarbeit von Lehrpersonen und Schulassistenzen im Unterricht als wirkungsvolle Entlastung und Unterstützung.

Das Pilotprojekt war auf die öffentliche Sekundarschule (Schulen der Sekundarstufe I ohne Gymnasien und private Sekundarschulen) beschränkt und umfasste die Teilnahme von neun Schulen. Die Umsetzung startete mit dem Schuljahr 2016/17 und endete am Ende des Schuljahrs 2018/19.

Theoretische Grundlagen und wissenschaftliche Verortung des Pilotprojekts

Im Pilotprojekt kam ein Massnahmenbündel zum Einsatz, das in seiner Gesamtheit dazu dienen sollte, über die Förderung professioneller Kompetenzen von Lehrpersonen und eine hohe Unterrichtsqualität den Lernerfolg der Schülerinnen und Schüler – insbesondere der leistungsschwächeren – in den Fächern Deutsch und Mathematik zu verbessern. Dieser Ansatz entspricht Erkenntnissen der empirischen Bildungsforschung, die zeigen, dass die Lehrperson und die Unterrichtsqualität bedeutende Grössen zur Beeinflussung des Lernerfolgs von Schülerinnen und Schülern sind (z.B. Hattie & Zierer, 2018; Lipowsky, 2006).

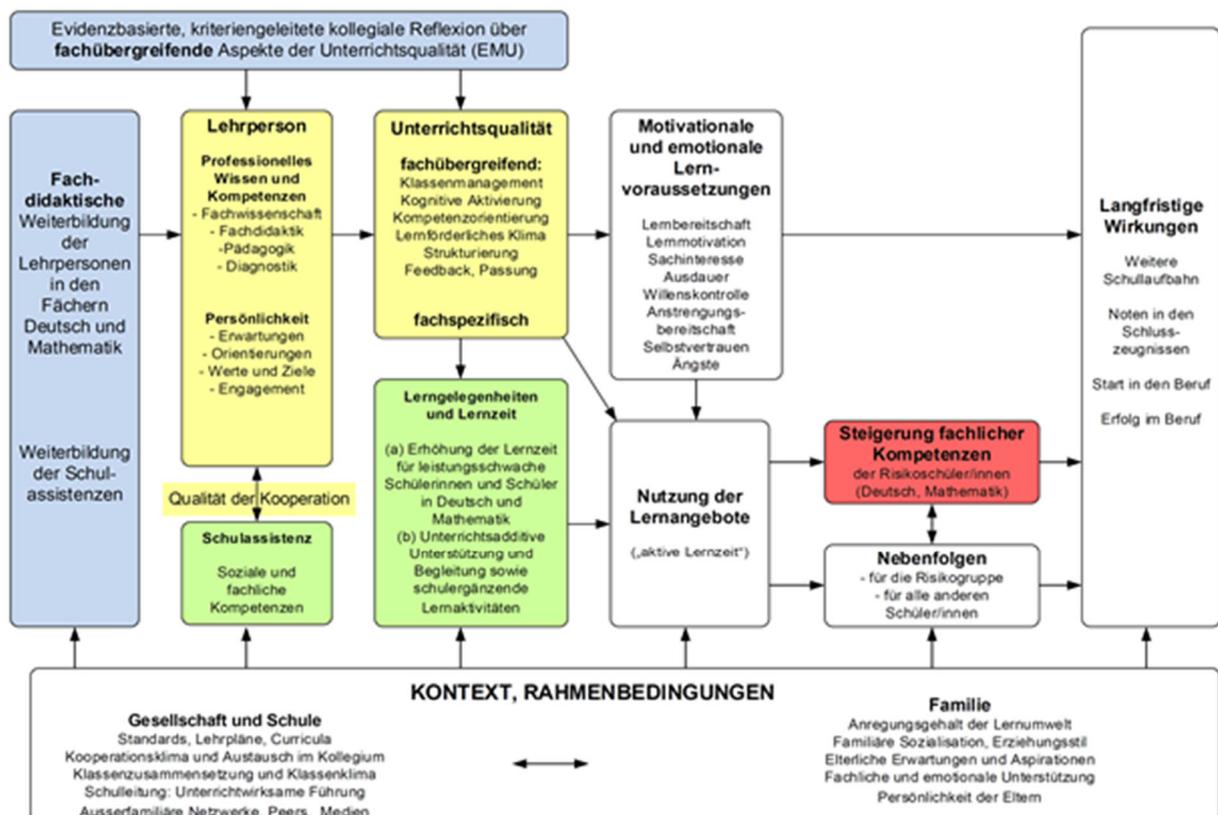
Dem Pilotprojekt lag ein heuristisches theoretisches Rahmenmodell zugrunde (Bildungsdirektion Kanton Zürich/Volksschulamt, 2014), in welchem vermutete Wirkmechanismen und Gelingensbe-

¹ Für weitere Informationen zu EMU siehe <http://unterrichtsdiagnostik.info>.

dingungen der verschiedenen Module systematisch dargestellt sind (vgl. Abbildung 1). Die angenommenen Wirkmechanismen sind in grüner Farbe, zentrale Gelingensbedingungen in gelber Farbe und der Einfluss der fachlichen und fachübergreifenden Weiterbildung in blauer Farbe veranschaulicht.

Diese konzeptuelle Grundlage folgt einem erweiterten Prozess-Produkt-Paradigma (Doyle, 1977; Winne, 1982) und integriert aktuelle Modelle des schulischen Lernens (Fend, 1998; Helmke, 2017; Krauss et al., 2004). Insbesondere berücksichtigt das Rahmenmodell (a) den Angebots-Nutzungs-Charakter von Lehr-Lernprozessen, (b) die hierarchische Organisation von Bildungsprozessen mit den Ebenen Weiterbildung, Lehrperson beziehungsweise Klasse/Unterricht sowie Schülerin und Schüler (Mehrebenenmodell), (c) die multiplen Bedingungsfaktoren und Kontextvariablen von Lehr-Lernprozessen, (d) Wechselwirkungen sowie (e) verschiedene Zielkriterien schulischen Lernens (z.B. Schülerleistungen in Deutsch und Mathematik, langfristige Wirkungen). Aus dem zu Grunde gelegten Rahmenmodell folgen verschiedene methodische Implikationen, die bei der Konzeption der Evaluation berücksichtigt wurden, wie beispielsweise die hierarchische Datenstruktur oder die Kontrolle alternativer Erklärungsgrößen.

Abbildung 1. Theoretisches Rahmenmodell im Pilotprojekt ALLE (Bildungsdirektion Kanton Zürich/ Volksschulamt, 2014)



Mit dem Pilotprojekt ALLE war diesem Modell folgend eine dreifache Wirksamkeitserwartung verbunden: Erstens sollten Weiterbildungs- und Beratungsangebote sowie Unterrichtsreflexionen (EMU) zum Aufbau professioneller Kompetenzen der Lehrpersonen (Professionalisierung) beitragen. Zweitens sollten die erworbenen Kompetenzen in der beruflichen Praxis zu einer hohen Unterrichtsqualität führen, die, unterstützt durch Schulassistenz und Aufgabenhilfe, drittens erfolgreiche Lernprozesse anstossen und so einen positiven Effekt auf den Lernerfolg der Schülerinnen und Schüler erbringen.

Ausgangspunkt und Anlage der Evaluation

Den oben skizzierten Wirksamkeitserwartungen entsprechend wurden in der Evaluation des Pilotprojekts drei Bereiche einer potenziellen Wirkung von ALLE fokussiert: die Wirkung auf die professionellen Kompetenzen der Lehrpersonen – in Orientierung an Baumert und Kunter (2006) wurden hierunter auch Überzeugungen und Werthaltungen der Lehrpersonen gefasst –, die Wirkung auf die (fachübergreifende) Qualität des Deutsch- und Mathematikunterrichts und die Wirkung auf die motivational-affektiven und fachlichen Lernergebnisse der Schülerinnen und Schüler. In diesen drei Bereichen wurden Zielvariablen bestimmt und mit standardisierten Kompetenztests und Fragebögen über insgesamt vier Messzeitpunkte erfasst. Sie stellen im vorliegenden Bericht die abhängigen Variablen dar, für welche mögliche Wirkungen des Pilotprojekts untersucht wurden.

Um die Wirkung des Pilotprojekts auf diese abhängigen Variablen zu bestimmen, wurde ein Vergleichsgruppendesign umgesetzt, das es ermöglichte, die Wirksamkeit des implementierten Massnahmenbündels durch einen Vergleich der Ergebnisse zwischen den am Pilotprojekt teilnehmenden Schulen und einer Vergleichsgruppe von Schulen aus dem Kanton Zürich zu untersuchen. Die Zugehörigkeit der Schulen zu der Gruppe der Schulen im Pilotprojekt gegenüber der Gruppe der Vergleichsschulen stellte im Rahmen der Evaluation die zentrale unabhängige Variable dar.

Neben der unmittelbaren Wirkung von ALLE auf die genannten abhängigen Variablen war schliesslich von grossem Interesse, wie die einzelnen Massnahmen an den Schulen des Pilotprojekts umgesetzt wurden, wie sie von den Beteiligten genutzt und hinsichtlich ihrer Verfügbarkeit, Nützlichkeit und Wirksamkeit eingeschätzt wurden. Um unterschiedliche Perspektiven berücksichtigen und einen vertieften Eindruck der Umsetzung der Massnahmen erhalten zu können, wurden mit allen zentralen Teilnehmendengruppen des Pilotprojekts – Lehrpersonen, schulische Heilpädagoginnen und Heilpädagogen, Schulassistenten, Schulleitungen und die Personen, welche im Pilotprojekt Beratungen und Weiterbildungen durchführten – Befragungen und Interviews durchgeführt.

Der vorliegende Bericht ist wie folgt gegliedert: In Kapitel 2 werden die Fragestellungen der Evaluation und die zugehörigen Datenquellen detailliert dargestellt. Kapitel 3 geht auf die methodischen Grundlagen der Evaluation ein. In Kapitel 4 erfolgt die Darstellung der Ergebnisse der Evaluation, die in Kapitel 5 zusammengefasst und bewertet werden.

2 Fragestellungen

Im Rahmenkonzept zum Pilotprojekt ALLE (Bildungsdirektion Kanton Zürich/Volksschulamt, 2014) werden die folgenden Kernfragen für eine wissenschaftliche Evaluation benannt:

- (1) Gelingt es den Lehrpersonen, ihren Unterricht lernwirksamer zu gestalten, insbesondere für die leistungsschwächeren Schülerinnen und Schüler?
- (2) Gelingt es, die Schulassistenten im Unterricht so einzusetzen, dass sich die Klassenlehrpersonen verstärkt den leistungsschwächeren Schülerinnen und Schülern zuwenden und sie lernwirksam fördern können?
- (3) Gelingt es den Lehrpersonen, die Aufgabenhilfe für den Kompetenzaufbau in Deutsch und Mathematik sowie für die Entwicklung wirksamer Lernstrategien zu nutzen?
- (4) Erreichen die Lernenden, insbesondere die leistungsschwächeren Schülerinnen und Schüler, einen höheren Lernfortschritt als Lernende einer Vergleichsgruppe mit herkömmlichem Unterricht?

Auf diesen Kernfragen aufbauend und unter Berücksichtigung der allgemeinen Konzeption des Pilotprojekts (vgl. Kap. 1) wurden in Abstimmung mit der Auftraggeberin die folgenden Fragebereiche für die wissenschaftliche Evaluation abgeleitet:

- Einsatz der Schulassistenten im Unterricht
- Einsatz der Aufgabenhilfe für die Fächer Deutsch und Mathematik
- Weiterbildungen für die Lehrpersonen und Schulleitungen
- Beratungen für die Lehrpersonen
- Einsatz der «Evidenzbasierten Methoden der Unterrichtsdiagnostik und -entwicklung» (EMU)
- Individualisierung im Fachunterricht
- (Fachübergreifende) Qualität des Unterrichts
- Professionelle Kompetenzen der Lehrpersonen
- Motivational-affektive und fachliche Lernergebnisse in Deutsch und Mathematik

Wie in Kapitel 1 erörtert wurde, kann die Weiterentwicklung der professionellen Kompetenzen der Lehrpersonen als wichtiger Ansatzpunkt verstanden werden, um die Lernwirksamkeit des Unterrichts und in der Folge das Lernen der Schülerinnen und Schüler zu verbessern. Daher wurde in Abstimmung mit der Auftraggeberin entschieden, diesen Bereich (noch) stärker zu berücksichtigen und die Veränderung der Überzeugungen sowie Aspekte der diagnostischen Kompetenz (personen- und aufgabenbezogene Urteilsakkuratheit) von Lehrpersonen durch das Pilotprojekt ALLE zu

untersuchen. Darüber hinaus wurde der ursprünglich in der Ausschreibung der Evaluation benannte Fragebereich zu «projektspezifischen Lernzielen» breiter gefasst und beinhaltet nun allgemeiner Aspekte einer stärkeren Individualisierung des Fachunterrichts. Dementsprechend sind die in Tabelle 1 formulierten Fragestellungen nicht vollständig in der dargestellten Form bearbeitet worden (vgl. die Anmerkungen zu Tabelle 1). Schliesslich wurde der in der Ausschreibung genannte Fragebereich «Eignung der förderorientierten Testinstrumente» in Abstimmung mit der Auftraggeberin gestrichen, da der Einsatz der Kompetenztests auf einen Vorschlag des Evaluations-teams hin grundsätzlich anders konzeptualisiert wurde als zunächst vorgesehen.

Alle oben genannten Fragebereiche mit den ihnen zugeordneten konkreten Fragestellungen sowie die verwendeten Datenquellen sind in Tabelle 1 aufgeführt. Die Reihenfolge der Darstellung entspricht dem Aufbau von Kapitel 4, in dem die Ergebnisse der Evaluation beschrieben werden. Zusätzlich zu den in Tabelle 1 aufgeführten Fragestellungen werden in Kapitel 4.11 die Ergebnisse einer explorativen Analyse berichtet, in der die Zusammenhänge zwischen Merkmalen der Umsetzung der ALLE-Massnahmen, der Unterrichtsqualität und den motivational-affektiven sowie fachlichen Lernergebnissen der Schülerinnen und Schüler untersucht wurden.

Tabelle 1. Fragebereiche mit Fragestellungen der wissenschaftlichen Evaluation und ihnen zugeordnete Datenquellen

Fragebereiche der Evaluation (mit zugehörigem Ergebniskapitel)	Fragestellungen	Datenquelle(n)				
		BP	LP	SA	SL	SuS
<i>Rahmenbedingungen des Pilotprojekts^a</i> (Kap. 4.1)	Wie gut gelingt es den Schulen, die für die Umsetzung der Massnahmen notwendigen finanziellen Mittel zu erhalten (Einsatz Schulassistenten, Angebot an Aufgabenhilfe, Weiterbildung, Beratung)?				x	
	Wie hat sich die Qualität der innerschulischen Kooperation, des kollegialen Austauschs über Unterricht etc. (siehe EMU) entwickelt?		x		x	
<i>Massnahmen im Pilotprojekt ALLE</i>						
Einsatz der Schulassistenten (Kap. 4.2)	Wie gut gelingt die Rekrutierung geeigneter Schulassistenten? ^b				x	
	Über welche Qualifikationen verfügen die Schulassistenten?			x		
	Wie und in welchem Ausmass werden die Schulassistenten im Unterricht in Deutsch und Mathematik eingesetzt?		x	x		
	Wie hoch ist die personale Konstanz bei den Schulassistenten über die Projektzeit hinweg?			x		
	Welche Rollenteilung besteht zwischen der Schulassistenten und der Lehrperson?		x	x		
	Wie ist die Qualität der Zusammenarbeit zwischen Schulassistenten und Lehrperson?		x	x		
	Wie nehmen die Lehrpersonen und die Schülerinnen und Schüler die Unterstützung durch die Schulassistenten wahr?		x			x
	Welche Veränderungen des Unterrichtsgeschehens haben sich durch den Beizug der Schulassistenten ergeben?		x			x

(Fortsetzung)

Tabelle 1 (Fortsetzung). Fragebereiche mit Fragestellungen der wissenschaftlichen Evaluation und ihnen zugeordnete Datenquellen

Fragebereiche der Evaluation (mit zugehörigem Ergebniskapitel)	Fragestellungen	Datenquelle(n)				
		BP	LP	SA	SL	SuS
Aufgabenhilfe in Deutsch und Mathematik (Kap. 4.3)	Wie gut gelingt die Integration der Aufgabenhilfe in den Stundenplan? ^b				x	
	Wie und in welchem Ausmass wird die Aufgabenhilfe für Deutsch und Mathematik eingesetzt?		x			x
	Wofür wird die Aufgabenhilfe eingesetzt?		x			
	Wer betreut die Aufgabenhilfe?		x	x		
	Wie kommt die Aufgabenhilfe bei den Lernenden an?					x
	Wird die Aufgabenhilfe als förderliche Unterstützung wahrgenommen?		x			x
Weiterbildung der Lehrpersonen (Kap. 4.4)	Welche Weiterbildungen wurden besucht?		x			
	Wie wird die Nützlichkeit der besuchten Weiterbildungen eingeschätzt?		x			
	Zu welchen Verbesserungen haben die Weiterbildungen beigetragen?		x			
Beratungen für Lehrpersonen (Kap. 4.5)	Wurde eine Beratung in Anspruch genommen?					
	Wenn ja, in welchem Ausmass, zu welcher Thematik?	x	x			
	War die Beratung hilfreich?	x	x			
	In welchen Bereichen hat die Beratung stattgefunden? ^c		x			
	Wie schätzen die Beraterinnen und Berater den Erfolg ihrer Beratung ein? ^c	x				

(Fortsetzung)

Tabelle 1 (Fortsetzung). Fragebereiche mit Fragestellungen der wissenschaftlichen Evaluation und ihnen zugeordnete Datenquellen

Fragebereiche der Evaluation (mit zugehörigem Ergebniskapitel)	Fragestellungen	Datenquelle(n)				
		BP	LP	SA	SL	SuS
Einsatz von EMU (Kap. 4.6)	Wie und in welchem Ausmass wird EMU eingesetzt? ^d		x		x	
	Welche Wirkungen lassen sich durch den Einsatz von EMU feststellen? ^d		x		x	
Individualisierung im Fachunterricht ^e (Kap. 4.7)	Sind die spezifischen Lernziele dem Leistungsstand der leistungsschwächeren Lernenden angepasst und decken sie die wichtigsten Lernbereiche ab?		x			
	Entsprechen die spezifischen Lernziele den Anforderungen der nachfolgenden Ausbildungen?		x			
<i>Professionelle Kompetenzen von Lehrpersonen^f</i> (Kap. 4.8)	Inwiefern verändern sich durch ALLE die Überzeugungen von Lehrpersonen zum Lehren und Lernen? ^f		x			
	Tragen die Massnahmen im Pilotprojekt ALLE zur Förderung von diagnostischen Kompetenzen (personen- und aufgabenbezogene Urteilsakkuratheit) der Lehrpersonen bei? ^f		x			
<i>Qualität des Unterrichts</i> (Kap. 4.9)	Wie nehmen Lernende den Unterricht hinsichtlich leistungsrelevanter Qualitätsmerkmale wahr?					x
	Wie verändert sich aus Sicht der Lernenden und der Lehrpersonen der Unterricht während des Projektverlaufs?		x			x
	Wie verändern sich lernrelevante Orientierungen und Motive seitens der Lernenden?					x

(Fortsetzung)

Tabelle 1 (Fortsetzung). Fragebereiche mit Fragestellungen der wissenschaftlichen Evaluation und ihnen zugeordnete Datenquellen

Fragebereiche der Evaluation (mit zugehörigem Ergebniskapitel)	Fragestellungen	Datenquelle(n)				
		BP	LP	SA	SL	SuS
<i>Lernergebnisse der Schülerinnen und Schüler</i> (Kap. 4.10)	Welchen Lernstand weisen die Lernenden auf bzw. welche Lernfortschritte zeigen sich? Sind die Aussagen zu differenzieren nach der Gruppe der Leistungsschwächeren und der Gruppe der Leistungsstärkeren?					x
	Sind die Lernfortschritte – insbesondere bei den Leistungsschwächeren – grösser als in einer leistungsmässig vergleichbaren Kontrollgruppe?					x

Anmerkungen. BP = Beratungs- und Weiterbildungspersonen, LP = Lehrpersonen, SA = Schülernassistenten, SL = Schulleitungen, SuS = Schülerinnen und Schüler. Die Fragestellungen beruhen in leicht gekürzter und in Teilen modifizierter Form auf der Ausschreibung der wissenschaftlichen Evaluation. Der Fragebereich «Eignung der förderorientierten Testinstrumente» wurde nicht bearbeitet (vgl. die Ausführungen im Text).

^a Der Begriff «Rahmenbedingungen» wird anstelle des in der Ausschreibung verwendeten Begriffs «Schulentwicklung» verwendet, da sich die diesem Fragenbereich zugeordneten Fragestellungen auf Bedingungen für die Umsetzung der zu implementierenden Massnahmen beziehen.

^b Die Fragestellungen waren in der Ausschreibung dem Fragebereich Rahmenbedingungen zugeordnet.

^c Die Fragestellungen waren in der Ausschreibung einem separaten Fragebereich zur Beratung aus Sicht der Beraterinnen und Berater zugeordnet.

^d Die Fragestellungen waren in der Ausschreibung dem Fragebereich Qualität des Unterrichts zugeordnet.

^e Der Begriff «Individualisierung im Fachunterricht» wird anstelle des in der Ausschreibung verwendeten Begriffs «Projektspezifische Lernziele» verwendet, da sich die durchgeführten Analysen allgemeiner auf die Umsetzung individueller Förderung beziehen. Entsprechend beziehen sich die berichteten Ergebnisse nur begrenzt bzw. weniger vertieft auf die Anpassung von Lernzielen. Die Fragestellung «Entsprechen die spezifischen Lernziele den Anforderungen der nachfolgenden Ausbildungen?» wurde (in Abstimmung mit der Auftraggeberin) nicht bearbeitet.

^f Der Fragebereich und die dazu gehörenden Fragestellungen wurden gegenüber der Ausschreibung ergänzt.

3 Methodische Grundlagen

In diesem Kapitel werden die methodischen Grundlagen der Evaluation des Pilotprojekts ALLE erläutert. Kapitel 3.1 beschreibt das Untersuchungsdesign der Evaluation, Kapitel 3.2 geht auf die Durchführung der Datenerhebungen im Rahmen der Evaluation ein. Kapitel 3.3 beschreibt die Stichproben, die den in Kapitel 4 berichteten Analysen zugrunde liegen. In Kapitel 3.4 werden die eingesetzten Erhebungsinstrumente dargestellt. Kapitel 3.5 beschreibt die Aufbereitung der Daten und das Vorgehen bei den Datenanalysen, Kapitel 3.6 gibt abschliessend Hinweise, welche für die Leserinnen und Leser bei der Interpretation der Ergebnisse hilfreich sein können.

3.1 Untersuchungsdesign

Die wissenschaftliche Evaluation des Pilotprojekts ALLE war als quasi-experimentelle Längsschnittstudie mit vier Messzeitpunkten angelegt. Datenerhebungen erfolgten zu Beginn der 7. Klassenstufe (Schuljahr 2016/2017) sowie jeweils am Ende der 7., 8. und 9. Klassenstufe (Abschluss: Schuljahr 2018/2019). Der Ausschreibung zur Evaluation folgend wurde ein Vergleichsgruppen-design realisiert, welches es ermöglicht, die besonders bedeutsame Frage der Wirksamkeit des implementierten Massnahmenbündels durch einen Vergleich zwischen den Ergebnissen der am Pilotprojekt ALLE teilnehmenden Schulen (im Folgenden: ALLE-Schulen) und einer Vergleichsgruppe ähnlicher Grösse (im Folgenden: Vergleichsschulen) zu untersuchen. Die Evaluation folgt mit ihrer Zusammenführung quantitativer und qualitativer Daten einem Mixed-Methods-Ansatz.

Tabelle 2 gibt einen Überblick über die Erhebungsinhalte und -instrumente zu den verschiedenen Messzeitpunkten, geordnet nach den zentralen Zielbereichen der Evaluation (Rahmenbedingungen des Pilotprojekts; Massnahmen des Pilotprojekts; professionelle Kompetenzen der Lehrpersonen; Unterrichtsqualität; Lernergebnisse der Schülerinnen und Schüler). Die Erhebungen im Rahmen der Evaluation umfassten Kompetenztests mit den Schülerinnen und Schülern, Online-Befragungen der am Pilotprojekt beteiligten Lehrpersonen, Schulassistenten, Schulleitungen und Schülerinnen und Schüler, die jeweils computergestützt und online durchgeführt wurden. Im Regelfall wurden dabei in den ALLE- und Vergleichsschulen dieselben Test- und Fragebogeninstrumente eingesetzt. Die Befragungen zu den Rahmenbedingungen und Massnahmen des Pilotprojekts erfolgten jedoch nur mit den Lehrpersonen, Schulassistenten, Schulleitungen und Schülerinnen und Schülern, die am Pilotprojekt teilnahmen. Ergänzt wurden diese Befragungen um Einzel- beziehungsweise Gruppeninterviews mit den am Pilotprojekt beteiligten Lehrpersonen, schulischen Heilpädagoginnen und Heilpädagogen, Schulassistenten, Schulleitungen und den Personen, welche im Pilotprojekt Beratungen und Weiterbildungen durchführten (vgl. Tabelle 2).

Tabelle 2. Übersicht zum Untersuchungsdesign der Evaluation ALLE: Erhebungsinhalte, Erhebungsinstrumente sowie Messzeitpunkte nach den Zielbereichen der Evaluation

Erhebungsinhalte und -instrumente	Messzeitpunkte															
	T1 (2016) 7. Klasse				T2 (2017) 7./8. Klasse				T3 (2018) 8. Klasse				T4 (2019) 9. Klasse			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
<i>Rahmenbedingungen Pilotprojekt^a</i>																
Fragebogen SL														■		
Einzelinterviews SL														■	■	
<i>Massnahmen im Pilotprojekt^a</i>																
Fragebogen LP														■		
Fragebogen SA														■		
Fragebogen SL														■		
Fragebogen SuS														■	■	
Einzelinterviews BP														■	■	
Einzelinterviews SA														■	■	
Einzelinterviews SL														■	■	
Gruppeninterviews LP/SHP														■	■	
<i>Professionelle Kompetenzen LP</i>																
Fragebogen LP (Überzeugungen/Urteilsakkuratheit)				■										■	■	
<i>Qualität des Unterrichts</i>																
Fragebogen LP				■										■	■	
Fragebogen SuS				■						■	■			■	■	
<i>Lernergebnisse SuS</i>																
Fragebogen SuS (motivational-affektive Merkmale)			■							■	■			■	■	
Kompetenztests D/M SuS ^b			■			■				■	■			■	■	
Test Texte Schreiben SuS								■								

Anmerkungen. Q1, Q2, Q3, Q4 = Jahresquartale. D/M = Deutsch und Mathematik. BP = Beratungs- und Weiterbildungspersonen, LP = Lehrpersonen, SHP = Schulische Heilpädagoginnen und Heilpädagogen, SA = Schulassistenzen, SL = Schulleitungen, SuS = Schülerinnen und Schüler. Alle Tests und Fragebögen wurden als Online-Instrumente erstellt und administriert.

^a Diese Erhebung wurde nur in den ALLE-Schulen durchgeführt.

^b Die Erhebung zu T3 wurde nur in den ALLE-Schulen durchgeführt.

Für die Evaluation war die Entwicklung der fachlichen Lernergebnisse der Schülerinnen und Schüler zentral. Daher wurde an allen Messzeitpunkten der individuelle Lernstand der Schülerinnen und Schüler in Deutsch und Mathematik mit computergestützten Kompetenztests erhoben (vgl. Tabelle 2), aufbauend auf den Aufgabenpools und Datenerhebungen für das Lernfördersystem *Lernpass plus* (vgl. Kap. 3.4.1). Während die ALLE-Schulen an allen vier Messzeitpunkten die Kompetenztests absolvierten, wurde in den Vergleichsschulen aufgrund der hohen Belastung der Schulen auf die Messung in der 8. Klassenstufe (T3; Schuljahr 2017/2018) nach Absprache mit der Auftraggeberin verzichtet. Zusätzlich für die Evaluation genutzt werden konnten die Ergebnisse des *Stellwerk*-Tests «Texte schreiben», der in der 8. Klassenstufe (November 2017) durchgeführt wurde (vgl. Kap. 3.4.1). Hinzu kamen mithilfe von Online-Fragebögen erfasste motivational-affektive Lernergebnisse der Schülerinnen und Schüler, die in den ALLE- und Vergleichsschulen am Anfang der 7., in der 8. und in der 9. Klassenstufe (T1, T3 und T4) erhoben wurden. Ergänzend in diesen Befragungen erfasst wurden wichtige lern- und leistungsrelevante Hintergrundmerkmale der Schülerinnen und Schüler (an T1 und T4). Dazu gehören insbesondere das Geschlecht, die soziale Herkunft sowie der Sprach- und Migrationshintergrund der Schülerinnen und Schüler. Diese Merkmale werden zur Beschreibung der Schülerstichproben herangezogen (vgl. Kap. 3.3.2.1); sie dienen zudem der statistischen Kontrolle von Ausgangsunterschieden der ALLE- und Vergleichsschulen, da keine randomisierte Zuweisung der Schulen zum Pilotprojekt beziehungsweise zur Vergleichsgruppe erfolgte (vgl. Kap. 3.3.1).

3.2 Durchführung der Datenerhebungen

Im Folgenden wird zunächst auf die Vorbereitung der Erhebungen an den Schulen eingegangen (Kap. 3.2.1). Darüber hinaus erfolgt die Beschreibung der Durchführungsmodalitäten der Kompetenztests sowie Befragungen (Kap. 3.2.2) und der Interviews (Kap. 3.2.3).

3.2.1 Vorbereitungen an den Schulen

In jeder der teilnehmenden ALLE- und Vergleichsschulen wurde eine Ansprechperson ernannt, die für die Kommunikation mit dem Evaluationsteam und die Organisation aller Aktivitäten innerhalb der Schule im Zusammenhang mit dem Pilotprojekt und seiner Evaluation verantwortlich war. Mit wenigen Ausnahmen übernahm die Schulleitung diese Aufgabe. Den Ansprechpersonen wurde jeweils vor den einzelnen Erhebungen detailliertes Informationsmaterial zur Verfügung gestellt, das die notwendigen Arbeiten zur Vorbereitung der Erhebung beschrieb. So unterstützten die Ansprechpersonen das Evaluationsteam bei der Erfassung der Hintergrundmerkmale der Schülerinnen und Schüler (z.B. Geburtsdatum, Geschlecht) und der Kontextmerkmale der Schulen und Klassen, insbesondere bei der Zuordnung der Schülerinnen und Schüler zu den Abteilungen beziehungsweise

Anforderungsstufen (Niveaus) sowie zu den Klassen beziehungsweise fachspezifischen Lerngruppen und deren Fachlehrpersonen. Ebenso unterstützten sie das Evaluationsteam bei der Durchführung der Erhebungen innerhalb der Schulen, indem sie sicherstellten, dass die technischen Voraussetzungen erfüllt und die Vorbereitungen durchgeführt wurden (z.B. Erfassung der Schülerinnen und Schüler auf den genutzten technischen Plattformen, vgl. Kap. 3.2.2; Bereitstellung notwendiger Geräte und Materialien). Zudem informierten sie die Lehrpersonen und weitere Teilnehmende der Evaluation über die Zeitfenster und Durchführungsmodalitäten der Erhebungen (z.B. Generierung der Logins für die Kompetenztests und Befragungen) und sorgten für die Information der teilnehmenden Schülerinnen und Schüler und deren Eltern. Während der Vorbereitungs- und Erhebungsphasen standen Mitarbeiterinnen und Mitarbeiter des Evaluationsteams telefonisch und per Email zur Verfügung, um die Ansprechpersonen und gegebenenfalls die Teilnehmenden selbst unmittelbar unterstützen zu können.

3.2.2 Durchführung der Kompetenztests und Befragungen

Kompetenztests und Befragungen mit den Schülerinnen und Schülern

Die Kompetenztests in Deutsch und Mathematik wurden an allen vier Messzeitpunkten durchgeführt (zum dritten Messzeitpunkt nur in den ALLE-Schulen; vgl. Kap. 3.1):

- Erster Messzeitpunkt (T1; Anfang 7. Klassenstufe): September 2016
- Zweiter Messzeitpunkt (T2; Ende 7. Klassenstufe): Mai 2017
- Dritter Messzeitpunkt (T3; Ende 8. Klassenstufe): Juni/Juli 2018
- Vierter Messzeitpunkt (T4; Ende 9. Klassenstufe): Mai/Juni 2019

Den genauen Zeitpunkt der Testdurchführung mit ihrer Klasse/Lerngruppe konnten die Lehrpersonen an jedem Messzeitpunkt innerhalb eines festgelegten Zeitraums selbst festsetzen. Dieser Zeitraum umfasste zum ersten Messzeitpunkt (T1) zunächst drei Wochen und wurde für die weiteren Messzeitpunkte (T2 bis T4) auf vier bis sechs Wochen verlängert, um den Schulen mehr Flexibilität bei der Terminplanung zu gewähren.

Die Durchführung der Kompetenztests in Deutsch und Mathematik erfolgte auf der technischen Plattform des Lernfördersystems *Lernpass plus*², welche von dem Unternehmen *Arcadix* entwickelt und bereitgestellt wird (siehe auch Kap. 3.4.1). Die Schulen mussten an allen Messzeitpunkten sicherstellen, dass die Schülerinnen und Schüler im System erfasst und korrekt den Klassen/Lerngruppen beziehungsweise Lehrpersonen zugeordnet waren. Dabei war besonders wichtig, den

² Für weitere Informationen zu Lernpass plus siehe die Webseite <https://lernpassplus.ch>.

Code zu erfassen, der jeder Schülerin/jedem Schüler vom Evaluationsteam zugewiesen und den Ansprechpersonen zur Verfügung gestellt wurde (vgl. Kap. 3.2.1), um dadurch die Identifizierung der Schülerinnen und Schüler zu ermöglichen. Darüber hinaus mussten die Lehrpersonen zu T1, T3 und T4 eine Niveauzuordnung – Niveau A/I, Niveau B/II, Niveau C/II – der Schülerinnen und Schüler im System vornehmen (jeweils für Deutsch und Mathematik), sodass die Schülerinnen und Schüler je nach ihrem Niveau das passende Testheft zugeteilt bekamen. Für die Zuordnung war das Niveau der besuchten Klasse/Lerngruppe ausschlaggebend. Beispielsweise wurden Schülerinnen und Schüler einer A-Klasse im Niveau II dem Niveau B/II zugeordnet, Schülerinnen und Schüler einer B-Klasse im Niveau I dem Niveau A/I, Schülerinnen und Schüler einer C-Klasse im Niveau II dem Niveau B/II. Am Testtag konnten sich die Schülerinnen und Schüler über einen für den individuellen Test erzeugten Code einloggen und den Test bearbeiten. Das Vorgehen zur Vorbereitung und Durchführung der Tests war den beteiligten Lehrpersonen im Allgemeinen gut bekannt, da die technische Plattform von Lernpass plus auch für die im Kanton Zürich obligatorischen Standortbestimmungen (*Stellwerk*) genutzt wird. Die Testdurchführung im Rahmen der Evaluation fand in der Verantwortung der jeweiligen Lehrperson (Fach- oder Klassenlehrperson) statt. In Absprache mit der Auftraggeberin wurden keine Testadministratorinnen beziehungsweise Testadministratoren an den Schulen eingesetzt, da dies mit zu hohem Aufwand verbunden gewesen wäre. Die Daten wurden von Arcadix für die weitere Bearbeitung exportiert und dem Evaluationsteam zur Verfügung gestellt.

Bei den Befragungen konnten die Lehrpersonen ebenfalls an jedem Messzeitpunkt selbst die Durchführung innerhalb eines vorgegebenen Zeitraums planen. Den Lehrpersonen standen während des Erhebungszeitfensters der Link zur Online-Befragung sowie die individuellen Login-Daten der Schülerinnen und Schüler zur Verfügung. Die Schülerinnen und Schüler konnten sich am Erhebungstag mit einem für den individuellen Fragebogen erstellten Benutzernamen und einem Passwort einloggen und den Fragebogen bearbeiten. Der jeweils zur Verfügung gestellte Zeitraum war weitgehend der gleiche wie bei den Kompetenztests. Eine Ausnahme war die Erfassung der Unterrichtsqualität zum ersten Messzeitpunkt. Die Schülerinnen und Schüler wurden zu ihrem Unterricht in Deutsch und Mathematik zwischen November 2016 und Februar 2017 (mit einer nachträglich gewährten Verlängerung von zwei Monaten) befragt (vgl. Tabelle 2). Der spätere Zeitpunkt dieser Befragung wurde bewusst gewählt, da die Klassen Anfang der 7. Klassenstufe (nach dem Wechsel in die Sekundarstufe I) neu zusammengestellt worden waren, die Schülerinnen und Schüler also erst ausreichend Gelegenheit haben mussten, den Unterricht ihrer Lehrperson kennenzulernen. Weiterhin wurde die Erhebung zum dritten Messzeitpunkt für einzelne Klassen auf Wunsch der Lehrpersonen in Absprache mit der Auftraggeberin verlängert und auf Mitte/Ende August 2018 (unmittelbar nach den Sommerferien der 8. Klasse bzw. zu Beginn der 9. Klasse) verschoben.

An allen Messzeitpunkten wurden die Befragungen computergestützt mithilfe des Online-Tools *Artologik* durchgeführt.³ Dieses Vorgehen ermöglichte die Erstellung leicht unterschiedlicher Fragebogenversionen für die ALLE- und Vergleichsschulen, wenn Fragen betreffend die ALLE-Massnahmen an die Schülerinnen und Schüler der ALLE-Schulen gestellt werden sollten. Beispielsweise wurden die Fragen zur Aufgabenhilfe und zur Schulassistenz zum letzten Messzeitpunkt mithilfe von Filterfragen am Ende des jeweils fachspezifischen Anteils des Fragebogens nur den Schülerinnen und Schülern der ALLE-Schulen vorgelegt.

Befragungen der Lehrpersonen, der Schulassistenten und der Schulleitungen

Die Befragungen der Lehrpersonen in den ALLE- und Vergleichsschulen sowie der Schulleitungen und Schulassistenten der ALLE-Schulen wurden ebenfalls über das Online-Tool *Artologik* durchgeführt. Zu Beginn des Erhebungszeitraums wurden den Befragten der Link zur Online-Befragung und ihre individuellen Zugangsdaten per E-Mail direkt über die Online-Plattform des Datenerhebungsprogramms zugesendet. Vorabinformationen zu den Erhebungsinhalten und den Durchführungsbedingungen wurden von den Ansprechpersonen der Schulen zur Verfügung gestellt.

Die Befragungen der Lehrpersonen in den ALLE- und Vergleichsschulen zum ersten und vierten Messzeitpunkt umfassten einen Fragebogen zu ihren fach- beziehungsweise domänenspezifischen Überzeugungen und zu ihrem Unterricht (vgl. Kap. 3.4.2). Zusätzlich wurde ein zweiter Fragebogen verwendet, um die Lehrpersonen zu befragen, wie ihre Schülerinnen und Schüler in den Kompetenztests der Evaluation ALLE abschneiden würden (vgl. Kap. 4.8.2). Während der vorgesehenen Zeitraum für den vierten Messzeitpunkt für beide Fragebögen derselbe war wie für die Kompetenztests und die Befragung der Schülerinnen und Schüler, wurden die beiden Fragebögen für den ersten Messzeitpunkt zwischen Anfang Dezember und Mitte Januar 2016/2017 administriert. Für den Fragebogen zu den Überzeugungen und zum Unterricht der Lehrpersonen wurden mithilfe des Online-Ehebungstools leicht unterschiedliche Fragebogenversionen für die Deutsch- und Mathematik-Lehrpersonen erzeugt, sodass eine fach- beziehungsweise domänenspezifische Erfassung der erwähnten Merkmale erfolgen konnte.

Um die Einschätzungen der Lehrpersonen zu erfassen, wurde für jede Lehrperson und jedes Fach ein einziger, individualisierter Fragebogen erstellt und verwaltet. Die Daten aus den einzelnen Befragungen wurden einzeln exportiert und bei der Datenaufbereitung zusammengeführt. Die Erstellung von individualisierten Fragebögen war notwendig, da die Lehrpersonen die Leistung jeder

³ Für weitere Informationen zu Artologik siehe <https://www.artologik.com/de/UmfrageSoftware.aspx>.

einzelnen Schülerin/jedes einzelnen Schülers beurteilen sollten. Darüber hinaus konnten so die erforderlichen Bilder und Audiodateien aus den Kompetenztests integriert werden.

Die Befragungen der Lehrpersonen, der Schulassistenzen und der Schulleitungen in den ALLE-Schulen zur Umsetzung der Massnahmen im Pilotprojekt (vgl. Tabelle 2) fanden zwischen Mitte Januar und Mitte März 2019 statt, wobei alle Schulassistenzen den Fragebogen bereits bis Mitte Februar bearbeitet hatten. Die Befragungen der Lehrpersonen und der Schulassistenzen waren ebenfalls fach- beziehungsweise domänenspezifisch konzipiert und wurden entsprechend administriert. Lehrpersonen und Schulassistenzen, die in Deutsch und Mathematik Schülerinnen und Schüler unterrichteten beziehungsweise im Unterricht beider Fächer eingesetzt wurden, bekamen die fach- beziehungsweise domänenspezifischen Fragen für beide Fächer vorgelegt.

3.2.3 Durchführung der Interviews

Die Befragungen der Lehrpersonen, Schulleitungen und Schulassistenzen in den ALLE-Schulen zur Umsetzung und Wirkung der erprobten Massnahmen wurden um Interviews ergänzt. Die Interviews fanden nach Absprache mit den beteiligten Personen im Zeitraum vom Anfang Juni bis Mitte Juli 2019 statt (vgl. Kap. 3.1), das heisst im Rahmen des vierten Messzeitpunkts gegen Ende der Evaluation. Mit den Fachlehrpersonen und den schulischen Heilpädagoginnen und Heilpädagogen wurden Gruppeninterviews durchgeführt, mit den Schulleitungen, Schulassistenzen sowie mit den Beratungs- und Weiterbildungspersonen Einzelinterviews. Die Gruppeninterviews fanden immer in der jeweiligen Schule statt. Vor oder nach einem Gruppeninterview wurden nach Möglichkeit die Einzelinterviews mit der Schulleitung und/oder der Schulassistentz der jeweiligen Schule durchgeführt. Wenn es nicht möglich war, das Interview in den jeweiligen Schulen durchzuführen (z.B. aufgrund terminlicher Unvereinbarkeiten), wurde auf ein telefonisches Interview ausgewichen. Mit den Beratungs- und Weiterbildungspersonen fanden ausschliesslich telefonische Interviews statt. Alle Interviews wurden durch wissenschaftliche Mitarbeitende des Evaluationsteams geführt. Für die Gruppeninterviews wurde neben der interviewenden Person jeweils eine zweite Person als Begleitperson engagiert, die sich auf die Bedienung des Aufzeichnungsgeräts konzentrierte, darauf achtete, welche Punkte im Interviewleitfaden noch nicht angesprochen worden waren und die interviewende Person gegebenenfalls darauf hinwies, und Auffälligkeiten notierte.

Am Tag des Interviews wurde nach der Begrüssung der Befragten ein einleitender Text vorgelesen, der über folgende Aspekte des anschliessenden Interviews informierte:

- Allgemeine Informationen über die Evaluation ALLE (Erläuterung der Massnahmen und der Ziele des Pilotprojekts sowie der Ziele der Interviews),
- Informationen zum Datenschutz,
- Informationen zur Aufnahme mithilfe eines Audio-Recorders,
- Erläuterungen zum Ablauf des Interviews.

Zusätzlich wurden in den Gruppeninterviews folgende Themen angesprochen:

- Wünsche und Erwartungen an die Teilnehmenden (z.B. Diskussion zwischen den Teilnehmenden ist erwünscht; Fragen sind immer an die ganze Gruppe zu richten),
- Gemeinsamkeiten in der interviewten Gruppe (z.B. Teilnahme am Pilotprojekt),
- Gesprächsregeln (z.B. Melden per Handzeichen und Nennung der den Befragten zugewiesenen Nummer, um bei der Transkription eine Zuordnung zu ermöglichen).

Anschliessend wurden die Interviews auf Basis eines halbstandardisierten Gesprächsleitfadens durchgeführt (siehe Kap. 3.4.3) und mithilfe eines Audio-Recorders aufgenommen.

3.3 Datengrundlage

In diesem Kapitel wird auf die im Rahmen der Evaluation ALLE untersuchten Stichproben aus den ALLE- und Vergleichsschulen eingegangen. In Kapitel 3.3.1 wird beschrieben, wie die am Pilotprojekt beziehungsweise an der Evaluation teilnehmenden Schulen gewonnen wurden. Zudem wird auf die Gewinnung der Teilnehmenden für die Interviews eingegangen. In Kapitel 3.3.2 wird die Datengrundlage näher beschrieben, welche den Analysen im vorliegenden Bericht zugrunde lag.

3.3.1 Gewinnung der Stichproben für die Evaluation ALLE

Rekrutierung der ALLE- und Vergleichsschulen

An der Evaluation ALLE nahmen neun Schulen der öffentlichen Sekundarschule teil, die am Pilotprojekt beteiligt waren und die darin vorgesehenen Massnahmen realisierten (ALLE-Schulen), sowie neun Schulen der öffentlichen Sekundarschule, die als Vergleichsschulen nicht am Pilotprojekt beteiligt waren (Vergleichsschulen). Die Gewinnung der Schulen, die am Pilotprojekt teilnahmen, erfolgte bereits vor der Evaluation über eine Ausschreibung der Bildungsdirektion des Kantons Zürich. Das folgende Zitat beschreibt die Voraussetzungen und Kriterien der Auswahl der ALLE-Schulen (Bildungsdirektion Kanton Zürich/Volksschulamt, 2016):

«Voraussetzung für die Teilnahme ist das Interesse an innerschulischer Kooperation, die Wertschätzung für den datenbasierten, kollegialen Austausch über Unterrichtsqualität, verbunden mit der Bereitschaft und Offenheit der beteiligten Lehrpersonen, die eigene Unterrichtspraxis zu überdenken und weiterzuentwickeln.»

Die Teilnahme am Pilotprojekt erfolgt nach folgenden Kriterien:

- *Teilnahme als Schuleinheit auf der Sekundarstufe;*
- *Zustimmung der Schulkonferenz und der Schulpflege;*
- *Umsetzung des Pilotprojekts auf der Grundlage des kantonalen Rahmenkonzepts;*
- *Einsetzung einer lokalen Projektorganisation in Verantwortung der Schulleitung;*

- *Mitwirkung der Schulleitung in der kantonalen Leitungsgruppe;*
- *Teilnahme an der Weiterbildung und Erprobung der Massnahmen im Unterricht;*
- *Teilnahme an den schulinternen und externen Evaluationsmassnahmen.*

Von Vorteil sind Erfahrungen in den Bereichen Aufgabenhilfe und/oder Schulassistenz.

Die Teilnahme am Pilotprojekt ist als Schuleinheit vorgesehen, d.h. für alle Klassen der jeweiligen Klassenstufe und nicht nur auf eine einzelne Abteilung oder Klasse beschränkt.»

Die ALLE-Schulen meldeten sich freiwillig für die Teilnahme und wurden von der Projektleitung des Pilotprojekts ausgewählt. Aufgrund dieses Auswahlprozesses handelt es sich bei den ALLE-Schulen um eine Gelegenheitsstichprobe von Schulen, für die keine Repräsentativität für die Grundgesamtheit der Sekundarschulen im Kanton Zürich angenommen werden kann.

Die Rekrutierung der Vergleichsschulen erfolgte in der ersten Phase der Evaluation ALLE. Die Zielsetzung sollte zunächst sein, Schulen für eine Teilnahme an der Evaluation zu gewinnen, die (a) an keiner der Massnahmen im Pilotprojekt partizipieren und (b) als gesamte Stichprobe eine möglichst hohe Vergleichbarkeit mit den ALLE-Schulen aufweisen. In Absprache mit der Auftraggeberin wurde dieses Vorgehen nochmals modifiziert. Der zunächst vorgesehene Vergleich zwischen den ALLE-Schulen und Schulen, die keine der Massnahmen durchführen, hätte keine Rückschlüsse auf die Wirkung spezifischer Massnahmen, sondern nur auf die Wirkung des gesamten ALLE-Massnahmenbündels erlaubt. Da ein Interesse bestand, die Wirkung zumindest einer Massnahme – der «Evidenzbasierten Methoden der Unterrichtsdiagnostik und –entwicklung» (EMU) – näher zu untersuchen, wurde entschieden, zwei Arten von Vergleichsschulen (je vier bis fünf Schulen) zu rekrutieren: Schulen, die *keine* der Massnahmen im Pilotprojekt durchführen, und Schulen, die *nur EMU* durchführen (als Teil ihrer bestehenden Unterrichtspraxis, soweit dies der Projektleitung für die einzelnen Schulen bekannt war).

Für die Analysen im vorliegenden Bericht wird jedoch – anders als zunächst vorgesehen – nicht zwischen Schulen differenziert, die nur EMU verwenden (EMU-Schulen), und solchen, die keine der Massnahmen im Pilotprojekt umsetzen. Stattdessen werden die teilnehmenden Schulen beiden Typs zu einer gemeinsamen Gruppe als Vergleichsschulen zusammengefasst. Begründet ist dies dadurch, dass im Verlauf der Evaluation deutlich wurde, dass die ausgewählten EMU-Schulen das Instrument sehr unterschiedlich und teilweise nur sporadisch einsetzten. Aufgrund des gewählten Untersuchungsdesigns, das detaillierte Erhebungen zur Umsetzung von Massnahmen nur in den ALLE-Schulen vorsah (vgl. Kap. 3.1), konnte der tatsächliche Einsatz von EMU in den EMU-Schulen zudem nicht hinreichend nachvollzogen werden. Vor diesem Hintergrund wird im vorliegenden Bericht in Absprache mit der Auftraggeberin auf eine Unterscheidung zwischen EMU-Schulen und sonstigen Vergleichsschulen verzichtet.

Da der Kernbestandteil der Evaluation ALLE – die Prüfung der Wirkungen der implementierten Massnahmen (vgl. Kap. 1) – auf dem Vergleich von Lernergebnissen in den ALLE-Schulen mit Lernergebnissen in den Vergleichsschulen beruht, sollten sich die Vergleichsschulen in Bezug auf lern- und leistungsrelevante Merkmale möglichst wenig von den ALLE-Schulen unterscheiden. Entsprechend wurden im Kanton Zürich Schulen der öffentlichen Sekundarschule gesucht, die den ALLE-Schulen hinsichtlich relevanter Hintergrundmerkmale (z.B. soziodemografische Zusammensetzung der Schülerschaft) ähnlich waren und eine vergleichbare Anzahl Schülerinnen und Schüler aufwiesen. Um diesbezüglich passende Schulen identifizieren zu können, wurden dem Evaluationsteam von der Bildungsdirektion Zürich relevante Begleitdaten der Schulen zur Verfügung gestellt. Daneben wurden Datensätze des Bundesamts für Statistik (BfS) einbezogen, in denen die Zugehörigkeit zu den wichtigsten räumlichen Gliederungen und Typologien auf Gemeindeebene zusammengefasst sind (Raumgliederungen). Basierend auf vorbereitenden Analysen und in Absprache mit der Auftraggeberin wurden für die Auswahl der Vergleichsschulen drei Kriterien berücksichtigt: der Teilnahmestatus am Programm QUIMS (Qualität in multikulturellen Schulen)⁴ des Kantons Zürich, der Grad der Urbanisierung der Schulgemeinde (*degree of urbanisation*; DEGUR-BA) aus den Statistiken des BfS sowie die Differenzierung nach Abteilungen in der betreffenden Schule (zweiteilig: A und B vs. dreiteilig: A, B und C). Für die Auswahl wurden zunächst die ALLE-Schulen gemäss diesen Kriterien klassifiziert. Für die verschiedenen Kombinationen der Kriterien, die bei den ALLE-Schulen vorkamen, wurde dann die Anzahl der zugeordneten ALLE-Schulen bestimmt und jeweils eine identische Anzahl von Schulen als potenzielle Vergleichsschulen zufällig gezogen. Zusätzlich wurde dabei differenziert, ob die Schule EMU durchführt oder nicht (vgl. oben), mit dem Ziel, fünf EMU-Schulen und fünf Schulen ohne Massnahmen gewinnen zu können.

In einem ersten Schritt wurden die Schulleitungen der gezogenen Schulen kontaktiert und für eine Teilnahme angefragt. Die Schulen mussten bereit sein, wie die ALLE-Schulen mit der gesamten Kohorte ihrer Siebtklässlerinnen und Siebtklässler im Schuljahr 2016/17 an der Evaluation teilzunehmen. Wenn bei einer Schule keine Bereitschaft zur Partizipation vorhanden war, wurde eine weitere zufällig gezogene Schule mit der gleichen Merkmalskombination (vgl. oben) angefragt. Um einen Anreiz für die Teilnahme zu schaffen, wurden den Lehrpersonen der Vergleichsschulen –

⁴ Die Bildungsdirektion des Kantons Zürich (Volksschulamt) fasst die Auswahl der Schulen für QUIMS wie folgt zusammen: «Über 130 Schulen im Kanton Zürich beteiligen sich gegenwärtig an QUIMS und erhalten kantonale Unterstützung, damit sie eine gute Schulqualität gewährleisten können. [...] Zur Beteiligung berechtigt und verpflichtet sind Schulen, die einen sogenannten Mischindex von 40% und mehr aufweisen. Dieser Index entspricht dem Mittel aus dem Anteil der Kinder nichtdeutscher Erstsprache und dem Anteil der Kinder ausländischer Nationalität (ohne Deutschland und Österreich).» (aus der QUIMS-Infobroschüre 2019, abgerufen am 07.09.2020 von <https://www.zh.ch/de/bildung/informationen-fuer-schulen/informationen-volksschule/volksschule-schulinfo-unterricht/volksschule-schulinfo-unterrichtsentwicklung/quims.html>).

analog zu den ALLE-Schulen – lehrpersonen- beziehungsweise klassenbezogene Ergebnismeldungen nach jedem Messzeitpunkt zugestellt (zu Resultaten in den Kompetenztests und in den Schülerinnen- und Schülerfragebögen). Darüber hinaus wurde als Dankeschön zu jedem Messzeitpunkt ein kleines Geschenk an jede teilnehmende Schülerin/jeden teilnehmenden Schüler sowohl der ALLE- als auch der Vergleichsschulen geschickt. Trotz dieser Bemühungen, die Teilnahme für die Vergleichsschulen attraktiv zu machen, ist festzuhalten, dass diese Schulen unmittelbar nur begrenzt von der Teilnahme profitieren konnten. Entsprechend mussten die – vollständig freiwillig teilnehmenden – Schulen ein vergleichsweise hohes Engagement aufbringen.

Im Vordergrund der Rekrutierung der Vergleichsschulen stand es, Schulen zu gewinnen, die im Hinblick auf wichtige lern- und leistungsrelevante Kontextbedingungen den ALLE-Schulen möglichst ähnlich waren (vgl. oben). Es war nicht die Zielsetzung, eine Stichprobe von Schulen zu erhalten, die in ihrer Zusammensetzung der Gesamtheit der Schulen der öffentlichen Sekundarstufe im Kanton Zürich möglichst ähnlich ist. Entsprechend sind auch die Vergleichsschulen als eine Gelegenheitsstichprobe von Schulen einzuordnen, für die keine Repräsentativität für die Grundgesamtheit der Schulen der öffentlichen Sekundarstufe angenommen werden kann.

Nach dem geschilderten Vorgehen konnten insgesamt neun Vergleichsschulen für die Teilnahme am Pilotprojekt gewonnen werden. Eine der Vergleichsschulen gab an, aufgrund der hohen Belastung zu Schuljahresbeginn nicht ab dem ersten Messzeitpunkt (September 2016), sondern erst danach an der Evaluation teilnehmen zu können. Um die Teilnahme dieser Vergleichsschule zu sichern, wurde dem Wunsch der Schule in Absprache mit der Auftraggeberin entsprochen. Nähere Informationen zu den beteiligten Vergleichsschulen bezogen auf die soziodemografischen Merkmale und die Eingangsleistung der an der Evaluation teilnehmenden Schülerinnen und Schüler können Kapitel 3.3.2.1 entnommen werden.

Rekrutierung der Teilnehmenden an den Interviews

Zur Gewinnung der Teilnehmenden an den Interviews wurden die Schulleitungen der neun ALLE-Schulen mit der Bitte kontaktiert, eine zuvor mit der Auftraggeberin abgesprochene Anzahl teilnehmender Lehrpersonen (für fünf Schulen vier bis fünf bzw. für vier Schulen drei Lehrpersonen, je nach Schultyp), je eine schulische Heilpädagogin/einen schulischen Heilpädagogen sowie eine Schulassistentin in eine Liste einzutragen und diese an das Evaluationsteam zu retournieren.

Für die Auswahl der Lehrpersonen erhielten die Schulen vier vorab definierte Kriterien, die möglichst vollständig berücksichtigt werden sollten:

- Die ausgewählten Lehrpersonen sind seit Beginn des Pilotprojekts an der Schule tätig und haben während dieser Zeit Deutsch und/oder Mathematik unterrichtet.
- Die ausgewählten Lehrpersonen haben die Weiterbildungen in Deutsch beziehungsweise Mathematik besucht und sind gut in das Pilotprojekt integriert.

- Zum Zeitpunkt des Interviews unterrichten diese Lehrpersonen eine Klasse beziehungsweise eine Anforderungsstufe in der 9. Klassenstufe in Deutsch und/oder Mathematik.
- In der Gruppe der Teilnehmenden sind Lehrpersonen aus allen Abteilungen beziehungsweise Anforderungsstufen (im Fach Mathematik) vertreten.

Betreffend die Schulassistenzen wurde den Schulen der Hinweis gegeben, dass die am Interview teilnehmende Schulassistenz möglichst in beiden Fächern eingesetzt worden sein sollte. Von drei Schulen erhielt das Evaluationsteam die Rückmeldung, dass keine Schulassistenz eingesetzt worden sei, und entsprechend keine Schulassistenz am Interview teilnehmen könne. An den Schulleitungsinterviews nahmen sämtliche neun zum Zeitpunkt der Interviews aktiven Schulleitungen der ALLE-Schulen teil. Zusätzlich wurden zwei mittlerweile abgelöste Schulleitungen von ALLE-Schulen interviewt, die während eines längeren Zeitraums zu Beginn des Pilotprojekts noch Schulleitungen gewesen waren. Von diesen abgelösten Schulleitungen wurden zusätzliche Informationen über die Umsetzung des Pilotprojekts gewonnen, die den aktuellen Schulleitungen nicht im vollen Umfang bekannt sein konnten. Sämtliche in das Pilotprojekt involvierte Weiterbildungs- und Beratungsfachpersonen wurden zu den Interviews eingeladen und nahmen an diesen teil.

3.3.2 Beschreibung der Stichproben

Im Folgenden wird auf die Stichproben eingegangen, welche den Analysen für das Kapitel 4 zugrunde lagen. In den Kapiteln 3.3.2.1 und 3.3.2.2 werden die Stichproben der Schülerinnen und Schüler beziehungsweise der Fachlehrpersonen in den ALLE- und Vergleichsschulen dargestellt, welche die Datenbasis für die Analysen in Kapitel 4.8 bis 4.10 sind. Dabei muss beachtet werden, dass die Beschreibung der Stichprobe der Schülerinnen und Schüler überwiegend auf den Rohdaten basiert (vor der Datenimputation; vgl. Kap. 3.5.3), um einen Überblick über die gewonnene Datenbasis zu geben. Nur für die soziodemografischen Merkmale der Schülerinnen und Schüler werden Statistiken mit imputierten Daten berichtet. In Kapitel 3.3.2.3 wird auf die Stichprobe aus den ALLE-Schulen eingegangen, die an den Befragungen zum Pilotprojekt und an den Interviews teilgenommen hat. Diese liegt den Kapiteln 4.1 bis 4.7 zugrunde.

3.3.2.1 Stichprobe der Schülerinnen und Schüler

Gesamtstichprobe und Teilnahme in den ALLE- und Vergleichsschulen

In die Evaluation ALLE wurden alle Schülerinnen und Schüler der ALLE- und Vergleichsschulen einbezogen, die an mindestens einer Erhebung der Evaluation ALLE (d.h. einem Messzeitpunkt) teilgenommen haben. An insgesamt 18 Schulen, davon neun ALLE- und neun Vergleichsschulen, nahmen 1'687 Schülerinnen und Schüler an mindestens einer Erhebung teil. Tabelle 3 gibt einen Überblick über diese Gesamtstichprobe der Schülerinnen und Schüler, unterteilt nach den ALLE-

und Vergleichsschulen. Für die Vergleichsschulen wird dabei zusätzlich nach ihren Teilnahmemustern über die vier Messzeitpunkte differenziert, da nicht alle Vergleichsschulen an allen Messzeitpunkten der Evaluation teilnahmen. Ein «x» in einer Spalte, die einem Messzeitpunkt zugeordnet ist, bedeutet die Teilnahme am jeweiligen Messzeitpunkt. Die Spalte «Anzahl Schulen» gibt an, auf wie viele Schulen das dargestellte Teilnahmemuster zutrifft. Die Spalten «Anzahl SuS» und «Prozent SuS» informieren zusätzlich über die Anzahl beziehungsweise die prozentualen Stichprobenanteile der Schülerinnen und Schüler an den Schulen mit unterschiedlichen Teilnahmemustern.

Tabelle 3. Gesamtstichprobe der Schülerinnen und Schüler und Teilnahmemuster der ALLE- und Vergleichsschulen

Schulgruppe	Messzeitpunkt				Anzahl Schulen	Anzahl SuS	Prozent SuS
	T1	T2	T3	T4			
ALLE-Schulen	x	x	x	x	9	970	57.5
Vergleichsschulen	x	x	x	x	6	510	30.2
	x	x	x		1	68	4.0
	x	x			1	57	3.4
		x	x	x	1	82	4.9
Gesamt					18	1687	100

Anmerkungen. SuS = Schülerinnen und Schüler. T1 = Anfang 7. Klassenstufe, T2 = Ende 7. Klassenstufe, T3 = Anfang 8. Klassenstufe, T4 = Ende 9. Klassenstufe. x = Teilnahme am jeweiligen Messzeitpunkt.

Wie aus Tabelle 3 zu entnehmen ist, hat die grosse Mehrheit der Schulen – sämtliche ALLE-Schulen und sechs der neun Vergleichsschulen – mit insgesamt 1'480 Schülerinnen und Schülern (87.7% der Gesamtstichprobe) an allen Messzeitpunkten teilgenommen. Dagegen nahmen von den neun Vergleichsschulen zwei Schulen (mit 7.4 Prozent der Gesamtstichprobe) nur an den ersten drei beziehungsweise zwei Messzeitpunkten teil (vorzeitig ausgeschiedene Vergleichsschulen), während eine Vergleichsschule (mit 4.9 Prozent der Gesamtstichprobe) erst ab dem zweiten Messzeitpunkt teilnahm (später einbezogene Vergleichsschule; vgl. Kap. 3.3.1).

Messzeitpunktspezifische Darstellung der Stichprobe

In Tabelle 4 und Tabelle 5 wird für die Fächer Deutsch und Mathematik aufgeschlüsselt, wie viele Schülerinnen und Schüler der ALLE- und Vergleichsschulen an den einzelnen Messzeitpunkten an der Evaluation teilnahmen. Zusätzlich wird nach den Abteilungen beziehungsweise Anforderungsstufen (Niveaus) differenziert, denen die Schülerinnen und Schüler zugeordnet waren.

Tabelle 4. Messzeitpunktspezifische Darstellung der Schülerstichprobe an den ALLE- und Vergleichsschulen mit Angaben zum Niveau für das Fach Deutsch

Schulen	T1		T2		T3		T4	
	N	Prozent	N	Prozent	N	Prozent	N	Prozent
<i>ALLE-Schulen^a</i>								
A/I	526	60.5	457	56.0	493	57.1	455	59.6
B/II	315	36.2	334	40.9	326	37.7	288	37.7
C/III	28	3.2	25	3.1	45	5.2	21	2.7
<i>Vergleichsschulen^b</i>								
A/I	315	55.1	389	61.4	318	58.0	261	57.9
B/II	225	39.3	207	32.6	208	38.0	166	36.8
C/III	32	5.6	38	6.0	22	4.0	24	5.3
<i>Je Niveau</i>								
A/I	841	58.4	846	58.3	811	57.4	716	58.9
B/II	540	37.5	541	37.3	534	37.8	454	37.4
C/III	60	4.2	63	4.3	67	4.7	45	3.7
Gesamt	1441	100	1450	100	1412	100	1215	100

Anmerkungen. T1 = Anfang 7. Klassenstufe, T2 = Ende 7. Klassenstufe, T3 = Anfang 8. Klassenstufe, T4 = Ende 9. Klassenstufe.

^a T3: fehlend: $N = 21$; T4: fehlend: $N = 97$. ^b T3: fehlend: $N = 38$; T4: fehlend: $N = 34$.

Die berichteten Kennwerte für die gesamte Stichprobe zeigen, dass die Anzahl der teilnehmenden Schülerinnen und Schüler bis zum dritten Messzeitpunkt etwa konstant bleibt, am vierten Messzeitpunkte jedoch abnimmt. Vergleicht man die Stichproben am ersten und letzten Messzeitpunkt mit Bezug zur Gesamtstichprobe aller Schülerinnen und Schüler ($N = 1'687$), geht in beiden Fächern der Anteil der Schülerinnen und Schüler, für die Daten verfügbar sind, von 85.4 auf 72.0 Prozent zurück. Die Verteilung der Schülerinnen und Schüler auf die Niveaus bleibt über die Messzeitpunkte in etwa konstant. Deutlichere Disparitäten zeigen sich allerdings bei den Vergleichsschulen im Fach Mathematik, bei denen die Zugehörigkeit zu Niveau A/I zwischen 42.6 und 52.6 Prozent und die Zugehörigkeit zu Niveau C/III zwischen 9.9 und 16.6 Prozent variiert. Dies ist zum Teil auf den Ein- und Ausstieg von Vergleichsschulen während der Evaluation zurückzuführen (vgl. Tabelle 3). Ein weiterer Grund kann eine zwischen den Niveaus unterschiedliche Teilnahmebereitschaft im Projektverlauf sein. Zusätzlich muss beachtet werden, dass die dargestellten Kennwerte primär auf den Niveaangaben basieren, die auf der technischen Plattform des Lernfördersystems Lernpass plus gemacht wurden. Die Erfassung und Aktualisierung der Angaben an jedem Messzeitpunkt lag in der Verantwortung der einzelnen Schulen (vgl. Kap. 3.2.2).

Tabelle 5. Messzeitpunktspezifische Darstellung der Schülerstichprobe an den ALLE- und Vergleichsschulen mit Angaben zum Niveau für das Fach Mathematik

Schulen	T1		T2		T3		T4	
	N	Prozent	N	Prozent	N	Prozent	N	Prozent
<i>ALLE-Schulen^a</i>								
A/I	442	50.9	458	53.5	455	51.8	382	50.3
B/II	316	36.4	315	36.8	327	37.2	298	39.3
C/III	111	12.8	83	9.7	96	10.9	79	10.4
<i>Vergleichsschulen^b</i>								
A/I	250	43.7	336	52.6	269	48.1	194	42.6
B/II	227	39.7	240	37.6	228	42.6	193	42.4
C/III	95	16.6	63	9.9	62	11.6	68	14.9
<i>Je Niveau</i>								
A/I	692	48.0	794	53.1	724	50.4	576	47.4
B/II	543	37.7	555	37.1	555	38.6	491	40.4
C/III	206	14.3	146	9.8	158	11.0	147	12.1
Gesamt	1441	100	1495	100	1437	100	1214	100

Anmerkungen. T1 = Anfang 7. Klassenstufe, T2 = Ende 7. Klassenstufe, T3 = Anfang 8. Klassenstufe, T4 = Ende 9. Klassenstufe.

^a T3: fehlend: $N = 7$; T4: fehlend: $N = 102$. ^b T3: fehlend: $N = 38$; T4: fehlend: $N = 30$.

Längsschnittliche Darstellung der Stichprobe

In Tabelle 6 wird für die Fächer Deutsch und Mathematik dargestellt, wie viele Schülerinnen und Schüler der ALLE- und Vergleichsschulen zwischen den einzelnen Messzeitpunkten in derselben Klasse und durch dieselbe Lehrperson unterrichtet wurden (d.h. die zeitlich stabile Schülerschaft in fachspezifischen Klassen bzw. Lerngruppen). Zudem wird die Anzahl der Klassen beziehungsweise Lerngruppen in den Fächern Deutsch und Mathematik angegeben, in denen die Schülerschaft zwischen den Messzeitpunkten, die von derselben Fachlehrperson unterrichtet wurden, stabil geblieben ist. Klassen beziehungsweise Lerngruppen mit einer stabilen Schülerschaft von $N < 5$ wurden nicht mitgezählt und weiter in die Analysen einbezogen. Ebenso wurden die dazugehörigen Schülerinnen und Schüler für die nachfolgende Darstellung nicht mitgezählt.

Tabelle 6. Längsschnittliche Darstellung der Schülerstichprobe an den ALLE- und Vergleichsschulen für die Fächer Deutsch und Mathematik

		ALLE-Schulen		Vergleichsschulen		Gesamt	
		Anzahl SuS	Anzahl Klassen	Anzahl SuS	Anzahl Klassen	Anzahl SuS	Anzahl Klassen
Deutsch	T1 = T2	785	45	574	35	1359	85
	T2 = T3	636	44	451	32	1087	76
	T3 = T4	419	28	415	28	834	56
	T1 = T4	369	23	301	21	670	44
Mathematik	T1 = T2	810	48	591	35	1401	83
	T2 = T3	534	39	411	32	945	71
	T3 = T4	484	43	407	28	891	71
	T1 = T4	347	23	300	23	647	46

Anmerkungen. SuS = Schülerinnen und Schüler. T1 = Anfang 7. Klassenstufe, T2 = Ende 7. Klassenstufe, T3 = Anfang 8. Klassenstufe, T4 = Ende 9. Klassenstufe.

Es zeigt sich, dass die Anzahl der Schülerinnen und Schüler, die in derselben Klasse und durch dieselbe Lehrperson zwischen zwei Messzeitpunkten unterrichtet werden, mit der Zeit abnimmt. Dies spiegelt sich auch in der Anzahl der konstanten Klassen beziehungsweise Lerngruppen wider. Am kleinsten ist (erwartungsgemäss) die zeitlich stabile Schülerschaft beziehungsweise die Anzahl zeitlich stabiler Klassen/Lerngruppen zwischen dem ersten (T1) und letzten (T4) Messzeitpunkt. Die über den gesamten Evaluationszeitraum stabile Stichprobe von Klassen/Lerngruppen liegt den Analysen in den Kapiteln 4.8 und 4.9 zugrunde. Vergleicht man die Kennwerte für die ALLE- und Vergleichsschulen, wird deutlich, dass in beiden Gruppen die Anzahl der stabil beteiligten Schülerinnen und Schüler sowie Klassen/Lerngruppen über den Evaluationszeitraum hinweg abnimmt. Darüber hinaus ist festzustellen, dass die zeitlich stabile Schülerschaft in den Vergleichsschulen schon zwischen den ersten Messzeitpunkten kleiner ist als in den ALLE-Schulen. Dies ist auf den Ein- beziehungsweise Ausstieg von Vergleichsschulen zurückzuführen (vgl. Tabelle 3).

Soziodemografische Merkmale und Eingangsleistungen der Schülerinnen und Schüler

Im Folgenden wird die Schülerstichprobe der Evaluation ALLE im Hinblick auf zentrale soziodemografische Merkmale beschrieben: Geschlecht, Alter, Migrationshintergrund, zuhause gesprochene Sprache und sozioökonomischer Status. Zudem wird dargestellt, inwieweit sich die Schülerinnen

und Schüler in den ALLE- und Vergleichsschulen hinsichtlich dieser soziodemografischen Merkmale und in ihren Deutsch- und Mathematik-Testleistungen zu Beginn der Evaluation (T1) unterschieden. Die ALLE- und Vergleichsschulen werden mithilfe von Regressionsmodellen in Bezug auf die oben genannten Merkmale verglichen und die Unterschiede auf statistische Signifikanz getestet. In den Regressionsmodellen werden die genannten Merkmale der Schülerinnen und Schüler als die zu erklärenden Zielvariablen (abhängige Variablen) verwendet. Als erklärende Variable (Prädiktoren bzw. unabhängige Variable) wird jeweils eine zweistufige (Dummy-)Variable herangezogen: ALLE-Schulen = 1, Vergleichsschulen = 0 (vgl. Kap. 3.6). Für zweistufige Merkmale wurde eine logistische, für mehrstufige Variablen eine multinomiale Regression berechnet.

In Tabelle 7 wird die Verteilung der einzelnen Merkmalen getrennt für die ALLE- und Vergleichsschulen dargestellt (Prozentanteile und absolute Häufigkeiten für kategoriale Variablen: Geschlecht, Migrationshintergrund und zuhause gesprochene Sprache; Mittelwert und Standardabweichung für die kontinuierlichen Variablen: sozioökonomischer Status, Deutsch- und Mathematikleistung). Als Ergebnisse wird aus den Regressionsmodellen berichtet, ob die Unterschiede zwischen den ALLE- und Vergleichsschulen statistisch signifikant sind und wie viel Varianz in den Merkmalen dabei erklärt wird (Bestimmtheitsmass R^2).

Bereits deskriptiv zeigt sich, dass die Disparitäten zwischen den ALLE- und Vergleichsschulen überwiegend gering waren. Statistisch bedeutsame Unterschiede finden sich einerseits beim Migrationshintergrund: Der Anteil an Schülerinnen und Schülern, bei denen ein Elternteil in der Schweiz geboren ist, fällt in den Vergleichsschulen etwas höher aus (Differenz: 3.4 Prozent). Auch der Anteil mit Migrationshintergrund (2. Generation) ist in den Vergleichsschulen etwas höher ausgeprägt (Differenz: 4.4 Prozent). Weiterhin ist der mittlere sozioökonomische Status in den Vergleichsschulen statistisch signifikant höher ausgeprägt. Bei allen diesen Unterschieden fällt jedoch der Anteil erklärter Varianz (R^2) im Vergleich zwischen ALLE- und Vergleichsschulen äusserst gering aus. Damit finden sich kaum Hinweise darauf, dass sich die Lernvoraussetzungen der Schülerinnen und Schüler aus den ALLE- und Vergleichsschulen zu Beginn des Evaluationszeitraums bedeutsam voneinander unterschieden haben.

Tabelle 7. Soziodemografische Merkmale und Eingangsleistungen in Deutsch und Mathematik der Schülerinnen und Schüler in den ALLE- und Vergleichsschulen

Kategoriale Merkmale	ALLE-Schulen	Vergleichsschulen	R^2 (Nagelkerke)
	Prozent (N)	Prozent (N)	
Geschlecht: weiblich	49.3 (478)	51.3 (368)	< 0.01
Migrationshintergrund			< 0.01
<i>Kein Migrationshintergrund (beide Eltern in CH geboren)</i>	36.0 (349)	30.8 (221)	
<i>Kein Migrationshintergrund (ein Elternteil in CH geboren)</i>	19.1 ^a (186)	22.5 ^a (161)	
<i>Migrationshintergrund (2. Generation)</i>	30.7 ^a (297)	35.1 ^a (252)	
<i>Migrationshintergrund (1. Generation)^b</i>	14.2 (138)	11.7 (84)	
Zuhause gesprochene Sprache			< 0.01
<i>Ausschliesslich Schulsprache</i>	58.0 (563)	54.6 (392)	
<i>Schulsprache und andere Sprache(n)</i>	1.3 (13)	2.2 (16)	
<i>Ausschliesslich andere Sprache(n)^b</i>	40.7 (395)	43.2 (310)	
Kontinuierliche Merkmale	M (SD)	M (SD)	R^2
Alter (am letzten Messzeitpunkt; T4)	15.9 (0.5)	15.9 (0.5)	< 0.01
Sozioökonomischer Status	-0.04 ^a	0.06 ^a	< 0.01
Deutsch-Gesamtleistung T1	500.7 (74.5)	500.0 (76.4)	< 0.01
Mathematik-Gesamtleistung T1	504.7 (77.2)	500.0 (75.9)	< 0.01

Anmerkungen. T1 = Anfang 7. Klassenstufe, T4 = Ende 9. Klassenstufe. ALLE-Schulen $N = 970$, Vergleichsschulen $N = 717$. Prozent (N): Die Zahl vor der Klammer gibt den Prozentanteil des betreffenden Merkmals in der Gesamtstichprobe der Schülerinnen und Schüler in der jeweiligen Gruppe (ALLE-, Vergleichsschulen) an, die Zahl in der Klammer die absolute Häufigkeit. M (SD): Die Zahl vor der Klammer gibt den Mittelwert des betreffenden Merkmals in der Gesamtstichprobe der Schülerinnen und Schüler in der jeweiligen Gruppe (ALLE-, Vergleichsschulen) an, die Zahl in der Klammer die Standardabweichung. Die Berechnung der erklärten Varianz (R^2) erfolgte mit dem ersten imputierten Datensatz (vgl. Kap. 3.5.3). Für genauere Angaben zu den soziodemografischen Merkmalen siehe Kap. 3.5.1.

^a Unterschied ist tendenziell signifikant mit $p < .10$ ($p > .05$).

^b Referenzkategorie im multinomialen Regressionsmodell.

3.3.2.2 Stichprobe der Lehrpersonen

In Tabelle 8 wird dargestellt, wie viele Lehrpersonen der ALLE- und Vergleichsschulen an den einzelnen Messzeitpunkten eine Klasse/Lerngruppe im Fach Deutsch beziehungsweise Mathematik unterrichtet haben.

Tabelle 8. Messzeitpunktspezifische Darstellung der Stichprobe der Lehrpersonen an den ALLE- und Vergleichsschulen für die Fächer Deutsch und Mathematik

Gruppe/Fach	Messzeitpunkt			
	T1	T2	T3	T4
ALLE-Schulen				
Deutsch	35	35	34	36
Mathematik	38	38	39	40
Vergleichsschulen				
Deutsch	28	27	24	21
Mathematik	26	26	24	20
Gesamt				
Deutsch	63	72	58	57
Mathematik	64	64	63	60

Anmerkungen. T1 = Anfang 7. Klassenstufe, T2 = Ende 7. Klassenstufe, T3 = Anfang 8. Klassenstufe, T4 = Ende 9. Klassenstufe.

Während die Anzahl der Lehrpersonen in den ALLE-Schulen über den Evaluationszeitraum relativ konstant bleibt, nimmt die Anzahl der Deutsch- und Mathematik-Lehrpersonen in den Vergleichsschulen ab. Dies ist vor allem darauf zurückzuführen, dass von den neun Vergleichsschulen zwei Schulen am dritten und vierten Messzeitpunkt vorzeitig ausgeschieden sind (vgl. Kapitel 3.3.2.1).

In Tabelle 9 wird für die Deutsch- und Mathematik-Lehrpersonen in den ALLE- und Vergleichsschulen dargestellt, wie viele Lehrpersonen zwischen den Messzeitpunkten dieselben Klassen/Lerngruppen unterrichtet haben (z.B. T1 = T4: Anzahl der Lehrpersonen, die zu T4 die gleichen Klassen/Lerngruppen unterrichtet haben wie zu T1). Wie sich zeigt, nimmt die Anzahl der Lehrpersonen, welche dieselben Schülerinnen und Schüler unterrichten, mit der Zeit ab. Diese Annahme ist an den ALLE- wie den Vergleichsschulen ersichtlich, für die Vergleichsschulen aber stärker ausgeprägt. Auch hier sind vor allem die später einbezogenen beziehungsweise früher ausgeschiedenen Vergleichsschulen als Grund zu sehen.

Tabelle 9. Längsschnittliche Darstellung der Stichprobe der Lehrpersonen an den ALLE- und Vergleichsschulen für die Fächer Deutsch und Mathematik

Fach		ALLE-	Vergleichs-	Gesamt
		Schulen	schulen	
		Anzahl LP	Anzahl LP	Anzahl LP
Deutsch	T1 = T2	33	27	60
	T2 = T3	29	23	52
	T3 = T4	22	19	41
	T1 = T4	20	18	38
Mathematik	T1 = T2	38	26	64
	T2 = T3	34	24	58
	T3 = T4	25	19	44
	T1 = T4	21	17	28

Anmerkungen. LP = Lehrpersonen. T1 = Anfang 7. Klassenstufe, T2 = Ende 7. Klassenstufe, T3 = Anfang 8. Klassenstufe, T4 = Ende 9. Klassenstufe.

In Tabelle 10 wird die Stichprobe der Lehrpersonen bezüglich der Merkmale Geschlecht, Alter und Berufserfahrung beschrieben (Prozentanteile und absolute Häufigkeiten). Die Kennwerte beruhen auf der Befragung der Lehrpersonen am ersten Messzeitpunkt; die Darstellung erfolgt getrennt für die Deutsch- und Mathematik-Lehrpersonen beziehungsweise die ALLE- und Vergleichsschulen.

Bei der Gegenüberstellung der Ergebnisse für die ALLE- und Vergleichsschulen erscheinen einige Unterschiede erwähnenswert: Für die Deutsch-Lehrpersonen ergibt sich ein höherer Anteil jüngerer Lehrpersonen (< 30 Jahre) und von Lehrpersonen mit vergleichsweise wenig Berufserfahrung (≤ 5 Jahre) in den ALLE-Schulen. Zugleich fällt der Anteil eher älterer Lehrpersonen (≥ 50 Jahre) und von Lehrpersonen mit sehr viel Berufserfahrung (> 25 Jahre) in den ALLE-Schulen etwas höher aus. Für die Mathematik-Lehrpersonen resultiert ein höherer Anteil von Frauen, von jüngeren Lehrpersonen (< 30 Jahre) und von Lehrpersonen ohne sehr lange Berufserfahrung (≤ 15 Jahre) in den ALLE-Schulen. Der Anteil von Lehrpersonen mit vergleichsweise wenig Berufserfahrung (≤ 5 Jahre) ist in beiden Gruppen jedoch ähnlich ausgeprägt.

Tabelle 10. Demografische Merkmale der Deutsch- und Mathematik-Lehrpersonen in den ALLE- und Vergleichsschulen

Merkmale	Deutsch-Lehrpersonen		Mathematik-Lehrpersonen	
	ALLE-Schulen	Vergleichsschulen	ALLE-Schulen	Vergleichsschulen
	Prozent (N)	Prozent (N)	Prozent (N)	Prozent (N)
Geschlecht: weiblich	62.9 (22)	63.6 (14)	51.6 (15)	36.8 (7)
Alter (am ersten Messzeitpunkt; T1)				
<i>Unter 25 Jahre alt</i>			3.2 (1)	
<i>25 bis 29 Jahre alt</i>	11.4 (4)	4.5 (1)	19.4 (6)	5.3 (1)
<i>30 bis 39 Jahre alt</i>	31.4 (11)	45.5 (10)	32.3 (10)	31.6 (6)
<i>40 bis 49 Jahre alt</i>	34.3 (12)	31.8 (7)	29.0 (9)	36.8 (7)
<i>50 bis 59 Jahre alt</i>	17.4 (6)	9.1 (2)	9.7 (3)	21.1 (4)
<i>60 Jahre oder älter</i>	5.7 (2)	9.1 (2)	6.5 (2)	5.3 (1)
Berufserfahrung				
<i>1 Jahr bis 5 Jahre</i>	34.3 (12)	22.7 (5)	29.0 (9)	31.6 (6)
<i>6 bis 15 Jahre</i>	28.6 (10)	45.5 (10)	48.4 (15)	15.8 (3)
<i>16 bis 25 Jahre</i>	11.4 (4)	13.6 (3)	9.7 (3)	26.3 (5)
<i>Über 25 Jahre</i>	25.7 (9)	18.2 (4)	12.9 (4)	26.3 (5)

Anmerkungen. ALLE-Schulen: Deutsch-Lehrpersonen $N = 35$; Mathematik-Lehrpersonen $N = 31$, fehlend $N = 7$. Vergleichsschulen: Deutsch-Lehrpersonen $N = 22$, fehlend $N = 6$; Mathematik-Lehrpersonen $N = 19$, fehlend: $N = 7$.

3.3.2.3 Stichproben in den Befragungen zum Pilotprojekt und in den Interviews

Die Online-Befragungen und Interviews zu den Rahmenbedingungen und Massnahmen des Pilotprojekts am vierten Messzeitpunkt (T4; vgl. Tabelle 2) wurden mit Lehrpersonen, schulischen Heilpädagoginnen und Heilpädagogen (nur Interviews), Schulleitungen, Schülerinnen und Schülern sowie Beratungs- und Weiterbildungspersonen (nur Interviews) durchgeführt. Einbezogen wurden nur Personen, welche am Pilotprojekt beteiligt waren, also keine Vertreterinnen oder Vertreter der Vergleichsschulen. Im Folgenden werden die Stichproben aus diesen Befragungen und Interviews dargestellt.

Lehrpersonen

Insgesamt 52 Lehrpersonen nahmen an der Online-Befragung zum Pilotprojekt teil. Von diesen Lehrpersonen waren 57.7 Prozent ($N = 30$) Frauen und 42.3 Prozent ($N = 22$) Männer. Für das

Alter der Lehrpersonen am vierten Messzeitpunkt (offen erfragt) resultierte ein Mittelwert von $M = 43.1$ und eine Streuung von $SD = 11.2$ (fehlende Angaben: $N = 6$). 38.5 Prozent ($N = 20$) der Befragten unterrichteten in der 9. Klassenstufe das Fach Deutsch, 50 Prozent ($N = 26$) das Fach Mathematik und 11.5 ($N = 6$) Prozent beide Fächer. Zwei Lehrpersonen gaben für den Zeitpunkt der Befragung an, weder das Fach Deutsch noch Mathematik in einer ALLE-Klasse zu unterrichten.

Schulassistenzen

Vierzehn Schulassistenzen nahmen an der Online-Befragung zum Pilotprojekt teil, davon waren 78.6 Prozent ($N = 11$) Frauen und 21.4 Prozent ($N = 3$) Männer. 28.6 Prozent ($N = 4$) der Befragten gaben an, im Fach Deutsch zur Unterstützung der an der Evaluation ALLE teilnehmenden Schülerinnen und Schüler (zu diesem Zeitpunkt: Neuntklässlerinnen und Neuntklässler) eingesetzt zu werden, 21.4 Prozent ($N = 3$) im Fach Mathematik und 50.0 Prozent ($N = 7$) in beiden Fächern. Von den 14 befragten Schulassistenzen waren 14.3 Prozent ($N = 2$) bis 29 Jahre alt, 28.6 Prozent ($N = 4$) 30 bis 39 Jahre alt, 14.3 Prozent ($N = 2$) 40 bis 49 Jahre alt und 42.9 Prozent ($N = 6$) 50 bis 59 Jahre alt.

Schulleitungen

An der Online-Befragung zum Pilotprojekt nahmen elf Schulleitungen von allen neun ALLE-Schulen teil, zwei davon waren ehemalige Schulleitungen. In Abstimmung mit der Auftraggeberin wurde entschieden, die Angaben der beiden ausgeschiedenen Schulleitungen mit einzubeziehen, da diese über vergleichsweise umfangreiche Erfahrungen mit und Kenntnisse über das Pilotprojekt verfügten. 18.2 Prozent ($N = 2$) der Schulleitungen waren Frauen und 81.8 Prozent ($N = 9$) Männer.

Stichproben der Interviews

Die an den Einzel- und Gruppeninterviews teilnehmende Personengruppe setzte sich wie folgt zusammen:

- Schulleitungen: $N = 9$ (inkl. zwei ehemalige Schulleitungen),
- Lehrpersonen: $N = 34$ (35 Personen waren eingeladen, davon 20 Frauen und 15 Männer; zwei dieser Personen konnten nicht teilnehmen, dafür kam eine Lehrperson hinzu),
- Schulische Heilpädagoginnen beziehungsweise Heilpädagogen: $N = 6$ (neun Personen waren eingeladen, davon sechs Frauen und drei Männer; drei dieser Personen konnten nicht teilnehmen),
- Schulassistenzen: $N = 6$ (fünf Frauen und ein Mann),
- Beratungs-/Weiterbildungspersonen: $N = 6$.

Von vier ALLE-Schulen wurden drei Lehrpersonen eingeladen, von drei Schulen vier Lehrpersonen und von zwei Schulen fünf Lehrpersonen. Von einer Schule mit fünf eingeladenen Personen fehlten

am Interview zwei Lehrpersonen (längere Krankheit, zweite Abwesenheit unerklärt). In einer Schule mit drei eingeladenen Lehrpersonen war beim Interview eine Lehrperson mehr anwesend als ursprünglich vorgesehen. Für jede Schule war zudem zu den Gruppeninterviews mit den Lehrpersonen eine schulische Heilpädagogin beziehungsweise ein schulischer Heilpädagoge eingeladen.

In einer Schule mit drei eingeladenen Lehrpersonen war im Gruppeninterview zusätzlich die neue Schulleitung präsent, welche daher nicht im Einzelinterview befragt wurde. Eine zweite Schulleitung, welche während des Pilotprojekts neu eine ALLE-Schule übernommen hatte, nahm an keinem Interview teil. Von insgesamt elf neuen und bereits länger tätigen Schulleitungen wurden entsprechend insgesamt neun im Einzelinterview befragt.

Weiterhin wurde von sechs der neun ALLE-Schulen je eine Schullassistentin im Einzelinterview befragt. Schliesslich wurden sämtliche sechs der in den Weiterbildungen und Beratungen im Pilotprojekt involvierten Fachpersonen einzeln interviewt.

3.4 Beschreibung der Erhebungsinstrumente

Nachfolgend wird detailliert auf die verwendeten Kompetenztests zur Erfassung der fachlichen Lernergebnisse der Schülerinnen und Schüler (Kap. 3.4.1), die in der Evaluation eingesetzten Fragebögen (Kap. 3.4.2) und die Inhalte der Einzel- und Gruppeninterviews (Kap. 3.4.3) eingegangen.

3.4.1 Kompetenztests

Kompetenztests in Deutsch und Mathematik

Die Durchführung der Kompetenztests in der Evaluation ALLE war eng mit dem Lernfördersystem *Lernpass plus* verknüpft, das vom Lehrmittelverlag St.Gallen angeboten wird. Lernpass plus ist ein System für den 3. Zyklus der Volksschule in den Fächern Deutsch, Englisch, Französisch, Mathematik sowie Natur + Technik. Auf der Grundlage von Orientierungstests ermöglicht es die Bestimmung des Lernstands der Schülerinnen und Schüler und darüber hinaus eine individuelle Standortbestimmung. Die Aufgabenpools von Lernpass plus sind auf die Erfassung der Kompetenzen nach dem Lehrplan 21 hin konzipiert. Die Aufgaben werden am Lehrmittelverlag St.Gallen entwickelt und von Fachexpertinnen und -experten der Pädagogischen Hochschule St.Gallen (PHSG) auf ihre Eignung geprüft. Im Fach Deutsch ermöglichen sie eine Messung in den Kompetenzbereichen «Lesen», «Hören» und «Sprache(n) im Fokus», im Fach Mathematik in den Kompetenzbereichen «Zahl und Variable», «Form und Raum» und «Grössen, Funktionen, Daten und Zufall».

In der Evaluation ALLE wurden zur Erfassung der fachlichen Lernergebnisse in Deutsch und Mathematik Aufgaben aus den Aufgabenpools von Lernpass plus verwendet. Die Aufgaben für die

Eingangserhebung zu Beginn der 7. Klassenstufe (T1) wurden vom Evaluationsteam selbst ausgewählt und zu Testheften zusammengestellt (siehe unten). Für die weiteren drei Messzeitpunkte war zunächst vorgesehen, auf Testdaten zurückzugreifen, die im Rahmen von regulären Erhebungen von Lernpass plus gewonnen wurden. Es war geplant, für die Erhebung am Ende der 7. Klassenstufe (T2) auf die Testergebnisse der an der Evaluation teilnehmenden Schülerinnen und Schüler bei der Normierungsstudie für Lernpass plus im Kanton Zürich, und für die Erhebungen am Ende der 8. und 9. Klassenstufe (T3, T4) jeweils auf die Ergebnisse der Schülerinnen und Schüler in der Standortbestimmung in Lernpass plus (*Stellwerk*) zurückzugreifen. Aufgrund von Verzögerungen in der (Weiter-)Entwicklung von Lernpass plus und im Interesse einer besseren Kontrolle der Untersuchungsbedingungen wurde jedoch im Einvernehmen mit der Auftraggeberin entschieden, die Verwendung von Daten aus Lernpass plus auf T2 zu beschränken und die Aufgabenauswahl und Zusammenstellung der Tests für T3 und T4 (wie bereits bei T1) in Eigenregie durchzuführen. Entsprechend wurden zu T1, T3 und T4 vom Evaluationsteam zusammengestellte Tests mit Aufgaben aus Lernpass plus verwendet, zu T2 die Testergebnisse der Normierungsstudie.

In den Tests wurden zu allen Messzeitpunkten Aufgaben aus den jeweils drei Kompetenzbereichen der beiden Fächer (siehe oben) vorgelegt. Tabelle 11 gibt einen Überblick über die Gesamtanzahl der eingesetzten Aufgaben pro Messzeitpunkt und Kompetenzbereich, die Anzahl der Aufgaben, welche von einer einzelnen Schülerin/einem einzelnen Schüler jeweils bearbeitet wurde, sowie die Testzeit, welche den Schülerinnen und Schülern zur Verfügung stand. Zudem werden die mittlere Schwierigkeit (Lösungsprozente), die mittlere Trennschärfe der Aufgaben und die Reliabilität für die Testwerte der Schülerinnen und Schüler angegeben.

Wie Tabelle 11 zeigt, variierte die Anzahl Aufgaben je Kompetenzbereich deutlich zwischen den Messzeitpunkten («Spalte Aufgaben pro KB gesamt»). Bei T3 und T4 wurde in Abstimmung mit der Auftraggeberin vom ursprünglichen Plan abgewichen, eine mit T1 vergleichbare Anzahl Aufgaben vorzulegen, um die Belastung der Schulen durch die Evaluation zu reduzieren. Entsprechend enthielten die Tests zu T3 und T4 deutlich weniger Aufgaben pro Kompetenzbereich. Beim Messzeitpunkt T2 war die Anzahl der Aufgaben weitgehend durch die Rahmenbedingungen der Normierungsstudie von Lernpass plus und das in dieser Studie eingesetzte Testheftdesign (siehe unten) vorgegeben. Allerdings war es möglich, die Testlänge der Normierungstests (14 Aufgaben je Kompetenzbereich) um jeweils sieben Aufgaben pro Kompetenzbereich zu ergänzen, um eine genauere (reliablere) Messung und bessere inhaltliche Abdeckung der Kompetenzen zu ermöglichen.

Tabelle 11. Zusammensetzung und statistische Kennwerte der Kompetenztests Deutsch und Mathematik für die Messzeitpunkte der Evaluation ALLE

MZP ^a	Aufgaben pro KB	Aufgaben pro KB gesamt ^b	Aufgaben pro Testheft	Testzeit ^c	Mittlere Lösungshäufigkeit (in Prozent)	Mittlere Trennschärfe	Reliabilität
T1	25	35 – 46	25	6 x 45 Min.	41.0/35.1	0.32/0.31	0.88/0.90
T2	21	je 183	21	6 x 45 Min.	44.8/30.7	0.35/0.34	0.88/0.89
T3	10	21 – 24	30	2 x 60 Min.	55.2/52.3	0.35/0.39	0.79/0.83
T4	15	27 – 31	45	2 x 90 Min.	48.8/52.6	0.38/0.44	0.89/0.91

Anmerkungen. MZP = Messzeitpunkt, KB = Kompetenzbereich. Angaben vor dem Schrägstrich beziehen sich auf das Fach Deutsch, Angaben nach dem Schrägstrich auf das Fach Mathematik.

^a T1: 7. Klassenstufe, September 2016; T2: 7. Klassenstufe, Mai 2017; T3: 8. Klassenstufe, Juni/Juli 2018; T4: 9. Klassenstufe, Mai/Juni 2019.

^b Angegeben je MZP ist die Anzahl (Spannweite) der insgesamt je Kompetenzbereich verwendeten Aufgaben. Diese Aufgaben wurden auf unterschiedliche Testhefte verteilt, je Schülerin/Schüler wurde die Anzahl Aufgaben vorgelegt, welche in der Spalte «Aufgaben pro KB» angeführt ist.

^c Angegeben ist die Anzahl der separaten Tests, die pro Messzeitpunkt absolviert wurden, sowie die Testzeit pro Messzeitpunkt. An T1 und T2 wurden die sechs Kompetenzbereiche in separaten Tests bearbeitet (= 6 Tests), an T3 und T4 zusammengefasst in je einem Test pro Fach (= 2 Tests).

An den einzelnen Messzeitpunkten erhielten nicht alle Schülerinnen und Schüler dieselben Aufgaben vorgelegt. Stattdessen wurde jeweils ein Multi-Matrix-Design mit unterschiedlichen Testheften eingesetzt (siehe z.B. Gonzalez & Rutkowski, 2010). In einem solchen Design enthalten die Testhefte einerseits unterschiedliche Aufgaben, sodass angesichts einer begrenzten Testzeit pro Schülerin/Schüler eine grössere Anzahl Aufgaben vorgelegt werden kann, und für unterschiedliche Schülergruppen jeweils passende Testhefte zusammengestellt werden können. Andererseits enthalten die Testhefte jeweils einen Anteil von Aufgaben, der auch in anderen Testheften eingesetzt wird («Link-Items»). Dadurch ist es möglich, die Ergebnisse der Schülerinnen und Schüler, die unterschiedliche Testhefte bearbeitet haben, gemeinsam auszuwerten und auf der gleichen Kompetenzskala abzubilden, sodass ihre Ergebnisse direkt vergleichbar sind. Bei den ALLE-Tests zu T1, T3 und T4 wurden Testhefte für die Niveaus A/I, B/II und C/III zusammengestellt (vgl. Kap. 3.2.2), die sich nicht hinsichtlich der Anzahl Aufgaben, aber der Aufgabenschwierigkeit unterschieden, um den unterschiedlichen Fähigkeiten der Schülerinnen und Schüler gerecht zu werden. Der Test zu T2 erfolgte auf Grundlage des Testheftdesigns der Normierung, welches je Kompetenzbereich 183 Testhefte mit je 14 Aufgaben umfasste (nach einem Youden-Square-Design auf Aufgabenebene; siehe z.B. Frey, Hartig & Rupp, 2009). Durch die Ergänzung um sieben Aufgaben je Testheft (vgl. oben) erhielten die ALLE-Schülerinnen und Schüler 183 Testhefte à 21 Aufgaben vorgelegt. Damit unterschied sich das Testdesign von T2 recht grundlegend von dem der anderen Messzeitpunkte, und auch die Durchführungsbedingungen (z.B. Instruktionen an die Schülerinnen

und Schüler) unterschieden sich zwischen der Normierungsstudie und den Erhebungen an den anderen, für die Evaluation ALLE spezifischen, Messzeitpunkten.

Die Auswahl der Aufgaben für die Tests zu T1, T3 und T4 erfolgte einerseits so, dass eine Verknüpfung der Messungen über die Messzeitpunkte (*Linking*; siehe Kap. 3.5.2) möglich war. Zum anderen wurde auf die Ausgewogenheit bezüglich der getesteten Inhalte und die Passung der Aufgabenschwierigkeiten zu den (vermuteten) Fähigkeiten der Schülerinnen und Schüler geachtet. Zu T1 konnten dazu Ergebnisse von Pilotierungsstudien aus Lernpass plus sowie bereits bekannte Aufgabeneigenschaften (z.B. Aufgabenschwierigkeiten) aus der Verwendung von vielen der Aufgaben im Vorläufer der Standortbestimmung in Lernpass plus – dem *Stellwerk*-Test – herangezogen werden. Zu T3 und T4 konnten die Ergebnisse von T1 sowie der Normierungsstudie (T2) berücksichtigt werden. An der Normierungsstudie nahmen Schülerinnen und Schüler der 7. bis 9. Klassenstufe teil. Dies ermöglichte es, die Fähigkeiten der ALLE-Schülerinnen und Schüler am Ende der 8. Klassenstufe (T3) und 9. Klassenstufe (T4) abzuschätzen und bezüglich ihrer Schwierigkeit passende Aufgaben auszuwählen.

Kompetenztest zum «Texte schreiben»

Ein wichtiger Kompetenzbereich im Fach Deutsch, der auf Basis der Aufgabenpools in Lernpass plus nicht getestet werden konnte, ist das Schreiben. Daher wurden zusätzlich zu den bereits beschriebenen Kompetenztests die Ergebnisse der ALLE-Schülerinnen und -Schüler im *Schreibauftrag*, einer Erweiterung der obligatorischen *Stellwerk*-Tests in der 8. Klassenstufe, herangezogen und analysiert (im vorliegenden Bericht kurz als Test zum «Texte schreiben» bezeichnet). Die Schreibaufträge basieren auf dem Lehrplan des Kantons Zürich. Die Entwicklung und Beurteilung der Schreibaufträge erfolgen durch das Institut für Bildungsevaluation IBE der Universität Zürich.⁵

3.4.2 Fragebogeninstrumente

Im Folgenden werden die Inhalte der eingesetzten Fragebögen für Schülerinnen und Schüler, Lehrpersonen, Schulleitungen und Schulassistenzen vorgestellt. Eine detaillierte Darstellung der erfassten Konstrukte und der entsprechenden Skalen oder einzelnen Fragen (Items) erfolgt im Rahmen der Darstellung der Ergebnisse in Kapitel 4.

Fragebögen für die Schülerinnen und Schüler

Fragebögen für die Schülerinnen und Schüler wurden an allen ausser dem zweiten Messzeitpunkt eingesetzt und gliederten sich jeweils in einen allgemeinen Teil und einen fach- beziehungsweise

⁵ Für zusätzliche Informationen siehe die Webseite des IBE: <https://www.ibe.uzh.ch/de.html>.

domänenspezifischen Teil. So wurden in den ALLE- und Vergleichsschulen demografische Merkmale (z.B. Alter, Geschlecht), Aspekte des familiären Hintergrunds (z.B. Migrationshintergrund, sozioökonomischer Status), gefolgt von lern- und leistungsrelevanten Merkmalen (z.B. domänenspezifische selbstbezogene Kognitionen und Motivationen) sowie die wahrgenommene Qualität des Deutsch- und Mathematikunterrichts (z.B. Schüler-Lehrer-Beziehung, Unterrichtsklima) erhoben. Während die lern- und leistungsrelevanten Schüler- und Unterrichtsqualitätsmerkmale zu drei Messzeitpunkten erhoben wurden (T1, T3 und T4), wurden die soziodemografischen Merkmale der Schülerinnen und Schüler zu zwei Messzeitpunkten erfasst (T1 und T4). Darüber hinaus wurden in den ALLE-Schulen zum vierten Messzeitpunkt Fragen zur Wahrnehmung der Schulassistenz und der Aufgabenhilfe sowie zur Nutzung des Schülerfeedbacks zum Unterricht durch die Fachlehrperson gestellt (vgl. auch Kap. 3.1).

Fragebögen für die Lehrpersonen

Die am ersten und letzten Messzeitpunkt (T1 und T4) eingesetzten Fragebögen für die Lehrpersonen der ALLE- und Vergleichsschulen waren ähnlich strukturiert wie die Fragebögen für die Schülerinnen und Schüler. Im ersten Teil wurden die Lehrpersonen zu ihrer Person (u.a. Geschlecht, Alter, Berufsausbildung, Berufserfahrung) befragt. Weiterhin wurden die Lehrpersonen zu ihren fach- beziehungsweise domänenspezifischen Überzeugungen zum Lehren und Lernen (Transmissionsorientierung, Konstruktionsorientierung, Schülerleistung als Begabung beziehungsweise angeborener Fähigkeit), zu Überzeugungen zu ihrer Person (Selbstwirksamkeit) und zu Merkmalen ihres Unterrichts in den Fächern Deutsch und/oder Mathematik befragt. Darüber hinaus wurden die Lehrpersonen zu beiden Messzeitpunkten dazu befragt, wie ihre Schülerinnen und Schüler in den Kompetenztests der Evaluation ALLE abschneiden würden (vgl. Kap. 4.8.2). Diese Befragung wurde mithilfe eines separaten Online-Fragebogens durchgeführt (vgl. Kap. 3.2.2). Auch dieser Fragebogen war in zwei Teile gegliedert. Im ersten Teil wurden die Lehrpersonen zu aufgabenbezogenen Einschätzungen auf Grundlage der in den Tests eingesetzten Aufgaben befragt, im zweiten Teil zu personenbezogenen Einschätzungen (vgl. Kap. 4.8.2.1). Der erste Teil war jedoch nur Bestandteil des Fragebogens am ersten Messzeitpunkt (vgl. Kap. 4.8.2.1). Um sicherzustellen, dass die Ausführung des Fragebogens zwischen den Messzeitpunkten vergleichbar war, wurden den Lehrpersonen dennoch auch am vierten Messzeitpunkt Testaufgaben präsentiert. Die Lehrpersonen mussten jedoch keine aufgabenbezogenen Beurteilungen vornehmen.

Ergänzend wurde in den ALLE-Schulen am vierten Messzeitpunkt (bereits etwa Mitte der 9. Klassenstufe, vgl. Tabelle 2) ein Fragebogen für Lehrpersonen eingesetzt, in dem nach der Umsetzung und wahrgenommenen Wirkung der ALLE-Massnahmen gefragt wurde. Der Fragebogen enthielt Fragen zu den Weiterbildungen und Beratungen, die im Rahmen von ALLE in Anspruch genommen wurden (z.B. Nutzung und wahrgenommene Nützlichkeit), zur Unterrichtsdiagnostik und Un-

terichtsreflexion mit den «Evidenzbasierten Methoden der Unterrichtsdiagnostik und -entwicklung» (EMU; z.B. Nutzung und wahrgenommene Nützlichkeit), zur kollegialen Zusammenarbeit in der Schule, zum Einsatz einer Schulassistentin (z.B. Qualität der Zusammenarbeit und Einsatzgebiete) und zur Aufgabenhilfe (z.B. Inanspruchnahme und Betreuung). Damit bildet er eine wichtige Datenquelle, um die Umsetzung der Massnahmen an den Schulen und deren wahrgenommene Wirkungen in der Förderung der Schülerinnen und Schüler zu beschreiben.

Fragebogen für die Schulleitungen

Der Fragebogen für die Schulleitungen wurde am vierten Messzeitpunkt (bereits etwa Mitte der 9. Klassenstufe) eingesetzt und gliederte sich in drei Teile. Im ersten Teil erhielten die Schulleitungen Fragen zu den Rahmenbedingungen der Umsetzung des Pilotprojekts, zum Einsatz von EMU sowie zu Voraussetzungen für die Schulentwicklung (u.a. Leitungshandeln, Innovative Grundhaltung). Der zweite Teil widmete sich der Gestaltung von Schulentwicklungsprozessen. Es wurden Daten zur Schaffung von Strukturen für Entwicklungsprozesse, zum Anstossen und zur Gestaltung innovativer Prozesse, zur Sicherung und Bewertung von Evaluationsergebnissen und zu wahrgenommenen Auswirkungen des Pilotprojekts ALLE erhoben. Im dritten Teil wurden die Schulleitungen zur Zusammenarbeit zwischen Schulleitung und Kollegium (z.B. Collective Self-efficacy), zur Zusammenarbeit im Schulleitungsteam und zur Kommunikation in der Schule befragt.

Fragebogen für Schulassistenten

Die im Schuljahr 2018/2019 eingesetzten Schulassistentinnen und Schulassistenten wurden am vierten Messzeitpunkt (T4; bereits etwa Mitte der 9. Klassenstufe) befragt. Sie machten Angaben zu persönlichen Daten (z.B. Geschlecht, Alter, Berufsausbildung, Berufserfahrung), ihren Weiterbildungen innerhalb und ausserhalb des Pilotprojekts ALLE, ihren Einsatzgebieten in der Schule und im Deutsch- beziehungsweise Mathematikunterricht sowie zu ihrer Zusammenarbeit mit den Deutsch- beziehungsweise Mathematik-Lehrpersonen.

3.4.3 Interviews

Leitfadenentwicklung

Für jede interviewte Personengruppe (Schulleitungen, Lehrpersonen etc.) wurde ein eigener Interviewleitfaden entwickelt. Die Entwicklung der Leitfäden für die Interviews erfolgte auf der Grundlage der Fragestellungen der Ausschreibung der wissenschaftlichen Evaluation und mit Blick auf nicht vollständig durch die quantitativen Daten abgedeckte Fragestellungen. Die Entwicklung der Leitfäden (Fragestellungen, Aufbau, Fragetypen, Dokumentation von Zusatzinformationen) orientierten sich unter anderem am «Handbuch VWA von A bis Z» (Bundesministerium Bildung, Wissenschaft und Forschung, 2018, S. 54-64) sowie bezüglich des Aufbaus insbesondere an Kruse

(2009, 2014). Jeder Leitfaden bestand aus einem standardisierten Einleitungstext, der zu Beginn des Interviews vorgelesen wurde. Danach wurden tabellarisch Hauptfragen in Themenblöcken kategorisiert und mit einer Spalte für Nachfragen versehen. In einer dritten Spalte wurden mögliche Anknüpfungspunkte für den Fortschritt des Interviews formuliert. Das Interview selbst folgte keiner festgelegten Reihenfolge der Fragen, um einen flexiblen Gesprächsverlauf zu ermöglichen.

Fragestellungen

Die Themenblöcke und Unterfragen wurden auf Grundlage der Fragestellungen der Evaluation formuliert und umfassten die in Tabelle 12 ersichtlichen Hauptfragestellungen.

Tabelle 12. Fragebereiche der Interviews mit Zuordnung zu den Personengruppen

Fragebereiche	Personengruppe			
	BP	LP	SA	SL
<i>Pilotprojekt ALLE allgemein</i>				
Veränderungen durch ALLE in der Schule und im Unterricht		x		
Wirkung von ALLE/Veränderungen an der Schule				x
Massnahmen im Projekt ALLE: Nutzen/Wirkung				x
Bilanzierung der Massnahmen				x
Team und ALLE (Akzeptanz von ALLE im Team)				x
Projektverlauf ALLE				x
Bilanzierung von ALLE (grösster Nutzen/Optimierung)				x
Pilotprojekt ALLE: generelles Erleben	x			
<i>EMU</i>				
Wirkungen/Veränderungen im Unterricht durch EMU		x		
Datenreflexion/Umgang mit Daten aus EMU		x		
Bilanzierung von EMU		x		
<i>Schulassistentz</i>				
Nutzen/Wirkung der Schulassistentz		x		
Zusammenarbeit mit der Schulassistentz		x		
Bilanzierung der Massnahme Schulassistentz		x		
Einsatz/Zusammenarbeit mit Lehrperson			x	
Lernwirksamer Nutzen der Zusammenarbeit			x	
Bilanzierung von ALLE, Einsatzoptimierung			x	

(Fortsetzung)

Tabelle 12 (Fortsetzung). Fragebereiche der Interviews mit Zuordnung zu den Personengruppen

Fragebereiche	Personengruppe			
	BP	LP	SA	SL
<i>Aufgabenhilfe</i>				
Wirkung der Aufgabenhilfe		x	x	
Durchführung der Aufgabenhilfe (Gestalten/Erleben)		x	x	
Bilanzierung der Massnahme der Aufgabenhilfe		x	x	
<i>Beratung</i>				
Beratungsangebote (in Anspruch genommene)		x		x
Bilanzierung der Beratungsangebote	x	x		
Inhalt/Durchführung der Beratungen	x			
Wirkungen der Beratungen	x			
<i>Weiterbildungen</i>				
Weiterbildungsangebote: Nutzen/Wirkungen		x	x	
Bilanzierung der Weiterbildungsangebote	x	x	x	
Angebotene Inhalte	x			
Wirkungen der angebotenen Weiterbildungen	x			
<i>Ergebnisrückmeldungen</i>				
Nutzen, Diskussion und Unterrichtsentwicklung		x		
Bilanzierung der Ergebnisrückmeldungen		x		
<i>Spezifische Wirkungen</i>				
Individuelle Lernförderung		x		
Zusammenarbeit im Kollegium		x		
Unterrichtswirksame Führung der Schulleitung		x		

Anmerkungen. BP = Beratungs- und Weiterbildungspersonen, LP = Lehrpersonen und schulische Heilpädagoginnen und Heilpädagogen, SA = Schulassistenzen, SL = Schulleitungen.

Abschliessend wurde in jedem Interview folgende Frage gestellt: «Von unserer Seite haben wir hiermit alle Fragen zum Projekt ALLE gestellt. Haben Sie noch zusätzliche Punkte, die noch nicht genannt/besprochen wurden und welche Sie gerne einbringen möchten?»

Nach jedem Interview wurden stichwortartig die Basisinformationen, Besonderheiten und Auffälligkeiten zur Interviewsituation, den interviewten Personen, dem Verlauf der Interviews sowie weitere

Ergänzungen notiert. Hinsichtlich möglicher Besonderheiten während der Interviewdurchführung wurden insbesondere folgende Aspekte protokolliert:

- Von einer Schule gaben nur eine beziehungsweise zwei der drei Lehrpersonen an, uneingeschränkt an der Evaluation ALLE beteiligt gewesen zu sein. Die übrigen Lehrpersonen gaben an, nur an der Pilotierungsphase von ALLE dabei gewesen zu sein.
- In einer Schule mussten Lehrpersonen während des Interviews den Raum zwischenzeitlich verlassen, aufgrund von parallel stattfindendem Unterricht. In dieser Schule traf eine Lehrperson zudem mit einer rund 20-minütigen Verspätung ein. Auch gaben zwei der Lehrpersonen an, erst seit ca. einem Jahr an der Schule tätig zu sein.
- In einer Schule nahm die neue Schulleiterin, welche nicht für ein Einzelinterview eingeplant war, zusätzlich am Gruppeninterview der Lehrpersonen/schulischen Heilpädagoginnen und Heilpädagogen teil.
- In einer Schule war eine Lehrperson mehr beim Gruppeninterview anwesend als eingeplant (vier statt drei).
- In einer Schule fand das Interview vor allem mit der schulischen Heilpädagogin/dem schulischen Heilpädagogen und einer Lehrperson statt. Die anderen Lehrpersonen, die vorwiegend Niveaustufe A unterrichteten, gaben an, dass ALLE für schwächere Schülerinnen und Schüler sei und sie am Pilotprojekt teilnehmen müssten, obwohl sie keine leistungsschwächeren Schülerinnen und Schüler hätten.

3.5 Datenaufbereitung und Datenanalyse

In diesem Kapitel werden wichtige Schritte in der Aufbereitung und Analyse der Daten näher erläutert. Dabei muss ein Grundverständnis für den Einsatz und die Interpretation quantitativer Analyseverfahren (z.B. beim Konzept statistischer Signifikanz) vorausgesetzt werden. In mancher Hinsicht gehen die eingesetzten Methoden jedoch über gängige Verfahren hinaus und bedürfen zusätzlicher Erläuterungen. In diesem Kapitel wird auf folgende Aspekte der Datenaufbereitung und Datenauswertung eingegangen:

- Aufbereitung soziodemografischer Daten (Kap. 3.5.1),
- Auswertung der Kompetenztests: IRT-Modelle und Linking-Methode (Kap. 3.5.2),
- Multiple Imputation von fehlenden Werten (Kap. 3.5.3),
- Erzeugung von Plausible Values für die Kompetenztestdaten (Kap. 3.5.4),
- Datenanalysen mit multipel imputierten Daten und Plausible Values (Kap. 3.5.5),
- Transkription der Interviews (Kap. 3.5.6),
- Kodierprozess und Kategoriensystem (Kap. 3.5.7).

Aus Platzgründen können nur die Grundzüge der Verfahren sowie ihre generelle Verwendung in der Evaluation ALLE dargestellt werden. Die konkret spezifizierten Analysen werden direkt in den Ergebniskapiteln kurz beschrieben (angelehnt an Eder, Altrichter, Hofmann & Weber, 2015).

3.5.1 Aufbereitung der soziodemografischen Merkmale

Die Indikatoren für die Bildung der Variablen zum soziodemografischen Hintergrund der Schülerinnen und Schüler stammten aus der Befragung zum vierten Messzeitpunkt (vgl. Kap. 3.4.2). Sie wurden zur Beschreibung der Schülerstichprobe (vgl. Kap. 3.3.2.1), für die Erzeugung der verwendeten Kompetenzschätzungen (vgl. Kap. 3.5.4) und in Regressionsmodellen als Kontrollvariablen herangezogen (vgl. Kap. 4). Waren Angaben zu den Indikatoren vom vierten Messzeitpunkt nicht vorhanden (z.B. aufgrund von Nichtteilnahme an der Erhebung), wurde auf die Angaben der Schülerinnen und Schüler vom ersten Messzeitpunkt zurückgegriffen. Danach noch fehlende Werte wurden imputiert (siehe Kap. 3.5.3). Die Bildung der Variablen zum soziodemografischen Hintergrund folgte mit wenigen Ausnahmen (siehe Erfassung des Migrationshintergrunds) etablierten Vorgehensweisen aus Large-Scale-Assessments im Bildungsbereich (wie PISA 2015, vgl. OECD, 2016; ÜGK 2017, vgl. Pham, Hebling, Verner & Ambrosetti, 2019).

Geschlecht. Das Geschlecht der Schülerinnen und Schüler wird durch eine zweistufige Variable abgebildet. Die weiblichen Lernenden erhalten auf dieser Variable den Wert 1, die männlichen Lernenden den Wert 0.

Alter. Die Variable bildet das Alter der Lernenden zum letzten Messzeitpunkt der Evaluation (bzw. am letzten Schultag des Schuljahres 2018/19) in Jahren ab.

Zuhause gesprochene Sprache. Die Variable unterscheidet zwischen Schülerinnen und Schüler, die (1) nur die Schulsprache (Deutsch oder deutschen Dialekt) zu Hause sprechen, (2) die Schulsprache (Deutsch oder deutschen Dialekt) und andere Sprache(n) zu Hause sprechen, und (3) die ausschliesslich andere Sprache(n) zu Hause sprechen.

Migrationshintergrund. Angelehnt an PISA 2015 (OECD, 2016) und ÜGK 2017 (Pham et al., 2019) hat eine Schülerin/ein Schüler im vorliegenden Bericht keinen Migrationshintergrund, wenn mindestens ein Elternteil in der Schweiz geboren ist. Auf Wunsch der Auftraggeberin wird die Gruppe der Schülerinnen und Schüler ohne Migrationshintergrund weiter unterschieden in: (1) Schülerinnen und Schüler, deren Eltern beide in der Schweiz geboren sind, und (2) Schülerinnen und Schüler, bei denen ein Elternteil in der Schweiz geboren ist. Ist die Schülerin/der Schüler in der Schweiz geboren, aber beide Elternteile sind ausserhalb der Schweiz geboren, gehört sie/er zur Gruppe der (3) Schülerinnen und Schüler mit Migrationshintergrund der zweiten Generation. Sind sowohl die Schülerin/der Schüler als auch beide Elternteile nicht in der Schweiz geboren, zählt sie/er zur Gruppe der (4) Schülerinnen und Schüler mit Migrationshintergrund der ersten Generation.

Sozioökonomischer Status. Zur Bestimmung des sozioökonomischen Status der Schülerinnen und Schüler wurde durch Anwendung eines generalisierten Partial-Credit-Modells (GPCM; Muraki, 1992) ein Index aus folgenden Variablen gebildet:

- dem höchsten Bildungsabschluss der Eltern,
- der Anzahl der Bücher im Elternhaus,
- dem Vorhandensein im Elternhaus von: für die Hausaufgaben hilfreichen Büchern; Büchern über Kunst, Musik oder Design; einem Musikinstrument (ohne Blockflöte); einer abonnierten Tageszeitung.

Der Index wurde z-standardisiert und weist entsprechend einen Mittelwert von 0 und eine Streuung (Standardabweichung) von 1 auf.

3.5.2 IRT-Skalierung der Kompetenztests und Linking

Die Kodierung der Schülerantworten auf die Testaufgaben in den Kompetenztests Deutsch und Mathematik – als richtig oder falsch oder gegebenenfalls nicht bearbeitet – erfolgte automatisiert auf der technischen Plattform von Lernpass plus. Zur weiteren Auswertung der kodierten Testantworten wurden Verfahren der Item-Response-Theorie (IRT; siehe z.B. Moosbrugger & Kelava, 2012) herangezogen. An allen Messzeitpunkten wurde ein dichotomes Rasch-Modell (Rasch, 1960) auf die Testdaten angepasst und zur Schätzung der Aufgabeneigenschaften (Aufgabenschwierigkeiten) und Personeneigenschaften (Testwerte der Schülerinnen und Schüler) verwendet. Von den Schülerinnen und Schülern übersprungene Aufgaben (*omitted items*) wurden als falsch und nicht erreichte Aufgaben (*not-reached items*) als fehlend gewertet (siehe z.B. Pohl, Gräfe & Rose, 2014, für einen Überblick über den Umgang mit fehlenden Testantworten in Bildungsstudien). Um die psychometrische Eignung der Aufgaben zu prüfen, wurde die Passung zum Rasch-Modell auf der Grundlage gängiger Fit-Indizes evaluiert. Als zentrales Kriterium diente der *Weighted Mean Square Index* (WMNSQ; siehe z.B. Wilson, 2005), der auf eine insgesamt gute Passung der Testaufgaben an allen Messzeitpunkten hinwies.

Da für die Evaluation die Entwicklung der Kompetenzen während des Pilotprojekts im Vordergrund stand, kam der Messung von Veränderungen eine besondere Bedeutung zu. Um Veränderungen in Kompetenzen erfassen zu können, wurden bestimmte Aufgaben an mehreren Messzeitpunkten eingesetzt («Link-Items»), um eine Verknüpfung (*Linking*; z.B. Kolen & Brennan, 2014) zwischen den Messzeitpunkten zu ermöglichen und die Testergebnisse auf einer gemeinsamen Kompetenzskala abzubilden. Dem Linking der Testergebnisse unter Verwendung von Link-Items liegt die Annahme zugrunde, dass diese Items über die Messzeitpunkte hinweg in beiden Stichprobengruppen (ALLE- und Vergleichsschulen) gleich funktionieren (vgl. Lord, 1980). Zur Überprüfung der Gültigkeit dieser Annahme wurden mehrere Analysen zum *Differential Item Functioning* (DIF; Holland & Wainer, 1993) durchgeführt. In beiden Fächern (Deutsch und Mathematik) konnte die Annahme

unter Berücksichtigung der ETS-Klassifizierung von DIF (Zwick, 2012) bestätigt werden. Ferner wurde überprüft, ob die in der Normierungsstudie von Lernpass plus (zu T2; vgl. Kap. 3.4.1) gewonnenen Aufgabenparameter für die Skalierung der Testergebnisse in der Evaluation ALLE übernommen werden können. Dafür wurde einerseits die Modellpassung von Skalierungsmodellen mit fixierten Aufgabenparametern aus der Normierungsstudie und Modellen mit frei geschätzten Aufgabenparametern verglichen, andererseits die DIF-Analysen zwischen den Messzeitpunkten und Stichprobengruppen (ALLE- und Vergleichsschulen) unter Verwendung der fixierten Aufgabenparameter aus der Normierungsstudie in beiden Fächern erneut durchgeführt. Die Ergebnisse dieser Prüfungen deuteten darauf hin, dass im Fach Deutsch die Aufgabenparameter aus der Normierungsstudie übernommen werden können, während dies im Fach Mathematik nicht der Fall ist. Entsprechend wurden im Fach Deutsch die Aufgabenparameter aus der Normierungsstudie für die Skalierung in beiden Stichprobengruppen und für das Linking der Ergebnisse zwischen den Messzeitpunkten übernommen. Im Fach Mathematik wurden alle Aufgabenparameter frei geschätzt, in einem simultanen Kalibrierungs- und Skalierungsmodell mit den Daten von allen Messzeitpunkten und aus beiden Stichprobengruppen. Die in diesen Skalierungsmodellen gewonnenen Personenparameter dienten als Startwerte, um die endgültigen Testwerte der Schülerinnen und Schüler für die Analysen zu schätzen (siehe Kap. 3.5.4).

3.5.3 Datenimputation

Fehlende Werte (*missing data*) sind ein typisches Problem sozialwissenschaftlicher Forschung, dem mit neueren statistischen Verfahren jedoch angemessen begegnet werden kann (siehe z.B. Lüdtke, Robitzsch, Trautwein & Köller, 2007). Eines dieser Verfahren ist die *Multiple Imputation*, bei der jeder fehlende Wert durch eine festgelegte Anzahl (z.B. fünf) geschätzter Werte ersetzt wird. Dafür wird Information über die Ausprägung der fehlenden Werte genutzt, die in anderen (verfügbaren) Daten einer Person, Schulklasse oder Schule enthalten sind. Multiple Imputation wurde in der Evaluation ALLE eingesetzt, um die fehlenden Werte aller teilnehmenden Schülerinnen und Schüler (bspw. in den Kompetenztests und Fragebögen) an allen vier Messzeitpunkten zu schätzen. Mit diesem Vorgehen wurde der Tatsache Rechnung getragen, dass bei den Variablen der Schülerinnen und Schüler fehlende Werte in nicht zu vernachlässigendem Umfang auftreten. Der Anteil fehlender Werte pro Variable betrug für die ALLE-Schulen zwischen 2.8 und 36.9 Prozent, für die Vergleichsschulen zwischen 3.6 und 48.5 Prozent. Aufgrund der eingeschränkten Stichprobengrößen konnte eine multiple Imputation für weitere Datenquellen (z.B. die Daten der Lehrpersonen) nicht umgesetzt werden.

Der Imputationsprozess folgte den Vorgehensweisen von Large-Scale-Assessments im Bildungsbereich, die in den nationalen Studien in Österreich (Robitzsch, Pham & Yanagida, 2016) oder in der Schweiz (Pham et al., 2019) eingesetzt werden. Dieser Prozess beruht auf der Annahme, dass

die fehlenden Werte im Datensatz *missing at random* (Rubin, 1987) sind. Dabei wird angenommen, dass das Auftreten fehlender Werte nach Berücksichtigung der vorhandenen Daten zufällig ist.

Für die Imputation der fehlenden Werte der Schülerinnen und Schüler wurden folgende Variablen berücksichtigt:

- *Schülerebene:*
 - Soziodemografische Daten der Schülerinnen und Schüler: Geschlecht, Alter, sozio-ökonomischer Status, zuhause gesprochene Sprache, Migrationshintergrund
 - Fach- und messzeitpunktspezifisch
 - Skalierte Kompetenztestwerte in beiden Fächern
 - Kompetenztestwerte zum „Texte schreiben“ (nur für T3 verfügbar)
 - Skalenwerte (Lernmotivation und eingeschätzte Unterrichtsqualität aus Schülersicht) und Einzelitems aus dem Schülerfragebogen
- *Unterrichtsebene:* fach- und messzeitpunktspezifisch
 - Abteilungszugehörigkeit
 - Aggregierte Klassen-/Lerngruppenmittelwerte aller Variablen auf Schülerebene
- *Schulebene:*
 - Gruppenzugehörigkeit: ALLE- vs. Vergleichsschulen
 - Aggregierte Schulmittelwerte aller Variablen auf Schülerebene

Vor der Bildung der Skalenwerte (Skalen aus dem Schülerfragebogen) wurden zunächst die Reliabilität, Mittelwerte und Standardabweichungen der Skalen berechnet. Die Werte für das Cronbachs Alpha der Skalen variierte zwischen 0.70 und 0.94 (vgl. Anhang 7.1). Bei der Bildung der individuellen Skalenwerte wurde die von Van Ginkel et al. (2007) vorgeschlagene Methode zur Korrektur unterschiedlicher Itemschwierigkeiten verwendet, die im R-Paket *miceadds* (Robitzsch & Grund, 2020) implementiert ist. Um eine zu hohe Modellkomplexität zu vermeiden, wurden anstelle der einzelnen Fragebogenitems jeweils die Skalenwerte imputiert.

Neben den oben genannten Variablen wurden auch die Interaktionen zwischen allen Variablen mit dem Gruppenstatus (ALLE- vs. Vergleichsschule) für die Imputation genutzt, um mögliche gruppenspezifischen Zusammenhänge in den Daten abzubilden. Ebenfalls wurden Interaktionen zwischen allen anderen Variablen im Imputationsmodell mit den soziodemografischen Variablen, mit den Kompetenztestdaten sowie mit der Abteilungszugehörigkeit mit berücksichtigt.

Daten, die in den Vergleichsschulen nicht gewonnen wurden, konnten nicht in die Imputation einbezogen werden. Dies betrifft insbesondere die zum vierten Messzeitpunkt (T4) erhobenen Fragebogendaten zu den Rahmenbedingungen und Massnahmen im Pilotprojekt. Daneben konnten die

mit dem Lehrerfragebogen zu T1 und T4 erhobenen Daten aufgrund häufig wechselnder Zuordnungen zwischen Lehrpersonen und Klassen/Lerngruppen nicht berücksichtigt werden. Dies ist für die Analysen mit den imputierten Daten jedoch unproblematisch, da die Variablen aus den Lehrerfragebögen nicht direkt mit den imputierten Daten in Beziehung gesetzt werden.

Die Imputation der fehlenden Werte erfolgte nach dem Verfahren der *Multiple Imputation by Chained Equations* (MICE; van Buuren, 2012). Dieser Prozess zerlegt die Imputation der Variablen in einzelne Imputationsschritte, in denen jede Variable separat imputiert wird. Die Imputation selbst basiert auf einer linearen Regression. Zur Imputation jeder einzelnen Variable wurden jeweils sämtliche verfügbare Daten aller anderen Variablen desselben Messzeitpunkts und aller anderen Messzeitpunkte sowie alle Interaktionen als Prädiktoren in Betracht gezogen. Aufgrund der sehr hohen Anzahl an Prädiktoren (jeweils > 500) waren jedoch Probleme wie verzerrte und stark streuende Parameterschätzungen zu erwarten (Hardt, Herke & Leonhart, 2012). Daher wurde die Anzahl der Prädiktoren schrittweise reduziert. Zuerst wurden Prädiktoren, die wenig Vorhersagekraft für alle anderen Variablen aufwiesen, nicht ins Imputationsmodell aufgenommen. Danach wurden aus den verbleibenden Prädiktoren mittels der *Partial-Least-Squares*-Methode (PLS; Abdi, 2010) schrittweise wenige unkorrelierte Faktoren extrahiert, mit denen die Kovarianz zwischen den Prädiktoren und den zu imputierenden Variablen maximiert wurde. Um sicherzustellen, dass die imputierten Werte nicht ausserhalb des gültigen Wertebereichs pro Variable sind, wurde als Imputationsmethode *Predictive Mean Matching* (PMM) gewählt (Robitzsch et al., 2016). Der Imputationsprozess wurde iterativ durchlaufen: Alle Variablen wurden nicht nur einmal, sondern mehrmals imputiert; die Ergebnisse eines Imputationsvorgangs dienten dabei als Startwerte für den jeweils nächsten Imputationsvorgang. Nach jedem Vorgang wurden Plausible Values für die Kompetenztestdaten generiert (siehe Kap. 3.5.4), die im nächsten Imputationsschritt wiederum als Prädiktoren zur Imputation der fehlenden anderen Daten verwendet wurden. Für die Analysen im Bericht wurden die imputierten Werte und Plausible Values der Kompetenztestdaten nach 40 Iterationen gespeichert und verwendet. Insgesamt wurden zehn vollständig imputierte Datensätze generiert. Für eine detaillierte Beschreibung des Prozesses siehe Robitzsch et al. (2016).

3.5.4 Schätzung der fachlichen Lernergebnisse: Plausible Values

Wie in vielen der grossen Bildungsstudien (z.B. PISA, vgl. OECD, 2020; ÜGK, vgl. Pham et al., 2019) wurden im vorliegenden Bericht *Plausible Values* (PVs; siehe Mislevy, 1991; von Davier, Gonzalez & Mislevy, 2009; Wu, 2005) verwendet, um in nachfolgenden Analysen messfehlerfreie und unverzerrte Parameterschätzungen (z.B. Korrelationen, Regressionskoeffizienten) zu erhalten. Zur optimalen Schätzung der PVs wird ein sogenanntes Hintergrundmodell benötigt, welches zumindest alle für spätere Analysen verwendeten Prädiktoren der Leistungsdaten umfassen sollte (vgl. Robitzsch et al., 2016). In der Evaluation ALLE wurden dafür sämtliche Daten aus den Schü-

lerfragebögen aller Messzeitpunkte sowie verschiedene Kontextdaten der Schulen und der Lerngruppen/Klassen einbezogen. Zur Ziehung der PVs war es erforderlich, dass die Variablen im Hintergrundmodell keine fehlenden Daten aufweisen. Daher wurde, wie in Kapitel 3.5.3 beschrieben, der Prozess der Imputation der Schülerdaten mit der Schätzung der PVs verknüpft. In diesem Prozess konnten – auf der Grundlage umfangreicher Daten von mehreren Messzeitpunkten – auch fehlende Antworten in den Schülerfragebögen und Kompetenztestdaten von Schülerinnen und Schülern, die an bestimmten Messzeitpunkten nicht an der Evaluation ALLE teilgenommen hatten, geschätzt werden. Insgesamt wurden zehn vollständig imputierte Datensätze mit Plausible Values der Kompetenztestdaten generiert. Sie dienen als Datenbasis für sämtliche Analysen in diesem Bericht, mit Ausnahme der explorativen Analysen, die in Kapitel 4.11 berichtet werden.

3.5.5 Datenanalysen mit PVs und multipel imputierten Datensätzen

Um mit multipel imputierten Datensätzen zu rechnen, wurden die Analysen gemäss den Regeln von Rubin (1987) durchgeführt. Dabei wird die Analyse zunächst mit jedem imputierten Datensatz umgesetzt, dann werden die Ergebnisse über alle imputierten Datensätze verrechnet. Zur Gewinnung der Standardfehler (Stichprobenfehler) in den berichteten Analysen – auf denen auch Aussagen zur statistischen Signifikanz basieren – wurde jeweils separat für die Daten der ALLE- und Vergleichsschulen die *Resampling*-Methode auf Basis von jeweils 1'000 Bootstrap-Stichproben verwendet (Efron & Tibshirani, 1986). Diese Methode ermöglicht die Schätzung der Standardfehler unter Berücksichtigung von hierarchischen Datenstrukturen (Schülerinnen und Schüler geschachtet in Klassen und Schulen), die ansonsten zu einer Unterschätzung der Standardfehler führen können (vgl. z.B. Hox, 2010). Die Berechnung der Standardfehler aller berichteten Schätzungen erfolgte (soweit nicht anders beschrieben) nach Standardverfahren in Large-Scale-Assessments (siehe Enders, 2010; vgl. Bruneforth, Oberwimmer & Robitzsch, 2016; Pham et al., 2019).

3.5.6 Transkription der Interviews

Die Transkription der Audioaufnahmen der Interviews (vgl. Kap. 3.2.3) erfolgte mittels festgelegter Regeln durch Mitarbeitende der Pädagogischen Hochschule St.Gallen (PHSG). Die Regeln wurden in einem früheren Projekt der PHSG⁶ verwendet und basieren auf der Literatur von Deppermann und Schütte (2008), Langer (2010) und Kuckartz (2010) sowie ergänzend auf Kuckartz,

⁶ Projekt «Kooperative Praxis von Assistenzpersonen und Lehrpersonen im Unterricht der Regelschule» (KoPrAss); siehe <https://www.phsg.ch/de/forschung/projekte/kooperative-praxis-von-assistenzpersonen-und-lehrpersonen-im-unterricht-der>

Dresing, Stefer und Rädiker (2007) sowie Dresing und Pehl (2011). Folgende grundsätzlichen Punkte wurden festgelegt:

- Formal:
 - Jede beziehungsweise jeder Sprechende erhält einen Absatz mit Code zur Identifikation, z.B. «I» für Interviewende.
 - Die Lehrpersonen erhalten eine aufsteigende Nummerierung mit dem Kürzel «LP», d.h. «LP1», «LP2», usw.
 - Es werden Zeilennummern für das Transkript gesetzt.
 - Bei einem Wechsel der Sprechenden Person wird eine Zeitmarke gesetzt.
- Sprachlich/inhaltlich:
 - «Small Talk» wird nicht transkribiert.
 - Die Transkription erfolgt wortwörtlich.
 - Begriffe aus dem Dialekt werden möglichst exakt in das Hochdeutsche übersetzt.
 - Wenn möglich wird ein Punkt statt eines Kommas gesetzt.
 - Wortverschleifungen (z.B. «so'n») werden geglättet (im Beispiel: in «so ein»).
 - Grammatikalische und syntaktische beziehungsweise reihenfolgebezogene Fehler werden transkribiert und nicht korrigiert.
 - Für Pausen, unverständliche Worte etc. wurden spezifische Zeichenfolgen definiert.

Die Transkription (und anschließende Kodierung) erfolgte mittels der Software *MAXQDA 2018 Portable for Windows* (VERBI Software, 2017). Den transkribierenden Personen wurde vor Beginn der Transkription eine Transkriptionsanleitung mit einem Schritt-für-Schritt-Vorgehen ausgehändigt. Weiter fand eine kurze Schulung mit Erklärung der Transkriptionsregeln statt, und es bestand die Möglichkeit, sich nach einer Probephase mit der Transkription nochmals mit Fragen und Unklarheiten an das Evaluationsteam zu wenden und diese zu besprechen.

3.5.7 Kodierprozess und Kategoriensystem

Die Auswertung der Interviewdaten erfolgt im vorliegenden Bericht mit Blick auf die Rahmenbedingungen des Pilotprojekts und die Umsetzung der Massnahmen an den ALLE-Schulen (Kapitel 4.1 bis 4.7). Gemeinsam mit der Auftraggeberin wurde beschlossen, dass für die Interviewdaten ein geeignetes Kategoriensystem entwickelt wird und die Aussagen in den Interviews auf Basis dieses Systems kodiert werden. Darüber hinaus wurde festgelegt, die kodierten Interviewpassagen im Bericht primär ergänzend zu den Ergebnissen aus den Fragebögen – im Sinne einer Untermauerung und Illustration der Fragebogenaussagen – zu verwenden. Entsprechend werden in Kapitel 4.1 bis 4.7 zunächst und schwerpunktmässig die Resultate aus der Auswertung der Fragebögen

dargestellt, die dann um jeweils passende Zitate aus den Interviews ergänzt werden. Eine vertiefende (z.B. quantitativ-inhaltsanalytische) Auswertung der Interviews erfolgte aufgrund des damit verbundenen hohen Aufwands für den vorliegenden Bericht nicht. Das Vorgehen bei der Aufbereitung und Kodierung des Interviewmaterials wird im Folgenden näher beschrieben.

Für jede der vier Interviewtengruppen (Beratungs-/Weiterbildungspersonen, Lehrpersonen und schulische Heilpädagoginnen und Heilpädagogen, Schulassistenten und Schulleitungen) wurde ein Kodiersystem erstellt, indem in einem ersten Schritt in einer deduktiven Vorgehensweise aus den Hauptfragestellungen und Nachfragen der jeweiligen Leitfäden entsprechende Überkategorien mit Unterkategorien abgeleitet und formuliert wurden. Anschliessend wurde eine Kodierschulung mit den insgesamt drei Kodierenden durchgeführt. Diese beinhaltete sowohl die Erklärung der Über- und Unterkategorien anhand von Interviewmaterial sowie die Kodierung mithilfe der Software *MAXQDA 2018 Portable for Windows* (VERBI Software, 2017).

Anschliessend wurde eine Phase der testweisen Kodierung durchgeführt und die Kodierenden erhielten die Möglichkeit, sich mit Unklarheiten und Fragen an das Evaluationsteam zu wenden. Im Rahmen dieser Phase wurden Kategorien, welche zu Missverständnissen und Unklarheiten führten, umbenannt und/oder zusammengeführt. Während des Kodierprozesses wurden zusätzlich zur deduktiven Kategorienbildung in einem induktiven Vorgehen (siehe auch Kuckartz, 2016) auf der Grundlage des Datenmaterials Kategorien neu hinzugefügt und bestehende Kategorien zusammengeführt. Die Interviews wurden für diese Kategorien nochmals kodiert.

Die Zuverlässigkeit der Kodierungen (Interraterreliabilität) wurde jeweils auf Basis von 10 Prozent des transkribierten Materials (jeweils einem zufällig ausgewählten Interview für jede Interviewtengruppe) bestimmt. Dieses Material wurde von einer zweiten kodierenden Person gegenkodiert. Zur Berechnung der Interraterreliabilität wurde Cohens Kappa (Cohen, 1960) in der Form nach Brennan und Prediger (1981) eingesetzt. Für die Interpretation der Koeffizienten können die Leitlinien nach Landis und Koch (1977) herangezogen werden. Demnach lassen sich Koeffizienten von .81 bis 1.00 als fast perfekt («almost perfect»), von .61 bis .80 als gut («substantial») und von .41 bis .60 als akzeptabel («moderate») einordnen.

Tabelle 13 gibt einen Überblick, wie viele der Oberkategorien (Ebene 1) aus der Kodierung der Interviews mit den verschiedenen Interviewtengruppen in die unterschiedlichen Kategorien dieser Klassifikation fallen. Aus der Tabelle wird ersichtlich, dass die Zuordnung zu einem grossen Teil der Kategorien (v.a. für die Lehrpersonen/SHP- und Schulassistenteninterviews) mit zumindest zufriedenstellender Zuverlässigkeit möglich war. Andererseits konnten die Aussagen der Interviewten nicht für alle Kategorien mit der wünschenswerten Genauigkeit zugeordnet werden, insbesondere für die Schulleitungsinterviews. Zu beachten ist allerdings, dass von allen als «schwach» klassifizierten Oberkategorien nur vier Kategorien fünf oder mehr Aussagen zugeordnet wurden. Es han-

delt sich somit fast durchweg um (sehr) schwach besetzte, praktisch wenig bedeutsame Kategorien. Zudem waren im vorliegenden Bericht die Kategorienzugehörigkeiten selbst nicht Gegenstand der Analysen (es wurden z.B. keine Häufigkeitsauszählungen durchgeführt). Wie zu Beginn dieses Kapitels erläutert, dienten die Kategorien stattdessen primär zur Orientierung für die Auswahl passender Zitate, welche im Ergebnisteil (Kap. 4) zur Ergänzung und Verdeutlichung der quantitativen Analysen herangezogen werden.

Tabelle 13. Verteilung der Interraterreliabilität (Cohens Kappa nach Brennan & Prediger, 1981) für die Oberkategorien in der Kodierung der Interviews nach Interviewtengruppen

Interview	Sehr gut	Gut	Akzeptabel	Schwach
Lehrpersonen/SHP	6	6	3	3
Schulassistenzen	1	4	3	2
Schulleitungen	0	8	4	10
Beratungs-/Weiterbildungspersonen	1	1	0	2

Anmerkungen. SHP = Schulische Heilpädagogin/schulischer Heilpädagoge. Einordnung nach Cohens Kappa nach den Empfehlungen von Landis und Koch (1977): .81 bis 1.00 sehr gut, .61 bis .80 gut, .41 bis .60 akzeptabel, bis .40 schwach.

3.6 Hinweise zur Interpretation der Ergebnisse

Wie sind die berichteten Testwerte aus den Kompetenztests zu interpretieren?

Um die Interpretation der berichteten fachlichen Lernergebnisse zu erleichtern, wurden die in den Analysen verwendeten Kompetenztestwerte (Plausible Values; vgl. Kap. 3.5.4) der Schülerinnen und Schüler auf eine einfach verständliche Metrik transformiert. Die Metrik der berichteten Punktwerte wurde so festgesetzt, dass der Mittelwert in den Vergleichsschulen zum ersten Messzeitpunkt der Evaluation ALLE (T1; Anfang 7. Klassenstufe) für jeden Kompetenzbereich 500 beträgt, die Streuung (Standardabweichung) 100. Demgemäss lassen sich Punktwerte über 500 im Sinne von fachlichen Kompetenzen interpretieren, die über dem Durchschnitt der Schülerinnen und Schüler in den Vergleichsschulen zu Beginn der Evaluation liegen. Punktwerte unter 500 liegen entsprechend unter diesem Durchschnitt. Der Gesamtwert im Fach (Deutsch bzw. Mathematik) wurde für jede Schülerin/jeden Schüler als Mittelwert der Testwerte in den zugehörigen Kompetenzbereichen (siehe Kap. 3.4.1) bestimmt. Daraus ergibt sich, dass der Mittelwert der Testwerte im gesamten Fach in den Vergleichsschulen zum ersten Messzeitpunkt ebenfalls 500 beträgt, die Standardabweichung jedoch kleiner als 100 ist.

Für den Test «Texte schreiben», dessen Ergebnisse ergänzend zu den in der Evaluation ALLE zusammengestellten Tests herangezogen wurden (vgl. Kap. 3.4.1), wurde in einer Normierungsstudie im Kanton Zürich ein Mittelwert von 500 und eine Standardabweichung von 100 festgelegt.

Wie erfolgt die Bewertung der statistischen Signifikanz?

Zur Bewertung der statistischen Signifikanz wird das gängige Fehlerniveau von $\alpha = 5\%$ ($p < 0.05$) herangezogen. Zusätzlich werden auch tendenziell signifikante Effekte mit $p < .10$ interpretiert. Dieses Vorgehen kann Weber und Kollegen (2015) folgend damit begründet werden, dass die Fallzahl auf der Ebene der Schulen verhältnismässig gering ist. Tendenziell signifikante Effekte werden jedoch im vorliegenden Bericht als weniger belastbar eingestuft, und es wird bei der Darstellung und Diskussion der Ergebnisse auf diese Einschränkung hingewiesen.

Statistische Signifikanztests werden im vorliegenden Bericht ausschliesslich zur Prüfung der für die Evaluation zentralen Unterschiede eingesetzt, das heisst für die Disparitäten zwischen den ALLE- und Vergleichsschulen. Zudem beschränkt sich der Einsatz von Signifikanztests auf die Analysen zur Unterrichtsqualität aus Sicht der Schülerinnen und Schüler (Kap. 4.9.2), zu den motivational-affektiven Lernergebnissen (Kap. 4.10.1) und den fachlichen Lernergebnissen (Kap. 4.10.2). Der Verzicht auf die Absicherung durch Signifikanztests in den anderen Abschnitten ist durch die vergleichsweise geringen Stichprobengrössen, insbesondere im Längsschnitt über den Evaluationszeitraum, bedingt (vgl. Kap. 3.3.2). Bei Analysen ohne Signifikanztests liegt der Schwerpunkt auf der vergleichenden Beschreibung für die vorliegenden Stichproben der ALLE- und Vergleichsschulen. Schlüsse über die möglichen kausalen Wirkungen von ALLE von den am Pilotprojekt teilnehmenden Schulen auf andere Schulen im Kanton Zürich können bei diesen Analysen nur auf der Grundlage zusätzlicher Annahmen erfolgen (siehe auch Kap. 5.4.1).

Wie erfolgt die Einordnung von Effektgrössen?

Zur Einordnung der praktischen Bedeutsamkeit von Mittelwertsveränderungen während des Evaluationszeitraums und von Mittelwertsunterschieden zwischen den ALLE- und Vergleichsschulen wird das *Effektstärkemass* d nach Cohen (1992) verwendet. Bei Vergleichen, die anhand von Signifikanztests geprüft werden, wird d nur im Fall eines statistisch signifikanten Ergebnisses interpretiert. Darüber hinaus wird d herangezogen, um bei deskriptiven Analysen beziehungsweise bei Analysen, für die keine Signifikanztests verwendet werden, Veränderungen oder Unterschiede *beschreibend* darzustellen. Effekte im Betrag zwischen 0.15 und 0.35 werden als schwach, Effekte im Betrag zwischen 0.35 und 0.55 als von mittlerer Stärke und Effekte im Betrag über 0.55 als stark bewertet. Die damit verbundene Reduktion der Anforderungen an schwache, mittlere und starke Effekte gegenüber den üblicheren Kriterien bei Cohen (1992; mit Grenzen bei 0.2, 0.5 und 0.8) erscheint angemessen, da Letztere für kontrollierte (experimentelle) Bedingungen entwickelt wurden (vgl. Weber et al., 2015, S. 247).

Wie werden die Unterschiede zwischen den ALLE- und Vergleichsschulen geprüft?

Um Unterschiede zwischen den ALLE- und Vergleichsschulen zu bestimmen und auf statistische Signifikanz zu prüfen, werden Regressionsmodelle eingesetzt. Details zur Spezifikation dieser Modelle werden in den entsprechenden Abschnitten des Ergebnisteils dargestellt. Bezüglich der Interpretation der Ergebnisse ist wichtig festzuhalten, wie die berichteten Regressionskoeffizienten (Regressionsgewichte) zu verstehen sind. Für alle Regressionsmodelle wird ausschliesslich der Koeffizient für den Unterschied zwischen ALLE- und Vergleichsschulen berichtet. Dieser Unterschied wird durch eine zweistufige (Dummy-)Variable abgebildet, welche in die Regressionsmodelle als erklärende Variable (Prädiktor bzw. unabhängige Variable) eingeht. Die Schülerinnen und Schüler in den ALLE-Schulen erhalten den Wert 1 auf dieser Variable, die Schülerinnen und Schüler in den Vergleichsschulen den Wert 0. Der Regressionskoeffizient für diese Variable gibt entsprechend an, um wie viele Einheiten sich der vorhergesagte Wert auf der zu erklärenden Zielvariable (abhängige Variable) ändert, wenn es sich um eine Schülerin oder einen Schüler aus einer ALLE-Schule – anstelle einer Schülerin oder eines Schülers aus einer Vergleichsschule – handelt. Positive Koeffizienten stehen dafür, dass die ALLE-Schülerinnen und Schüler eine höhere Ausprägung auf der Zielvariable aufweisen. Negative Koeffizienten stehen für eine höhere Ausprägung in den Vergleichsschulen. Ein Beispiel sind die Regressionsmodelle, in denen untersucht wird, ob sich die ALLE- und Kontrollschulen darin unterscheiden, wie viel die Schülerinnen und Schüler im Evaluationszeitraum dazugelernt haben. In diesen Modellen dient die Veränderung in der Testleistung zwischen dem ersten und letzten Messzeitpunkt der Evaluation als die zu erklärende Zielvariable. Um die Unterschiede zwischen den ALLE- und Vergleichsschulen zu bestimmen, wird die oben genannte Dummy-Variable (ALLE-Schulen = 1, Vergleichsschulen = 0) als erklärende Variable (Prädiktor) verwendet. Resultiert für die Dummy-Variable ein positiver Regressionskoeffizient, zeigt dies einen höheren Lernzuwachs in den ALLE-Schulen an. Resultiert ein negativer Regressionskoeffizient, weist dies auf einen höheren Lernzuwachs in den Vergleichsschulen hin.

Wie bereits an diesem Beispiel veranschaulicht, dient in den meisten berichteten Regressionsmodellen die *Veränderung* beziehungsweise der *Zuwachs* auf der interessierenden Variable (z.B. der Testleistung im Fach Deutsch) als die zu erklärende Zielvariable. Die Veränderung wird als Differenz zwischen der Variable zum letzten Messzeitpunkt der Evaluation (Ende der 9. Klassenstufe; T4) und dem ersten Messzeitpunkt der Evaluation (Anfang der 7. Klassenstufe; T1) berechnet. Durch die Verwendung der Veränderung als Zielvariable werden individuelle Unterschiede der Schülerinnen und Schüler zum ersten Messzeitpunkt berücksichtigt. Bezogen auf den Vergleich zwischen den ALLE- und Vergleichsschulen wird in Rechnung gestellt, dass die Schülerinnen und Schüler der beiden Gruppen sich bereits zu Beginn der Evaluation unterscheiden haben können. Damit wird vermieden, dass allfällige Unterschiede zwischen den ALLE- und Vergleichsschulen, die am Ende des Evaluationszeitraums festgestellt werden, irrtümlich auf die ALLE-Massnahmen

zurückgeführt werden, obwohl sie tatsächlich unterschiedlichen Ausgangsvoraussetzungen der Schülerinnen und Schüler in beiden Gruppen geschuldet sind.

In verschiedenen berichteten Modellen werden neben der bereits beschriebenen Dummy-Variable (ALLE-Schulen = 1, Vergleichsschulen = 0) auch leistungsrelevante soziodemografische Hintergrundmerkmale der Schülerinnen und Schüler als weitere erklärende Variablen («Kontrollvariablen») berücksichtigt (vgl. Kap. 3.5.1). Damit wird der Tatsache Rechnung getragen, dass sich die Schülerschaft der ALLE- und Vergleichsschulen in ihrer Zusammensetzung bezüglich dieser Schülermerkmale unterscheiden könnte. Der Einbezug dieser Merkmale in die Regressionsmodelle verändert die Interpretation des Koeffizienten für den Unterschied zwischen den ALLE- und Vergleichsschulen. Dieser bildet nun nicht mehr den vorhergesagten «Brutto-Unterschied» zwischen den beiden Gruppen ab, sondern den «Netto-Unterschied», wenn man die Hintergrundmerkmale statistisch kontrolliert beziehungsweise konstant hält – das heisst, wenn alle Schülerinnen und Schüler gleich ausgeprägte Hintergrundmerkmale hätten. Auf diesem Wege lässt sich bestimmen, ob sich die ALLE- und Vergleichsschulen auch dann unterscheiden, wenn der Einfluss der Hintergrundmerkmale auf die Zielvariable statistisch berücksichtigt wird.

Wie werden Einschränkungen durch das Ausscheiden von Schulen beurteilt?

Zwei Vergleichsschulen schieden im Lauf der Evaluation vorzeitig aus (vgl. Kap. 3.3.2.1). Eine weitere Vergleichsschule sagte zu Beginn der Evaluation ihre geplante Teilnahme ab; für diese wurde eine zusätzliche Schule gewonnen, die an der Evaluation ab dem zweiten Messzeitpunkt teilnahm. Demnach fehlten für drei Schulen für Messzeitpunkte am Ende beziehungsweise Anfang der Evaluation sämtliche Daten. Auf Basis der umfangreich vorhandenen Daten von den anderen Messzeitpunkten wurden für diese Schulen mit multipler Imputation die fehlenden Werte auf Schülerebene durch geschätzte Werte ersetzt (vgl. Kap. 3.5.3). Entsprechend werden bei der Verwendung von Variablen auf der Schülerebene die Ergebnisse von Analysen berichtet, bei denen für die genannten Vergleichsschulen die Daten einzelner Messzeitpunkte vollständig imputiert wurden.

Das Verfahren der multiplen Imputation gilt auch bei grösseren Anteilen fehlender Daten als relativ zuverlässig (Enders, 2017). Nichtsdestotrotz stellt die Imputation einzelner Messzeitpunkte für ganze Schulen eine erhebliche Herausforderung dar, da nur Daten anderer Messzeitpunkte Informationen über die fehlenden Werte liefern können. Um die Belastbarkeit der Ergebnisse zu prüfen, werden daher alle Berechnungen mit imputierten Daten zusätzlich auch unter Ausschluss der frühzeitig ausgeschiedenen beziehungsweise später einbezogenen Vergleichsschulen durchgeführt. Es wird insbesondere geprüft, ob sich statistisch signifikante Unterschiede auch in dieser reduzierten Stichprobe nachweisen lassen. Ergibt sich dabei ein gegenteiliger Befund, wird dies im Sinne einer geringeren Belastbarkeit des Ergebnisses interpretiert und berichtet. Dies betrifft die Analysen zur Unterrichtsqualität aus Sicht der Schülerinnen und Schüler (Kap. 4.9.2), zu den motivational-affektiven Lernergebnissen (Kap. 4.10.1) und zu den fachlichen Lernergebnissen (Kap. 4.10.2).

4 Ergebnisse

In diesem Kapitel werden die Ergebnisse der Evaluation ALLE nach den Fragestellungen (vgl. Kap. 2) geordnet dargestellt. Kapitel 4.1 befasst sich mit den Rahmenbedingungen des Pilotprojekts, die Kapitel 4.2 bis 4.7 mit den im Pilotprojekt umgesetzten Massnahmen und die Kapitel 4.8 bis 4.10 mit den Wirkungen des Pilotprojekts auf die professionellen Kompetenzen der Lehrpersonen, die Qualität des Unterrichts und die Lernergebnisse der Schülerinnen und Schüler. Kapitel 4.11 geht auf die Ergebnisse einer zusätzlichen explorativen Analyse ein, welche die Identifikation von Gelingensbedingungen des Pilotprojekts in den ALLE-Schulen zum Ziel hatte. Die verschiedenen Stichproben, die den einzelnen Kapiteln zugrunde liegen, werden in Kapitel 3.3.2 erläutert; die Datengrundlage, die der angesprochenen explorativen Analyse zugrunde liegt, wird zudem in Kapitel 4.11 thematisiert. Zusätzliche Kennwerte zu den Fragebogenangaben, welche in die nachfolgend berichteten Analysen eingingen, können dem Anhang (siehe Kap. 7) entnommen werden.

4.1 Rahmenbedingungen des Pilotprojekts ALLE

Das Pilotprojekt ALLE umfasste die Implementation eines breiten Spektrums an potenziell lernförderlichen Massnahmen. Es ergibt sich die Frage, ob dies die Schulen vor grössere Herausforderungen stellte, die finanziellen, personellen und organisatorischen Voraussetzungen für eine gelingende Umsetzung der Massnahmen zu schaffen. In den Fragebögen sowie den begleitenden Interviews wurden die beteiligten Personengruppen zu den Bedingungen befragt, unter denen das Pilotprojekt durchgeführt wurde. Die Auskünfte sind wichtig für ein Verständnis möglicher Ursachen von erwartungsgemässen und weniger erwartungsgemässen Ergebnissen des Pilotprojekts, und können so auch für die zukünftige Umsetzung vergleichbarer Vorhaben nützlich sein.

Bei der Analyse der Rahmenbedingungen werden die folgenden Aspekte in den Blick genommen:

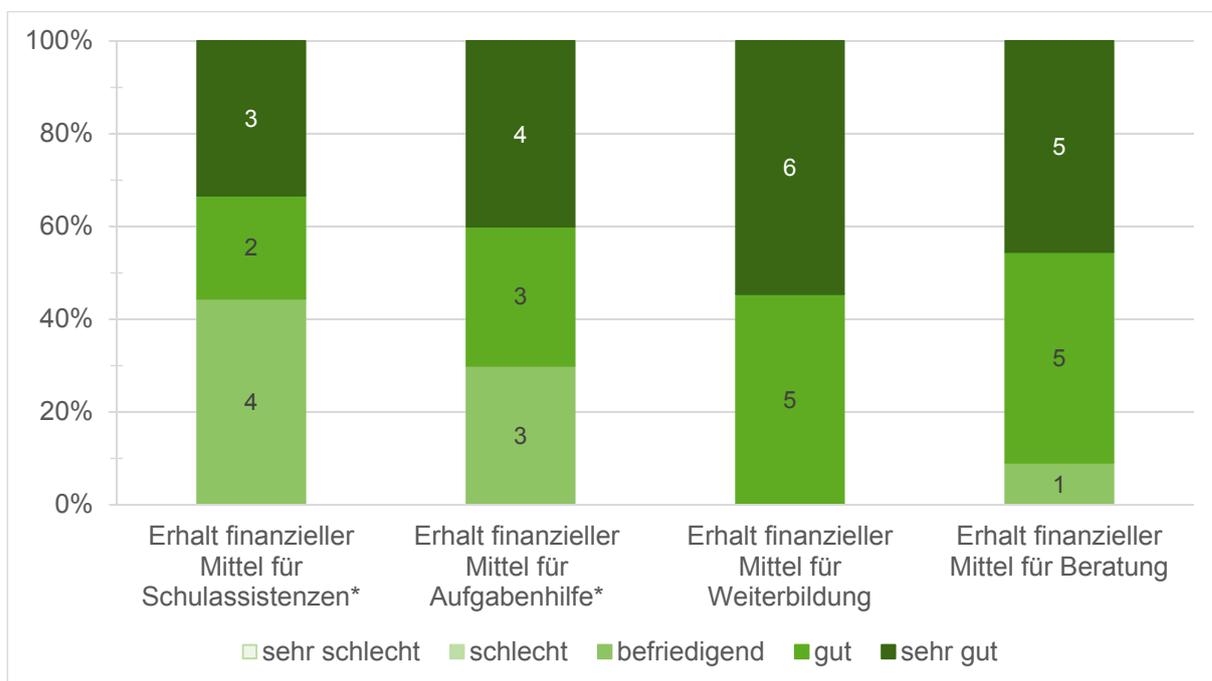
- die finanziellen Voraussetzungen zur Umsetzung der Massnahmen (Kap. 4.1.1),
- die Qualität der innerschulischen Kooperation und des kollegialen Austausches (Kap. 4.1.2).

4.1.1 Finanzielle Voraussetzungen zur Umsetzung der Massnahmen

Wie Abbildung 2 zeigt, berichten die Schulleitungen von überwiegend günstigen finanziellen Rahmenbedingungen. Die Bereitstellung finanzieller Mittel im Pilotprojekt war demzufolge weitgehend unproblematisch. Aus den zusätzlich erfragten Anmerkungen der Schulleitungen ergeben sich Hinweise auf vereinzelte Einschränkungen, beispielsweise:

- «Durch die Festlegung der Budgets, zum Beispiel Ende Schuljahr 2018/19 für das Kalenderjahr 2020, sind teilweise Ressourcen gebunden und können im Nachhinein, wenn der Bedarf da ist, nicht mehr so einfach gesprochen werden.»
- «Wir arbeiten mit Zivis als Assistenten und in der Aufgabenbetreuung. Das ist einerseits eine äusserst kostengünstige Sache, andererseits ist es anspruchsvoll, geeignete Zivis zu finden.»

Abbildung 2. Angaben der Schulleitungen zum Erhalt finanzieller Mittel als Rahmenbedingung für die Umsetzung des ALLE-Massnahmenbündels



Anmerkungen. N = 11; *Items von links nach rechts: N = 9, fehlend: N = 2; N = 10, fehlend: N = 1.

Weitere Hinweise zu den finanziellen Voraussetzungen der Umsetzung von ALLE enthalten die Einzelinterviews mit den Schulleitungen. So bewertet eine Schulleitung die Ressourcenverfügbarkeit allgemein und auch die finanziellen Ressourcen als positiv:

«Viele Ressourcen standen uns zur Verfügung. Wir haben uns an Sitzungen getroffen, an welchen der Austausch wirklich gut stattgefunden hat. Auch materiell, die Entschädigung usw. für diesen zusätzlichen Aufwand vom Kanton, diese Wertschätzung war da. Ich glaube, es wurde auch gut informiert und begleitet. Hierzu kann ich eigentlich nur gute Attribute mitteilen.»

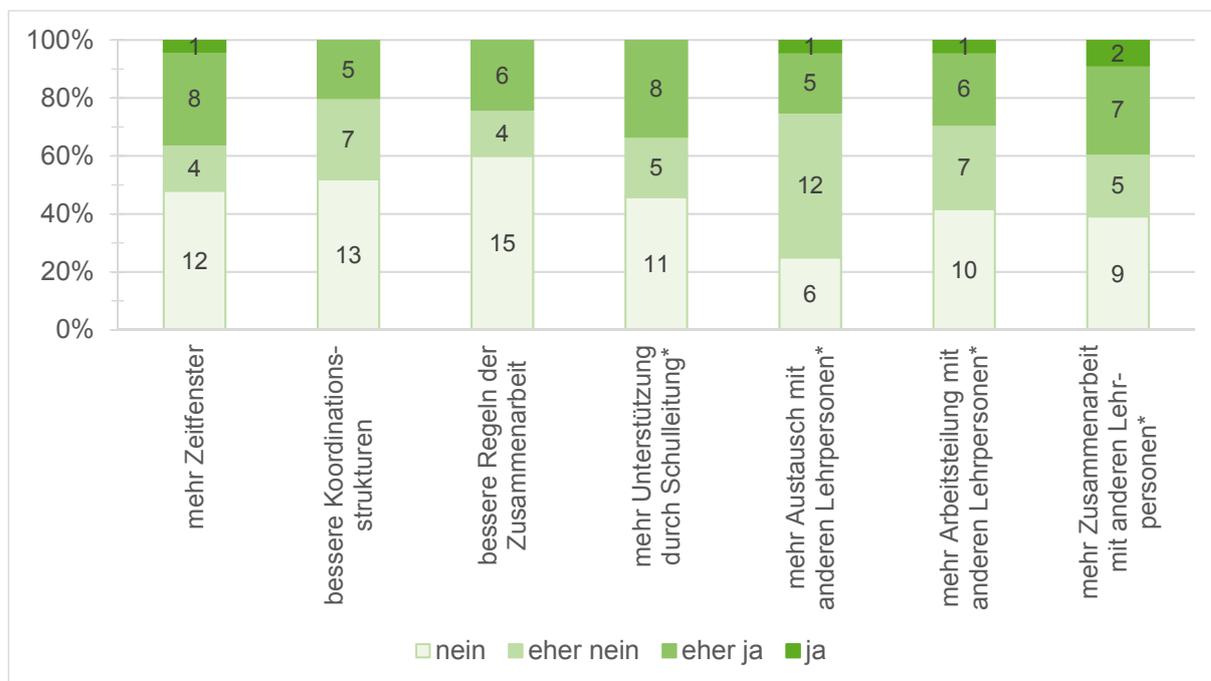
Eine zweite Schulleitung äussert hingegen, dass das Pilotprojekt mit weiteren Ressourcen noch gewinnbringender hätte sein können:

«Ich fand, es war ein gutes Projekt. Es sollte unbedingt weitergeführt werden. Was ich schade finde, dass es keine zusätzlichen Ressourcen dafür gibt, weil da könnte man noch eine ganze Menge mehr herausholen.»

4.1.2 Qualität der innerschulischen Kooperation

In den Befragungen und Interviews mit den Lehrpersonen wurde erfragt, ob sich die innerschulische Kooperation zwischen den Lehrpersonen im Zuge des Pilotprojekts verbessert hat. Eine solche Verbesserung kann einerseits als *Wirkung* von ALLE auf die Zusammenarbeit innerhalb der ALLE-Schulen angesehen werden. Angesichts der Herausforderungen von ALLE an die gemeinsame kollegiale Arbeit können entsprechende Entwicklungen aber auch als *Gelingsbedingung* für den Erfolg der ALLE-Massnahmen angesehen werden. Abbildung 3 und Abbildung 4 zeigen die Zustimmung der Deutsch- und Mathematik-Lehrpersonen zu sieben Aspekten möglicher Verbesserungen in der innerschulischen Kooperation.

Abbildung 3. Verbesserung der Lehrerinnen- und Lehrerkoooperation nach Angaben der Deutsch-Lehrpersonen



Anmerkungen. N = 25, fehlend: N = 1; *Items von links nach rechts: N = 24, fehlend: N = 2; N = 24, fehlend: N = 2; N = 24, fehlend: N = 2; N = 23, fehlend: N = 3.

Abbildung 4. Verbesserung der Lehrerinnen- und Lehrerk Kooperation nach Angaben der Mathematik-Lehrpersonen



Anmerkungen. N = 32.

In beiden Fächern stimmen die Lehrpersonen einer Verbesserung der Lehrerinnen- und Lehrerk Kooperation überwiegend (eher) nicht zu. Noch am häufigsten wird geäußert, dass das Pilotprojekt zu einem vermehrten Austausch mit anderen Lehrpersonen geführt habe, dies beschränkt sich jedoch auf das Fach Mathematik (*eher ja* oder *ja*: Mathematik 56.3 Prozent). Den weiteren Aussagen stimmen jeweils rund 20 bis 40 Prozent der Lehrpersonen (eher) zu, wobei die Anteile in beiden Fächern vergleichbar sind.

Dies bedeutet allerdings nicht, dass die einzelnen Lehrpersonen durchweg keine Verbesserungen wahrnehmen. Individuell kann sich dieser Eindruck auch nur auf einzelne der erfragten Aspekte beschränken. Wertet man die Antworten pro Lehrperson aus, so zeigt sich, dass nur etwa ein Drittel der Deutsch-Lehrpersonen (36.0 Prozent, N = 9 [von 25]) und ein Fünftel der Mathematik-Lehrpersonen (18.8 Prozent, N = 6 [von 32]) über alle erfragten Aspekte hinweg angeben, dass keine Verbesserung der Lehrerinnen- und Lehrerk Kooperation stattgefunden habe⁷.

Eine Betrachtung der Lehrpersonen-Äusserungen für die einzelnen Schulen ergibt weiterhin, dass sich die Auffassungen innerhalb der Kollegien durchaus unterscheiden. So findet sich keine

⁷ Vier Deutsch- und vier Mathematik-Lehrpersonen mit Antwort «nein» auf alle Fragen; fünf Deutsch- und zwei Mathematik-Lehrpersonen mit gemischten Antworten «nein» oder «eher nein».

Schule, in der alle beteiligten Lehrpersonen durchgehend keine Verbesserungen der Kooperation wahrnehmen. Die neun Deutsch-Lehrpersonen, die bei allen erfragten Kooperationsaspekten keine Verbesserungen berichten, verteilen sich auf acht der neun ALLE-Schulen, die sechs Mathematik-Lehrpersonen mit derselben Wahrnehmung auf insgesamt fünf Schulen.

Die Äusserungen der Lehrpersonen in den Gruppeninterviews legen zudem nahe, dass geringe wahrgenommene Verbesserungen der Lehrerinnen- und Lehrerkooperation unterschiedliche Ursachen haben können. So berichtet eine Lehrperson zur Zusammenarbeit mit den anderen Lehrpersonen im Kollegium, dass sich keine Veränderungen ergeben hätten, da eine intensive Zusammenarbeit bereits über einen längeren Zeitraum bestanden habe:

[Interviewende/r]: «Hat sich etwas in der Zusammenarbeit bei euch im Kollegium verändert?»

[Lehrperson:] «Nein. Also wir arbeiten sowieso zusammen. Also man muss zusammenarbeiten, das ist klar oder es macht wenig Sinn, wenn man nicht zusammenarbeitet. Aber wir arbeiten so oder so zusammen. Das ist auch nicht das erste Projekt und auch nicht das zweite, [das] wir in den letzten sechs bis sieben Jahren zusammen gemacht haben. Das gehört bei uns dazu.»

Eine Deutsch-Lehrperson berichtet von einer besonders starken Zusammenarbeit im Fachteam, die allerdings nach Abschluss des Pilotprojekts wieder nachgelassen habe:

«Sicher bei uns im Deutsch, oder, im Fachteam: dass wir sehr intensiv zusammengearbeitet haben. Ich finde, das war eine gute Arbeit, die wir dort gemacht haben. Ein guter Austausch, welchen wir jetzt ein bisschen vernachlässigen. Ich denke, in dieser intensiven Phase war es eine sehr, sehr gute und intensive Zusammenarbeit. Das finde [ich], das hat etwas gebracht.»

Aussagen der Lehrpersonen in Richtung einer ausbleibenden Verbesserung in der Kooperation sind demnach nicht pauschal als ungünstig zu bewerten, sondern bedürfen gegebenenfalls einer weiterführenden Betrachtung mit Fokus auf die einzelne Schule beziehungsweise die einzelne Lehrperson.

4.2 Einsatz der Schulassistenten

Eine der im Pilotprojekt ALLE angebotenen Massnahmen sah den Einsatz von Schulassistenten vor. Durch den Einsatz von Schulassistenten sollen die Lehrpersonen sowohl im Unterricht als auch bei unterrichtsergänzenden Aufgaben entlastet werden. Die zweckgemässe Aufteilung der Aufgaben soll es ermöglichen, dass sowohl Lehrpersonen als auch Schulassistenten vermehrt auf die individuellen Bedürfnisse der Lernenden eingehen können. Nachfolgend werden unterschiedliche Aspekte des Einsatzes der Schulassistenten im Rahmen des Pilotprojekts in den Blick genommen. Konkret wird dabei auf die folgenden Fragestellungen eingegangen:

- Unter welchen Voraussetzungen erfolgte die Rekrutierung der Schulassistenzen? Wie hoch ist die personale Konstanz über die Projektzeit hinweg? Wie wird der Einsatz der Schulassistenzen von der Schulleitung unterstützt? (Kap. 4.2.1)
- Über welche Qualifikationen verfügen die Schulassistenzen? (Kap. 4.2.2)
- Wie und in welchem Umfang werden die Schulassistenzen innerhalb und ausserhalb des Unterrichts in Deutsch und Mathematik eingesetzt? (Kap. 4.2.3)
- Wie ist die Qualität der Zusammenarbeit der Schulassistenzen mit den Lehrpersonen? (Kap. 4.2.4)
- Wie nehmen die Lehrpersonen und die Schülerinnen und Schüler die Unterstützung durch die Schulassistenzen wahr? (Kap. 4.2.5)
- Welche Veränderungen des Unterrichts haben sich durch den Beizug der Schulassistenzen ergeben? (Kap. 4.2.6)
- Wie zufrieden sind die Lehrpersonen und die Schulleitungen mit der Umsetzung der Massnahme Schulassistenten an ihrer Schule? (Kap. 4.2.7)

4.2.1 Rekrutierung, personale Konstanz und Unterstützung durch die Schulleitung

Rekrutierung der Schulassistenzen

Die Bedingungen für die Rekrutierung von qualifizierten Schulassistenzen werden lediglich von einer der zehn befragten Schulleitungen als «schlecht» bewertet (10.0 Prozent). Der überwiegende Anteil der Schulleitungen hingegen sieht den Erfolg bei der Rekrutierung der Schulassistenzen als «gut» bis «sehr gut» (jeweils 40 Prozent bzw. $N = 4$), eine einzelne Schulleitung (10.0 Prozent) als «befriedigend» an. Die Gewinnung geeigneter Schulassistenzen ist somit aus Sicht der Schulleitungen überwiegend zufriedenstellend gelungen.

Personale Konstanz

Die personale Konstanz beim Einsatz der Schulassistenzen ist relativ hoch. So geben zwölf der 14 befragten Schulassistenzen (85.7 Prozent) an, seit mehr als einem Jahr tätig zu sein. Je einmal ist eine Person seit neun beziehungsweise seit vier Monaten als Schulassistenten tätig (je 7.1 Prozent).

Im Einzelinterview beschreibt eine Schulassistentin den Nutzen eines solchen längeren Engagements folgendermassen:

«Das finde ich etwas sehr Wertvolles, dass ich ganze Jahrgänge begleiten kann. Es gibt Schulen, die haben Klassenassistenten einfach im Rahmen von Zivildienst. Das finde ich immer schade. Es braucht immer gerade etwa ein halbes Jahr bis ein Jahr, bis eine Beziehung aufgebaut ist. Wenn dann diese Person wieder geht, dann ... Die Beziehungsarbeit finde ich schon sehr wertvoll in dieser Aufgabe.»

Unterstützung durch die Schulleitung

Die Schulleitungen wurden befragt, inwieweit sie vier verschiedene Möglichkeiten nutzen, um den Einsatz von Schulassistenten an ihrer Schule zu ermöglichen. Die Ergebnisse zeigt Tabelle 14.

Tabelle 14. Nutzung von Möglichkeiten zur Unterstützung des Einsatzes von Schulassistenten durch die Schulleitung nach eigenen Angaben

Unterstützungsmöglichkeiten	Anzahl	in Prozent
Einstellung von qualifizierten Schulassistenten	9	90.0
Schnellstmögliche Suche einer Vertretung bei Ausfall einer Schulassistentz	4	40.0
Erkundigung bei den Lehrpersonen, wie nutzbringend sie die Schulassistenten einsetzen können	9	90.0
Organisation des Austausches der Lehrpersonen untereinander zum erfolgreichen Einsatz der Schulassistenten	7	70.0

Anmerkungen. $N = 10$, fehlend: $N = 1$. Antwortformat pro Item: ja – nein.

Die erfragten Unterstützungsmöglichkeiten werden von einem Grossteil beziehungsweise fast allen Schulleitungen genutzt. Eine Ausnahme ist die schnellstmögliche Suche nach einer Vertretung bei Ausfall einer Schulassistentz, von der nur rund die Hälfte der Schulleitungen berichtet.

4.2.2 Qualifikationen der Schulassistenten

Ausbildungsabschlüsse der Schulassistenten

Nach ihren Ausbildungsabschlüssen gefragt, geben alle 14 befragten Schulassistenten an, die obligatorische Schulzeit abgeschlossen zu haben (vgl. Tabelle 15). Bei den weiteren Abschlüssen werden am häufigsten eine Berufsausbildung, gefolgt von der Matura sowie anderen Abschlüssen genannt. Jede der Schulassistenten gibt dabei zusätzlich zur obligatorischen Schulzeit mindestens einen konkreten (d.h. nicht unter «anderer Abschluss» angeführten) Ausbildungsabschluss an.

Für den jeweils höchsten angegebenen Ausbildungsabschluss ergibt sich die folgende Verteilung:

- Berufsausbildung: $N = 7$ (50.0 Prozent),
- Fachmittelschule: $N = 2$ (14.3 Prozent),
- Matura: $N = 1$ (7.1 Prozent),
- Pädagogische Hochschule: $N = 2$ (14.3 Prozent),
- Nicht-pädagogische Hochschule: $N = 2$ (14.3 Prozent).

Tabelle 15. Ausbildungsabschlüsse der Schulassistenzen nach eigenen Angaben (mehrfache Nennungen möglich)

Ausbildungsabschluss	Anzahl	in Prozent
Obligatorische Schulzeit	14	100.0
Berufsausbildung	10	71.4
Fachmittelschule	2	14.3
Matura	5	35.7
Pädagogische Hochschule	2	14.3
Nicht-pädagogische Hochschule	2	14.3
Anderer Abschluss	5	35.7

Anmerkungen. $N = 14$.

Die Frage nach den vor Beginn der Schulassistenten ausgeübten Tätigkeiten weist auf eine hohe Bandbreite bei den Vorerfahrungen hin. Die folgende Aufzählung enthält die 13 einzelnen Angaben der Schulassistenten im Wortlaut (Angabe einer Schulassistenten fehlend):

- Arbeitsagoge
- Betriebsassistentin PTT
- Dentalassistentin, Bankangestellte CS, Altenpflegerin SRK, Polizistin Stadtpolizei Zürich, Schulpflege, Assistenz Kindergarten
- Detailhandel
- kaufm. Angestellte, Spielgruppenleiterin, Aufgabenhilfe
- Personalassistentin/Kauffrau
- Sekretärin
- Réceptionassistentin
- Polygrafin, Berufsbildnerin
- Schreinerlehre EFZ und 3 Jahre als Schreiner (ausgelernt) gearbeitet
- Monteur / Student / Barkeeper usw.
- Hausfrau und Mittagstisch-Betreuung für OS und PS
- Student

Im Einzelinterview äussert sich eine Schulassistentin zu möglichen Vorteilen verschiedener Ausbildungsabschlüsse folgendermassen:

«Auch mit deinem KV-Hintergrund [Anm.: kaufmännische Grundbildung] kannst Du den Schülerinnen und Schülern wirklich helfen. Die Lehrpersonen waren ja meistens nie in einem anderen Berufsleben gestanden, sage ich jetzt mal, ohne es zu werten.»

Eine weitere Schulassistentin beschreibt die Möglichkeiten, ihre beruflichen Vorerfahrungen nutzbringend in der Schule einzusetzen, wie folgt:

«Ursprünglich bin ich gelernte Polygrafin und habe Lehrlinge ausgebildet. Ich persönlich werde jetzt auch oft in der Berufswahl eingesetzt, dass ich mit Schülern Bewerbungsgespräche übe, ihre Bewerbungsunterlagen durchgehe mit ihnen [...]. Zum Beispiel für so etwas werde ich im Deutschunterricht auch eingesetzt oder ganz spezifisch für eine Periode des Jahres. Wenn sie finden, ok, das wäre jetzt schön, wenn ich mit jedem Schüler ein Telefongespräch üben kann oder ein Bewerbungsgespräch.»

Teilnahme an Weiterbildungen

Für die Schulassistenten wurde im Rahmen des Pilotprojekts eine freiwillige Weiterbildung mit dem Titel «Gelingende Kooperation Lehrperson – Schulassistentin» angeboten, welche zusammen mit dem Schulteam inklusive der Schulleitung durchgeführt wurde. Ein Besuch von anderen Weiterbildungen konnte aber auch bereits vor dem Pilotprojekt oder während der Projektlaufzeit stattgefunden haben (siehe unten). Es lag daher im Ermessen der Schulleitung, ob die thematisch spezifische ALLE-Weiterbildung – mit Fokus auf gute Kommunikation und Abstimmung der Aufgabenteilung sowie den Umgang mit allfälligen Konflikten zwischen Schulassistentin und Lehrperson – wahrgenommen werden sollte oder nicht. Für die Teilnahme an der ALLE-Weiterbildung beziehungsweise anderen Weiterbildungen ergibt sich folgende Verteilung:

- nur die ALLE-Weiterbildung: $N = 1$ (50.0 Prozent),
- ALLE-Weiterbildung + andere Weiterbildung(en): $N = 3$ (21.4 Prozent),
- nur andere Weiterbildung(en): $N = 5$ (35.7 Prozent),
- keine Weiterbildung: $N = 5$ (35.7 Prozent).

Die von acht Schulassistenten besuchten zusätzlichen Weiterbildungen fanden mehrheitlich an anderen, nicht näher benannten Institutionen ($N = 5$), bei privaten Anbietern ($N = 2$) oder an der Pädagogischen Hochschule Zürich ($N = 1$) statt. Dabei handelte es sich um die folgenden Kurse (jeder Aufzählungspunkt entspricht der Einzelaussage einer Schulassistentin im Wortlaut):

- AAI IP- Erziehung
- Grundkurs für Klassenassistentinnen AAI IP-Erziehung
- Interne Schulung durch Alfred Adler Institut
- Klassenassistentin und Klassenassistenten füllen ihren Ideenkoffer/ZAL, 1x1 Heilpädagogik für Klassenassistenten/HfH, und noch versch. Kurse
- Onlinekurs: 1x1 der Heilpädagogik (HfH Zürich)
- Schulassistentin Schulsupport Zürich

- Schulsupport Firma Rudin ZH
- Zwei Kurswochen des Zivildienstes

Auf mögliche Missverständnisse in der Kommunikation, für wen die Weiterbildung im Pilotprojekt bestimmt war, deutet die folgende Begründung einer Schulassistentin für ihre Nichtteilnahme aus dem Einzelinterview hin:

«Nein, war ich nicht, und zwar weil es nicht für uns Assistentinnen ist. Es war nur für die Lehrpersonen.»

Eine weitere Assistentin führt fehlende zeitliche Ressourcen als Grund der Nichtteilnahme an:

«Nein, ich war nicht dabei, bin aber auch selber noch als Jobcoach in dieser Ausbildung und hätte keine Kapazitäten dafür.»

Von den vier Schulassistenten, welche die Weiterbildung im Pilotprojekt besucht haben, schätzen zwei diese als nützlich ein. Mögliche Verbesserungsvorschläge für den Weiterbildungskurs macht eine der Schulassistenten. Ihrer Ansicht nach hätte der Kurs *«situationsbezogener, wertschätzender, dem Bildungsstand der Schulassistentin angemessener»* gestaltet werden können.

4.2.3 Umfang und Art des Einsatzes

Umfang des Einsatzes der Schulassistenten

Zum Erhebungszeitpunkt (Ende 9. Klassenstufe) werden die befragten Schulassistenten durchschnittlich knapp neun Stunden pro Woche eingesetzt, wobei der Umfang beträchtlich variiert und zwischen einer und 30 Stunden liegt ($M = 8.8$, $SD = 8.7$; $N = 12$ [von 14]). Im Fach Deutsch sind elf (78.6 Prozent), im Fach Mathematik zehn (71.4 Prozent) der 14 Schulassistenten tätig. Die Hälfte der Schulassistenten berichtet von einem Einsatz in beiden Fächern ($N = 7$ [von 14]).

Im Hinblick auf einzelne auffällig hohe Einsatzzeiten von bis zu 30 Stunden pro Woche äussert sich eine Schulassistentin, die 28 Stunden pro Woche arbeitet, im Einzelinterview so:

«Dann kam ich in diese Klasse und bin drei Jahre mit dieser Klasse mitgegangen. Ich habe mit einer ISR-Schülerin [Anm.: Schülerin der Integrierten Sonderschulung] 28 Stunden in der Woche. Das heisst wirklich von Montag bis Freitag gehe ich mit dieser Schülerin. Das heisst, ich gehe mit der Klasse mit.»

Eine weitere Schulassistentin führt zu eher niedrigeren Einsatzzeiten die Abhängigkeit von Entscheidungen der Lehrperson an:

«Ich sage mal, es gibt Schüler, die einen Zivi oder Klassenassistenten noch mehr brauchen könnten und es noch nicht immer eingesetzt wird. Wie gesagt, das ist aber auch ein wenig lehrerabhängig.»

Auch aus der Befragung der Lehrpersonen geht hervor, dass der Umfang des Einsatzes von Schulassistenzen insgesamt recht unterschiedlich ausfällt. Neben der Anzahl der von den Lehrpersonen gegebenen Unterrichtslektionen variiert auch die Anzahl der Lektionen mit Schulassistenz deutlich (vgl. Tabelle 16 und Tabelle 17).

Tabelle 16. Anzahl der Lektionen mit Schulassistenz im Verhältnis zur Anzahl unterrichteter Lektionen pro Woche im Fach Deutsch nach Angaben der Lehrpersonen

Unterrichtslektionen pro Woche	Anzahl LP	Unterrichtslektionen mit Schulassistenz pro Woche				
		0	1	2	3	4
0	1	1				
2	2	2				
4	10	5	1 (25.0%)	1 (50.0%)		3 (100.0%)
5	3	1	2 (20.0%)			
6	1	1				
8	6	5	1 (12.5%)			
9	1	1				
10	1	1				
12	1					1 (33.3%)
Gesamt	26	17	4	1	0	4

Anmerkungen. LP = Lehrpersonen. $N = 26$. Angaben in offenem Antwortfeld. *Beispiel für die Interpretation:* Die Zeile mit der Eintragung «8» in der Spalte «Unterrichtslektionen pro Woche» berichtet, wie viele Lehrpersonen acht Lektionen pro Woche unterrichteten (siehe Spalte «Anzahl LP»: dies war bei sechs Lehrpersonen der Fall), und in wie vielen dieser acht Lektionen eine Schulassistenz zur Verfügung stand (siehe Spalten zu «Unterrichtslektionen mit Schulassistenz pro Woche»). Von den sechs Lehrpersonen mit acht Unterrichtslektionen pro Woche hatten fünf Lehrpersonen keine Schulassistenz (Eintragung «5» in Spalte «0»), eine Lehrperson hatte eine Assistenz für eine Lektion pro Woche (Eintragung «1» in Spalte «1»). *Prozentangabe in Klammern:* Anteil der Lektionen mit Schulassistenz an der Gesamtzahl unterrichteter Lektionen. Beispiel: Zehn Lehrpersonen geben an, vier Lektionen zu unterrichten. Von diesen gibt eine Lehrperson an, in zwei Lektionen eine Assistenz zu haben. Der Prozentanteil der Lektionen mit Assistenz liegt für diese Lehrperson entsprechend bei 50% (Eintrag: «1 (50.0%)» = eine Lehrperson, die in 50% ihrer Lektionen eine Assistenz hat).

Für die Deutsch-Lehrpersonen liegen die Angaben mit null bis zwölf unterrichteten Lektionen ($M = 5.5$, $SD = 2.8$) und null bis vier Lektionen mit Schulassistenz ($M = 0.8$, $SD = 1.5$) etwas unter denjenigen der Mathematik-Lehrpersonen (siehe unten). Von den 26 befragten Deutsch-Lehrper-

sonen berichten lediglich neun (34.6 Prozent), zum Zeitpunkt der Erhebung von einer Schulassistenz unterstützt zu werden. Ein Drittel der Lehrpersonen mit Schulassistenz wird von dieser in allen unterrichteten Stunden unterstützt. Den meisten der restlichen Lehrpersonen steht die Schulassistenz für einen eher kleinen Anteil ihrer Unterrichtszeit zur Verfügung.

Tabelle 17. Anzahl der Lektionen mit Schulassistenz im Verhältnis zur Anzahl unterrichteter Lektionen pro Woche im Fach Mathematik nach Angaben der Lehrpersonen

Unterrichts- lektionen pro Woche	Anzahl LP	Unterrichtslektionen mit Schulassistenz pro Woche						
		0	1	2	3	4	5	6
2	2	2						
3	1	1						
4	7	4		1 (50.0%)			2 (100%)	
6	10	4	1 (16.7%)			3 (66.7%)	1 (83.3%)	1 (100%)
8	1	1						
10	2	2						
11	1	1						
12	7	5		1 (16.7%)			1 (33.3%)	
18	1				1 (16.7%)			
Gesamt	32	20	1	2	1	6	1	1

Anmerkungen. LP = Lehrpersonen. $N = 32$. Angaben in offenem Antwortfeld. Für Erläuterungen zur Interpretation der Tabelleneinträge siehe die Anmerkungen zu Tabelle 16.

Für die Mathematik-Lehrpersonen liegt die Anzahl unterrichteter Lektionen pro Woche zwischen zwei und 18 ($M = 7.4$, $SD = 3.9$), die Anzahl der Lektionen pro Woche mit Schulassistenz zwischen null und sechs ($M = 1.3$, $SD = 1.9$), und damit etwas höher als bei den Deutsch-Lehrpersonen. Von den 32 Mathematik-Lehrpersonen gibt eine Minderheit von zwölf Befragten (37.5 Prozent) an, in ihren Lektionen über eine Schulassistenz zu verfügen. Einem Viertel der Mathematik-Lehrpersonen mit Schulassistenz steht diese in allen unterrichteten Lektionen zur Verfügung. Etwas weniger als die Hälfte der Befragten (41.6 Prozent, $N = 5$) wird zumindest in jeder zweiten Lektion (50.0 Prozent bis 83.3 Prozent der Lektionen) von einer Schulassistenz unterstützt.

Auffällig ist, dass damit in beiden Fächern weniger als die Hälfte der befragten Lehrpersonen berichtet, Lektionen mit einer Schulassistenz zu haben. Betrachtet man diese Angaben schulweise,

so gibt es zwei Schulen, für die alle befragten Lehrpersonen verneinen, Lektionen mit einer Schulassistentin zu haben (vgl. Tabelle 18, Schulen A und B). Auf der anderen Seite findet sich eine Schule, in der alle Lehrpersonen über Lektionen mit Schulassistentin berichten (Schule I).

Tabelle 18. Lehrpersonen pro Schule (fächerübergreifend) mit Angabe «0 Lektionen mit Schulassistentin», geordnet nach Prozentanteil ohne Schulassistentin

Schule	Anzahl Lehrpersonen	Lehrpersonen mit Angabe «0 Lektionen mit Schulassistentin»	
		Anzahl	in Prozent
A	6	6	100.0
B	2	2	100.0
C	6	5	83.3
D	6	5	83.3
E	13	9	69.2
F	6	4	66.7
G	5	3	60.0
H	3	1	33.3
I	5	0	0.0
Gesamt*	52	35	66.2

Anmerkungen. $N = 52$, fehlend: $N = 2$. *Spalte «in Prozent»: Durchschnittswert der Prozentanteile je Schule. Sechs Lehrpersonen unterrichten beide Fächer, ihre fachspezifischen Angaben wurden für die Auswertung zusammengefasst.

Im Folgenden beziehen sich die Aussagen von Lehrpersonen zum Einsatz der Schulassistentin jeweils auf die Stichproben derjenigen Lehrpersonen in den Fächern Deutsch und Mathematik, die angeben, *Lektionen mit Schulassistentin* zu unterrichten.

Einsatz der Schulassistenten für den Unterricht

Bezogen auf den Deutsch- beziehungsweise Mathematikunterricht wurden die Schulassistenten nach den prozentualen Anteilen ihres Einsatzes in bestimmten Tätigkeitsbereichen gefragt.

Tabelle 19 zeigt deskriptive Ergebnisse zu den Prozentangaben für das Fach Deutsch, Tabelle 20 zeigt die Ergebnisse für das Fach Mathematik. Die Standardabweichungen und Spannweiten (Min-Max) weisen für beide Fächer darauf hin, dass der Einsatz der Schulassistenten recht unterschiedlich gehandhabt wurde. Für jeden der Tätigkeitsbereiche werden Anteile zwischen 0 und 100 Prozent angegeben, das heisst die Tätigkeiten wurden von einigen Schulassistenten überhaupt nicht, von anderen hauptsächlich oder ausschliesslich ausgeübt. Die im Mittel höchsten Prozentanteile

werden dabei in beiden Fächern für das Unterrichten im Fach sowie für andere Tätigkeiten berichtet. Die Vor- und Nachbereitung von Unterricht spielt dagegen eine deutlich untergeordnete Rolle.

Tabelle 19. Prozentuale Anteile des Einsatzes der Schulassistenzen in Tätigkeitsbereichen im Fach Deutsch nach eigenen Angaben

Tätigkeitsbereich	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>
Im Fach Deutsch unterrichten	42.6	38.3	0	100
Vorbereitung des Deutschunterrichts	8.2	12.1	0	33
Nachbereitung des Deutschunterrichts	7.3	10.1	0	33
Andere Tätigkeiten	42.0	46.9	0	100

Anmerkungen. $N = 10$, fehlend: $N = 1$.

Tabelle 20. Prozentuale Anteile der Schulassistenzen in Tätigkeitsbereichen im Fach Mathematik nach eigenen Angaben

Tätigkeitsbereich	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>
Im Fach Mathematik unterrichten	30.3	35.4	0	80
Vorbereitung des Mathematikunterrichts	1.7	3.3	0	10
Nachbereitung des Mathematikunterrichts	3.9	6.6	0	20
Andere Tätigkeiten	64.1	38.7	0	100

Anmerkungen. $N = 9$, fehlend: $N = 1$.

Vier der Schulassistenzen machen zudem Angaben zu anderen Tätigkeiten im Rahmen des Deutschunterrichts (Wortlaut im offenen Antwortfeld):

- Andere Fächer
- Berufswahl
- Klassenassistenz im Deutschunterricht vorwiegend für ISR Schüler [Anm.: Schülerinnen und Schüler der Integrierten Sonderschulung]
- Unterstützung und Hilfestellung bei Bedarf, Begleitung von ISR Schülern

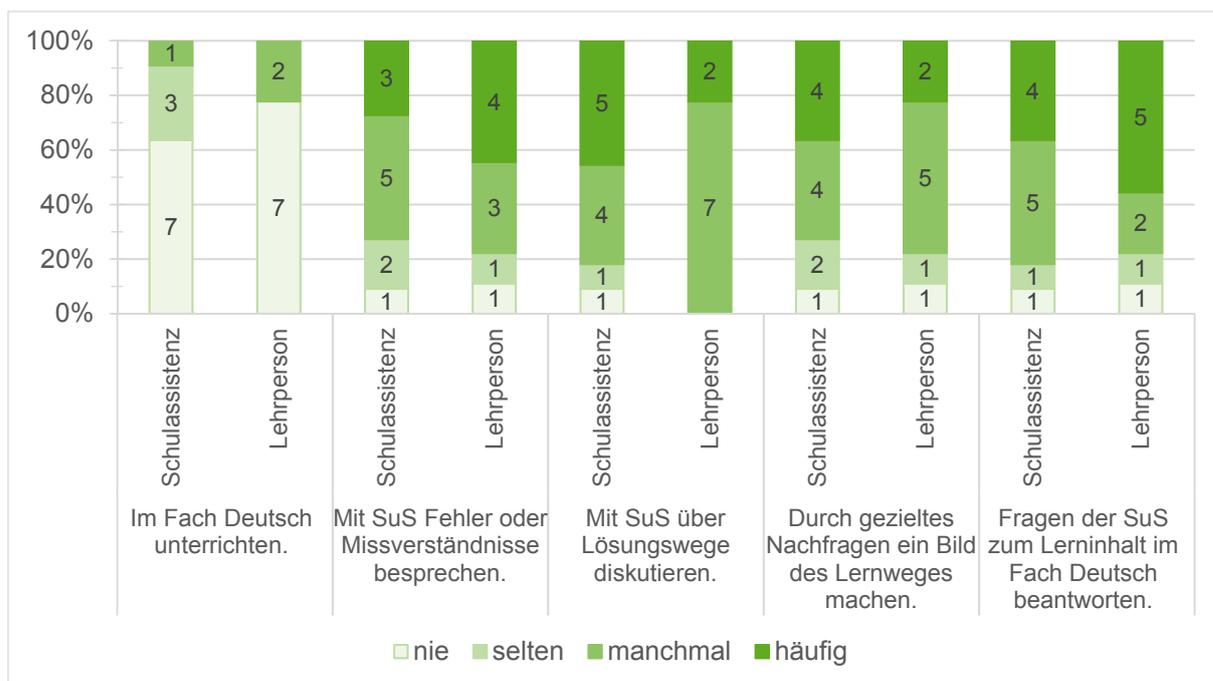
Sechs der Schulassistenzen berichten andere Tätigkeiten im Mathematikunterricht:

- Betreuung und Unterstützung der ISR-Schüler und Niveau 3 Schüler
- Hilfestellung geben im Unterricht
- Korrekturarbeiten, administrative Aufträge wie Dossiers kopieren, Prüfungen ausdrucken und Kopieren
- Korrekturen, Zusammenstellungen von Lernmaterialien

- Unterstützung
- Unterstützung als Klassenassistent in Deutsch und Mathematik während Unterrichtszeiten und im Office (selbständiges Arbeiten)

Ausser zu ihrer Arbeit in übergeordneten Tätigkeitsbereichen wurden die Schulassistenzen zu konkreten Tätigkeiten während des Unterrichts befragt. Auch die Lehrpersonen sollten angeben, wie häufig die Schulassistenzen die entsprechenden Tätigkeiten im Unterricht ausüben. In Abbildung 5 bis Abbildung 10 werden die Ergebnisse jeweils gemeinsam für die beiden Befragtengruppen dargestellt, zunächst für das Fach Deutsch, dann das Fach Mathematik. Unterschieden werden Tätigkeiten, die sich auf fachinhaltliche Aspekte beziehen, Tätigkeiten eher unterrichtsorganisatorischer Art sowie Tätigkeiten zur Unterstützung spezifischer Schülergruppen.

Abbildung 5. Fachinhaltliche Tätigkeiten der Schulassistenzen während des Unterrichts im Fach Deutsch nach eigenen Angaben und nach Angaben der Lehrpersonen

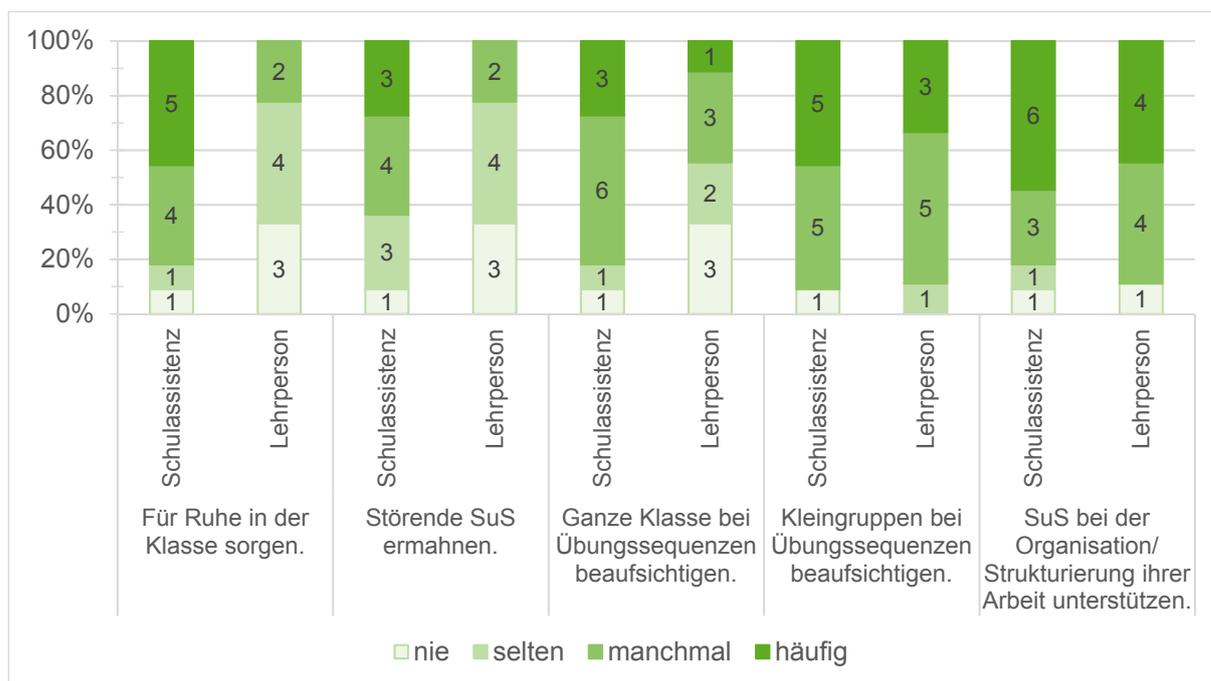


Anmerkungen. SuS = Schülerinnen und Schüler. Schulassistenzen: N = 11; Lehrpersonen: N = 9.

Bei den fachinhaltlichen Tätigkeiten im Deutschunterricht (vgl. Abbildung 5) geben die Schulassistenzen vergleichsweise am häufigsten an, mit den Schülerinnen und Schülern über Lösungswege zu diskutieren (*häufig*: 45.5 Prozent). Auch die weiteren Tätigkeiten stellen häufige oder zumindest gelegentliche Aufgaben der meisten Schulassistenzen dar. Eine Ausnahme ist das Unterrichten im

Fach, das gemäss der Mehrheit der Schulassistenzen nicht ausgeübt wird (*nie*: 63.6 Prozent).⁸ Die Lehrpersonen sehen die häufigste fachinhaltliche Tätigkeit, für die sie die Schulassistentin einsetzen, im Beantworten von Fragen zum Lerninhalt (*häufig*: 55.6 Prozent). Meist sind die Angaben der Lehrpersonen jenen der Schulassistenzen ähnlich. Nur beim Diskutieren der Lösungswege weicht ihr Urteil deutlicher ab, die Kategorie «häufig» wird seltener gewählt, die Kategorien «nie» und «selten» – anders als bei den Schulassistenzen – gar nicht.

Abbildung 6. Unterrichtsorganisatorische Tätigkeiten der Schulassistenzen während des Unterrichts im Fach Deutsch nach eigenen Angaben und nach Angaben der Lehrpersonen



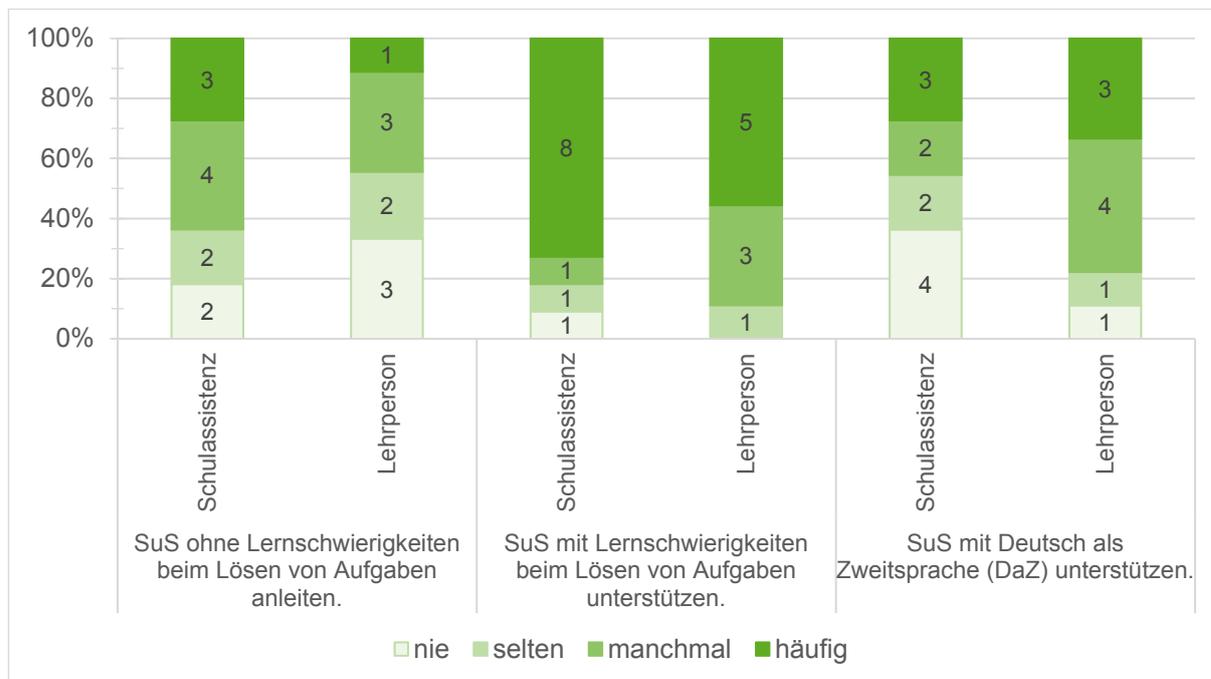
Anmerkungen. SuS = Schülerinnen und Schüler. Schulassistenzen: N = 11; Lehrpersonen: N = 9.

Im unterrichtsorganisatorischen Bereich im Deutschunterricht (vgl. Abbildung 6) steht nach Angaben der Schulassistenzen am häufigsten die Unterstützung der Schülerinnen und Schüler bei der Organisation und Strukturierung ihrer Arbeit im Vordergrund (*häufig*: 54.5 Prozent), gefolgt von der Beaufsichtigung von Kleingruppen bei Übungssequenzen und dem Herstellen von Ruhe in der Klasse (*häufig*: jeweils 45.5 Prozent). Die Befragung der Lehrpersonen ergibt für diese Tätigkeiten ein vergleichbares Bild. Eine Ausnahme stellt das Herstellen von Ruhe in der Klasse dar, das die

⁸ Der geringe Stellenwert des Unterrichts, der sich ähnlich für das Fach Mathematik ergibt (vgl. Abbildung 8), widerspricht scheinbar den zuvor berichteten vergleichsweise hohen prozentualen Angaben für das Unterrichten im Fach (vgl. Tabelle 19 und Tabelle 20). Anders als zuvor wird hier jedoch detaillierter nach einzelnen Tätigkeiten im Unterricht gefragt. Dabei wird erkennbar, dass das eigentliche Unterrichten eine eher nachgeordnete Rolle für die Schulassistenzen gespielt hat.

Lehrpersonen deutlich weniger häufig berichten – ebenso wie eine damit zusammenhängende Tätigkeit, das Ermahnen störender Schülerinnen und Schüler.

Abbildung 7. Unterstützung spezifischer Schülergruppen durch die Schulassistenzen während des Unterrichts im Fach Deutsch nach eigenen Angaben und nach Angaben der Lehrpersonen



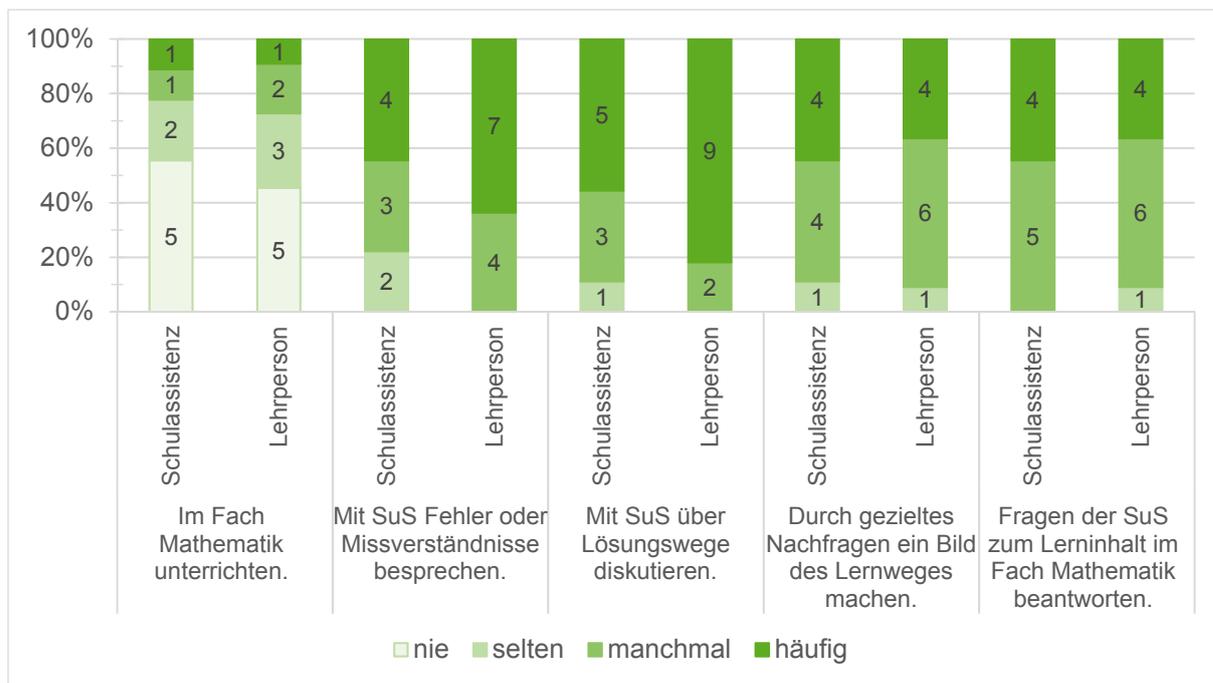
Anmerkungen. SuS = Schülerinnen und Schüler. Schulassistenzen: N = 11; Lehrpersonen: N = 9.

Als besonders häufige Tätigkeit bei der Unterstützung spezifischer Schülergruppen im Fach Deutsch (vgl. Abbildung 7) berichten die Schulassistenzen die Unterstützung von Schülerinnen und Schülern mit Lernschwierigkeiten beim Lösen von Aufgaben (*häufig*: 72.7 Prozent). Diese wird von allen Tätigkeiten – über alle erfragten Tätigkeitsbereiche hinweg – am häufigsten angegeben. Auch von den Lehrpersonen wird die Unterstützung dieser Schülergruppe als das häufigste Einsatzgebiet der Schulassistenzen berichtet (*häufig*: 55.6 Prozent). Eine Arbeit mit Schülerinnen und Schülern ohne Lernschwierigkeiten wird von den Schulassistenzen häufiger, mit Schülerinnen und Schülern mit Deutsch als Zweitsprache seltener angegeben als von den Lehrpersonen.

Die fachinhaltlichen Tätigkeiten im Mathematikunterricht ähneln denen im Deutschunterricht (vgl. Abbildung 8). Auch hier geben die Schulassistenzen an, vergleichsweise am häufigsten mit den Schülerinnen und Schülern über Lösungswege zu diskutieren (*häufig*: 55.6 Prozent). Die Lehrpersonen betonen diese Tätigkeit noch mehr, neun der elf Lehrpersonen (81.8 Prozent) berichten, dies geschehe «häufig». Auch das Besprechen von Fehlern/Missverständnissen, das gezielte Nachfragen bei den Schülerinnen und Schülern und das Beantworten von Fragen der Schülerinnen

und Schüler werden fast durchgehend von allen Befragten als gelegentliche oder häufige Tätigkeiten berichtet. Wie im Deutschunterricht wird das Unterrichten im Fach am wenigsten genannt, von den Schulassistenzen (*nie*: 55.6 Prozent) wie von den Lehrpersonen (*nie*: 45.5 Prozent).

Abbildung 8. Fachinhaltliche Tätigkeiten der Schulassistenzen während des Unterrichts im Fach Mathematik nach eigenen Angaben und nach Angaben der Lehrpersonen

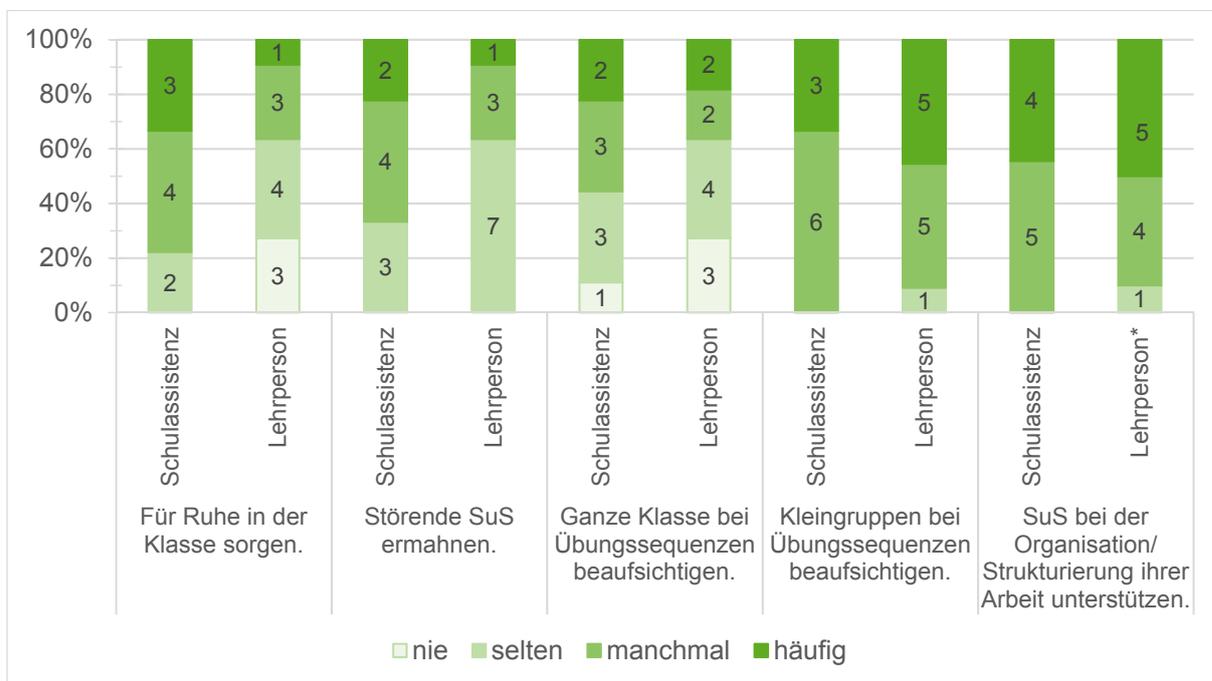


Anmerkungen. SuS = Schülerinnen und Schüler. Schulassistenzen: N = 9, fehlend: N = 1; Lehrpersonen: N = 11, fehlend: N = 1.

Auch im unterrichtsorganisatorischen Bereich zeigt sich im Fach Mathematik ein ähnliches Bild wie im Fach Deutsch (vgl. Abbildung 9). Von den Schulassistenzen am häufigsten genannt wird die Unterstützung der Schülerinnen und Schüler bei der Organisation und Strukturierung ihrer Arbeit (*häufig*: 44.4 Prozent), gefolgt von der Beaufsichtigung von Kleingruppen bei Übungssequenzen und dem Herstellen von Ruhe in der Klasse (*häufig*: jeweils 33.3 Prozent). Die Angaben der Lehrpersonen bestätigen dieses Tätigkeitsprofil weitgehend; wie bereits im Fach Deutsch berichten sie das Herstellen von Ruhe – und ähnlich, das Ermahnen störender Schülerinnen und Schüler – seltener.

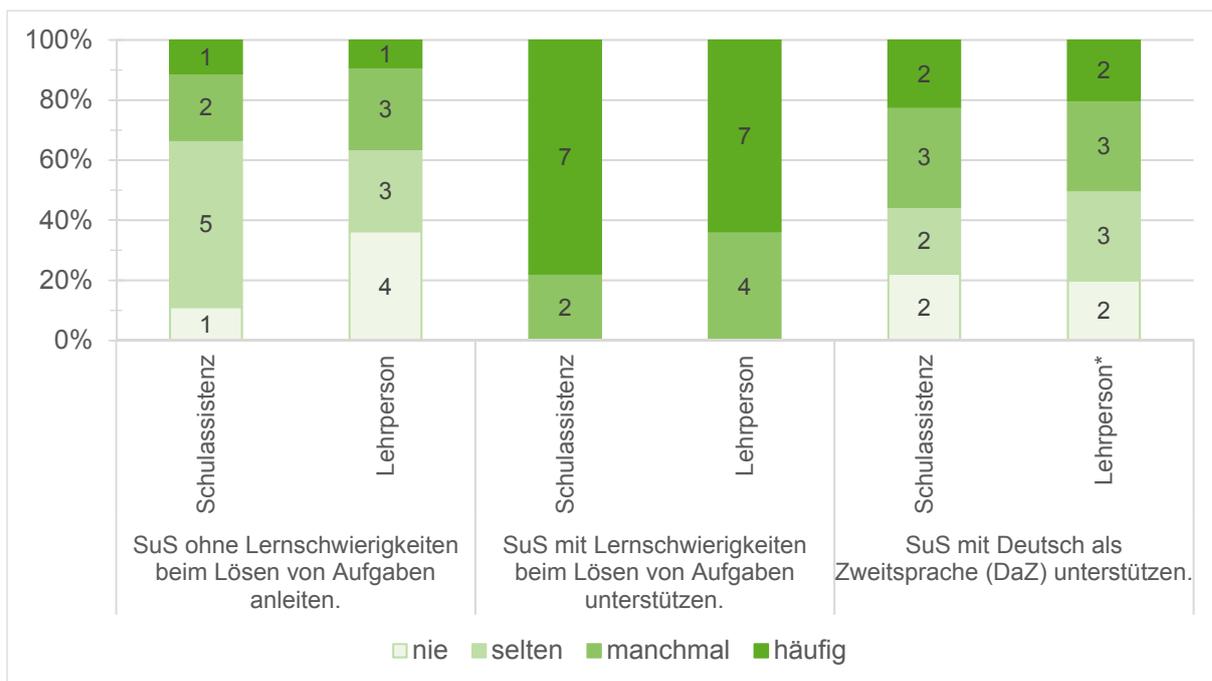
Nicht nur im Bereich der Unterstützung spezifischer Schülergruppen, sondern über alle erfragten Tätigkeiten hinweg, wird auch im Fach Mathematik (vgl. Abbildung 10) von Schulassistenzen und Lehrpersonen als häufigste Tätigkeit die Unterstützung von Schülerinnen und Schülern mit Lernschwierigkeiten beim Lösen von Aufgaben angegeben (*häufig*: Schulassistenzen 77.8 Prozent; Lehrpersonen 63.6 Prozent). Bei der Unterstützung von Schülerinnen und Schülern ohne Lernschwierigkeiten beziehungsweise mit Deutsch als Zweitsprache zeigt sich wie im Fach Deutsch, dass der Einbezug der Schulassistenzen individuell deutlich verschieden ist.

Abbildung 9. Unterrichtsorganisatorische Tätigkeiten der Schulassistenten während des Unterrichts im Fach Mathematik nach eigenen Angaben und nach Angaben der Lehrpersonen



Anmerkungen. SuS = Schülerinnen und Schüler. Schulassistenten: N = 9, fehlend: N = 1; Lehrpersonen: N = 11, fehlend: N = 1; *N = 10, fehlend: N = 2.

Abbildung 10. Unterstützung spezifischer Schülergruppen durch die Schulassistenten während des Unterrichts im Fach Mathematik nach eigenen Angaben und nach Angaben der Lehrpersonen



Anmerkungen. SuS = Schülerinnen und Schüler. Schulassistenten: N = 9, fehlend: N = 1; Lehrpersonen: N = 11, fehlend: N = 1; *N = 10, fehlend: N = 2.

Die grosse Vielfalt ihrer Tätigkeiten wird von einer Schulasistenz im Einzelinterview anschaulich beschrieben:

«Ja, ich habe ganz viele Funktionen. Das geht von Deutschstunden und in der Mathe drin sein, mit einem autistischen Kind mit ins Kochen oder eben auch ins Deutsch, wo er Unterstützung braucht, oder dann ist da mein Platz im Förderzentrum. Da kommen viele Kinder, aus dem Ausland, wo ich dann das DAZ auch unterstütze im Deutsch, oder auch muss ich dort Polizistin sein. Es werden auch Kinder ins Förderzentrum geschickt, welche einen «Seich» [Anm.: Unsinn] gemacht haben oder die Lehrerin einfach sagt, dich ertrage ich jetzt nicht mehr. Bevor etwas eskaliert, schicke ich dich zum Arbeiten ins Förderzentrum. Einfach für alles. Auch wenn es einen Unfall gibt, rufen sie mich an, ob ich bitte kommen kann. Es ist sehr vielseitig, auch beim Kochen, Französisch, dort, wo es gerade jemanden braucht.»

Die Schulasistzenzen und Lehrpersonen wurden darüber hinaus nach der Häufigkeit von drei weiteren Tätigkeiten der Schulasistzenzen für den Unterricht gefragt: dem Schlichten von Konflikten zwischen Schülerinnen und Schülern (z.B. bei Mobbing), der Korrektur von Hausaufgaben/Prüfungen und der Durchführung von Projekten. Wie sich zeigt, wurden diese Tätigkeiten eher sporadisch von den Schulasistzenzen übernommen. Noch am häufigsten wird von den Schulasistzenzen das Schlichten von Konflikten als Tätigkeit berichtet, allerdings nur im Fach Deutsch (*manchmal*: Deutsch 36.4 Prozent, $N = 4$ [von 11]). Im Fach Mathematik geben acht der neun Schulasistzenzen (88.9 Prozent) an, dies *nie* oder *selten* zu tun. Zudem ergibt sich – vergleichbar den auf Klassenführung bezogenen unterrichtsorganisatorischen Tätigkeiten (vgl. Abbildung 6) – eine Diskrepanz zu den Angaben der Deutsch-Lehrpersonen, die das Schlichten von Konflikten kaum als Aufgabe der Schulasistzenzen betrachten (*nie* oder *selten*: 85.7 Prozent, $N = 6$ [von 7]). Daneben berichten einzelne Schulasistzenzen, Korrekturarbeiten für Hausaufgaben/Prüfungen zu übernehmen (*manchmal* oder *häufig*: Deutsch 27.3 Prozent, $N = 3$ [von 11]; Mathematik 33.3 Prozent, $N = 3$ [von 9]), wobei auch bei dieser Tätigkeit die Einschätzung der Deutsch-Lehrpersonen deutlich zurückhaltender ausfällt (*nie*: 85.7 Prozent, $N = 6$ [von 7]). Das Durchführen von Projekten spielt als Tätigkeit der Schulasistzenzen eine untergeordnete Rolle.

Einsatz der Schulasistzenzen ausserhalb des Unterrichts

Um die Einsatzgebiete der Schulasistzenzen möglichst vollständig abzubilden, wurde zusätzlich nach drei Tätigkeiten ausserhalb des Unterrichts gefragt: der Teilnahme an Teambesprechungen, dem Begleiten von Ausflügen und der Planung von Klassenveranstaltungen und Schulanlässen. Von den Schulasistzenzen in beiden Fächern wird das Begleiten von Ausflügen als häufigste der Tätigkeiten genannt (*manchmal* oder *häufig*: Deutsch 81.8 Prozent, $N = 9$ [von 11]; Mathematik 100.0 Prozent, $N = 9$ [von 9]). Auch die Planung von Veranstaltungen gehört nach Angaben einer Mehrheit der Schulasistzenzen zu den regelmässigeren Tätigkeiten (*manchmal* oder *häufig*:

Deutsch 72.7 Prozent, $N = 8$ [von 11]; Mathematik 55.6 Prozent, $N = 5$ [von 9]). Die geringste Häufigkeit ergibt sich für die Teilnahme an Teambesprechungen (*nie* oder *selten*: Deutsch 72.7 Prozent, $N = 8$ [von 11]; Mathematik 55.6 Prozent, $N = 5$ [von 9]). Die Angaben der Lehrpersonen zeigen ein insgesamt vergleichbares Bild, allerdings berichten zwei Drittel der Mathematik-Lehrpersonen von einer regelmässigeren Teilnahme der Schulassistenten an Teambesprechungen (*manchmal* oder *häufig*: 66.7 Prozent, $N = 8$ [von 12]).

Zusammenfassende Betrachtung der Tätigkeitsbereiche

Betrachtet man, welche der 21 erfragten Tätigkeiten die einzelnen Schulassistenten nach eigenen Angaben «häufig» ausüben, wird deutlich, dass die Aufgabengebiete individuell sehr unterschiedlich sind. Es lassen sich – auch aufgrund der geringen Stichprobengrösse – keine für die Schulassistenten typischen Tätigkeitsprofile erkennen. Die Anzahl der pro Schulassistentin als «häufig» berichteten Tätigkeiten variiert deutlich, im Fach Deutsch zwischen null und 13 ($M = 5.6$, $SD = 4.2$), im Fach Mathematik zwischen einer und 17 ($M = 6.0$, $SD = 5.0$) der 21 erfragten Tätigkeiten.

Jedoch lassen sich Tätigkeiten identifizieren, für die alle Schulassistenten angeben, dass diese nicht die Regel sind und nur sporadisch ausgeübt werden. In Tabelle 21 werden die Tätigkeiten aufgeführt, für welche *alle* Schulassistenten berichten, sie *höchstens manchmal* durchzuführen.

Tabelle 21. Sporadisch ausgeübte Tätigkeiten der Schulassistenten nach eigenen Angaben

Deutsch	Mathematik
Lernangebote (Lerninhalt und Aktivitäten) für den Deutschunterricht gestalten	Projekte durchführen
Aufgaben und Materialien für den Deutschunterricht bereitstellen (z.B. Unterlagen kopieren, Medien bereitstellen)	Klassenveranstaltungen und Schulanlässe planen
Im Fach Deutsch unterrichten	
Konflikte zwischen Schülerinnen und Schülern schlichten (z.B. bei Mobbing)	
Hausaufgaben/Prüfungen für das Fach Deutsch korrigieren	
Projekte durchführen	
An Teambesprechungen teilnehmen	

Anmerkungen. Deutsch: $N = 11$; Mathematik: $N = 9$, fehlend: $N = 1$.

Einsatz der Schulassistenten in weiteren Lernfeldern

Neben Einsätzen im Regelunterricht in den Fächern Deutsch und Mathematik wurden die Schulassistenten in weiteren Lernfeldern eingesetzt. Wie aus Tabelle 22 ersichtlich wird, arbeiteten knapp drei Viertel der Schulassistenten zusätzlich im Bereich der Integrierten Sonderschulung (ISR) und je knapp zwei Drittel in der Aufgabenhilfe und der Berufsberatung.

Tabelle 22. Einsatz der Schulassistenzen in weiteren Lernfeldern nach eigenen Angaben

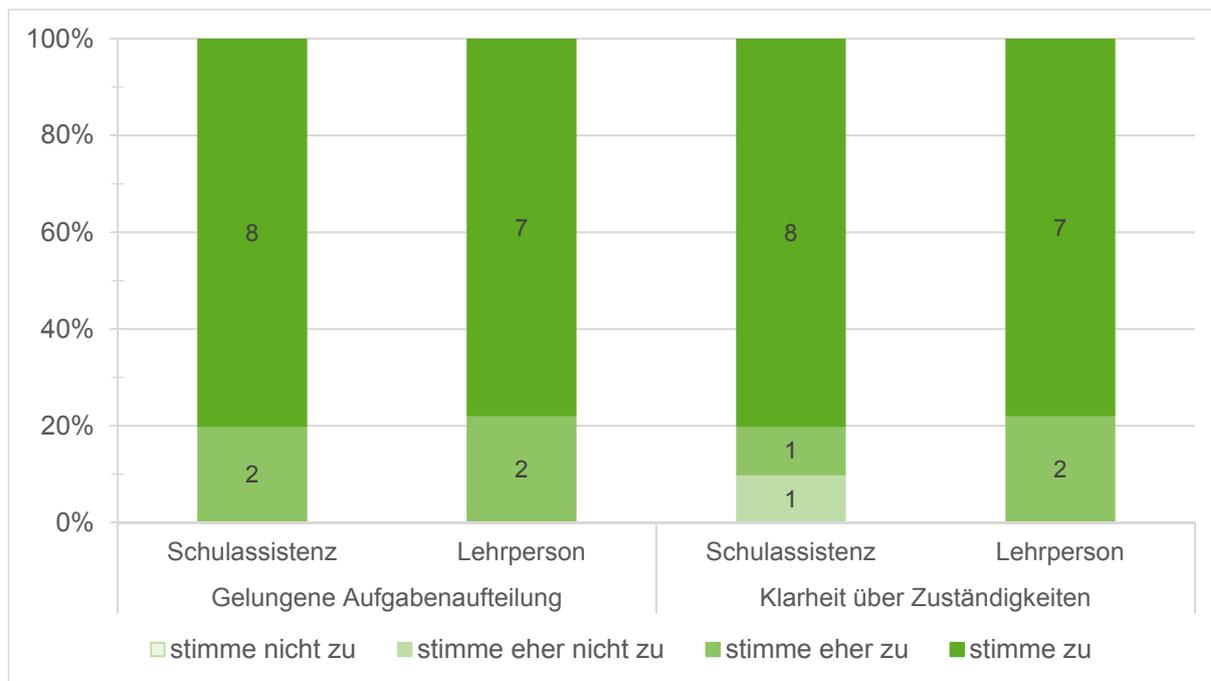
Lernfeld	Anzahl	in Prozent
ISR	10	71.4
Aufgabenhilfe	9	64.3
in der Berufsberatung	9	64.3
bei Elternanlässen	5	35.7
in einem anderen Lernfeld*	3	27.3

Anmerkungen. ISR = Integrierte Sonderschulung. N = 14; *N = 11, fehlend: N = 3. Antwortformat pro Item: ja – nein.

4.2.4 Qualität der Zusammenarbeit

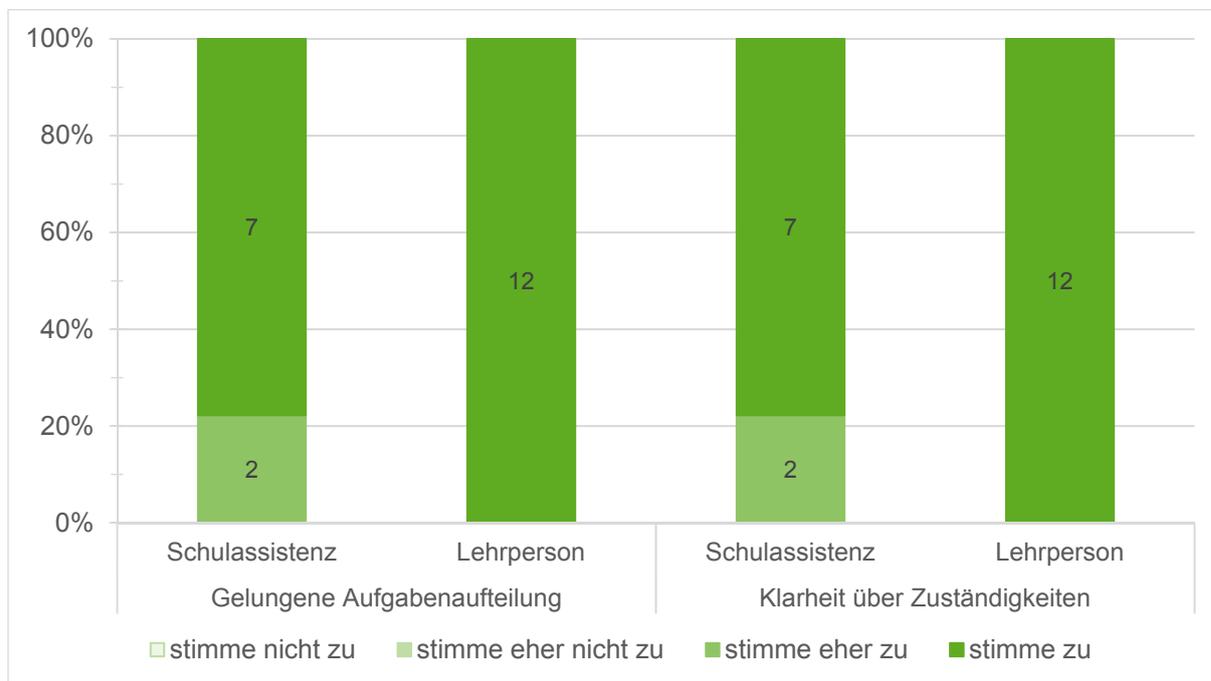
Die Schulassistenzen und Lehrpersonen wurden um eine Einschätzung gebeten, wie gut die Qualität der Aufgabenteilung zwischen ihnen gelungen ist, und wie klar die Zuständigkeiten zwischen ihnen geregelt sind. Die Ergebnisse lassen sich Abbildung 11 und Abbildung 12 für die Fächer Deutsch beziehungsweise Mathematik entnehmen.

Abbildung 11. Einschätzungen zur Aufgabenteilung und zu den Zuständigkeiten im Fach Deutsch nach Angaben der Schulassistenzen und der Lehrpersonen



Anmerkungen. Schulassistenzen: N = 10, fehlend: N = 1; Lehrpersonen: N = 9.

Abbildung 12. Einschätzungen zur Aufgabenteilung und zu den Zuständigkeiten im Fach Mathematik nach Angaben der Schulklassen und der Lehrpersonen



Anmerkungen. Schulklassen: $N = 9$, fehlend: $N = 1$; Lehrpersonen: $N = 12$.

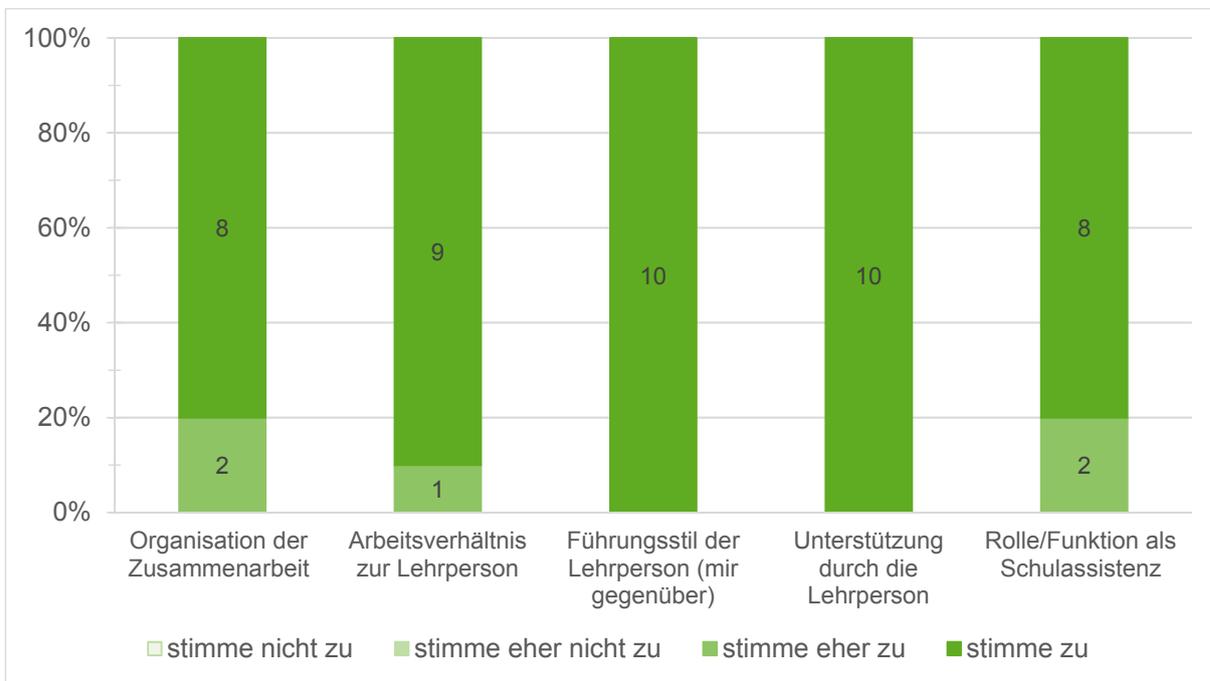
Die Aufgabenteilung zwischen Lehrperson und Schulassistent sehen in beiden Fächern mehr als drei Viertel der befragten Schulklassen als vollständig gelungen und klar kommuniziert an (*stimme zu*: Deutsch jeweils 80.0 Prozent; Mathematik jeweils 77.8 Prozent). Das Urteil der Lehrpersonen, die über eine Schulassistent verfügen, fällt gleichermassen positiv aus. Im Fach Mathematik stimmen sogar alle befragten Lehrpersonen den Aussagen vollständig zu.

Ein Beispiel für eine gelungene Zusammenarbeit beschreibt eine Schulassistent im Einzelinterview wie folgt:

«Was ich super fand, in den ersten drei Jahren hatte ich eine ISR-Schülerin [Anm.: Schülerin der Integrierten Sonderschulung]. Miteinander haben wir diese super begleitet, die Lehrperson und der Heilpädagoge. Am Schluss erhielt sie ein Praktikum. Wir haben fest miteinander gearbeitet und sie musste wirklich auch Tests machen, vor allem auch Mathe, wo sie Schwierigkeiten hatte, und sie hat das sehr gut bestanden. [...]»

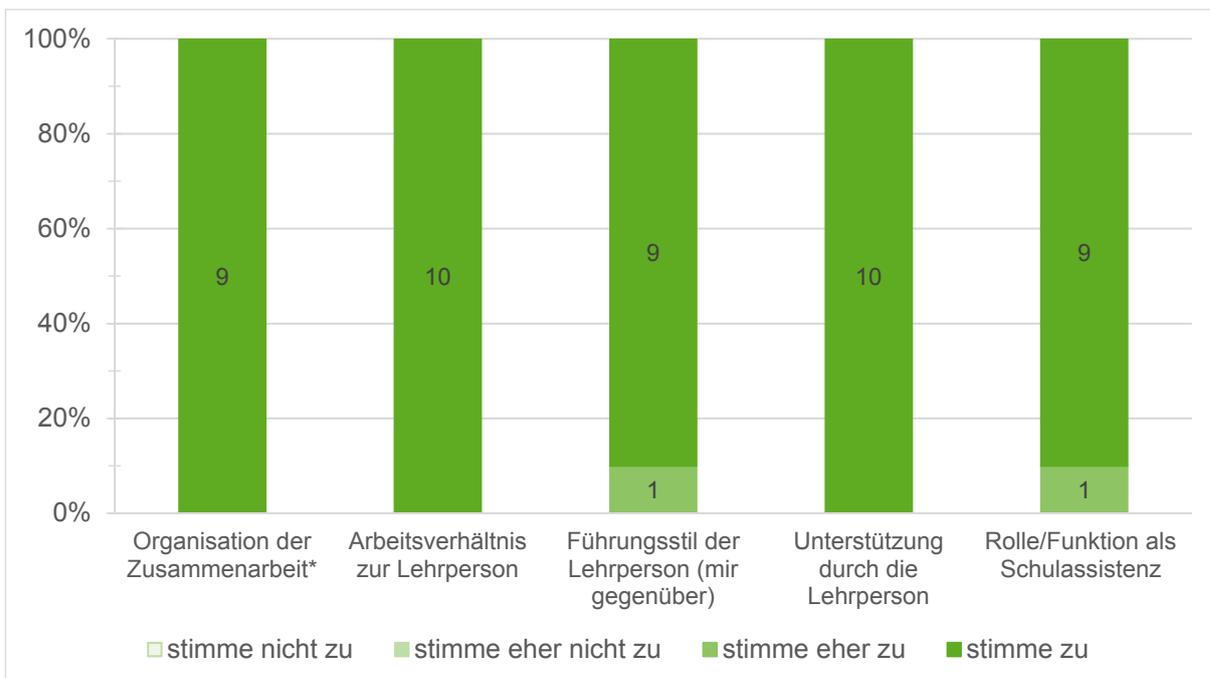
Diesem positiven Bild entsprechend äussern sich die befragten Schulklassen insgesamt sehr zufrieden mit der Führung durch die Lehrpersonen und stimmen bei den erfragten Aspekten des Führungsverhaltens fast immer vollständig zu (vgl. Abbildung 13 und Abbildung 14).

Abbildung 13. Zufriedenheit mit der Führung durch die Lehrperson im Fach Deutsch nach Angaben der Schulassistenzen



Anmerkungen. N = 10, fehlend: N = 1.

Abbildung 14. Zufriedenheit mit der Führung durch die Lehrperson im Fach Mathematik nach Angaben der Schulassistenzen



Anmerkungen. N = 10; *N = 9, fehlend: N = 1.

Die Bedeutung einer guten Führung durch die Lehrperson hebt eine Schulasistenz im Einzelinterview wie folgt hervor:

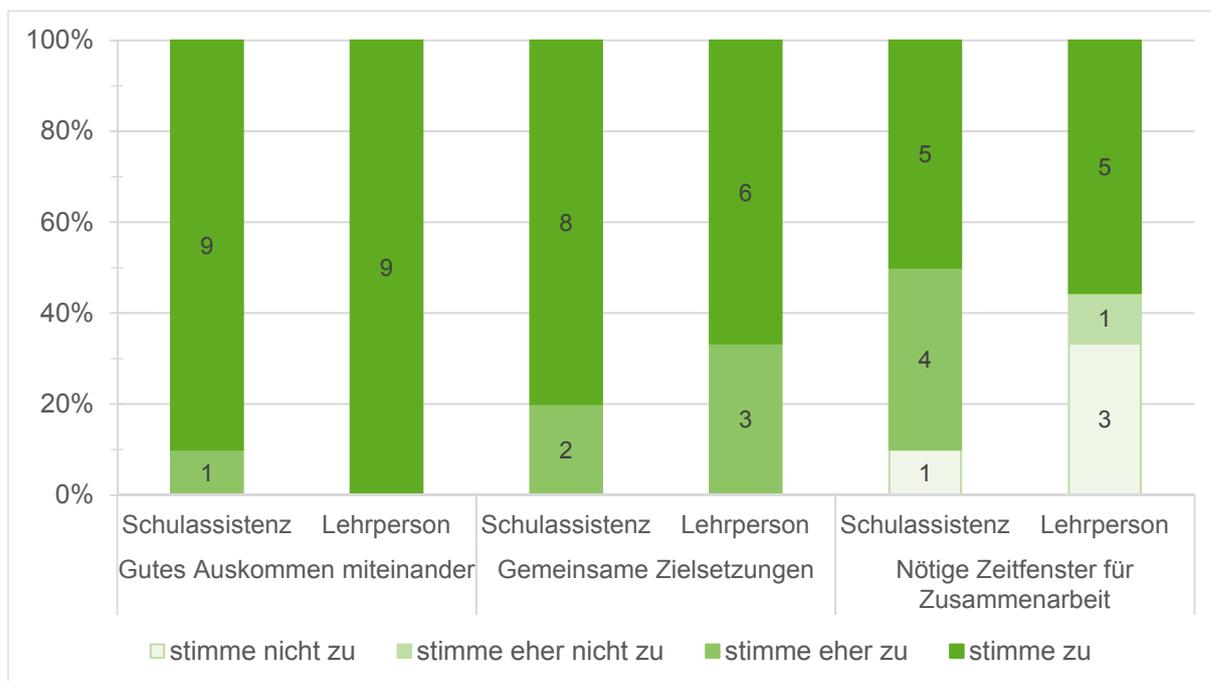
«Ich habe das Gefühl, für mich ist es wichtig, dass ich weiss, was diese bestimmte Lehrperson, mit der ich zusammenarbeite, was sie von mir erwartet, wie sie mich gerne einsetzt, dass ich das weiss. Sicher die Kommunikation ist ganz ein wichtiger Punkt.»

Eine andere Schulasistenz äussert sich allgemein zufrieden zur Ausgestaltung der Zusammenarbeit mit den Lehrpersonen und betont das «Geben und Nehmen»:

«Ich habe es eigentlich mit allen Lehrern sehr positiv erlebt. [...] Es hat wirklich super funktioniert. Es war ein grosses Geben und Nehmen. Ich habe mich sowieso mega flexibel eingestellt. Das macht es für die Lehrer sicher auch ein wenig einfacher. Zum Teil müssen sie ziemlich nach Plan und sind mit den Stunden gebunden. Wenn ich dann mal etwas für sie machen kann, sind sie bestimmt auch froh.»

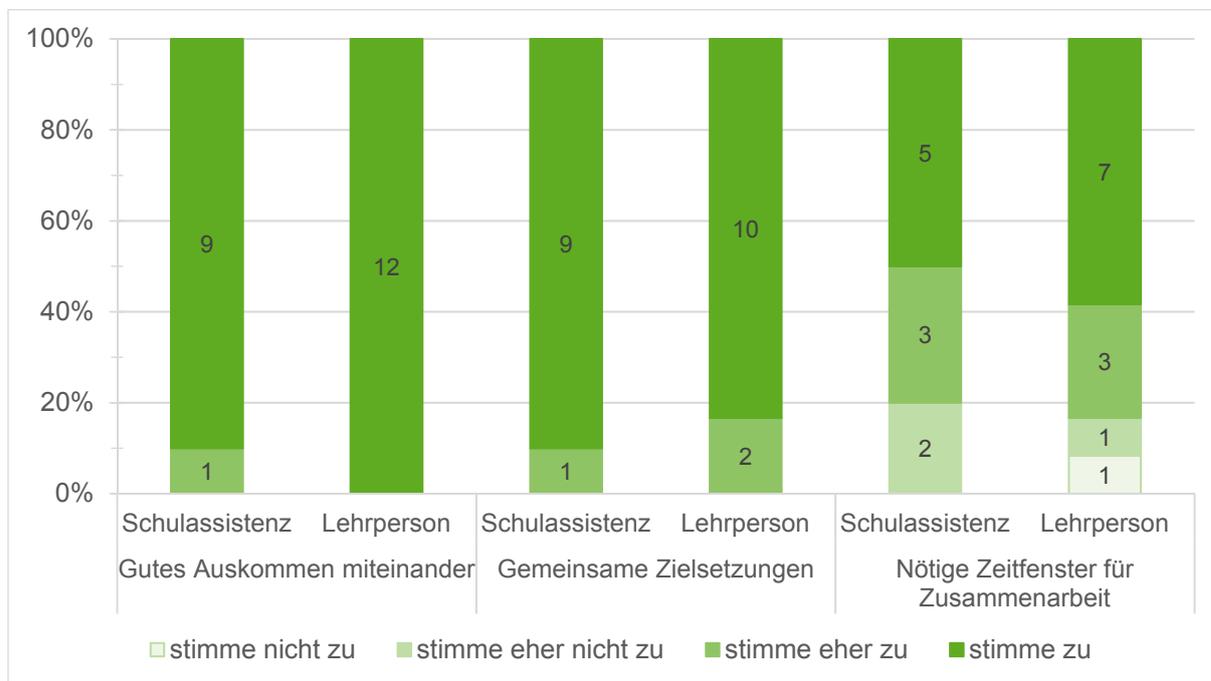
Auch bezüglich wichtiger persönlicher Aspekte der Zusammenarbeit äussern sich sowohl Schulasistenzen als auch Lehrpersonen zufrieden (vgl. Abbildung 15 und Abbildung 16). Fast alle befragten Schulasistenzen stimmen zu, mit der Lehrperson ein «gutes Auskommen» zu haben und sich über Zielsetzungen einig zu sein (*stimme zu*: Deutsch/Mathematik je 90.0 Prozent). Die Aussagen der Lehrpersonen bestätigen die positive Wahrnehmung der Schulasistenzen.

Abbildung 15. Einschätzungen zur Zusammenarbeit im Fach Deutsch nach Angaben der Schulasistenzen und der Lehrpersonen



Anmerkungen. Schulasistenzen: N = 10, fehlend: N = 1; Lehrpersonen: N = 9.

Abbildung 16. Einschätzungen zur Zusammenarbeit im Fach Mathematik nach Angaben der Schulassistenzen und der Lehrpersonen



Anmerkungen. Schulassistenzen: N = 10; Lehrpersonen: N = 12.

Dass die Einpassung der Schulassistenzen in den schulischen Betrieb dennoch in einzelnen Fällen nicht reibungsfrei verlaufen sein dürfte, wird aus der Äusserung einer Schulleitung deutlich:

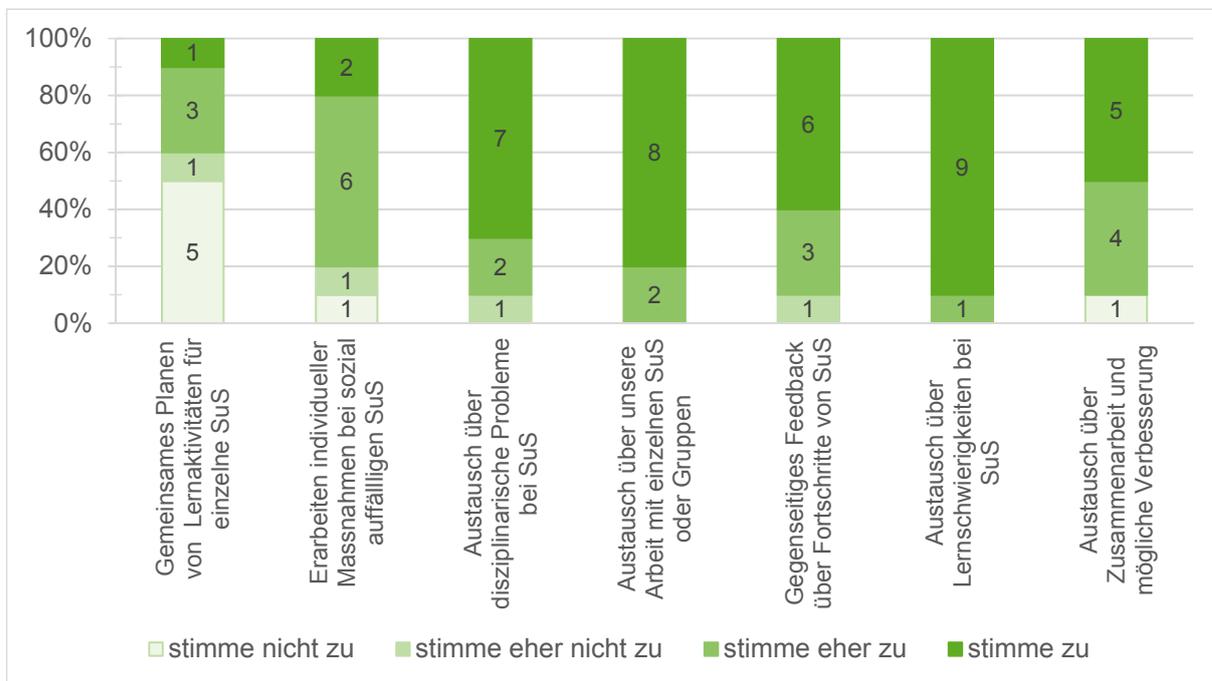
«Die Lehrpersonen haben sich wie Katzen gegen das Wasser gesträubt, eine Schulassistentenz in ihr Zimmer zu lassen [...]. Egal wie gleichgültig oder gut die Person gewesen ist, oder wäre [...] und die Klassenassistentenzen haben sich geweigert zu einzelnen Lehrpersonen zu gehen [...]»

In diesem Zusammenhang ist nochmals hervorzuheben, dass die in diesem Kapitel berichteten Angaben von den eher wenigen Lehrpersonen stammen, welche berichten, zum Zeitpunkt der Befragung gegen Ende des Pilotprojekts eine Schulassistentenz zu haben. Möglich ist, dass Lehrpersonen, welche der Massnahme besonders kritisch gegenüberstanden (vgl. das Zitat der Schulleitung), zu diesem Zeitpunkt keine Schulassistentenz (mehr) hatten oder an der Befragung nicht teilnahmen. Inwieweit dies der Fall war, lässt sich anhand der vorliegenden Daten nicht feststellen.

Von den Schulassistentenzen und Lehrpersonen nicht immer günstig beurteilt werden die zeitlichen Rahmenbedingungen (vgl. Abbildung 15 und Abbildung 16). Am deutlichsten wird dies im Urteil der Deutsch-Lehrpersonen: Knapp die Hälfte berichtet, dass die nötigen Zeitfenster für die Zusammenarbeit (eher) nicht vorhanden sind (*stimme nicht zu* oder *stimme eher nicht zu*: 44.4 Prozent).

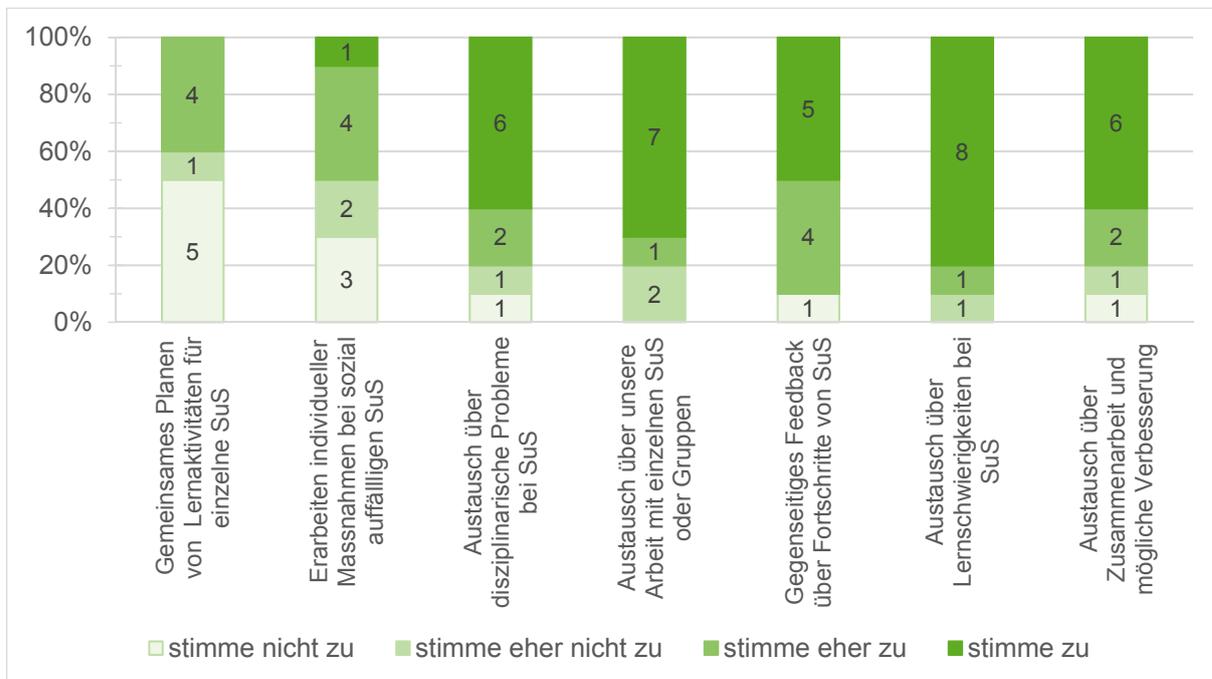
Die Schulassistentenzen wurden weiterhin gefragt, inwieweit einzelne Tätigkeiten kooperativen Arbeitens mit den Lehrpersonen für den Unterricht vorkommen. Die Ergebnisse werden in Abbildung 17 und Abbildung 18 dargestellt.

Abbildung 17. Kooperatives Arbeiten mit der Lehrperson im Fach Deutsch nach Angaben der Schulasstiszenzen



Anmerkungen. SuS = Schülerinnen und Schüler. N = 10, fehlend: N = 1.

Abbildung 18. Kooperatives Arbeiten mit der Lehrperson im Fach Mathematik nach Angaben der Schulasstiszenzen



Anmerkungen. SuS = Schülerinnen und Schüler. N = 10.

In beiden Fächern zeigt sich für die meisten erfragten Aktivitäten eine deutliche Zustimmung der Schül*innen. Für das Fach Deutsch stimmen sie am meisten zu, sich mit den Lehrpersonen über Lernschwierigkeiten von Schül*innen und Schülern (*stimme zu*: 90.0 Prozent), die Arbeit mit einzelnen Lernenden oder Gruppen (*stimme zu*: 80.0 Prozent) und disziplinarische Probleme mit Schül*innen und Schülern (*stimme zu*: 70.0 Prozent) auszutauschen. Am wenigsten häufig geben sie an, in die Planung gemeinsamer Lernaktivitäten für einzelne Schül*innen und Schülern involviert zu sein (*stimme nicht zu*: 50.0 Prozent). Im Fach Mathematik ergeben sich vergleichbare Befunde. Allerdings wird seltener von einer gemeinsamen Erarbeitung individueller Massnahmen bei sozial auffälligen Schül*innen und Schülern berichtet (*stimme nicht zu* oder *stimme eher nicht zu*: Deutsch 20.0 Prozent, Mathematik 50.0 Prozent).

Insgesamt steht in beiden Fächern der Austausch beziehungsweise das wechselseitige Feedback zwischen Schül*innen und Lehrpersonen im Vordergrund. Dies betrifft sowohl das Lernen als auch das disziplinarische Verhalten der Schül*innen und Schülern sowie die Kommunikation über die Zusammenarbeit selbst. Weniger relevant ist dagegen, insbesondere im Fach Mathematik, die konkrete Planung von Massnahmen.

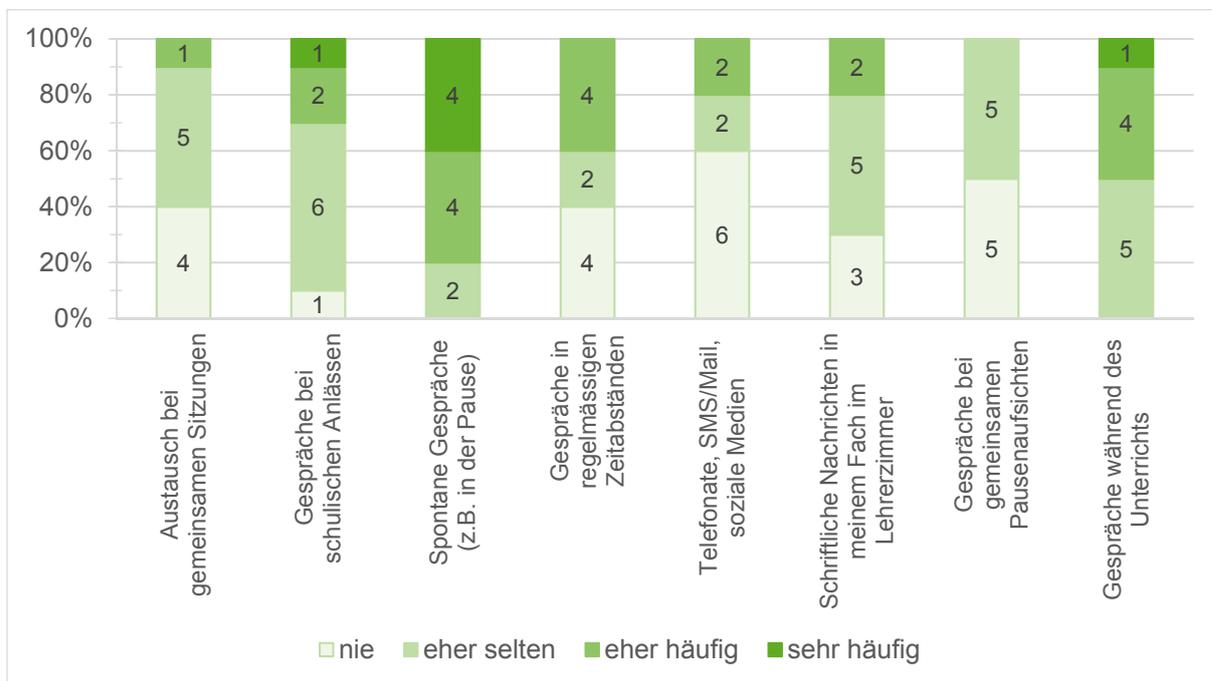
Zur Qualität und zur Bedeutung des Austausches mit der Lehrperson äussert sich eine Schül*in im Einzelinterview wie folgt:

«Für mich ist eine gute Zusammenarbeit, dass ich immer rückfragen kann oder sagen kann, dass ich anstehe und wie fahren wir weiter? Oder dass ich auch Infos erhalte, welche für mich mit dem Zusammenarbeiten mit diesen Kindern wichtig sind. Das ist wirklich schön, war auch immer schön.»

Und weiter: «Vielleicht dann mehr so in den Fächern ‹Machst du mit ihnen das und das und das›, ‹Schau, das geht und das andere lassen wir›. Ich werde da recht gut unterstützt. Auf der anderen Seite kann ich aber auch wieder kommen und sagen, das geht jetzt wirklich nicht, das ist zu schwierig. Da müssen wir schauen, dass wir etwas Anderes machen können.»

Eine letzte Frage an die Schül*innen bezog sich auf die Häufigkeit bestimmter Kommunikationsformen zwischen ihnen und den Lehrpersonen. Die entsprechenden Angaben lassen sich Abbildung 19 und Abbildung 20 für die Fächer Deutsch und Mathematik entnehmen. Wie sich zeigt, spielen in beiden Fächern spontane Gespräche, zum Beispiel in der Pause, eine besondere Rolle (*sehr häufig*: Deutsch 40.0 Prozent; Mathematik 60.0 Prozent). Weitere Gesprächsformen mit einiger Häufigkeit sind Gespräche während des Unterrichts (*eher häufig* oder *sehr häufig*: Deutsch 50.0 Prozent; Mathematik 40.0 Prozent) sowie Gespräche in regelmässigen Zeitabständen (*eher häufig* oder *sehr häufig*: Deutsch 40.0 Prozent; Mathematik 50.0 Prozent). Im Fach Mathematik treten zudem Gespräche bei gemeinsamen Pausenaufsichten vergleichsweise häufig auf (*eher häufig* oder *sehr häufig*: 40.0 Prozent, $N = 4$).

Abbildung 19. Kommunikationsformen von Lehrperson und Schulasistenz im Fach Deutsch nach Angaben der Schulasstizenzen



Anmerkungen. N = 10, fehlend: N = 1.

Abbildung 20. Kommunikationsformen von Lehrperson und Schulasistenz im Fach Mathematik nach Angaben der Schulasstizenzen



Anmerkungen. N = 10.

Die eher geringen Häufigkeiten dieser Kommunikationsformen legen die Frage nahe, ob manche Schulassistenzen nur sehr selten im Austausch mit den Lehrpersonen standen. Um dem nachzugehen, wurde je Schulassistentz ausgezählt, für wie viele Kommunikationsformen die Angabe «nie» oder «eher selten» erfolgte. Im Fach Deutsch berichtet eine Schulassistentz, dass alle Kommunikationsformen zwar vorkommen, aber «eher selten» sind; eine weitere benennt vier Kommunikationsformen, die «eher selten» stattfinden (z.B. Austausch während Sitzungen, spontane Gespräche), während die anderen gar nicht vorkommen. Im Fach Mathematik berichtet eine Schulassistentz, dass drei Kommunikationsformen «eher selten» sind (spontane Gespräche, Gespräche in regelmässigen Zeitabständen, Gespräche während des Unterrichts) und die anderen nicht vorkommen.

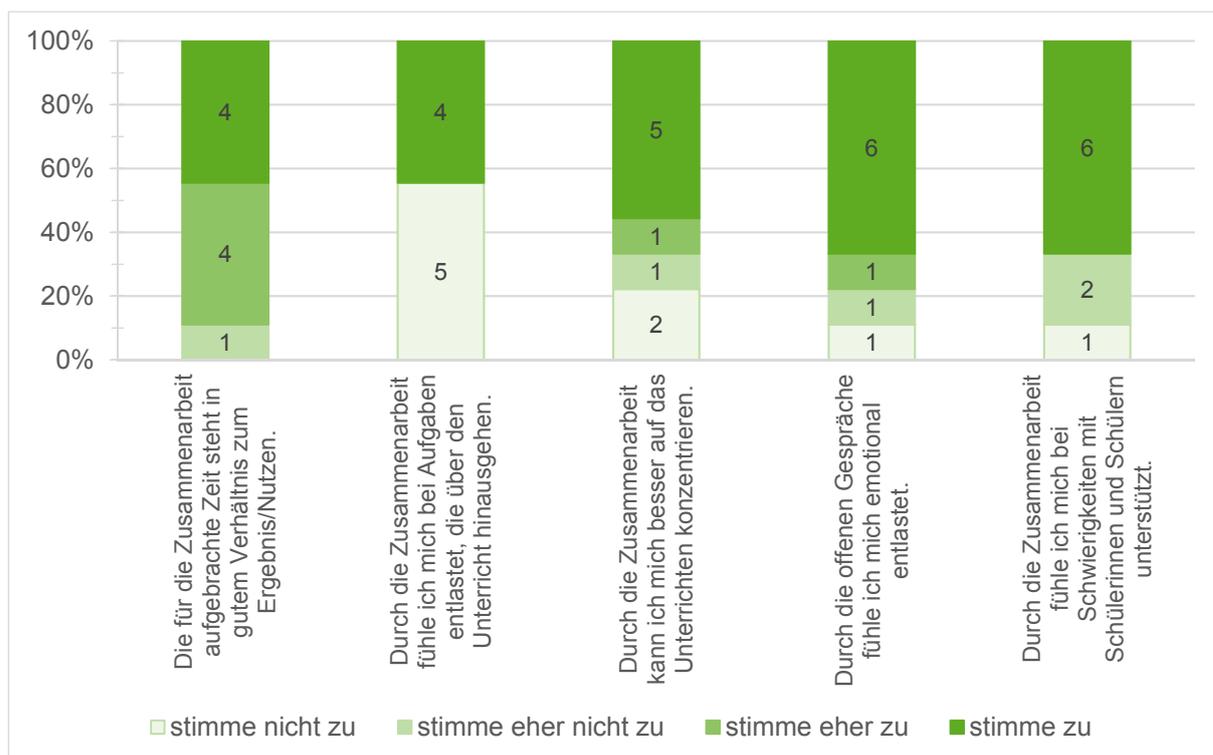
4.2.5 Wahrgenommene Unterstützung durch die Schulassistentzen

Lehrpersonen sowie Schülerinnen und Schüler wurden dazu befragt, ob sie den Einsatz der Schulassistentzen im Unterricht als unterstützend erleben. Im Folgenden wird zunächst die Wahrnehmung der Lehrpersonen, dann diejenige der Lernenden dargestellt.

Wahrnehmung der Lehrpersonen

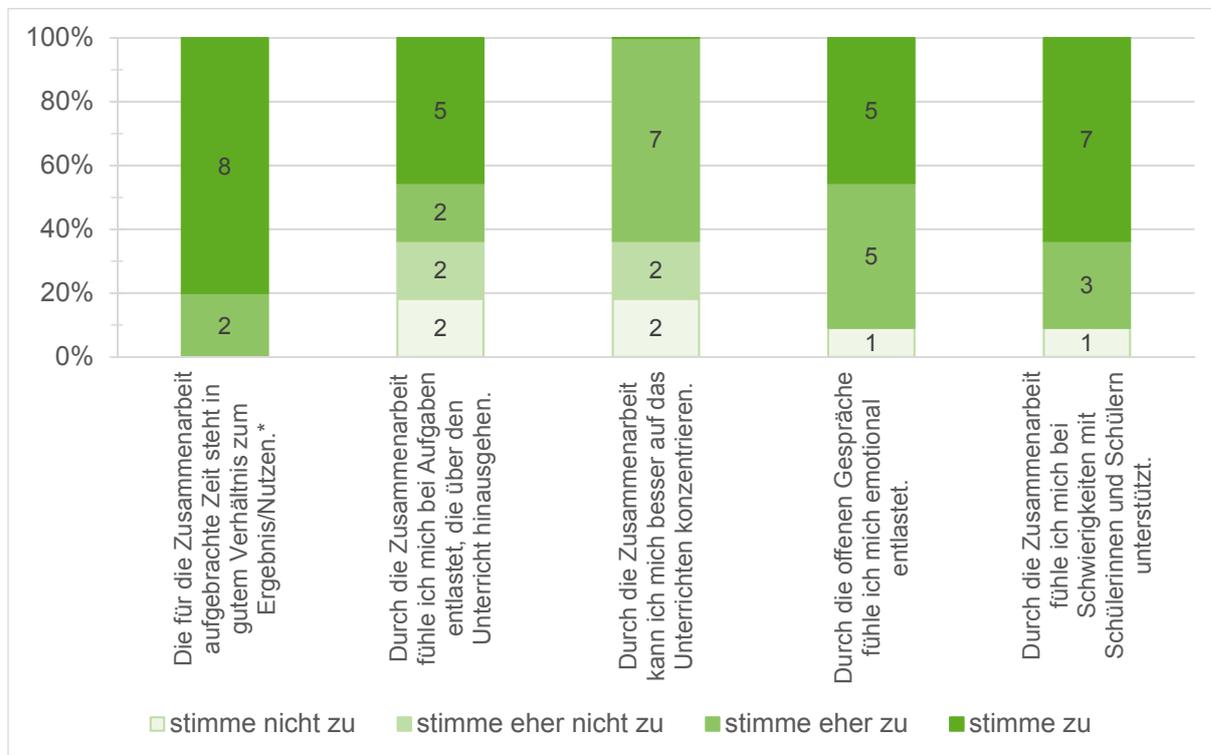
Abbildung 21 und Abbildung 22 zeigen, inwieweit die Deutsch- und Mathematik-Lehrpersonen die Zusammenarbeit mit den Schulassistentzen in verschiedener Hinsicht als unterstützend erlebten.

Abbildung 21. Wahrgenommene Unterstützung durch die Schulassistentz im Fach Deutsch nach Angaben der Lehrpersonen



Anmerkungen. N = 9.

Abbildung 22. Wahrgenommene Unterstützung durch die Schulassistenten im Fach Mathematik nach Angaben der Lehrpersonen



Anmerkungen. N = 11, fehlend: N = 1; *N = 10, fehlend: N = 2.

Ein Beispiel, wie sich eine Lehrperson von der Schulassistenten entlastet fühlt, findet sich in den Gruppeninterviews mit den Lehrpersonen:

«Dann ist es mit dem Zivi finde ich super, also mit der Schulassistenten klappt das gut, weil ich habe noch 19 andere, welche ich betreuen kann. Umgekehrt aber auch, er kann für Ruhe sorgen, wenn Fragen auftauchen, und ich kann mich auf die Schwachen konzentrieren. Wir setzen ihn auch in den Niveaus ein. Dort ist es auch ähnlich, einfach mehr Zeit, Raum zu schaffen für Schwächere, damit man gut mit ihnen arbeiten kann.»

Ähnlich äussert sich eine Lehrperson an einer anderen Schule:

«[...] und [die Schulassistenten] hat mir sehr viel abgenommen. Mit ihr konnte ich sagen, schau, jetzt machen wir heute das, ich wäre froh, wenn du das machen würdest.»

Dass die Zusammenarbeit mit der Schulassistenten auch ambivalent erlebt werden kann, wenn sie den Erwartungen der Lehrperson nicht ganz gerecht wird, zeigt folgende Äusserung:

«Einerseits war es eine Entlastung [...], wo man Schüler auch einfach mal aus dem Unterricht nehmen und draussen arbeiten lassen konnte und sie waren beaufsichtigt. Andererseits waren es aber auch Punkte, welche [die andere Lehrperson] angesprochen hat. Man hat schon die Erwartung, dass ein Klassenassistent dann schaut, dass es leise ist. Wenn das nicht passiert, kann es unter Umständen ein weiterer störender Punkt im Unterricht sein, wo man gucken muss.»

In der Äusserung einer weiteren Lehrperson wird eine Entlastung – oder allfällige zusätzliche Belastung – der Lehrperson in Beziehung zu den (fachlichen) Fähigkeiten der Schulassistentz gesetzt:

[Lehrperson A:] «Ich finde, es kommt darauf an, was diese Klassenassistentz an Kompetenzen mitbringt. Wenn ich einen zusätzlichen Schüler habe, dann bringt es mir nichts.»

[Lehrperson B:] «Ich denke, das ist ganz wichtig, was [Lehrperson A] gesagt hat. Gerade auf nämlich jetzt eine A-Klasse, einen zusätzlichen Schüler brauche ich nicht.»

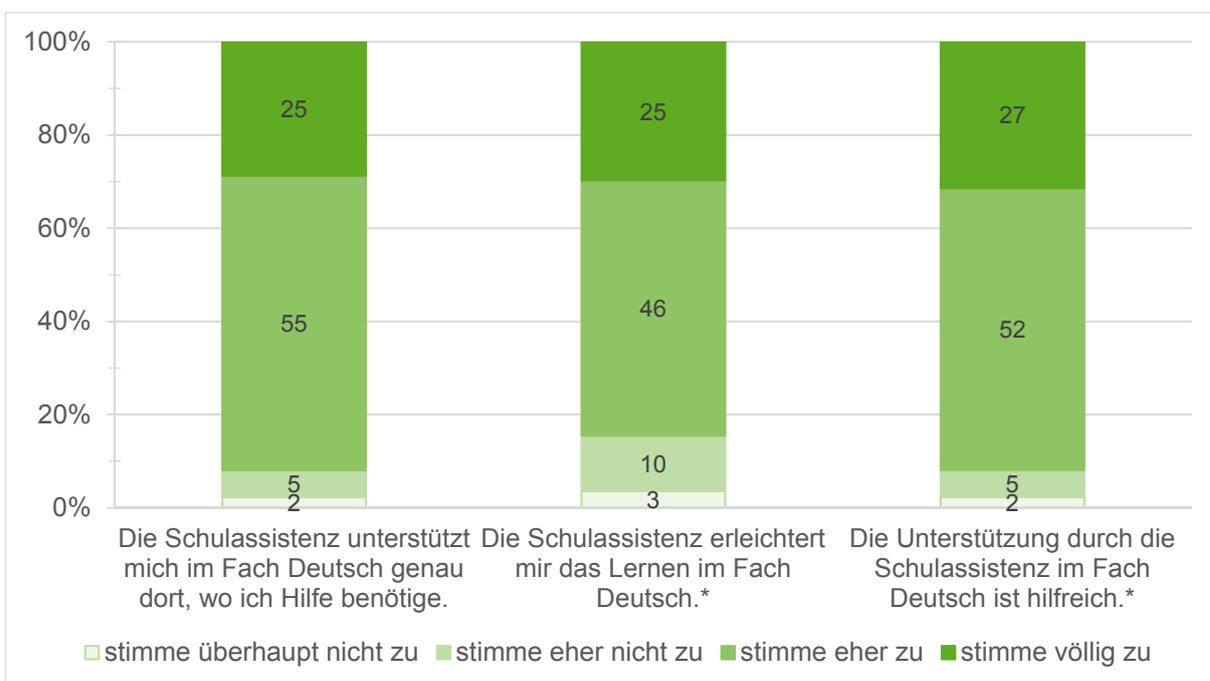
Auch diese Lehrperson betont, dass Schulassistentzen bestimmte Fähigkeiten mitbringen müssen:

»Ich würde einfach Wert darauf legen bei den Schulassistentzen, dass es [...] professionelle Leute sind, weil [...], ich finde auch der ganze Lehrberuf wird abgewertet, in dem Moment, indem ich sage jeder x-beliebige Typ kann da rein und den Lehrer unterstützen und das stimmt einfach nicht.»

Wahrnehmung der Schülerinnen und Schüler

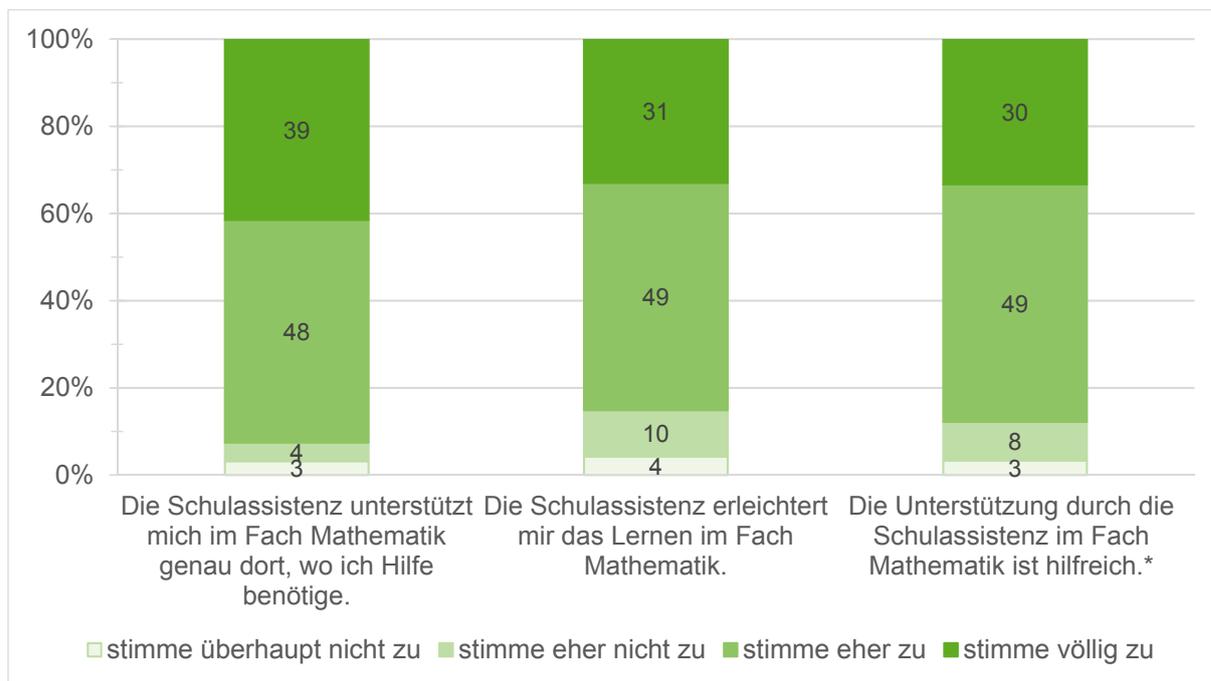
Im Fach Deutsch gaben 18.8 Prozent der Schülerinnen und Schüler (N = 91 [von 485]) an, Unterstützung durch eine Schulassistentz zu erhalten, im Fach Mathematik 20.3 Prozent (N = 97 [von 478]). Die Fragen zur Wahrnehmung dieser Unterstützung wurden nur vorgelegt, wenn diese erste Frage bejaht wurde. Die Einschätzungen zur Unterstützung durch die Schulassistentz fallen in beiden Fächern fast durchweg positiv aus (vgl. Abbildung 23 und Abbildung 24).

Abbildung 23. Wahrgenommene Unterstützung durch die Schulassistentz im Fach Deutsch nach Angaben der Schülerinnen und Schüler



Anmerkungen. N = 87, fehlend: N = 4; *Items von links nach rechts: N = 84, fehlend: N = 7; N = 86, fehlend: N = 5.

Abbildung 24. Wahrgenommene Unterstützung durch die Schullassistenten im Fach Mathematik nach Angaben der Schülerinnen und Schüler



Anmerkungen. $N = 94$, fehlend: $N = 3$; * $N = 90$, fehlend: $N = 7$.

Allen drei Aussagen stimmt ein Grossteil der Schülerinnen und Schüler in beiden Fächern zu, wobei die verhaltenere Antwort «stimme eher zu» jeweils häufiger gewählt wird. Die Unterstützung durch die Schullassistenten wird demnach mehrheitlich als hilfreich, lernförderlich und den Bedürfnissen der Schülerinnen und Schüler angemessen beurteilt.

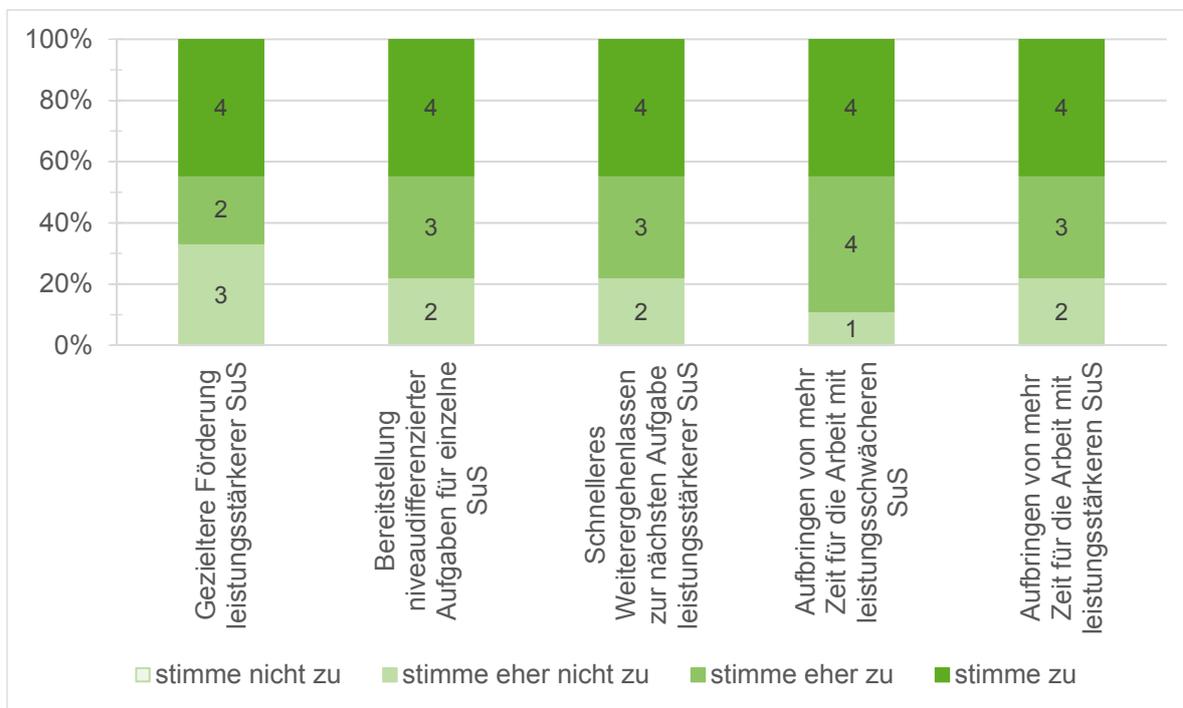
4.2.6 Veränderungen im Unterrichtsgeschehen

Im Folgenden wird der Frage nachgegangen, welche Veränderungen sich im Unterrichtsgeschehen aus Sicht der Lehrpersonen sowie der Schülerinnen und Schüler durch den Beizug der Schullassistenten ergeben haben.

Wahrnehmung der Lehrpersonen

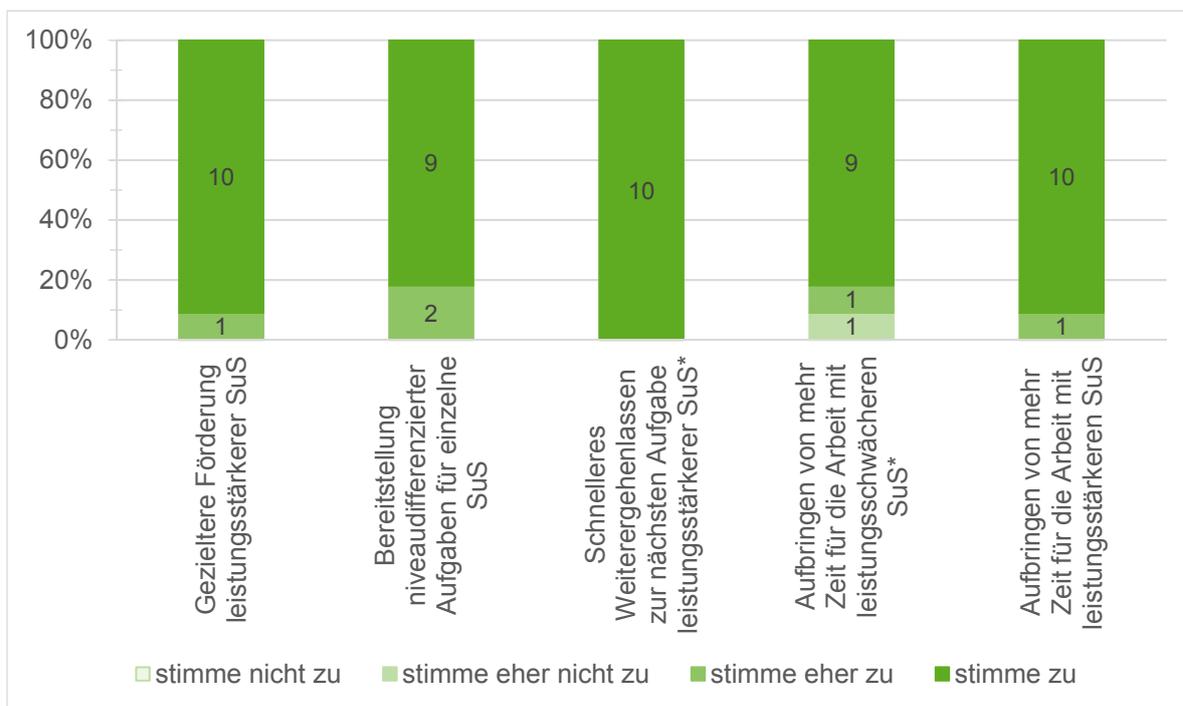
Die Lehrpersonen beantworteten fünf Fragen dazu, inwieweit durch die Schullassistenten bestimmte Massnahmen zur Individualisierung im Unterricht besser genutzt werden konnten. Abbildung 25 und Abbildung 26 zeigen die Ergebnisse für die Fächer Deutsch und Mathematik. In beiden Fächern resultieren überwiegend zustimmende Äusserungen, die sich bei den Deutsch-Lehrpersonen fast gleichmässig auf die Kategorien «stimme eher zu» und «stimme zu» verteilen, während bei den Mathematik-Lehrpersonen die volle Zustimmung dominiert. Die Verteilung der Antworten bei den einzelnen Massnahmen fällt dabei jeweils sehr ähnlich aus.

Abbildung 25. Verbesserte Individualisierung im Unterricht durch den Einsatz der Schullasistenz im Fach Deutsch nach Angaben der Lehrpersonen



Anmerkungen. SuS = Schülerinnen und Schüler. N = 9.

Abbildung 26. Verbesserte Individualisierung im Unterricht durch den Einsatz der Schullasistenz im Fach Mathematik nach Angaben der Lehrpersonen



Anmerkungen. SuS = Schülerinnen und Schüler. N = 11, fehlend: N = 1; *N = 10, fehlend: N = 2.

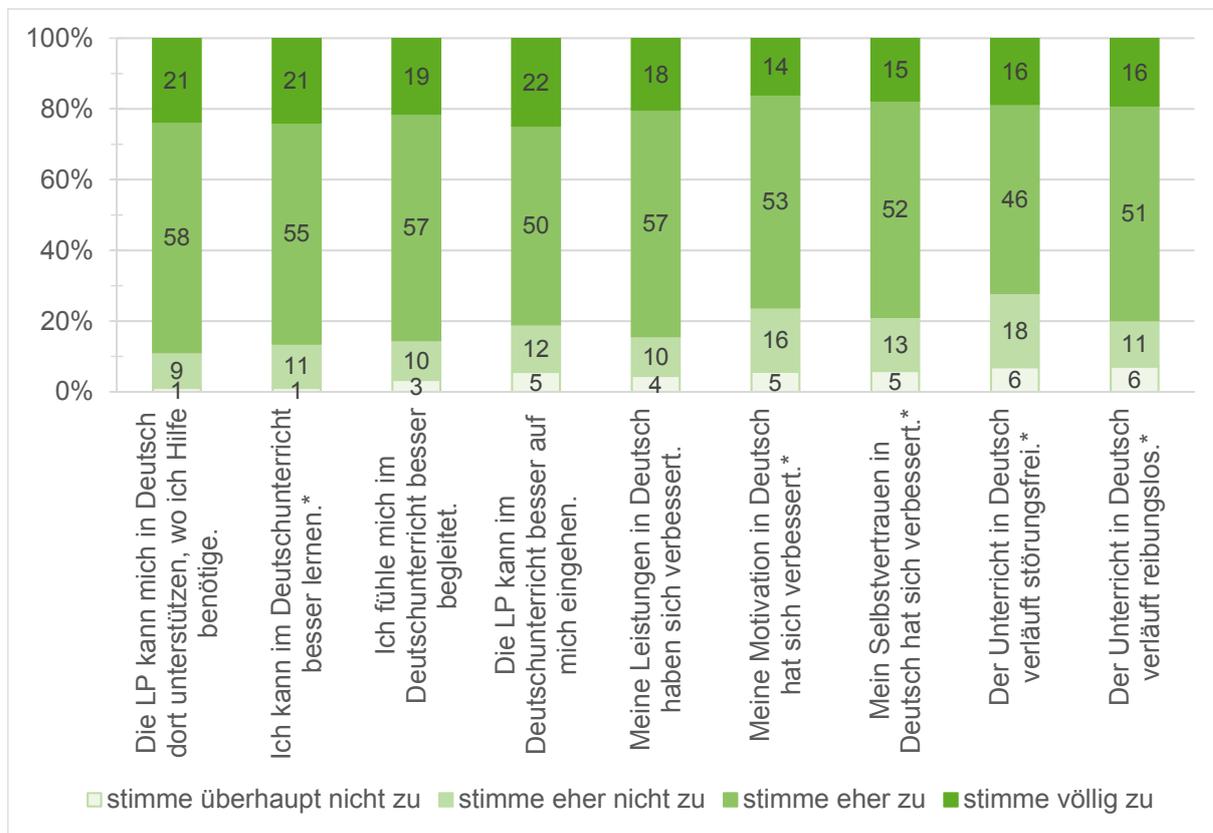
Im Gruppeninterview beschreiben zwei Deutsch-Lehrpersonen die Aufteilung leistungsstärkerer und leistungsschwächerer Schülerinnen und Schüler mit Hilfe der Schulassistentz wie folgt:

- «Ich hatte am Anfang von der ersten Klasse auch eine Klassenassistentz und hatte sie in Deutsch. Ich habe dort [...] darauf geachtet, dass ich mit den Schwächen gearbeitet habe und sie mit den Stärken. Also weil ich auch eine kombinierte Klasse hatte, also alle drei Klassen. Das war eigentlich auch die Idee oder die Weisung von der Schulleitung [...], dass der pädagogische Teil bei uns ist. Das heisst, wir nehmen uns denen an, die mehr Unterstützung brauchen. Und so können wir die Klassenassistentz auch einsetzen zum Schauen bei denen, die vielleicht weniger Unterstützung brauchen, damit es gut läuft. Vielleicht geht es dann auch nur um Material oder was auch immer.»
- «Also im Deutsch mache ich mit den Starken, die schon weiter sind, die haben eigene Aufgaben, ein kleines eigenes Projekt, wo sie mit einer Klassenassistentz arbeiten können. Und hier sind sie sehr motiviert dafür.»

Wahrnehmung der Schülerinnen und Schüler

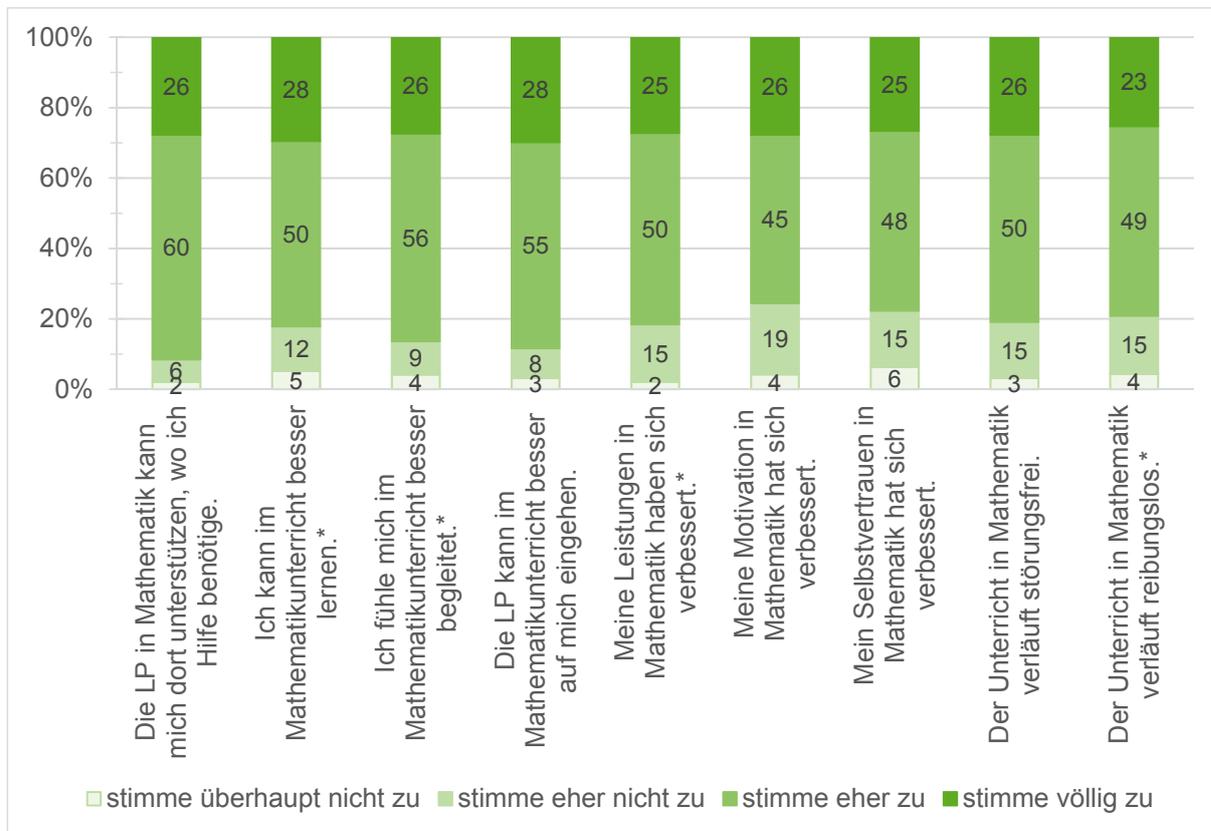
Die Schülerinnen und Schüler wurden zu Veränderungen ihres Deutsch- und Mathematikunterrichts durch den Einsatz einer Schulassistentz befragt (vgl. Abbildung 27 und Abbildung 28).

Abbildung 27. Veränderungen durch die Schulassistentz (Fach Deutsch) nach Schülerangaben



Anmerkungen. LP = Lehrperson. N = 89, fehlend: N = 2; *Items von links nach rechts: N = 88, fehlend: N = 3; N = 88, fehlend: N = 3; N = 85, fehlend: N = 6; N = 86, fehlend: N = 5; *N = 84, fehlend: N = 7.

Abbildung 28. Veränderungen durch die Schulassistentz (Fach Mathematik) nach Schülerangaben



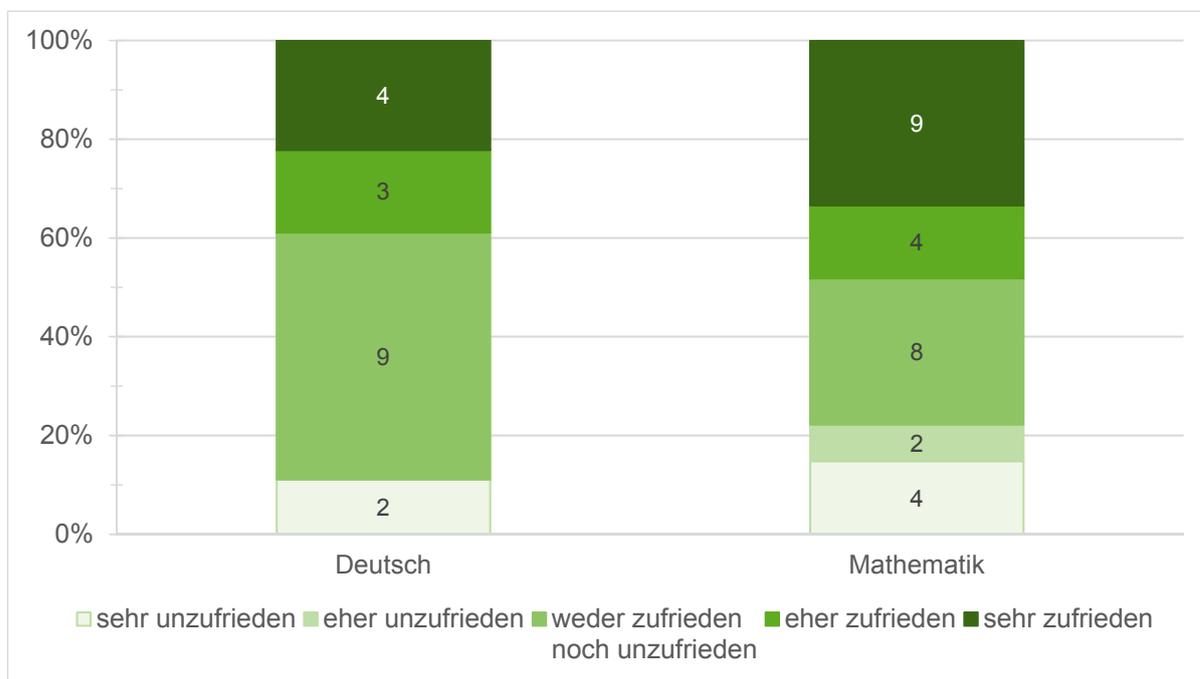
Anmerkungen. LP = Lehrperson. N = 94, fehlend: N = 3; *Items von links nach rechts: N = 95, fehlend: N = 2; N = 95, fehlend: N = 2; N = 92, fehlend: N = 5; N = 91, fehlend: N = 6.

In beiden Fächern stimmen mindestens drei von vier Schülerinnen und Schüler einer positiven Veränderung durch den Einsatz der Schulassistenten (eher) zu, wobei eine volle Zustimmung im Fach Mathematik jeweils häufiger erfolgt als im Fach Deutsch. Am häufigsten zurückgewiesen wird im Fach Deutsch, dass der Unterricht durch die Schulassistentz störungsfrei verlaufe (*stimme überhaupt nicht zu* oder *stimme eher nicht zu*: 27.9 Prozent). Im Fach Mathematik wird am ehesten abgelehnt, dass sich die eigene Motivation durch den Einsatz der Schulassistentz verbessert habe (*stimme überhaupt nicht zu* oder *stimme eher nicht zu*: 24.5 Prozent). Insgesamt kann festgehalten werden, dass die grosse Mehrheit der Schülerinnen und Schüler positive Veränderungen durch den Einsatz der Schulassistenten wahrgenommen hat.

4.2.7 Allgemeine Zufriedenheit mit der Umsetzung der Massnahme

Alle an der Befragung zum Pilotprojekt teilnehmenden Lehrpersonen wurden abschliessend gefragt, wie zufrieden sie ganz allgemein mit der Umsetzung der einzelnen Massnahmen des Pilotprojekts an ihrer Schule sind. Für den Einsatz der Schulassistenten im Unterricht sind die Ergebnisse in Abbildung 29 dargestellt.

Abbildung 29. Allgemeine Zufriedenheit mit der Umsetzung der Massnahme Schulassistenz im Rahmen des Pilotprojekts nach Angaben der Lehrpersonen



Anmerkungen. Deutsch $N = 18$, fehlend: $N = 8$ (befragt: alle Lehrpersonen, die angeben, das Fach Deutsch im Rahmen von ALLE zu unterrichten); Mathematik $N = 27$, fehlend: $N = 5$ (befragt: alle Lehrpersonen, die angeben, das Fach Mathematik im Rahmen von ALLE zu unterrichten).

Fasst man die Kategorien «eher zufrieden» und «sehr zufrieden» zusammen, so zeigen sich etwas mehr als ein Drittel aller am Pilotprojekt beteiligten Deutsch-Lehrpersonen (38.9 Prozent) sowie knapp die Hälfte der Mathematik-Lehrpersonen (48.1 Prozent) mit der Umsetzung des Einsatzes der Schulassistenzen an ihrer Schule zufrieden. Ein Grossteil der Lehrpersonen, insbesondere im Fach Deutsch, äussert sich neutral (*weder zufrieden noch unzufrieden*: Deutsch 50.0 Prozent; Mathematik 29.6). Sehr unzufrieden mit der Umsetzung zu sein, geben zwei der Deutsch-Lehrpersonen (11.1 Prozent) sowie vier der Mathematik-Lehrpersonen (14.8 Prozent) an.

Da viele Lehrpersonen nach eigenen Angaben nicht über eine Schulassistenz verfügten (vgl. Kap. 4.2.3), stellt sich die Frage, wie zufrieden die Lehrpersonen waren, welche die Gelegenheit zur Kooperation mit einer Schulassistenz – zumindest zum Zeitpunkt der Befragung – hatten. Die Verteilung auf die Antwortkategorien zeigt Tabelle 23. Wie sich zeigt, äussert sich die grosse Mehrheit der Lehrpersonen, welche mit einer Schulassistenz arbeiten, eher oder sogar sehr zufrieden mit der Umsetzung der Massnahme an ihrer Schule.

Tabelle 23. Allgemeine Zufriedenheit mit der Umsetzung der Massnahme Schulassistenz bei Lehrpersonen, die eine Schulassistenz hatten

Antwortkategorie	Deutsch	Mathematik
Sehr zufrieden	3	8
Eher zufrieden	3	3
Weder zufrieden noch unzufrieden	1	0
Eher unzufrieden	0	0
Sehr unzufrieden	1	1

Anmerkungen. Deutsch $N = 8$; fehlend: $N = 1$; Mathematik $N = 12$.

Die Schulleitungen wurden ebenfalls um eine allgemeine Einschätzung gebeten und dazu befragt, für wie gelungen sie die Umsetzung der Massnahmen im Pilotprojekt an ihrer Schule halten. Die Ergebnisse für den Einsatz der Schulassistenz sind Tabelle 24 zu entnehmen.

Tabelle 24. Allgemeine Zufriedenheit mit der Umsetzung der Massnahme Schulassistenz bei den Schulleitungen

Antwortkategorie	Deutsch	Mathematik
Sehr gut	2	2
Gut	4	7
Befriedigend	4	0
Schlecht	0	1
Sehr schlecht	0	0

Anmerkungen. $N = 10$; fehlend: $N = 1$.

Die meisten Schulleitungen äussern sich demnach zufrieden mit der Umsetzung der Massnahme an ihrer Schule. Lediglich eine Schulleitung bewertet den Einsatz der Schulassistenzen im Mathematikunterricht als nicht gelungen. Für das Fach Deutsch wählen vier Schulleitungen die moderate Kategorie «befriedigend».

4.3 Aufgabenhilfe in Deutsch und Mathematik

Eine weitere Massnahme im Rahmen des Pilotprojekts ALLE umfasste eine Aufgabenhilfe für die Fächer Deutsch und Mathematik. Diese sollte als verbindliches Lernangebot die individuelle Lernzeit leistungsschwächerer Schülerinnen und Schüler erhöhen und diese bei der Bearbeitung von

Aufgaben in den beiden Fächern unterstützen. In den nachfolgenden Kapiteln wird auf folgende Fragestellungen zur Umsetzung der Aufgabenhilfe eingegangen:

- Wie gut ist die Integration der Aufgabenhilfe in den Stundenplan gelungen? Wie wird der Einsatz der Aufgabenhilfe von der Schulleitung unterstützt? (Kap. 4.3.1)
- In welchem Ausmass und in welchen Einsatzgebieten wird die Aufgabenhilfe in den Fächern Deutsch und Mathematik genutzt? Wer betreut die Aufgabenhilfe? (Kap. 4.3.2)
- Wie kommt die Aufgabenhilfe bei den Lernenden an? Wird die Aufgabenhilfe als Unterstützung wahrgenommen? (Kap. 4.3.3)
- Wie zufrieden sind die Lehrpersonen und die Schulleitungen mit der Umsetzung der Massnahme Aufgabenhilfe an ihrer Schule? (Kap. 4.3.4)

4.3.1 Integration der Aufgabenhilfe und Unterstützung durch die Schulleitung

Integration der Aufgabenhilfe in den Stundenplan

Die Schulleitungen der ALLE-Schulen wurden gefragt, wie gut an ihren Schulen die Integration der Aufgabenhilfe in den Stundenplan gelungen ist. Die Antworten fallen zwar vergleichsweise heterogen, aber überwiegend positiv aus. So bezeichnen acht der elf befragten Schulleitungen (72.8 Prozent) die Integration der Aufgabenhilfe als «gut» oder «sehr gut», die restlichen drei Schulleitungen bewerten diese als «befriedigend» (27.3 Prozent). Wie die nachfolgenden Zitate aus den Einzelinterviews mit den Schulleitungen illustrieren, erfolgte die Integration der Aufgabenhilfe in den Stundenplan je nach Schule in unterschiedlicher Art und Weise.

Eine Schulleitung schildert, wie die Aufgabenhilfe an ihrer Schule an vier Wochentagen jeweils mittags eingesetzt wurde:

«Das ist einfach vier Tage pro Woche jeweils über den Mittag und dort kommen sie, schreiben sich jeweils ein und arbeiten an den Hausaufgaben, zumindest da, wo ich drin war, da habe ich erlebt, die haben gearbeitet. Teilweise habe ich sie dann ein wenig gecoach.»

Eine weitere Schulleitung berichtet von der täglichen Durchführung einer Aufgabenhilfe in Form einer Zusatzstunde:

«[...] wir haben im Stundenplan jeden Tag eine Stunde für die Aufgabenhilfe eingetragen. Wir nennen das aber Lernstunde.»

Eine dritte Schulleitung beschreibt die Integration in das bestehende Gefäss eines Förderzentrums:

«Aufgabenhilfe, die haben wir institutionalisiert im Förderzentrum. Das läuft alles über das Förderzentrum. Und zwar gibt es fixe Stunden, aber die Schülerinnen und Schüler können irgendwann ins Förderzentrum, um Aufgaben zu machen.»

Unterstützung durch die Schulleitung

Die Schulleitungen wurden dazu befragt, inwieweit sie vier verschiedene Möglichkeiten nutzen, um den Einsatz der Aufgabenhilfe an ihrer Schule zu ermöglichen. Die Ergebnisse zeigt Tabelle 25.

Tabelle 25. Nutzung von Möglichkeiten zur Unterstützung des Einsatzes der Aufgabenhilfe durch die Schulleitung nach eigenen Angaben

Unterstützungsmöglichkeiten	Anzahl	in Prozent
Organisation von qualifiziertem Lehrpersonal in der Aufgabenhilfe	8	88.9
Betonung der Wichtigkeit der Aufgabenhilfe gegenüber den Lehrpersonen und den Schülerinnen und Schülern	8	88.9
Besuche der Aufgabenhilfe	5	55.6
Organisation des Austausches der Lehrpersonen untereinander zum erfolgreichen Einsatz der Aufgabenhilfe	3	33.3

Anmerkungen. $N = 9$, fehlend: $N = 2$. Antwortformat pro Item: ja – nein.

In erster Linie schaffen die Schulleitungen die erforderlichen Rahmenbedingungen für den Einsatz der Aufgabenhilfe, indem sie qualifiziertes Lehrpersonal in der Aufgabenhilfe einsetzen und die Wichtigkeit der Aufgabenhilfe betonen. Besuche der Aufgabenhilfe durch die Schulleitung oder ein von der Schulleitung initiiertes Austausch der Lehrpersonen finden demgegenüber seltener statt.

4.3.2 Verfügbarkeit, Einsatz und Durchführung der Aufgabenhilfe

Verfügbarkeit der Aufgabenhilfe an den Schulen

Für beide Fächer gibt nur etwa die Hälfte der Lehrpersonen an, dass an ihrer Schule zum Zeitpunkt der Befragung für ihre Schülerinnen und Schüler das Angebot der Aufgabenhilfe bestehe (Deutsch: 57.7 Prozent, $N = 15$ [von 26]; Mathematik: 46.9 Prozent, $N = 15$ [von 32]). Betrachtet man diese Angaben schulweise, so wird deutlich, dass sich die Angaben der Lehrpersonen einer Schule voneinander unterscheiden können (vgl. Tabelle 26). Beispielsweise berichten zwei der vier befragten Deutsch-Lehrpersonen in Schule D, dass eine Aufgabenhilfe zur Verfügung stehe, die beiden anderen Lehrpersonen weisen dies zurück. An zwei Schulen gibt keine der befragten Lehrpersonen an, dass für ihre Schülerinnen und Schüler eine Aufgabenhilfe angeboten wird (vgl. Schulen B und I), in einer weiteren Schule (vgl. Schule H) berichtet dies keine der Mathematik-Lehrpersonen.

Tabelle 26. Verfügbarkeit einer Aufgabenhilfe in den einzelnen Schulen für die Schülerinnen und Schüler im Pilotprojekt nach Angaben der Lehrpersonen

Schule	Aufgabenhilfe Deutsch				Aufgabenhilfe Mathematik			
	Nein	Ja	Gesamt	Ja (in %)	Nein	Ja	Gesamt	Ja (in %)
F	0	4	4	100.0	0	5	5	100.0
G	0	2	2	100.0	2	1	3	33.3
C	0	3	3	100.0	0	4	4	100.0
H	0	1	1	100.0	2	0	2	0.0
D	2	2	4	50.0	1	2	3	66.7
E	4	2	6	33.3	5	3	8	37.5
A	2	1	3	33.3	3	0	3	0.0
I	2	0	2	0.0	3	0	3	0.0
B	1	0	1	0.0	1	0	1	0.0
Gesamt	11	15	26	57.7	17	15	32	46.9

Anmerkungen. Deutsch $N = 26$; Mathematik $N = 32$. Absteigend sortiert nach der Verfügbarkeit im Fach Deutsch («Ja» in Prozent).

Die nachfolgenden Ergebnisdarstellungen beziehen sich auf die Angaben derjenigen Lehrpersonen, welche berichtet hatten, dass an ihrer Schule eine Aufgabenhilfe für ihre Schülerinnen und Schüler angeboten werde (pro Fach jeweils $N = 15$). Für mögliche Erklärungen zur geringen berichteten Häufigkeit einer Aufgabenhilfe und den uneinheitlichen Angaben der Lehrpersonen innerhalb der Schulen siehe Kapitel 5.2.2.

Ausmass der Nutzung der Aufgabenhilfe

Von den Lehrpersonen, die von einer Aufgabenhilfe an ihrer Schule berichten, beschreiben etwas mehr als die Hälfte diese als freiwilliges, nicht verpflichtendes Angebot (je Fach: 60.0 Prozent, $N = 9$ [von 15]). Für das Fach Deutsch gibt ein Drittel Lehrpersonen an, dass die Aufgabenhilfe für *ausgewählte* Schülerinnen und Schüler verpflichtend sei (33.3 Prozent, $N = 5$). Eine einzelne Deutsch-Lehrperson (6.7 Prozent) bezeichnet das Angebot der Aufgabenhilfe als für alle Schülerinnen und Schüler verpflichtend. Für das Fach Mathematik geben sechs Lehrpersonen (40.0 Prozent) an, die Aufgabenhilfe sei für ausgewählte Schülerinnen und Schüler verpflichtend.

Betrachtet man die tatsächliche Nutzung der Aufgabenhilfe, das heisst die von den Lehrpersonen berichtete Anzahl ihrer Schülerinnen und Schüler im Pilotprojekt, welche die Aufgabenhilfe besuchen, ergibt sich je nach Fach ein etwas anderes Bild. So wird die Aufgabenhilfe nach Angabe der

Lehrpersonen im Fach Deutsch pro Woche von maximal 13 der von ihnen unterrichteten Schülerinnen und Schülern besucht, im Mittel etwa zwei Schülerinnen und Schüler ($M = 2.1$, $SD = 3.5$). Im Fach Mathematik besuchen maximal drei Schülerinnen und Schüler die Aufgabenhilfe, im Mittel rund eine Schülerin/ein Schüler ($M = 0.9$, $SD = 1.1$). Jeweils etwa die Hälfte der Lehrpersonen beider Fächer gibt an (Deutsch: 50.0 Prozent, $N = 7$ [von 14]; Mathematik: 53.3 Prozent, $N = 8$ [von 15]), dass keine Schülerin und kein Schüler die Aufgabenhilfe besucht.

Für die Lehrpersonen, welche eine Nutzung der Aufgabenhilfe durch ihre Schülerinnen und Schüler berichten, ergeben sich die folgenden prozentualen Anteile an Schülerinnen und Schülern mit Aufgabenhilfe im Verhältnis zu allen von der jeweiligen Lehrperson unterrichteten Schülerinnen und Schülern (in Prozent; Vorkommen jeweils einmal):

- Deutsch: 100.0 – 60.0 – 10.5 – 9.8 – 8.3 – 5.4 – 4.3
- Mathematik: 33.3 – 21.4 – 20.0 – 8.7 – 5.9 – 4.3 – 2.2

In den meisten Fällen besuchte also ein eher kleiner Anteil der von einer Lehrperson unterrichteten Schülerinnen und Schüler die Aufgabenhilfe. Lediglich zwei Deutsch-Lehrpersonen geben an, dass mehr als die Hälfte ihrer Schülerschaft (100.0 bzw. 60.0 Prozent) das Angebot wahrnahm.

Die Lehrpersonen wurden zusätzlich gefragt, wie viele Lektionen Aufgabenhilfe ihren Schülerinnen und Schülern pro Woche zur Verfügung stehen, und wie viele Lektionen Aufgabenhilfe die Schülerinnen und Schüler pro Woche konkret besuchen. Nach Angaben der Lehrpersonen beider Fächer können pro Woche bis zu acht Lektionen Aufgabenhilfe in Anspruch genommen werden, im Durchschnitt drei bis vier (Deutsch: $M = 3.4$, $SD = 2.5$; Mathematik: $M = 4.3$, $SD = 1.8$). Besucht werden davon in Deutsch maximal drei, in Mathematik bis zu fünf Lektionen pro Schülerin oder Schüler, im Mittel in beiden Fächern rund eine Lektion pro Schülerin oder Schüler (Deutsch: $M = 1.1$, $SD = 1.2$; Mathematik: $M = 0.9$, $SD = 1.4$). Auffallend ist, dass – vor allem im Fach Mathematik – Lehrpersonen angeben, dass für ihre Lernenden Lektionen in der Aufgabenhilfe zur Verfügung stehen, gleichzeitig aber berichten, dass keine Lektionen von ihren Schülerinnen und Schülern besucht werden (Deutsch: 22.2 Prozent, $N = 2$ [von 9]; Mathematik 53.3 Prozent, $N = 8$ [von 15]).

Eine Schwierigkeit bei der Nutzung der Aufgabenhilfe – die Sicherstellung des tatsächlichen Besuchs durch die Lernenden – beschreibt eine Lehrperson im Gruppeninterview wie folgt:

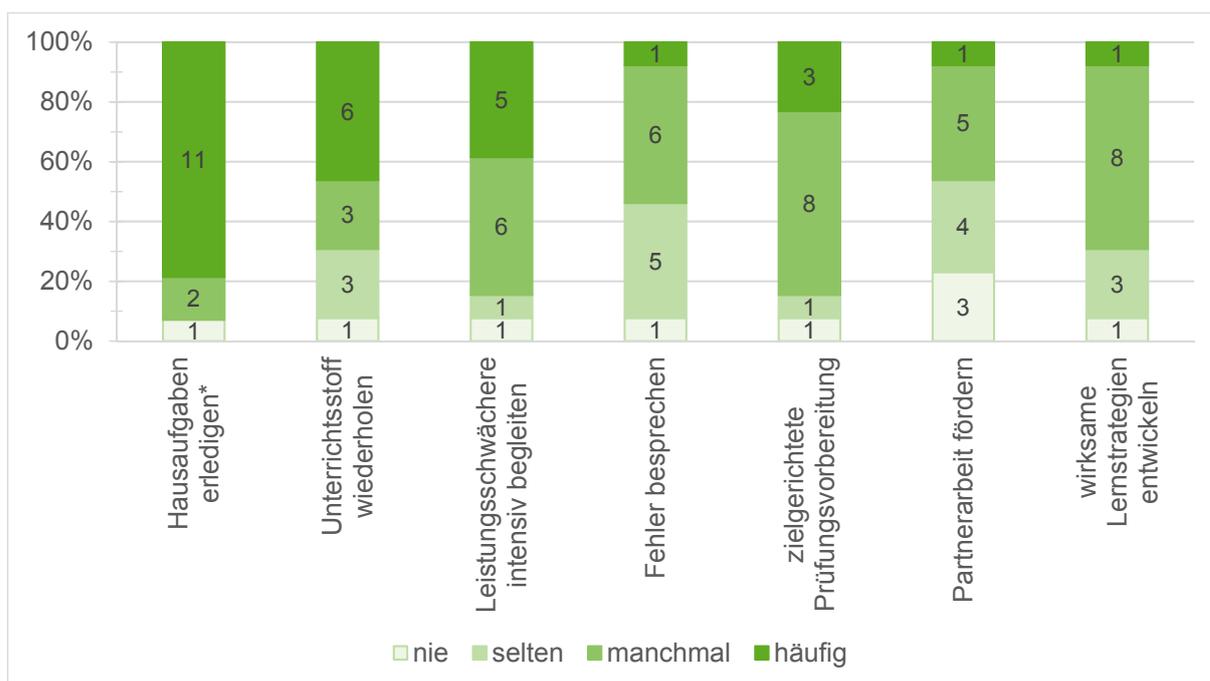
«Genau, ich habe das auch auf dem Lerncoaching mit den Schülern vereinbart, dass sie in die Hausaufgabenbetreuung gehen. Aber bei demotivierten, schwächeren Schülern ist das dann versandet. Die motivierten, schwachen Schüler sind dann schon gegangen, es waren einfach wenig. Und dann auch nicht so, wie man es vereinbart hat, weil es eben auf freiwilliger Basis ist, war das noch schwierig. Das kann ich auch nicht kontrollieren, eine Sanktion gibt es dann ja auch nicht wirklich. Ja, ich war im Zwiespalt, wie gut ich das jetzt finde.»

Die Äusserungen von zwei weiteren Lehrpersonen weisen in eine ähnliche Richtung:
 [Lehrperson A]: «Also bei uns im C, wir sind beide C-Lehrer, [Lehrperson B] und ich, haben verschiedene Niveaus, ist es ganz interessant, weil die Schüler nehmen das Angebot so gut wie gar nicht wahr, weil es freiwillig ist und das Engagement nicht so hoch ist.» [Lehrperson B]: «[...] Es soll ja keine Strafe sein auf der einen Seite, und der C-Schüler nimmt das eben nicht als Belohnung wahr, dass er jetzt am Mittag kommen soll und Hausaufgaben machen soll.»

Einsatzgebiete der Aufgabenhilfe

Die Lehrpersonen wurden weiterhin gefragt, wie häufig die Aufgabenhilfe bezogen auf sieben Aufgaben- oder Tätigkeitsbereiche in Anspruch genommen wird. Die Ergebnisse sind Abbildung 30 und Abbildung 31 zu entnehmen.

Abbildung 30. Einsatzgebiete der Aufgabenhilfe für das Fach Deutsch nach Angaben der Lehrpersonen

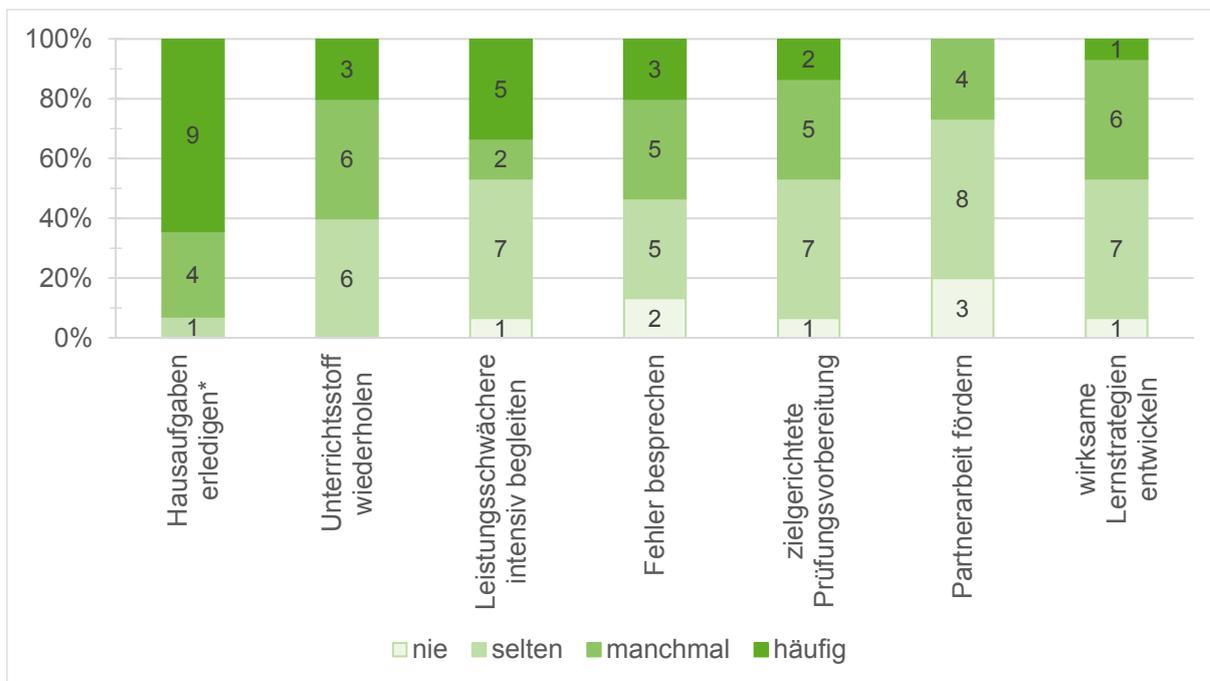


Anmerkungen. N = 13, fehlend: N = 2; *N = 14, fehlend: N = 1.

Ein wichtiger Schwerpunkt des Einsatzes der Aufgabenhilfe in beiden Fächern ist die Unterstützung bei der Erledigung von Hausaufgaben (*häufig*: Deutsch 78.6 Prozent; Mathematik 64.3 Prozent). Im Fach Deutsch sind darüber hinaus der Einsatz für die Wiederholung von Unterrichtsstoff (*häufig*: 46.2 Prozent) und für die intensive Begleitung von leistungsschwächeren Schülerinnen und Schülern (*häufig*: 38.5 Prozent) vergleichsweise häufig. Alle weiteren Tätigkeiten kommen nach Auskunft der Deutsch-Lehrpersonen zumindest gelegentlich vor. Für das Fach Mathematik

werden insgesamt geringere Häufigkeiten berichtet, aber auch hier kommen nach Angaben der meisten Lehrpersonen alle Tätigkeiten zumindest «selten» oder «manchmal» vor.

Abbildung 31. Einsatzgebiete der Aufgabenhilfe für das Fach Mathematik nach Angaben der Lehrpersonen



Anmerkungen. $N = 13$, fehlend: $N = 2$; * $N = 14$, fehlend: $N = 1$.

Durchführung der Aufgabenhilfe

Danach gefragt, wer *in der Regel* die Aufgabenhilfe durchführt, äussern sich die Deutsch-Lehrpersonen ($N = 15$) wie folgt:

- andere Lehrperson (60.0 Prozent, $N = 9$),
- Lehrperson für schulische Heilpädagogik (SHP) (26.7 Prozent, $N = 4$),
- andere Person (ohne genauere Bezeichnung; 6.7 Prozent, $N = 1$),
- die befragte Lehrperson selbst (6.7 Prozent, $N = 1$).

Ein ähnliches Bild zeigt sich bei der Befragung der Mathematik-Lehrpersonen ($N = 15$):

- andere Lehrperson (53.3 Prozent, $N = 8$),
- Lehrperson für schulische Heilpädagogik (SHP) (33.3 Prozent, $N = 5$),
- andere Person (ohne genauere Bezeichnung; 13.3 Prozent, $N = 2$).

In beiden Fächern sind somit am häufigsten andere Lehrpersonen (als die befragte Lehrperson selbst) sowie die schulische Heilpädagogik in die Betreuung der Aufgabenhilfe involviert. Betrachtet man die Angaben der Lehrpersonen separat für die einzelnen Schulen, ergeben sich Hinweise

auf die schulspezifische Ausgestaltung der Betreuung. Tabelle 27 und Tabelle 28 berichten für die Fächer Deutsch und Mathematik, welche Personengruppen nach den Angaben der Lehrpersonen in den einzelnen Schulen an der Aufgabenhilfe mitwirken. Dargestellt ist, wie viele Lehrpersonen in der jeweiligen Schule angeben, dass die betreffende Personengruppe (a) in der Regel die Aufgabenhilfe betreut beziehungsweise (b) darüber hinaus an der Durchführung beteiligt ist.

Tabelle 27. Durchführung der Aufgabenhilfe in den einzelnen Schulen nach Angaben der Deutsch-Lehrpersonen

Schule	In der Regel durch ...			Weitere Durchführende				Beteiligte
	LP*	SHP	andere	LP*	SHP	SA	Peers	
H	1							LP
A	1				1			LP, SHP
F	4			4	1			LP, SHP
C		3		1	2			LP, SHP
G	2					1		LP, SA
D	1	1			1	1		LP, SHP, SA
E	1		1	1			1	LP, Peers, andere

Anmerkungen. LP = Lehrperson, SHP = schulische Heilpädagogin/schulischer Heilpädagoge, SA = Schulassistenz. N = 15. Die Spalte «Beteiligte» listet alle Personengruppen auf, die in der Schule an der Durchführung der Aufgabenhilfe beteiligt sind. Die fett markierten Gruppen sind schwerpunktmässig («in der Regel») beteiligt. *die befragte Lehrperson selbst oder eine andere Lehrperson

Tabelle 28. Durchführung der Aufgabenhilfe in den einzelnen Schulen nach Angaben der Mathematik-Lehrpersonen

Schule	In der Regel durch ...			Weitere Durchführende				Beteiligte
	LP*	SHP	andere	LP*	SHP	SA	Peers	
F	5			4	2			LP, SHP
C		4		2	2			LP, SHP
D	1	1		1	1			LP, SHP
G	1					1		LP, SA
E	1		2					LP, andere

Anmerkungen. LP = Lehrperson, SHP = schulische Heilpädagogin/schulischer Heilpädagoge, SA = Schulassistenz. N = 15. Die Spalte «Beteiligte» listet alle Personengruppen auf, die in der Schule an der Durchführung der Aufgabenhilfe beteiligt sind. Die fett markierten Gruppen sind schwerpunktmässig («in der Regel») beteiligt. *die befragte Lehrperson selbst oder eine andere Lehrperson

In beiden Fächern deutet sich an, dass in der Mehrzahl der Schulen die Lehrpersonen die zentrale Rolle bei der Durchführung der Aufgabenhilfe eingenommen haben. Von den weiteren Personengruppen tragen insbesondere die schulischen Heilpädagoginnen und Heilpädagogen zur Betreuung der Aufgabenhilfe bei. In einer der Schulen sind sie in der Regel für die Durchführung der Aufgabenhilfe zuständig (Schule C). Bei der Interpretation ist zu beachten, dass nicht aus allen neun ALLE-Schulen Angaben von Lehrpersonen vorliegen. Die vorhandenen Angaben pro Schule stammen zudem oft nur von ein bis zwei Lehrpersonen.

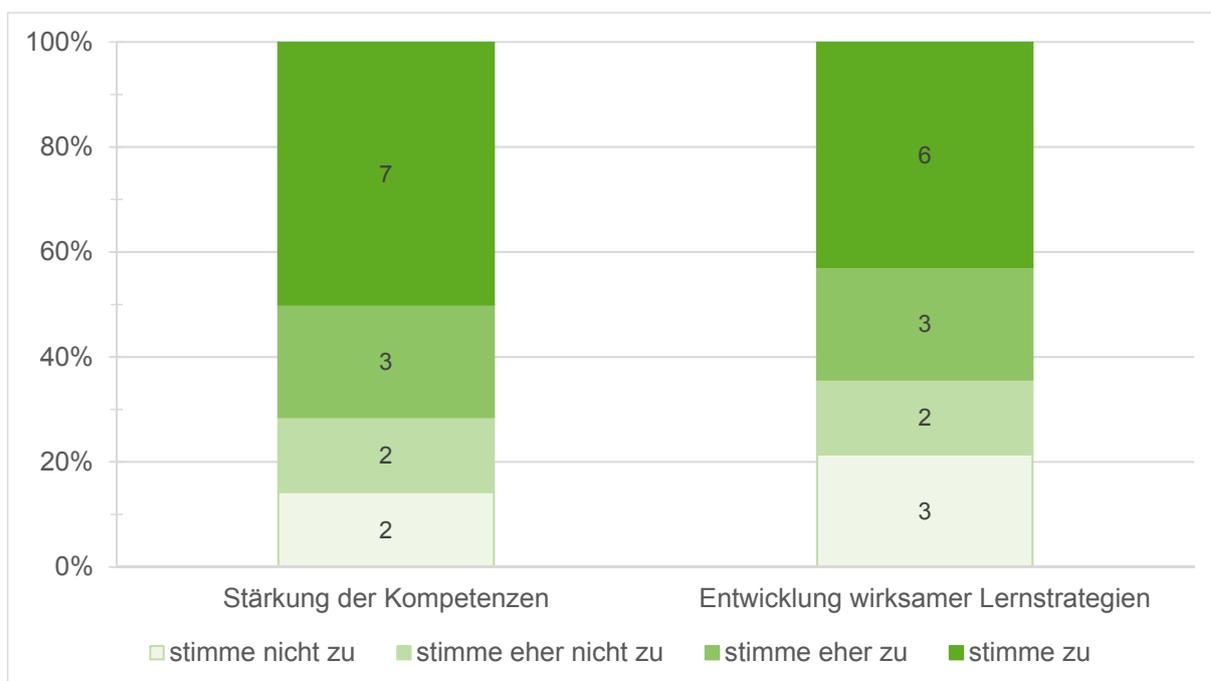
4.3.3 Wahrgenommene Unterstützung durch die Aufgabenhilfe

Sowohl Lehrpersonen als auch Schülerinnen und Schüler wurden dazu befragt, inwieweit sie die Aufgabenhilfe als Unterstützung des Lernens wahrnehmen. Es folgen zunächst die Ergebnisse für die Lehrpersonen, daran anschliessend die Ergebnisse für die Schülerinnen und Schüler.

Wahrnehmung der Lehrpersonen

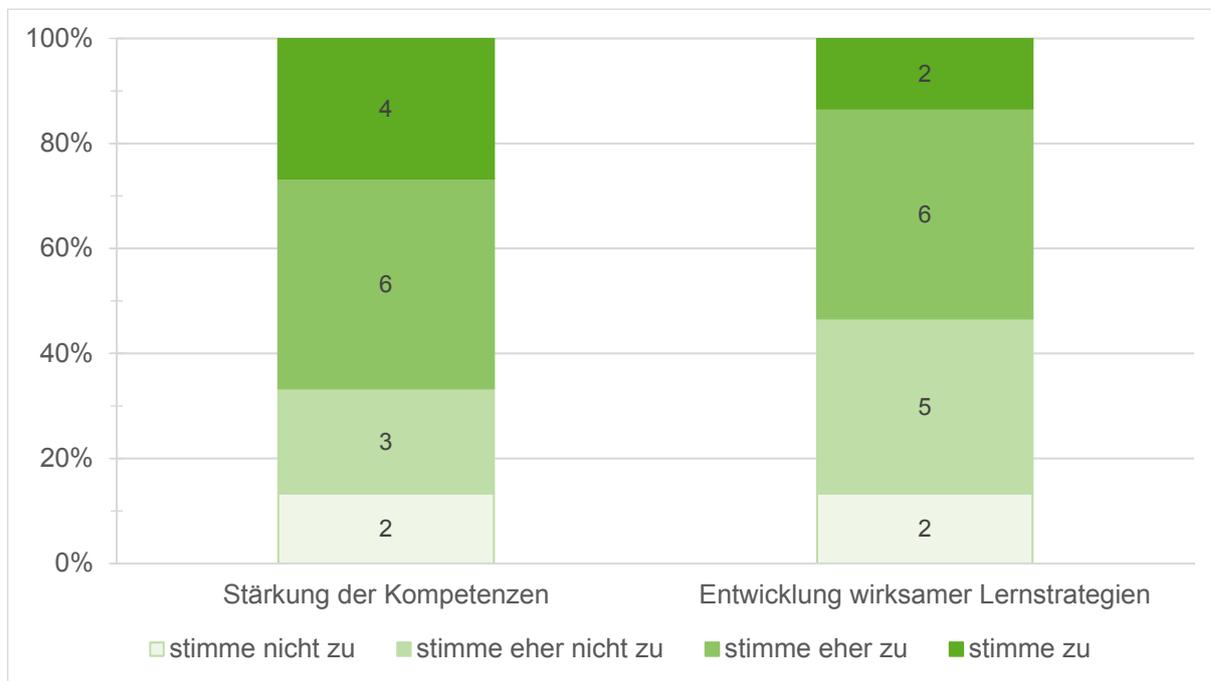
Lehrpersonen, die angaben, dass von ihnen unterrichtete Schülerinnen und Schüler an der Aufgabenhilfe teilnehmen, wurden dazu befragt, inwieweit sie die Aufgabenhilfe als Unterstützung für das Lernen im Fach wahrnehmen (vgl. Abbildung 32 und Abbildung 33).

Abbildung 32. Wahrnehmung der Aufgabenhilfe als lernförderliche Unterstützung im Fach Deutsch nach Angaben der Lehrpersonen



Anmerkungen. N = 14, fehlend: N = 1.

Abbildung 33. Wahrnehmung der Aufgabenhilfe als lernförderliche Unterstützung im Fach Mathematik nach Angaben der Lehrpersonen



Anmerkungen. N = 15.

Die Lehrpersonen beider Fächern stimmen zu einem Grossteil (eher) zu, dass der Einsatz der Aufgabenhilfe zu einer Stärkung der Kompetenzen der teilnehmenden Schülerinnen und Schüler geführt habe (*stimme eher zu* oder *stimme zu*: Deutsch 71.4 Prozent; Mathematik 66.7 Prozent). Weniger Lehrpersonen – aber noch immer die Mehrheit – äussern, dass die Aufgabenhilfe zu einer Entwicklung wirksamer Lernstrategien beigetragen habe (*stimme eher zu* oder *stimme zu*: Deutsch: 64.3 Prozent; Mathematik 53.3 Prozent).

Wie die Aufgabenhilfe zu einer Entwicklung von Lernstrategien beitragen kann, beschreibt eine Lehrperson in den Gruppeninterviews wie folgt:

«Eindeutig. Das ist jetzt mehrmals erwähnt worden. Zum Beispiel Strategien, wie kann man eine Fremdsprache lernen, wie kann ich mich für eine Matheprüfung vorbereiten oder wie kann ich mir einen Stoff aneignen, welchen ich nicht so gut verstehe. Das ist auch wieder subjektiv, aber ich habe den Eindruck, dass es Diverse hat in meiner Matheklasse, die enorm davon profitieren, weil sie Sicherheit bekommen und sich ein Zutrauen aneignen, weil sie begleitet sind von einer Lehrperson.»

Wahrnehmung der Schülerinnen und Schüler

Insgesamt gibt nur ein Bruchteil der Schülerinnen und Schüler an, die Aufgabenhilfe Deutsch (2.8 Prozent, $N = 18$ [von 651]) beziehungsweise Mathematik (2.2 Prozent, $N = 14$ [von 631]) zu besuchen (für mögliche Erklärungen siehe Kap. 5.2.2). Diese Schülerinnen und Schüler verteilen sich wie in Tabelle 29 dargestellt auf sechs der neun ALLE-Schulen. Für drei der Schulen berichtet keine Schülerin/kein Schüler von einem Besuch der Aufgabenhilfe. Die nachfolgenden Analysen basieren ausschliesslich auf den Angaben der Schülerinnen und Schüler, welche angeben, die Aufgabenhilfe zu besuchen.

Tabelle 29. Schülerinnen und Schüler mit Aufgabenhilfe nach Fach und Schule nach eigenen Angaben

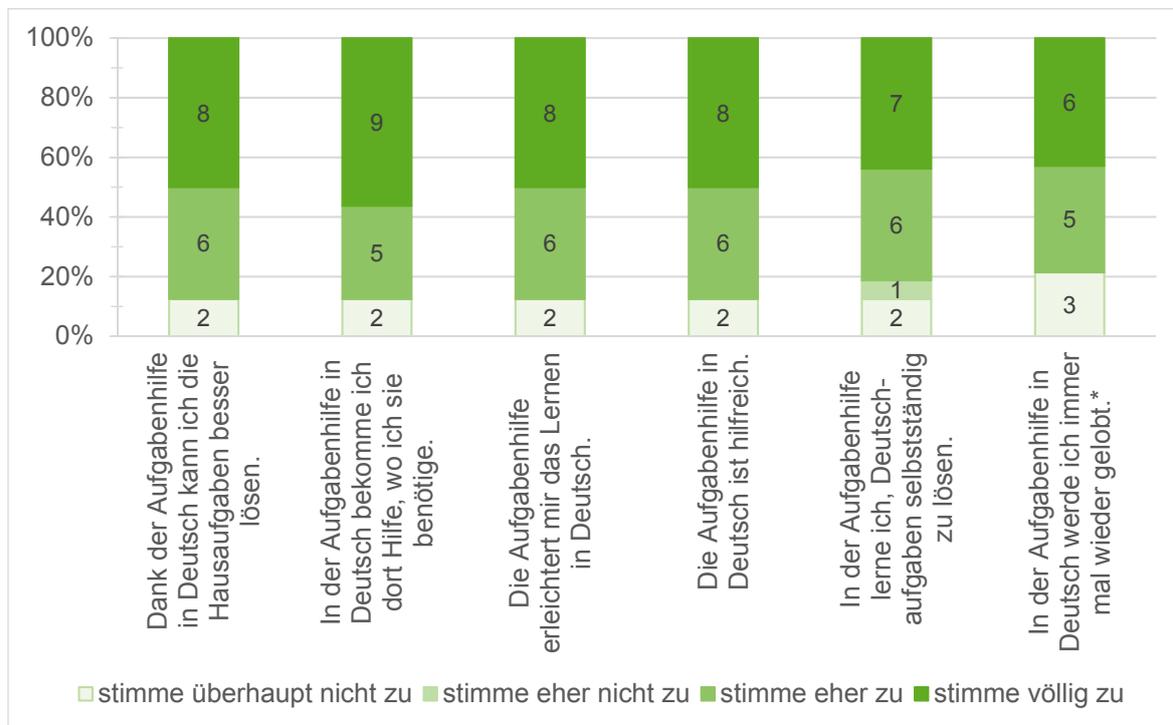
Schule	Deutsch	Mathematik
E	2 (1.7%)	4 (3.4%)
A	3 (7.5%)	1 (2.4%)
C	2 (1.5%)	5 (4.0%)
I	3 (5.0%)	1 (1.6%)
B	3 (6.0%)	2 (3.8%)
D	5 (4.9%)	1 (0.9%)
Gesamt	18	14

Anmerkungen. Deutsch $N = 18$; Mathematik: $N = 14$. Prozentangabe: Anteil der Schülerinnen und Schüler mit Aufgabenhilfe an allen befragten Schülerinnen und Schüler der jeweiligen Schule.

Wie aus Abbildung 34 und Abbildung 35 ersichtlich ist, wird die Aufgabenhilfe überwiegend positiv wahrgenommen. So stimmen in beiden Fächern jeweils mehr als drei Viertel der Schülerinnen und Schüler den erfragten Aspekten «eher» oder «völlig» zu. Die positive Bewertung fällt dabei für die Aufgabenhilfe Deutsch insgesamt etwas schwächer aus als für die Aufgabenhilfe Mathematik.

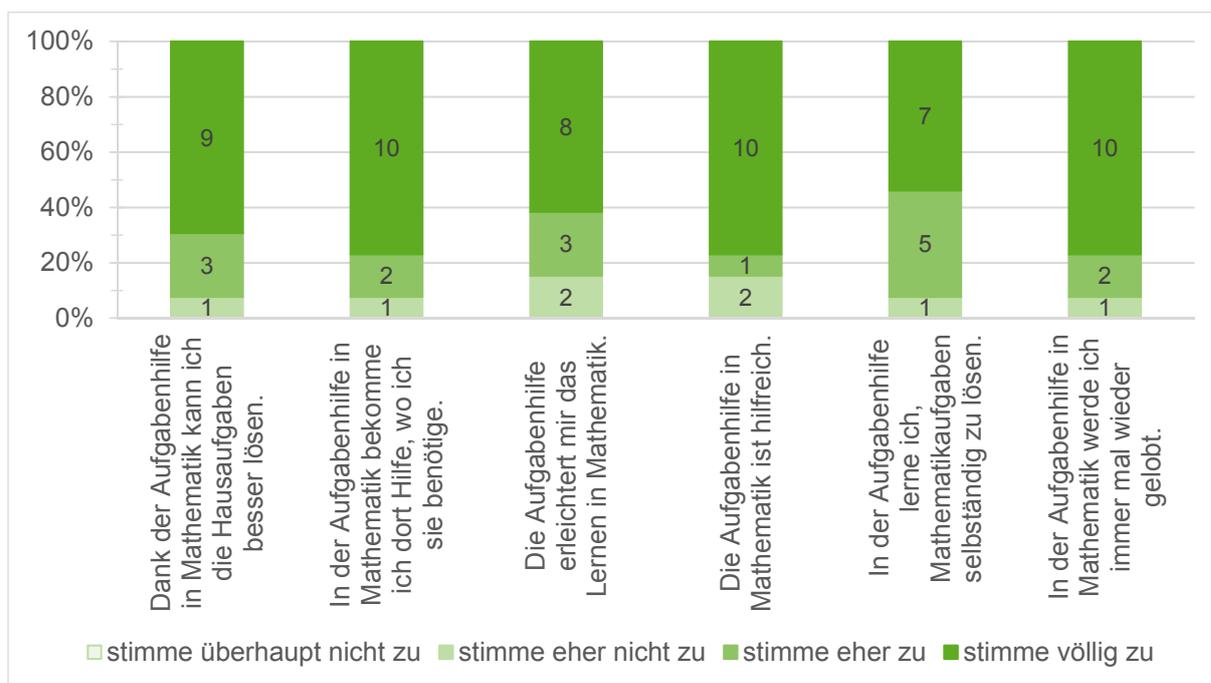
Nach positiven Veränderungen bezogen auf den Unterricht gefragt, welche die Aufgabenhilfe bewirkt hat, findet sich ebenfalls für beide Fächer eine breite Zustimmung bei den Schülerinnen und Schülern (vgl. Abbildung 36 und Abbildung 37, S. 114).

Abbildung 34. Wahrgenommene Unterstützung durch die Aufgabenhilfe im Fach Deutsch nach Angaben der Schülerinnen und Schüler



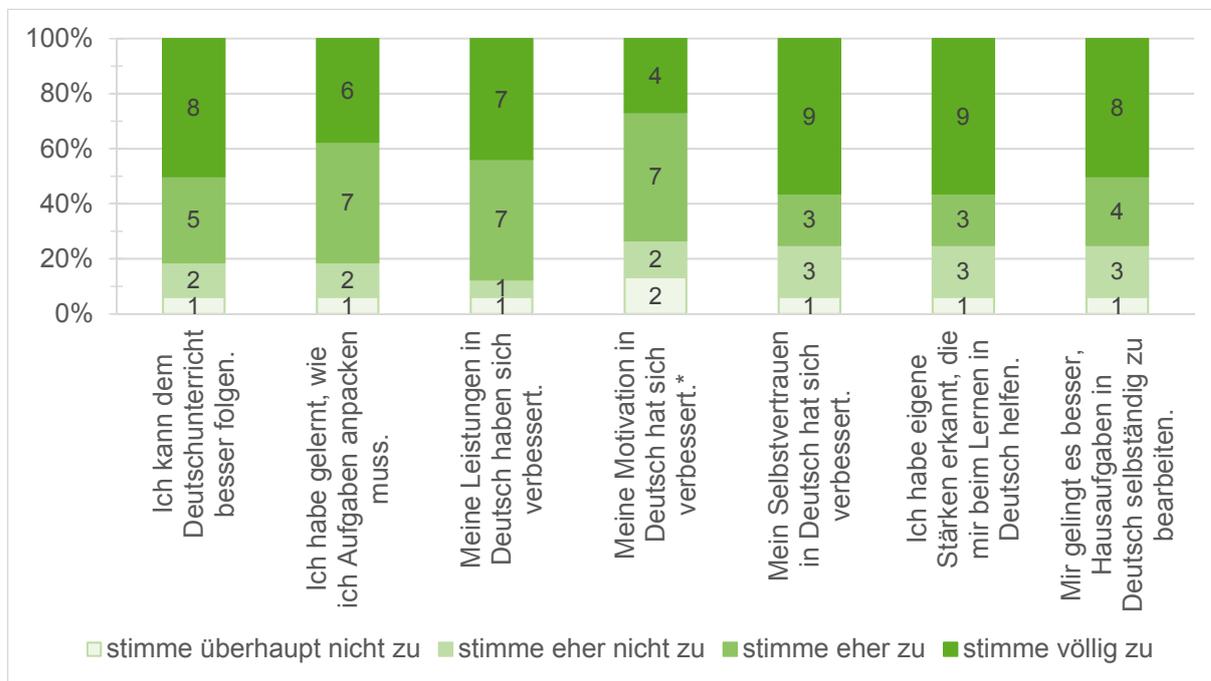
Anmerkungen. N = 16, fehlend: N = 2; *N = 14, fehlend: N = 4.

Abbildung 35. Wahrgenommene Unterstützung durch die Aufgabenhilfe im Fach Mathematik nach Angaben der Schülerinnen und Schüler



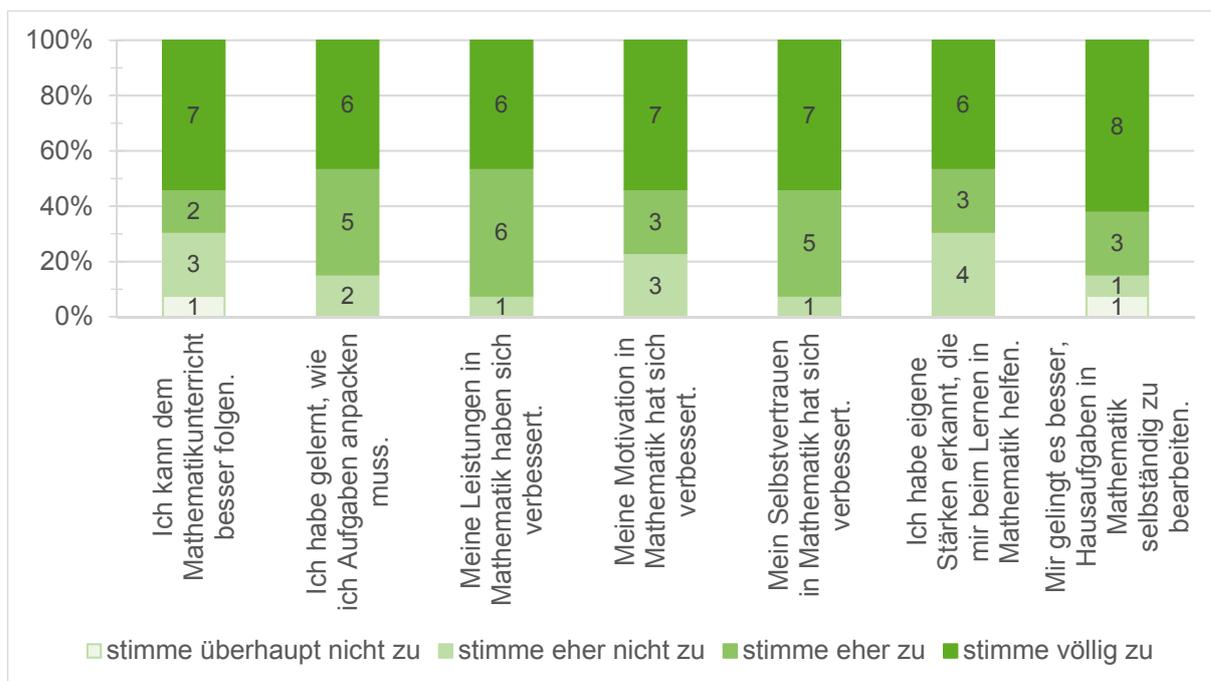
Anmerkungen. N = 13, fehlend: N = 1.

Abbildung 36. Veränderungen durch die Aufgabenhilfe bezogen auf den Deutschunterricht nach Angaben der Schülerinnen und Schüler



Anmerkungen. N = 16, fehlend: N = 2; *N = 15, fehlend: N = 3.

Abbildung 37. Veränderungen durch die Aufgabenhilfe bezogen auf den Mathematikunterricht nach Angaben der Schülerinnen und Schüler



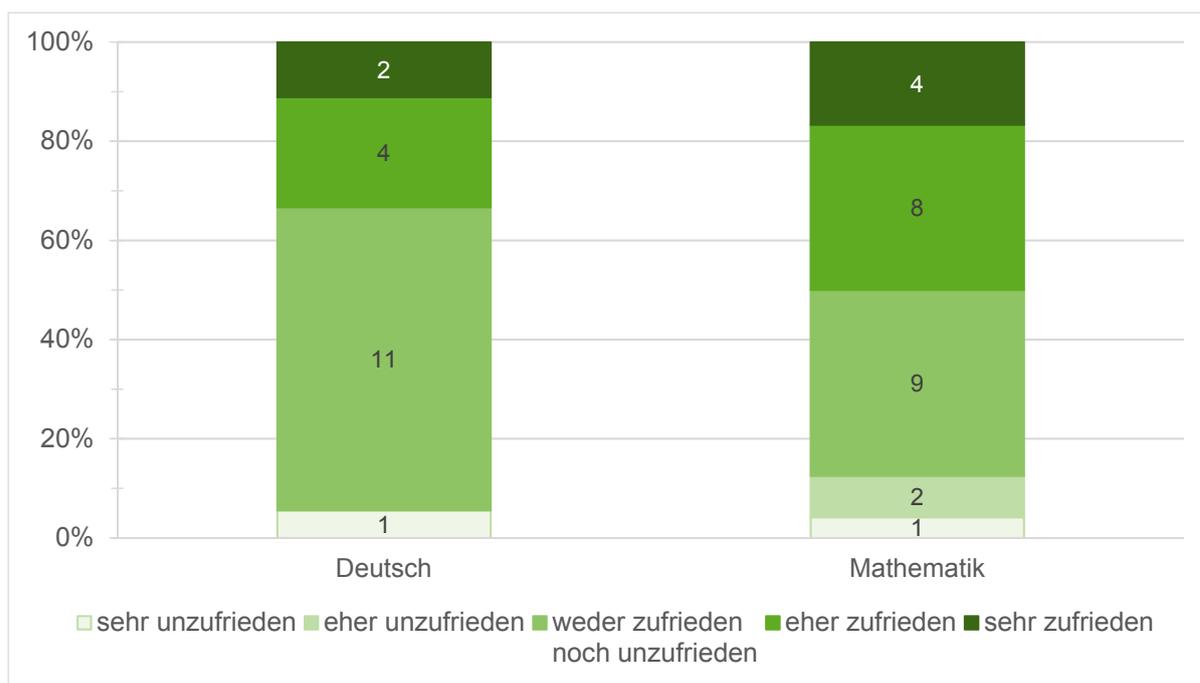
Anmerkungen. N = 13, fehlend: N = 1.

Zu bedenken ist bei der Beurteilung aller Ergebnisse, dass diese – wie bereits oben beschrieben – nur von einem sehr kleinen Anteil der befragten Schülerinnen und Schüler stammen, da nur diese angegeben Schülerinnen und Schüler hatten, die Aufgabenhilfe zu besuchen.

4.3.4 Allgemeine Zufriedenheit mit der Umsetzung der Massnahme

Alle an der Befragung teilnehmenden Lehrpersonen wurden zusätzlich nach ihrer allgemeinen Zufriedenheit damit gefragt, wie die Massnahme Aufgabenhilfe an ihrer Schule umgesetzt wurde. Wie die Ergebnisse in Abbildung 38 zeigen, stellt in beiden Fächern die neutrale Kategorie «weder zufrieden noch unzufrieden» die am häufigsten gewählte Antwort dar (Deutsch 61.1 Prozent; Mathematik 37.5 Prozent). Die weiteren Lehrpersonen äussern sich tendenziell zufrieden (*eher zufrieden* oder *sehr zufrieden*): Deutsch 33.3 Prozent; Mathematik 50.0 Prozent).

Abbildung 38. Allgemeine Zufriedenheit mit der Umsetzung der Massnahme Aufgabenhilfe im Rahmen des Pilotprojekts nach Angaben der Lehrpersonen



Anmerkungen. Deutsch $N = 18$, fehlend: $N = 8$ (befragt: alle Lehrpersonen, die angaben, das Fach Deutsch im Rahmen von ALLE zu unterrichten); Mathematik $N = 24$, fehlend: $N = 8$ (befragt: alle Lehrpersonen, die angaben, das Fach Mathematik im Rahmen von ALLE zu unterrichten).

Bezieht man nur die Lehrpersonen ein, die angeben, an ihrer Schule habe es eine Aufgabenhilfe gegeben (vgl. Kap. 4.3.2), resultiert die in Tabelle 30 gezeigte Verteilung auf die Antwortkategorien.

Tabelle 30. Allgemeine Zufriedenheit mit der Umsetzung der Massnahme Aufgabenhilfe bei Lehrpersonen, die berichten, dass eine Aufgabenhilfe für ihre Schülerinnen und Schüler angeboten wird

Antwortkategorie	Deutsch	Mathematik
Sehr zufrieden	2	4
Eher zufrieden	2	6
Weder zufrieden noch unzufrieden	8	3
Eher unzufrieden	0	1
Sehr unzufrieden	0	1

Anmerkungen. Deutsch $N = 12$, fehlend: $N = 3$; Mathematik $N = 15$.

Der Anteil der eher oder sehr zufriedenen Lehrpersonen fällt noch etwas höher aus als in der zuvor berichteten Analyse, insbesondere im Fach Mathematik bildet er die deutliche Mehrheit (66.7 Prozent). Im Fach Deutsch geben zwei Drittel der Lehrpersonen das Urteil «weder zufrieden noch unzufrieden» ab.

Nachfolgend zwei exemplarische Aussagen aus den Gruppeninterviews, in denen sich eine Lehrperson sehr positiv, eine andere ambivalent beziehungsweise stärker differenzierend zum Angebot der Aufgabenhilfe äussert:

- *«Ich finde es mega, wie die einen Schüler teilweise in diesen Stunden selbständig arbeiten können.»*
- *«Ich haben auch Schüler, bei denen es nicht funktioniert hat, die bei mir im Zimmer sitzen, nach der Schule oder über den Mittag und dann auch nicht arbeiten wollen, es kommt ganz darauf an, wie sie darauf anspringen und das ist wirklich sehr individuell.»*

Die von der zweiten Lehrperson beschriebene grosse Heterogenität der Wirkungen der Aufgabenhilfe könnte auch ein Grund dafür sein, dass ein grösserer Anteil der Lehrpersonen sich letztlich als «weder zufrieden noch unzufrieden» mit der Aufgabenhilfe sah.

Abschliessend soll ein Interviewausschnitt zitiert werden, der auf Unklarheiten im Verständnis des Begriffs der Aufgabenhilfe hindeutet:

[Interviewende/r]: «Zur Aufgabenhilfe. Aufgabenhilfe ist auch eine solche mögliche Massnahme gewesen, welche man im Rahmen von ALLE umsetzen konnte. Jetzt ist die Frage....» [Zwischenfrage von mehreren Personen.]: «Was ist das?»

Im weiteren Verlauf des Interviews ergeben sich die folgenden Aussagen:

[Interviewende/r:] «Was ist das, habt ihr das schon vorher gehabt?» [Lehrperson A:] «Bei uns gibt es das gar nicht.» [Interviewende/r:] «Bei euch gibt es das gar nicht? Also Aufgaben- oder Hausaufgabenhilfe?» [Lehrperson A:] «Also die klassische Hausaufgabenhilfe hat man offiziell abgeschafft bei uns. Also bei uns im Schulhaus. Das hat man nachher integriert in diese Lernlandschaften, in diese Gefässe, wie die auch immer aussehen.» [Lehrperson B:] «Bei uns im Zentrum ist es anders. Wir haben das im Angebot. Und zwar Montag, Dienstag, Donnerstag würde ich meinen, jeweils von vier bis sechs. Ja, in der Mediathek.»

Die Begrifflichkeit der «Aufgabenhilfe» als eine der im Rahmen des Pilotprojekts realisierten Massnahmen scheint somit nicht allen Lehrpersonen – auch innerhalb desselben Schulhauses – gleichermassen vertraut gewesen zu sein.

In der Befragung der Schulleitungen, wie sie das Gelingen der Massnahme Aufgabenhilfe an ihrer Schule einschätzen, ergeben sich die in Tabelle 31 dargestellten Antworten.

Tabelle 31. Allgemeine Zufriedenheit mit der Umsetzung der Massnahme Aufgabenhilfe bei den Schulleitungen

Antwortkategorie	Deutsch	Mathematik
Sehr gut	1	2
Gut	6	7
Befriedigend	4	1
Schlecht	0	0
Sehr schlecht	0	0

Anmerkungen. Deutsch $N = 11$; Mathematik $N = 10$, fehlend $N = 1$.

Demnach sehen alle befragten Schulleitungen den Einsatz der Aufgabenhilfe für beide Fächer als gelungen, in der Mehrheit als «gut» umgesetzt an.

4.4 Weiterbildung der Lehrpersonen

Als eine weitere Massnahme wurden im Rahmen des Pilotprojekts Weiterbildungen angeboten, die den Lehrpersonen Hinweise und Anreize für die Weiterentwicklung ihres Unterrichts geben sollten. Der Besuch war zum Teil sowohl für die Schulleitungen als auch für die Lehrpersonen verpflichtend (fachspezifische Weiterbildung; Einführung zu EMU [Evidenzbasierte Methoden der Unterrichtsdiagnostik und -entwicklung; vgl. Kap. 4.6]) und zum Teil freiwillig (Kooperation zwischen Lehr-

person und Schulassistent; Lerncoaching). Die Weiterbildungen wurden schulintern vor Ort durchgeführt, und die Schulleitungen meldeten jeweils die Lehrpersonen für die Teilnahme an. In diesem Kapitel wird auf folgende Fragestellungen eingegangen:

- Welche und wie viele Weiterbildungen wurden besucht? (Kap. 4.4.1)
- Wie wird die Nützlichkeit der Weiterbildungen beurteilt? Welche Verbesserungen im Fachunterricht infolge der Weiterbildungen nehmen die Lehrpersonen wahr? (Kap. 4.4.2)
- Wie zufrieden sind die Lehrpersonen und die Schulleitungen mit der Umsetzung der Weiterbildungsmaßnahmen an ihrer Schule? (Kap. 4.4.3)
- Welche Möglichkeiten zur Optimierung des Weiterbildungsangebots sehen die Weiterbilderinnen und Weiterbildner? (Kap. 4.4.4)

4.4.1 Besuchte Weiterbildungen

Wie Tabelle 32 zeigt, berichtet eine Mehrheit der Lehrpersonen beider Fächer, an den verpflichtenden Weiterbildungen (im Fach bzw. zu EMU) teilgenommen zu haben. Die freiwillige Weiterbildung zum Lerncoaching besuchte je etwa die Hälfte der Lehrpersonen. Die ebenfalls freiwillige Weiterbildung zur gelingenden Kooperation zwischen Lehrperson und Schulassistent wurde von weniger als einem Fünftel der Befragten besucht.

Tabelle 32. Besuchte Weiterbildungen im Rahmen des Pilotprojekts nach Lehrpersonen-Angaben

Besuchte Weiterbildungen	Deutsch-Lehrpersonen		Mathematik-Lehrpersonen		Gesamt
	Anzahl	Prozent aller LP	Anzahl	Prozent aller LP	
Fach Deutsch / Mathematik	19	73.1	28	87.5	47
Einführung zu EMU	19	73.1	23	71.9	42
Gelingende Kooperation zwischen LP und SA*	3	12.5	5	16.7	8
Lerncoaching**	13	52.0	14	46.7	27

Anmerkungen. LP = Lehrperson, SA = Schulassistent. Deutsch: $N = 26$; $*N = 24$, fehlend: $N = 2$; $**N = 25$, fehlend: $N = 1$; Mathematik: $N = 32$; $*N = 30$, fehlend: $N = 2$; $**N = 30$, fehlend: $N = 2$. Antwortformat pro Item: ja – nein.

Tabelle 33 gibt einen Überblick, wie viele Weiterbildungen von den Lehrpersonen insgesamt besucht wurden. Das Ausmass, in dem das Weiterbildungsangebot wahrgenommen wurde, ist demnach individuell deutlich verschieden. Die Mehrheit der Deutsch-Lehrpersonen nahm an einer bis zwei, rund die Hälfte der Mathematik-Lehrpersonen an zwei Weiterbildungen teil. Alle vier Weiter-

bildungen besuchten zwei Deutsch- und vier Mathematik-Lehrpersonen. Drei Mathematik-Lehrpersonen geben an, keine Weiterbildung besucht zu haben. Diese Lehrpersonen übernahmen jedoch erst im Lauf des Pilotprojekts eine ALLE-Klasse, zwei von ihnen erst zum letzten Messzeitpunkt.

Tabelle 33. Anzahl der insgesamt besuchten Weiterbildungen im Rahmen des Pilotprojekts pro Lehrperson nach eigenen Angaben

Anzahl besuchter Weiterbildungen	Deutsch-Lehrpersonen		Mathematik-Lehrpersonen		Gesamt
	Anzahl	Prozent	Anzahl	Prozent	
0	0	0.0	3	9.4	3
1	9	34.6	3	9.4	12
2	8	30.8	15	46.9	23
3	7	26.9	7	21.9	14
4	2	7.7	4	12.5	6
Gesamt	26	100.0	32	100.0	

Anmerkungen. Deutsch: $N = 26$; Mathematik: $N = 32$. Anzahl besuchter Weiterbildungen: Auszählung der Ja-Antworten für die einzelnen Weiterbildungen pro Lehrperson.

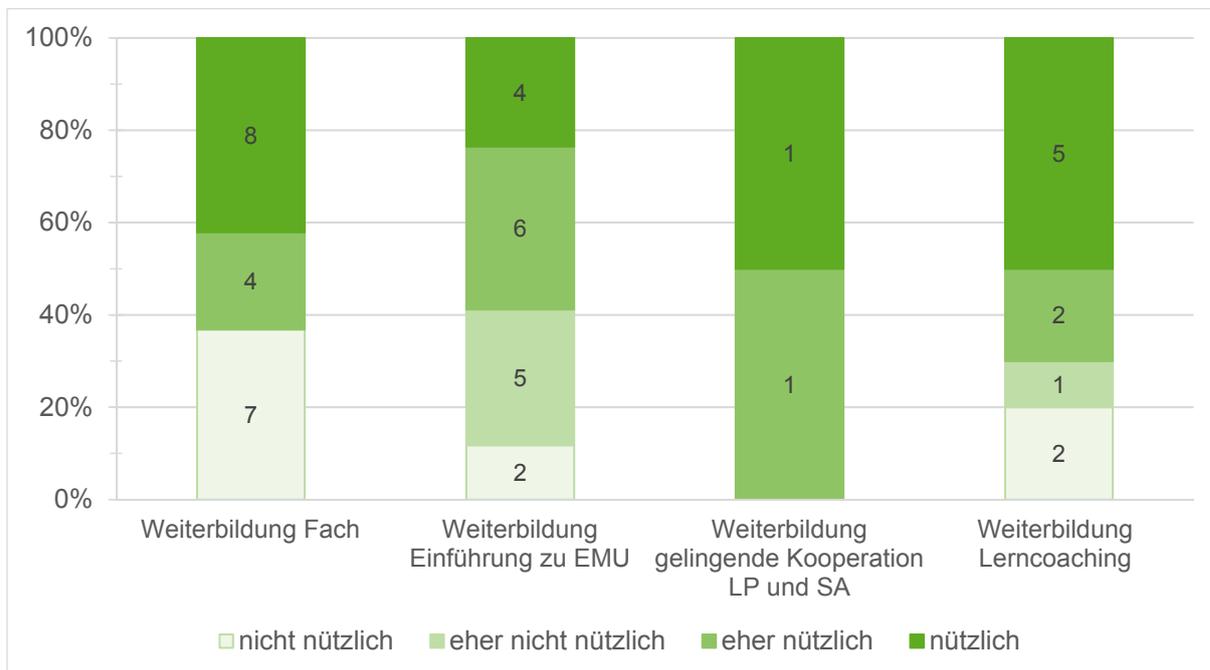
4.4.2 Wahrgenommene Nützlichkeit und Verbesserungen im Fachunterricht

Die Lehrpersonen wurden danach gefragt, wie nützlich sie die Teilnahme an den einzelnen Weiterbildungen für die konkrete Umsetzung im Unterricht (Praxistransfer) empfanden. Die Ergebnisse sind in Abbildung 39 und Abbildung 40 für die Deutsch- und Mathematik-Lehrpersonen dargestellt. Es werden nur die Angaben derjenigen Lehrpersonen einbezogen, welche berichten, an der jeweiligen Weiterbildung teilgenommen zu haben.

Fast alle Weiterbildungen werden hinsichtlich ihres Nutzens für die Umsetzung im Unterricht mehrheitlich positiv beurteilt. Eine Ausnahme stellt die fachspezifische Weiterbildung im Fach Deutsch dar, für die sich eine deutliche Polarisierung zeigt: acht von 19 Lehrpersonen schätzen diese als «nützlich» ein (42.1 Prozent), sieben (36.8 Prozent) dagegen als «nicht nützlich». Auch bei den meisten anderen Weiterbildungen zeigen die Urteile eine gewisse Heterogenität, eine durchgehend günstige Beurteilung als «eher nützlich» oder «nützlich» ergibt sich für die Weiterbildung zur Kooperation Lehrperson/Schulassistenz (im Urteil der Deutsch-Lehrpersonen) und die Weiterbildung zum Lerncoaching (im Urteil der Mathematik-Lehrpersonen).

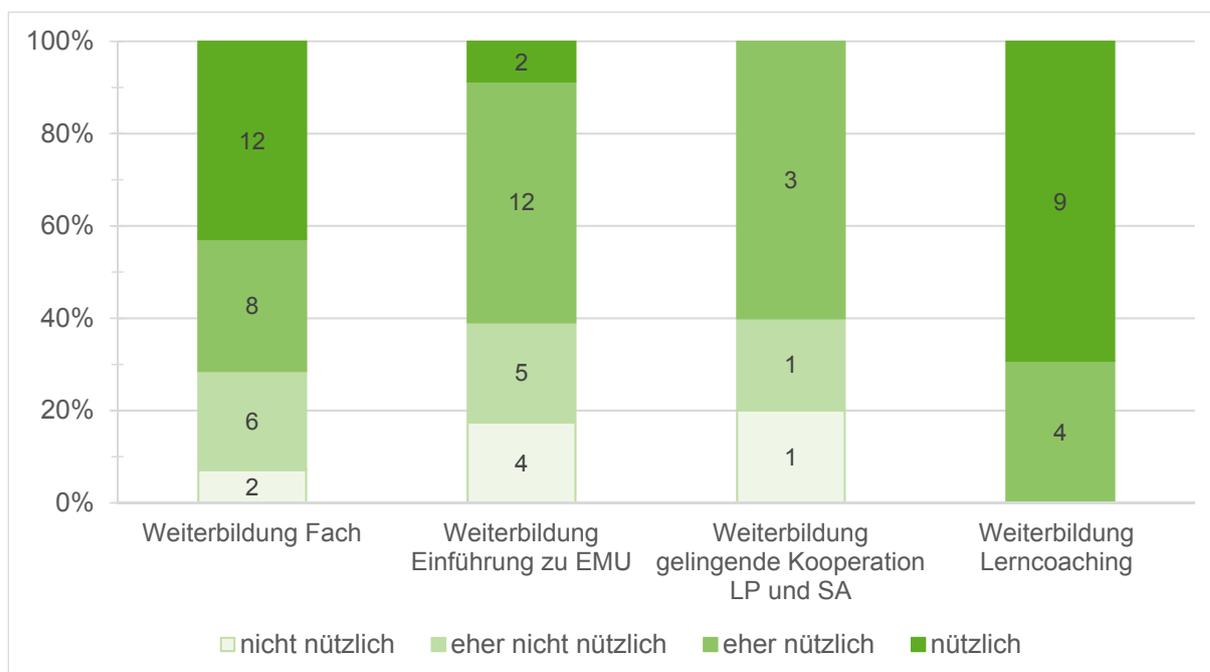
Ein nochmals positiveres Urteil wird von den Schulleitungen abgegeben, welche ebenfalls zu den Weiterbildungen der Lehrpersonen befragt wurden. Die elf befragten Schulleitungen schätzen die Nützlichkeit für die konkrete Umsetzung im Unterricht allesamt als «eher nützlich» (27.3 Prozent, $N = 3$) oder «nützlich» (72.7 Prozent, $N = 8$) ein.

Abbildung 39. Nützlichkeit der Weiterbildungen nach Angaben der Deutsch-Lehrpersonen



Anmerkungen. LP = Lehrperson, SA = Schuassistenz. Weiterbildung Fach: N = 19; Weiterbildung Einführung zu EMU: N = 17, fehlend: N = 2; Weiterbildung gelingende Kooperation LP und SA: N = 2, fehlend: N = 1; Weiterbildung Lerncoaching: N = 10, fehlend: N = 3. «Fehlend» sind Lehrpersonen, die angeben, die jeweilige Weiterbildung besucht zu haben, aber kein Urteil zur Nützlichkeit abgeben.

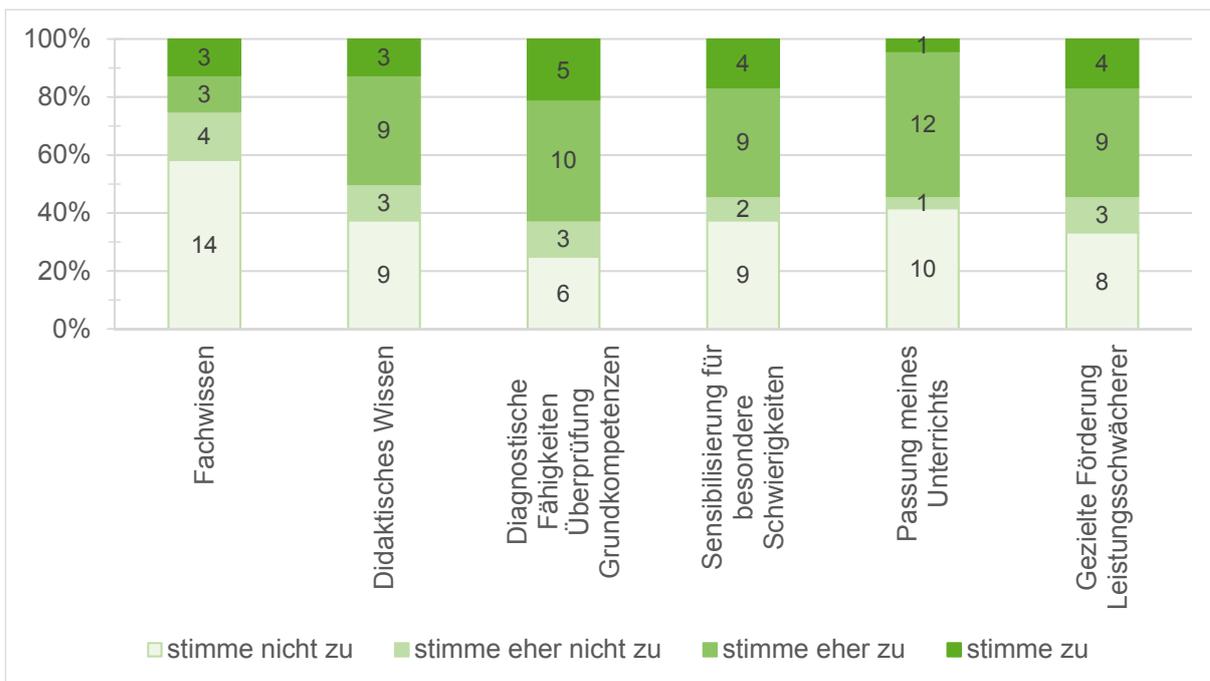
Abbildung 40. Nützlichkeit der Weiterbildungen nach Angaben der Mathematik-Lehrpersonen



Anmerkungen. LP = Lehrperson, SA = Schuassistenz. Weiterbildung Fach: N = 28; Weiterbildung Einführung zu EMU: N = 23; Weiterbildung gelingende Kooperation LP und SA: N = 4, fehlend: N = 1; Weiterbildung Lerncoaching: N = 13, fehlend: N = 1. «Fehlend» sind Lehrpersonen, die angeben, die jeweilige Weiterbildung besucht zu haben, aber kein Urteil zur Nützlichkeit abgeben.

Im Folgenden wird der Frage nachgegangen, inwiefern die Lehrpersonen eine Verbesserung von Aspekten mit Bezug zu ihrem Fachunterricht infolge des Besuchs der Weiterbildungen sehen. Die Resultate für die Deutsch- und Mathematik-Lehrpersonen zeigen Abbildung 41 und Abbildung 42.

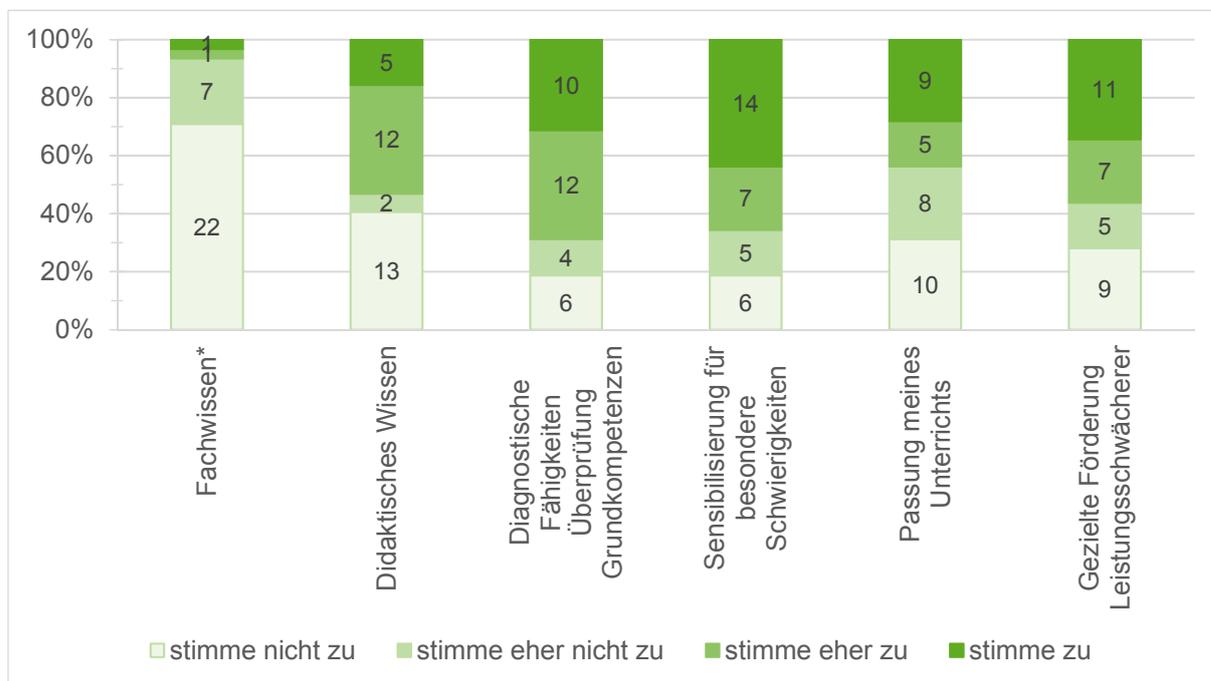
Abbildung 41. Verbesserungen in Bezug auf den Fachunterricht durch die besuchten Weiterbildungen nach Angaben der Deutsch-Lehrpersonen



Anmerkungen. N = 24, fehlend: N = 2.

Die geringste Zustimmung findet sich für beide Fächer in Bezug auf Verbesserungen des eigenen Fachwissens (*stimme eher zu* oder *stimme zu*: Deutsch: 25.0 Prozent; Mathematik: 6.5 Prozent). Zu berücksichtigen ist dabei, dass die Verbesserung des Fachwissens nur am Rande Ziel der Weiterbildungen war. Vor allem sollte eine Stärkung der diagnosegeleiteten Förderung und der Sensibilisierung für bestimmte (Lern-)Schwierigkeiten erfolgen. Entsprechend zeigt sich eine höhere Zustimmung für Aspekte der diagnostischen Fähigkeiten der Lehrpersonen: Jeweils rund zwei Drittel der Befragten gibt an, dass sich die Fähigkeit zur Prüfung von Grundkompetenzen (eher) gesteigert habe (*stimme eher zu* oder *stimme zu*: Deutsch: 62.5 Prozent; Mathematik: 68.8 Prozent). In beiden Fächern stimmt etwas mehr als die Hälfte (eher) zu, dass sich die auf Grundkompetenzen ausgerichtete Förderung leistungsschwächerer Schülerinnen und Schüler verbessert (*stimme eher zu* oder *stimme zu*: Deutsch: 54.2 Prozent; Mathematik: 56.3 Prozent) und die Sensibilisierung für besondere Schwierigkeiten von Lernenden zugenommen habe (*stimme eher zu* oder *stimme zu*: Deutsch: 54.2 Prozent; Mathematik: 65.6 Prozent). Eine Steigerung des didaktischen Wissens (*stimme eher zu* oder *stimme zu*: Deutsch: 50.0 Prozent; Mathematik: 53.1 Prozent) und eine verbesserte Passung des Unterrichts (*stimme eher zu* oder *stimme zu*: Deutsch: 54.2 Prozent; Mathematik: 43.8 Prozent) berichten ebenfalls etwa die Hälfte der Lehrpersonen.

Abbildung 42. Verbesserungen in Bezug auf den Fachunterricht durch die besuchten Weiterbildungen nach Angaben der Mathematik-Lehrpersonen



Anmerkungen. N = 32; *N = 31, fehlend: N = 1.

Zwei Beispiele aus den Gruppeninterviews mit den Lehrpersonen illustrieren, wie sich als Folge der Weiterbildungen eine erhöhte Passung des Mathematikunterrichts ergeben kann:

- [Interviewende/r:] «Bleiben wir doch bei der Weiterbildung im Bereich Mathe. Habt ihr diese Matheweiterbildungen besucht?» [Lehrperson:] «Ja, [Name des Weiterbildungsdozenten] war bei uns und hat die Weiterbildungen hier gemacht. Alle Mathelehrer fanden das gut, dass wir hier einen Einblick hatten. Sie haben sich auch noch überlegt, was machen wir in der Mathe mit unseren Plänen, welche Themen nehmen wir dazu, welche können wir gemäss [Name des Weiterbildungsdozenten] auch auslassen. Wir haben dort diese Pläne überarbeitet.»
- «Es gab schon noch ein paar Anregungen, welche wir weiter brauchen konnten. Zum Beispiel für das Niveau I diese zehn Aufgaben in fünf Minuten, welche wir auch schulintern weiterentwickelt haben und noch mehr davon gemacht haben.»

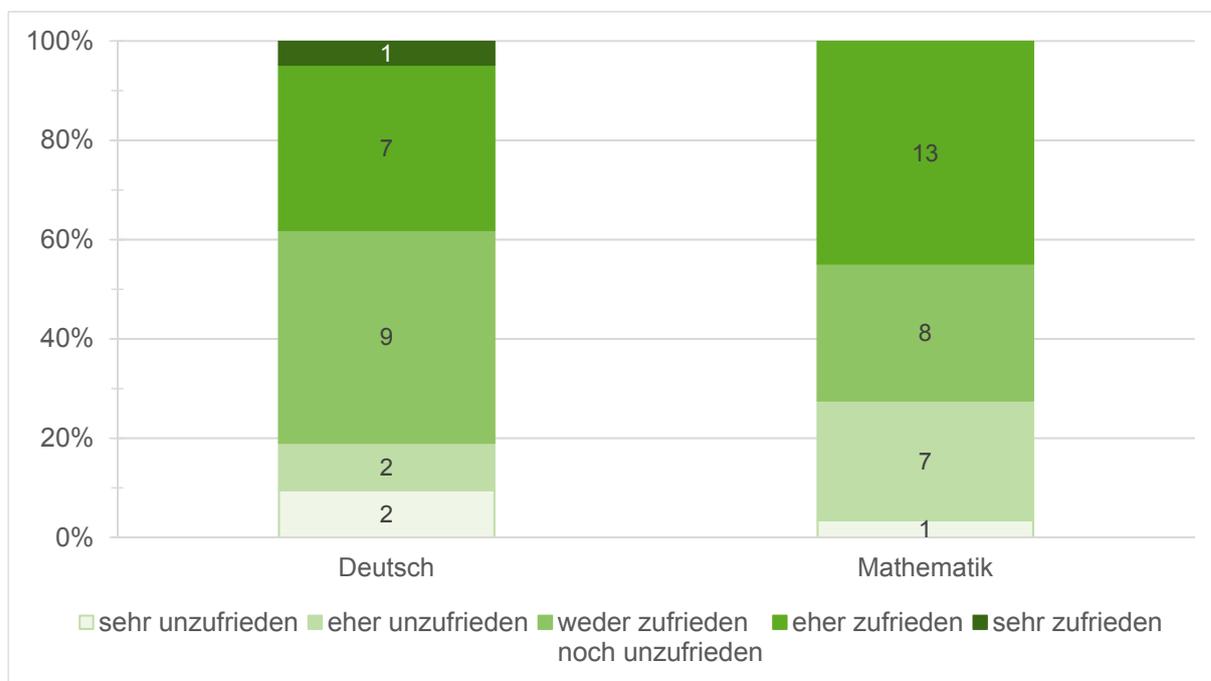
Um die Breite der Wirkungen der Weiterbildungen einschätzen zu können, ist ebenfalls bedeutsam, inwiefern die einzelnen Lehrpersonen Verbesserungen in Bezug auf ihren Unterricht wahrnehmen. Ungünstig wäre es, wenn eine grössere Anzahl Lehrpersonen *keine* Wirkungen der besuchten Weiterbildungen auf ihren Unterricht wahrnehmen würden. Dies ist in beiden Fächern bei weniger als einem Fünftel der Lehrpersonen der Fall: vier von 24 Deutsch-Lehrpersonen und sechs von 32 Mathematik-Lehrpersonen stimmen bei jedem der erfragten Aspekte zum Fachunterricht (vgl. Abbildung 41 und Abbildung 42) «nicht zu», dass es eine Verbesserung gegeben habe.

4.4.3 Allgemeine Zufriedenheit mit der Umsetzung der Massnahme

Abschliessend wurde von den Lehrpersonen bewertet, wie zufrieden sie generell mit der Umsetzung der Massnahme Weiterbildungen in der eigenen Schule sind. Dabei ist zu berücksichtigen, dass es sich um ein globales Urteil über alle angebotenen Weiterbildungen handelt. Die Frage wurde nicht nach den einzelnen Weiterbildungen getrennt gestellt, und es wurden auch keine Qualitätsmerkmale, wie zum Beispiel Inhalt oder Struktur, separat bewertet. Zudem können die Urteile auch Erfahrungen widerspiegeln, welche eher die Umsetzung an der jeweiligen Schule als die besuchten Weiterbildungen selbst betreffen (z.B. die Organisation seitens der Schulleitung).

Wie Abbildung 43 zeigt, fällt das Urteil der Lehrpersonen eher heterogen aus. In Deutsch beträgt der Anteil an (eher) zufriedenen Lehrpersonen und an Lehrpersonen, die «weder zufrieden noch unzufrieden» sind, jeweils rund 40 Prozent (*eher zufrieden* oder *sehr zufrieden*: 38.1 Prozent; *weder zufrieden noch unzufrieden*: 42.9 Prozent). In Mathematik sind annähernd die Hälfte der Lehrpersonen (eher) zufrieden (44.8 Prozent), knapp ein Drittel äussert sich «weder zufrieden noch unzufrieden» (27.6 Prozent). Rund jede fünfte Deutsch-Lehrperson (19.0 Prozent) und jede vierte Mathematik-Lehrperson (27.6 Prozent) berichtet, (eher) unzufrieden mit der Umsetzung der Weiterbildungen zu sein.

Abbildung 43. Allgemeine Zufriedenheit mit der Umsetzung der Durchführung der Weiterbildungen im Rahmen des Pilotprojekts nach Angaben der Lehrpersonen



Anmerkungen. Deutsch $N = 21$, fehlend: $N = 5$ (befragt: alle Lehrpersonen, die angaben, das Fach Deutsch im Rahmen von ALLE zu unterrichten); Mathematik $N = 29$, fehlend: $N = 3$ (befragt: alle Lehrpersonen, die angaben, das Fach Mathematik im Rahmen von ALLE zu unterrichten).

In den Gruppeninterviews äussert sich eine Lehrperson, welche angibt, (eher) unzufrieden mit den Weiterbildungen gewesen zu sein, wie folgt:

«Im Bereich Deutsch war es so, dass wir begonnen haben und im Bereich Schreiben einen sozialen Schreib Anlass gestaltet haben. Diesen haben wir auch mit dem ganzen Jahrgang durchgeführt. Das konnten wir dann auch weiterziehen. [...] hätten wir gerne noch mehrere Anregungen erhalten. Letztenendes ist das Ganze dann versendet, weil wir kein Feedback und keine weiteren Vorgehensaufträge mehr erhalten haben. [...] Dort hörten wir dann auch nichts. Seit 2 Jahren oder 1 1/2 Jahren warten wir auf neue Aufträge, welche wir durchführen könnten und um einen Lernerfolg zu erzielen. Dies hat uns ins Stocken gebracht.»

Auch die Äusserung einer weiteren Lehrperson weist auf fehlende Anknüpfungspunkte hin:

«Auch, dass das ganze Team, das bei der Weiterbildung dabei war, unzufrieden war mit: Wie weiter? Jetzt im Fach Deutsch, das habe ich wirklich mitgekriegt.»

Bei einigen Lehrpersonen könnte die Unzufriedenheit auch auf individuelle (nicht erfüllte) Erwartungen zurückgehen. So äussert sich eine Lehrperson im Gruppeninterview wie folgt:

«Deutsch nein, wirklich gar nichts, Entschuldigung, es ist, ja vom ganzen Aufbau her, ich habe etwas ganz anderes erwartet. Ich bin ernüchtert gewesen [...] und auch, also vom Inhalt her bin ich mir nicht ernst genommen vorgekommen.»

Demgegenüber gibt es auch positive Stimmen zu den Weiterbildungen:

«Aber im Deutsch, das hat mir total viel gebracht. Also auch einen neuen Blick darauf zu bekommen, wie man [...] an Text herangeht, wie man Texte beurteilt. Ich meine jetzt nicht einfach nur notenmässig, sondern wo man eigentlich den Fokus darauf legt, was man als wichtig erachtet. Das habe ich total spannend gefunden.»

Wie Impulse aus den Weiterbildungen in den Unterricht mitgenommen wurden, zeigt exemplarisch diese Äusserung einer Lehrperson:

«Der Schwerpunkt ist ja auf dem sozialen Schreiben gelegen. Dort hat sich vor allem bei mir ein wenig mehr getan, dass ich den Schwerpunkt ein bisschen verlegt habe im Fach Deutsch.»

Danach gefragt, wie gut ihrer Ansicht nach die Umsetzung der Weiterbildungsmassnahmen an der eigenen Schule gelungen ist, geben zehn Schulleitungen folgende Antworten:

- Sehr gut: 3 Schulleitungen
- Gut: 6 Schulleitungen
- Befriedigend: 1 Schulleitung
- Schlecht: keine Schulleitung
- Sehr schlecht: keine Schulleitung

Alle zehn Schulleitungen, die geantwortet haben, sehen die Umsetzung der Weiterbildungen an ihrer Schule demnach als positiv, die Mehrheit als «gut» gelungen an.

4.4.4 Überlegungen der Weiterbildungspersonen zu Optimierungen des Angebots

Hinweise auf mögliche Optimierungen des Weiterbildungsangebots lassen sich den Interviews mit den Weiterbildungspersonen entnehmen. Diese sehen vor allem Perspektiven

- in der genaueren Erfassung und Berücksichtigung des Vorwissens der Teilnehmenden bei der Planung und inhaltlichen Gestaltung der Weiterbildungen;
- in der Vereinfachung und stärkeren Zentrierung der Weiterbildungsinhalte;
- in der parallelen Entwicklung von Unterrichtsmaterialien und Kursangeboten, um diese besser aufeinander abzustimmen;
- in der Ausweitung der Kursinhalte auf die Förderung der leistungsstärkeren Schülerinnen und Schüler;
- in der Ergänzung der Weiterbildungen durch systematische und verbindliche Unterrichtsbeobachtungen, um den Blick stärker auf den eigenen Unterricht zu lenken;
- im Angebot von individuellen Coachings statt allgemeiner Gruppen-Weiterbildungen, um besser auf die Bedürfnisse einzelner Lehrpersonen eingehen zu können (z.B. bei Schwierigkeiten, sich oder den eigenen Unterricht umzustellen/anzupassen) oder um in Kleingruppen an konkreten Themen zu arbeiten;
- in der Bereitstellung von umfangreicheren Ressourcen für mehr und fortlaufende Weiterbildungstage (z.B. zu Beginn jedes Schuljahres zu zentralen Themen des jeweiligen Schuljahres), um die kontinuierliche Weiterarbeit an den Themen sowie das Behandeln komplexerer Aspekte sicherzustellen beziehungsweise zu ermöglichen;
- in der (strukturellen) Ermöglichung von Unterrichts- und Schulentwicklung, um dauerhafte Änderungen des Unterrichts und der Förderung der Schülerinnen und Schüler erreichen zu können.

Die nachfolgenden Ausschnitte aus den Interviews mit den Weiterbildungspersonen sollen die angesprochenen Aspekte verdeutlichen.

Berücksichtigung des Vorwissens

«[...] Manchmal kann man sagen, hätte man vielleicht noch ein wenig länger in dieser Vorwissenszeit, dort ein wenig länger verweilen [können]. Also, das noch präziser erfassen. Einfach um ihnen auch zu zeigen, ja, was sie alles eigentlich schon machen. Vielleicht noch nicht so systematisch, wo man auch sieht, das könnte man noch ein Stück weit verbindlicher machen. Und auch vielleicht festschreiben.»

Inhalte: Zentrierung, Anpassung an Unterrichtsmaterialien, Ausweitung auf Leistungsstarke

- *[Interviewende/r:] «Haben Sie das Gefühl, dass das Angebot, so wie Sie das jetzt machen konnten, ist das gross genug, breit gewesen in dem Projekt? Also hat das genug Gewicht erhalten?»
[Weiterbildner/in:] «Ich meine ja, es hat genug Gewicht erhalten. Ich meine, das, was wir angeboten haben, hätte den Rahmen ein wenig gesprengt, im Sinne von, es sei fast überfüllt gewesen. Man hätte es vielleicht vereinfachen und stärker zentrieren können. Und dann wäre die Wirksamkeit, falls es überhaupt eine hat, die wäre stärker geworden. Da bin ich überzeugt.»*
- *«Ja, es würde sich jetzt in Zukunft ändern, weil zum Beispiel die Fördermaterialien, welche wir damals vorgestellt haben, mit welchen die Schulen gearbeitet haben, die werden in einem guten Jahr abgelöst durch ein Fördermaterial, welches der Lehrmittelverlag Zürich herstellt, welches besser aufs Lehrmittel bezogen ist. Also, würde man vermutlich dann eher mit jenem Material arbeiten, als mit dem, welches man damals hatte. Also, das wäre jetzt sicher etwas ... halt, dass das Materialienangebot, das es gibt, sich verbessert. So, dass man auch die Kurse überarbeiten würde.»*
- *«Es wäre auch mal noch so ein Teil vom Projekt gewesen, also Förderung von ALLE heisst ja eigentlich auch, wie man die Guten fördern könnte. Da haben wir in der Weiterbildung ein wenig etwas gemacht. Das ist aber zu wenig gewesen. Da hat glaube ich, hat ... ist jetzt nicht viel daraus erwachsen.»*

Ergänzungen der Weiterbildungen: verbindliche Unterrichtsbeobachtungen, Coachings

- *«[...] Das ist richtig, was Sie sagen, weil [...] man ist ja mit sich selbst unterwegs, und ist nicht selbst geeignet, immer beide Rollen auch als Forschungsobjekt noch grad zu führen. [...] Wir sind eigentlich für Introspektion nicht so geeignet. Also, weil es einfach nicht geht, ausser eben, man verwendet Hilfsmittel, also eine Triangulation, mit Beobachtern und Video. Darum sind wir immer wieder so stark da drauf ausgewiesen (haben wir immer wieder so stark darauf gepocht). Aber, ob sie das machen, das kann ich Ihnen jetzt nicht sagen, weil es einfach aufwändig ist, und sie sich dem nicht ganz stellen können oder wollen.»*
- *«[...] Und es haben auch nicht alle die Videobänder gemacht. Also das ist jetzt so eine Umsetzstelle, wo ich finde, die hätte man noch ein wenig energischer verlangen müssen.»*
- *«Und es ist eben so, bei [...] defensiv stark engagierten Lehrpersonen, sollte man dann eben nicht mit Weiterbildung arbeiten, sondern mit einem individuellen Coaching, wo die Lehrperson zur Schulleitung muss, und die Schulleitung das mit ihr anschaut. Ich bin der Meinung, dass man dort Weiterbildung und Beratung kippen muss in ein Einzelgespräch [...]. Weil es kann sein, dass diese Person gute persönliche Gründe hat, wieso sie etwas nicht kann oder nicht mag oder nicht will.»*

- *«Ich habe so das Gefühl, es müsste eigentlich ... eher in kleinen Gruppen gleich direkt am Thema in Form von Zusammensitzen... und ich hätte jetzt gedacht, ja, so als Coach dort noch dabei sein zu können.»*
- *«Eben, ich bin eigentlich, ich bin eher überzeugt von Coaching, also von kleineren Gruppen [...]»*

Strukturelle Aspekte: fortlaufende Weiterbildungen, verbesserte Ressourcen, Änderungen in der Schulstruktur

- *«[...] und ich könnte mir vorstellen, dass man [...] anfangs zweite Sek, anfangs dritte Sek, dort nochmals irgendwie Thematiken aufgreifen [würde], also konkret an zentralen Themen, welche im Zweitklassstoff kommen oder welche im Drittklassstoff kommen, nochmals dort eine Weiterbildung [dazu] machen [würde]. Einfach mehr auch nochmal, um dran zu bleiben und ja, so ein wenig ein Start ins neue Jahr irgendwo, so ein wenig eine Grundlage schaffen, und ja, wie man das anpacken könnte.»*
- *«Also, ich glaube, was man schon ein wenig überdenken müsste, wenn man wirklich nochmals so grundsätzlich überlegen will, ist die Struktur: Also, zwei Weiterbildungstage als guter Starter, und dann ist nur noch freiwilliges Coaching. Also, ich würde jetzt eben, im Gesamteindruck, sagen, die zwei Weiterbildungstage haben viel ausgelöst, haben auch zur Schulentwicklung geführt, zur Teamentwicklung, zur Unterrichtsentwicklung. Nachdem, sage ich jetzt mal, fulminanten Anfang, dem ersten halben Jahr, würde ... ist eigentlich so mein Eindruck, ohne das jetzt so voll [Anm.: Unterbruch] hat das abgeflacht. Man hat das eben eingeführt für die Siebtklässler und so weiter, und dann hat man es wieder ein wenig aus den Augen verloren, abgesehen von den paar wenigen Coachingsachen. Also, ich glaube, man müsste mehr Ressourcen haben: Zwei Tage am Anfang ist in Ordnung, aber es bräuchte auch mehr ... Gefässe, welche über die Länge sicherstellen, dass man dran bleibt. Also, dass man irgendwie ein Jahr später schauen könnte, oder zwei Jahre ... [...]»*
- *«Sie haben es gerade angesprochen, also dass mehr Tage immer besser wären, um noch mehr anzubieten oder Interessanteres anzubieten.»*
- *«Es gibt andere Themen, welche wir aufgrund von nur zwei Weiterbildungstagen nicht so richtig etablieren konnten. Also da würde ich jetzt zum Beispiel sagen: langfristige Förderung der Basiskompetenzen, also nicht nur im 7. Schuljahr einmal gut hinschauen und Lücken füllen, sondern auch schauen, ist das im 8. und im 9. auch noch da? Da würde ich jetzt sagen, haben wir nicht so viel vermitteln können, weil es nur zwei Tage sind.»*

- *«Über die Art, wie das dann umgesetzt worden ist, dort finde ich ja [...] ... gibt es noch Entwicklungsmöglichkeiten. Also mein Eindruck ist, dass man am Regelunterricht nicht viel geändert hat, und dann einfach mehr so ein wenig ein Förderzentrum gemacht hat [...]. Und für mich ist das jetzt noch nicht genügend. [...] Und ich glaube, es braucht dann auch Schulentwicklung, um die Strukturen machen zu können, damit man wirklich einen differenzierten Unterricht machen kann. Ich glaube, das ist noch ein sehr langer Weg auf der Sek-Stufe, weil einfach die Strukturen und die Kultur, welche auf der Sek herrschen ..., das ist – empfinde ich – ist speziell, und es macht es schwieriger auf der Sek-Stufe da eine Entwicklung anstossen zu können.»*

4.5 Beratungen für Lehrpersonen

Die am Pilotprojekt teilnehmenden Lehrpersonen hatten die Möglichkeit, Beratungen von Fachdidaktikerinnen und Fachdidaktikern in Anspruch zu nehmen. Diese sollten vor allem Impulse für die Weiterentwicklung des Unterrichts und damit für eine Steigerung der Unterrichtsqualität geben. Im Folgenden wird auf die nachfolgenden Fragestellungen zu den Beratungen eingegangen:

- In welchem Ausmass und zu welchen Themen wird das Beratungsangebot von den Lehrpersonen wahrgenommen? (Kap. 4.5.1)
- Wie zufrieden sind die Lehrpersonen und die Schulleitungen mit der Umsetzung des Beratungsangebots? (Kap. 4.5.2)
- Wie schätzen die Beratungspersonen den Erfolg ihrer Beratungen ein? Welche Möglichkeiten zur Optimierung des Angebots sehen sie? (Kap. 4.5.3)

4.5.1 Umfang und Art der in Anspruch genommenen Beratungen

Insgesamt wurde das Beratungsangebot nach Angaben der befragten Lehrpersonen nur wenig frequentiert. So geben lediglich vier der 26 Deutsch-Lehrpersonen (16.0 Prozent) und zwei der 32 Mathematik-Lehrpersonen (6.3 Prozent) an, Beratungen im Verlauf des Pilotprojekts in Anspruch genommen zu haben. Im Fach Deutsch wurden zwei Lehrpersonen einmalig, eine Lehrperson zweimal und eine fünfmal beraten. Die beide Mathematik-Lehrpersonen nahmen jeweils eine Beratung in Anspruch.

Für das Fach Deutsch beinhalteten die Beratungen nach Angaben der beteiligten Lehrpersonen folgende Themen (im Wortlaut):

- Orthographie
- Schreibschwäche
- Umsetzung

Für das Fach Mathematik werden die folgenden Angaben genannt (im Wortlaut):

- Basiskompetenzen der 2. und 3. Sek
- Wichtigkeit der Themenbereiche

Insgesamt scheinen die Themen sehr verschieden und für den Beratungsanlass spezifisch ausgefallen zu sein. So erwähnen die Beratungspersonen im Rahmen der mit ihnen geführten Interviews die folgenden Themen und Aspekte der in Anspruch genommenen Beratungen:

- Eher Weiterbildung als Beratung zum Thema Ablösung vom zählenden Rechnen
- Eingrenzung von Stoff auf Kernthemen, die gefestigt werden sollten
- Gestaltung von Lernaufträgen für selbstorganisiertes Lernen
- Jahresplanung
- Soziale Praxis für Schülerinnen und Schüler mit Deutsch als Zweitsprache und noch geringen Sprachkenntnissen
- Erstellung von Leitfäden für die 8. oder 9. Klassenstufe mit zentralen Aufgaben und Lernzielen

Als möglicher Grund für die geringe berichtete Inanspruchnahme des Beratungsangebotes deuten sich in den Gruppeninterviews mit den Lehrpersonen fehlende Zeitressourcen an:

«Das ist natürlich, ich finde eine mega Bereicherung. Wir hätten noch viele Fragen, das Problem ist einfach immer der Zeitfaktor. Dann ist das weg und dann haben wir tausend andere Sachen – und dann geht es vergessen. Ich denke, [...] so eine Fachperson da zu haben, ist immer eine Bereicherung. Aber wann, und organisatorisch geht das dann vielleicht einfach vergessen, kann ich mir vorstellen.»

Auch eine Beratungsperson äussert sich im Interview in diese Richtung:

«Ich vermute, der Haupthinderungsgrund ist, dass durch den Alltag und alles, was sonst läuft, hat man keine Zeit. Ich glaube nicht, dass man keine Lust hätte.»

Weitere Überlegungen der Beratungspersonen zur geringen Inanspruchnahme des Beratungsangebots beziehen sich vor allem auf die Inhalte der Beratung. Angesprochen werden eine (fehlende) Aktualität oder Dringlichkeit der Themen, der mögliche Mangel an Vertrautheit mit dem Angebot und die mit diesem verbundenen Ansprüche sowie Möglichkeiten, in mehr direktem Austausch mit den Lehrpersonen geeignetere Angebote zu schaffen (vgl. auch Kap. 4.5.3):

- *«Ich glaube, wenn einfach so der Schulalltag läuft, dann, wenn man nicht grad auf ein Thema stösst, bei dem man gleich denkt, oh, das wäre super, wenn mal jemand von aussen schauen würde – aha wir haben ja noch ALLE, komm das ist doch super – da holen wir gleich jemanden. Wenn es nicht so läuft, dann wird es glaube ich einfach ein wenig überspült vom Alltag.»*

- *«Das ist schwer zu sagen. Sie haben deutlich mehr, soweit ich das mitbekommen habe, Beratungen in der Mathematik in Anspruch genommen. Oder dann halt mehr im Zusammenhang mit EMU. Weniger jetzt Deutsch, Deutschdidaktik. Ich denke, das ist ihnen vielleicht auch weniger vertraut, dass es diese Möglichkeit gäbe. Und wenn man mehr in eine Beratung gehen will, müsste man sich wahrscheinlich auch eher ein Konzept überlegen, wo man vielleicht auch näher ran kommt.»*
- *«Also, wenn die Beratung bleibt, im Sinne von <es ist freiwillig, man kann sie beziehen oder auch nicht>, dann könnte ich mir vorstellen, wenn man mit Lehrpersonen oder jemandem aus dem Matheteam oder so gemeinsam überlegen würde, was für ein Coaching denn Sinn machen würde [...] – von den Themen – dann könnte ich mir vorstellen, könnte man mehr Interesse erzeugen [...].»*

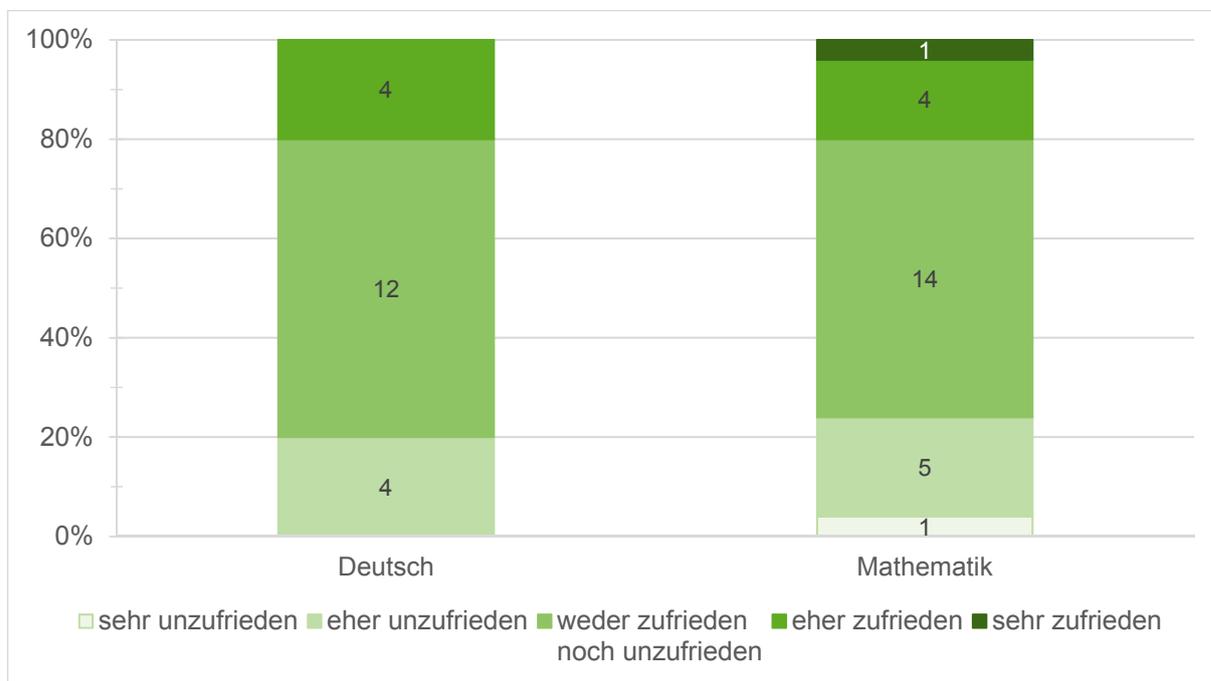
Ein weiteres Hemmnis wird von den Beratungspersonen in mangelndem «Mut» oder in der mangelnden Bereitschaft gesehen, anderen Einblick in den eigenen Unterricht zu gewähren:

- *«Also, ich weiss es nicht, wie man es attraktiver machen könnte. Weil es ..., eben, es tut natürlich auch weh. Also, wenn ich dann sehe, [...] wie viele Fehler ich ... Das sieht man ja dann, wenn man sich wieder anschaut und wenn die anderen einem das auch noch sagen, wo man Fehler macht: Das ist nicht so lustig.»*
- *«Und dann schon Prozessberatung (für die Nachhaltigkeit), also, sich getrauen, jemanden kommen zu lassen. Oberstufenschulen sind nicht die, welche gerne Leute haben, Fremde im System. Das ist ja auch positiv, weil die machen dann Sachen selber, kann man hoffen. Aber, wenn es jetzt zum Beispiel wichtig ist, dass sie sich das leisten.»*

4.5.2 Allgemeine Zufriedenheit mit der Umsetzung der Massnahme

Alle an der Befragung teilnehmenden Lehrpersonen wurden darüber hinaus nach ihrer Zufriedenheit gefragt, wie die Massnahme Beratungen an ihrer Schule umgesetzt wurde. Die Mehrheit der Lehrpersonen beider Fächer sieht sich als «weder zufrieden noch unzufrieden» (Deutsch 60.0 Prozent; Mathematik 56.0 Prozent; vgl. Abbildung 44). Weiterhin äussern sich die Lehrpersonen beider Fächer zu je etwa gleichen Teilen als «eher unzufrieden» oder «eher zufrieden». Im Fach Mathematik gibt jeweils eine Lehrperson an, «sehr zufrieden» beziehungsweise «sehr unzufrieden» mit der Umsetzung der Beratungen zu sein.

Abbildung 44. Allgemeine Zufriedenheit mit der Durchführung der Beratungen im Rahmen des Pilotprojekts nach Angaben der Lehrpersonen



Anmerkungen. Deutsch $N = 20$, fehlend: $N = 6$ (befragt: alle Lehrpersonen, die angaben das Fach Deutsch im Rahmen von ALLE zu unterrichten); Mathematik $N = 25$, fehlend: $N = 7$ (befragt: alle Lehrpersonen, die angaben das Fach Mathematik im Rahmen von ALLE zu unterrichten).

Betrachtet man nur die Angaben der Lehrpersonen, die berichten, (mindestens) eine Beratung in Anspruch genommen zu haben, zeigt sich eine Deutsch-Lehrperson als «eher zufrieden» und zwei Deutsch-Lehrpersonen als «weder zufrieden noch unzufrieden». Im Fach Mathematik urteilt eine Lehrperson, mit der Beratung «eher zufrieden» gewesen zu sein. Von je einer Lehrperson pro Fach wird keine Angabe zu ihrer Zufriedenheit gemacht.

Neben den Lehrpersonen wurden auch die Schulleitungen gefragt, wie sie die Durchführung der Beratungen an ihrer Schule bewerten. Die meisten der elf Befragten sehen diese als «gut» bis «sehr gut» gelungen an. Zwei Schulleitungen bewerten die Umsetzung an ihrer Schule als «befriedigend», eine Schulleitung als «schlecht»:

- Sehr gut: 4 Schulleitungen
- Gut: 4 Schulleitungen
- Befriedigend: 2 Schulleitungen
- Schlecht: 1 Schulleitung
- Sehr schlecht: keine Schulleitung

4.5.3 Überlegungen der Beratungspersonen zum Beratungserfolg und zu Optimierungen

Die Interviews mit den Beratungspersonen wurden zusätzlich daraufhin ausgewertet, inwieweit die Beratungen als erfolgreich wahrgenommen und welche Möglichkeiten zur Optimierung des Angebots geäußert wurden.

Eher unsichere beziehungsweise verhaltene Einschätzungen in Bezug auf eine grundlegende und längerfristige Wirkung der durchgeführten Beratungen geben drei der Beratungspersonen ab:

- *«Bei der einen Schule nicht, weiss ich nicht, was sie nachher damit gemacht haben. Meistens geht es ja dann um so Beratungsgeschichten, wo man irgendwie Ideen kreiert, wie man etwas machen kann. Und das ist dann anregend während des Coachings, ob sie es dann tatsächlich machen, das weiss ich nicht.»*
- *«Das weiss ich auch nicht, ob sie das dann auch weiterführen, oder? Man könnte das ja quasi als ein Qualitätsinstrument anschauen, das man regelmässig macht. [...]. Das weiss ich nicht, ob das jetzt eine Schule, also ich weiss, dass viele Schulen es machen, aber ob sie quasi Qualität weiterhin überprüfen, indem sie sich auf Video anschauen, Rückmeldungen geben, das weiss ich nicht.»*
- *«Ja, also da, [...] wo wir zum Beispiel Videos angeschaut haben. Das ist natürlich hochinteressant gewesen [...] nach jeder Runde sagt die Lehrperson, das hat es mir gebracht oder eure Rückmeldungen haben bei mir zu dem geführt. Also, insofern hat es nach jedem Video ein kleines Feedback gegeben von der Lehrperson an die anderen, die ihr eine Rückmeldung gegeben haben [...] und da passiert auch etwas, und das ist auch total interessant, aber das reicht ... Ja, das ist einfach schwierig, dass das neben dem übrigen Alltag dann auch noch, ja, bedeutungsvoll wird.»*

Eine andere Beratungsperson berichtet von einer geringen Nachfrage bei Massnahmen, die sie als «hoch wirksam» beschreibt. Sie betont in diesem Zusammenhang den Bedarf nach Professionalisierung und einer stärkeren Etablierung der Massnahmen an den Schulen:

«Ich weiss, dass viele Schulen das [Anm.: Coaching/Beratungen(?)] machen, also das eingeführt haben, aber wenn man so ein wenig die Tiefenstrukturen genauer anschauen würde ... Ich glaube, in einer Schule habe ich dann angeregt, dass sie mal Videoaufnahmen von den Coachings machen, oder in zwei Schulen, und dann haben wir miteinander die Videos in Gruppen angeschaut. Das ist aber jetzt etwas, das schwach nachgefragt wurde, wo aber wieder ..., wo man sehr genau weiss, das wäre hoch wirksam. [...] Aber, dass das so [...] als professioneller Akt eingeführt würde [...] glaube ich, da sind wir noch ziemlich weit weg. Das hätte ich mir bei der Konzeption vom ALLE-Projekt eben ein wenig gewünscht, dass man da nochmals wirklich in den tieferen Strukturen arbeiten könnte. Das ist nicht häufig vorgekommen.»

Eine tendenziell positivere Beurteilung der Wirkung der Massnahme geht aus den nachfolgenden Äusserungen von zwei Beratungspersonen hervor. Erkennbar werden auch die Vorteile einer eher kontinuierlichen Begleitung der Schulen:

- *«[...] dort haben wir beim zweiten Coaching, haben wir Sachen gemacht, welche thematisch Bezug haben zum ersten. Da habe ich gesehen, dass das, was wir im ersten Coaching erarbeitet haben, unterdessen in eine schriftliche Form gebracht wurde und auf dem Schulserver vorhanden ist. Und sie haben es gleich gefunden, das heisst [lacht], sie haben es vermutlich zwischendurch auch mal gebraucht. Von dem her hat man also wie gesehen, ja, das liegt jetzt vor, und das ist ihnen in einem gewissen Sinne vertraut, also hat es vermutlich etwas gebracht in dem Sinn, mit dem zu arbeiten.»*
- *«Und bei den Schulen, welche jetzt im Projekt sind, wo ich zum Beispiel solche Sachen wie so ein wenig Begleitung mache, muss ich sagen, ist der Stand schon relativ gut. Also, sie haben das Verständnis, sind nicht mehr so normativ [...] unterwegs mit den Schülern oder einfach hilflos im Umgang mit Schwierigkeiten. Sondern sie haben so ein Interesse bekommen, und ich könnte mir vorstellen, an dieser Stelle hat sich wie das professionelle Interesse oder auch die Aufgabe verschoben oder erweitert.»*

Als eine weitere positive Wirkung sehen diese zwei Beratungspersonen die Verbesserung der Beziehungen zwischen Lehrpersonen und Lernenden:

- *«In diesen Fallseminaren, welche die Schwierigkeiten von den Schülern oder das Verhalten der Schüler angesprochen haben, ist die Grenze verschoben worden, wenn es glücklich gegangen ist. Die Grenze verschoben worden zwischen einem reinen <ich sage jetzt, wie er es machen soll> zu einem <hast du dir Alternativen überlegt, und was könntest du denn sonst noch machen, wenn es dir so und so geht>. Also zu einer Art prozeduralen Hilfe [...].»*
- *«Es gibt auch Beispiele von Schülern, welche – glaube ich – jetzt zum ersten Mal etwas wie [eine] Beziehung zu ihrer Lehrperson aufgebaut haben, weil sie plötzlich gemerkt haben, nur der Fakt schon, er hat Zeit für mich, er spricht mit mir, weil er das nie gemacht hat bis jetzt. [...]. Also einfach nur, nur das Interessiertsein, Sich-Zuwenden, ist schon so ein Fakt. [...] Das sind jetzt nicht dramatische Sachen, welche eine riesige Weiterbildung brauchen, wenn man das mal verstanden hat, kann man das machen.»*

Konkrete Optimierungsmöglichkeiten für das Beratungsangebot sehen die Beratungspersonen unter anderem

- in mehr Informationen für die Lehrpersonen darüber, was die Beratung beinhaltet;
- in einem stärkeren Einbezug der Lehrpersonen und ihrer Bedürfnisse in die Planung und Festlegung von Inhalten für die Beratung;
- in der Anwesenheit und Teilnahme der Schulleitung an der Beratung;

- in der Festlegung von regelmässigen Terminen beziehungsweise Nachfolgeterminen;
- in einer Kombination aus Weiterbildung und Beratung;
- in einer intensiveren gemeinsamen Auseinandersetzung mit dem Unterricht.

Eine Auswahl entsprechender Zitate aus den Interviews mit den Beratungspersonen wird nachfolgend gegeben.

Mehr Informationen über die Beratungen:

«Vielleicht müsste man es ein wenig genauer beschreiben. Also [...], dass man Schülerfälle besprechen kann oder Unterricht vorbereiten. Dass es ein wenig klarer ist, was das ist [...].»

Stärkerer Einbezug der Lehrpersonen in die Planung:

- *«Und die andere, die kleine Schule, dort haben sie so ein wenig mehr gesagt [...], aber sie hätten das Thema [...] das ist so vom Schulleiter initiiert gewesen, weil der das einfach wollte, und ich hatte dort schon den Eindruck von den Teilnehmenden her, dass das nicht wirklich ihr Bedürfnis ist.»*
- *«Von dem her wäre es ein wenig so wie irgendwie mithelfen, um zu einem guten Thema zu kommen oder sich einfach mal Zeit zu nehmen ... Also, ich weiss jetzt auch nicht, wenn das zum Beispiel an einem, wie heisst es, Konvent oder irgendwie an eine Versammlung ginge an der Schule, und mit denen zusammen 10 Minuten überlegen würde, man möchte ein Coaching machen, was wäre denn jetzt ein interessantes Thema? Ich glaube, dann würde man schon Sachen finden, die sie auch interessant fänden und bereit wären.»*

Teilnahme der Schulleitung:

«Also, was mir noch aufgefallen ist, für mich hat die Schulleitung ja auch eine zentrale Rolle. Und in meinen Weiterbildungen war die Schulleitung nicht anwesend. Und ich fände das eigentlich noch wichtig, dass die Schulleitung auch anwesend ist. Weil das, ich habe den Eindruck, das ist ein Signal. Und es gäbe so ein wenig wie [...], wir bleiben an diesem Thema dran [...].»

Festlegung von regelmässigen beziehungsweise Nachfolgeterminen:

- *«Vielleicht irgendwie so nach einem Jahr oder nach einem halben Jahr, vielleicht auch nur ein Halbtag oder so, dass man einfach so ein wenig wie dranbleibt [...].»*
- *«Ich weiss, ich weiss es nicht. Ich glaube, [...] das weiss man aus der Wissenschaft, die kurzfristigen Angebote haben sowieso wenig Wirkung. Jetzt wenn man [...] in einer Schule sagen könnte, wir vereinbaren jetzt wirklich mal, ich sage jetzt etwas, sechs Sitzungen, und wir schauen von jeder Person mindestens zwei Videos an und besprechen diese.»*

Kombination von Weiterbildung und Beratung:

«Und, dass das so ein wenig eine Mischform wäre zwischen Weiterbildung und Beratung, weil ich eben das Gefühl habe, die Beratung alleine, das wird irgendwie ... [...], dass es meistens nicht gut ankommt, wenn die Lehrpersonen gleich zuerst selbst etwas machen müssen. Also, man muss immer zuerst irgendwie als Weiterbildner etwas liefern, und dann kann man in den Austausch kommen. Also, ich könnte mir so eine Kombination noch vorstellen.»

Intensivere gemeinsame Auseinandersetzung mit dem Unterricht:

- «Und dann denke ich, müsste es in eine Richtung, sage ich jetzt mal, videobasiertes Coaching oder so, gehen. Damit man wirklich den Unterricht mal anschauen kann. Weil von dem, was sie jeweils erzählen, oder wenn ich nur die Texte der Schüler sehe, weiss ich immer noch nicht, was im Unterricht passiert ist.»
- «Jawohl, also eben das, was ich schon erwähnt habe, dass man Filmmaterial zur Verfügung stellen würde. [...] Ihr filmt eine Lektion, welche wir abgesprochen haben, und gebt sie mir danach, und ich schneide Stellen raus, welche wir besprechen. Wenn man das wie als Element reinnehmen könnte, würde das sicher einen Gewinn bringen, weil man mehr am Unterricht selbst sein kann. Die Gefahr ist, [...] dass man sonst immer sehr auf einer allgemeineren Ebene bleibt und eigentlich ... wie setze ich das jetzt um, was heisst das in dieser Stunde, in diesen fünf Minuten. Das man dort immer ausweichen kann, wenn man so allgemein redet [...].»

4.6 Einsatz von EMU

Eine weitere im Pilotprojekt eingesetzte Massnahme waren die «Evidenzbasierten Methoden der Unterrichtsdiagnostik und -entwicklung», kurz EMU. Orientiert an Kriterien der Lernwirksamkeit unterstützt EMU den datenbasierten professionellen Austausch von Lehrpersonen. Ziel ist es, Stärken und Schwächen des eigenen Unterrichts zu erkennen und diesen auf der Grundlage von Feedback der Kolleginnen und Kolleginnen sowie der Schülerinnen und Schüler weiterzuentwickeln. In den folgenden Kapiteln wird auf diese Fragestellungen zum Einsatz von EMU eingegangen:

- Wie unterstützt die Schulleitung den Einsatz von EMU? Wie häufig und auf welche Art und Weise wird EMU von den Lehrpersonen eingesetzt? (Kap. 4.6.1)
- Welche Wirkungen im Unterricht werden durch den Einsatz von EMU wahrgenommen? (Kap. 4.6.2)
- Wie zufrieden sind die Lehrpersonen und die Schulleitungen mit der Umsetzung der Massnahme EMU an ihrer Schule? (Kap. 4.6.3)

Da EMU fachübergreifend angelegt ist, wurden die Fragen nicht bezogen auf einzelne Fächer gestellt. Entsprechend wird im Folgenden keine Trennung zwischen den Deutsch- und Mathematik-

Lehrpersonen vorgenommen. Eine Ausnahme stellt die Frage an die Lehrpersonen zur allgemeinen Zufriedenheit mit der Massnahme in Kapitel 4.6.3 dar.

4.6.1 Umsetzung von EMU

Rahmenbedingungen seitens der Schulleitung

Wichtig für die Implementation von EMU an den ALLE-Schulen war es, dass die Schulleitungen zur Schaffung der erforderlichen Rahmenbedingungen beitragen (z.B. für die von EMU vorgesehenen kollegialen Hospitationen). In der Befragung der Schulleitungen sollten diese angeben, inwieweit sie fünf verschiedene Möglichkeiten zur Unterstützung des Einsatzes von EMU an ihrer Schule nutzen. Die Ergebnisse werden in Tabelle 34 dargestellt.

Tabelle 34. Nutzung von Möglichkeiten zur Unterstützung von EMU durch die Schulleitung nach eigenen Angaben

Unterstützungsmöglichkeiten	Anzahl	in Prozent
Organisation von Vertretungslösungen bei kollegialer Hospitation*	5	50.0
Eingabe von Schülerfeedback durch die SuS selbst*	4	40.0
Videografie und Selbstreflektion des eigenen Unterrichts in einer Fachgruppe (virtuelle Hospitation)*	5	50.0
Übernahme von Unterrichtsstunden, um Lehrpersonen Hospitationen zu ermöglichen	1	9.1
Verankerung von EMU im Schulprogramm	9	81.8

Anmerkungen. SuS = Schülerinnen und Schüler. $N = 11$; $*N = 10$, fehlend: $N = 1$. Antwortformat pro Item: ja – nein.

Weit verbreitet war demnach vor allem die Verankerung von EMU im Schulprogramm. Jeweils (rund) die Hälfte der Schulleitungen gibt an, dass sie an ihrer Schule Vertretungslösungen bei kollegialen Hospitationen organisiert, virtuelle Hospitationen in einer Fachgruppe etabliert und die Eingabe von Schülerfeedback durch die Lernenden selbst ermöglicht hat. Von einer Übernahme von Unterrichtsstunden, um den Lehrpersonen Hospitationen bei Kolleginnen und Kollegen zu ermöglichen, berichtet nur eine einzige Schulleitung.

Wie zentral die Unterstützung der Lehrpersonen durch die Schulleitung ist, um EMU gewinnbringend einsetzen zu können, betont diese Schulleitung:

«EMU ist natürlich ... Ich merke, dass dies für die Lehrpersonen ein zusätzlicher Aufwand ist. Wenn sie es machen, finden sie es aber gut. Es ist etwas, woran man als Schulleitung etwas dranbleiben und einfordern muss. Es ist nicht so, dass ich finde, klar, cool, machen wir EMU.»

Eine weitere Schulleitung beschreibt anders gelagerte Schwierigkeiten, Akzeptanz für den Einsatz von EMU zu schaffen:

«Dann EMU – da ist der Widerstand ein bisschen grösser gewesen, weil die Lehrpersonen zwar durchaus offen gewesen sind für Rückmeldungen und Kritik und und und ... aber nicht gewillt gewesen sind, auf Kritik von Schülerinnen und Schülern oder Eltern einzugehen. Da war unser Team wirklich zu verbohrt. Es ist mir auch bis zum Schluss nicht wirklich gelungen, eine offene Fehlerkultur daraus zu implementieren.»

Häufigkeit und Art des Einsatzes

Die Lehrpersonen wurden gefragt, wie häufig sie verschiedene Bestandteile von EMU in den der Befragung vorausgehenden beiden Jahren eingesetzt haben. Die Ergebnisse zeigt Tabelle 35.

Tabelle 35. Häufigkeit des Einsatzes von EMU in den der Befragung vorausgehenden beiden Jahren nach Angaben der Lehrpersonen

Häufigkeit	Kollegiale Hospitation		Feedback durch hospitierende Kolleg/-innen		Feedback durch Schüler/-innen		Summe eingesetzter EMU-Bestandteile	
	Anzahl LP	in Prozent	Anzahl LP	in Prozent	Anzahl LP	in Prozent	Anzahl LP	in Prozent
0	4	7.4	5	9.3	2	3.7		
1	9	16.7	8	14.8	9	16.7	2	3.7
2	20	37.0	17	31.5	14	25.9	2	3.7
3	14	25.9	15	27.8	11	20.4	3	5.6
4	3	5.6	5	9.3	11	20.4	1	1.9
5	2	3.7	1	1.9	2	3.7	4	7.4
6					1	1.9	10	18.5
7							2	3.7
8	2	3.7	2	3.7	1	1.9	9	16.7
9							10	18.5
10			1	1.9	2	3.7	1	1.9
12							5	9.3
14							1	1.9
≥ 20					1	1.9	4	7.4
<i>M</i>	2.4		2.6		3.2		8.2	
<i>SD</i>	1.6		1.9		3.0		4.8	

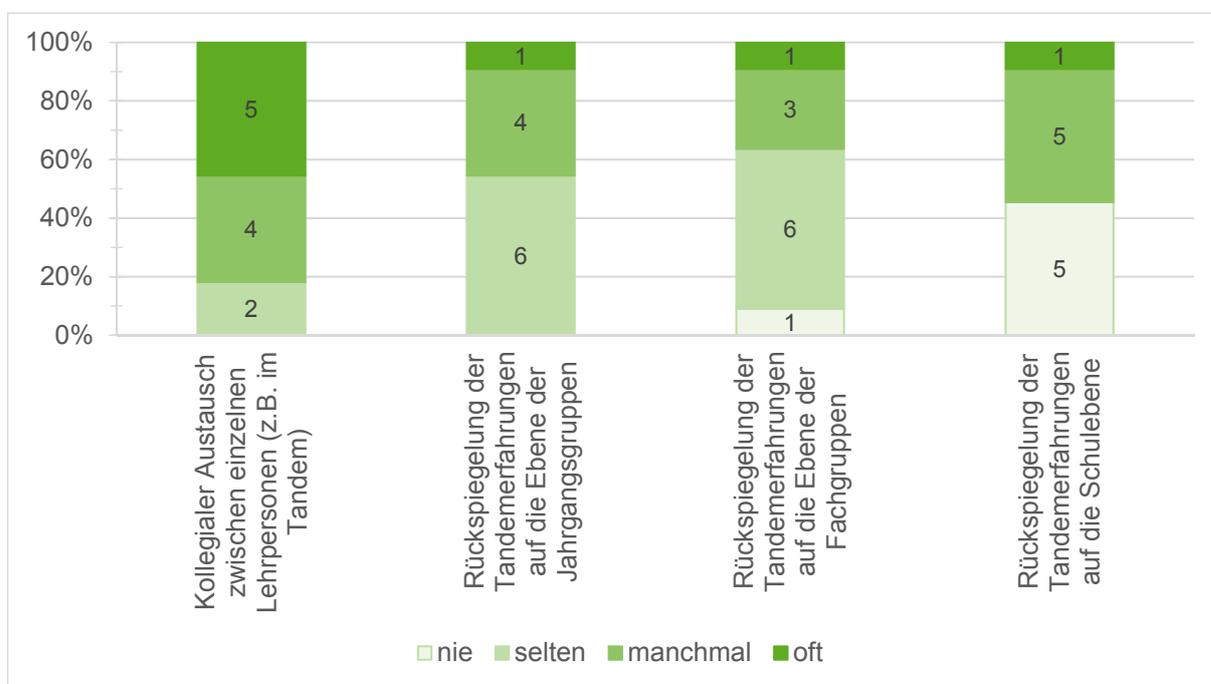
Anmerkungen: LP = Lehrpersonen. N = 54.

Unterrichtsbesuche bei anderen Lehrpersonen (Kollegiale Hospitationen) wurden in den letzten beiden Jahren meist zwei- bis dreimal durchgeführt. Vier Lehrpersonen (7.4 Prozent) geben an,

keine Hospitationen durchgeführt zu haben, sieben Lehrpersonen (13.0 Prozent) berichten von mehr als drei Unterrichtsbesuchen. Feedback durch hospitierende Kolleginnen und Kollegen wurde ebenfalls meist zwei- bis dreimal eingeholt. Kein Feedback eingeholt haben fünf Lehrpersonen (9.3 Prozent), andererseits berichten neun Lehrpersonen (16.7 Prozent), mehr als dreimal Feedback erhalten zu haben. Etwas häufiger war Feedback durch die Schülerinnen und Schüler zum eigenen Unterricht. Zwei Drittel der Befragten (66.7 Prozent) gibt an, Schülerfeedback zwei- bis viermal angefordert zu haben, einen völligen Verzicht darauf berichten nur zwei Lehrpersonen. Summiert man individuell auf, wie häufig EMU-Bestandteile eingesetzt wurden, zeigt sich, dass alle Lehrpersonen zumindest eines der Elemente mindestens einmal anwendeten (vgl. rechte Tabellenspalte). Drei Lehrpersonen (5.6 Prozent) geben an, nur Feedback der Schülerinnen und Schülern eingeholt zu haben. Alle restlichen Lehrpersonen berichten, mindestens zwei der drei Elemente angewendet zu haben. Die Mehrheit der Lehrpersonen (57.4 Prozent) berichtet von einem sechs- bis neunmaligen Einsatz von EMU-Elementen. Ein Einsatz von durchschnittlich weniger als zwei Elementen pro Jahr (Summe < 4) findet sich bei sieben der Lehrpersonen (13.0 Prozent). Demgegenüber ergibt sich für alle drei Bestandteile von EMU in einzelnen Fällen ein sehr gehäufter Einsatz.

Die Schulleitungen wurden dazu befragt, inwiefern im Rahmen des Pilotprojekts ein regelmässiger Erfahrungsaustausch zu EMU stattgefunden hat. Die Ergebnisse zeigt Abbildung 45.

Abbildung 45. Formen des Erfahrungsaustauschs zu EMU nach Angaben der Schulleitungen



Anmerkungen. N = 11.

Die Aussagen der Schulleitungen deuten darauf hin, dass ein Erfahrungsaustausch zu EMU eher wenig oder gelegentlich stattgefunden hat. Am häufigsten kam es noch zu einem kollegialen Austausch zwischen einzelnen Lehrpersonen (*oft*: 45.5 Prozent). Die Rückspiegelung der Tandemerfahrungen auf die Ebenen der Jahrganggruppen, der Fachgruppen oder der Schule erfolgte dagegen eher sporadisch. Für die Einschätzungen zur Rückspiegelung auf die Schulebene fällt auf, dass sich die Mehrzahl der Befragten in zwei gleich grosse Gruppen aufteilt (je 45.5 Prozent), von denen eine von gelegentlichem (*manchmal*), die andere von keinem Austausch zu EMU berichtet.

Weitere Einblicke zur Umsetzung von EMU an den Schulen lassen sich aus den nachfolgenden Äusserungen der Lehrpersonen und Schulleitungen aus den Interviews gewinnen.

Kollegiale Hospitation und Feedback:

Äusserungen der Lehrpersonen

- *«Man hatte den Auftrag von der Schulleitung, das jahrgangsweise zu organisieren. Und jeder Jahrgang hat dies wieder anders gemacht.»*
- *«Und wir haben das gemacht, in unserer Lernlandschaft. Glaube dieses Mal Anfang Jahr so. Und wir sagen jedes Mal, wenn wir so eine Runde gemacht haben: Eigentlich sollte man es mehr machen.»*
- *«Bei uns läuft das kollegiale Feedback vor allem über das Teamteaching [...]. Wir haben ganz viele Momente, in denen wir zu zweit oder zu dritt sogar arbeiten, mit zwei oder mit drei Klassen.»*
- *«Wir haben so ein Rundherum – wir waren fünf gewesen, also zu fünft dieses Jahr – und haben so einen Rundherumbesuch gemacht im Unterrichtsbesuch. Und das Spannende war, dass wir auch unsere Klassen bei jemand anderem in einem anderen Fach gesehen haben. Das war mega spannend.»*
- *«Wir haben uns gegenseitig im Unterricht besucht. Wir haben das mal zusammen gemacht, sie ist zweimal zu mir gekommen und ich kam zweimal zu ihr.»*
- *«Das war halt das erste Jahr, wo es wirklich [Anm.: unverständlich] also bewusst. Da haben wir uns auch wirklich besucht und uns Rückmeldungen gegeben.»*

Äusserungen der Schulleitungen

- *«Schlussendlich haben wir das EMU so abgewandelt, dass es im Prinzip Hospitation ist. Dieses Jahr haben sie Videohospitationen gemacht. Sie haben sich selber im Unterricht aufgenommen und ein Teil aus deiner Aufnahme ausgewählt, den sie den Kollegen im Jahrgangsteam zeigen möchten. Im Jahrgangsteam hat man dann miteinander alle diese Sequenzen angeschaut und das besprochen. Das haben alle sehr gut gefunden.»*

- *«Zudem haben wir noch gesagt [...] wir haben ganz viele iPads an unserer Schule: Es ist auch eine Möglichkeit, mal das iPad für solche Videosequenzen und Tonsprechanteile von Lehrpersonen einzusetzen. So kann man mit neuen Möglichkeiten und neuen Medien Rückmeldungen geben. Dies wurde auch so umgesetzt und wir erhielten gute Rückmeldungen.»*
- *«Wir haben verbindlich festgelegt, dass wir pro Jahr eine Unterrichtshospitation durchführen und pro Jahr ein Schülerfeedback, alternierend mit Elternfeedback. [...]. Jetzt hat sich das mit EMU, mit ALLE, noch einmal intensiviert.»*
- *«Ja, nicht permanent, weil das ist eine Überforderung. Eine absolute Überforderung. Das ist nicht machbar im gesamten Schulbetrieb, aber man kann so einzelne Elemente von EMU fokussieren und punktuell angehen. Das funktioniert.»*

Die Zitate belegen, dass die Möglichkeit zur kollegialen Hospitation und zur Gewinnung von Feedback durch Kolleginnen und Kollegen als sehr gewinnbringend erlebt werden konnte (siehe auch Kap. 4.6.2). Es deutet sich dabei an, dass die Umsetzung in den Schulen, in den Kollegien (z.B. «in unserer Lernlandschaft», «über das Teamteaching», «Rundherumbesuch») und selbst zwischen Jahrgängen («jeder Jahrgang hat dies wieder anders gemacht») unterschiedlich organisiert war. Die Äusserungen der Schulleitungen zeigen auf, wie neuere technische Möglichkeiten genutzt wurden, um zum Beispiel Hospitationen auch ohne einen direkten Besuch in der Klasse zu ermöglichen («Videohospitationen») und gezielt Feedback einzuholen («ein Teil [der] Aufnahme ausgewählt, den sie den Kollegen im Jahrgangsteam zeigen möchten»). Die Äusserung der letzten zitierten Schulleitung verweist auf ein mögliches Hindernis für eine noch intensivere Nutzung: eine erlebte «Überforderung», zumindest bei «permanent[em]» Einsatz.

Feedback durch die Schülerinnen und Schüler (Äusserungen der Lehrpersonen):

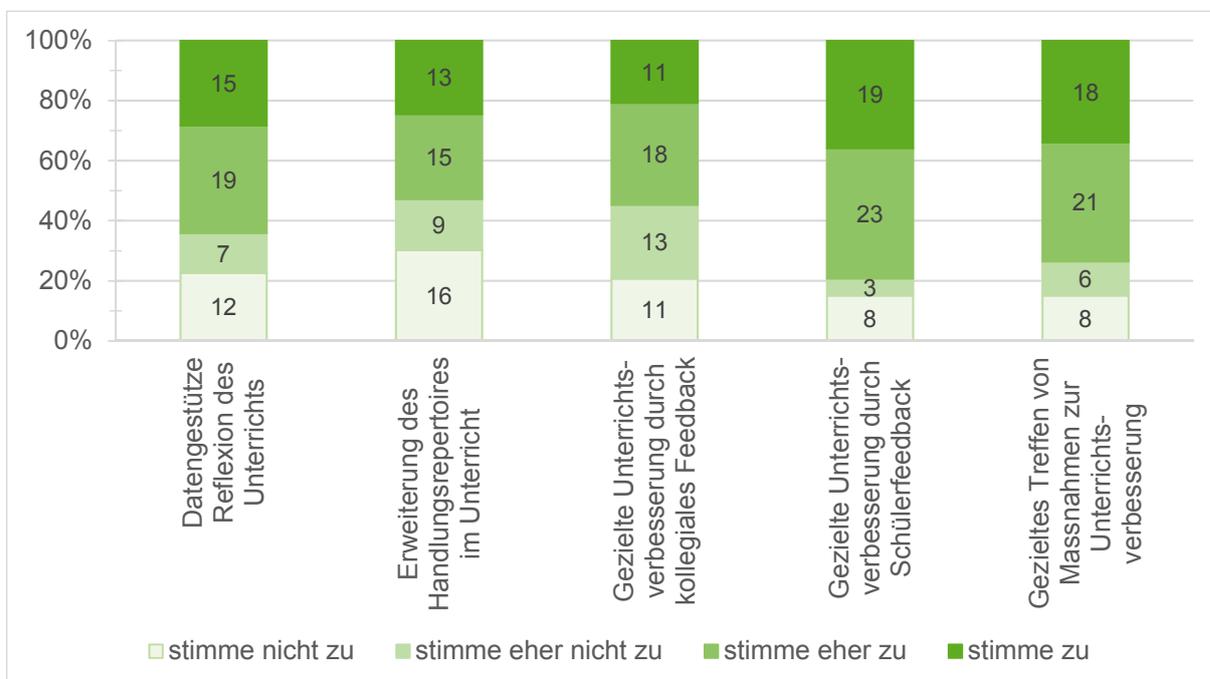
- *«Also ich habe es, glaube ich, ein oder zweimal eingesetzt – die Fragebögen –, um Rückmeldungen über den eigenen Unterricht zu bekommen.»*
- *«Schülerfeedback haben wir sowieso halbjährlich drin. Das ist bei uns im Lernbuch verankert.»*
- *«Ich mache solche Feedbackrunden regelmässig seit fünf bis sechs Jahren. Das auch in verschiedenen Fächern. Ich sehe dort auch, mögen sie mich oder mögen sie meinen Unterricht nicht. Je mehr Vertrauen sie zu einem haben, also irgendwie in der 2. oder 3. Sek, desto nützlicher sind dann diese Feedbacks.»*
- *«Also wenn man es das erste Mal mit den Schülern macht, finden sie es cool. Und je öfter man es macht desto mehr finden sie es ... [...] Wir haben uns nur noch durchgeklickt, es ist immer das Gleiche, es ist nicht spannend.»*

Zum Schülerfeedback resultieren insgesamt weniger Aussagen als zur kollegialen Hospitation beziehungsweise zum kollegialen Feedback. Die zitierten Äusserungen zeigen, dass Lehrpersonen dieses Instrument zum Teil bereits vor ALLE intensiv genutzt haben (ggf. auch als auf Schulebene etablierte Massnahme; vgl. die oben zitierte Äusserung einer Schulleitung: «Wir haben verbindlich festgelegt [...] pro Jahr ein Schülerfeedback»). Aus der Aussage einer Lehrperson wird deutlich, dass der erlebte Wert des Feedbacks auch von Voraussetzungen seitens der Schülerschaft abhängen kann («Je mehr Vertrauen sie zu einem haben [...], desto nützlicher sind dann diese Feedbacks»). Die abschliessende Aussage weist darauf hin, dass das Instrument beim wiederholten Einsatz auch an Reiz verlieren kann – möglicherweise als Folge einer zu wenig fokussierten oder zu routinemässigen Verwendung, was sich für diesen Fall jedoch nicht sicher feststellen lässt.

4.6.2 Wahrgenommene Wirkungen von EMU

In der Lehrpersonenbefragung wurde die Wirkung von EMU auf verschiedene Aspekte des unterrichtlichen Handelns erfragt. Unterscheiden lassen sich dabei Wirkungen von EMU auf den eigenen Unterricht, auf die eigene Wahrnehmung als Lehrperson und auf den Schulalltag im Allgemeinen. Die Ergebnisse finden sich in Abbildung 46 bis Abbildung 48.

Abbildung 46. Wirkungen von EMU auf den eigenen Unterricht nach Angaben der Lehrpersonen



Anmerkungen. N = 53, fehlend N = 1.

Wirkungen von EMU auf den eigenen Unterricht (vgl. Abbildung 46) sehen die Lehrpersonen vor allem in der gezielten Verbesserung des eigenen Unterrichts durch das Feedback ihrer Schülerinnen und Schüler (*stimme eher zu* oder *stimme zu*: 79.2 Prozent) und die Möglichkeit, gezielte

Massnahmen zur Unterrichtsverbesserung zu treffen (*stimme eher zu* oder *stimme zu*: 73.6 Prozent). Bei den weiteren Aspekten teilen sich die Befragten jeweils in zwei etwa gleich grosse Gruppen, die sich (eher) zustimmend beziehungsweise (eher) ablehnend äussern.

Die Schulleitung einer Schule, in der Unterrichtshospitationen und Feedback bereits vor ALLE etabliert waren, berichtet, wie EMU genutzt werden konnte, um die Umsetzung der Massnahmen noch mehr zu fokussieren:

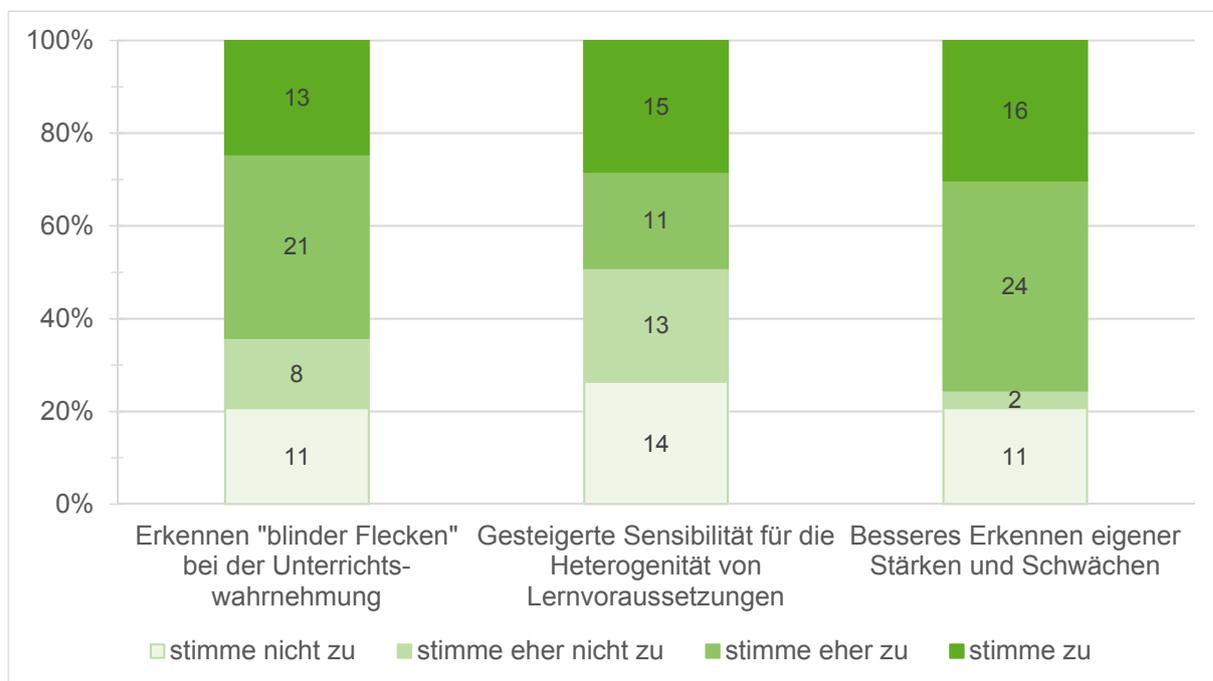
«Wir haben gesagt, ihr lasst Unterrichtsqualitätsmerkmale von EMU bewusst in die Unterrichtshospitation einfliessen. Das heisst, ihr könnt nicht einfach allgemein Rückmeldung geben, sondern ihr müsst von EMU her Unterrichtsqualitätsmerkmale aussuchen und im Hinblick auf diese Merkmale diese Hospitation abwickeln. Sie müssen also genau auf diese Unterrichtsqualitätsmerkmale Feedback geben.»

Diese Lehrpersonen heben hervor, dass der erlebte Gewinn durch Unterrichtsbesuche und das Feedback von Kolleginnen und Kollegen auch personen- und kontextabhängig ist:

- *«Aber was ich finde, so der gegenseitige Unterrichtsbesuch ist halt bei uns einfach ein wenig ... Jetzt ist er wieder stufenübergreifend dieses Jahr, und ich frage mich dann manchmal schon: Bringt mir das viel, wenn ich jetzt bei einer Erstklass-Lehrperson schauen gehe? Also ich kann für meinen Unterricht dann nicht sehr viel mitnehmen.»*
- *«Das haben wir auch gemacht. Aber es sind auch ... eben je nachdem wer einen besuchen kommt, wie man miteinander kommuniziert, weiss man schon Einiges, und es ist nicht, dass durch das neue Tatsachen eigentlich hervorkommen.»*

Bezüglich der Verbesserungen der eigenen Wahrnehmung (vgl. Abbildung 47) berichtet jeweils die klare Mehrheit der Lehrpersonen von einem besseren Erkennen eigener Stärken und Schwächen (*stimme eher zu* oder *stimme zu*: 75.5 Prozent) sowie vom Erkennen «blinder Flecken» bei der Wahrnehmung des eigenen Unterrichts (*stimme eher zu* oder *stimme zu*: 64.2 Prozent). Das Urteil, ob durch EMU die Sensibilität für heterogene Lernvoraussetzungen gesteigert wurde, fällt dagegen recht unterschiedlich aus, mit je etwa gleichen Graden der Zustimmung und Ablehnung.

Abbildung 47. Wirkungen von EMU auf die eigene Wahrnehmung im Unterricht nach Angaben der Lehrpersonen



Anmerkungen. N = 53, fehlend: N = 1.

Wie Unterrichtsfeedback dazu beiträgt, die eigenen «blinden Flecken» besser zu erkennen, beschreiben zwei Lehrpersonen wie folgt:

- *«Ich schätze das sehr, wenn ich bei den Kollegen mal in den Unterricht reinschauen kann. Das ist neuer Input. Man erhält neue Kritik, wo man selber vielleicht nicht mehr so darauf achtet.»*
- *«Ich habe auch darüber gestaunt, wie präzise Schüler beurteilen können, fair, aber präzise, und sie haben die blinden Flecken, die berühmten, haben sie gesehen, und sie haben dann das Kreuz schon am richtigen Ort gemacht. Das habe ich sehr aufschlussreich gefunden.»*

Wie dies auch relativ alltägliche Aspekte betreffen kann, zeigt das Beispiel dieser Lehrperson:

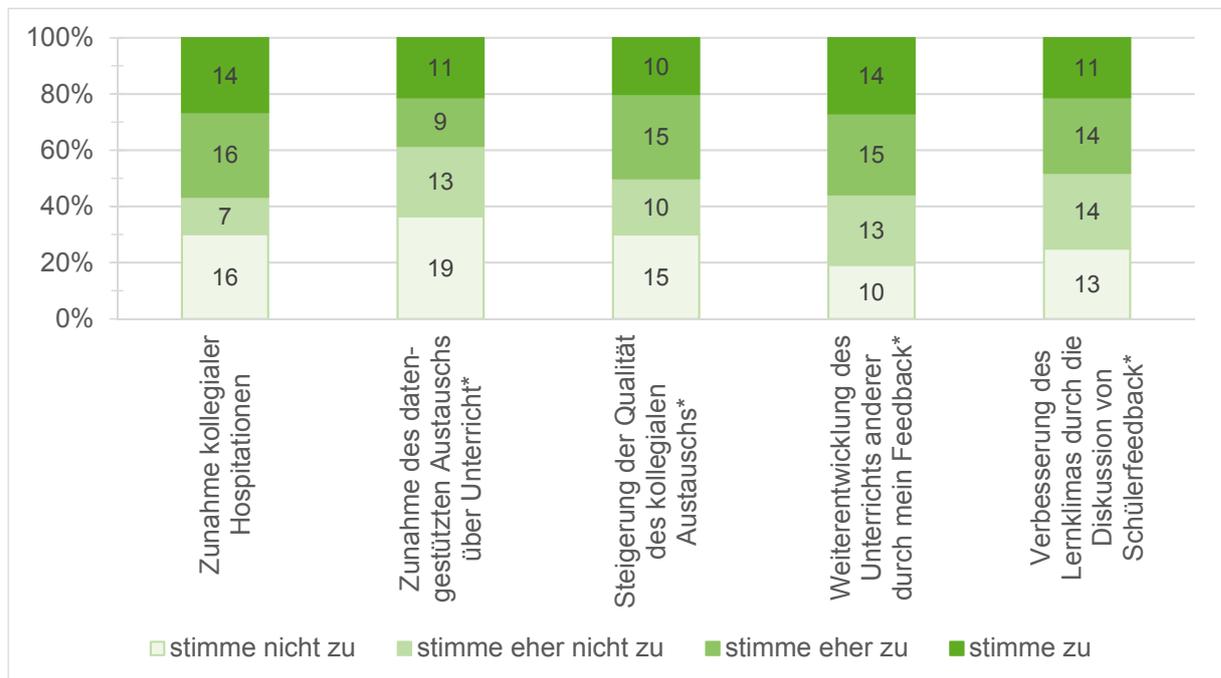
«Ich bin zum Beispiel immer so vor dem Projektor gestanden und man sah es gar nicht richtig, weil ich den Projektor blöd platziert habe. Das hat man gerade auch gemerkt, ich selber habe das gar nicht gemerkt.»

Dass der praktische Umgang mit dem Schülerfeedback auch eine Herausforderung an die Lehrpersonen darstellt, wird aus dieser Äusserung deutlich:

«Es ist wirklich herausgekommen, dass die Schüler finden, teilweise möchten sie noch mehr individueller gefördert oder gefordert werden. Und ich studiere natürlich immer wieder daran herum, an dieser Umsetzung oder suche immer wieder Gelegenheiten. Ich finde das eine grosse Herausforderung.»

Im Hinblick auf Veränderungen im Schulalltag durch EMU (vgl. Abbildung 48) deuten die Antworten der Lehrpersonen jeweils mehrheitlich auf eine Zunahme kollegialer Hospitationen (*stimme eher zu* oder *stimme zu*: 56.6 Prozent) und auf Weiterentwicklungen des Unterrichts anderer Lehrpersonen durch das eigene Feedback (*stimme eher zu* oder *stimme zu*: 55.8 Prozent) hin. Die Urteile über die Steigerung der Qualität des kollegialen Austausches und des Lernklimas fallen geteilt aus. Relativ am wenigsten zugestimmt wird einer Zunahme von datengestütztem Austausch über den Unterricht (*stimme nicht zu* oder *stimme eher nicht zu*: 61.5 Prozent).

Abbildung 48. Wirkungen von EMU auf den Schulalltag nach Angaben der Lehrpersonen



Anmerkungen. N = 53, fehlend: N = 1; *Items von links nach rechts: N = 52, fehlend: N = 2; N = 50, fehlend: N = 4; N = 52, fehlend: N = 2; N = 52, fehlend: N = 2.

Wie die kollegiale Hospitation in EMU dazu beitragen kann, die Qualität des Austausches im Kollegium zu verbessern, zeigt diese Aussage einer Schulleitung:

«Ich glaube aber, dass bei den Lehrpersonen, welche die kollegiale Hospitation schon bereits vorher geschätzt haben oder gefunden haben: das würde mir aber gut tun – dort hat EMU insofern Qualitätsveränderungen gebracht, weil es eigentlich das Erforschen des Unterrichtes in den Vordergrund gestellt hat und nicht das Kritisieren.»

Um einen Eindruck zu erhalten, wie viele der einzelnen Lehrpersonen eine Wirkung von EMU wahrnehmen, wurden abschliessend alle 13 erfragten Wirkungen von EMU gemeinsam betrachtet. Es wurde ausgezählt, für wie viele der Fragen die Lehrpersonen einer Wirkung zumindest eher zustimmen. Tabelle 36 zeigt das Ergebnis.

Tabelle 36. Anzahl der von den Lehrpersonen berichteten Wirkungen von EMU

Anzahl Wirkungen	Anzahl LP	Prozent
0 (keine Wirkung)	4	7.5
1 – 4	10	18.9
5 – 8	14	26.4
9 – 12	17	32.1
13 (alle Wirkungen)	8	15.1

Anmerkungen. LP = Lehrpersonen. $N = 53$, fehlend: $N = 1$.

Demnach berichtet nur ein geringer Anteil der Lehrpersonen (7.5 Prozent), dass EMU keine der erfragten Wirkungen erzielt hat. Fast die Hälfte der Lehrpersonen hingegen gibt an, dass durch EMU ein Grossteil der erfragten Wirkungen (≥ 9) ausgelöst wurde. Rund jede sechste Lehrperson (15.1 Prozent) sieht sogar, dass EMU sich bezüglich aller erfragten Aspekte positiv ausgewirkt hat.

4.6.3 Allgemeine Zufriedenheit mit der Umsetzung der Massnahme

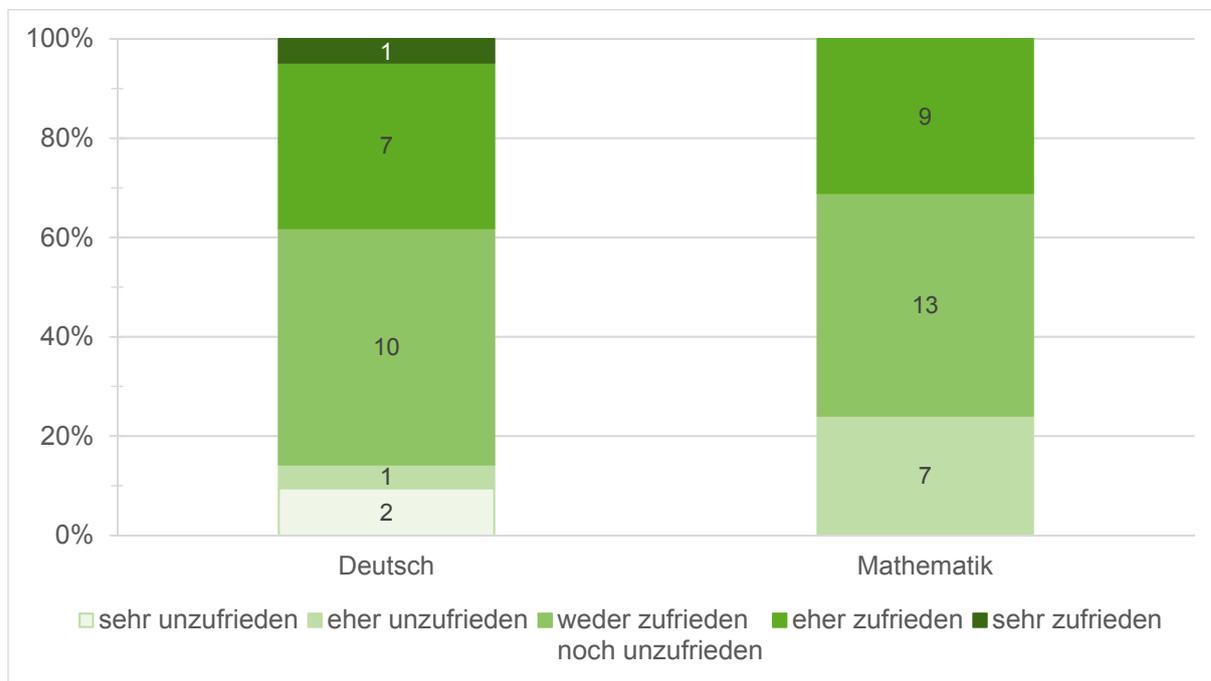
Die Lehrpersonen wurden nach ihrer allgemeinen Zufriedenheit mit der Umsetzung jeder Massnahme des Pilotprojekts an ihrer Schule gefragt. Die Ergebnisse für die Massnahme EMU sind in Abbildung 49 dargestellt.

Als eher zufrieden oder – im Fall einer Mathematik-Lehrperson – sehr zufrieden mit der Umsetzung von EMU zeigt sich je etwa ein Drittel der befragten Lehrpersonen (Deutsch 33.3 Prozent; Mathematik 31.0 Prozent). Etwas weniger als die Hälfte der Befragten beider Fächer äussert sich neutral (*weder zufrieden noch unzufrieden*: Deutsch 47.6 Prozent; Mathematik 44.8 Prozent). (Eher) unzufrieden zu sein berichten drei Deutsch-Lehrpersonen (14.3 Prozent) und sieben Mathematik-Lehrpersonen (24.1 Prozent).

Die Einschätzung der Schulleitungen zur Umsetzung der Massnahme an der eigenen Schule wurde ebenfalls erhoben. Für EMU sieht die Mehrheit der elf Befragten diese als «gut» oder zumindest «befriedigend» gelungen an:

- Sehr gut: 1 Schulleitung
- Gut: 3 Schulleitungen
- Befriedigend: 5 Schulleitungen
- Schlecht: 2 Schulleitungen
- Sehr schlecht: keine Schulleitung

Abbildung 49. Allgemeine Zufriedenheit mit der Umsetzung der Massnahme EMU im Rahmen des Pilotprojekts nach Angaben der Lehrpersonen



Anmerkungen. Deutsch $N = 21$, fehlend: $N = 5$ (befragt: alle Lehrpersonen, die angaben das Fach Deutsch im Rahmen von ALLE zu unterrichten); Mathematik $N = 29$, fehlend: $N = 3$ (befragt: alle Lehrpersonen, die angaben das Fach Mathematik im Rahmen von ALLE zu unterrichten).

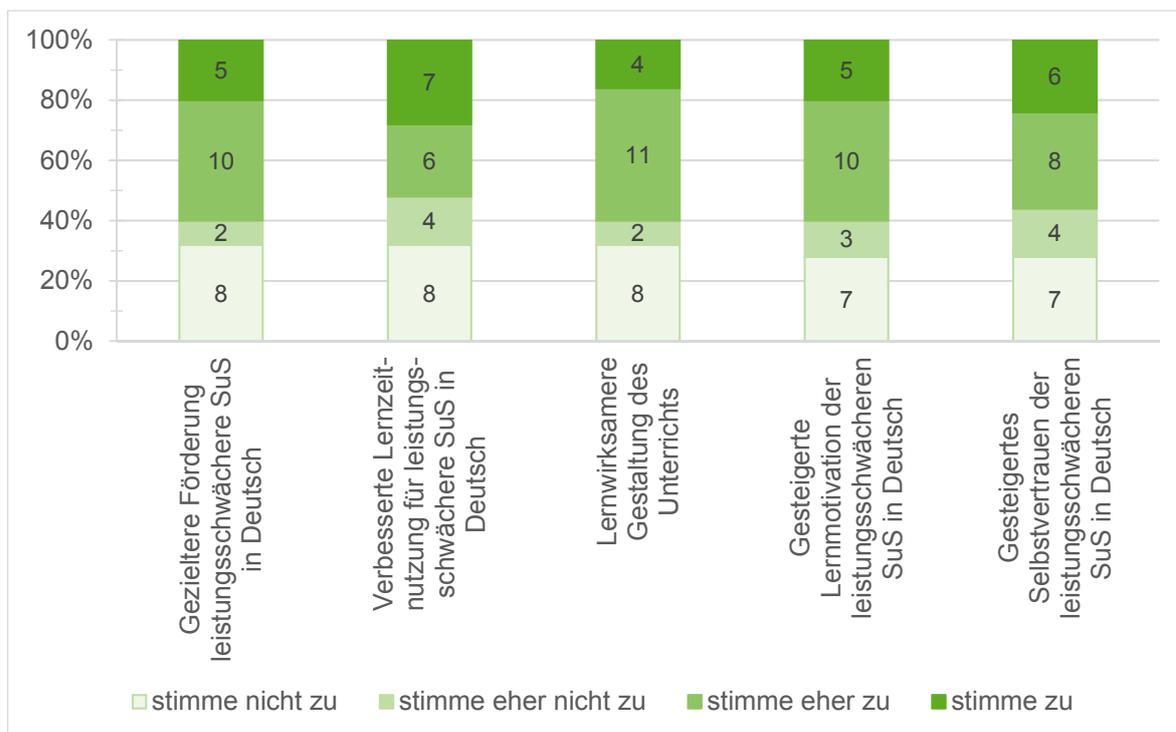
4.7 Individualisierung im Fachunterricht

Die Anpassung des Unterrichts an die Bedürfnisse von Schülerinnen und Schülern mit unterschiedlichen Lernvoraussetzungen und unterschiedlichem Lernstand stellte im Pilotprojekt einen wichtigen Ansatzpunkt zur Förderung dar. Im Rahmen der Befragungen von Lehrpersonen, Schulleitungen und Schulassistenten wurden daher Fragen zu unterschiedlichen Aspekten einer verbesserten Individualisierung des Unterrichts gestellt. Im Vordergrund standen dabei die gezielte Förderung insbesondere auch der leistungsschwächeren Schülerinnen und Schüler und die wahrgenommenen Wirkungen auf diese Schülergruppe. Nachfolgend werden zunächst die Ergebnisse aus der Befragung der Lehrpersonen berichtet, dann wird auf die Einschätzungen der Schulleitungen und der Schulassistenten eingegangen.

Die Lehrpersonen sollten Einschätzungen abgeben, inwieweit sie bezogen auf fünf verschiedene Aspekte wahrnehmen, dass das Pilotprojekt zu einer besseren individuellen Förderung vor allem leistungsschwächerer Schülerinnen und Schüler und zu positiven Wirkungen in dieser Schülergruppe geführt hat. Wie Abbildung 50 zeigt, fällt die Wahrnehmung der Deutsch-Lehrpersonen über die verschiedenen Aspekte hinweg sehr ähnlich aus: Jeweils etwa zwei Drittel der Befragten stimmen (eher) zu, dass das Pilotprojekt zu einer verbesserten individuellen Förderung (z.B. gezielte

Förderung leistungsschwächerer Schülerinnen und Schüler), lernwirksamerem Unterricht und positiven Wirkungen bei Leistungsschwächeren (z.B. Steigerung der Lernmotivation) geführt habe.

Abbildung 50. Veränderungen durch das Pilotprojekt bezogen auf die Individualisierung im Unterricht nach Angaben der Deutsch-Lehrpersonen

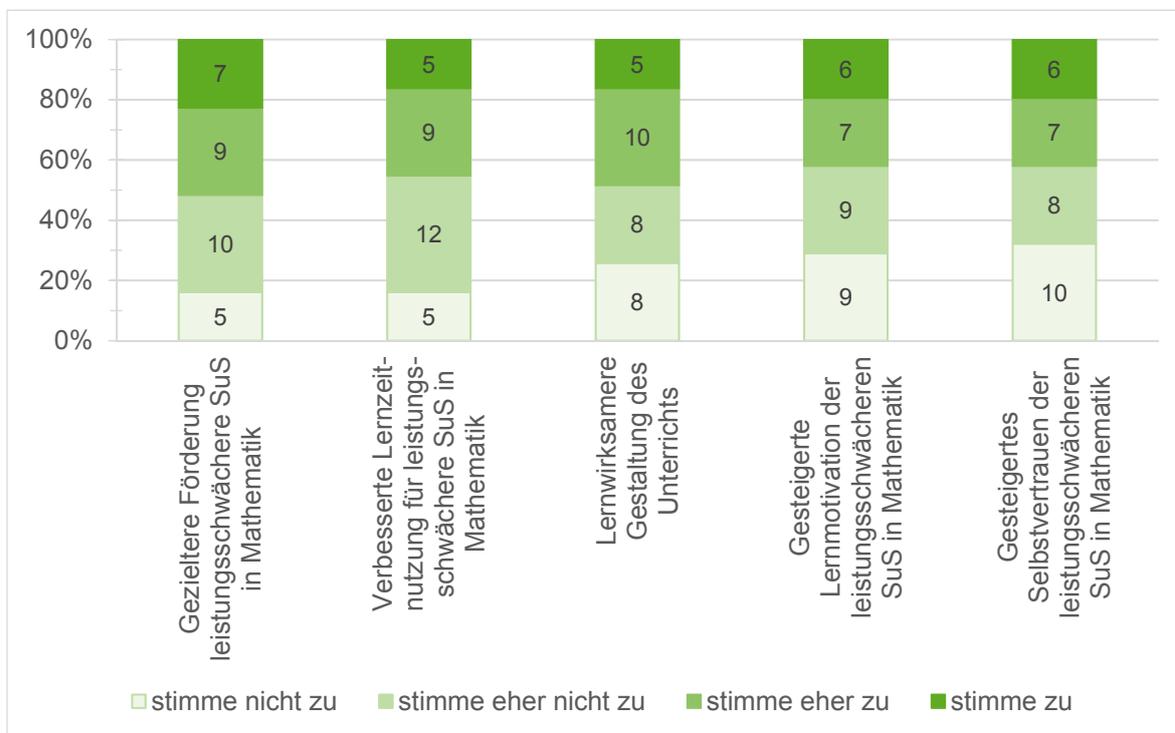


Anmerkungen. SuS = Schülerinnen und Schüler. N = 25, fehlend: N = 1.

Bei den Mathematik-Lehrpersonen fällt die Zustimmung zur Wirkung des Pilotprojekts etwas verhaltener aus (vgl. Abbildung 51). Für die erfragten Aspekte sieht je knapp die Hälfte der Lehrpersonen Veränderungen infolge des Pilotprojekts als (eher) gegeben an. Aspekte, denen relativ stärker zugestimmt wird, sind die gezielte Förderung leistungsschwächerer Schülerinnen und Schüler im Mathematikunterricht, ferner eine verbesserte Lernzeitnutzung und Lernwirksamkeit des Unterrichts. Eine höhere Lernmotivation und mehr Selbstvertrauen bei den leistungsschwächeren Schülerinnen und Schülern nehmen demgegenüber nur je rund 40 Prozent der Lehrpersonen wahr.

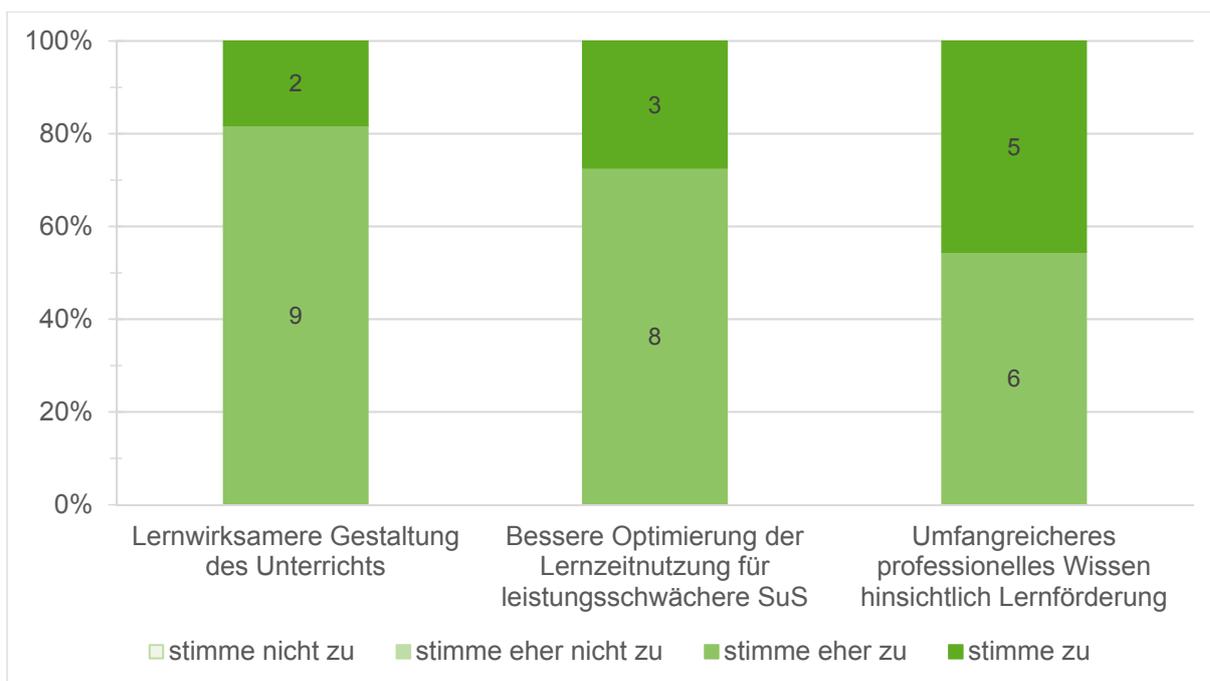
In der Schulleitungsbefragung wurde zum einen nach wahrgenommenen Verbesserungen in der Lernwirksamkeit des Unterrichts und der Lernzeitnutzung für leistungsschwächere Schülerinnen und Schüler gefragt. Zum anderen wurde erhoben, ob die Schulleitungen eine Zunahme des professionellen Wissens zur Lernförderung bei den Lehrpersonen wahrnehmen. Wie aus Abbildung 52 ersichtlich wird, fallen die Einschätzungen durchweg positiv aus: alle Schulleitungen stimmen den erfragten Veränderungen (eher) zu. Während bei der Lernwirksamkeit und der Lernzeitnutzung die etwas verhaltenerere Antwort überwiegt, stimmt knapp die Hälfte der Schulleitungen der Zunahme professionellen Wissens uneingeschränkt zu (*stimme zu*: 45.5 Prozent).

Abbildung 51. Veränderungen durch das Pilotprojekt bezogen auf die Individualisierung im Unterricht nach Angaben der Mathematik-Lehrpersonen



Anmerkungen. SuS = Schülerinnen und Schüler. N = 31, fehlend: N = 1.

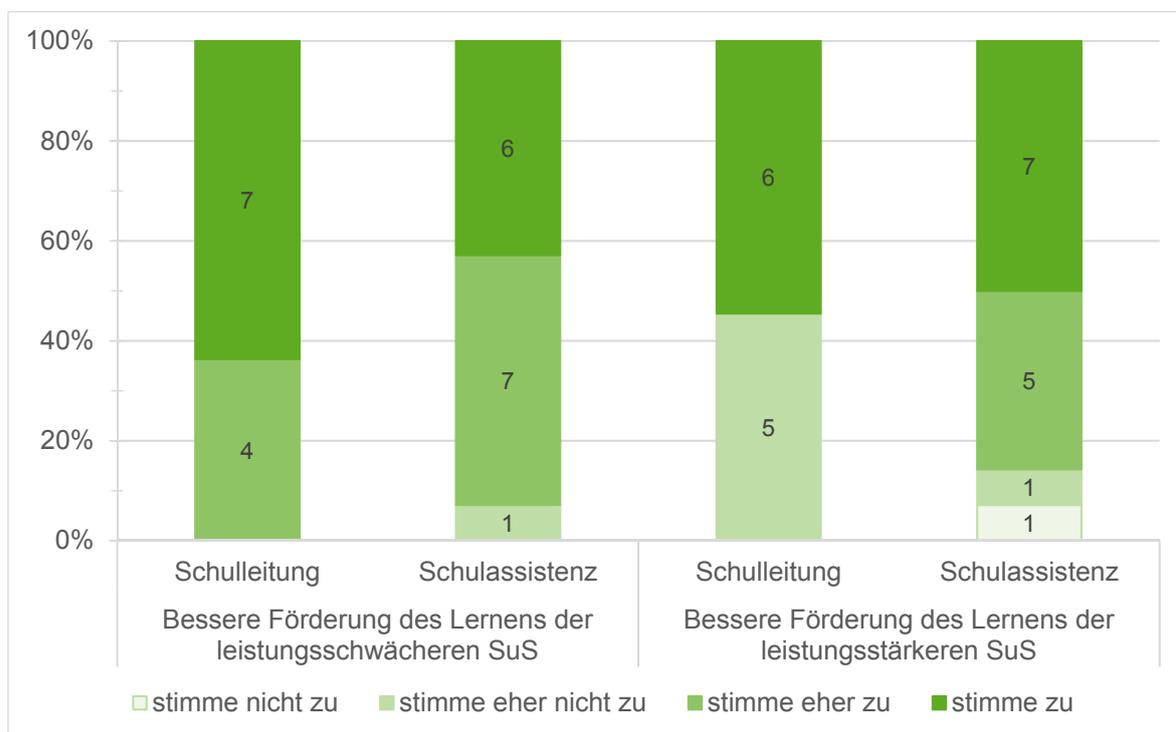
Abbildung 52. Veränderungen durch das Pilotprojekt bezogen auf die Individualisierung des Unterrichts nach Angaben der Schulleitungen



Anmerkungen. SuS = Schülerinnen und Schüler. N = 11.

Abschliessend werden die Einschätzungen der Schulleitungen sowie Schulassistenzen im Hinblick auf eine verbesserte Förderung des Lernens sowohl leistungsschwächerer als auch leistungsstärkerer Schülerinnen und Schüler durch die Lehrpersonen gegenübergestellt (vgl. Abbildung 53). Dies ermöglicht es, zwei Perspektiven darauf zu vergleichen, wie sich der Unterricht der teilnehmenden Lehrpersonen als Folge des Pilotprojekts verändert hat.

Abbildung 53. Verbesserte Förderung des Lernens leistungsschwächerer und leistungsstärkerer Schülerinnen und Schüler durch die Lehrpersonen nach Angaben der Schulleitungen und Schulassistenzen



Anmerkungen. SuS = Schülerinnen und Schüler. Schulleitungen: N = 11; Schulassistenzen: N = 14.

Während einer verbesserten Förderung des Lernens *leistungsschwächerer* Schülerinnen und Schüler von fast allen Befragten (eher) zugestimmt wird, fällt die Einschätzung für die *leistungsstärkeren* Schülerinnen und Schüler heterogener aus. So sehen bei diesen je etwa die Hälfte der Schulleitungen eine bessere Lernförderung als vollständig gegeben (*stimme zu*: 54.5 Prozent) oder eher nicht gegeben (*stimme eher nicht zu*: 45.5 Prozent) an. Die Antworten der Schulassistenzen fallen dagegen überwiegend positiv aus (*stimme eher zu* oder *stimme zu*: 85.7 Prozent).

Weiterführende Aussagen zur Individualisierung im Unterricht finden sich in den Interviews mit den Lehrpersonen und Schulleitungen. Konkret werden darin von Teilnehmenden die folgenden Massnahmen und Wirkungen in den ALLE-Schulen als Folge des Pilotprojekts angesprochen:

- Verwendung unterschiedlicher Lehrmittel und Lernmaterialien innerhalb einer Klasse je nach Lernstand,
- spezielle Lernmaterialien und/oder Lektionen für leistungsschwächere Schülerinnen und Schüler,
- Unterricht von A-, B- und C-Schülerinnen und -Schülern innerhalb einer Klasse mit Werkstattunterricht und individuellem Lernen,
- neue Zeitgefässe für die Förderung und Vertiefung,
- individuelle Anpassung von Lernzielen und Entwicklung von entsprechenden Lernkontrollen an die unterschiedlichen Leistungsniveaus innerhalb einer Klasse/Schule,
- individuelleres Eingehen auf und eine verbesserte Wahrnehmung von Schülerinnen und Schülern im Unterricht,
- Sensibilisierung für die Bedürfnisse leistungsschwächerer Schülerinnen und Schüler.

Die folgenden Interviewzitate veranschaulichen die wahrgenommenen Wirkungen des Pilotprojekts bezogen auf eine allgemein stärkere Individualisierung des Unterrichts, eine stärkere Förderung von leistungsschwachen Schülerinnen und Schülern und die gezielte Anpassung von Lernzielen im Unterricht.

Stärkere Individualisierung des Unterrichts:

Äusserungen der Lehrpersonen

- *«Dann war es so gewesen, dass wir uns überlegt haben – jetzt mit diesem Jahrgang, der in die zweite gekommen ist –, hat man bei den schwächeren Schülern [Anm.: Schülerinnen und Schüler in Abteilung B] überlegt, zum Beispiel das Thema Brüche, machen wir jetzt nicht mit unserem Standardlehrmittel, sondern beim Thema Brüche integrieren wir *«Mathe sicher können»* jetzt im normalen Unterricht [...].»*
- *«Ich habe neue Zeitgefässe. Und ich kann diese Zeitgefässe sehr bewusst nutzen. Ich mache das so, dass, es hat gewisse Personen, die müssen kommen – jetzt im Bezug auf Mathe zum Beispiel [...], aber ich habe immer auch genug Kapazität, um zu sagen *«wer will denn kommen?»*. Und das ist mittlerweile ein recht attraktives Angebot, das intensiv genutzt wird.»*
- *«[...] in unseren Lernlandschaften war es ein wenig ein Knorz gewesen – in unserem Mini-Team – um irgendwie mehr zu individualisieren. Also wir haben uns echt ein wenig schwer getan. Und ALLE ist so [...] ein wenig der Input gewesen. So dass wir dann gesehen haben, dass das eigentlich völlig okay ist [...].»*

Äusserungen der Schulleitungen

- *«Mehr Wert auf Individualisierung im Unterricht gelegt, also noch stärker. Die Lehrpersonen, die sich überhaupt nicht vorstellen konnten, dass Niveaudurchmischung funktionieren kann, haben*

sich freiwillig dazu bereit erklärt eine A-B-Klasse zu führen, die dann am Ende zwar immer noch A-B heissen, aber eigentlich eine A-B-C-Klasse ist und sind hellauf begeistert. Das hat sich schon geändert. Werkstattunterricht und individuelles Lernen hat einen höheren Wert erhalten, auch in der Wahrnehmung der Lehrpersonen.»

- *«Sie [Anm.: die Lehrpersonen] haben sich mit Lösungen auseinander gesetzt, was man mit den Schülern an den verschiedenen Ecken machen kann [...].»*

Stärkere Förderung Leistungsschwächerer:

Äusserungen der Lehrpersonen

- *«Und dann hast du eben die, die halt eben sonst normalerweise hinten runterfallen, auf die ist natürlich schon der Fokus zu richten. Gerade in so einem System, welches teilweise sehr offene Aufgabenstellung[en] hat. [...]. Und das braucht Ressourcen und es braucht das Know-how eben. Und das ist schon auch durch ALLE noch mehr pointiert worden, aufgedeckt worden [...]. Ja, und das ist wichtig, ich glaube die schwachen Schüler sind besser unterwegs und besser gefördert als vorher.»*
- *«Unser Heilpädagoge hat da etwas zusammengestellt, in relativ grossem Aufwand, und mit dem habe ich jetzt gearbeitet und habe eigentlich gute Erfahrungen gemacht. Es sind so vier, fünf Schüler [...], die mit diesem Dossier gearbeitet haben. Und das ist auch speziell gemacht für Schüler mit einer Schreibschwäche.»*
- *«In der Mathe haben wir wirklich die, bei denen man Defizite in den ersten zwei Teilen – Zahlen- und Operationsverständnis – gesehen hat, in einer zusätzlichen Stunde; sei das Hausaufgabenstunde oder sonst eine zusätzliche Stunde, wo sie kommen und arbeiten. [...].»*

Anpassung von Lernzielen:

Äusserungen der Lehrpersonen

- *«Von der Mathe-Weiterbildung her kann ich mich noch erinnern: Mut zur Lücke. Ich habe grosszügiger angefangen, ich habe jetzt angefangen, vier Programme zu machen. Etwas für jemanden, der niemals in die Regellernziele hineinkommen wird. Etwas für zwei, die hineinkommen könnten. Etwas für vier, die auf relativ hohem Level arbeiten sollten und dann noch mit Aufgaben, welche in das nächsthöhere Niveau hineinleiten könnten. Aber das ist halt auch, weil ich nur sieben Schüler habe.»*
- *«Und in Deutsch habe ich wirklich einen Schritt zurück gemacht und versucht «weniger ist mehr», und habe das Gefühl gehabt, doch sie haben im Schreiben recht schnell Fortschritte gemacht – in den zwei Jahren.»*

- *«Wirklich zuerst die Defizite abholen und nicht einfach durchrasseln und dann schauen, das hat er schon wieder nicht gekonnt. Wirklich mega zurückbrechen, und dann, im Verlauf dieser drei Jahre, bin ich wirklich einen Schritt zurückgegangen.»*
- *«Wir haben in Anlehnung zu diesen Stützkursen, welche es in Mathe gab, in Deutsch angesetzt. Vor allem bei der A-Stufe wollten wir das Thema vom Fördern mehr gewichtet haben, und zwar in der Sprachrichtigkeit. Also nicht einfach eben sich im Sozialen über die Schriftlichkeit verständigen und kommunizieren können, sondern auch die Sprachrichtigkeit. So haben wir in eigener Initiative eine Gruppe gebildet, in welcher jetzt ein Jahr gearbeitet wurde, wie wir eine Lernstandserhebung für dieses Thema machen können.»*

Äusserungen der Schulleitungen

- *«In Deutsch, aus dem heraus, hat die Deutschgruppe gefunden, wir möchten nicht nur auf die Schwächeren [achten], sondern auch etwas für die Stärkeren. Dort ist die Idee, dass wir bestimmte Lernstandserhebungen machen in punkto Deutsch.»*
- *«Ja, doch. Gerade, dass man nicht einfach strikt nach dem Buch weiterläuft, also zum Beispiel der Matheunterricht, sondern man schaut erst einmal: das fehlt und das. Dann machen wir mal eine Auszeit vom Buch und sagen, jetzt machen wir eine Förderphase. Dies hat sich schon verändert.»*

4.8 Professionelle Kompetenzen der Lehrpersonen

Dieses Kapitel thematisiert, wie sich ausgewählte professionelle Kompetenzen der am Pilotprojekt ALLE teilnehmenden Lehrpersonen im Evaluationszeitraum entwickelt haben. Im Rahmen von ALLE kann die Weiterentwicklung der professionellen Kompetenzen der Lehrpersonen als wichtiger Ansatzpunkt verstanden werden, um die Lernwirksamkeit des Unterrichts zu verbessern und als Folge davon die Lernergebnisse der Schülerinnen und Schüler zu fördern. Professionelle Kompetenzen wurden in der Evaluation bezogen auf zwei wichtige Konstrukte in den Blick genommen (vgl. z.B. Baumert & Kunter, 2011), die Überzeugungen zum Lehren und Lernen sowie die diagnostischen Kompetenzen der Lehrpersonen. Konkret werden in Kapitel 4.8.1 Veränderungen im Evaluationszeitraum in den Überzeugungen im Sinne einer *Transmissionsorientierung*, *Konstruktionsorientierung* sowie *Schülerleistung als Begabung beziehungsweise angeborener Fähigkeit* analysiert. Anschliessend wird in Kapitel 4.8.2 auf die Merkmale der *Urteilsakkuratheit* und *Urteilsicherheit* von Lehrpersonen eingegangen, bezogen auf ihre Fähigkeit, die Leistung ihrer Schülerinnen und Schüler zutreffend zu beurteilen und Leistungsanforderungen richtig einzuschätzen.

4.8.1 Überzeugungen zum Lehren und Lernen

Die Überzeugungen zum Lehren und Lernen wurden mithilfe von Fragen an die Lehrpersonen zum Erwerb sprachlichen beziehungsweise mathematischen Wissens (Transmissions- und Konstruktionsorientierung) sowie zum sprachlichen beziehungsweise mathematischen Lernen (Schülerleistung als Begabung bzw. angeborene Fähigkeit) erhoben. Die Merkmale wurden zum ersten Messzeitpunkt (Anfang der 7. Klassenstufe) und vierten Messzeitpunkt (Ende der 9. Klassenstufe) immer mit den gleichen Fragen (Fragebogen-Items) erfasst. Dadurch ist es möglich, die Veränderung der Überzeugungen während des Evaluationszeitraums zu beschreiben. Die Überzeugungen wurden jeweils getrennt für die Fächer Deutsch und Mathematik und mit einer auf das Fach angepassten Formulierung (z.B. «im Fach Deutsch» vs. «im Fach Mathematik») erfasst. Die zugehörigen Items wurden für jedes Merkmal zu einer Skala zusammengefasst, das heisst aus den Antworten der Lehrpersonen auf die Items wurde jeweils der Mittelwert gebildet. Tabelle 37 zeigt die Anzahl Items je Skala sowie ein Beispielitem für jede Skala.

Tabelle 37. Skalen zur Erfassung der Überzeugungen zum Lehren und Lernen der Lehrpersonen

Merkmal	Anzahl Items	Beispielitem
Transmissionsorientierung	4	Effektive/gute Lehrpersonen zeigen den richtigen Weg vor, wie eine Aufgabe im Fach Deutsch/eine Mathematikaufgabe zu lösen ist.
Konstruktionsorientierung	5	Es hilft Schülerinnen und Schülern, Inhalte im Fach Deutsch/Mathematik zu begreifen, wenn man sie ihre eigenen Ideen/Lösungswege diskutieren lässt.
Schülerleistung als Begabung bzw. angeborene Fähigkeit	5	Um im Fach Deutsch/Mathematik gut zu sein, muss man eine besondere sprachliche/mathematische Begabung haben.

Anmerkungen. Beurteilung auf einer sechsstufigen Antwortskala (1, «stimme überhaupt nicht zu» bis 6, «stimme völlig zu»).

Höhere Werte auf der Skala *Transmissionsorientierung* entsprechen einer stärkeren Zustimmung zu der Vorstellung, dass Wissen durch einen gerichteten Vermittlungsprozess von der Lehrperson an die Schülerinnen und Schüler weitergegeben wird. Höhere Werte auf der Skala *Konstruktionsorientierung* indizieren eine Zustimmung zu der Sichtweise, dass Wissen seitens der Schülerinnen und Schüler durch aktive und individuelle Konstruktionsprozesse aufgebaut werden muss. Bei der Skala *Schülerleistung als Begabung beziehungsweise angeborene Fähigkeit* weisen höhere Werte auf die Annahme seitens der Lehrpersonen hin, dass die Fähigkeiten der Schülerinnen und Schüler in Deutsch beziehungsweise in Mathematik in hohem Masse durch ihre Anlagen vorbestimmt sind.

Im Folgenden werden Befunde zu den einzelnen Messzeitpunkten und zur Veränderung im Evaluationszeitraum präsentiert. Dabei werden die Resultate beschränkt auf diejenigen Lehrpersonen

dargestellt, die an beiden Messzeitpunkten teilgenommen haben. Diese Einschränkung ist wichtig, da ein grösserer Anteil der Lehrpersonen nur an einem Zeitpunkt befragt werden konnte. So wird vermieden, dass Entwicklungen konstatiert werden, die möglicherweise nur auf eine sich verändernde Stichprobenszusammensetzung zurückgehen. Aufgrund der resultierenden eingeschränkten Stichprobengrössen wird auf eine statistische Prüfung von Unterschieden zwischen den ALLE- und Vergleichsschulen mit Signifikanztests verzichtet (vgl. Kap. 3.6), und die Darstellung erfolgt rein deskriptiv. Zur Verdeutlichung der praktischen Bedeutsamkeit der Unterschiede wird das Effektstärkemass d nach Cohen (1992) verwendet (vgl. ebd.). Wegen der kleinen Stichproben und der fehlenden statistischen Absicherung muss die Interpretation jedoch zurückhaltend erfolgen.

Tabelle 38 und Tabelle 39 berichten die Resultate für die Fächer Deutsch und Mathematik getrennt für die ALLE- und Vergleichsschulen. Für beide Fächer und sowohl für die ALLE- als auch die Vergleichsschulen zeigt sich eine hohe Zustimmung der Lehrpersonen zur konstruktionsorientierten Perspektive (die Schülerinnen und Schüler erwerben Wissen durch aktive und individuelle Konstruktionsprozesse). Die Zustimmung zur transmissionsorientierten Perspektive auf das Lernen der Schülerinnen und Schüler (Wissen wird durch einen gerichteten Vermittlungsprozess von der Lehrperson an die Schülerinnen und Schüler weitergegeben) fällt geringer aus, liegt aber stets nahe oder über dem theoretischen Mittelwert der Antwortskala von 3.5. Für die Zuschreibung von Leistungen auf die Begabung beziehungsweise die angeborene Fähigkeit der Schülerinnen und Schüler liegen alle Mittelwerte dagegen deutlich unter dem theoretischen Mittelwert der Antwortskala, die damit verbundenen Aussagen werden also tendenziell abgelehnt.

Die Gegenüberstellung der ALLE- und Vergleichsschulen im Fach Deutsch zeigt ein inkonsistentes Bild (vgl. Tabelle 38).

Tabelle 38. Mittelwerte für die Überzeugungen zum Lehren und Lernen in den ALLE- und Vergleichsschulen im Fach Deutsch für die Messzeitpunkte der ALLE-Evaluation

Merkmal	Messzeitpunkt	ALLE-Schulen		Vergleichsschulen		Cohens d
		M	SD	M	SD	
Transmissionsorientierung	T1	3.63	0.67	4.00	0.72	-0.49
	T4	3.71	0.75	3.42	0.80	0.39
Konstruktionsorientierung	T1	4.69	0.56	4.92	0.62	-0.39
	T4	4.72	0.62	4.98	0.65	-0.44
Schülerleistung als Begabung bzw. angeborene Fähigkeit	T1	2.96	0.86	2.42	0.48	0.78
	T4	2.64	0.93	2.85	0.91	-0.24

Anmerkungen. T1 = Anfang 7. Klasse, T4 = Ende 9. Klasse. ALLE-Schulen $N = 17$ und Vergleichsschulen $N = 12$. Die Ergebnisse beruhen auf den Angaben derjenigen Lehrpersonen, die an beiden Befragungen (T1 und T4) teilgenommen haben (vgl. Einleitung Kap. 4.8.1).

Zum ersten Messzeitpunkt (T1) fällt die Zustimmung an den Vergleichsschulen sowohl bei der Transmissionsorientierung als auch der Konstruktionsorientierung höher aus als in den ALLE-Schulen. Die Mittelwertsunterschiede sind mit $d = -0.49$ beziehungsweise $d = -0.39$ von mittlerer Ausprägung. Anders am vierten Messzeitpunkt (T4): Für die Lehrpersonen der Vergleichsschulen ergeben sich weiterhin höhere Mittelwerte bei der Konstruktionsorientierung ($d = -0.44$), sie stimmen den Fragen zur Transmissionsorientierung nun aber weniger zu als die Lehrpersonen der ALLE-Schulen ($d = 0.39$). Dies ist vor allem durch eine abnehmende Zustimmung zur Transmissionsperspektive in den Vergleichsschulen bedingt, in den ALLE-Schulen bleiben die Mittelwerte beider Merkmale nahezu unverändert. Der Überzeugung, dass Schülerleistungen auf einer Gabe oder angeborenen Fähigkeit beruhen, stimmen die Lehrpersonen in den ALLE-Schulen am ersten Messzeitpunkt deutlich stärker zu ($d = 0.78$), am vierten Messzeitpunkt dagegen etwas weniger stark zu ($d = -0.24$) als die Lehrpersonen in den Vergleichsschulen.

Ein insgesamt ähnliches Bild ergibt sich für das Fach Mathematik (vgl. Tabelle 39). Hinsichtlich der Transmissionsorientierung fällt allerdings auf, dass zumeist eine geringere Zustimmung als im Fach Deutsch resultiert. Zum ersten Messzeitpunkt (T1) zeigen sich bei diesem Merkmal deutlich höhere Mittelwerte in den Vergleichsschulen als in den ALLE-Schulen ($d = -0.64$). Zum vierten Messzeitpunkt (T4) ist diese Disparität aber nicht mehr festzustellen ($d = 0.07$). Demgegenüber ergeben sich für die Konstruktionsorientierung zum ersten Messzeitpunkt deutlich höhere Mittelwerte in den ALLE-Schulen ($d = 0.56$). Dieser Unterschied bleibt – bei allerdings erheblich geringerer Ausprägung ($d = 0.16$) – auch am vierten Messzeitpunkt bestehen. Dass Schülerleistung vor allem von der Begabung abhängt beziehungsweise eine angeborene Fähigkeit darstellt, nehmen die Lehrpersonen der ALLE-Schulen an beiden Messzeitpunkten in etwas geringerem Umfang an als die Lehrpersonen der Vergleichsschulen ($d = 0.23$ bzw. $d = 0.13$).

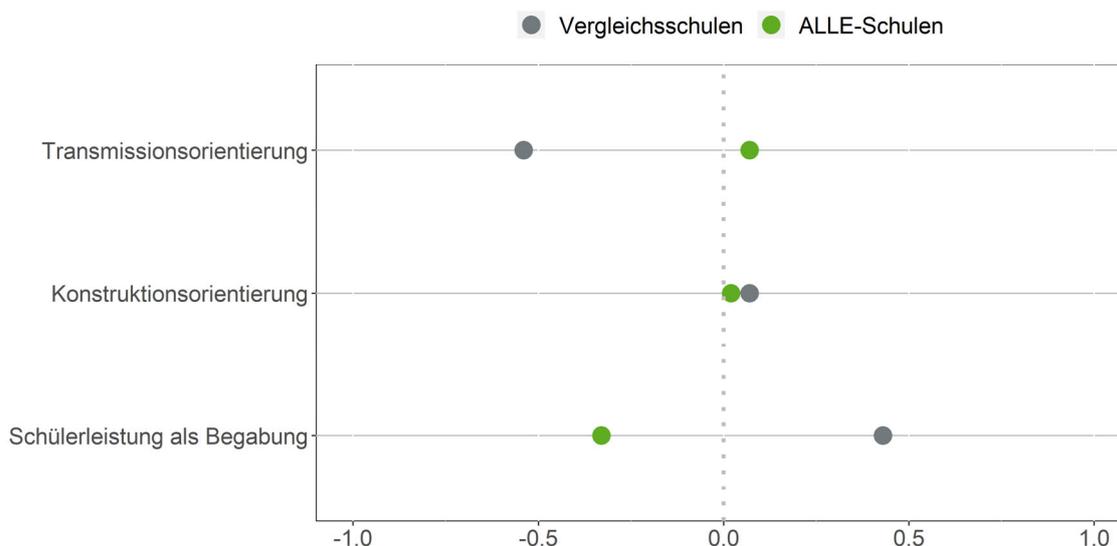
Tabelle 39. Mittelwerte für die Überzeugungen zum Lehren und Lernen in den ALLE- und Vergleichsschulen im Fach Mathematik für die Messzeitpunkte der ALLE-Evaluation

Merkmal	Messzeitpunkt	ALLE-Schulen		Vergleichsschulen		Cohens d
		M	SD	M	SD	
Transmissionsorientierung	T1	3.22	0.72	3.65	0.67	-0.64
	T4	3.42	0.88	3.67	0.44	0.07
Konstruktionsorientierung	T1	4.88	0.62	4.38	1.23	0.56
	T4	5.06	0.62	4.97	0.43	0.16
Schülerleistung als Begabung bzw. angeborene Fähigkeit	T1	2.73	0.90	2.55	0.55	0.23
	T4	2.75	0.80	2.65	0.72	0.13

Anmerkungen. T1 = Anfang 7. Klasse, T4 = Ende 9. Klasse. ALLE-Schulen $N = 21$ und Vergleichsschulen $N = 13$. Die Ergebnisse beruhen auf den Angaben derjenigen Lehrpersonen, die an beiden Befragungen (T1 und T4) teilgenommen haben (vgl. Einleitung Kap. 4.8.1).

Abbildung 54 und Abbildung 55 veranschaulichen die Veränderungen der Überzeugungen für die Fächer Deutsch und Mathematik zwischen dem ersten und letzten Messzeitpunkt der Evaluation. Bei der Transmissionsorientierung fallen die Unterschiede zwischen den ALLE- und Vergleichsschulen vergleichsweise deutlich aus. In den ALLE-Schulen nimmt die Zustimmung zur Transmissionsperspektive zu, im Fach Deutsch allerdings nur minimal und im Fach Mathematik moderat. Demgegenüber nimmt sie in den Vergleichsschulen erkennbar ab, vor allem im Fach Deutsch. Dabei ist zu beachten, dass die Transmissionsorientierung in den ALLE-Schulen bereits zu Beginn der Evaluation deutlich geringer ausgeprägt war als in den Vergleichsschulen. Durch den erheblichen Rückgang in den Vergleichsschulen ergibt sich allerdings am Ende der Evaluation eine vergleichsweise stärkere Transmissionsorientierung in den ALLE-Schulen (vgl. Tabelle 38).

Abbildung 54. Veränderungen der Überzeugungen zum Lehren und Lernen in den ALLE- und Vergleichsschulen im Fach Deutsch

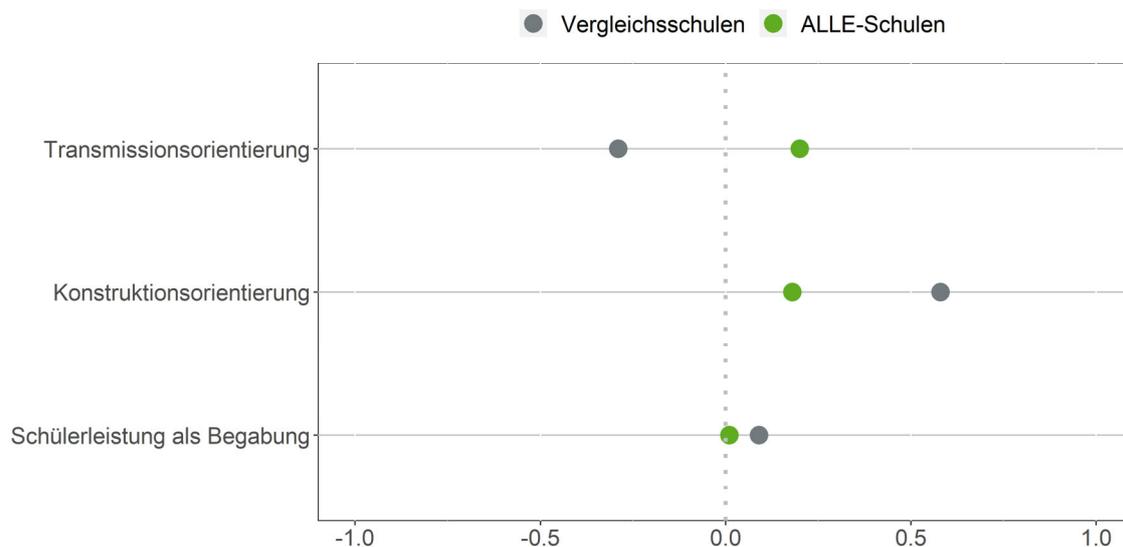


Anmerkungen. ALLE-Schulen $N=17$ und Vergleichsschulen $N=12$. Eingetragen sind Mittelwertsdifferenzen auf der sechsstufigen Antwortskala (1, «stimme überhaupt nicht zu» bis 6, «stimme völlig zu») zwischen dem vierten und ersten Messzeitpunkt in den ALLE- bzw. Vergleichsschulen, basierend auf den Angaben der Lehrpersonen, welche seit Beginn der Evaluation ALLE (ab dem ersten Messzeitpunkt) dieselben Schülerinnen und Schüler unterrichtet haben. Im Betrag höhere Werte entsprechen einer stärkeren Veränderung; positive Werte entsprechen einer Zunahme, negative Werte einer Abnahme der Zustimmung.

Die Konstruktionsorientierung nimmt im Fach Deutsch sowohl in den ALLE- als auch in den Vergleichsschulen geringfügig zu. Im Fach Mathematik ist die Zunahme deutlich stärker, insbesondere in den Vergleichsschulen. Allerdings war die Zustimmung zur Konstruktionsorientierung in den ALLE-Schulen bereits zu Beginn der Evaluation deutlich höher ausgeprägt als in den Vergleichsschulen. Am Ende des Evaluationszeitraums unterscheiden sich ALLE- und Vergleichsschulen

kaum noch (vgl. Tabelle 39). Bei der Überzeugung zur Schülerleistung als Begabung beziehungsweise angeborene Fähigkeit ergibt sich im Fach Deutsch ein gegenläufiger Befund für die ALLE- und Vergleichsschulen. In den ALLE-Schulen ist eine klare Abnahme der Zustimmung festzustellen, während diese in den Vergleichsschulen noch zunimmt. Im Fach Mathematik resultiert in den ALLE-Schulen keine Veränderung, in den Vergleichsschulen ist wieder eine – im Umfang allerdings minimale – Zunahme der Zustimmung zu dieser Überzeugung erkennbar.

Abbildung 55. Veränderungen der Überzeugungen zum Lehren und Lernen in den ALLE- und Vergleichsschulen im Fach Mathematik



Anmerkungen. ALLE-Schulen $N = 21$ und Vergleichsschulen $N = 13$. Eingetragen sind Mittelwertsdifferenzen auf der sechsstufigen Antwortskala (1, «stimme überhaupt nicht zu» bis 6, «stimme völlig zu») zwischen dem vierten und ersten Messzeitpunkt in den ALLE- bzw. Vergleichsschulen, basierend auf den Angaben der Lehrpersonen, welche seit Beginn der Evaluation ALLE (ab dem ersten Messzeitpunkt) dieselben Schülerinnen und Schüler unterrichtet haben. Im Betrag höhere Werte entsprechen einer stärkeren Veränderung; positive Werte entsprechen einer Zunahme, negative Werte einer Abnahme der Zustimmung.

4.8.2 Urteilsakkuratheit und Urteilssicherheit der Lehrpersonen

Thema dieses Abschnitts sind die Urteilsakkuratheit⁹ und Urteilssicherheit der am Pilotprojekt ALLE teilnehmenden Lehrpersonen. Die Urteilsakkuratheit (oder Veridikalität, ein Aspekt der Validität) erfasst den Grad der Übereinstimmung subjektiver Urteile über lern- und unterrichtsrelevante Merkmale mit den bei Schülerinnen und Schülern objektiv gemessenen Merkmalsausprägungen

⁹ Im Folgenden wird der Begriff Urteilsakkuratheit verwendet, da Urteilsgenauigkeit (bzw. Diagnosegenauigkeit) nach Praetorius und Südkamp (2017) zu sehr mit Reliabilitätsaspekten assoziiert wird.

(vgl. Helmke, Hosenfeld & Schrader, 2004; Schrader & Praetorius, 2018). In Abhängigkeit vom zu beurteilenden Merkmal kann die Akkuratheit der Urteile in die personenbezogene, aufgabenbezogene und aufgabenspezifische Urteilsakkuratheit differenziert werden (Schrader, 1989). Die *personenbezogene* Urteilsakkuratheit bezieht sich auf die Einschätzung leistungsbezogener Merkmale der Schülerinnen und Schüler, die *aufgabenbezogene* Urteilsakkuratheit auf die Einschätzung der Leistung einer Klasse beim Lösen einer Aufgabe (Lösungshäufigkeit der Aufgabe in der Klasse bzw. Aufgabenschwierigkeit). Der *aufgabenspezifischen* Urteilsakkuratheit liegen Einschätzungen der Leistungen unterschiedlicher Lernender bei jeder einzelnen Aufgabe zugrunde (als richtig oder falsch gelöst). Im Rahmen der Evaluation ALLE wurden nur die personen- und aufgabenbezogenen Urteile der Lehrpersonen erhoben und untersucht. Im Folgenden werden zunächst die verwendeten Indikatoren der Urteilsakkuratheit und Urteilssicherheit erläutert (Kap. 4.8.2.1). Anschliessend werden die Befunde zur Urteilsakkuratheit und Urteilssicherheit und ihrer Veränderung während des Pilotprojekts dargestellt (Kap. 4.8.2.2).

4.8.2.1 Indikatoren der Urteilsakkuratheit und Urteilssicherheit

In zwei Erhebungen in der 7. Klassenstufe (Dezember/Januar) sowie am Ende der 9. Klassenstufe wurden Lehrpersonen dazu befragt, wie ihre Schülerinnen und Schüler in den Kompetenztests der Evaluation ALLE abschneiden würden. Erhoben wurden Urteile bezogen auf die jeweils drei Kompetenzbereiche im Fach Deutsch beziehungsweise Mathematik, die in der Evaluation getestet wurden (Deutsch: Lesen, Hören, Sprache(n) im Fokus; Mathematik: Zahl und Variable, Form und Raum, Grössen, Funktionen, Daten und Zufall). In der ersten Erhebung wurden zusätzlich Einschätzungen über die Lösungshäufigkeit in der Klasse bei Aufgaben erfasst, die den drei Kompetenzbereichen in den Fächern Deutsch beziehungsweise Mathematik zugeordnet waren (Deutsch: 21 Aufgaben; Mathematik: 30 Aufgaben; 7 bzw. 10 Aufgaben pro Kompetenzbereich). Darüber hinaus wurden die Lehrpersonen danach gefragt, wie sicher sie sich ihrer personen- beziehungsweise aufgabenbezogenen Einschätzungen in den einzelnen Kompetenzbereichen seien.

Die personenbezogenen Einschätzungen der Lehrpersonen wurden mittels einer 10-Punkte-Zustimmungsskala erfasst. Die niedrigste und höchste Antwortkategorie der Skala wurden wie folgt gekennzeichnet: «0-10 %, die 10 % leistungsschwächsten Schüler/-innen» und «90-100 %, die 10% leistungsstärksten Schüler/-innen». Die Lehrpersonen wurden gebeten, das Testergebnis jeder einzelnen Schülerin und jedes einzelnen Schülers ihrer Klasse für jeden der Kompetenzbereiche im Vergleich mit allen Schülerinnen und Schülern im Kanton Zürich (Anfang 7. Klassenstufe) einzuordnen. Folgende Instruktion wurde beispielsweise zum ersten Messzeitpunkt für den Kompetenzbereich Lesen verwendet: «Bitte kreuzen Sie für jede Schülerin/jeden Schüler an, wie sie/er nach Ihrer Einschätzung im Test für den Bereich LESEN im Vergleich mit allen Schüler/-innen im Kanton Zürich (Anfang 7. Klassenstufe) abgeschnitten hat.» Zur Erfassung am Ende der

Evaluation wurden dieselben Formulierungen eingesetzt, diese waren jedoch leicht an die entsprechende Klassenstufe angepasst (z.B. «gegen Ende 9. Klassenstufe»). Anschliessend an die Einschätzung der Leistung aller Schülerinnen und Schüler im jeweiligen Kompetenzbereich wurden die Lehrpersonen gebeten, auf einer sechsstufigen Zustimmungsskala anzugeben, wie sicher sie sich ihrer Einschätzung seien (1, «sehr unsicher» bis 6, «sehr sicher»). Zur Erfassung der aufgabenbezogenen Einschätzungen wurden die Lehrpersonen gebeten, für jede der sieben beziehungsweise zehn Aufgaben jedes sprachlichen beziehungsweise mathematischen Kompetenzbereichs anzugeben, wie viele ihrer Schülerinnen und Schüler, die am Test teilgenommen hatten, die dargestellte Aufgabe korrekt gelöst hatten. Analog zu den personenbezogenen Einschätzungen wurden die Lehrpersonen zur Sicherheit ihrer aufgabenbezogenen Einschätzungen befragt.

Als Indikator der Urteilsakkuratheit wurde nach Schrader (1989; siehe auch Schrader & Helmke, 1987) je Kompetenzbereich die *Rangordnungskomponente* bestimmt, die oft als zentrale Facette der Urteilsakkuratheit aufgefasst wird. Sie wurde berechnet als die Produkt-Moment-Korrelation nach Pearson (a) zwischen den Einschätzungen der Lehrpersonen zu den Leistungen der Schülerinnen und Schüler ihrer Klasse und den tatsächlich erzielten Testleistungen (*personenbezogene* Akkuratheit) beziehungsweise (b) zwischen den Einschätzungen der Lehrpersonen zur Schwierigkeit der Testaufgaben und den tatsächlich erzielten Lösungshäufigkeiten (*aufgabenbezogene* Akkuratheit). Die Rangordnungskomponente erfasst die Fähigkeit der Lehrperson, die relative Leistungsposition einzelner Schülerinnen und Schüler beziehungsweise die relative Schwierigkeit einer Aufgabe korrekt einzuschätzen. Im Fall einer perfekt akkuraten Einschätzung läge ihr Wert bei 1.

4.8.2.2 Ergebnisse zu Urteilsakkuratheit und Urteilssicherheit

Im Folgenden werden zunächst Ergebnisse zur *aufgabenbezogenen* Urteilsakkuratheit und Urteilssicherheit berichtet. Da für die aufgabenbezogenen Masse keine Längsschnittdaten verfügbar waren, für die vorliegende Evaluation jedoch vor allem die Veränderung während des Evaluationszeitraums im Vordergrund stand, steht die Beschreibung im Vordergrund, auf Signifikanztests wird dagegen verzichtet (vgl. Kap. 3.6). Zur Verdeutlichung der beobachteten Unterschiede zwischen ALLE- und Vergleichsschulen wird das Effektstärkemass d nach Cohen (1992) verwendet. Im Anschluss werden zur *personenbezogenen* Urteilsakkuratheit und Urteilssicherheit Befunde zu den einzelnen Messzeitpunkten und zu Unterschieden in der Veränderung zwischen ALLE- und Vergleichsschulen über den Evaluationszeitraum präsentiert. Die Resultate werden beschränkt auf diejenigen Lehrpersonen dargestellt, welche an beiden Messzeitpunkten teilgenommen haben. Für diese Lehrpersonen wird der Datensatz auf diejenigen Schülerinnen und Schüler reduziert, die an beiden Messzeitpunkten von ihnen unterrichtet wurden. Die Beschränkung auf diese Stichprobe ist notwendig, da ein grösserer Anteil der Lehrpersonen nur an einem Zeitpunkt an der Befragung

teilnahm, und Veränderungen in der Stichprobenzusammensetzung im Projektverlauf sonst zu einem falschen Eindruck der Entwicklung in den ALLE- und Vergleichsschulen führen könnten. Da der Schwerpunkt des Interesses auf den Unterschieden zwischen ALLE- und Vergleichsschulen in der Veränderung über den Evaluationszeitraum liegt, werden bei den messzeitpunktspezifischen Analysen keine Signifikanztests durchgeführt (vgl. Kap. 3.6). Zur Verdeutlichung der deskriptiv dargestellten Unterschiede wird wiederum das Effektstärkemass d verwendet.¹⁰

Um Unterschiede in der Veränderung in der personenbezogenen Urteilsakkuratheit und Urteilsicherheit zwischen ALLE- und Vergleichsschulen weiter zu analysieren, werden Regressionsanalysen eingesetzt (vgl. Kap. 3.6). Insgesamt werden vier Gruppen von Regressionsmodellen berechnet, in denen (1) die Urteilsakkuratheit im Fach Deutsch, (2) die Urteilssicherheit im Fach Deutsch, (3) die Urteilsakkuratheit im Fach Mathematik und (4) die Urteilssicherheit im Fach Mathematik fokussiert werden. Die zu erklärenden Zielvariablen (abhängige Variablen) stellen dabei die Urteilsakkuratheit beziehungsweise Urteilssicherheit in den einzelnen Kompetenzbereichen am Ende des Evaluationszeitraums (vierter Messzeitpunkt) dar. Als erklärende Variable (Prädiktor bzw. unabhängige Variable) wird neben einer zweistufigen (Dummy-)Variable (ALLE-Schulen = 1, Vergleichsschulen = 0) die jeweils interessierende Zielvariable zu Beginn des Evaluationszeitraums (erster Messzeitpunkt) berücksichtigt. Diese statistische Kontrolle des Ausgangswerts der jeweiligen Zielvariable erlaubt es, den Regressionskoeffizienten der Variable ALLE- vs. Vergleichsschulen vorsichtig im Sinne einer Wirkung von ALLE zu interpretieren. Zwar werden an dieser Stelle auch Signifikanztests eingesetzt. Die Ergebnisse sind aber mit Vorsicht zu interpretieren, da nur ein moderater Anteil der Lehrpersonen am Anfang und Ende der Evaluation Urteile zu den Schülerleistungen abgegeben hat, die Stichprobe also entsprechend klein ist.

Um die Höhe der berichteten Kennwerte zur Urteilsakkuratheit einzuordnen, können diese mit anderen Studien verglichen werden, welche die Urteilsakkuratheit ebenfalls basierend auf der Korrelation zwischen den Urteilen der Lehrpersonen und Testleistungen gemessen haben. Für die aufgabenbezogene Urteilsakkuratheit können die Befunde von McElvany et al. (2009) herangezogen werden, die eine Korrelation zwischen den Lehrpersonenurteilen und der Lösungshäufigkeit von Aufgaben von $r = 0.50$ (Streuung zwischen den Lehrpersonen: $SD = 0.31$) berichten. Hinsichtlich der personenbezogenen Urteilsakkuratheit zeigt die Meta-Analyse von Südkamp et al. (2012), dass Lehrpersonen testbasierte Schülerleistungen meist relativ genau, mit einer mittleren Korrelation

¹⁰ Die Berechnung von d basiert in diesem Kapitel auf Korrelationskoeffizienten, mit denen eine Fisher-Z-Transformation durchgeführt wurde. Daher kann eine Anwendung der gängigen Formel für Cohens d auf die in den Tabellen dargestellten mittleren Korrelationskoeffizienten zu abweichenden Ergebnissen führen. Auch in den anschliessend dargestellten Regressionsmodellen zur Veränderung in der Urteilsakkuratheit wurden Fisher-Z-transformierte Korrelationskoeffizienten herangezogen.

zwischen Lehrpersonenurteilen und Testergebnissen von $r = 0.63$, einschätzen können (Bandbreite der Akkuratheit über die Studien: $-0.03 \leq r \leq 1.18$; Fisher-Z transformierte Korrelationen).

Tabelle 40 und Tabelle 41 berichten die Resultate zur *aufgabenbezogenen* Urteilsakkuratheit und Urteilssicherheit für die Fächer Deutsch und Mathematik für den ersten Messzeitpunkt (7. Klassenstufe). Dargestellt wird getrennt für ALLE- und Vergleichsschulen, wie hoch die Urteilsakkuratheit und Urteilssicherheit im Mittel über die Lehrpersonen ausfallen, und wie stark sich diese zwischen den Lehrpersonen unterscheiden (Streuung bzw. Standardabweichung).

Tabelle 40. Aufgabenbezogene Urteilsakkuratheit und Urteilssicherheit von Lehrpersonen in den ALLE- und Vergleichsschulen im Fach Deutsch für den ersten Messzeitpunkt (7. Klassenstufe)

Variable	ALLE-Schulen			Vergleichsschulen			Cohens <i>d</i>
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	
<i>Urteilsakkuratheit</i>							
Lesen	35	0.44	0.40	18	0.35	0.51	0.22
Hören	32	0.22	0.43	18	0.32	0.50	-0.23
Sprache(n) im Fokus	34	0.17	0.41	17	0.00 ^a	0.34	0.44
<i>Urteilssicherheit</i>							
Lesen	35	3.37	0.84	18	3.48	1.03	-0.04
Hören	31	3.19	0.70	17	3.41	1.12	-0.02
Sprache(n) im Fokus	35	3.09	0.87	17	3.00	0.93	0.10

Anmerkungen. Für jede Lehrperson wurde die Korrelation zwischen ihren Urteilen und den tatsächlichen Lösungshäufigkeiten bei den einzelnen Aufgaben berechnet (7 Aufgaben pro Kompetenzbereich). Die Korrelationen wurden Fisher-Z-transformiert, über die Lehrpersonen gemittelt und wieder in einen Korrelationskoeffizienten transformiert (Spalte *M*). Die Berechnung der Standardabweichung der Urteilsakkuratheit (Spalte *SD*) erfolgte analog zur Berechnung des Mittelwerts. Für die Berechnung von Cohens *d* wurden die Fisher-Z-transformierten Korrelationskoeffizienten verwendet.

^a Für etwa die Hälfte der Lehrpersonen ergab sich eine *negative* Korrelation zwischen ihren Urteilen und den Lösungshäufigkeiten der Schülerinnen und Schüler bei den Aufgaben. Diese Lehrpersonen schätzten die schwierigeren Aufgaben also tendenziell als eher leicht ein, leichtere Aufgaben tendenziell als eher schwierig. Im Mittel über die (positiven und negativen) Korrelationskoeffizienten der Lehrpersonen resultierte eine Korrelation von 0.00.

Die mittleren Korrelationen in den ALLE- und Vergleichsschulen weisen auf eine eher schwache bis moderate Fähigkeit der Lehrpersonen hin, die Leistung ihrer Klasse beim Lösen von Deutsch- und Mathematikaufgaben einzuschätzen. Innerhalb der Fächer deuten sich beträchtliche Unterschiede der Urteilsakkuratheit zwischen den Kompetenzbereichen an, wobei die höchsten Ausprägungen in Deutsch im Bereich «Lesen», in der Mathematik im Bereich «Form und Raum» erzielt werden. Die berichteten Standardabweichungen indizieren darüber hinaus deutliche Unterschiede

in der Akkuratheit zwischen den Lehrpersonen. Bei der Urteilssicherheit ergeben sich im Fach Mathematik geringere Werte als im Fach Deutsch. Die Mittelwerte der Lehrpersonen in beiden Fächern liegen durchgehend unterhalb des theoretischen Mittelwerts der Antwortskala von 3.5 und weisen damit eher auf Unsicherheiten bei der Einschätzung der Aufgabenschwierigkeit hin.

Im Fach Deutsch deutet sich für die Bereiche «Lesen» sowie «Sprache(n) im Fokus» an, dass die Lehrpersonen in den ALLE-Schulen die Leistung ihrer Klasse beim Lösen von Aufgaben akkurater einschätzen können als jene in den Vergleichsschulen (in geringem bis mittlerem Umfang; $d = 0.22$ bzw. $d = 0.44$). Im Bereich «Hören» weisen hingegen die Lehrpersonen der Vergleichsschulen eine etwas höhere Urteilsakkuratheit auf ($d = -0.23$). Bei der Urteilssicherheit resultiert für keinen der Bereiche ein praktisch bedeutsamer Unterschied. Für das Fach Mathematik ergibt sich im Bereich «Zahl und Variable» ein kleiner Unterschied zugunsten der ALLE-Schulen ($d = 0.34$). In den beiden weiteren Bereichen sind die Unterschiede minimal. Bei der Urteilssicherheit weisen die ALLE-Schulen in den Bereichen «Zahl und Variable» sowie «Grössen, Funktionen, Daten und Zufall» minimal höhere Werte als die Vergleichsschulen auf ($d = 0.13$ bzw. $d = 0.11$), im Bereich «Form und Raum» ergibt sich ein geringfügiger Unterschied zugunsten der Vergleichsschulen ($d = -0.15$).

Tabelle 41. Aufgabenbezogene Urteilsakkuratheit und Urteilssicherheit von Lehrpersonen in den ALLE- und Vergleichsschulen im Fach Mathematik für den ersten Messzeitpunkt (7. Klassenstufe)

Variable	ALLE-Schulen			Vergleichsschulen			Cohens d
	N	M	SD	N	M	SD	
<i>Urteilsakkuratheit</i>							
Zahl und Variable	35	0.41	0.46	21	0.27	0.42	0.34
Form und Raum	35	0.51	0.32	20	0.52	0.31	-0.03
Grössen, Funktionen, Daten und Zufall	35	0.29	0.34	21	0.32	0.47	-0.10
<i>Urteilssicherheit</i>							
Zahl und Variable	34	3.12	0.88	21	3.00	1.00	0.13
Form und Raum	35	2.91	0.89	20	3.05	0.94	-0.15
Grössen, Funktionen, Daten und Zufall	35	2.91	0.89	20	2.76	0.89	0.11

Anmerkungen. Für jede Lehrperson wurde die Korrelation zwischen ihren Urteilen und den tatsächlichen Lösungshäufigkeiten bei den einzelnen Aufgaben berechnet (10 Aufgaben pro Kompetenzbereich). Die Korrelationen wurden Fisher-Z-transformiert, über die Lehrpersonen gemittelt und wieder in einen Korrelationskoeffizienten transformiert (Spalte M). Die Berechnung der Standardabweichung der Urteilsakkuratheit (Spalte SD) erfolgte analog zur Berechnung des Mittelwerts. Für die Berechnung von Cohens d wurden die Fisher-Z-transformierten Korrelationskoeffizienten verwendet.

Für diese Befunde ist nochmals hervorzuheben, dass auf eine statistische Absicherung der Unterschiede zwischen ALLE- und Vergleichsschulen verzichtet wurde, die berichteten Disparitäten also rein deskriptiver Natur sind (vgl. Kap. 3.6; siehe auch die Einleitung zu Kap. 4.8.2.2). Zudem ist zu beachten, dass die aufgabenbezogene Urteilsakkuratheit ausschliesslich zu Beginn der Evaluation (in der 7. Klassenstufe) erhoben wurde. Entsprechend sind die beobachteten – insgesamt geringen – Unterschiede zwischen ALLE- und Vergleichsschulen nicht als Wirkung von ALLE, sondern allenfalls im Sinne von Unterschieden in den Eingangsvoraussetzungen in beiden Gruppen zu interpretieren (d.h. als Unterschiede in der Fähigkeit der Lehrpersonen, eine akkurate Einschätzung abzugeben, die bereits zu Beginn des Evaluationszeitraums bestanden). Der im vorliegenden Kontext möglicherweise bedeutsamere Befund besteht in den erkennbar hohen Streuungen zwischen den Lehrpersonen. Soweit sich die Urteilsakkuratheit als Indikator der diagnostischen Kompetenzen von Lehrpersonen verallgemeinern lässt, bringen diese deutlich verschiedene Voraussetzungen mit, um beispielsweise die (schwierigkeitsbezogene) Eignung von Aufgaben in bestimmten didaktischen Settings oder in der individuellen Förderung einzuschätzen.

Tabelle 42 und Tabelle 43 zeigen die Ergebnisse zur *personenbezogenen* Akkuratheit und Urteilsicherheit für die Fächer Deutsch und Mathematik. Für beide Fächer und sowohl für die ALLE- als auch die Vergleichsschulen zeigt sich, dass die Übereinstimmung zwischen Lehrpersonenurteil und Schülerleistung in den einzelnen Kompetenzbereichen und zu beiden Messzeitpunkten zwar deutlich zwischen den Lehrpersonen variiert, aber im Durchschnitt etwa den in der Literatur berichteten typischen Werten entspricht (Hoge & Coladarci, 1989; Südkamp et al., 2012). Je nach Kompetenzbereich resultiert eine moderate bis eher hohe Ausprägung (z.B. im Bereich «Zahl und Variable» zum vierten Messzeitpunkt) personenbezogener Akkuratheit. Daneben hängt die Urteilsakkuratheit auch vom Messzeitpunkt ab: Im Vergleich zum Beginn des Evaluationszeitraums resultiert an dessen Ende eine tendenziell höhere Genauigkeit bei der Beurteilung der Schülerleistungen. Noch deutlicher wird dies für die Sicherheit der Lehrpersonen bei ihrem Urteil: Anders als in der ersten Erhebung liegt am Ende der Evaluation die mittlere Urteilssicherheit fast durchweg über dem theoretischen Mittelwert der Antwortskala von 3.5 und damit im Zustimmungsbereich.

Betrachtet man deskriptiv die Unterschiede zwischen ALLE- und Vergleichsschulen zu den einzelnen Messzeitpunkten, zeigt sich für die Deutsch-Lehrpersonen der Vergleichsschulen zu Beginn der Evaluation eine deutlich höhere Urteilsakkuratheit in den beiden Kompetenzbereichen «Lesen» ($d = -0.63$) und «Hören» ($d = -1.43$) als für die Lehrpersonen der ALLE-Schulen (vgl. Tabelle 42). Lediglich für den Bereich «Sprache(n) im Fokus» resultiert ein kleiner Unterschied zugunsten der ALLE-Schulen ($d = 0.18$). Am Ende des Evaluationszeitraums weisen die Lehrpersonen der Vergleichsschulen eine im Mittel durchweg höhere Akkuratheit auf. Der Unterschied zugunsten der Vergleichsschulen ist je nach Kompetenzbereich eher klein (Hören: $d = -0.16$) oder von mittlerer (Lesen: $d = -0.42$) bis grosser (Sprache[n] im Fokus: $d = -0.64$) Ausprägung. Auch bei der Urteilsicherheit zeigt sich eine Tendenz zugunsten der Vergleichsschulen. Deren Lehrpersonen fühlen

sich in allen Bereichen und zu beiden Zeitpunkten sicherer als jene der ALLE-Schulen. Der Unterschied fällt im «Hören» eher klein aus (T1: $d = -0.29$; T4: $d = -0.19$), ist in den Bereichen «Lesen» (T1: $d = -1.31$; T4: -1.12) und «Sprache(n) im Fokus» ($d = -1.08$; $d = -0.66$) dagegen erheblich.

Tabelle 42. Personenbezogene Urteilsakkuratheit und Urteilssicherheit von Lehrpersonen in den ALLE- und Vergleichsschulen im Fach Deutsch

Variable	Messzeitpunkt	ALLE-Schulen			Vergleichsschulen			Cohens d
		N	M	SD	N	M	SD	
<i>Urteilsakkuratheit</i>								
Lesen	T1	14	0.43	0.28	7	0.55	0.19	-0.63
	T4	14	0.50	0.31	7	0.58	0.20	-0.42
Hören	T1	13	0.36	0.14	9	0.56	0.22	-1.43
	T4	13	0.51	0.29	9	0.54	0.26	-0.16
Sprache(n) im Fokus	T1	13	0.51	0.20	9	0.49	0.21	0.18
	T4	13	0.48	0.31	9	0.61	0.23	-0.64
<i>Urteilssicherheit</i>								
Lesen	T1	12	3.17	0.72	7	4.00	0.58	-1.31
	T4	12	3.83	0.39	7	4.29	0.49	-1.12
Hören	T1	10	3.00	0.82	8	3.25	1.04	-0.29
	T4	10	3.90	0.32	8	4.00	0.76	-0.19
Sprache(n) im Fokus	T1	10	2.50	0.85	7	3.43	1.62	-1.08
	T4	10	3.60	0.70	7	4.14	0.90	-0.66

Anmerkungen. T1 = Anfang 7. Klasse, T4 = Ende 9. Klasse. Für jede Lehrperson wurde die Korrelation zwischen ihren Urteilen und den Testleistungen der Schülerinnen und Schüler berechnet. Die Korrelationen wurden Fisher-Z-transformiert, über die Lehrpersonen gemittelt und wieder in einen Korrelationskoeffizienten transformiert (Spalte M). Die Berechnung der Standardabweichung der Urteilsakkuratheit (Spalte SD) erfolgte analog zur Berechnung des Mittelwerts. Für die Berechnung von Cohens d wurden die Fisher-Z-transformierten Korrelationskoeffizienten verwendet.

Bei der Urteilsakkuratheit im Fach Mathematik ergeben sich zu Beginn der Evaluation Unterschiede mittlerer Grösse zugunsten der ALLE-Schulen in den Bereichen «Grössen, Funktionen, Daten und Zufall» ($d = 0.37$) sowie «Zahl und Variable» ($d = 0.52$). Für den Bereich «Form und Raum» lässt sich keine Divergenz feststellen ($d = 0.04$). Am Ende der Evaluation fällt die Akkuratheit für den Bereich «Zahl und Variable» an den Vergleichsschulen etwas höher aus ($d = -0.27$). Demgegenüber zeigen sich keine Unterschiede für die Bereiche «Form und Raum» ($d = 0.05$) und «Grössen, Funktionen, Daten und Zufall» ($d = -0.03$). Bei der Urteilssicherheit weisen zunächst die ALLE-

Schulen in den Bereichen «Zahl und Variable» ($d = 0.33$) sowie «Grössen, Funktionen, Daten und Zufall» ($d = 0.25$) etwas höhere Werte auf als die Vergleichsschulen. Im Bereich «Form und Raum» unterscheiden sich die Lehrpersonen der ALLE- und Vergleichsschulen hingegen nicht nennenswert in ihrer Urteilssicherheit ($d = -0.06$). Demgegenüber ergibt sich am Ende der Evaluation für den Bereich «Grössen, Funktionen, Daten und Zufall» eine etwas höhere Urteilssicherheit an den Vergleichsschulen ($d = -0.17$), während für die weiteren beiden Kompetenzbereiche praktisch keine Unterschiede resultieren (Form und Raum: $d = 0.07$; Zahl und Variable: $d = 0.08$).

Tabelle 43. Personenbezogene Urteilsakkuratheit und Urteilssicherheit von Lehrpersonen in den ALLE- und Vergleichsschulen im Fach Mathematik

Variable	Messzeitpunkt	ALLE-Schulen			Vergleichsschulen			Cohens d
		N	M	SD	N	M	SD	
<i>Urteilsakkuratheit</i>								
Zahl und Variable	T1	15	0.49	0.25	12	0.35	0.40	0.52
	T4	15	0.50	0.42	12	0.58	0.45	-0.27
Form und Raum	T1	15	0.43	0.32	12	0.42	0.26	0.04
	T4	15	0.47	0.37	12	0.46	0.31	0.05
Grössen, Funktionen, Daten und Zufall	T1	15	0.51	0.23	11	0.43	0.30	0.37
	T4	15	0.51	0.41	11	0.51	0.32	-0.03
<i>Urteilssicherheit</i>								
Zahl und Variable	T1	15	3.33	0.62	10	3.10	0.88	0.33
	T4	15	3.87	0.92	10	3.80	0.79	0.08
Form und Raum	T1	14	3.07	0.73	9	3.11	0.78	-0.06
	T4	14	3.50	1.02	9	3.44	0.53	0.07
Grössen, Funktionen, Daten und Zufall	T1	14	3.00	0.78	9	2.78	1.09	0.25
	T4	14	3.64	0.74	9	3.78	0.97	-0.17

Anmerkungen. T1 = Anfang 7. Klasse, T4 = Ende 9. Klasse. Für jede Lehrperson wurde die Korrelation zwischen ihren Urteilen und den Testleistungen der Schülerinnen und Schüler berechnet. Die Korrelationen wurden Fisher-Z-transformiert, über die Lehrpersonen gemittelt und wieder in einen Korrelationskoeffizienten transformiert (Spalte M). Die Berechnung der Standardabweichung der Urteilsakkuratheit (Spalte SD) erfolgte analog zur Berechnung des Mittelwerts. Für die Berechnung von Cohens d wurden die Fisher-Z-transformierten Korrelationskoeffizienten verwendet.

Abschliessend werden die Resultate von Regressionsmodellen berichtet, in denen untersucht wird, ob sich die ALLE- und Vergleichsschulen in ihrer Entwicklung der Urteilsakkuratheit und Urteils-sicherheit im Laufe der Evaluation unterscheiden. In den Regressionsmodellen werden die Urteils-akkuratheit und Urteilssicherheit in den Kompetenzbereichen der Fächer Deutsch und Mathematik am Ende der Evaluation als die zu erklärenden Zielvariablen (abhängige Variablen) herangezogen (z.B. Urteilsakkuratheit im Lesen zu T4). Als erklärende Variablen (Prädiktoren bzw. unabhängige Variablen) werden die bereits erwähnte Dummy-Variable (ALLE-Schulen = 1, Vergleichs-schulen = 0) sowie die jeweilige Zielvariable zu Beginn der Evaluation (erster Messzeitpunkt, z.B. Urteilsakkuratheit im Lesen zu T1) genutzt (vgl. Einleitung Kap. 4.8.2.2; vgl. auch Kap. 3.6). Tabelle 44 zeigt die Ergebnisse. Dargestellt wird der Unterschied in der Veränderung der Urteilsakkuratheit und Urteilssicherheit zwischen den ALLE- und Vergleichsschulen unter statistischer Kontrolle der jeweils gleichen Variable zu Beginn der Evaluation (Spalte *Wert*).

Tabelle 44. Unterschiede zwischen ALLE- und Vergleichsschulen in der personenbezogenen Urteilsakkuratheit und Urteilssicherheit am Ende der Evaluation (9. Klassenstufe; T4)

Zielvariable zu T4	Urteilsakkuratheit		Urteilssicherheit	
	Wert	(SE)	Wert	(SE)
<i>Deutsch</i>				
Lesen	-0.04	(0.13)	-0.20	(0.21)
Hören	0.03	(0.15)	-0.01	(0.22)
Sprache(n) im Fokus	-0.20	(0.12)	-0.22	(0.36)
<i>Mathematik</i>				
Zahl und Variable	-0.22	(0.17)	-0.06	(0.32)
Form und Raum	0.01	(0.13)	0.07	(0.35)
Grössen, Funktionen, Daten und Zufall	-0.04	(0.16)	-0.23	(0.33)

Anmerkungen. Deutsch: ALLE-Schulen $N = 10 - 14$ und Vergleichsschulen $N = 7 - 9$; Mathematik: ALLE-Schulen $N = 14 - 15$ und Vergleichsschulen $N = 9 - 12$.

(*) $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$

Wie aus Tabelle 44 ersichtlich wird, kann ein Unterschied zwischen den ALLE- und Vergleichsschulen in der Urteilsakkuratheit oder Urteilssicherheit am Ende der Evaluation – unter Kontrolle der Urteilsakkuratheit beziehungsweise Urteilssicherheit zu Beginn der Evaluation – für keinen der Kompetenzbereiche in Deutsch oder Mathematik statistisch nachgewiesen werden. Dabei ist zu beachten, dass die zur Verfügung stehenden Stichproben sehr klein waren und die Schlussfolgerungen mit entsprechend hoher Unsicherheit behaftet sind. Sicherlich lässt sich auch keine der

berichteten Disparitäten im Sinne einer Tendenz zugunsten der ALLE-Schulen interpretieren, da die Koeffizienten entweder sehr nahe bei null liegen oder der Richtung nach zugunsten der Vergleichsschulen ausfallen (jedoch durchweg nicht signifikant sind).

4.9 Qualität des Unterrichts

Ein wesentliches Ziel des Pilotprojekts ALLE war die Verbesserung der Qualität des Unterrichts, um so mittelbar auch positive Wirkungen auf die Lernergebnisse der Schülerinnen und Schüler zu erzielen. Um Rückschlüsse auf die Entwicklung des Unterrichts zu ziehen, wurden die Schülerinnen und Schüler sowie Lehrpersonen an den ALLE- und Vergleichsschulen zu wichtigen Aspekten der Unterrichtsqualität befragt. Nachfolgend wird thematisiert, wie sich die wahrgenommene Unterrichtsqualität aus Sicht der Schülerinnen und Schüler sowie der Lehrpersonen während des Evaluationszeitraums verändert hat, und inwieweit sich eine günstigere Entwicklung in den ALLE-Schulen gegenüber den Vergleichsschulen zeigen lässt. Die Analysen beschränken sich dabei auf diejenigen Schülerinnen und Schüler in Klassen beziehungsweise Lerngruppen, die ab Beginn der Evaluation (erster Messzeitpunkt) von derselben Fachlehrperson unterrichtet wurden. Damit wird der Tatsache Rechnung getragen, dass Unterricht auf dyadischen Interaktionen zwischen Lehrpersonen und Schülerinnen und Schülern basiert (Ditton, 2002). Entsprechend ist es sinnvoll, diejenigen Schülerinnen und Schüler im Längsschnitt zu betrachten, welche in zeitlich konstanten Dyaden mit ihren Lehrpersonen unterrichtet wurden. Analog werden in den Analysen zur Unterrichtswahrnehmung der Lehrpersonen diejenigen Lehrpersonen in den Blick genommen, welche seit dem Beginn der Evaluation dieselben Schülerinnen und Schüler unterrichtet haben.

Im Folgenden werden zunächst die Indikatoren der Unterrichtsqualität dargestellt, die auf Basis der Fragebogenangaben der Schülerinnen und Schüler und Lehrpersonen gebildet und für die Analysen verwendet wurden (Kap. 4.9.1). Anschliessend wird die wahrgenommene Unterrichtsqualität zunächst aus der Sicht der Schülerinnen und Schüler (Kap. 4.9.2), dann aus Lehrpersonensicht (Kap. 4.9.3) thematisiert. In beiden Abschnitten werden zunächst die Ergebnisse pro Messzeitpunkt in den ALLE-Schulen und Vergleichsschulen einander gegenübergestellt. Danach wird untersucht, inwieweit sich die ALLE-Schulen und Vergleichsschulen in den Veränderungen während des Evaluationszeitraums unterscheiden – und sich Hinweise darauf ergeben, dass die Teilnahme am Pilotprojekt mit günstigen Wirkungen auf die Unterrichtsqualität einherging. Details zur Vorgehensweise bei den Analysen werden zu Beginn der beiden Abschnitte dargestellt.

4.9.1 Indikatoren der wahrgenommenen Unterrichtsqualität

In der ersten Erhebung (Anfang 7. Klassenstufe), dritten Erhebung (Ende 8. Klassenstufe) und vierten Erhebung (Ende 9. Klassenstufe) der Evaluation wurden die Schülerinnen und Schüler zu

16 Unterrichtsmerkmalen mit identischen Aussagen (Fragebogen-Items) befragt. Zusätzlich wurden in der ersten Erhebung (Anfang 7. Klassenstufe) und vierten Erhebung (Ende 9. Klassenstufe) die Fachlehrpersonen zu acht Qualitätsmerkmalen, ebenfalls mit den gleichen Aussagen zu beiden Zeitpunkten, befragt. Dadurch ist es möglich, Veränderungen in der Wahrnehmung des Unterrichts durch die Schülerinnen und Schüler und durch die Lehrpersonen zu beschreiben.

Die Unterrichtsmerkmale wurden für die Fächer Deutsch und Mathematik getrennt und mit einer auf das Fach angepassten Formulierung (z.B. «Deutschunterricht» vs. «Mathematikunterricht») erfasst. Aus den Antworten der Schülerinnen und Schüler beziehungsweise der Lehrpersonen auf die einzelnen Items jedes Unterrichtsmerkmals wurden Mittelwerte gebildet, das heisst die Items wurden zu Skalen zusammengefasst. Die Anzahl der Items sowie ein Beispielitem je Skala sind in Tabelle 45 (Befragung der Lehrpersonen) und Tabelle 46 (Befragung der Schülerinnen und Schüler) enthalten. Eine Kurzbeschreibung der erfassten Merkmale lässt sich Tabelle 46 entnehmen (alle bei den Lehrpersonen erhobenen Merkmale wurden auch bei den Schülerinnen und Schülern gemessen). Da nicht alle Items zu Skalen zusammengefasst werden konnten, und einige Items aufgrund ungünstiger Messeigenschaften ausgeschlossen werden mussten, werden für einige Merkmale Ergebnisse für einzelne Items berichtet (in den Tabellen dadurch gekennzeichnet, dass die Anzahl Items gleich 1 ist). Diese Merkmale konnten entsprechend nur in Bezug auf eine einzelne Facette – und damit relativ eingeschränkt – erfasst werden.

Tabelle 45. Skalen zur Erfassung der Unterrichtsqualität aus Sicht der Lehrpersonen

Unterrichtsmerkmal	Anzahl Items pro Fach	Beispielitem
Verständlichkeit und Klarheit	3	Ich bringe gute Beispiele, damit die Schülerinnen und Schüler den Stoff besser verstehen.
Kognitive Aktivierung	3	Mir ist es wichtig, dass Behauptungen auch gut begründet werden.
Thematische Motivierung	1	Ich finde das Fach Mathematik sehr spannend.
Leistungserwartungen	1	Für besonders gute Beiträge zum Mathematikunterricht lobe ich die Schülerinnen und Schüler.
Unterrichtsklima	5	Mit den Beiträgen der Schülerinnen und Schüler gehe ich wertschätzend um.
Individuelle Förderung	3	Ich gebe die Möglichkeit, zwischen verschiedenen Aufgaben zu wählen.
Störungsprävention	1	Bei Störungen gebe ich ein klares STOPP-Signal.
Nutzung der Lernzeit (Time on Task)/Zeitmanagement	1	Ich regle vieles mit kurzen Blicken und knappen Zeichen.

Anmerkungen. Beurteilung auf einer sechsstufigen Antwortskala (1, «stimme überhaupt nicht zu» bis 6, «stimme völlig zu»). Die Anzahl Items pro Fach bezieht sich auf die für die längsschnittlichen Analysen verwendeten Items. Für eine Kurzbeschreibung der Unterrichtsmerkmale siehe Tabelle 46.

Tabelle 46. Skalen zur Erfassung der Unterrichtsqualität aus Sicht der Schülerinnen und Schüler

Unterrichtsmerkmal	Itemzahl	Beschreibung des Merkmals	Beispielitem
Verständlichkeit und Klarheit	6	Die Lehrperson kann gut erklären, drückt sich klar und deutlich aus, bringt klare und verständliche Aufgabenstellungen sowie gute Beispiele. Sie wiederholt bereits Gelerntes aus vergangenen Lektionen und fasst am Ende der Lektion das Wichtigste zusammen.	Meine Lehrperson bringt gute Beispiele, so dass ich den Stoff besser verstehe.
Lernzielorientierung	3	Die Lehrperson gibt den Schülerinnen und Schülern (SuS) zu Beginn eines neuen Themas/einer Lektion die Lernziele bekannt und weist diese immer wieder auf die Lernziele hin.	Meine Lehrperson weist immer wieder auf die Lernziele hin.
Konsolidierung	4	Die SuS bekommen die Gelegenheit, den Unterrichtsstoff ausreichend oft und in angemessener Weise (d.h. nicht mechanisch, sondern mit Transferangeboten) zu üben/wiederholen/vertiefen.	Im Deutschunterricht (Bereiche Rechtschreibung und Grammatik)/Mathematikunterricht kann ich das Gelernte üben und wiederholen, bis es sitzt.
Kognitive Aktivierung	5/7	Die Lehrperson fördert das aktive, selbstständige Lernen von SuS: Ermutigung zum Schreiben eigener Texte und zur Äusserung eigener Meinungen, zum gründlichen Nachdenken auffordernde Fragen und Einfordern von Begründungen/Erklärungen. Die Lehrperson achtet darauf, dass im Unterricht möglichst alle drankommen. Anspuchsvolle, herausfordernde Aufgaben.	Meine Lehrperson stellt uns Aufgaben, bei denen ich gründlich nachdenken muss.
Thematische Motivierung	4	Die Lehrperson macht einen spannenden Unterricht und kann die SuS für die Unterrichtsthemen begeistern.	Meine Lehrperson kann mich richtig für die Unterrichtsthemen begeistern.
Instrumentelle Motivierung	3	Die Lehrperson akzentuiert die Wichtigkeit und Nützlichkeit des Lernstoffs, beispielsweise für den Alltag, für eine gute Lehrstelle/Ausbildung und für den Beruf.	Meine Lehrperson betont, dass gutes Deutsch/Mathematik in vielen Berufen eine grosse Rolle spielt.
Leistungserwartungen	4	Die Lehrperson fördert die Lern- und Anstrengungsbereitschaft der SuS durch hohe Leistungserwartungen.	Meine Lehrperson traut mir zu, dass ich meine Leistungen noch verbessern kann.
Unterrichtsklima	7	Das Klima ist durch wechselseitige Unterstützung gekennzeichnet. Es herrschen eine entspannte Unterrichts Atmosphäre, die auch mal durch Lachen aufgelockert wird, und Wertschätzung/Respektierung der SuS vor, was etwa dadurch zum Ausdruck kommt, dass SuS ausreden können, ohne unterbrochen zu werden und dass ihre Vorschläge ernst genommen werden.	Mit meinen Beiträgen geht die Lehrperson wertschätzend um.

(Fortsetzung)

Tabelle 46 (Fortsetzung). Skalen zur Erfassung der Unterrichtsqualität aus Sicht der Schülerinnen und Schüler

Unterrichtsmerkmal	Itemzahl	Beschreibung des Merkmals	Beispielitem
Positive Fehlerkultur	2	Die Lehrperson geht konstruktiv mit Fehlern der SuS um und sieht diese als Lernchance. Sie ist geduldig mit SuS, die im Unterricht etwas nicht verstehen.	Meine Lehrperson findet, dass Fehlermachen zum Deutschunterricht/Mathematikunterricht dazu gehört.
Lehrer-Schüler-Beziehung	5	Die SuS kommen mit der Lehrperson gut aus. Die Lehrperson gibt den SuS die Möglichkeit zur Mitsprache/zum Ausdruck der eigenen Meinung. Sie behandelt die SuS fair, und es ist ihr wichtig, dass sich die SuS wohl fühlen.	Ich komme mit meiner Lehrperson gut aus.
Diagnostische Kompetenz	5	Die Lehrperson merkt, wenn der Unterricht zu schwierig ist oder einzelne SuS nicht mehr mitkommen. Sie weiss, bei welchen Aufgaben die SuS Mühe haben und wie sie diesen beim Lernen helfen kann.	Meine Lehrperson merkt, wenn ich etwas nicht richtig verstanden habe.
Individuelle Förderung	7	Die SuS können zwischen verschiedenen Aufgaben wählen und in ihrem eigenen Tempo lernen. Wenn sie etwas nicht verstehen, bekommen sie Hilfe von ihrer Lehrperson. Die Lehrperson gibt den SuS regelmässig Rückmeldungen zu den Lernfortschritten.	Meine Lehrperson gibt mir die Möglichkeiten, zwischen verschiedenen Aufgaben zu wählen.
Allgegenwärtigkeit	3	Die Lehrperson merkt, wenn die SuS nicht mitarbeiten oder nicht aufpassen. Sie weiss, was in der Klasse passiert.	Meine Lehrperson merkt sofort, wenn ich nicht aufpasse.
Regelorientierung	3	Den SuS ist klar, was sie machen dürfen und was nicht. Es gibt Regeln, an die sie sich halten müssen und klare Konsequenzen, wenn die Regeln nicht eingehalten werden.	Bei uns im Deutschunterricht/Mathematikunterricht gibt es klare Regeln, an die ich mich halten muss.
Störungsprävention	1	Die Lehrperson ergreift Massnahmen, um in der Klasse Störungen vorzubeugen bzw. diese zu unterbinden.	Bei uns im Deutschunterricht/Mathematikunterricht gibt es bei Störungen ein klares STOPP-Signal.
Nutzung der Lernzeit (Time on Task)/Zeitmanagement	6	Der Unterricht beginnt pünktlich, und es wird keine Zeit vertrödelt. Die SuS beginnen sofort zu arbeiten, wenn die Lehrperson einen Auftrag erteilt. Die SuS sind während des Unterrichts aufmerksam und konzentriert.	Ich bin während des Deutschunterrichts/Mathematikunterrichts aufmerksam und konzentriert.

Anmerkungen. Beurteilung auf einer vierstufigen Antwortskala (1, «stimme überhaupt nicht zu» bis 4, «stimme völlig zu»). Die Anzahl Items beim Merkmal «Kognitive Aktivierung» vor dem Schrägstrich bezieht sich auf das Fach Deutsch, die Angabe nach dem Schrägstrich auf das Fach Mathematik. Alle anderen Angaben beziehen sich auf beide Fächer. Die Anzahl Items pro Fach bezieht sich auf die für die längsschnittlichen Analysen verwendeten Items.

In den nachfolgenden Analysen bedeuten höhere Werte jeweils eine positivere Beurteilung des betreffenden Unterrichtsmerkmals durch die Schülerinnen und Schüler beziehungsweise Lehrpersonen. Es ist jedoch zu beachten, dass aufgrund der unterschiedlichen Antwortformate in den Befragungen (sechsstufige vs. vierstufige Antwortskala; vgl. die Anmerkungen zu Tabelle 45 und Tabelle 46) identische Mittelwerte in den Ergebnissen der beiden Zielgruppen eine unterschiedliche Bedeutung haben und daher nicht direkt vergleichbar sind.

4.9.2 Unterrichtsqualität aus Sicht der Schülerinnen und Schüler

Die Resultate zur Qualität des Unterrichts aus Schülersicht werden zunächst bezogen auf die einzelnen Messzeitpunkte beschrieben. Die Unterschiede zwischen den ALLE-Schulen und Vergleichsschulen werden dabei je Messzeitpunkt auf statistische Signifikanz geprüft. Dazu wird jedes Unterrichtsmerkmal zum betreffenden Messzeitpunkt als die zu erklärende Zielvariable (abhängige Variable) in einem Regressionsmodell analysiert. Als erklärende Variable (Prädiktor bzw. unabhängige Variable) dient eine zweistufige (Dummy-)Variable (ALLE-Schulen = 1, Vergleichsschulen = 0), welche den Unterschied zwischen den ALLE- und Vergleichsschulen abbildet. Ein signifikantes Resultat für diese Variable deutet auf einen Unterschied zwischen ALLE- und Vergleichsschulen hin. Der Lesbarkeit halber beschränken sich die Ausführungen im Text auf die statistisch abgesicherten Unterschiede, auf Einzelheiten aus den Ergebnissen der Regressionsmodelle wird verzichtet. Im Anschluss wird auf die *Veränderungen* in den Schülerwahrnehmungen der Unterrichtsqualität zwischen dem ersten (Anfang 7. Klassenstufe) und letzten (Ende 9. Klassenstufe) Messzeitpunkt der Evaluation eingegangen. Im Mittelpunkt steht, inwieweit sich Unterschiede in der Veränderung der Unterrichtswahrnehmung zwischen ALLE- und Vergleichsschulen feststellen lassen, die auf eine positive Wirkung des ALLE-Massnahmenbündels hindeuten. Zur Prüfung der statistischen Signifikanz wird die Veränderung jedes Unterrichtsmerkmals im Evaluationszeitraum – das heisst, die Differenz zwischen dem letzten und dem ersten Messzeitpunkt – als zu erklärende Zielvariable in einem Regressionsmodell analysiert. Als erklärende Variable dient wiederum die oben genannte Dummy-Variable (ALLE-Schulen = 1, Vergleichsschulen = 0). Ein signifikantes Ergebnis für diese Variable weist darauf hin, dass sich die ALLE- und Vergleichsschulen in der Entwicklung der Schülerwahrnehmungen für das betreffende Unterrichtsmerkmal unterscheiden. Weitere Details zur Umsetzung der Regressionsmodelle können Kapitel 3.6 entnommen werden.

Tabelle 47 zeigt die Mittelwerte und Streuungen (Standardabweichung) der Schülerwahrnehmungen in den ALLE- und Vergleichsschulen für das Fach Deutsch an den verschiedenen Messzeitpunkten. Zusätzlich berichtet wird Cohens d als Mass der Effektstärke des Unterschieds zwischen den ALLE- und Vergleichsschulen (vgl. Kap. 3.6).

Tabelle 47. Unterrichtsqualität im Fach Deutsch aus Sicht der Schülerinnen und Schüler in den ALLE-Schulen und Vergleichsschulen für die Messzeitpunkte der ALLE-Evaluation

Unterrichtsmerkmal	Messzeitpunkt	ALLE-Schulen		Vergleichsschulen		Cohens <i>d</i>
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Verständlichkeit und Klarheit	T1	3.26	0.56	3.20	0.49	0.12
	T3	3.17	0.53	3.06	0.56	0.21
	T4	2.98	0.78	2.92	0.70	0.09
Lernzielorientierung	T1	2.99	0.67	2.84	0.72	0.22
	T3	2.82	0.76	2.73	0.72	0.11
	T4	2.73	0.87	2.57	0.82	0.19
Konsolidierung	T1	3.14	0.55	3.04	0.53	0.19
	T3	3.03	0.54	3.04	0.56	0.00
	T4	2.95	0.79	2.92	0.68	0.05
Kognitive Aktivierung	T1	3.04	0.61	2.99	0.51	0.09
	T3	3.04	0.51	2.99	0.55	0.10
	T4	2.87	0.77	2.88	0.62	-0.01
Thematische Motivierung	T1	3.00	0.63	2.90	0.62	0.15
	T3	2.89	0.62	2.78	0.65	0.19
	T4	2.69	0.81	2.70	0.70	-0.02
Instrumentelle Motivierung	T1	3.19	0.71	3.12	0.72	0.10
	T3	3.26	0.67	3.12	0.68	0.20
	T4	3.07	0.83	2.99	0.78	0.10
Leistungserwartungen	T1	3.08	0.65	3.15	0.56	-0.10
	T3	3.19	0.57	3.02	0.61	0.28
	T4	3.03	0.72	2.95	0.68	0.12
Unterrichtsklima	T1	3.30	0.54	3.25	0.50	0.10
	T3	3.30	0.50	3.13	0.54	0.33
	T4	3.11	0.70	3.04	0.64	0.11
Positive Fehlerkultur	T1	3.26	0.77	3.30	0.69	-0.04
	T3	3.30	0.61	3.12	0.71	0.28
	T4	3.10	0.80	3.01	0.74	0.11

(Fortsetzung)

Tabelle 47 (Fortsetzung). Unterrichtsqualität im Fach Deutsch aus Sicht der Schülerinnen und Schüler in den ALLE-Schulen und Vergleichsschulen für die Messzeitpunkte der ALLE-Evaluation

Unterrichtsmerkmal	Messzeitpunkt	ALLE-Schulen		Vergleichsschulen		Cohens <i>d</i>
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Lehrer-Schüler-Beziehung	T1	3.33	0.72	3.35	0.61	-0.03
	T3	3.33	0.58	3.17	0.62	0.27
	T4	3.11	0.81	3.03	0.71	0.10
Diagnostische Kompetenz	T1	3.06	0.60	2.93	0.65	0.22
	T3	2.93	0.61	2.75	0.65	0.29
	T4	2.90	0.74	2.72	0.72	0.25
Individuelle Förderung	T1	3.11	0.58	3.00	0.57	0.19
	T3	3.05	0.55	2.81	0.61	0.41
	T4	2.83	0.78	2.75	0.67	0.10
Allgegenwärtigkeit	T1	3.21	0.69	3.13	0.70	0.12
	T3	3.13	0.67	2.98	0.72	0.22
	T4	2.84	0.91	2.86	0.80	-0.02
Regelorientierung	T1	3.47	0.59	3.42	0.57	0.09
	T3	3.38	0.58	3.36	0.63	0.04
	T4	3.05	0.93	3.19	0.75	-0.16
Nutzung der Lernzeit/ Zeitmanagement	T1	3.03	0.59	2.99	0.54	0.07
	T3	3.07	0.50	2.93	0.60	0.26
	T4	2.97	0.71	2.87	0.70	0.15
Störungsprävention	T1	2.91	1.00	2.92	0.93	-0.01
	T3	2.98	0.92	3.09	0.86	-0.12
	T4	3.04	0.93	2.98	0.92	0.06

Anmerkungen. T1 = Anfang 7. Klasse, T3 = Ende 8. Klasse, T4 = Ende 9. Klasse. *N* = 369 (23 Klassen bzw. Lerngruppen), Vergleichsschulen *N* = 302 (23 Klassen bzw. Lerngruppen). Die Ergebnisse beruhen auf den Angaben derjenigen Schülerinnen und Schüler, die ab T1 von derselben Lehrperson unterrichtet wurden (vgl. Einleitung Kap. 4.9). Die Ergebnisse für die messzeitpunktspezifischen Stichproben können Anhang 7.3 entnommen werden.

Wie sich zeigt, liegen die Mittelwerte aller Merkmale über dem theoretischen Mittelwert der Antwortskala von 2.5. Dies deutet darauf hin, dass die Schülerinnen und Schüler in den ALLE- und Vergleichsschulen den Deutschunterricht über alle Merkmale hinweg überwiegend positiv beurteilen.

len. Stellt man die ALLE- und Vergleichsschulen gegenüber, ergeben sich zum ersten Messzeitpunkt (T1) signifikant höhere Mittelwerte zugunsten der ALLE-Schulen für die Merkmale «Lernzielorientierung» und «Diagnostische Kompetenz». Die Unterschiede sind jedoch als eher gering einzuordnen (jeweils $d = 0.22$). Tendenziell signifikant ist zudem die in den ALLE-Schulen geringfügig höhere Ausprägung des Merkmals «Konsolidierung» ($d = 0.19$). Zum dritten Messzeitpunkt (T3) wird der Unterricht in den ALLE-Schulen bezüglich der Merkmale «Verständlichkeit und Klarheit», «Thematische Motivierung», «Instrumentelle Motivierung», «Leistungserwartungen», «Unterrichtsklima», «Positive Fehlerkultur», «Lehrer-Schüler-Beziehung», «Diagnostische Kompetenz der Lehrpersonen», «Individuelle Förderung», «Allgegenwärtigkeit» und «Nutzung der Lernzeit» signifikant günstiger wahrgenommen. Die Unterschiede fallen im Umfang durchwegs eher gering aus (zwischen $d = 0.19$ und $d = 0.33$), nur für das Merkmal «Individuelle Förderung» lässt sich eine Disparität mittlerer Ausprägung feststellen ($d = 0.41$). Zum vierten Messzeitpunkt (T4) resultiert ein einziger (tendenziell) signifikanter Unterschied: In den ALLE-Schulen weist das Merkmal «Diagnostische Kompetenz» einen etwas höheren Mittelwert auf ($d = 0.25$). Damit resultieren zwar mehrere signifikante Unterschiede zugunsten der ALLE-Schulen, doch treten diese mit Ausnahme des Merkmals «Diagnostische Kompetenz» nur an einzelnen Messpunkten auf und sind stets von kleiner oder höchstens (bei der «Diagnostischen Kompetenz» zu T3) mittlerer Grösse.

In Tabelle 48 werden die Ergebnisse für das Fach Mathematik dargestellt. Wie im Fach Deutsch werden alle Unterrichtsmerkmale sowohl in den ALLE-Schulen als auch den Vergleichsschulen tendenziell positiv, das heisst oberhalb des theoretischen Mittelwerts der Antwortskala, beurteilt. Zum ersten Messzeitpunkt (T1) resultiert in den ALLE-Schulen eine tendenziell signifikant niedrigere Ausprägung des Merkmals «Positive Fehlerkultur» als in den Vergleichsschulen. Der Unterschied ist jedoch als gering einzuordnen ($d = -0.18$). Zum dritten Messzeitpunkt (T3) weisen die ALLE-Schulen signifikant höhere Werte bei den Merkmalen «Lernzielorientierung» ($d = 0.19$) und «Allgegenwärtigkeit» ($d = 0.17$) auf, wobei die Unterschiede auch hier als eher gering einzustufen sind. Für die Merkmale «Verständlichkeit und Klarheit» ($d = 0.15$) und «Konsolidierung» ($d = 0.15$) resultieren sehr geringe, tendenziell signifikante Unterschiede zugunsten der ALLE-Schulen. Am vierten Messzeitpunkt (T4) ergibt sich ein eher geringer, signifikanter Unterschied der wahrgenommenen «Regelorientierung» zugunsten der Vergleichsschulen ($d = -0.30$). Damit resultieren, anders als im Fach Deutsch, Disparitäten der wahrgenommenen Unterrichtsqualität, die teilweise zugunsten der ALLE-Schulen, teilweise zugunsten der Vergleichsschulen ausfallen. Wie im Fach Deutsch erweisen sich die beobachteten Unterschiede als inkonsistent über die Messzeitpunkte und fallen in der Ausprägung durchweg gering aus.

Tabelle 48. Unterrichtsqualität im Fach Mathematik aus Sicht der Schülerinnen und Schüler in den ALLE-Schulen und Vergleichsschulen für die Messzeitpunkte der ALLE-Evaluation

Unterrichtsmerkmal	Messzeitpunkt	ALLE-Schulen		Vergleichsschulen		Cohens <i>d</i>
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Verständlichkeit und Klarheit	T1	3.23	0.70	3.34	0.54	-0.17
	T3	3.12	0.66	3.02	0.67	0.15
	T4	2.89	0.87	2.97	0.65	-0.10
Lernzielorientierung	T1	3.07	0.77	3.01	0.73	0.07
	T3	2.89	0.78	2.74	0.83	0.19
	T4	2.95	0.89	2.75	0.83	0.24
Konsolidierung	T1	3.25	0.63	3.26	0.56	-0.01
	T3	3.12	0.65	3.03	0.64	0.15
	T4	2.89	0.84	3.02	0.65	-0.16
Kognitive Aktivierung	T1	3.29	0.54	3.24	0.52	0.11
	T3	3.22	0.55	3.15	0.55	0.11
	T4	3.04	0.78	3.08	0.63	-0.06
Thematische Motivierung	T1	3.23	0.66	3.24	0.63	-0.01
	T3	3.01	0.70	2.96	0.67	0.07
	T4	2.89	0.82	2.87	0.71	0.03
Instrumentelle Motivierung	T1	3.17	0.72	3.14	0.72	0.05
	T3	3.13	0.71	3.06	0.75	0.10
	T4	2.99	0.97	2.97	0.72	0.02
Leistungserwartungen	T1	3.23	0.68	3.27	0.60	-0.05
	T3	3.21	0.62	3.14	0.62	0.12
	T4	2.93	0.90	3.02	0.71	-0.11
Unterrichtsklima	T1	3.32	0.55	3.35	0.52	-0.05
	T3	3.16	0.59	3.22	0.53	-0.10
	T4	3.12	0.72	3.08	0.65	0.05
Positive Fehlerkultur	T1	3.29	0.74	3.41	0.61	-0.18
	T3	3.17	0.70	3.20	0.74	-0.05
	T4	3.03	0.82	3.08	0.74	-0.06

(Fortsetzung)

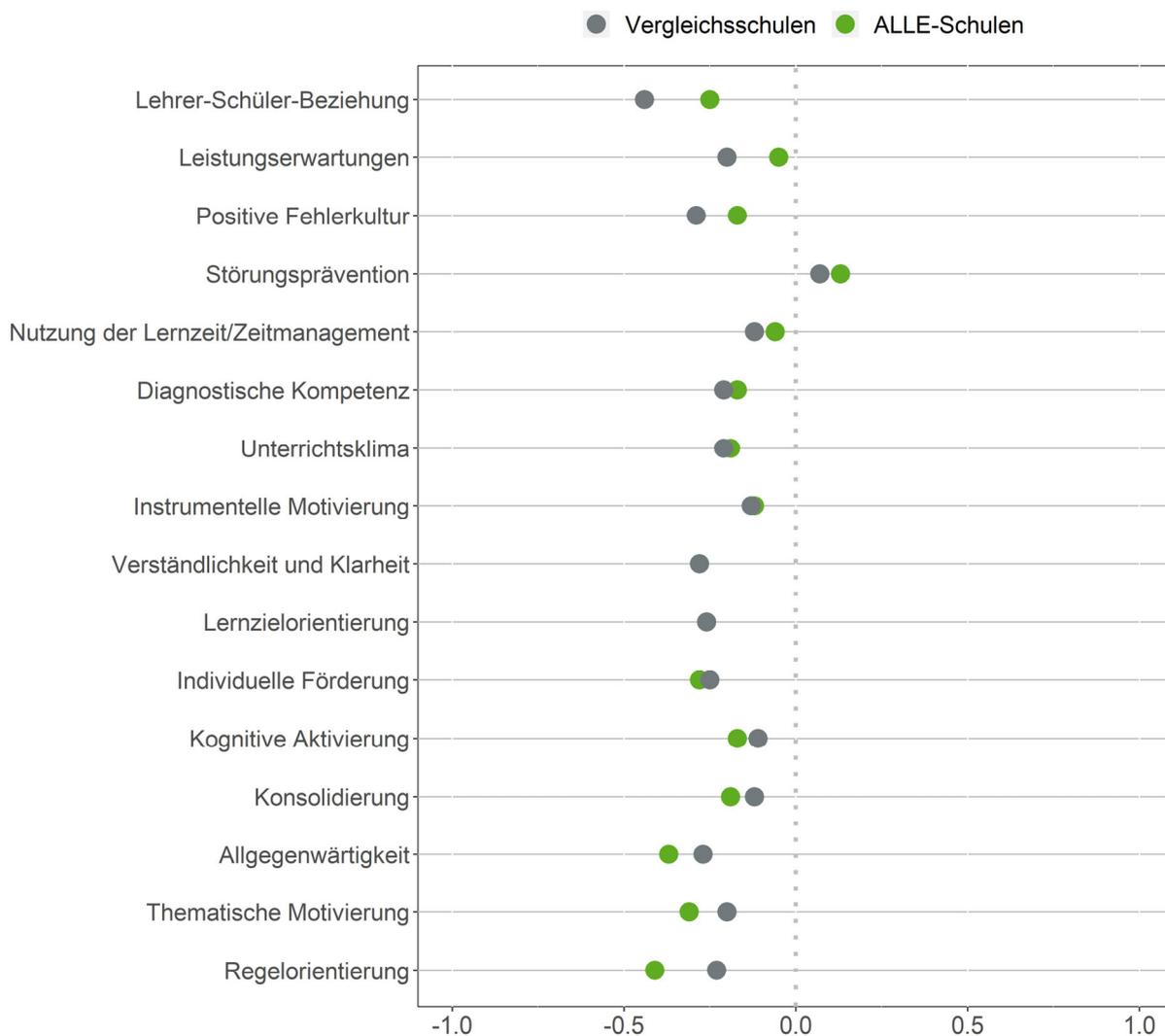
Tabelle 48 (Fortsetzung). Unterrichtsqualität im Fach Mathematik aus Sicht der Schülerinnen und Schüler in den ALLE-Schulen und Vergleichsschulen für die Messzeitpunkte der ALLE-Evaluation

Unterrichtsmerkmal	Messzeitpunkt	ALLE-Schulen		Vergleichsschulen		Cohens <i>d</i>
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Lehrer-Schüler-Beziehung	T1	3.39	0.63	3.45	0.56	-0.09
	T3	3.23	0.66	3.22	0.61	0.02
	T4	3.15	0.79	3.08	0.71	0.10
Diagnostische Kompetenz	T1	3.18	0.64	3.13	0.64	0.09
	T3	3.01	0.66	2.92	0.70	0.13
	T4	2.83	0.90	2.87	0.74	-0.05
Individuelle Förderung	T1	3.17	0.60	3.18	0.55	-0.02
	T3	3.02	0.57	2.94	0.62	0.14
	T4	2.88	0.79	2.88	0.67	0.00
Allgegenwärtigkeit	T1	3.21	0.68	3.17	0.68	0.06
	T3	3.15	0.66	3.03	0.74	0.17
	T4	2.87	0.91	2.93	0.74	-0.07
Regelorientierung	T1	3.38	0.65	3.47	0.60	-0.13
	T3	3.32	0.62	3.32	0.65	-0.01
	T4	2.90	0.99	3.16	0.73	-0.30
Nutzung der Lernzeit/ Zeitmanagement	T1	3.17	0.61	3.23	0.53	-0.11
	T3	3.11	0.59	3.08	0.56	0.04
	T4	2.92	0.82	2.99	0.64	-0.09
Störungsprävention	T1	3.03	0.95	3.12	0.88	-0.10
	T3	3.15	0.88	3.13	0.86	0.02
	T4	3.07	0.93	3.03	0.84	0.05

Anmerkungen. T1 = Anfang 7. Klasse, T3 = Ende 8. Klasse, T4 = Ende 9. Klasse. ALLE-Schulen $N = 347$ (23 Klassen bzw. Lerngruppen), Vergleichsschulen $N = 300$ (23 Klassen bzw. Lerngruppen). Die Ergebnisse beruhen auf den Angaben derjenigen Schülerinnen und Schüler, die ab T1 von derselben Lehrperson unterrichtet wurden (vgl. Einleitung Kap. 4.9). Die Ergebnisse für die messzeitpunkt-spezifischen Stichproben können Anhang 7.3 entnommen werden.

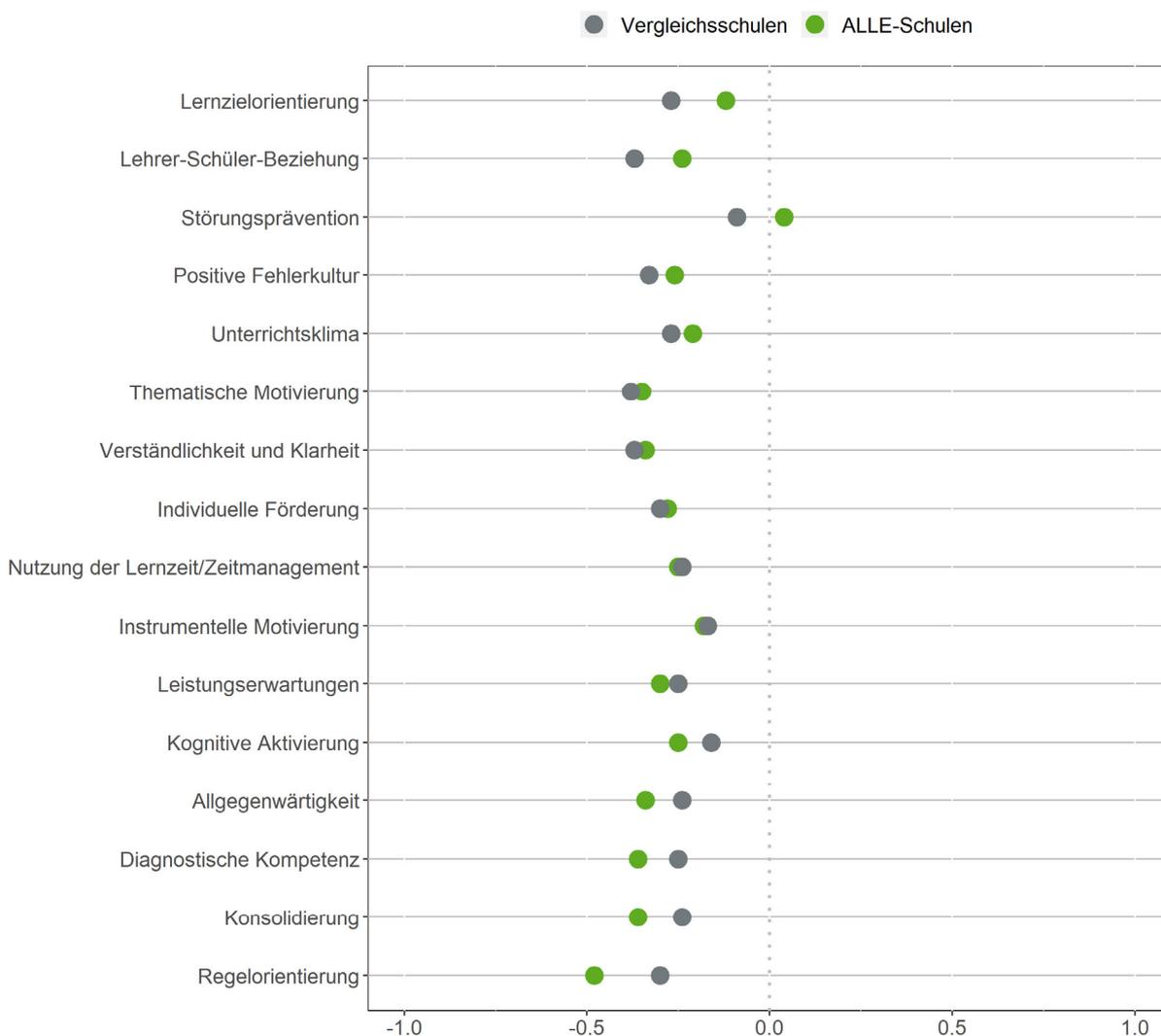
Abbildung 56 und Abbildung 57 veranschaulichen die Veränderungen der wahrgenommenen Unterrichtsqualität für die Fächer Deutsch und Mathematik zwischen dem ersten und letzten Messzeitpunkt (Ende 9. vs. Anfang 7. Klassenstufe).

Abbildung 56. Veränderung der Qualitätsmerkmale des Unterrichts aus Sicht der Schülerinnen und Schüler in den ALLE-Schulen und Vergleichsschulen im Fach Deutsch



Anmerkungen. ALLE-Schulen $N = 369$ (23 Klassen bzw. Lerngruppen), Vergleichsschulen $N = 302$ (23 Klassen bzw. Lerngruppen). Eingetragen sind Mittelwertsdifferenzen auf der vierstufigen Antwortskala (1, *stimme überhaupt nicht zu* bis 4, *stimme völlig zu*) zwischen dem vierten und ersten Messzeitpunkt in den ALLE-Schulen bzw. Vergleichsschulen, basierend auf den Angaben der Schülerinnen und Schüler, welche seit Beginn der Evaluation ALLE (ab dem ersten Messzeitpunkt) von derselben Fachlehrperson unterrichtet wurden. Die Angaben sind nach der Grösse der resultierenden Differenzen zugunsten der ALLE-Schulen sortiert. Im Betrag höhere Werte entsprechen einer stärkeren Veränderung; positive Werte entsprechen einer Zunahme, negative Werte einer Abnahme der Zustimmung.

Abbildung 57. Veränderungen der Qualitätsmerkmale des Unterrichts aus Sicht der Schülerinnen und Schüler in den ALLE-Schulen und Vergleichsschulen im Fach Mathematik



Anmerkungen. ALLE-Schulen: $N = 347$ (23 Klassen bzw. Lerngruppen), Vergleichsschulen: $N = 300$ (23 Klassen bzw. Lerngruppen). Eingezeichnet sind Mittelwertsdifferenzen auf der vierstufigen Antwortskala (1, *stimme überhaupt nicht zu* bis 4, *stimme völlig zu*) zwischen dem vierten und ersten Messzeitpunkt in den ALLE-Schulen bzw. Vergleichsschulen, basierend auf den Angaben der Schülerinnen und Schüler, welche seit Beginn der Evaluation ALLE (ab dem ersten Messzeitpunkt) von derselben Fachlehrperson unterrichtet wurden. Die Angaben sind nach der Grösse der resultierenden Differenzen zugunsten der ALLE-Schulen sortiert. Im Betrag höhere Werte entsprechen einer stärkeren Veränderung; positive Werte entsprechen einer Zunahme, negative Werte einer Abnahme der Zustimmung.

Über alle Unterrichtsmerkmale hinweg lässt sich feststellen, dass zum Ende der 9. Klassenstufe die Zustimmung der Schülerinnen und Schüler zu einer hohen Qualität deskriptiv leicht abnimmt, sowohl bei den ALLE-Schulen als auch den Vergleichsschulen und in beiden Fächern. Allerdings

bewegt sich die Bewertung der Unterrichtsmerkmale relativ zum theoretischen Mittelwert der Antwortskala weiterhin im positiven Bereich. Zwischen den ALLE- und Vergleichsschulen deuten sich dabei zwar Unterschiede in der Veränderung in den Unterrichtsqualitätsmerkmalen an, diese sind aber in ihrer Tendenz – zugunsten oder zu Ungunsten der ALLE-Schulen – inkonsistent und zudem durchweg nicht statistisch signifikant.

4.9.3 Unterrichtsqualität aus Sicht der Lehrpersonen

Um einen umfassenden Eindruck vom Unterricht in den teilnehmenden Schulen zu erhalten, wurden nicht nur die Schülerinnen und Schüler, sondern auch ihre Fachlehrpersonen am ersten und letzten Messzeitpunkt der Evaluation (Anfang 7. und Ende 9. Klassenstufe) zu wichtigen Aspekten der Unterrichtsqualität befragt. Im Folgenden werden Befunde zu den einzelnen Messzeitpunkten und zur Veränderung über den Evaluationszeitraum analog zur Darstellung der Befunde für die Schülerinnen und Schüler präsentiert. Anders als in der vorherigen Ergebnisdarstellung wurde aufgrund der eingeschränkten Stichprobengrösse auf die Prüfung von Unterschieden zwischen den ALLE-Schulen und Vergleichsschulen mit Signifikanztests verzichtet (vgl. Kap. 3.6). Zudem war die interne Konsistenz von drei der Skalen («Verständlichkeit und Klarheit» im Fach Deutsch zum ersten Messzeitpunkt; «Individuelle Förderung» in beiden Fächern zum vierten Messzeitpunkt) nicht zufriedenstellend. Ergebnisse werden auch für diese Skalen dargestellt, doch muss aufgrund dieser Einschränkung die Interpretation nochmals zurückhaltender erfolgen.

Tabelle 49 gibt die Ergebnisse für das Fach Deutsch getrennt für die ALLE-Schulen und Vergleichsschulen wieder. Die Mittelwerte aller Unterrichtsmerkmale liegen durchweg über dem theoretischen Mittel der Antwortskala von 3.5, die Deutsch-Lehrpersonen beschreiben ihren Unterricht hinsichtlich der ausgewählten Qualitätsmerkmale also überwiegend positiv. Fokussiert man auf Unterschiede von zumindest kleiner Effektstärke ($d \geq 0.15$; vgl. Kap. 3.6), resultiert zum ersten Messzeitpunkt (T1) ein höherer Mittelwert zugunsten der ALLE-Schulen für das Merkmal «Kognitive Aktivierung» ($d = 0.38$), ein höherer Mittelwert zugunsten der Vergleichsschulen beim Merkmal «Thematische Motivierung» ($d = -0.19$). Zum vierten Messzeitpunkt (T4) ergeben sich höhere Mittelwerte in den ALLE-Schulen für die Merkmale «Individuelle Förderung» ($d = 0.72$), «Verständlichkeit und Klarheit» ($d = 0.63$), «Kognitive Aktivierung» ($d = 0.44$), «Nutzung der Lernzeit» ($d = 0.23$) und «Leistungserwartungen» ($d = 0.20$). Für die Störungsprävention ergibt sich ein höherer Wert in den Vergleichsschulen ($d = -0.39$). Dies deutet darauf hin, dass die Lehrpersonen der ALLE-Schulen ihren Unterricht insbesondere am Ende des Evaluationszeitraums positiver wahrnehmen als die Lehrpersonen der Vergleichsschulen.

Tabelle 49. Unterrichtsqualität im Fach Deutsch aus Sicht der Lehrpersonen in den ALLE-Schulen und Vergleichsschulen für die Messzeitpunkte der ALLE-Evaluation

Unterrichtsmerkmal	Messzeitpunkt	ALLE-Schulen		Vergleichsschulen		Cohens <i>d</i>
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Verständlichkeit und Klarheit	T1	5.08	0.43	5.14	0.44	-0.13
	T4	4.92	0.48	4.64	0.42	0.63
Kognitive Aktivierung	T1	5.27	0.49	5.08	0.57	0.38
	T4	5.10	0.45	4.90	0.55	0.44
Thematische Motivierung	T1	5.35	0.79	5.50	0.80	-0.19
	T4	5.41	0.62	5.42	0.67	-0.01
Leistungserwartungen	T1	5.18	1.07	5.25	0.87	-0.08
	T4	5.53	0.51	5.42	0.67	0.20
Unterrichtsklima	T1	5.40	0.46	5.48	0.72	-0.13
	T4	5.26	0.48	5.28	0.57	-0.05
Individuelle Förderung	T1	4.60	0.86	4.61	0.68	-0.01
	T4	4.86	0.44	4.56	0.41	0.72
Störungsprävention	T1	5.25	0.51	5.17	0.83	0.10
	T4	4.88	0.89	5.17	0.58	-0.39
Nutzung der Lernzeit/ Zeitmanagement	T1	5.25	0.68	5.17	0.84	0.12
	T4	5.31	0.60	5.17	0.72	0.23

Anmerkungen. T1 = Anfang 7. Klasse, T4 = Ende 9. Klasse. ALLE-Schulen $N_{T1} = 16 - 17$, $N_{T4} = 16 - 17$ und Vergleichsschulen: $N_{T1} = 12$, $N_{T4} = 12$. Die Ergebnisse beruhen auf den Angaben derjenigen Lehrpersonen, die ab T1 dieselbe Klasse bzw. Lerngruppe unterrichtet haben (vgl. Einleitung Kap. 4.9). Die Ergebnisse für die messzeitpunktspezifischen Stichproben können Anhang 7.3 entnommen werden.

Tabelle 50 enthält die Resultate für das Fach Mathematik. Es zeigt sich, dass auch die Mathematik-Lehrpersonen alle Unterrichtsqualitätsmerkmale tendenziell positiv, oberhalb des theoretischen Mittelwerts der Antwortskala von 3.5, beurteilen. Betrachtet man die Ergebnisse zum ersten Messzeitpunkt (T1), stellt man fest, dass sich für die Merkmale «Thematische Motivierung» ($d = 2.26$), «Verständlichkeit und Klarheit» ($d = 0.81$), «Individuelle Förderung» ($d = 0.71$) und «Unterrichtsklima» ($d = 0.56$) grosse Unterschiede zugunsten der ALLE-Schulen ergeben. Demgegenüber zeigen sich für die Merkmale «Nutzung der Lernzeit» ($d = -1.55$), «Störungsprävention» ($d = -0.67$) und «Kognitive Aktivierung» ($d = -0.28$) im Umfang grosse bis eher geringe Mittelwertsdifferenzen zugunsten der Vergleichsschulen. Zum vierten Messzeitpunkt (T4) resultieren dagegen in den ALLE-Schulen höhere Mittelwerte für fast alle Unterrichtsmerkmale. Die Unterschiede für die Merk-

male «Thematische Motivierung» ($d = 2.07$), «Verständlichkeit und Klarheit» ($d = 1.99$), «Individuelle Förderung» ($d = 1.18$), «Kognitive Aktivierung» ($d = 0.94$) und «Störungsprävention» ($d = 0.61$) sind als gross einzuordnen. Mittlere Unterschiede sind bei den Merkmalen «Leistungserwartungen» ($d = 0.53$) und «Unterrichtsklima» ($d = 0.49$) festzustellen. Für die «Nutzung der Lernzeit» ergibt sich ein höherer Wert in den Vergleichsschulen ($d = -1.35$). Damit finden sich wie im Fach Deutsch Anzeichen, dass die Lehrpersonen der ALLE-Schulen ihren Unterricht insbesondere am Ende der Evaluation besser beurteilen als die Lehrpersonen der Vergleichsschulen.

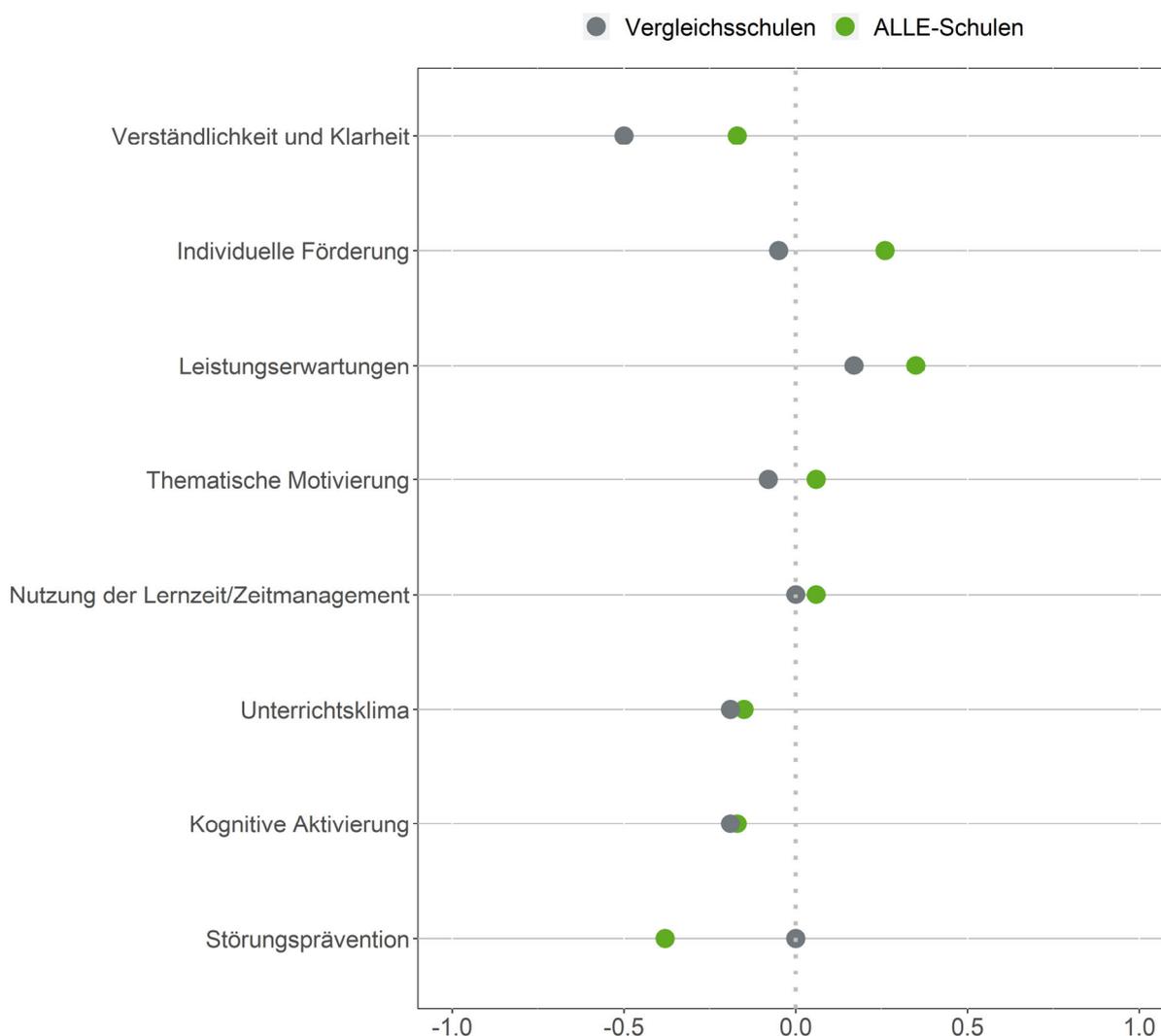
Tabelle 50. Unterrichtsqualität im Fach Mathematik aus Sicht der Lehrpersonen in den ALLE-Schulen und Vergleichsschulen für die Messzeitpunkte der ALLE-Evaluation

Unterrichtsmerkmal	Messzeitpunkt	ALLE-Schulen		Vergleichsschulen		Cohens d
		M	SD	M	SD	
Verständlichkeit und Klarheit	T1	5.07	0.72	4.42	1.10	0.81
	T4	5.13	0.66	4.08	0.50	1.99
Kognitive Aktivierung	T1	5.07	0.76	5.25	0.74	-0.28
	T4	5.15	0.78	4.44	0.94	0.94
Thematische Motivierung	T1	6.00	0.00	4.75	0.96	2.26
	T4	5.80	0.45	4.00	1.41	2.07
Leistungserwartungen	T1	5.40	0.89	5.50	1.00	-0.12
	T4	5.60	0.55	5.25	0.96	0.53
Unterrichtsklima	T1	5.48	0.52	5.05	1.18	0.56
	T4	5.40	0.62	5.05	1.02	0.49
Individuelle Förderung	T1	4.77	0.57	4.33	0.82	0.71
	T4	5.20	0.57	4.50	0.79	1.18
Störungsprävention	T1	5.00	0.71	5.50	1.00	-0.67
	T4	5.20	0.84	4.50	1.73	0.61
Nutzung der Lernzeit/ Zeitmanagement	T1	5.40	0.55	6.00	0.00	-1.55
	T4	5.20	0.84	6.00	0.00	-1.35

Anmerkungen. T1 = Anfang 7. Klasse, T4 = Ende 9. Klasse. ALLE-Schulen $N_{T1} = 5$, $N_{T4} = 5$ und Vergleichsschulen $N_{T1} = 3 - 4$, $N_{T4} = 3 - 4$. Die Ergebnisse beruhen auf den Angaben derjenigen Lehrpersonen, die ab T1 dieselbe Klasse bzw. Lerngruppe unterrichtet haben (vgl. Einleitung Kap. 4.9). Die Ergebnisse für die messzeitpunktspezifischen Stichproben können Anhang 7.3 entnommen werden.

Abbildung 58 und Abbildung 59 zeigen die Veränderung der von den Lehrpersonen wahrgenommenen Unterrichtsqualität für die Fächer Deutsch und Mathematik zwischen dem ersten und letzten Messzeitpunkt (Ende 9. vs. Anfang 7. Klassenstufe). Im Fach Deutsch lässt sich in den ALLE-Schulen bei je der Hälfte der Merkmale eine wahrgenommene Zunahme beziehungsweise Abnahme der Qualität zum Ende der 9. Klasse feststellen. Das Urteil in den Vergleichsschulen fällt tendenziell kritischer aus, nur beim Merkmal «Störungsprävention» wird die Entwicklung von den Lehrpersonen der Vergleichsschulen günstiger beurteilt als von den Lehrpersonen der ALLE-Schulen.

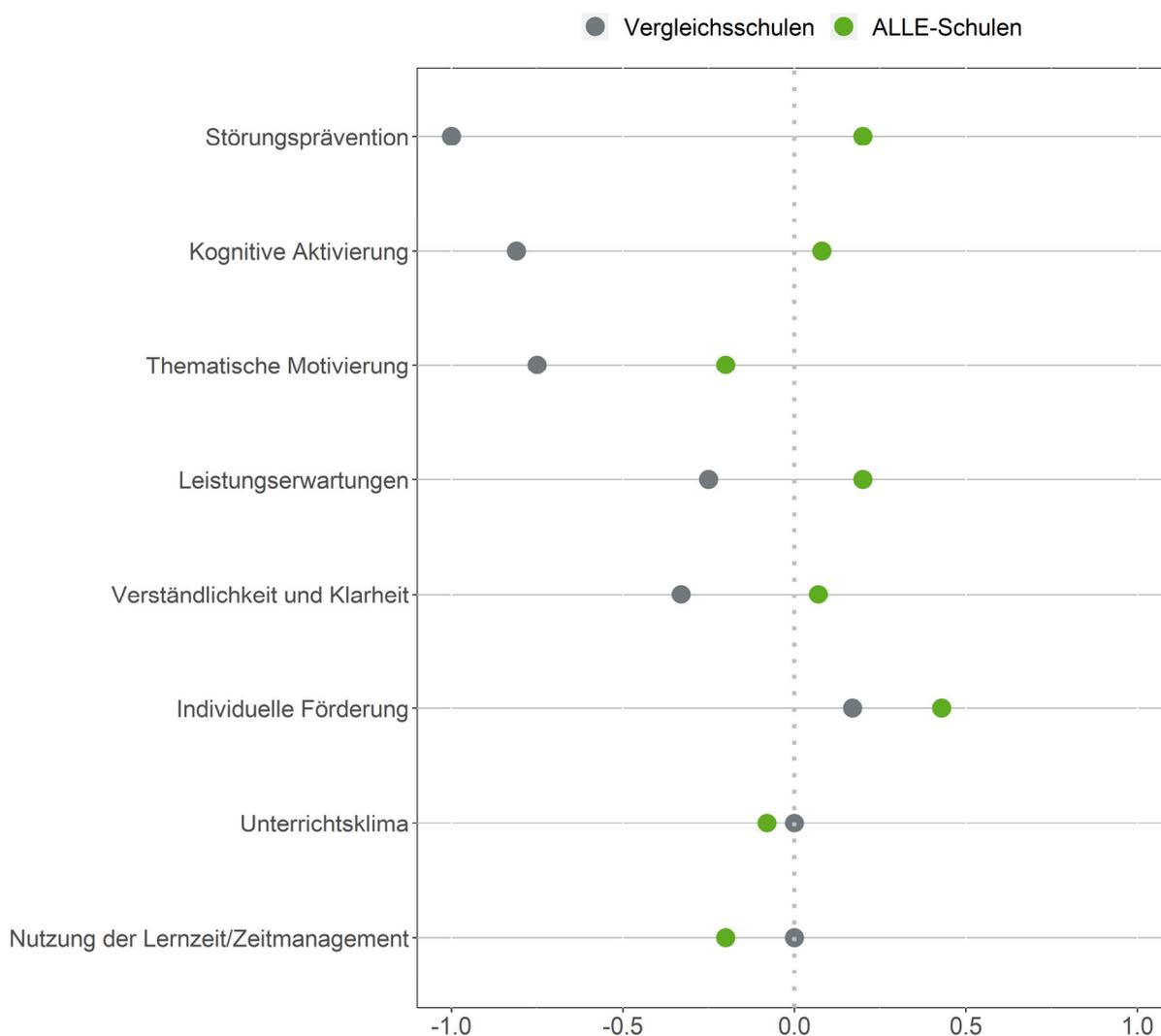
Abbildung 58. Veränderungen der Qualitätsmerkmale des Unterrichts aus Sicht der Lehrpersonen in den ALLE-Schulen und Vergleichsschulen im Fach Deutsch



Anmerkungen. ALLE-Schulen $N_{T1} = 16 - 17$, $N_{T4} = 16 - 17$ und Vergleichsschulen $N_{T1} = 12$, $N_{T4} = 12$. Eingetragen sind Mittelwertsdifferenzen auf der sechsstufigen Antwortskala (1, *stimme überhaupt nicht zu* bis 6, *stimme völlig zu*) zwischen dem vierten und ersten Messzeitpunkt in den ALLE- bzw. Vergleichsschulen, basierend auf den Angaben der Lehrpersonen, die seit dem ersten Messzeitpunkt dieselben Schülerinnen und Schüler unterrichtet haben. Die Angaben sind nach Grösse der Differenzen zugunsten der ALLE-Schulen sortiert. Im Betrag höhere Werte entsprechen einer stärkeren Veränderung; positive Werte entsprechen einer Zunahme, negative Werte einer Abnahme der Zustimmung.

Ein ähnliches Bild resultiert für das Fach Mathematik. Für die Vergleichsschulen zeigt sich jedoch bei mehreren Merkmalen (z.B. «Störungsprävention», «Kognitive Aktivierung») ein stärkerer Rückgang der Zustimmung zu einer hohen Unterrichtsqualität. Die Entwicklung der Merkmale «Unterrichtsklima» und «Nutzung der Lernzeit» wird zwar in den Vergleichsschulen günstiger beurteilt, der Unterschied zu den ALLE-Schulen ist jedoch gering. Allerdings ist der Stichprobenumfang äusserst beschränkt (< 10 Personen), zuverlässige Rückschlüsse auf die Gesamtheit der an ALLE beteiligten Mathematik-Lehrpersonen sind auf dieser Grundlage kaum möglich.

Abbildung 59. Veränderungen der Qualitätsmerkmale des Unterrichts aus Sicht der Lehrpersonen in den ALLE-Schule und Vergleichsschulen im Fach Mathematik



Anmerkungen. ALLE-Schulen $N_{T1} = 5$, $N_{T4} = 5$ und Vergleichsschulen $N_{T1} = 3 - 4$, $N_{T4} = 3 - 4$. Eingezeichnet sind Mittelwertsdifferenzen auf der sechsstufigen Antwortskala (1, *stimme überhaupt nicht zu* bis 6, *stimme völlig zu*) zwischen dem vierten und ersten Messzeitpunkt in den ALLE- bzw. Vergleichsschulen, basierend auf den Angaben der Lehrpersonen, die seit dem ersten Messzeitpunkt dieselben Schülerinnen und Schüler unterrichtet haben. Die Angaben sind nach Grösse der Differenzen zugunsten der ALLE-Schulen sortiert. Im Betrag höhere Werte entsprechen einer stärkeren Veränderung; positive Werte entsprechen einer Zunahme, negative Werte einer Abnahme der Zustimmung.

4.10 Lernergebnisse der Schülerinnen und Schüler

Die Frage, inwieweit mithilfe des an den ALLE-Schulen implementierten Massnahmenbündels positive Wirkungen im Hinblick auf die Lernergebnisse der Schülerinnen und Schüler erzielt werden konnten, zählt zu den zentralen Gegenständen der vorliegenden Evaluation. Ihr wird in den nachfolgenden Abschnitten zunächst bezogen auf wichtige motivational-affektive Merkmale nachgegangen (Kap. 4.10.1) dann bezogen auf die fachlichen Lernergebnisse der Schülerinnen und Schüler (Kap. 4.10.2), das heisst ihre Leistungen in den Deutsch- und Mathematiktests.

In beiden Abschnitten werden die *Veränderungen beziehungsweise Zuwächse* dieser Merkmale im Evaluationszeitraum untersucht. Dazu werden die Unterschiede (Differenzen) zwischen den Messzeitpunkten auf den verwendeten Punktwertskalen bestimmt und analysiert. Dies erfolgt zunächst deskriptiv und getrennt für die ALLE- und Vergleichsschulen. Auf eine statistische Prüfung der Zuwächse innerhalb der beiden Gruppen wird verzichtet, da der Fokus auf dem nachfolgenden Vergleich der ALLE- und Vergleichsschulen liegt (vgl. Kap. 3.6). Nach der deskriptiven Darstellung wird beschrieben und anhand von statistischen Signifikanztests geprüft, inwieweit sich die ALLE- und Vergleichsschulen in den Veränderungen im Evaluationszeitraum unterscheiden – und sich Belege dafür finden, dass die Teilnahme am Pilotprojekt ALLE mit günstigen Wirkungen auf die Lernergebnisse der Schülerinnen und Schüler einherging.

Diese Analysen gliedern sich in zwei Schritte: Im ersten Schritt wird der Unterschied zwischen den ALLE- und Vergleichsschulen in der Veränderung des betreffenden Merkmals bestimmt und statistisch geprüft. Dazu wird die Veränderung im Evaluationszeitraum als die zu erklärende Zielvariable (abhängige Variable) in einem Regressionsmodell analysiert. Die Zielvariable ist die Punktwertdifferenz zwischen dem letzten Messzeitpunkt (T4) und dem ersten Messzeitpunkt (T1) der Evaluation. Als erklärende Variable (Prädiktor bzw. unabhängige Variable) dient eine zweistufige (Dummy-)Variable (ALLE-Schulen = 1, Vergleichsschulen = 0), welche den Unterschied zwischen den ALLE- und Vergleichsschulen abbildet. Resultiert für diese Variable ein *positiver* Regressionskoeffizient, spricht dies für eine Veränderung *zugunsten der ALLE-Schulen*. Ein negativer Regressionskoeffizient spricht für eine Veränderung zugunsten der Vergleichsschulen.

Im zweiten Schritt wird untersucht, ob allfällige Differenzen in der Veränderung zwischen den ALLE- und Vergleichsschulen auch dann fortbestehen, wenn Unterschiede in leistungsrelevanten Hintergrundmerkmalen der Schülerinnen und Schüler statistisch berücksichtigt werden. Dafür werden die Hintergrundmerkmale als zusätzliche erklärende Variablen in die Regressionsmodelle einbezogen. Damit wird der Tatsache Rechnung getragen, dass günstigere Veränderungen in den ALLE-Schulen auch günstigeren Lernvoraussetzungen der Schülerinnen und Schüler (z.B. besseren Sprachkenntnissen) geschuldet sein können. Weitere Erläuterungen zur methodischen Vorgehensweise finden sich in Kapitel 3.6.

Bei den berichteten Analysen wird zunächst jeweils die gesamte verfügbare Stichprobe der Evaluation ALLE herangezogen. In beiden Abschnitten werden aber zusätzliche Auswertungen dargestellt, in denen auf die leistungsschwächeren Schülerinnen und Schüler fokussiert wird. Die berichteten Resultate sollen Aufschluss darüber geben, inwieweit ein wichtiges Ziel von ALLE – die leistungsschwächeren Schülerinnen und Schüler besonders zu fördern – erreicht werden konnte.

4.10.1 Motivational-affektive Lernergebnisse

In der ersten (Anfang der 7. Klassenstufe) und vierten Erhebung (Ende der 9. Klassenstufe) der Evaluation ALLE wurden wichtige motivationale Orientierungen – die intrinsische Motivation, die anerkennungs- und leistungsbezogene Motivation und die zukunftsbezogene instrumentelle Motivation – sowie die Selbstwirksamkeit der Schülerinnen und Schüler immer mit identischen Fragen (Fragebogen-Items) erhoben. Die intrinsische Motivation wurde zusätzlich in der dritten Erhebung (Ende der 8. Klassenstufe) erfasst. Dadurch ist es möglich, die Merkmale in ihrer Veränderung zwischen der 7. und 9. Klassenstufe zu beschreiben.

Die motivationalen Orientierungen und die Selbstwirksamkeit der Schülerinnen und Schüler wurden jeweils getrennt für die Fächer Deutsch und Mathematik erfasst. Tabelle 51 zeigt die Anzahl der Fragebogen-Items je Skala sowie Beispielitems der Skalen zur Erfassung der motivationalen Orientierungen und der Selbstwirksamkeit (für weitere Erläuterungen siehe unten).

Tabelle 51. Skalen zu den motivationalen Orientierungen und zur Selbstwirksamkeit der Schülerinnen und Schüler

Merkmal	Anzahl Items pro Fach	Beispielitem
Intrinsische Motivation ^a	4	Ich habe Spass an Deutsch/Mathematik.
Anerkennungs- und leistungsbezogene Motivation ^a	5	Ich möchte in Deutsch/Mathematik gut sein, damit ich ein gutes Zeugnis bekomme.
Zukunftsbezogene instrumentelle Motivation ^a	4	Ich strengte mich im Deutsch-/Mathematikunterricht an, weil es für meinen späteren Beruf nützlich sein wird.
Selbstwirksamkeit ^b	7/6	Texte lesen und verstehen/für einen Würfel mit gegebenem Rauminhalt (z.B. 1 m ³) die Kantenlänge berechnen

Anmerkungen. Die Selbstwirksamkeit wurde in jedem Fach anhand von bereichs- oder aufgabenspezifischen Items erfasst (siehe die Angaben im Text).

^a Beurteilung auf einer vierstufigen Antwortskala (1, *stimme überhaupt nicht zu* bis 4, *stimme völlig zu*).

^b Im Fach Deutsch: Beurteilung auf einer sechsstufigen Skala (1, *ungenügend* bis 4, *sehr gut*). Im Fach Mathematik: Beurteilung auf einer vierstufigen Skala (1, *überhaupt nicht sicher* bis 4, *sehr sicher*).

Die Selbstwirksamkeit wurde in jedem Fach anhand von bereichs- oder aufgabenspezifischen Items erfasst. In Deutsch haben die Schülerinnen und Schüler auf einer sechsstufigen Skala (*ungenügend, mangelhaft, ausreichend, befriedigend, gut, sehr gut*) ihre Fähigkeiten in sieben Bereichen des Fachs eingeschätzt. Die Bereiche waren (a) deutliche und korrekte Aussprache, (b) Texte schreiben und überarbeiten, (c) Rechtschreibung und Zeichensetzung, (d) richtige Grammatik beim Sprechen, (e) richtige Grammatik beim Schreiben, (f) gesprochene Texte verstehen und (g) Texte lesen und verstehen. In Mathematik beurteilten die Schülerinnen und Schüler auf einer vierstufigen Skala (*überhaupt nicht sicher, eher nicht sicher, eher sicher, sehr sicher*), wie sicher sie sind, dass sie folgende sechs Mathematikaufgaben lösen können: (a) für einen Würfel mit gegebenem Rauminhalt (z.B. 1 m^3) die Kantenlänge berechnen, (b) zwei Linien in einem rechten Winkel zu einem Rechteck ergänzen, (c) für eine gegebene dreistellige Zahl bestimmen, ob diese durch 2, 3 und 4 teilbar ist, (d) mehrere vorgegebene Brüche der Grösse nach ordnen, (e) bestimmen, wie viele dreistellige Zahlen mit Endziffer 4 es gibt, (f) das Gewicht 239'000 kg in Tonnen umrechnen.

In den folgenden beiden Kapiteln werden Resultate zu den motivational-affektiven Lernergebnissen der Schülerinnen und Schüler für die ALLE- und Vergleichsschulen berichtet. Im ersten Kapitel werden die Lernergebnisse der gesamten Stichprobe aus den ALLE- und Vergleichsschulen einander gegenübergestellt (Kap. 4.10.1.1). Im zweiten Abschnitt erfolgt der Vergleich zwischen ALLE- und Vergleichsschulen mit Blick auf diejenigen Schülerinnen und Schüler, welche zu Beginn der Evaluation ALLE (erster Messzeitpunkt) über die geringsten fachlichen Kompetenzen, das heisst eine ungünstige Lernausgangslage, verfügten (Kap. 4.10.1.2). Um die Ergebnisse zu präsentieren, wurden aus den Antworten der Schülerinnen und Schüler auf die einzelnen Fragebogen-Items jeder Skala Mittelwerte gebildet. Höhere Werte entsprechen einer höheren Motivation beziehungsweise höheren Selbstwirksamkeit der Schülerinnen und Schüler.

4.10.1.1 Lernergebnisse in der gesamten Evaluationsstichprobe

Tabelle 52 sowie Abbildung 60 bis Abbildung 63 berichten die Resultate für die motivational-affektiven Lernergebnisse im Fach Deutsch getrennt für die ALLE- und Vergleichsschulen. Tabelle 52 enthält neben den Mittelwerten (Spalte *M*) auch die Standardabweichungen (Spalte *SD*) in der jeweiligen Stichprobe, das heisst das Ausmass der Streuung in den Merkmalsausprägungen der Schülerinnen und Schüler. Abbildung 60 bis Abbildung 63 veranschaulichen zusätzlich die Veränderungen in den mittleren Merkmalsausprägungen über die Messzeitpunkte.

Vergleichbar mit den Befunden anderer Untersuchungen (siehe z.B. die Resultate der Zürcher Lernstandserhebung; Angelone, Keller & Moser, 2013) ist am Ende der 9. Klassenstufe in den ALLE- und Vergleichsschulen ein Rückgang bei den motivationalen Orientierungen der Schülerinnen und Schülern feststellbar, wohingegen die Selbstwirksamkeit weitgehend unverändert bleibt. In den ALLE-Schulen zeigt sich ein im Umfang geringes Absinken der intrinsischen Motivation von

0.31 Punkten (2.78 Punkte – 2.47 Punkte; $d = -0.25$), in den Vergleichsschulen eine etwas stärkere Abnahme von 0.46 Punkten (2.71 Punkte – 2.25 Punkte; $d = -0.35$). Die anerkennungs- und leistungsbezogene Motivation sowie die zukunftsbezogene instrumentelle Motivation sinken in den ALLE-Schulen moderat um 0.40 Punkte (3.34 Punkte – 2.94 Punkte; $d = -0.38$) beziehungsweise 0.42 Punkte (3.29 Punkte – 2.88 Punkte; $d = -0.37$). Ein ähnliches Bild ergibt sich in den Vergleichsschulen bezüglich der zukunftsbezogenen instrumentellen Motivation mit einem Rückgang von 0.29 Punkten (3.20 Punkte – 2.91 Punkte, $d = -0.23$), während eine deutlich grössere Abnahme um 0.69 Punkte bei der anerkennungs- und leistungsbezogenen Motivation festzustellen ist (3.33 Punkte – 2.64 Punkte, $d = -0.61$). Hinsichtlich der Selbstwirksamkeit lassen sich sowohl in den ALLE-Schulen ($d = -0.04$) als auch den Vergleichsschulen ($d = -0.18$) nur geringfügige Veränderungen ausmachen.

Tabelle 52. Motivational-affektive Lernergebnisse in den ALLE- und Vergleichsschulen im Fach Deutsch für die Messzeitpunkte der ALLE-Evaluation

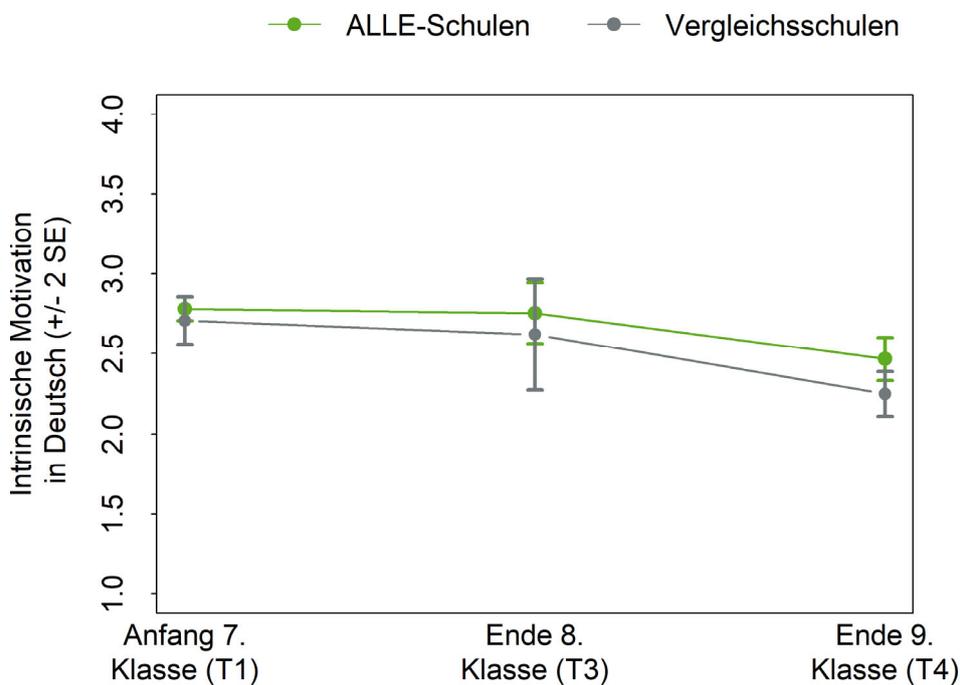
Merkmal	Messzeitpunkt	ALLE-Schulen			Vergleichsschulen		
		<i>M</i>	(<i>SE</i>)	<i>SD</i>	<i>M</i>	(<i>SE</i>)	<i>SD</i>
Intrinsische Motivation ^a	T1	2.78	(0.04)	0.73	2.71	(0.07)	0.85
	T3	2.75	(0.10)	0.82	2.62	(0.17)	0.86
	T4	2.47	(0.07)	0.90	2.25	(0.07)	0.93
Anerkennungs- und leistungsbezogene Motivation ^a	T1	3.34	(0.03)	0.56	3.33	(0.07)	0.64
	T4	2.94	(0.08)	0.78	2.64	(0.13)	0.86
Zukunftsbezogene instrumentelle Motivation ^a	T1	3.29	(0.03)	0.64	3.20	(0.11)	0.83
	T4	2.88	(0.10)	0.88	2.91	(0.10)	0.79
Selbstwirksamkeit ^b	T1	4.52	(0.05)	0.89	4.42	(0.10)	1.01
	T4	4.47	(0.08)	1.02	4.14	(0.10)	1.27

Anmerkungen. T1 = Anfang 7. Klasse, T3 = Ende 8. Klasse, T4 = Ende 9. Klasse. ALLE-Schulen $N = 970$ und Vergleichsschulen $N = 717$. Die Erfassung der anerkennungs- und leistungsbezogenen Motivation, der zukunftsbezogenen instrumentellen Motivation und der Selbstwirksamkeit erfolgte nur zum ersten und vierten Messzeitpunkt (T1 und T4).

^a Beurteilung auf einer vierstufigen Antwortskala (1, *stimme überhaupt nicht zu* bis 4, *stimme völlig zu*).

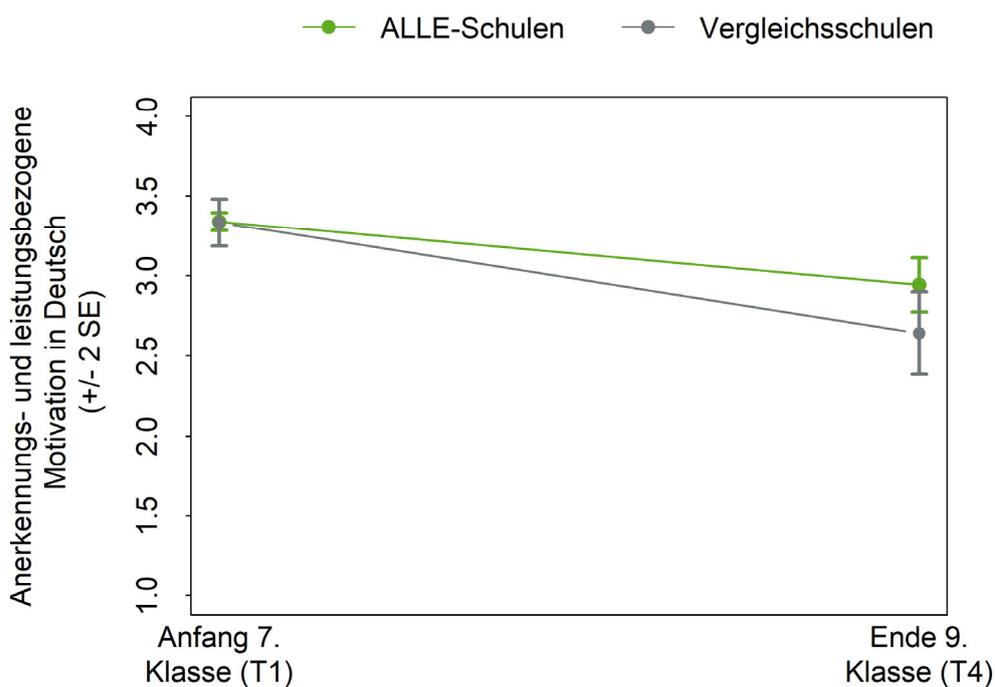
^b Beurteilung auf einer sechsstufigen Antwortskala (1, *ungenügend* bis 6, *sehr gut*).

Abbildung 60. Veränderung der intrinsischen Motivation in den ALLE- und Vergleichsschulen im Fach Deutsch



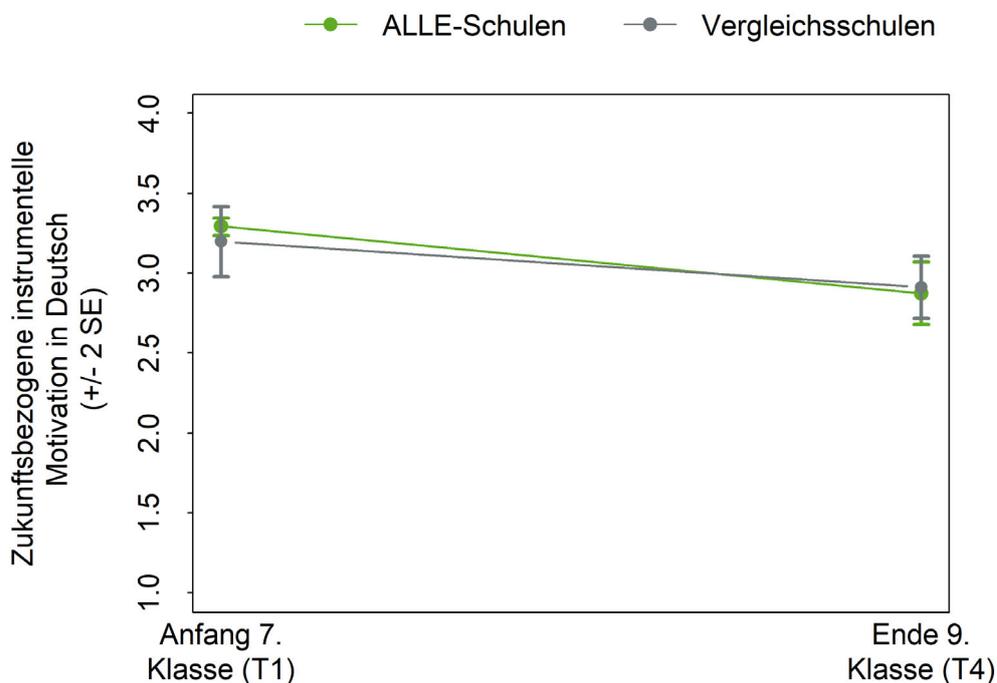
Anmerkungen. ALLE-Schulen $N = 970$ und Vergleichsschulen $N = 717$.

Abbildung 61. Veränderung der anerkennungs- und leistungsbezogenen Motivation in den ALLE- und Vergleichsschulen im Fach Deutsch



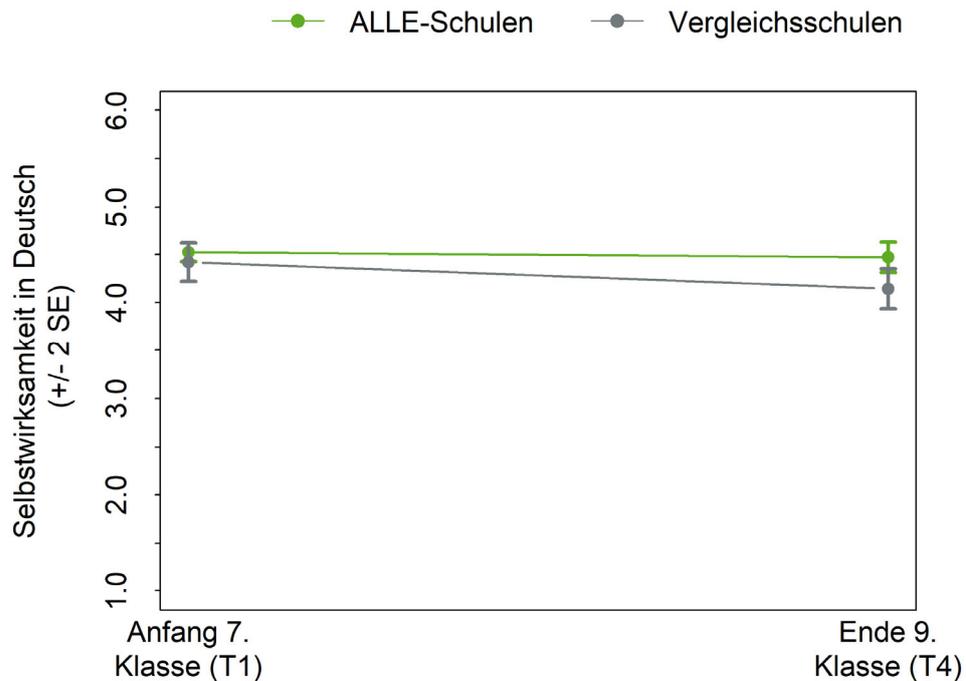
Anmerkungen. ALLE-Schulen $N = 970$ und Vergleichsschulen $N = 717$.

Abbildung 62. Veränderung der zukunftsbezogenen instrumentellen Motivation in den ALLE- und Vergleichsschulen im Fach Deutsch



Anmerkungen. ALLE-Schulen $N = 970$ und Vergleichsschulen $N = 717$.

Abbildung 63. Veränderung der Selbstwirksamkeit in den ALLE- und Vergleichsschulen im Fach Deutsch



Anmerkungen. ALLE-Schulen $N = 970$ und Vergleichsschulen $N = 717$.

Tabelle 53 und Abbildung 64 bis Abbildung 67 berichten analog zur vorherigen Darstellung die Resultate für das Fach Mathematik. Es zeigen sich in den ALLE- und Vergleichsschulen mit dem Fach Deutsch vergleichbare Rückgänge bei den motivationalen Orientierungen der Schülerinnen und Schüler. Wie im Fach Deutsch ist im Fach Mathematik kein nennenswerter Rückgang bei der Selbstwirksamkeit der Schülerinnen und Schüler festzustellen.

Tabelle 53. Motivational-affektive Lernergebnisse in den ALLE- und Vergleichsschulen im Fach Mathematik für die Messzeitpunkte der ALLE-Evaluation

Merkmal	Messzeitpunkt	ALLE-Schulen			Vergleichsschulen		
		<i>M</i>	(<i>SE</i>)	<i>SD</i>	<i>M</i>	(<i>SE</i>)	<i>SD</i>
Intrinsische Motivation ^a	T1	2.69	(0.05)	0.78	2.57	(0.07)	0.86
	T3	2.80	(0.07)	0.94	2.65	(0.13)	0.97
	T4	2.55	(0.07)	0.95	2.26	(0.07)	1.01
Anerkennungs- und leistungsbezogene Motivation ^a	T1	3.34	(0.02)	0.54	3.21	(0.07)	0.62
	T4	2.92	(0.06)	0.82	2.54	(0.09)	0.99
Zukunftsbezogene instrumentelle Motivation ^a	T1	3.20	(0.04)	0.62	2.95	(0.09)	0.84
	T4	2.92	(0.08)	0.91	2.65	(0.08)	0.95
Selbstwirksamkeit ^b	T1	2.91	(0.04)	0.60	2.71	(0.06)	0.80
	T4	2.98	(0.06)	0.87	2.72	(0.09)	0.99

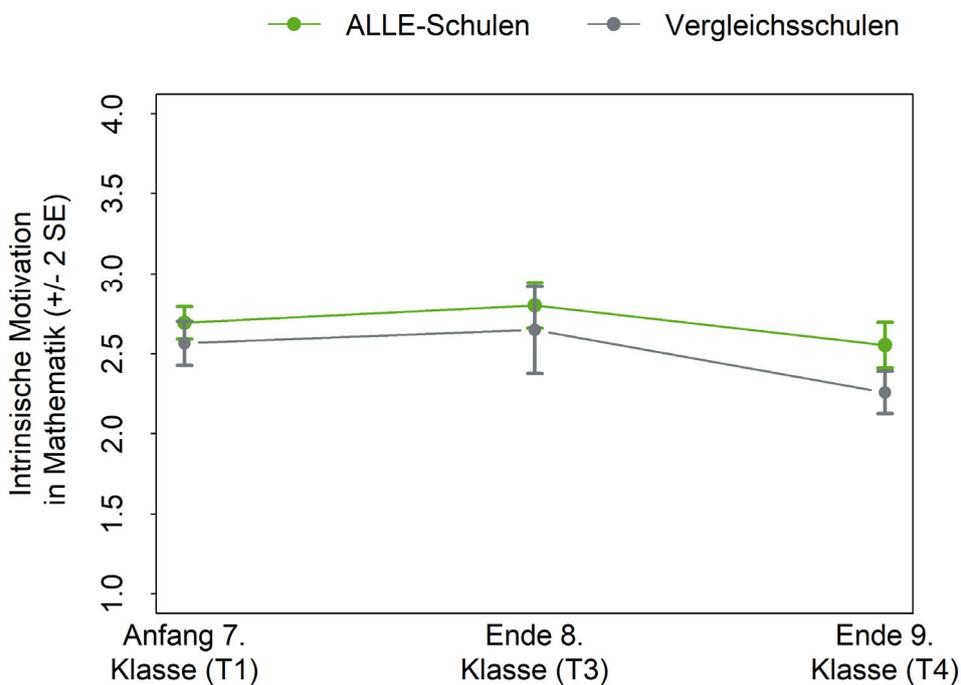
Anmerkungen. T1 = Anfang 7. Klasse, T3 = Ende 8. Klasse, T4 = Ende 9. Klasse. ALLE-Schulen $N = 970$ und Vergleichsschulen $N = 717$. Die Erfassung der anerkennungs- und leistungsbezogenen Motivation, der zukunftsbezogenen instrumentellen Motivation und der Selbstwirksamkeit erfolgte nur zum ersten und vierten Messzeitpunkt (T1 und T4).

^a Beurteilung auf einer vierstufigen Antwortskala (1, *stimme überhaupt nicht zu* bis 4, *stimme völlig zu*).

^b Beurteilung auf einer vierstufigen Antwortskala (1, *überhaupt nicht sicher* bis 4, *sehr sicher*).

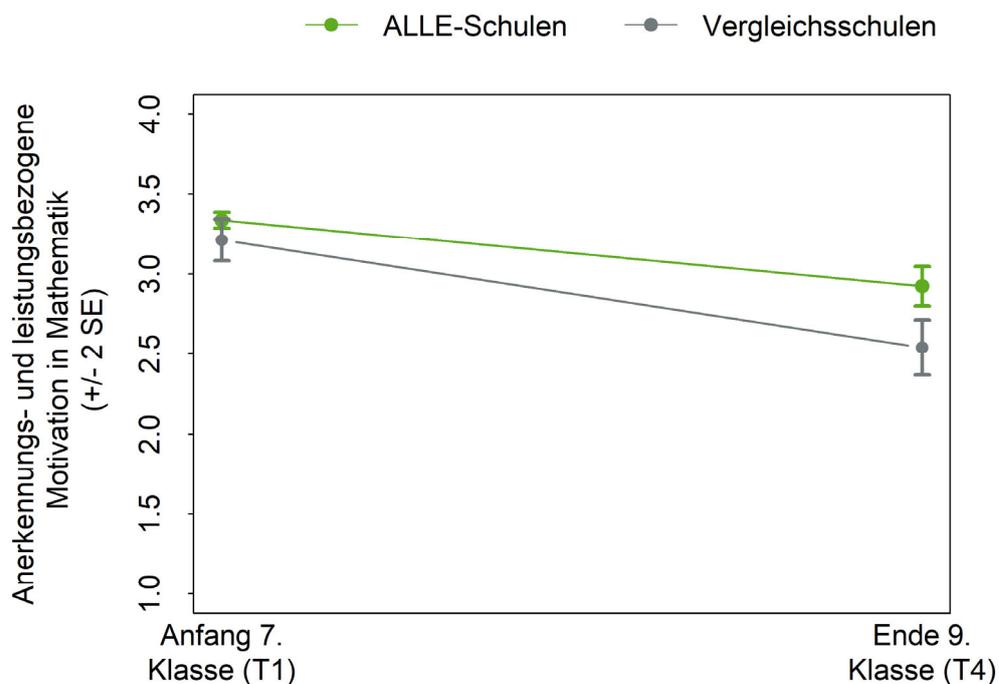
In den ALLE-Schulen zeigt sich zum Ende der 9. Klassenstufe ein geringer Rückgang der intrinsischen Motivation um 0.14 Punkte (2.69 Punkte – 2.55 Punkte; $d = -0.10$), der in den Vergleichsschulen mit 0.31 Punkten (2.57 Punkte – 2.26 Punkte, $d = -0.22$) etwas stärker ausgeprägt ist. Die anerkennungs- und leistungsbezogene sowie die zukunftsbezogene instrumentelle Motivation sinken in den ALLE-Schulen moderat um 0.42 Punkte (3.34 Punkte – 2.92 Punkte; $d = -0.39$) und 0.28 Punkte (3.20 Punkte – 2.92 Punkte; $d = -0.25$). In den Vergleichsschulen findet sich ein vergleichbarer Rückgang von 0.30 Punkten (2.95 Punkte – 2.65 Punkte, $d = -0.23$) für die zukunftsbezogene instrumentelle Motivation. Dagegen zeigt sich – ähnlich wie im Fach Deutsch – eine grössere Abnahme von 0.67 Punkten (3.21 Punkte – 2.54 Punkte; $d = -0.57$) bei der anerkennungs- und leistungsbezogenen Motivation. Für die Selbstwirksamkeit resultieren in den ALLE-Schulen ($d = 0.06$) und in den Vergleichsschulen ($d < 0.01$) jeweils nur minimale Veränderungen.

Abbildung 64. Veränderung der intrinsischen Motivation in den ALLE- und Vergleichsschulen im Fach Mathematik



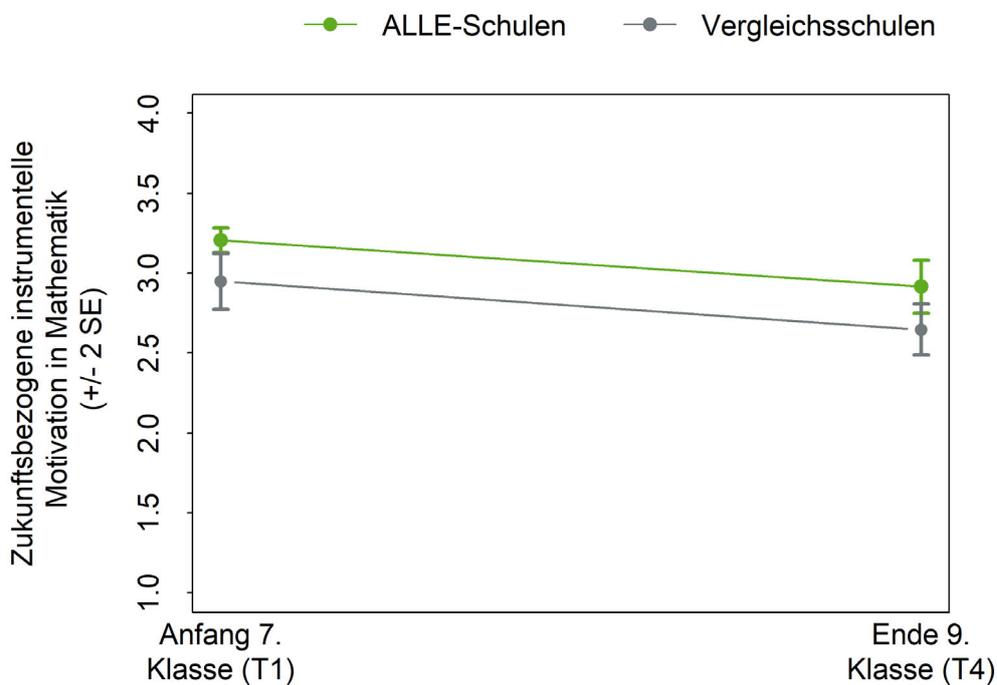
Anmerkungen. ALLE-Schulen $N = 970$ und Vergleichsschulen $N = 717$.

Abbildung 65. Veränderung der anerkennungs- und leistungsbezogenen Motivation in den ALLE- und Vergleichsschulen im Fach Mathematik



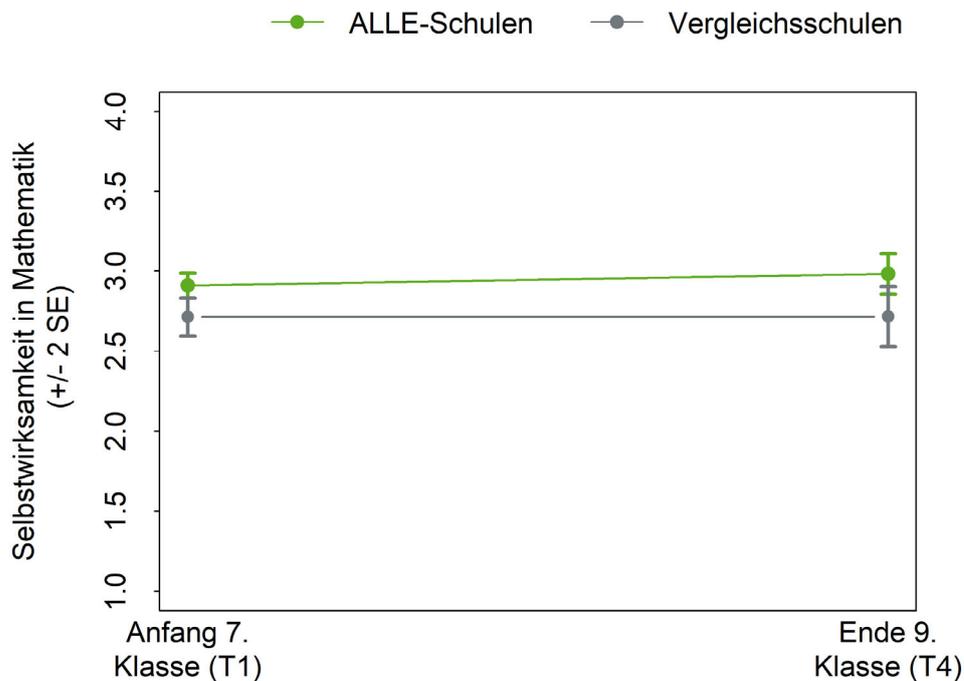
Anmerkungen. ALLE-Schulen $N = 970$ und Vergleichsschulen $N = 717$.

Abbildung 66. Veränderung der zukunftsbezogenen instrumentellen Motivation in den ALLE- und Vergleichsschulen im Fach Mathematik



Anmerkungen. ALLE-Schulen $N = 970$ und Vergleichsschulen $N = 717$.

Abbildung 67. Veränderung der Selbstwirksamkeit in den ALLE- und Vergleichsschulen im Fach Mathematik



Anmerkungen. ALLE-Schulen $N = 970$ und Vergleichsschulen $N = 717$.

Nachfolgend werden die Veränderungen in den motivational-affektiven Lernergebnissen in den ALLE- und Vergleichsschulen mithilfe von Regressionsmodellen verglichen und die Unterschiede auf statistische Signifikanz geprüft (vgl. Kap. 3.6). Tabelle 54 zeigt die Ergebnisse für die motivational-affektiven Lernergebnisse im Fach Deutsch. Die Resultate vor und nach Kontrolle der Hintergrundmerkmale (besuchte Abteilung zum ersten Messzeitpunkt, Geschlecht, Alter, Migrationshintergrund, zuhause gesprochene Sprache, sozioökonomischer Status) der Schülerinnen und Schüler fallen für alle Lernergebnisse jeweils ähnlich aus. Es zeigt sich für keine der analysierten Differenzen ein statistisch bedeutsamer Unterschied zwischen ALLE- und Vergleichsschulen.

Tabelle 54. Unterschiede in der Veränderung zwischen ALLE- und Vergleichsschulen für die gesamte Evaluationsstichprobe in den motivational-affektiven Lernergebnissen im Fach Deutsch

Merkmal	Kontrolle von Hintergrundmerkmalen	Differenz der Veränderung	(SE)	Cohens <i>d</i>
Intrinsische Motivation	vor Kontrolle	0.14	(0.12)	0.13
	nach Kontrolle	0.13	(0.12)	
Anerkennungs- und leistungsbezogene Motivation	vor Kontrolle	0.29	(0.19)	0.33
	nach Kontrolle	0.28	(0.19)	
Zukunftsbezogene instrumentelle Motivation	vor Kontrolle	-0.13	(0.16)	-0.12
	nach Kontrolle	-0.14	(0.16)	
Selbstwirksamkeit	vor Kontrolle	0.22	(0.13)	0.17
	nach Kontrolle	0.22	(0.13)	

Anmerkungen. ALLE-Schulen $N = 970$ und Vergleichsschulen $N = 717$. Grundlage für den Vergleich der Veränderung zwischen ALLE- und Vergleichsschulen ist eine Dummy-Variable: Schülerinnen und Schüler in den ALLE-Schulen = 1, Schülerinnen und Schüler in den Vergleichsschulen = 0. Positive Differenzen bedeuten daher Unterschiede in der Veränderung zugunsten der ALLE-Schulen. Als Hintergrundmerkmale kontrolliert wurden die Abteilung (A vs. B/C), Geschlecht, Alter, Migrationshintergrund, zuhause gesprochene Sprache, sozioökonomischer Status.

(*) $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$

Tabelle 55 zeigt die Ergebnisse für die motivational-affektiven Lernergebnisse im Fach Mathematik. Ähnlich wie im Fach Deutsch fallen die Resultate für die Berechnung mit und ohne Berücksichtigung der Hintergrundmerkmale der Schülerinnen und Schüler ähnlich aus, sodass eine gemeinsame Interpretation der Befunde erfolgen kann. Für die intrinsische Motivation resultiert ein tendenziell signifikanter Unterschied von 0.16 Punkten in der Veränderung (vor Kontrolle) zugunsten der ALLE-Schulen. Die Differenz ist mit $d = 0.15$ jedoch als eher gering einzustufen. Sie ist zudem nach Ausschluss der später einbezogenen beziehungsweise frühzeitig ausgeschiedenen Vergleichsschulen nicht signifikant ($p > .10$), was auf eine eingeschränkte Belastbarkeit des Befunds hindeutet (vgl. Kap. 3.6). Für die aner kennungs- und leistungsbezogene Motivation ergibt sich ein statistisch bedeutsamer Unterschied ($p < .05$) von 0.26 Punkten (vor Kontrolle) zugunsten

der ALLE-Schulen, der auch nach Ausschluss der später einbezogenen oder vorzeitig ausgeschiedenen Vergleichsschulen signifikant bleibt. Damit resultiert eine günstigere Entwicklung der Anerkennung- und leistungsbezogenen Motivation in den ALLE- gegenüber den Vergleichsschulen, wobei der Unterschied hinsichtlich der Effektstärke als moderat einzustufen ist ($d = 0.26$). Aufgrund der insgesamt beobachteten Abnahme in der Anerkennung- und leistungsbezogenen Motivation im Fach Mathematik (vgl. Tabelle 53 bzw. Abbildung 65) bedeutet dies, dass der Rückgang bei dieser Motivationsfacette in den ALLE-Schulen geringer ausfällt als in den Vergleichsschulen. Für die weiteren motivational-affektiven Merkmale lassen sich keine signifikanten Unterschiede zwischen den ALLE- und Vergleichsschulen feststellen.

Tabelle 55. Unterschiede in der Veränderung zwischen ALLE- und Vergleichsschulen für die gesamte Evaluationsstichprobe in den motivational-affektiven Lernergebnissen im Fach Mathematik

Merkmalsname	Kontrolle von Hintergrundmerkmalen	Differenz der Veränderung	(SE)	Cohens d
Intrinsische Motivation	vor Kontrolle	0.16 ^(*)	(0.09)	0.15
	nach Kontrolle	0.16 ^(*)	(0.09)	
Anerkennungs- und leistungsbezogene Motivation	vor Kontrolle	0.26*	(0.12)	0.26
	nach Kontrolle	0.26*	(0.12)	
Zukunftsbezogene instrumentelle Motivation	vor Kontrolle	0.01	(0.10)	0.01
	nach Kontrolle	0.01	(0.10)	
Selbstwirksamkeit	vor Kontrolle	0.07	(0.10)	0.06
	nach Kontrolle	0.07	(0.11)	

Anmerkungen. ALLE-Schulen $N = 970$ und Vergleichsschulen $N = 717$. Grundlage für den Vergleich der Veränderung zwischen ALLE- und Vergleichsschulen ist eine Dummy-Variable: Schülerinnen und Schüler in den ALLE-Schulen = 1, Schülerinnen und Schüler in den Vergleichsschulen = 0. Positive Differenzen bedeuten daher Unterschiede in der Veränderung zugunsten der ALLE-Schulen. Als Hintergrundmerkmale kontrolliert wurden die Abteilung (A vs. B/C), Geschlecht, Alter, Migrationshintergrund, zuhause gesprochene Sprache, sozioökonomischer Status.

(*) $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$

4.10.1.2 Lernergebnisse bei ungünstiger Lernausgangslage

Als Schülerinnen und Schüler mit ungünstiger Lernausgangslage wurde für die nachfolgende Auswertung jener Anteil der Schülerschaft definiert, welcher im Kompetenztest am Anfang der 7. Klassenstufe (erster Messzeitpunkt) zu den 25 Prozent der leistungsschwächsten Schülerinnen und Schülern (d.h. dem unteren Quartil) gehörten. Für die Schülerinnen und Schüler mit ungünstiger Lernausgangslage wurden die in Kapitel 4.10.1.1 berichteten Analysen zu Unterschieden in den motivational-affektiven Lernergebnissen zwischen ALLE- und Vergleichsschulen nochmals separat

durchgeführt. Ziel dieser Analysen war es, festzustellen, inwieweit sich bei den leistungsschwachen Schülerinnen und Schülern als Folge der ALLE-Massnahmen vergleichsweise günstigere Entwicklungen bei den motivational-affektiven Merkmalen zeigen lassen (vgl. Einleitung Kap. 4.10).

Tabelle 56 und Tabelle 57 zeigen die in Regressionsmodellen (vgl. Einleitung Kap. 4.10) ermittelten Unterschiede zwischen ALLE- und Vergleichsschulen für die Fächer Deutsch und Mathematik. Im Fach Deutsch resultiert für kein Merkmal ein statistisch bedeutsamer Unterschied in der Veränderung über den Evaluationszeitraum. Die negative Differenz bei der zukunftsbezogenen instrumentellen Motivation deutet der Richtung nach auf eine ungünstigere Entwicklung in den ALLE-Schulen hin, ist jedoch statistisch nicht signifikant. Für das Fach Mathematik zeichnet sich ein ähnliches Bild ab wie für die gesamte Evaluationsstichprobe (vgl. Kap. 4.10.1.1). Die Unterschiede in der Veränderung der anerkennungs- und leistungsbezogenen Motivation und der intrinsischen Motivation sind statistisch signifikant beziehungsweise tendenziell signifikant mit mittlerer bis kleiner Effektgröße. Sie deuten auf eine günstigere Entwicklung der Merkmale bei leistungsschwachen Schülerinnen und Schülern in den ALLE-Schulen hin. Allerdings sind die Unterschiede nach Ausschluss der später einbezogenen oder frühzeitig ausgeschiedenen Vergleichsschulen nicht signifikant ($p > .10$), was auf eine eingeschränkte Belastbarkeit der Ergebnisse hinweist (vgl. Kap. 3.6).

Tabelle 56. Unterschiede in der Veränderung zwischen ALLE- und Vergleichsschulen bei Schülerinnen und Schülern mit ungünstiger Lernausgangslage in den motivational-affektiven Lernergebnissen im Fach Deutsch

Merkmal	Kontrolle von Hintergrundmerkmalen	Differenz der Veränderung	(SE)	Cohens d
Intrinsische Motivation	vor Kontrolle	0.11	(0.18)	0.10
	nach Kontrolle	0.11	(0.17)	
Anerkennungs- und leistungsbezogene Motivation	vor Kontrolle	0.27	(0.25)	0.29
	nach Kontrolle	0.26	(0.24)	
Zukunftsbezogene instrumentelle Motivation	vor Kontrolle	-0.04	(0.19)	-0.04
	nach Kontrolle	-0.06	(0.19)	
Selbstwirksamkeit	vor Kontrolle	0.09	(0.24)	0.07
	nach Kontrolle	0.12	(0.24)	

Anmerkungen. ALLE-Schulen $N = 242$ und Vergleichsschulen $N = 177$. Grundlage für den Vergleich der Veränderung zwischen ALLE- und Vergleichsschulen ist eine Dummy-Variable: Schülerinnen und Schüler in den ALLE-Schulen = 1, Schülerinnen und Schüler in den Vergleichsschulen = 0. Positive Differenzen bedeuten daher Unterschiede in der Veränderung zugunsten der ALLE-Schulen. Als Hintergrundmerkmale kontrolliert wurden die Abteilung (A vs. B/C), Geschlecht, Alter, Migrationshintergrund, zuhause gesprochene Sprache, sozioökonomischer Status.

(*) $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$

Tabelle 57. Unterschiede in der Veränderung zwischen ALLE- und Vergleichsschulen bei Schülerinnen und Schülern mit ungünstiger Lernausgangslage in den motivational-affektiven Lernergebnissen im Fach Mathematik

Merkmal	Kontrolle von Hintergrundmerkmalen	Differenz der Veränderung	(SE)	Cohens <i>d</i>
Intrinsische Motivation	vor Kontrolle	0.26 ^(*)	(0.14)	0.23
	nach Kontrolle	0.26 ^(*)	(0.14)	
Anerkennungs- und leistungsbezogene Motivation	vor Kontrolle	0.39 [*]	(0.18)	0.37
	nach Kontrolle	0.39 [*]	(0.17)	
Zukunftsbezogene instrumentelle Motivation	vor Kontrolle	0.15	(0.15)	0.13
	nach Kontrolle	0.14	(0.15)	
Selbstwirksamkeit	vor Kontrolle	0.18	(0.23)	0.16
	nach Kontrolle	0.18	(0.23)	

Anmerkungen. ALLE-Schulen $N = 235$ und Vergleichsschulen $N = 183$. Grundlage für den Vergleich der Veränderung zwischen ALLE- und Vergleichsschulen ist eine Dummy-Variable: Schülerinnen und Schüler in den ALLE-Schulen = 1, Schülerinnen und Schüler in den Vergleichsschulen = 0. Positive Differenzen bedeuten daher Unterschiede in der Veränderung zugunsten der ALLE-Schulen. Als Hintergrundmerkmale kontrolliert wurden die Abteilung (A vs. B/C), Geschlecht, Alter, Migrationshintergrund, zuhause gesprochene Sprache, sozioökonomischer Status.

(*) $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$

4.10.2 Fachliche Lernergebnisse

Im Folgenden werden die Resultate zu den fachlichen Lernergebnissen der Schülerinnen und Schüler berichtet. Im ersten Abschnitt werden die Lernergebnisse der gesamten Evaluationsstichprobe aus den ALLE- und Vergleichsschulen einander gegenübergestellt (Kap. 4.10.2.1). Im zweiten Abschnitt erfolgt der Vergleich zwischen den ALLE- und Vergleichsschulen mit Blick auf diejenigen Schülerinnen und Schüler, welche zu Beginn der Evaluation ALLE (erster Messzeitpunkt) über die geringsten fachlichen Kompetenzen, das heisst eine ungünstige Lernausgangslage, verfügten (Kap. 4.10.2.2). Im dritten Abschnitt werden Unterschiede zwischen den ALLE- und Vergleichsschulen in den Lernergebnissen der Schülerinnen und Schüler fokussiert, welche den mittleren und unteren Leistungsniveaus im jeweiligen Fach (Deutsch: Abteilungen B und C; Mathematik: Abteilungen B und C / Anforderungsstufen II und III; vgl. Kap. 3.2.2) zugewiesen waren (Kap. 4.10.2.3). Abschliessend wird auf die Ergebnisse der ALLE- und Vergleichsschulen in der Lernstandserhebung zum «Texte schreiben» eingegangen (Kap. 4.10.2.4).

In jedem der Abschnitte werden zunächst die Lernergebnisse in den ALLE- und Vergleichsschulen zu den verschiedenen Messzeitpunkten im Verlauf dargestellt (vgl. Einleitung Kap. 4.10). Ein besonderes Augenmerk liegt darauf, wie die zwischen den Messzeitpunkten erzielten Veränderungen – im Folgenden als *Lernzuwächse* bezeichnet – im Vergleich mit Ergebnissen aus anderen Studien

einzuordnen sind. Im Vordergrund steht hier, ähnlich wie bei den Analysen zu den motivational-affektiven Merkmalen (vgl. Kap. 4.10.1.1), die deskriptive Beschreibung der Veränderungen. Um zu bestimmen, inwieweit das ALLE-Massnahmenbündel mit Wirkungen auf die Lernergebnisse der Schülerinnen und Schüler in Deutsch und Mathematik einherging, werden danach Unterschiede zwischen den ALLE- und Vergleichsschulen in den Lernzuwächsen betrachtet und statistisch geprüft (vgl. Einleitung Kap. 4.10). Die Analysen werden für die gesamte Evaluationsstichprobe (Kap. 4.10.2.1) für die Fächer (Deutsch bzw. Mathematik) und die einzelnen Kompetenzbereiche (z.B. Deutsch: Lesen, Hören, Sprache[n] im Fokus) durchgeführt und berichtet. In den nachfolgenden Abschnitten (ab Kap. 4.10.2.2) beschränken sich die Auswertungen auf die Ebene der Fächer. Informationen zur Bedeutung der Punktwerte aus den Kompetenztests können Kapitel 3.6 entnommen werden.

4.10.2.1 Lernergebnisse in der gesamten Evaluationsstichprobe

Nachfolgend werden zunächst die Lernergebnisse der Schülerinnen und Schüler an den verschiedenen Messzeitpunkten getrennt für die ALLE- und Vergleichsschulen dargestellt. Das Augenmerk liegt dabei auf den Lernzuwächsen der Schülerinnen und Schüler, welche zwischen den Klassenstufen 7 bis 9 erzielt wurden (vgl. Einleitung Kap. 4.10). Tabelle 58 und Abbildung 68 bis Abbildung 71 berichten die Resultate für das Fach Deutsch auf Fachebene beziehungsweise in den einzelnen Kompetenzbereichen. Tabelle 58 enthält neben den Mittelwerten (Spalte *M*) auch die Standardabweichungen (Spalte *SD*) in der jeweiligen Stichprobe, das heisst das Ausmass der Streuung in den Merkmalsausprägungen der Schülerinnen und Schüler. Die Abbildungen veranschaulichen zusätzlich die Veränderungen in den mittleren Lernergebnissen über die Messzeitpunkte.

Betrachtet man den gesamten Evaluationszeitraum (Anfang der 7. Klassenstufe bis Ende der 9. Klassenstufe), wird auf Fachebene in den ALLE-Schulen ein mittlerer Lernzuwachs von 111 Punkten erzielt (612 Punkte – 501 Punkte), in den Vergleichsschulen ein Lernzuwachs von 98 Punkten (598 Punkte – 500 Punkte). Mit Blick auf die einzelnen Kompetenzbereiche fällt der in beiden untersuchten Gruppen relativ höhere Lernzuwachs im Bereich «Sprache(n) im Fokus» auf, die in beiden Gruppen geringste Punktwertdifferenz wird im Bereich «Hören» erzielt.

In den ALLE-Schulen wurden am Ende der 8. Klassenstufe (d.h. am dritten Messzeitpunkt; T3) ebenfalls Tests in Deutsch und Mathematik durchgeführt, nicht aber in den Vergleichsschulen. Zwischen der 8. und 9. Klassenstufe zeigen sich in den ALLE-Schulen geringe Lernzuwächse, zum Teil werden am Ende der 9. Klassenstufe sogar etwas niedrigere Testleistungen erzielt. Dies entspricht Befunden, die für das Ende der Sekundarstufe I und weiterführend die Sekundarstufe II nur noch minimale beziehungsweise nicht signifikante Lernzuwächse berichten (Angelone et al., 2013; Gailberger & Willenberg, 2008; vgl. die nachfolgende Erläuterung zu erwarteten Lernzuwächsen).

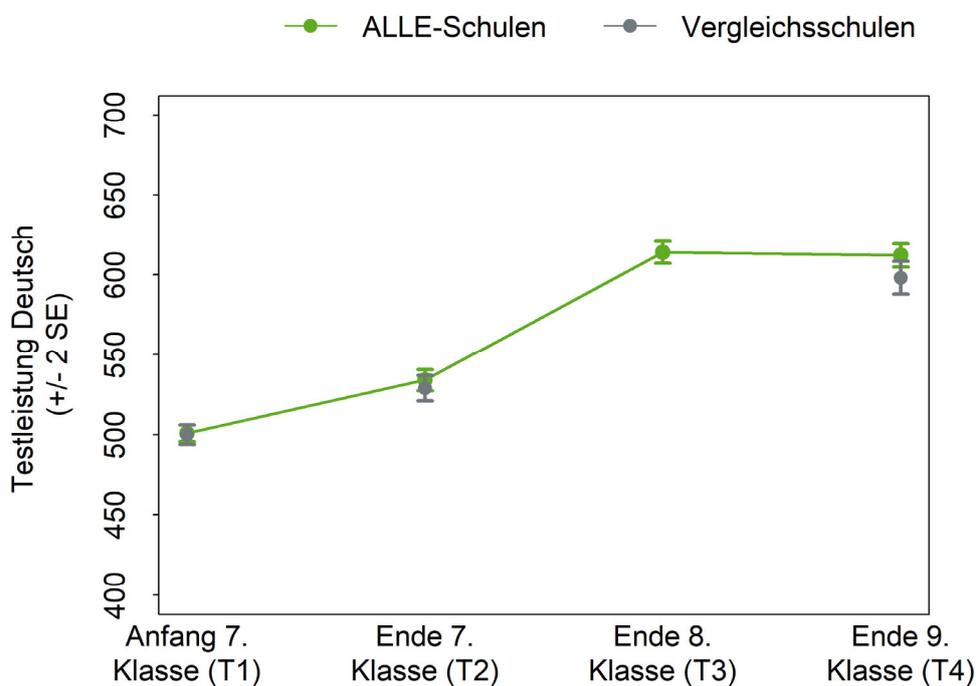
Die Streuung der Testleistungen (Standardabweichung) nimmt in beiden Gruppen für das gesamte Fach und in allen Kompetenzbereichen über den Evaluationszeitraum zu, das heisst, die Unterschiede in den Kompetenzen zwischen den Schülerinnen und Schülern verstärken sich zwischen dem Anfang der 7. und dem Ende der 9. Klassenstufe. Dabei fällt auf, dass das Ausmass der Leistungsunterschiede zwischen den Schülerinnen und Schülern in den ALLE-Schulen weniger stark zunimmt als in den Vergleichsschulen.

Tabelle 58. Lernergebnisse in den ALLE- und Vergleichsschulen im Fach Deutsch für die Messzeitpunkte der ALLE-Evaluation

Fach/Kompetenzbereich	Messzeitpunkt	ALLE-Schulen			Vergleichsschulen		
		<i>M</i>	(<i>SE</i>)	<i>SD</i>	<i>M</i>	(<i>SE</i>)	<i>SD</i>
Deutsch	T1	501	(2.6)	74	500	(3.0)	76
	T2	534	(3.3)	93	529	(4.0)	102
	T3	614	(3.4)	80			
	T4	612	(3.6)	94	598	(5.1)	104
Lesen	T1	501	(3.5)	99	500	(4.1)	100
	T2	533	(4.9)	133	523	(5.9)	141
	T3	612	(5.3)	123			
	T4	609	(5.8)	126	591	(7.3)	140
Hören	T1	502	(3.9)	98	500	(4.8)	100
	T2	524	(5.0)	127	522	(6.0)	134
	T3	613	(5.0)	113			
	T4	596	(5.8)	125	578	(7.6)	144
Sprache(n) im Fokus	T1	499	(3.8)	92	500	(4.0)	100
	T2	543	(3.6)	103	541	(4.4)	108
	T3	615	(3.7)	99			
	T4	631	(4.1)	112	626	(5.5)	126

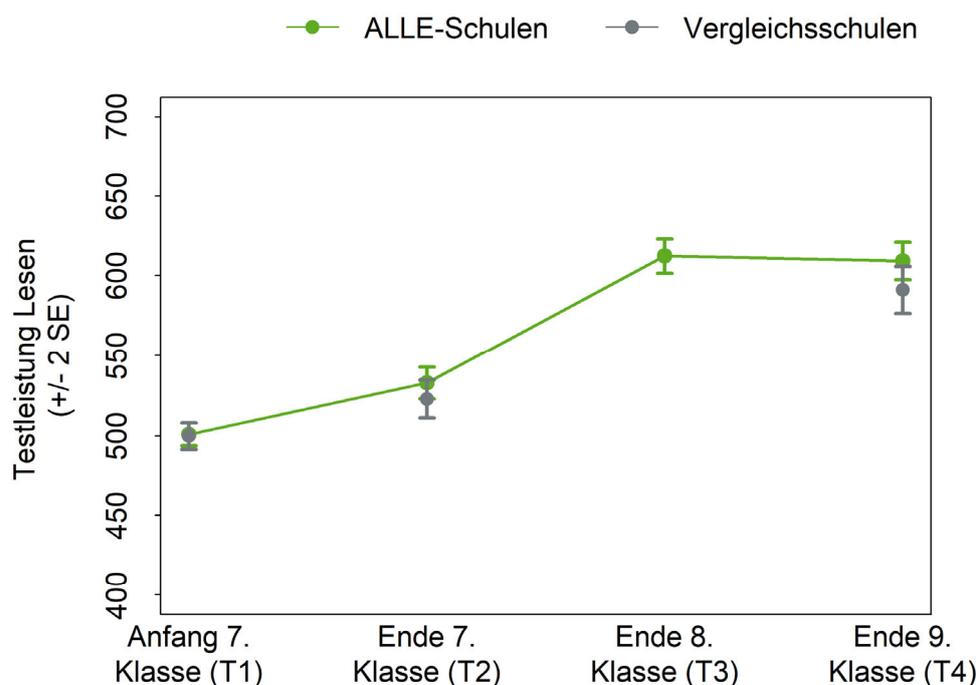
Anmerkungen. T1 = Anfang 7. Klasse, T2 = Ende 7. Klasse, T3 = Ende 8. Klasse, T4 = Ende 9. Klasse. ALLE-Schulen *N* = 970 und Vergleichsschulen *N* = 717. Die Erhebung am dritten Messzeitpunkt (T3) wurde nur in den ALLE-Schulen durchgeführt (vgl. Kap. 3.1).

Abbildung 68. Entwicklung der Testleistungen in den ALLE- und Vergleichsschulen im Fach Deutsch für die Messzeitpunkte der ALLE-Evaluation



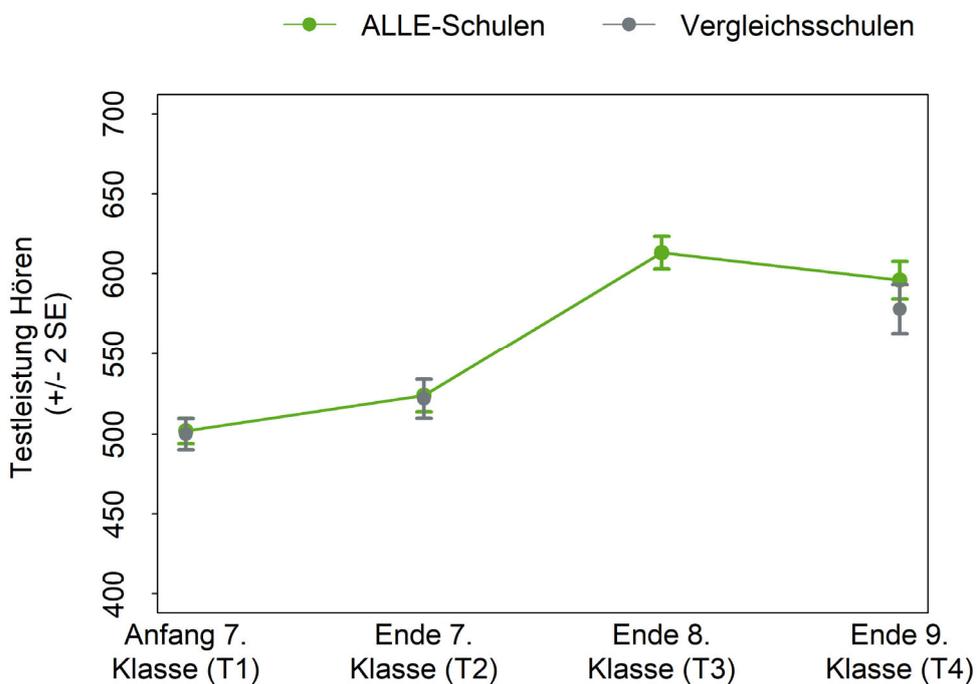
Anmerkungen. ALLE-Schulen $N = 970$ und Vergleichsschulen $N = 717$. Die Erhebung am dritten Messzeitpunkt (T3) wurde nur in den ALLE-Schulen durchgeführt.

Abbildung 69. Entwicklung der Testleistungen in den ALLE- und Vergleichsschulen im Kompetenzbereich Lesen für die Messzeitpunkte der ALLE-Evaluation



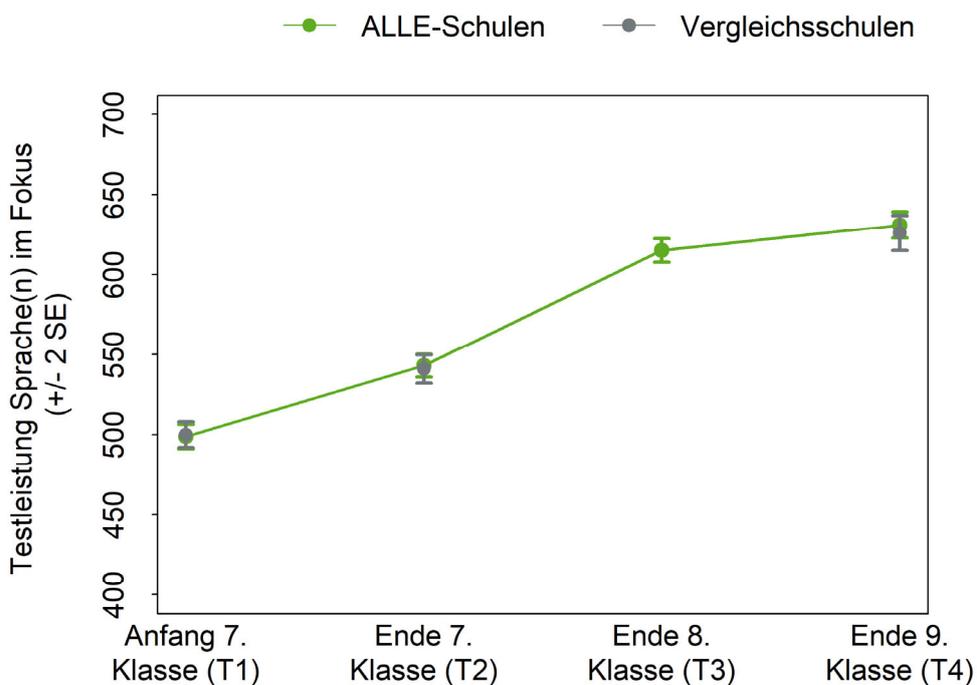
Anmerkungen. ALLE-Schulen $N = 970$ und Vergleichsschulen $N = 717$. Die Erhebung am dritten Messzeitpunkt (T3) wurde nur in den ALLE-Schulen durchgeführt.

Abbildung 70. Entwicklung der Testleistungen in den ALLE- und Vergleichsschulen im Kompetenzbereich Hören für die Messzeitpunkte der ALLE-Evaluation



Anmerkungen. ALLE-Schulen $N = 970$ und Vergleichsschulen $N = 717$. Die Erhebung am dritten Messzeitpunkt (T3) wurde nur in den ALLE-Schulen durchgeführt.

Abbildung 71. Entwicklung der Testleistungen in den ALLE- und Vergleichsschulen im Kompetenzbereich Sprache(n) im Fokus für die Messzeitpunkte der ALLE-Evaluation



Anmerkungen. ALLE-Schulen $N = 970$ und Vergleichsschulen $N = 717$. Die Erhebung am dritten Messzeitpunkt (T3) wurde nur in den ALLE-Schulen durchgeführt.

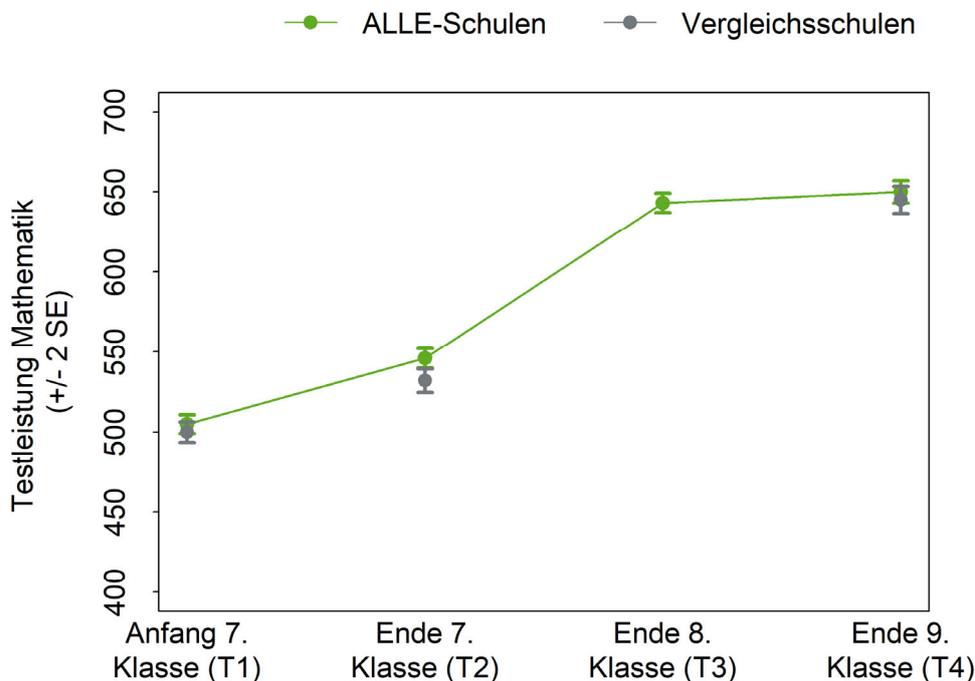
Tabelle 59 und Abbildung 72 bis Abbildung 75 zeigen die Resultate in Mathematik für das gesamte Fach sowie die einzelnen Kompetenzbereiche. Über den gesamten Evaluationszeitraum wird auf der Fachebene sowohl in den ALLE-Schulen als auch in den Vergleichsschulen ein mittlerer Lernzuwachs von 145 Punkten erzielt (650 Punkte – 505 Punkte bzw. 645 Punkte – 500 Punkte). Bei den Kompetenzbereichen entfällt der höchste Lernzuwachs auf den Bereich «Form und Raum» (ALLE-Schulen) beziehungsweise «Zahl und Variable» (Vergleichsschulen), der jeweils geringste Zuwachs findet sich im Bereich «Grössen, Funktionen, Daten und Zufall». Anders als im Fach Deutsch sind die Veränderungen auch zwischen dem Ende der 8. und 9. Klassenstufe durchweg positiv, im Umfang allerdings eher gering – wiederum im Einklang mit früheren Befunden am Ende der Sekundarstufe I (vgl. z.B. Angelone et al., 2013). Vergleichbar mit dem Fach Deutsch nimmt die Streuung in den Testleistungen in beiden Gruppen auf der Fachebene und in allen Kompetenzbereichen über den Evaluationszeitraum zu. Dagegen ergibt sich kein nennenswerter Unterschied in der Zunahme der Leistungsunterschiede zwischen den ALLE- und Vergleichsschulen.

Tabelle 59. Lernergebnisse in den ALLE- und Vergleichsschulen im Fach Mathematik für die Messzeitpunkte der ALLE-Evaluation

Fach/Kompetenzbereich	Messzeitpunkt	ALLE-Schulen			Vergleichsschulen		
		<i>M</i>	(<i>SE</i>)	<i>SD</i>	<i>M</i>	(<i>SE</i>)	<i>SD</i>
Mathematik	T1	505	(2.9)	77	500	(3.2)	76
	T2	546	(3.0)	86	532	(3.6)	87
	T3	643	(3.0)	85			
	T4	650	(3.5)	105	645	(4.2)	103
Zahl und Variable	T1	508	(3.9)	98	500	(4.5)	100
	T2	535	(4.2)	115	527	(4.6)	115
	T3	640	(4.2)	109			
	T4	649	(4.4)	124	653	(5.7)	122
Form und Raum	T1	500	(3.7)	99	500	(4.8)	100
	T2	564	(3.9)	107	545	(5.1)	109
	T3	655	(4.9)	131			
	T4	665	(4.9)	147	650	(6.7)	148
Grössen, Funktionen, Daten und Zufall	T1	506	(4.4)	99	500	(4.9)	100
	T2	538	(4.0)	110	525	(5.0)	109
	T3	634	(3.4)	95			
	T4	635	(4.8)	115	631	(5.5)	126

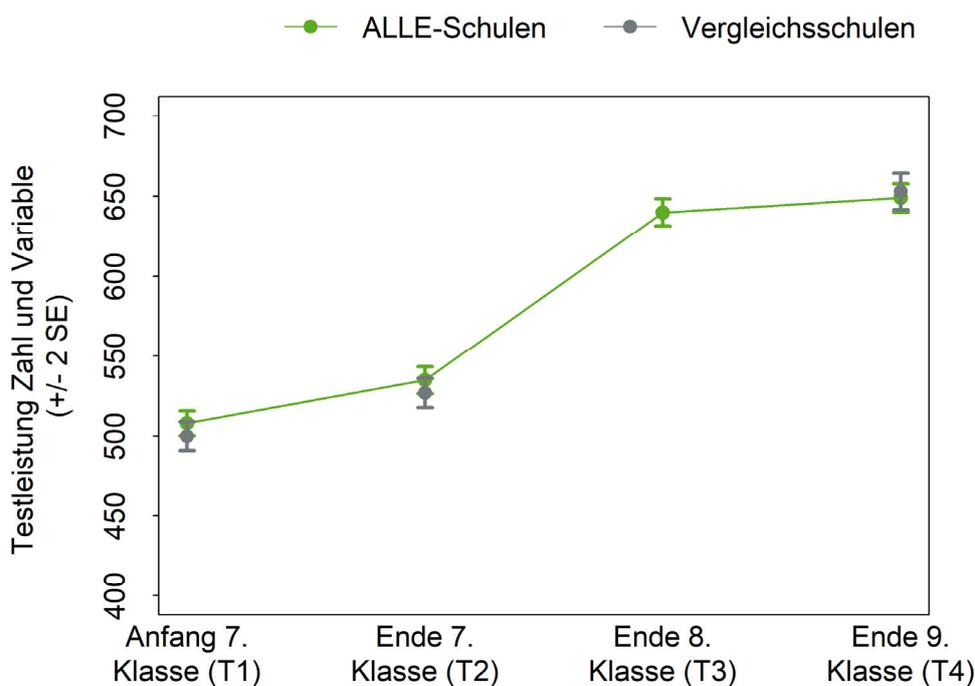
Anmerkungen. T1 = Anfang 7. Klasse, T2 = Ende 7. Klasse, T3 = Ende 8. Klasse, T4 = Ende 9. Klasse. ALLE-Schulen $N = 970$ und Vergleichsschulen $N = 717$. Die Erhebung am dritten Messzeitpunkt (T3) wurde nur in den ALLE-Schulen durchgeführt (vgl. Kap. 3.1).

Abbildung 72. Entwicklung der Testleistungen in den ALLE- und Vergleichsschulen im Fach Mathematik für die Messzeitpunkte der ALLE-Evaluation



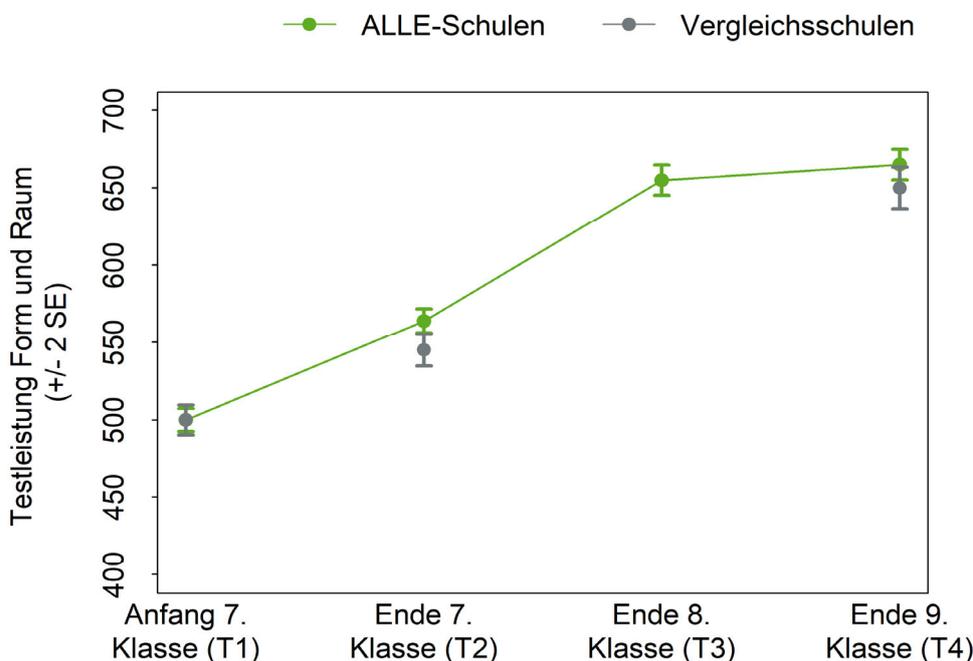
Anmerkung. ALLE-Schulen $N = 970$ und Vergleichsschulen $N = 717$. Die Erhebung am dritten Messzeitpunkt (T3) wurde nur in den ALLE-Schulen durchgeführt.

Abbildung 73. Entwicklung der Testleistungen in den ALLE- und Vergleichsschulen im Kompetenzbereich Zahl und Variable für die Messzeitpunkte der ALLE-Evaluation



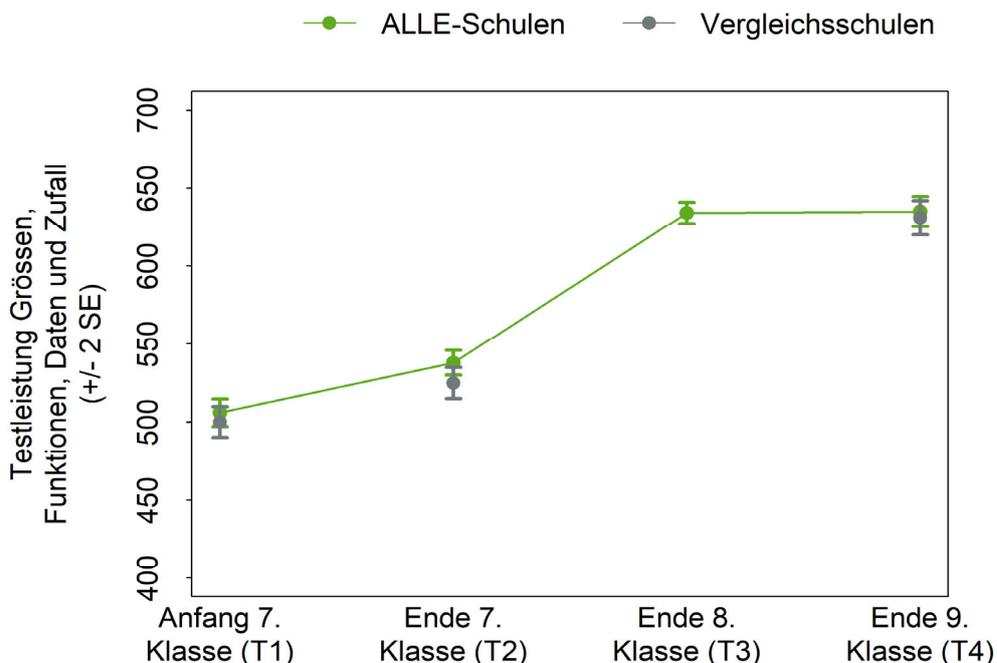
Anmerkung. ALLE-Schulen $N = 970$ und Vergleichsschulen $N = 717$. Die Erhebung am dritten Messzeitpunkt (T3) wurde nur in den ALLE-Schulen durchgeführt.

Abbildung 74. Entwicklung der Testleistungen in den ALLE- und Vergleichsschulen im Kompetenzbereich Form und Raum für die Messzeitpunkte der ALLE-Evaluation



Anmerkung. ALLE-Schulen $N = 970$ und Vergleichsschulen $N = 717$. Die Erhebung am dritten Messzeitpunkt (T3) wurde nur in den ALLE-Schulen durchgeführt.

Abbildung 75. Entwicklung der Testleistungen in den ALLE- und Vergleichsschulen im Kompetenzbereich Grössen, Funktionen, Daten und Zufall für die Messzeitpunkte der ALLE-Evaluation



Anmerkung. ALLE-Schulen $N = 970$ und Vergleichsschulen $N = 717$. Die Erhebung am dritten Messzeitpunkt (T3) wurde nur in den ALLE-Schulen durchgeführt.

Eine Interpretation bezüglich der Grösse der erzielten Lernzuwächse kann durch den Vergleich mit bisherigen Studien erfolgen, welche die Lernfortschritte von Schülerinnen und Schülern im Verlauf von einem oder mehreren Schuljahren quantifiziert haben. Die in Studien aus dem deutschsprachigen Raum berichteten Lernfortschritte in der Sekundarstufe I bewegen sich für Deutsch und Mathematik in der Regel zwischen $d = 0.20$ und $d = 0.40$ je Schuljahr (z.B. Baumert et al., 1997; DESI-Konsortium, 2008; Ehmke, Blum, Neubrand, Jordan & Ulfig, 2006; Lehmann, Peek, Gänsfuß & Husfeldt, 2002). Die erzielten Lernzuwächse nehmen dabei von Schuljahr zu Schuljahr tendenziell ab. Differenzen einer ähnlichen Grössenordnung berichten Hill, Bloom, Black und Lipsey (2008) aus einer umfangreichen Untersuchung von US-amerikanischen Datenquellen (Lesen: $d = 0.26$ zwischen den Klassenstufen 7 und 8, $d = 0.24$ zwischen den Klassenstufen 8 und 9; Mathematik: $d = 0.32$ bzw. $d = 0.22$ in den gleichen Zeiträumen). Für den vorliegenden Bericht besonders relevant sind die Resultate der Zürcher Lernstandserhebung, die auf einer repräsentativen Stichprobe im Kanton Zürich basieren. Angelone, Keller und Moser (2013) berichten jährliche Lernzuwächse in der Sekundarstufe I von $d = 0.31$ für Deutsch und $d = 0.14$ für Mathematik. Während die Lernfortschritte für Deutsch im erwarteten Bereich liegen, fallen die vergleichsweise geringen Zuwächse im Fach Mathematik auf. Als mögliche Erklärung diskutieren Angelone et al. (2013) Besonderheiten des Mathematikunterrichts auf der Sekundarstufe I.

Legt man diese Befundlage zugrunde, deuten die Ergebnisse darauf hin, dass sowohl in den ALLE- als auch den Vergleichsschulen durchweg Lernzuwächse erzielt wurden, die zumindest dem erwarteten Umfang entsprechen. Im Mittel über die drei untersuchten Schuljahre liegen diese in Deutsch auf der Fachebene für die ALLE-Schulen bei $d = 0.44$ je Schuljahr, für die Vergleichsschulen bei $d = 0.36$ je Schuljahr.¹¹ In Mathematik betragen diese $d = 0.53$ für die ALLE- sowie die Vergleichsschulen je Schuljahr. Damit deutet sich für das Fach Deutsch ein höherer Lernzuwachs in den ALLE-Schulen an; die statistische Bedeutsamkeit dieses Unterschieds wird in den nachfolgend berichteten Regressionsmodellen geprüft. Zudem liegt der Lernzuwachs in Mathematik eher oberhalb der Erwartungen, wenn man die Resultate mit bisherigen Studien – insbesondere auch den Resultaten der Zürcher Lernstandserhebung – vergleicht. Dabei ist einschränkend zu beachten, dass sich die Differenzen in Standardabweichungseinheiten (d) auf die Streuung der jeweiligen Stichprobe(n) beziehen. Diese kann in repräsentativen Stichproben wie jenen der Zürcher Lernstandserhebung grösser ausfallen als in möglicherweise selektiven Stichproben wie jener der vorliegenden Evaluation. Relativiert an der Streuung in der Gesamtheit der Zürcher Schulen könnten

¹¹ Die Berechnung erfolgte anhand der Mittelwertsdifferenz der Testleistungen zwischen T1 und T4 und der zusammengefassten («gepoolten») Standardabweichung, geteilt durch 3 für die drei Schuljahre. Eine alternative Kalkulation auf Grundlage der zwischen den einzelnen Testzeitpunkten erzielten und dann gemittelten Lernzuwächse führte zu sehr ähnlichen Ergebnissen.

die Lernzuwächse entsprechend moderater sein. Die Zuwächse je Schuljahr in den Kompetenzbereichen fallen in Deutsch – analog zu den in Tabelle 58 und Tabelle 59 berichteten Ergebnissen – für «Sprache(n) im Fokus» höher aus als in den anderen Bereichen (Bsp. ALLE-Schulen: $d = 0.32$ in «Lesen», $d = 0.28$ in «Hören», $d = 0.43$ in «Sprache[n] im Fokus»), in Mathematik sind die Zuwächse in den Bereichen vergleichsweise ähnlich (Bsp. ALLE-Schulen: $d = 0.42$ in «Zahl und Variable», $d = 0.44$ in «Form und Raum», $d = 0.40$ in «Grössen, Funktionen, Daten und Zufall»).¹² Für alle Resultate gilt, dass diese nicht auf den Kanton Zürich verallgemeinert werden dürfen, sondern nur für die in der Evaluation des Pilotprojekts ALLE untersuchte Stichprobe gültig sind.

Nachfolgend werden die Lernzuwächse in den ALLE- und Vergleichsschulen auf der Grundlage von Regressionsmodellen verglichen und die Unterschiede auf statistische Bedeutsamkeit geprüft (vgl. Einleitung Kap. 4.10). Tabelle 60 zeigt die Ergebnisse für das Fach Deutsch als Ganzes sowie für die drei im Fach Deutsch getesteten Kompetenzbereiche.

Tabelle 60. Unterschiede im Lernzuwachs zwischen den ALLE- und Vergleichsschulen für die gesamte Evaluationsstichprobe im Fach Deutsch

Fach/Kompetenzbereich	Kontrolle von Hintergrundmerkmalen	Differenz im Lernzuwachs	(SE)	Cohens d
Deutsch	vor Kontrolle	12.6 ^(*)	(6.8)	0.13
	nach Kontrolle	12.5 ^(*)	(6.9)	
Lesen	vor Kontrolle	16.8	(10.6)	0.11
	nach Kontrolle	16.9	(10.8)	
Hören	vor Kontrolle	15.4	(11.1)	0.10
	nach Kontrolle	15.4	(11.2)	
Sprache(n) im Fokus	vor Kontrolle	5.5	(7.6)	0.04
	nach Kontrolle	5.2	(7.6)	

Anmerkungen. ALLE-Schulen $N = 970$ und Vergleichsschulen $N = 717$. Grundlage für den Vergleich der Veränderung zwischen ALLE- und Vergleichsschulen ist eine Dummy-Variable: Schülerinnen und Schüler in den ALLE-Schulen = 1, Schülerinnen und Schüler in den Vergleichsschulen = 0. Positive Differenzen bedeuten daher Unterschiede im Lernzuwachs zugunsten der ALLE-Schulen. Als Hintergrundmerkmale kontrolliert wurden die Abteilung (A vs. B/C), Geschlecht, Alter, Migrationshintergrund, zuhause gesprochene Sprache, sozioökonomischer Status. ^(*) $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$

¹² Der insgesamt höhere Lernzuwachs auf Fachebene als in den einzelnen Kompetenzbereichen ergibt sich durch die Bildung der Testwerte auf Fachebene als Mittelwerte der Kompetenzbereiche und der resultierenden geringeren Standardabweichung auf Fachebene (vgl. Tabelle 58 und Tabelle 59). Dies ist rechnerisch korrekt, verweist aber darauf, dass der Vergleich mit d -Werten aus anderen Studien vor allem auf Konstrukte mit demselben inhaltlichen Auflösungsgrad bezogen werden sollte.

Die Resultate vor und nach Kontrolle der Hintergrundmerkmale der Schülerinnen und Schüler fallen jeweils sehr ähnlich aus, sodass für die Interpretation nicht zwischen den beiden Bedingungen differenziert werden muss. Für das Fach Deutsch resultiert (vor Kontrolle) eine Differenz im Lernzuwachs von 12.6 Punkten zugunsten der ALLE-Schulen, welche tendenziell signifikant ist. Damit deutet sich an, dass die ALLE-Schulen im Evaluationszeitraum einen höheren Lernzuwachs erzielt haben als die Vergleichsschulen. Der Unterschied im Lernzuwachs ist hinsichtlich der Effektgrösse ($d = 0.13$) im unteren Bereich einzuordnen. Für die einzelnen Kompetenzbereiche fällt die Disparität zwischen den ALLE- und Vergleichsschulen augenscheinlich verschieden aus, mit deutlich grösseren Unterschieden den Bereichen «Lesen» und «Hören» als für den Bereich «Sprache(n) im Fokus». Allerdings lässt sich der Unterschied im Lernzuwachs zwischen den ALLE- und Vergleichsschulen für keinen der Kompetenzbereiche statistisch absichern.

Tabelle 61 zeigt die Ergebnisse für das Fach Mathematik als Ganzes sowie für die drei getesteten Kompetenzbereiche. Auch hier fallen die Resultate für die Berechnung mit und ohne Berücksichtigung der Hintergrundmerkmale der Schülerinnen und Schüler sehr ähnlich aus, sodass eine gemeinsame Interpretation der Befunde erfolgen kann.

Tabelle 61. Unterschiede im Lernzuwachs zwischen den ALLE- und Vergleichsschulen für die gesamte Evaluationsstichprobe im Fach Mathematik

Fach/Kompetenzbereich	Kontrolle von Hintergrundmerkmalen	Differenz im Lernzuwachs	(SE)	Cohens d
Mathematik	vor Kontrolle	0.3	(5.5)	<0.01
	nach Kontrolle	<0.1	(5.5)	
Zahl und Variable	vor Kontrolle	-11.8	(9.3)	-0.09
	nach Kontrolle	-12.1	(9.3)	
Form und Raum	vor Kontrolle	14.6 ^(*)	(8.3)	0.09
	nach Kontrolle	14.0 ^(*)	(8.2)	
Grössen, Funktionen, Daten und Zufall	vor Kontrolle	-2.1	(9.8)	-0.02
	nach Kontrolle	-1.6	(9.8)	

Anmerkungen. ALLE-Schulen $N = 970$ und Vergleichsschulen $N = 717$. Grundlage für den Vergleich der Veränderung zwischen ALLE- und Vergleichsschulen ist eine Dummy-Variable: Schülerinnen und Schüler in den ALLE-Schulen = 1, Schülerinnen und Schüler in den Vergleichsschulen = 0. Positive Differenzen bedeuten daher Unterschiede im Lernzuwachs zugunsten der ALLE-Schulen. Als Hintergrundmerkmale kontrolliert wurden die Abteilung (A vs. B/C) bzw. Anforderungsstufe (I vs. II/III), Geschlecht, Alter, Migrationshintergrund, zuhause gesprochene Sprache, sozioökonomischer Status.

(*) $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$

Auf Ebene des gesamten Fachs resultieren keine signifikanten Unterschiede zwischen ALLE- und Vergleichsschulen. Die geschätzte Punktwertdifferenz im Lernzuwachs ist mit 0.3 Punkten (vor Kontrolle) äusserst gering. In den einzelnen Kompetenzbereichen resultiert für den Bereich «Form

und Raum» ein tendenziell signifikanter Unterschied von 14.6 Punkten im Lernzuwachs (vor Kontrolle) zugunsten der ALLE-Schulen. Die Differenz ist mit $d = 0.09$ jedoch als geringfügig einzustufen. Die Unterschiede in den weiteren Kompetenzbereichen weisen ein umgekehrtes Vorzeichen auf (zugunsten der Vergleichsschulen), erreichen jedoch keine statistische Bedeutsamkeit.

Insgesamt deuten die berichteten Resultate darauf hin, dass das ALLE-Massnahmenbündel zumindest punktuell mit günstigen Wirkungen auf die fachlichen Lernergebnisse verbunden ist. Hinweise darauf ergeben sich für das Fach Deutsch (Gesamttest) und den Kompetenzbereich «Form und Raum» im Fach Mathematik, wobei für das Fach Mathematik als Ganzes keine Unterschiede festzustellen sind. Im Umfang sind die beobachteten Unterschiede im Lernzuwachs zwischen den ALLE- und Vergleichsschulen als eher gering einzustufen.

4.10.2.2 Lernergebnisse bei ungünstiger Lernausgangslage

Als Schülerinnen und Schüler mit ungünstiger Lernausgangslage wurde für die nachfolgende Auswertung jener Anteil der Schülerschaft definiert, welcher im Kompetenztest am Anfang der 7. Klassenstufe (erster Messzeitpunkt) zu den 25 Prozent der leistungsschwächsten Schülerinnen und Schülern (d.h. dem unteren Quartil) gehörten. Für diese Schülergruppen wurden die in Kapitel 4.10.1.1 berichteten Analysen zum Unterschied im Lernzuwachs zwischen ALLE- und Vergleichsschulen nochmals separat durchgeführt. Ziel war es festzustellen, inwieweit die leistungsschwachen Schülerinnen und Schüler in den ALLE-Schulen durch die ALLE-Massnahmen einen höheren Lernfortschritt erzielen konnten. Eine zufriedenstellend stabile Schätzung der Punktwertdifferenzen im Lernzuwachs erwies sich wegen der deutlich verringerten Stichprobengrösse jedoch als schwierig. Es zeigte sich, dass die Resultate unter anderem deutlich von der Berücksichtigung der später einbezogenen beziehungsweise frühzeitig ausgeschiedenen Vergleichsschulen abhängig waren (vgl. Kap. 3.6). Dieses Problem trat für die Kompetenzbereiche noch verstärkt auf. Anders als für die Gesamtstichprobe beschränkt sich die Darstellung daher auf die Ebene der Fächer (Deutsch, Mathematik), auch die Resultate auf Fachebene sollten aber mit Vorsicht interpretiert werden.

Tabelle 62 stellt die Unterschiede im Lernzuwachs zwischen ALLE- und Vergleichsschulen bei den Schülerinnen und Schülern mit ungünstiger Lernausgangslage für die Fächer Deutsch und Mathematik dar, die auf der Grundlage von Regressionsmodellen ermittelt und auf statistische Bedeutsamkeit geprüft wurden (vgl. Einleitung Kap. 4.10). Für kein Fach lässt sich die Punktwertdifferenz im Lernzuwachs statistisch absichern. Die negative Differenz in Mathematik deutet der Richtung nach auf eine ungünstigere Leistungsentwicklung in den ALLE-Schulen hin. Jedoch ist diese Differenz nicht statistisch signifikant und liegt bei Ausschluss der später einbezogenen beziehungsweise frühzeitig ausgeschiedenen Vergleichsschulen knapp im positiven Bereich. Entsprechend sollte das Resultat nicht zu Ungunsten der ALLE-Schulen interpretiert werden. Stattdessen deuten die Ergebnisse daraufhin, dass sich die ALLE- und Vergleichsschulen hinsichtlich ihres Erfolgs bei

der Förderung von Schülerinnen und Schülern mit ungünstiger Lernausgangslage nicht unterscheiden. Inwieweit dies gleichbedeutend mit einer gelungenen Förderung fachlichen Lernens in beiden Gruppen ist, lässt sich aufgrund der fehlenden Vergleichsgrundlage nicht beurteilen.

Tabelle 62. Unterschiede im Lernzuwachs zwischen ALLE- und Vergleichsschulen für Schülerinnen und Schüler mit ungünstiger Lernausgangslage im Fach Deutsch und im Fach Mathematik

Fach	Kontrolle von Hintergrundmerkmalen	Differenz im Lernzuwachs	(SE)	Cohens <i>d</i>
Deutsch	vor Kontrolle	0.5	(11.9)	<0.01
	nach Kontrolle	-3.4	(12.0)	
Mathematik	vor Kontrolle	-13.1	(11.7)	-0.13
	nach Kontrolle	-13.5	(12.2)	

Anmerkungen. Deutsch: ALLE-Schulen $N = 242$ und Vergleichsschulen $N = 177$, Mathematik: ALLE-Schulen $N = 235$ und Vergleichsschulen $N = 183$. Grundlage für den Vergleich der Veränderung zwischen ALLE- und Vergleichsschulen ist eine Dummy-Variable: Schülerinnen und Schüler in den ALLE-Schulen = 1, Schülerinnen und Schüler in den Vergleichsschulen = 0. Positive Differenzen bedeuten daher Unterschiede im Lernzuwachs zugunsten der ALLE-Schulen. Als Hintergrundmerkmale kontrolliert wurden die Abteilung (A vs. B/C), Geschlecht, Alter, Migrationshintergrund, zuhause gesprochene Sprache, sozioökonomischer Status.

(*) $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$

Die dargestellte Analyse beschränkt sich gemäss den Zielen der Evaluation auf das Quartil der Schülerinnen und Schüler mit der ungünstigsten Lernausgangslage. Eine häufig geäusserte Befürchtung ist, dass Fördermassnahmen für leistungsschwächere Schülerinnen und Schüler zulasten der Lernentwicklung bei leistungsstärkeren Schülerinnen und Schülern geht. Hierzu ist festzuhalten, dass sich in weiterführenden Analysen auch für das Quartil der Schülerinnen und Schüler mit der günstigsten Lernausgangslage keine Unterschiede im Lernzuwachs zwischen ALLE- und Vergleichsschulen nachweisen liessen.

4.10.2.3 Lernergebnisse im mittleren und unteren Leistungsniveau

Neben den Schülerinnen und Schülern mit ungünstiger Lernausgangslage wurden auch die Schülerinnen und Schüler, welche zu Beginn der 7. Klassenstufe dem mittleren Leistungsniveau (Abteilung B bzw. Anforderungsstufe II; im Folgenden B/II) oder unteren Leistungsniveau (Abteilung C bzw. Anforderungsstufe III; im Folgenden C/III) zugewiesen wurden, in eine separate Analyse einbezogen. Die Zuweisung zu den Niveaus soll grundsätzlich vor allem auf den Leistungen der Schülerinnen und Schüler beruhen, wie vom Volksschulamt (VSA) des Kantons Zürich festgehalten wird: «Die Schülerinnen und Schüler werden am Ende der 6. Primarklasse einer der Abteilungen und Anforderungsstufen zugeteilt. Die Zuteilung zu einer Abteilung erfolgt aufgrund einer umfassenden

Beurteilung der Leistungen und Fähigkeiten in allen Fächern (Gesamtbeurteilung) im Gespräch mit den Eltern und der Schülerin beziehungsweise dem Schüler. Wird keine Einigung erzielt, entscheidet die für die Sekundarschule zuständige Schulpflege über die Zuteilung. Die Zuteilung zu einer Anforderungsstufe erfolgt aufgrund einer Leistungsbeurteilung im entsprechenden Fachbereich.»¹³

Eine getrennte Analyse der Lernzuwächse vor allem im unteren Leistungsniveau (C/III) erschiene angemessen, da mit dem Pilotprojekt eine verstärkte Förderung gerade auch der leistungsschwächeren Schülerinnen und Schüler beabsichtigt war. Allerdings konnte das Niveau C/III aufgrund der sehr geringen Stichprobengrösse zu Beginn der 7. Klassenstufe (Deutsch: $N = 60$; Mathematik: $N = 206$; vgl. Kap. 3.3.2.1) nicht für sich betrachtet werden, sodass im Folgenden die Zusammenfassung mit dem Niveau B/II erfolgt. Daraus resultiert eine Einschränkung der Aussagekraft in Bezug auf leistungsschwächere Schülerinnen und Schüler, da nicht wenige Schülerinnen und Schüler im Niveau B/II durchschnittliche Leistungen erzielen oder sogar die Leistungen vieler Schülerinnen und Schüler in Niveau A/I übertreffen. Trotzdem sind die Analysen interessant, da sich die Förderung in den ALLE-Schulen von jener in den Vergleichsschulen auch bei der Differenzierung von zwei Niveaus unterscheiden könnte. Eine weitere Einschränkung resultiert daraus, dass viele Schülerinnen und Schüler über den Evaluationszeitraum nicht konstant im gleichen Niveau beschult wurden, sondern gemäss der prinzipiell hohen Durchlässigkeit zwischen den Niveaus ein- oder mehrmals umgeteilt wurden (vgl. Kap. 3.3.2.1). Für die nachfolgende Analyse wurde – im Sinne der Berücksichtigung der Lernausgangslage vor Projektbeginn – das Niveau zugrunde gelegt, in das die Schülerinnen und Schüler zum ersten Messzeitpunkt (Beginn der 7. Klassenstufe) zuteilung worden waren, unabhängig von möglichen Niveauwechseln.

Tabelle 63 stellt die Unterschiede im Lernzuwachs zwischen den ALLE- und Vergleichsschulen für die Fächer Deutsch und Mathematik dar, die basierend auf Regessionsmodellen berechnet und auf statistische Bedeutsamkeit geprüft wurden (vgl. Einleitung Kap. 4.10). Die Punktwertdifferenz im Lernzuwachs ist für keines der Fächer statistisch signifikant. Demnach resultieren keine Hinweise darauf, dass sich die ALLE- und Vergleichsschulen in ihrem Erfolg bei der Förderung von Schülerinnen und Schülern aus den Niveaus B/II und C/III unterscheiden haben – mit der bereits beschriebenen Einschränkung, dass differenzierende Aussagen für die beiden Niveaus nicht möglich sind. Dies bedeutet nicht, dass eine Förderung fachlichen Lernens in diesen Niveaus nicht gelingt, sondern lediglich, dass sich der Lernzuwachs in den zusammengefassten Niveaus B/II und C/III nicht nachweisbar zwischen ALLE- und Vergleichsschulen unterscheidet.

¹³ Information verfügbar auf: <https://www.zh.ch/de/bildung/schulen/volksschule/sekundarschule.html> (abgerufen am 07.09.2020).

Tabelle 63. Unterschiede im Lernzuwachs zwischen ALLE- und Vergleichsschulen für Schülerinnen und Schüler in den Niveaus B/II und C/III (zusammengefasst) in den Fächern Deutsch und Mathematik

Fach	Kontrolle von Hintergrundmerkmalen	Differenz im Lernzuwachs	(SE)	Cohens <i>d</i>
Deutsch	vor Kontrolle	11.8	(10.1)	0.11
	nach Kontrolle	11.2	(10.1)	
Mathematik	vor Kontrolle	-0.8	(8.5)	-0.01
	nach Kontrolle	-1.3	(8.4)	

Anmerkungen. Deutsch: ALLE-Schulen $N = 406$ und Vergleichsschulen $N = 319$, Mathematik: ALLE-Schulen $N = 492$ und Vergleichsschulen $N = 387$. Grundlage für den Vergleich der Veränderung zwischen ALLE- und Vergleichsschulen ist eine Dummy-Variable: Schülerinnen und Schüler in den ALLE-Schulen = 1, Schülerinnen und Schüler in den Vergleichsschulen = 0. Positive Differenzen bedeuten daher Unterschiede im Lernzuwachs zugunsten der ALLE-Schulen. Als Hintergrundmerkmale kontrolliert wurden Geschlecht, Alter, Migrationshintergrund, zuhause gesprochene Sprache, sozioökonomischer Status.

(*) $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$

4.10.2.4 Ergebnisse der Lernstandserhebung «Texte schreiben»

Neben den im Rahmen der Evaluation ALLE durchgeführten Tests konnten die Ergebnisse der Lernstandserhebung zum «Texte schreiben» herangezogen werden, die im Kanton Zürich etwa Mitte der 8. Klassenstufe flächendeckend durchgeführt wird. Da es sich um eine einmalige Testdurchführung handelt, liegen für den Kompetenzbereich «Texte schreiben» keine Längsschnittdaten vor. Um dennoch Rückschlüsse auf die Wirkung des ALLE-Massnahmenbündels in diesem Bereich ziehen zu können, wurde in den nachfolgend berichteten Analysen neben den bereits bekannten Hintergrundmerkmalen der Schülerinnen und Schüler die Deutsch-Gesamtleistung zum ersten Messzeitpunkt (T1; Anfang 7. Klasse) aus den Tests der Evaluation ALLE als erklärende Variable (Prädiktor) berücksichtigt. Konkret wurden zwei Regressionsmodelle berechnet, um zu untersuchen, welche Variablen die Testleistung im «Texte schreiben» erklären können. Als erklärende Variablen wurden neben der Dummy-Variable ALLE- vs. Vergleichsschule (vgl. Einleitung Kap. 4.10) im ersten Modell (1) nur die Deutsch-Gesamtleistung zum ersten Messzeitpunkt, im zweiten Modell (2) zusätzlich die Hintergrundmerkmale der Schülerinnen und Schüler eingeführt. Aufgrund der breiten Anlage der Deutschtests in der Evaluation ALLE wird davon ausgegangen, dass mit diesem Vorgehen annäherungsweise auch Ausgangsunterschiede zwischen den Schülerinnen und Schülern in der Kompetenz zum Verfassen von Texten berücksichtigt werden konnten. Dennoch müssen die nachfolgenden Befunde zurückhaltend interpretiert werden.

Für den Test «Texte schreiben» wurde im Rahmen einer Normierungsstudie im Kanton Zürich ein Mittelwert von 500 und eine Standardabweichung von 100 festgelegt. Der Mittelwert der ALLE-

Schulen liegt mit 494 Punkten vergleichsweise nah an diesem Gesamtmittelwert, der Mittelwert der Vergleichsschulen deutlicher darunter mit 481 Punkten. Die Differenz von 13.1 Punkten ($SE = 10.2$ Punkte, $d = 0.12$) erreicht jedoch keine statistische Bedeutsamkeit. Tabelle 64 stellt den vorhergesagten Unterschied zwischen den ALLE- und Vergleichsschulen im «Texte schreiben» in den beiden Regressionsmodellen dar.

Tabelle 64. Unterschied in der Testleistung zwischen ALLE- und Vergleichsschulen für den Kompetenzbereich «Texte schreiben»

Kontrollierte Variablen	Differenz in der Testleistung	(SE)
Deutsch-Gesamtleistung T1	12.7	(10.2)
Deutsch-Gesamtleistung T1 + Hintergrundmerkmale	13.7	(9.4)

Anmerkungen. ALLE-Schulen $N = 970$ und Vergleichsschulen $N = 717$. Grundlage für den Vergleich der Leistung im «Texte schreiben» zwischen ALLE- und Vergleichsschulen ist eine Dummy-Variable: Schülerinnen und Schüler in den ALLE-Schulen = 1, Schülerinnen und Schüler in den Vergleichsschulen = 0. Positive Differenzen bedeuten daher Unterschiede zugunsten der ALLE-Schulen.

(*) $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$

Die positiven Koeffizienten in beiden Modellen deuten in die Richtung einer günstigeren Leistungsentwicklung in den ALLE-Schulen, doch ist diese Differenz nicht statistisch signifikant und liegt bei Ausschluss der später einbezogenen beziehungsweise frühzeitig ausgeschiedenen Vergleichsschulen nur noch knapp im positiven Bereich. Damit ergibt sich kein schlüssiger Hinweis auf eine günstigere Entwicklung der Schreibkompetenz in den ALLE-Schulen.

4.11 Explorative Analyse: Gelingensbedingungen von ALLE

Durch Vergleiche zwischen den ALLE- und Vergleichsschulen mit Blick auf verschiedene Lernergebnisse der Schülerinnen und Schüler wurden bereits Indikatoren zum Erfolg des ALLE-Massnahmenbündels gewonnen. Weitere Befunde deuten allerdings darauf hin, dass sich die Durchführungsbedingungen und die eigentliche Durchführung der Massnahmen zwischen den ALLE-Schulen unterschieden. Aus einer Untersuchung der Merkmale der Umsetzung, welche die Schulen mit besseren Ergebnissen kennzeichnen, könnten sich wertvolle Hinweise auf die Gelingensbedingungen beziehungsweise Erfolgsfaktoren der Schulen im Pilotprojekt ergeben. In den folgenden explorativen Analysen wird versucht, Anhaltspunkte für solche Gelingensbedingungen zu finden. Dazu werden Verfahren eingesetzt, welche möglichst informative und zugleich belastbare Resultate auch mit kleinen Stichproben liefern. Im ersten Schritt werden in Kapitel 4.11.1 Analysen dar-

gestellt, die dazu dienen, die relevante Analyseebene (Schule vs. Klasse/Lerngruppe) für die weitere Untersuchung zu identifizieren. Darauf aufbauend wird die Datenbasis für die nachfolgenden Analysen beschrieben. In Kapitel 4.11.2 wird ausführlicher auf die Analysestrategie eingegangen, die zur Aufdeckung potenziell bedeutsamer Erfolgsfaktoren eingesetzt wird. Abschliessend werden in Kapitel 4.11.3 die verschiedenen Analyseschritte und die Ergebnisse der Analysen dargestellt.

4.11.1 Bestimmung der Analyseebene und Datenbasis

In einem ersten Schritt erfolgt die Klärung, auf welcher Ebene (Schule vs. Klassen-/Lerngruppen) eine Suche nach Gelingensbedingungen der ALLE-Massnahme mit Blick auf statistische Kriterien eher lohnenswert erscheint (Kap. 4.11.1.1). Im Anschluss wird darauf eingegangen, welche Stichprobe den Analysen zugrunde liegt (Kap. 4.11.1.2) und welche Variablen aus verschiedenen verfügbaren Datenquellen in welcher Form in die Analysen einbezogen wurden (Kap. 4.11.1.3).

4.11.1.1 Klärung der Analyseebene

Bisherige Befunde der Schul- und Unterrichtsforschung legen nahe, dass Unterschiede in den Unterrichtsprozessen und Lernergebnissen der Schülerinnen und Schüler zwischen Klassen der gleichen Schule meist stärker ausgeprägt sind als zwischen den Schulen. Für die Suche nach Faktoren, welche den Erfolg der ALLE-Massnahmen beeinflusst haben könnten, lohnt es sich demnach zu prüfen, auf welcher Ebene die Unterschiede in Bezug auf interessierende Kriterien (abhängige Variablen) hauptsächlich angesiedelt sind. Findet sich auf einer der beiden Ebenen (Schule bzw. Klasse/Lerngruppe) kaum entsprechende Variation, sind weitere Analysen auf dieser Ebene wenig erfolgversprechend, und die Analysen sollten auf die andere Ebene fokussieren.

Der *Intraklassen-Korrelationskoeffizient* (ICC, engl. *intraclass correlation coefficient*; Fisher, 1958) drückt aus, welcher Anteil der Gesamtvarianz einer Variable durch Unterschiede zwischen Gruppen (Schulen und Lerngruppen/Klassen) erklärt werden kann (Raudenbush & Bryk, 2002; Snijders & Bosker, 2012). Wie sich für die ALLE-Schulen zeigt, sind die ICCs der Unterrichtsprozess- und Ergebnisvariablen auf Klassenebene – mit einer Ausnahme beim Lernzuwachs im Gesamttestwert in Mathematik – wesentlich höher ausgeprägt als die ICCs auf der Schulebene, die unter praktischen Gesichtspunkten vernachlässigbar gering sind ($ICCs \leq 0.05$). Entsprechend fokussieren die nachfolgend berichteten Analysen auf die Erklärung von Unterschieden auf der Klassenebene. In Tabelle 65 werden die Varianzanteile auf den beiden Ebenen für diejenigen Variablen dargestellt, die in die nachfolgenden Analysen als abhängige Variablen eingingen (vgl. Kap. 4.11.1.3).

Tabelle 65. ICCs der Unterrichtsprozess- und Ergebnisvariablen auf Schul- und Klassenebene

Variable	Deutsch		Mathematik	
	Schulebene	Klassen- ebene	Schulebene	Klassen- ebene
Lernerfolg (Zuwachs in der Schülerleistung)	.02	.08	.05	.02
Zuwachs in anerkennungs- und leistungsbezogener Motivation	.04	.24	.02	.29
Veränderung in Verständlichkeit und Klarheit	.03	.41	.00	.22
Veränderung in Lernzielorientierung	.04	.35	.01	.22
Veränderung in Konsolidierung	.04	.35	.02	.21
Veränderung in Kognitive Aktivierung	.03	.34	.02	.20
Veränderung in Thematische Motivierung	.04	.35	.02	.19
Veränderung in Instrumentelle Motivierung	.04	.34	.02	.19
Veränderung in Leistungserwartungen	.04	.34	.03	.18
Veränderung in Unterrichtsklima	.03	.34	.03	.17
Veränderung in Positive Fehlerkultur	.03	.34	.03	.17
Veränderung in Lehrer-Schüler-Beziehung	.03	.34	.03	.17
Veränderung in Diagnostische Kompetenz	.03	.34	.03	.17
Veränderung in Individuelle Förderung	.03	.34	.03	.17
Veränderung in Allgegenwärtigkeit	.04	.34	.03	.17
Veränderung in Regelerorientierung	.05	.35	.03	.17
Veränderung in Nutzung der Lernzeit (Time on Task)/ Zeitmanagement	.05	.35	.03	.17

Anmerkungen. Die ICCs auf Schul- und Klassenebene pro Variable pro Fach wurden jeweils anhand eines Dreiebenen-Nullmodells mit der betreffenden Variable als abhängige Variable berechnet. Zugrunde lag der Berechnung die auf Klassen- und Schülerebene zwischen dem ersten und letzten Messzeitpunkt stabile Schülerschaft in den ALLE-Schulen (siehe Beschreibung in Kap. 4.11.1.2).

4.11.1.2 Stichprobe

Damit die Ergebnisse der Zusatzanalysen im Sinne einer Wirkung der umgesetzten Massnahmen auf Klassenebene interpretierbar sind, müssen sich die erhobenen Daten zur Umsetzung von ALLE

zuverlässig mit Veränderungen von abhängigen Variablen *in einzelnen Klassen/Lerngruppen* in Beziehung setzen lassen. So können beispielsweise Änderungen in Einschätzungen der Unterrichtsqualität einer Klasse nicht auf die Umsetzung der ALLE-Massnahmen zurückgeführt werden, wenn es zwischenzeitlich einen Lehrpersonenwechsel gegeben oder sich die Schülerzusammensetzung in der Klasse zwischen dem Anfang und dem Ende des Pilotprojekts stark geändert hat. Diesbezüglich zeigen sich aber deutliche Einschränkungen der Daten.

So gibt es nur 23 von insgesamt 46 Klassen/Lerngruppen in Deutsch, bei denen es zwischen dem ersten Messzeitpunkt (T1) und dem letzten Messzeitpunkt (T4) der Evaluation keinen Lehrpersonenwechsel gab *und* in denen mindestens 80 Prozent der Schülerinnen und Schüler – was hier als Indikator für eine ungefähr stabile Schülerschaft dienen soll – gleich blieben. Insgesamt gibt es 369 Schülerinnen und Schüler in diesen 23 Lerngruppen im Fach Deutsch, die von 19 Lehrpersonen unterrichtet wurden. Im Fach Mathematik sind es nach den gleichen Kriterien 23 beziehungsweise 46 Lerngruppen (jedoch ist die Zuordnung der Schülerinnen und Schüler zu den Lerngruppen anders als im Fach Deutsch) mit insgesamt 347 Schülerinnen und Schülern, die von 21 Mathematik-Lehrpersonen unterrichtet wurden. Diese fachspezifischen Stichproben liegen den nachfolgend dargestellten Analysen zugrunde.

Es ist wichtig zu beachten, dass diese Stichproben gegenüber den ALLE-Gesamtstichproben im Umfang stark reduziert sind. So gehört pro Fach jeweils eine ganze Schule *nicht* in die Stichproben dieses Kapitels, da es in diesen keine Klassen/Lerngruppen gab, die beide oben genannten Bedingungen – keinen Lehrpersonenwechsel und eine vergleichsweise stabile Schülerschaft – erfüllten. Insgesamt können Daten von rund 50 Prozent der teilnehmenden Klassen/Lerngruppen und von etwa einem Drittel der ALLE-Schülerstichprobe pro Fach einbezogen werden. Zwischen den Schulen variieren zudem die Anteile von berücksichtgbaren Daten auf Schüler- und Klassenebene stark. Sie liegen zwischen elf und 89 Prozent auf der Ebene der Schülerinnen und Schüler und zwischen 25 und 50 Prozent auf der Ebene der Klassen/Lerngruppen.

4.11.1.3 Variablenauswahl

Erfolgsindikatoren (abhängige Variablen)

Auf der Seite der Erfolgsindikatoren werden zum einen die Veränderungen von leistungsbezogenen und motivationalen Lernergebnissen während des Evaluationszeitraums berücksichtigt.

- Als *leistungsbezogener Erfolgsindikator* wird der Mittelwert des fachspezifischen Kompetenzzuwachses – das heisst, die Differenz der Leistungen in den Gesamttests Deutsch beziehungsweise Mathematik zwischen dem letzten (T4) und ersten (T1) Messzeitpunkt der Evaluation – auf der Klassenebene herangezogen. Anders als in anderen Teilen des Berichts werden dafür die WLE-Schätzer (Warm, 1989) und nicht die Plausible Values (PVs;

vgl. Kap. 3.5.5) der Schülerleistungen herangezogen, da sie die besseren Schätzer für einzelne Klassen darstellen, während PVs für die Schätzung in einer gesamten Population besser geeignet sind (vgl. Robitzsch et al., 2016). Im Fach Mathematik wird nachfolgend aufgrund der äusserst geringen ICC auf Klassenebene ($ICC = 0.02$) keine Analyse mit dem leistungsbezogenen Erfolgsindikator durchgeführt.

- Als *lernmotivationsbezogener Erfolgsindikator* wird die Veränderung der anerkennungs- und leistungsbezogenen Motivation der Schülerinnen und Schüler zwischen T1 und T4, wiederum auf der Klassenebene, ausgewählt. Für diese Motivationsfacette resultiert in den Regressionsmodellen zur Wirkung des Pilotprojekts in beiden Fächern der im Betrag stärkste Koeffizient, zudem konnte nur für diese Facette ein statistisch bedeutsamer Unterschied zwischen ALLE- und Vergleichsschulen gezeigt werden (im Fach Mathematik; vgl. Kap. 4.10.1.1).

Darüber hinaus wird die Veränderung der *Unterrichtsqualität aus Schülersicht* sowohl als Zielkriterium als auch als vorhersagende Variable (Prädiktor) in die Analysen einbezogen. In der Evaluation erhoben wurden 15 fachspezifische Skalen zu Schülereinschätzungen der Unterrichtsqualität (vgl. Kap. 4.9.1). Ergebnisse von explorativen Faktorenanalysen deuten darauf hin, dass bereits ein einziger Faktor den grössten Teil der Variation in der Veränderung der schülerseitig wahrgenommenen Unterrichtsqualität auf sich vereinen kann. Um die Unterrichtsqualität dennoch differenziert abzubilden, wird den nachfolgenden Analysen eine Faktorenlösung zugrunde gelegt, die sich gut mit den drei Basisdimensionen der Unterrichtsqualität in Deckung bringen lässt (Klassenführung, Lernklima, kognitive Aktivierung; vgl. Praetorius et al., 2018). In den Faktor *Klassenführung* gehen die drei Skalen «Allgegenwärtigkeit», «Regelorientierung» und «Nutzung der Lernzeit» ein. Der Faktor *Lernklima* umfasst die Skalen «Unterrichtsklima», «Positive Fehlerkultur» und «Lehrer-Schüler-Beziehung». Die *kognitive Aktivierung* wird über ein einzelnes Item abgebildet. Als abhängige Variable für die nachfolgenden Analysen hinzugenommen wird aufgrund der Bedeutung für das Pilotprojekt zusätzlich die Veränderung bezüglich der individuellen Förderung (ebenfalls als Einzelitem), die zu keinem der drei genannten Faktoren passt. Zur Gewinnung der Indikatoren der Unterrichtsqualität für die nachfolgenden Analysen werden die Klassenmittelwerte in der Veränderung der betreffenden Variablen (d.h. der Differenzen zwischen T4 und T1) berechnet.

Prädiktoren

Grundsätzlich wurden in die Analysen alle verfügbaren quantitativen Daten einbezogen, die sich auf die Qualität der Umsetzung der ALLE-Massnahmen sowie auf die Merkmale beziehen, welche diese direkt beeinflussen können:

- Variablen aus dem Schulleitungsfragebogen
- Variablen aus den Lehrpersonen-Fragebögen

- Variablen aus den Schülerfragebögen
- Kontext- und Prozessvariablen

Da die verfügbaren Stichprobengrößen die mögliche Komplexität der Analysen jedoch beträchtlich einschränken, mussten Schritte unternommen werden, um die Anzahl der Prädiktoren zu reduzieren beziehungsweise Variablen auszuschliessen, von denen aufgrund ihrer psychometrischen Eigenschaften kaum Erklärungsgewinne zu erwarten waren.

Aus dem *Schulleitungsfragebogen* (SLFB) wurden alle einzelnen Items sowie Skalenwerte als Prädiktoren für die nachfolgenden Analysen ausgewählt, die folgende Kriterien erfüllten:

- Ausreichende Varianz in den Antworten zwischen den Schulleitungen war vorhanden (ausgeschlossen wurden alle Items, bei denen maximal die Hälfte der vorgegebenen Antwortoptionen gewählt wurden; dies führte zum Ausschluss von 42 Items).
- Items, die theoretisch zu einem Konstrukt gehören, hatten eine gute Trennschärfe (alle Items mit einer Trennschärfe kleiner als 0.3 wurden eliminiert; dies betraf weitere 9 Items).
- Alle Skalen, bei denen auf der Grundlage dieser Kriterien mehr als die Hälfte der Items ausgeschlossen wurden, wurden nicht mehr berücksichtigt (betraf 9 Skalen).
- Alle Skalen mit einem Cronbachs Alpha kleiner 0.7 sollten eliminiert werden; nach Berücksichtigung anderer Kriterien wurde jedoch keine Skala wegen dieses Kriteriums eliminiert.

Insgesamt resultierten auf dieser Grundlage sieben Skalen und zehn einzelne Items aus dem SLFB als Prädiktoren für die nachfolgenden Analysen. Bei der Bildung der individuellen Skalenwerte wurde die von Van Ginkel et al. (2007) vorgeschlagene Methode zur Korrektur unterschiedlicher Itemschwierigkeiten verwendet, die im R-Paket *miceadds* (Robitzsch & Grund, 2020) implementiert ist. In zwei Schulen lagen Antworten von mehr als einer Schulleitung vor, da im Verlauf des Pilotprojekts ein Schulleitungswechsel stattgefunden hatte. Nur die Antworten der Schulleitungen, die während der Projektlaufzeit länger in den Schulen waren, wurden weiter berücksichtigt.

Die Datenaufbereitung für den *Lehrpersonenfragebogen* (LFB) erfolgte nach Fächern getrennt und analog dem Vorgehen für den SLFB. Anders als beim SLFB entfielen im LFB nur wenige Items (in Mathematik: zwei Items – ein Item aus einer Skala und ein einzelnes Item; in Deutsch: vier Items – drei Items aus drei verschiedenen Skalen und ein einzelnes Item). Alle zehn Skalen im LFB wurden dagegen beibehalten, sowohl in Deutsch als auch in Mathematik. Die individuellen Skalenwerte wurden ebenfalls nach der Methode von Van Ginkel et al. (2007) berechnet.

Aus den LFB-Angaben wurden Prozess- und Kontextdaten auf der Ebene der Klassen/Lerngruppen sowie der Schulen gewonnen und in die nachfolgenden Analysen einbezogen. Dazu wurden den Klassen/Lerngruppen pro Fach (mit vergleichsweise stabiler Schülerschaft und ohne Lehrpersonenwechsel zwischen T1 und T4, vgl. Kap. 4.11.1.2) die Lehrpersonendaten zugeordnet. Für

alle (verbliebenen, vgl. oben) Skalen und Items des LFB wurden zudem die Schulmittelwerte berechnet, die als Kontextdaten auf Schulebene in die nachfolgenden Analysen eingingen. Pro Schule gab es zwar in der Regel Angaben von mehreren Lehrpersonen pro Fach, die zu einem Schulmittelwert zusammengefasst werden konnten. Wichtig zu betonen ist jedoch, dass es sich dabei nicht um eine Vollerhebung beziehungsweise um eine repräsentative Stichprobe der Lehrpersonen in den ALLE-Schulen handelte. Die Anzahl der einbezogenen Lehrpersonen variierte zudem zwischen einer und sieben pro Schule und Fach.

Weitere Kontextdaten auf der Schul- und Klassen-/Lerngruppenebene wurden aus den Schülerdaten gewonnen. Dafür wurden die Mittelwerte sowie die Standardabweichungen auf der Schul- und Klassenebene aller Leistungs- und Motivationsvariablen sowie der Unterrichtsqualitätsskalen am Anfang des Projekts (T1) berechnet. Darüber hinaus wurden Mittelwert und Standardabweichung des sozioökonomischen Status der Schülerinnen und Schüler auf Schul- und Klassenebene einbezogen.

Pro Fach wurde anschliessend ein Gesamtdatensatz mit allen gewonnenen Variablen aus SLFB, LFB und Schülerdaten erstellt. Insgesamt resultierten daraus 84 Variablen pro Fach, die für die nachfolgend dargestellten Zusatzanalysen herangezogen wurden.

Datenimputation

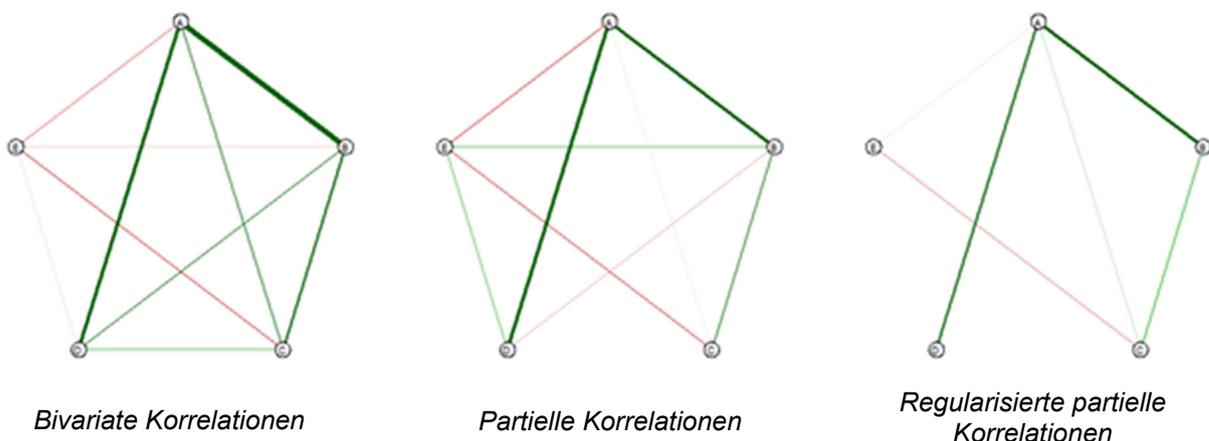
Um mit den fehlenden Werten auf den Analysevariablen umzugehen, wurde eine Imputation basierend auf *Chained Equations* durchgeführt (van Buuren, 2012; vgl. Kapitel 3.5.3). Aufgrund des explorativen Charakters der Analysen wurde pro Fach nur ein vollständig imputierter Datensatz erstellt, der als Basis für die nachfolgenden Analysen diente.

4.11.2 Analysestrategie

Mit einer sehr kleinen Stichprobe auf Klassenebene ($N = 23$) und einer grossen Anzahl von Variablen (> 80) sind gängige frequentistische Verfahren zur Datenanalyse und Inferenz auf zugrunde liegende Populationen nicht einsetzbar. Ziel der dargestellten Zusatzanalysen sollte es jedoch nicht sein, verallgemeinerbare Aussagen über die Gelingensbedingungen von Schul- und Unterrichtsentwicklungsmassnahmen zu treffen, sondern möglichst gut gestützte Hypothesen dazu zu generieren, welche Faktoren zum grösseren oder weniger grossen Erfolg einzelner ALLE-Schulen beigetragen haben könnten. Dafür wurden Analysestrategien herangezogen, die für das «grosses p , kleines N » (grosse Anzahl von Prädiktoren bei kleiner Stichprobe) Problem speziell geeignet sind. Ein solches Verfahren ist die regularisierte Methode unter Verwendung von *lasso* (*least absolute shrinkage and selection operator*; Tibshirani, 1996), welche häufig in Forschungsfeldern mit typischerweise kleinen Stichproben und einer extrem grossen Anzahl von Prädiktoren – wie beispielsweise der Genforschung (Tutz, 2012) – eingesetzt wird.

Zur Identifikation von Gelingensbedingungen der ALLE-Schulen wurden mit *lasso* die Zusammenhänge zwischen verschiedenen potenziellen Erfolgsfaktoren (z.B. Qualitätsmerkmale der Durchführung von Massnahmen) untereinander und mit den ausgewählten abhängigen Variablen (z.B. durchschnittliche Leistungszuwächse) explorativ analysiert. In der dafür verwendeten graphischen Netzwerkanalyse mit *lasso* werden die Variablen als Knoten und die Korrelationen beziehungsweise partiellen Korrelationen zwischen diesen als Kanten betrachtet (Epskamp, Cramer, Waldorp, Schmittmann & Borsboom, 2012). Mit Hilfe des R Paket *qgraph* (Epskamp et al., 2012) kann die Zusammenhangsstruktur zwischen den Konstrukten analysiert und visualisiert werden. Genutzt werden dafür entweder die bivariaten Korrelationen, die partiellen Korrelationen oder die optimierten partiellen Korrelationen, welche mittels der graphischen *lasso*-Methode berechnet werden (Friedman, Hastie & Tibshirani, 2008). Abbildung 76 zeigt ein Beispiel für ein Zusammenhangsmuster zwischen fünf Konstrukten A, B, C, D und E unter Verwendung dieser drei Methoden.

Abbildung 76. Zusammenhangsstruktur zwischen fünf beispielhaften Konstrukten mittels graphischer Netzwerkanalyse



Anmerkungen: grüne Linie = positiver Zusammenhang; roter Linie = negativer Zusammenhang; helle Farbe = Korrelation, die im absoluten Betrag geringer als 0.3 ist; je stärker die Linie, desto höher ist die Korrelation zwischen zwei Konstrukten.

Unter Verwendung der partiellen Korrelationsmatrix ist es möglich, Hinweise auf «Scheinzusammenhänge» – welche sich nur aus der Beziehung zu dritten Variablen ergeben – zu erhalten, sodass diese entsprechend schwächer dargestellt werden (zwischen A und C, B und D, und C und D). Dies stellt einen entscheidenden Vorteil gegenüber Analysen dar, bei denen lediglich Gruppen (z.B. unterschiedlich erfolgreiche Schulen) im Hinblick auf die Ausprägung bei interessierenden Merkmalen (z.B. Aspekte der Umsetzungsqualität) verglichen werden. Mit der *lasso*-Methode werden zudem manche schwachen partiellen Korrelationen auf null gesetzt und die anderen Korrelationen in ihrem absoluten Betrag reduziert. Dadurch wird die partielle Korrelationsmatrix sparsamer

und besser interpretierbar. Ausserdem wird eine mögliche Überschätzung der Zusammenhänge durch Messfehler zum Teil korrigiert.

Es ist mit diesen Methoden also möglich, sich den wahren Zusammenhängen zwischen Variablen anzunähern, und man erhält (bezogen auf ALLE) eine zuverlässigere Schätzung der Beziehung zwischen Variablen zur Massnahmenumsetzung und den Ergebnisvariablen. Zur Interpretation der Ergebnisse sollten nur die theoretisch begründbaren und plausiblen Zusammenhänge berücksichtigt werden, um Fehlschlüsse aufgrund von Messfehlern zu vermeiden. Alle Zusammenhänge sind zudem ungerichtet, die Annahme kausaler Zusammenhänge bedarf einer plausiblen Begründung.

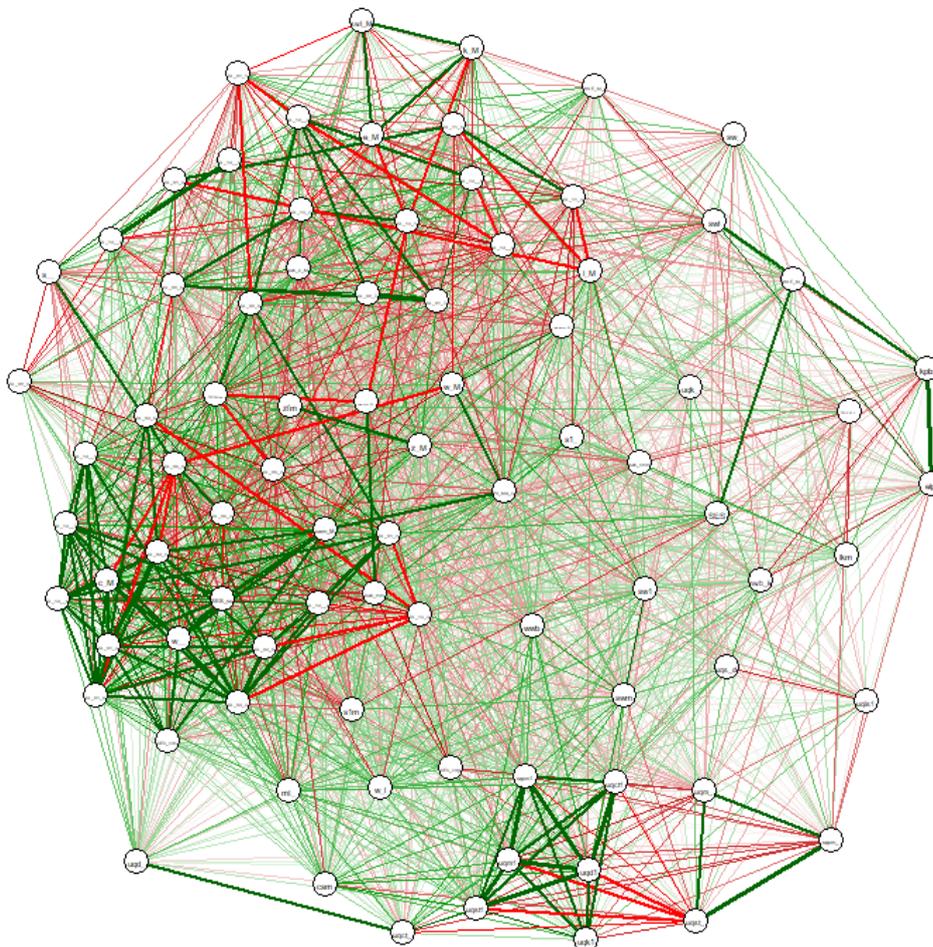
4.11.3 Schritte der Datenanalyse und Ergebnisse

Schritte der Datenanalyse

Zur Reduktion der Anzahl der Variablen zwecks verbesserter Interpretierbarkeit der Ergebnisse wurden zunächst die nicht bedeutsamen Prädiktoren ausgeschlossen. Dies sind Variablen, die entweder mit der abhängigen Variable praktisch nicht beziehungsweise nur schwach zusammenhängen, oder deren Vorhersagekraft für die abhängige Variable durch andere Variablen weitgehend erklärt werden kann. Dafür wurden zuerst die bivariaten Korrelationen zwischen allen Variablen im Datensatz pro Fach berechnet. Alle Korrelationen zwischen allen Variablen im Fach Mathematik werden exemplarisch in Abbildung 77 dargestellt. Als Kandidaten für bedeutsame Prädiktoren jeder abhängigen Variable wurden nur diejenigen Variablen berücksichtigt, die mindestens mittelstark ($|r| \geq 0.3$; Kategorisierung nach Cohen, 1988) mit der abhängigen Variable korrelieren.

Im nächsten Schritt gingen die Prädiktoren, die mindestens mittelstark mit der jeweiligen abhängigen Variable zusammenhängen, in eine *lasso*-Regressionsanalyse ein (Tibshirani, 1996). Mit dieser Methode können Prädiktoren mit Scheinzusammenhängen mit der abhängigen Variable (auch bei mehreren Prädiktoren und kleiner Stichprobe, hier $N = 23$) identifiziert werden. Das heisst, es können Variablen bestimmt werden, deren Zusammenhänge mit der abhängigen Variable fast komplett durch andere Prädiktoren erklärt werden können. Diese erhalten einen *lasso*-Regressionskoeffizienten von null. Nach diesem Schritt verbleiben nur noch Prädiktoren, die wesentlich zur Erklärung der abhängigen Variable beitragen. Dies sind alle Variablen mit einem *lasso*-Regressionskoeffizienten, der im Betrag von null abweicht (siehe Pham, Robitzsch, George & Freunberger, 2016, für eine genaue Beschreibung der Methode). Im Bildungsbereich wurde die *lasso*-Regression für den Fairen Vergleich des nationalen Ergebnisberichts der Bildungsstandardüberprüfung in Österreich (Pham et al., 2016) und in Bildungsstudien mit kleinen Stichproben angewendet (McNeish, 2015; Pham, 2018).

Abbildung 77. Bivariate Korrelationen zwischen allen 84 berücksichtigten Variablen im Fach Mathematik



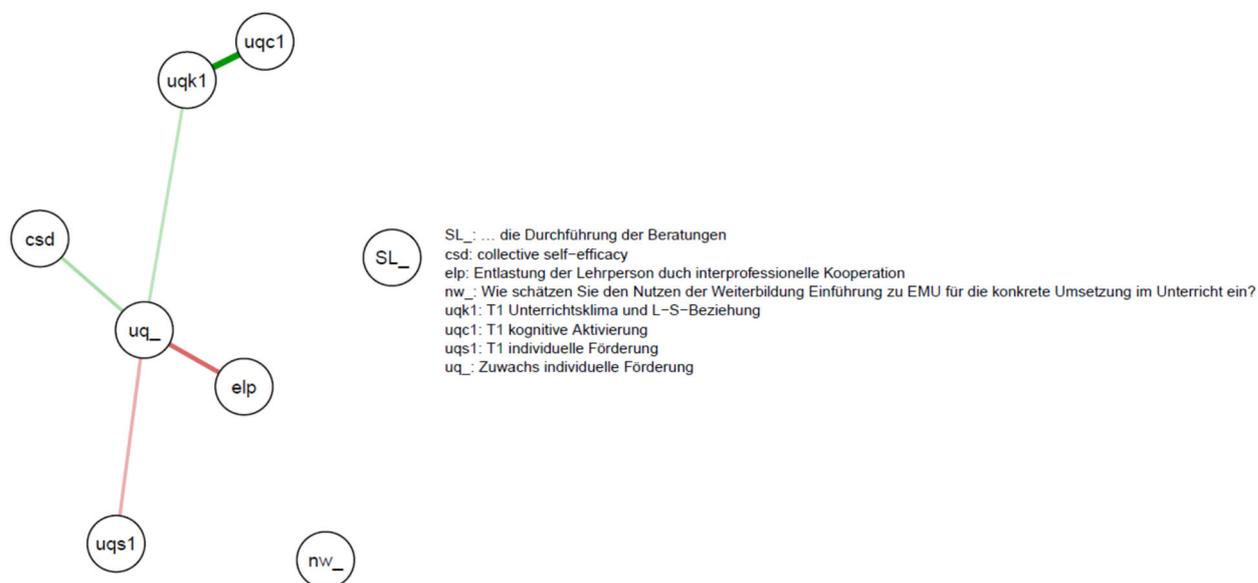
Anschliessend wurde pro abhängige Variable ein partielles Korrelationsnetzwerk mit *lasso* berechnet, um die Zusammenhänge zwischen allen Prädiktoren mit der abhängigen Variable sowie miteinander zu visualisieren. Im Modell wurden alle Prädiktoren mit einem Nicht-Null-Koeffizienten aus dem vorherigen Schritt sowie die abhängige Variable berücksichtigt.

Ergebnisse: Korrelationsnetzwerke

Im Folgenden werden die Ergebnisse der Bildung von Korrelationsnetzwerken zusammenfassend dargestellt, zunächst für das Fach Deutsch, dann das Fach Mathematik. Dabei wird auf die zentralen und für das vorliegende Forschungsinteresse relevanten Befunde fokussiert, es erfolgt daher keine Darstellung jedes einzelnen Netzwerks..

Fokussiert man zunächst auf den *Unterricht im Fach Deutsch*, so finden sich relevante Zusammenhänge für die Veränderung der von den Schülerinnen und Schülern wahrgenommenen „Individuellen Unterstützung“. Die Ergebnisse werden in Abbildung 78 dargestellt.

Abbildung 78. Korrelationsnetzwerks mit der Veränderung der Unterrichtsqualitätsvariable Individuelle Unterstützung (Variable «uq_») im Fach Deutsch



Anmerkungen. Rote Linien stellen negative Zusammenhänge dar, grüne Linien stellen positive Zusammenhänge dar. Je dicker und stärker eine Linie ist, desto höher ist der Zusammenhang im absoluten Wert. Hellgrüne und hellrote Linien stellen Zusammenhänge dar, die einen absoluten Wert kleiner als 0.1 haben.

Es wird ersichtlich, dass vier Variablen einen nennenswerten direkten Zusammenhang mit der abhängigen Variable haben, unter Berücksichtigung aller anderen Prädiktoren im Modell. Ein positiver Zusammenhang besteht zwischen der wahrgenommenen individuellen Unterstützung und der Lehrpersonen-Skala zur Collective Self-efficacy («csd») in der Schule. Die Collective Self-efficacy beschreibt das Ausmass, in dem eine Gruppe (hier: das Kollegium der Lehrpersonen) von ihrer Fähigkeit zum erfolgreichen gemeinschaftlichen Handeln überzeugt ist, und beeinflusst ihre Zielsetzungen, ihre kollektiven Anstrengungen und ihr Durchhaltevermögen, falls Hindernisse auftreten. Der gefundene Zusammenhang stimmt mit bisherigen Forschungsbefunden überein, dass sich eine hohe Collective Self-efficacy günstig auf die Unterrichtsqualität – in diesem Fall die individuelle Förderung der Schülerinnen und Schüler – auswirken kann. Erklärbar ist zudem der negative Zusammenhang zwischen dem Ausgangswert zur Individuellen Förderung («uqs1»; Schülereinschätzungen zu T1) und der Veränderung im Zeitverlauf. Er lässt sich als statistisches Artefakt («Regression zur Mitte») interpretieren, könnte aber auch darauf hindeuten, dass eine Steigerung bei der individuellen Förderung schwierig ist, wenn diese bereits zu Beginn des Pilotprojekts auf hohem Niveau stattgefunden hat.

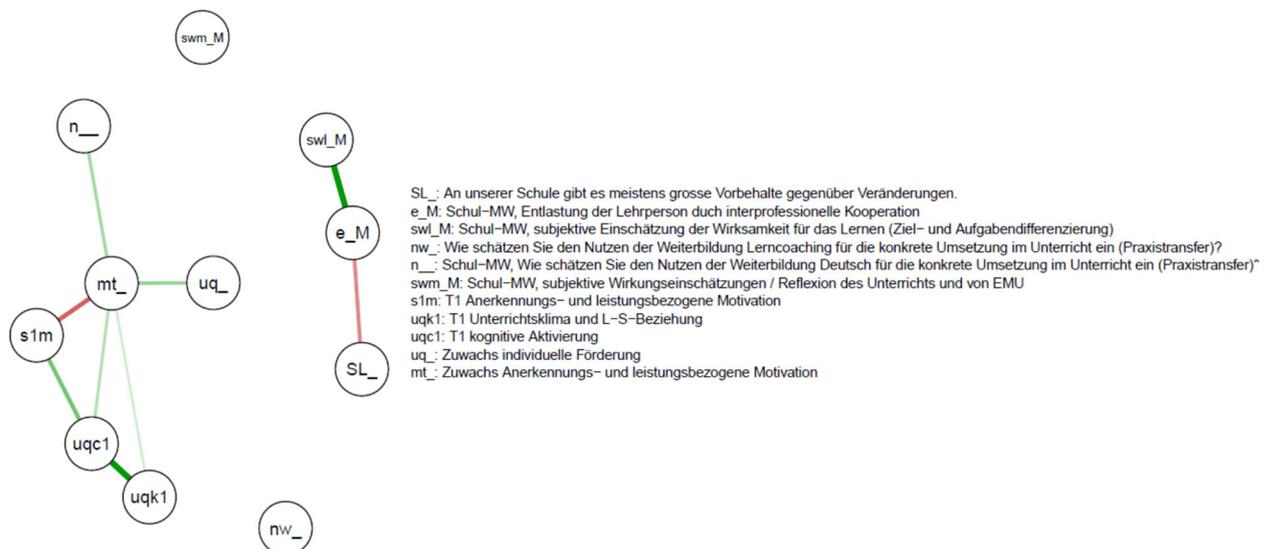
Ein positiver Zusammenhang besteht weiterhin zwischen dem Anfangswert der Schülereinschätzungen zum Lernklima («uqk1») und der Veränderung in der individuellen Förderung, der sich vielleicht am ehesten dahingehend erklären lässt, dass sich eine intensivere individuelle Förderung bei positivem Unterrichtsklima und guten Lehrer-Schüler-Beziehungen eher realisieren lässt. Im Betrag am stärksten, aber nicht leicht zu erklären ist der negative Zusammenhang mit der Lehrpersonen-Skala «Entlastung der Lehrperson durch interprofessionelle Kooperation» («elp»). Eine Interpretationsmöglichkeit wäre, dass die individuelle Förderung in manchen Klassen/Lerngruppen verstärkt von den Schulassistenzen übernommen wurde, und die Schülerinnen und Schüler die Förderung durch die Lehrperson selbst infolgedessen als weniger unterstützend erlebten.

Für den klassenspezifischen Zuwachs in den *Schülerleistungen im Fach Deutsch* resultiert kein Prädiktor, der eine gute Vorhersage leistet (keine Prädiktoren mit Nicht-Null-Koeffizienten nach *lasso*-Regression). Die Ergebnisse zur Veränderung der *aner kennungs- und leistungsbezogenen Motivation* der Schülerinnen und Schüler («mt_») werden in Abbildung 79 dargestellt. Als Prädiktor mit einer positiven Beziehung zur Veränderung der Motivation resultiert zum einen die Veränderung der Variable «Individuelle Förderung» («uq_»; vgl. oben). Dies deutet darauf hin, dass eine verbesserte individuelle Förderung mit einer günstigeren Entwicklung dieser Motivationsfacette einhergeht. Eine positive Beziehung der Veränderung der anerkennungs- und leistungsbezogenen Motivation besteht zudem mit der kollektiven Einschätzung der Lehrpersonen auf Schulebene bezüglich des Nutzens der fachbezogenen Weiterbildung für die Unterrichtsentwicklung («n_»). Daneben finden sich Zusammenhänge mit den anfänglichen Schülereinschätzungen zu den Unterrichtsqualitätsskalen Kognitive Aktivierung und Lernklima («uqc1» bzw. «uqk1»; Skalenwerte zu T1). Schliesslich ergibt sich wenig überraschend, dass die Anfangswerte (T1) und Veränderung der anerkennungs- und leistungsbezogenen Motivation negativ zusammenhängen (vgl. oben).

Für den *Unterricht im Fach Mathematik* zeigt sich nur ein direkter Zusammenhang: Die anfängliche Schülereinschätzung zur Strukturierung des Unterrichts (T1) steht in einer positiven Entwicklung mit der Veränderung der kognitiven Aktivierung aus Schülersicht (Netzwerk nicht dargestellt).

Im Fach Mathematik wurde zudem – wie bereits erwähnt – kein Analysemodell für den Zuwachs in der *Schülerleistung* auf Klassenebene berechnet, da kaum Unterschiede zwischen den Klassen festgestellt wurden (sehr geringe Intraklassen-Korrelation; vgl. Kap. 4.11.1.1).

Abbildung 79. Partielles Korrelationsnetzwerk mit der Veränderung der anerkennungs- und leistungsbezogenen Motivation im Fach Deutsch (Variable «mt_») als abhängige Variable

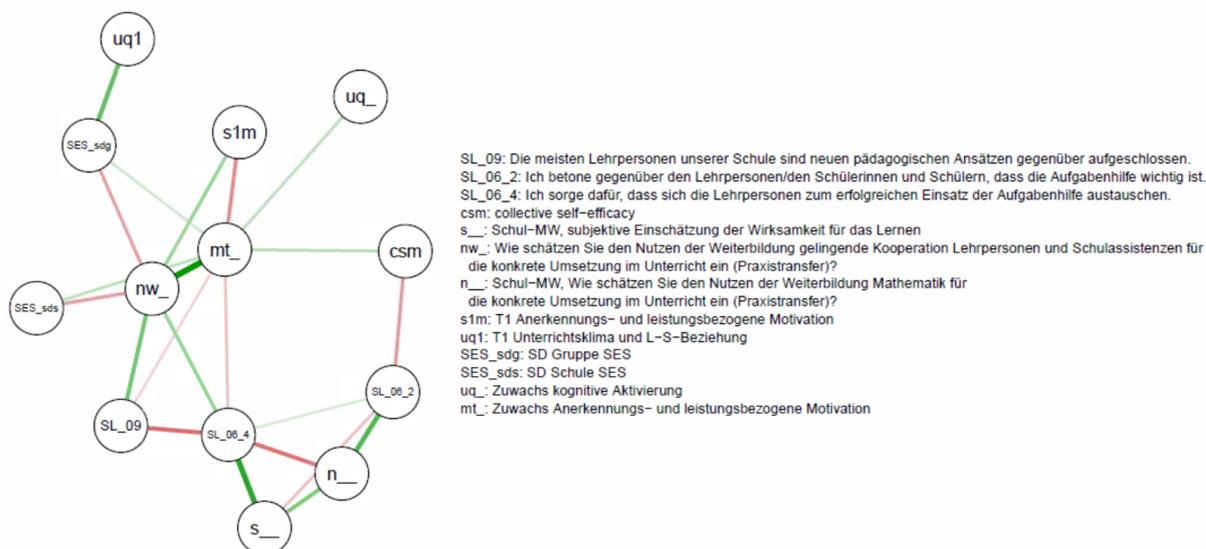


Anmerkungen. Rote Linien stellen negative Zusammenhänge dar, grüne Linien stellen positive Zusammenhänge dar. Je dicker und stärker eine Linie ist, desto höher ist der Zusammenhang im absoluten Wert. Hellgrüne und hellrote Linien stellen Zusammenhänge dar, die einen absoluten Wert kleiner als 0.1 haben.

Abbildung 80 zeigt die Ergebnisse für die *anerkennungs- und leistungsbezogene Motivation* im Fach Mathematik («mt_»). Als stärkster positiver Prädiktor resultiert hier die Einschätzung der Lehrpersonen zum Nutzen der Weiterbildung „Gelingende Kooperation Lehrpersonen und Schulassistenten“ («nw_»). Eine mögliche Interpretation wäre, dass die Weiterbildung zu einer besseren Zusammenarbeit von Lehrpersonen und Schulassistenten, und indirekt – beispielsweise durch mehr positives Feedback – zu einem Motivationszuwachs der Schülerinnen und Schüler beiträgt. Als weitere schwache positive Prädiktoren ergeben sich die Lehrpersonen-Einschätzung zur Collective Self-efficacy der Schule («csm»), die Veränderung der Unterrichtsqualitätsskala Kognitive Aktivierung aus Schülersicht («uq_») sowie die Heterogenität der Schülerschaft auf Klassen- und Schulebene («SES_sdg» bzw. «SES_sds»).

Zwei Variablen, die schwach negativ (partielle Korrelationen mit absolutem Wert kleiner als 0.1) mit der Entwicklung der anerkennungs- und leistungsbezogenen Motivation zusammenhängen, sind die Einschätzung der Schulleitung «Die meisten Lehrpersonen unserer Schule sind neuen pädagogischen Ansätzen gegenüber aufgeschlossen» («SL_09») und «Ich Sorge dafür, dass sich die Lehrpersonen zum erfolgreichen Einsatz der Aufgabenhilfe austauschen» («SL_06_4»). Da diese negativen Zusammenhänge inhaltlich nicht plausibel und statistisch sehr schwach sind, sollte auf eine Interpretation verzichtet werden.

Abbildung 80. Partielles Korrelationsnetzwerk mit der Veränderung der anerkennungs- und leistungsbezogenen Motivation im Fach Mathematik (Variable «mt_») als abhängige Variable



Anmerkungen. Rote Linien stellen negative Zusammenhänge dar, grüne Linien stellen positive Zusammenhänge dar. Je dicker und stärker eine Linie ist, desto höher ist der Zusammenhang im absoluten Wert. Hellgrüne und hellrote Linien stellen Zusammenhänge dar, die einen absoluten Wert kleiner als 0,1 haben.

Die berichteten Analysen liefern interessante, wenn auch nur punktuelle Hinweise auf Faktoren, welche zu einer gelingenden Umsetzung der ALLE-Massnahmen beigetragen haben könnten. Aufgrund der sehr kleinen Stichproben (auf Klassen/Lerngruppen- bzw. Schulebene) und des rein explorativen Charakters der Analysen lassen sich diese Befunde jedoch lediglich als *Hypothesen* hinsichtlich möglicher Gelingensbedingungen auffassen.

Als mögliche Gelingensbedingungen im Fach Deutsch für eine günstige Entwicklung der anerkennungs- und leistungsbezogenen Motivation lassen sich – mit den genannten deutlichen Einschränkungen – eine verbesserte individuelle Förderung im Unterricht, qualitativ hochwertige (fachbezogene) Weiterbildungen und eine bereits zu Beginn hohe Unterrichtsqualität identifizieren. Eine verbesserte Collective Self-efficacy lässt sich mit einer verbesserten individuellen Förderung (aus Schülersicht) im Deutschunterricht sowie mit einer günstigeren Entwicklung der anerkennungs- und leistungsbezogenen Motivation im Fach Mathematik in Beziehung bringen. Zu Letzterer trägt, wie sich in den Analysen andeutet, auch die (Weiterbildung für) gelingende Kooperation zwischen Lehrpersonen und Schulassistenzen bei.

Bezüglich der Kompetenzzuwächse im Fach Mathematik konnten keine nennenswerten Unterschiede zwischen den Klassen festgestellt werden, sodass auf eine Untersuchung von möglichen Gelingensbedingungen verzichtet wurde. Unterschiede zeigen sich stärker auf der Ebene der

Schulen (vgl. Kap. 4.11.1.1), welche jedoch nicht Gegenstand der hier berichteten Analysen ist. Über die Gründe dieser geringen Disparitäten kann nur spekuliert werden. Auffallend ist, dass sich die Klassen zwar bezüglich einzelner Unterrichtsqualitätsmerkmale durchaus unterschiedlich entwickelten, diese Unterschiede jedoch deutlich geringer ausfallen als im Fach Deutsch (vgl. Tabelle 65, S. 213). Möglich ist auch, dass diese Unterschiede bei einzelnen Merkmalen mit keiner konsistent günstigen (bzw. weniger günstigen) Entwicklung des Unterrichts in den einzelnen Klassen einherging. Im Fach Deutsch waren die Unterschiede in den Kompetenzzuwächsen der Klassen zwar grösser, doch ergaben sich keine Hinweise auf spezifische Gelingensbedingungen. Dabei ist zu beachten, dass die Identifikation entsprechender Zusammenhänge mit der vorliegenden Datenlage (geringe Anzahl von Klassen) und aufgrund des gewählten Vorgehens bei der Analyse (wechselseitige statistische Kontrolle potenziell erklärender Merkmale) herausfordernd war. Die Resultate sollten dahingehend interpretiert werden, dass für keines der untersuchten Merkmale die Vermutung eines Zusammenhangs mit den Deutschleistungen besonders naheliegt (i.S. der Hypothesenbildung, vgl. oben), nicht als Nachweis, dass diese Merkmale tatsächlich irrelevant sind.

5 Zusammenfassung, Bewertung und Entwicklungsperspektiven

Im Folgenden werden die zentralen Befunde der Evaluation des Pilotprojekts ALLE zusammenfassend bewertet, bezogen auf die Rahmenbedingungen des Pilotprojekts (Kap. 5.1), die im Pilotprojekt realisierten Massnahmen (Kap. 5.2) und die Wirkungen des Pilotprojekts (Kap. 5.3). Daran anschliessend wird auf die Grenzen der Aussagekraft der vorliegenden Evaluation eingegangen (Kap. 5.4). Abschliessend werden für den weiteren Umgang mit den Ergebnissen Handlungsfelder abgeleitet, und es werden Empfehlungen für weiterführende Massnahmen gegeben (Kap. 5.5).

5.1 Rahmenbedingungen des Pilotprojekts ALLE

Das Wirkungspotenzial eines Schul- und Unterrichtsentwicklungsprogramms wie ALLE hängt zunächst davon ab, in welchem Umfang und mit welcher Intensität die Angebote von den Zielpersonen – insbesondere den Lehrpersonen – tatsächlich genutzt werden konnten. Hierfür sind nicht zuletzt die finanziellen, personellen und organisatorischen Rahmenbedingungen entscheidend, unter denen das Projekt an den einzelnen Schulen realisiert wurde.

Die für die ALLE-Massnahmen erforderlichen finanziellen Ressourcen wurden auf Antrag der Schulleitung durch die Schulbehörden gesprochen. Damit waren die ALLE-Schulen davon abhängig, dass die für die Umsetzung benötigten Mittel rechtzeitig und im benötigten Umfang zur Verfügung gestellt wurden. Die den Schulen im Rahmen des Pilotprojekts bereitgestellten Mittel wurden von den Schulleitungen insgesamt als gut bis sehr gut bewertet. Eine zurückhaltendere Einschätzung ergibt sich an einigen Schulen bei den Mitteln für die Schulassistenzen und die Aufgabenhilfe. In den Interviews mit den Schulleitungen wurde darüber hinaus vereinzelt eine eingeschränkte Flexibilität bei der Mittelverwendung angesprochen, die zumindest teilweise aus einer zu frühzeitigen Festlegung des Mitteleinsatzes resultierte. Informationen, die dem Evaluationsteam zusätzlich seitens der ALLE-Projektleitung zur Verfügung gestellt wurden, weisen zudem darauf hin, dass für die Umsetzung der Massnahmen erforderliche Mittel teilweise zu spät oder für andere Zwecke gesprochen wurden. Daraus ergab sich nach Auskunft der Projektleitung, dass an einigen Schulen Massnahmen wie die Schulassistenten oder die Aufgabenhilfe erst verzögert umgesetzt werden konnten. Auch die Intensität, mit welcher die Massnahmen implementiert werden konnten, sei in den Schulen je nach finanzieller Ressourcenlage verschieden gewesen.

Von einer weiteren Problematik berichteten vor allem Lehrpersonen in den Interviews: nicht ausreichende Zeitkontingente, um die Massnahmen in ALLE noch umfassender oder gewinnbringender zu nutzen. Konkrete Hinweise hierauf ergaben sich für die Teilnahme an den Weiterbildungsangeboten, die Inanspruchnahme von Beratungsangeboten und den Austausch zwischen Lehrpersonen und Schulassistenten (vgl. die nachfolgenden Ausführungen zu den entsprechenden Massnahmen). Allerdings lässt sich aus den vorliegenden Daten nicht mit Gewissheit ableiten, ob fehlende zeitliche Ressourcen in grösserem Umfang oder nur in Einzelfällen zu Einschränkungen bei der Umsetzung von Massnahmen geführt haben könnten.

Die Anlage des Pilotprojekts sah vor, dass die Schulen – zumindest bezogen auf die Klassen der teilnehmenden Schülerinnen und Schüler – jeweils als Einheit mit der Umsetzung der Massnahmen betraut waren. Daraus resultierte ein hoher Bedarf an wechselseitigen Absprachen und an Kooperation insbesondere im Fachkollegium, sodass die innerschulische Zusammenarbeit im Rahmen von ALLE als wichtige Gelingensbedingung des Projekts verstanden werden kann. Dies entspricht Befunden zur Wirkung von Schulentwicklungsmassnahmen, welche die Bedeutung der Qualität der Zusammenarbeit im Kollegium hervorheben (vgl. Hattie & Zierer, 2018). Ferner lässt sich eine verbesserte Kooperation in den Schulen auch als ein mögliches wünschenswertes Ergebnis von ALLE auffassen. Dementsprechend wurde im Rahmen der Evaluation vor allem nach Verbesserungen der Zusammenarbeit als Folge des Pilotprojekts gefragt.

Die Aussagen der Lehrpersonen deuten zwar darauf hin, dass das Pilotprojekt bezüglich der erfragten Aspekte (z.B. Etablierung besserer Koordinationsstrukturen, bessere Regeln zur Zusammenarbeit) eher selten zu grundlegenden Veränderungen der kollegialen Kooperation führte. Jedoch berichtete die deutliche Mehrheit der Lehrpersonen in beiden Fächern, dass sich zumindest in einzelnen Bereichen der Zusammenarbeit positive Veränderungen ergeben hätten. Äusserungen der Lehrpersonen in den Interviews deuten zudem darauf hin, dass eine intensive Zusammenarbeit in einigen der Schulen bereits vor Umsetzung des ALLE-Massnahmenbündels etabliert war. Tatsächlich war das Interesse an innerschulischer Zusammenarbeit eine Voraussetzung, um am Pilotprojekt teilnehmen zu können (vgl. Kap. 3.3.1). Dementsprechend könnten begrenzte Verbesserungen teilweise durch eine positive Vorselektion der Schulen bezüglich der innerschulischen Kooperation erklärt werden. In anderen Schulen scheinen Kooperationen vor allem projektbezogen intensiviert worden zu sein, mit einem entsprechenden Rückgang gegen Ende des Projekts.

5.2 Das Massnahmenbündel im Pilotprojekt ALLE

Im Pilotprojekt ALLE wurde ein umfassendes Massnahmenbündel konzipiert, das von den teilnehmenden Schulen teils obligatorisch umgesetzt wurde und teils freiwillig in Anspruch genommen werden konnte. Nachfolgend werden die Befunde zum Einsatz der Schulassistenten (Kap. 5.2.1),

der Aufgabenhilfe (Kap. 5.2.2), den Weiterbildungen für Lehrpersonen (Kap. 5.2.3), den Beratungsangeboten für Lehrpersonen (Kap. 5.2.4), zum Einsatz von EMU (Kap. 5.2.5) und zur Individualisierung im Fachunterricht (Kap. 5.2.6) zusammenfassend dargestellt.

5.2.1 Schulassistenzen

Die Rekrutierung von Schulassistenzen durch die Schulen war gemäss den Aussagen fast aller Schulleitungen unproblematisch. Der Erfolg bei der Rekrutierung spiegelt sich auch in der insgesamt hohen zeitlichen Konstanz in der Beschäftigung der Schulassistenzen wider. Die Schulleitungen berichteten, bei der Rekrutierung auf die Einstellung von qualifizierten Schulassistenzen zu achten, und mit den Lehrpersonen im Gespräch über den Einsatz der Schulassistenzen zu stehen. Die Mehrheit der Schulleitungen gab zudem an, die Umsetzung der Massnahme durch die Organisation von Austauschmöglichkeiten zwischen den Lehrpersonen über den erfolgreichen Einsatz der Schulassistenzen zu unterstützen.

Hinsichtlich der Qualifikationen der Schulassistenzen lässt sich feststellen, dass die Mehrheit die in den Empfehlungen des Kantons Zürich festgelegten Mindestanforderungen erfüllten (siehe Bildungsdirektion Kanton Zürich/Volksschulamt, 2018). Zugleich zeigt sich ein breites Spektrum an Vorerfahrungen, welche die Schulassistenzen aufgrund der vor ihrer Beschäftigung als Schulassistentin ausgeübten Tätigkeiten mitbrachten. Diese Vorerfahrungen dürften für bestimmte Einsatzbereiche (z.B. Verwaltungstätigkeiten) von Vorteil gewesen sein – Interviewaussagen von Schulassistenten weisen entsprechend darauf hin, dass diese ihre beruflichen Vorerfahrungen durchaus als Gewinn in der Arbeit mit den Schülerinnen und Schülern erlebten. Andererseits können fehlende Vorqualifikationen im pädagogischen Bereich zu Schwierigkeiten beim Einsatz im Unterricht führen, insbesondere wenn die Unterstützung der Schülerinnen und Schüler beim Bewältigen fachlicher Anforderungen im Vordergrund steht (z.B. Bacher, Pfaffenberger & Pöschko, 2007). Die Angaben der Schulassistenten legen nahe, dass sie auf ihre Funktion in unterschiedlichem Umfang und auf Basis einer Vielfalt von Weiterbildungsmassnahmen vorbereitet waren, und zudem nur wenige Schulassistenten die empfohlene Weiterbildung für die Zusammenarbeit zwischen Lehrpersonen und Schulassistenten besuchten (vgl. auch Bildungsdirektion Kanton Zürich/Volksschulamt, 2018). In der Summe könnte dies dazu beigetragen haben, dass die Integration innerhalb der Schulen und die Akzeptanz der Schulassistenten bei den Lehrpersonen nicht immer vollständig gegeben waren (siehe unten).

Aus den Befragungen ergab sich weiterhin, dass das Ausmass, in dem Schulassistenten in den beiden Fächern eingesetzt wurden, in den Schulen beziehungsweise Klassen unterschiedlich war. Die Anzahl an Lektionen mit Schulassistentenz variierte zudem erheblich. Zum Zeitpunkt der Befragung (gegen Ende des Evaluationszeitraums) wurden einige Lehrpersonen von einer Schulassistentenz bei *allen* Unterrichtslektionen des Faches unterstützt; bei einem Grossteil der Lehrpersonen

war die Unterstützung im Unterricht dagegen auf einzelne Stunden beschränkt oder *gar nicht* gegeben. In sieben der neun ALLE-Schulen berichtete weniger als die Hälfte der befragten Lehrpersonen, von einer Schulassistenz unterstützt zu werden; in zwei dieser Schulen gab keine der Lehrpersonen an, über eine Schulassistenz zu verfügen. Damit übereinstimmend zeigt eine von den Schulen zusätzlich bereitgestellte Dokumentation, welche das Evaluationsteam von der ALLE-Projektleitung erhielt, dass in manchen Schulen relativ zur Anzahl der Lehrpersonen beziehungsweise Schülerinnen und Schüler deutlich mehr Schulassistenzen zur Verfügung standen als in anderen (wobei über die konkreten Pensen der Schulassistenzen keine Informationen vorlagen). Die Dokumentation deutet ausserdem darauf hin, dass in einigen Schulen noch im zweiten Jahr des Pilotprojekts die Schulassistenzen nicht oder nur mit geringen Kapazitäten zur Verfügung standen. In der Gesamtschau legen diese Befunde nahe, dass ein zentrales Instrument von ALLE – der Einsatz von Schulassistenzen, um den Lehrpersonen die Umsetzung eines lernwirksameren Unterrichts zu ermöglichen – nicht durchweg konsequent umgesetzt wurde. Im Fall mancher Schulen dürfte dies darauf zurückzuführen sein, dass die notwendigen finanziellen Mittel erst spät oder nicht in vollem Umfang zur Verfügung gestellt wurden (vgl. Kap. 5.1). Daneben könnten zumindest vereinzelt auch Vorbehalte innerhalb der Schulen gegenüber der Eignung von Schulassistenzen für die vorgesehenen Aufgaben von Bedeutung gewesen sein (siehe unten).

Eine beträchtliche Heterogenität zeigte sich darin, wie der Einsatz der Schulassistenzen in unterschiedlichen möglichen Handlungsfeldern gehandhabt wurde. Die Aussagen der Schulassistenzen weisen darauf hin, dass diese in beiden Fächern an verschiedenen Aktivitäten beteiligt waren, welche für eine gezielte individuelle Förderung der Schüler wichtig sind (z.B. Feststellung von Lernschwierigkeiten und Lernfortschritten). Dagegen waren die Assistenzen weniger in die gemeinsame Planung von Lernaktivitäten für Schülerinnen und Schüler eingebunden. Dies mag zunächst nicht überraschen, da die Schulassistenzen meist nicht über die notwendigen didaktischen Kompetenzen verfügen dürften. Andererseits wurden sie mehrheitlich eingesetzt, um einzelne Schülerinnen und Schüler beim Lösen von Aufgaben zu unterstützen (z.B. Fehler oder Missverständnisse besprechen; über Lösungswege diskutieren). Auch fällt auf, dass die häufigsten Einsatzgebiete der Schulassistenzen nach Angaben der Schulassistenzen und Lehrpersonen im Bereich der Unterstützung der leistungsschwächeren Schülerinnen und Schüler lagen, während die Arbeit mit Schülerinnen und Schülern ohne Lernschwierigkeiten vergleichsweise selten berichtet wurde.

Dies wirft die Frage auf, wie die Assistenzen auf ihre Tätigkeiten vorbereitet waren, wenn sie aufgrund ihrer Vorqualifikationen nicht unbedingt das notwendige professionelle Wissen mitbrachten. Zwar lassen es die Empfehlungen des Volksschulamts Zürich grundsätzlich zu, dass Schulassistenzen Lernprozesse begleiten oder Lernangebote methodisch-didaktisch gestalten (vgl. Bildungsdirektion Kanton Zürich/Volksschulamt, 2018). Dies muss jedoch angeleitet sein und kann nur in eng begrenztem Rahmen erfolgen: «Die Schulassistenz ist keine ausgebildete Fachperson und

kann deshalb im Handlungsfeld Unterricht nicht in professionell herausfordernden Situationen eingesetzt werden. Vielmehr unterstützt sie die Lehrperson darin, die Arbeitsfähigkeit der Klasse aufrechtzuerhalten» (ebd., S. 8). Die Grundidee der Massnahme Schulassistentenz besteht darin, die Lehrpersonen zu entlasten (als «Helping Hands»; Meier & Wolf, 2015), damit diese sich den pädagogischen Bedürfnissen der einzelnen – insbesondere leistungsschwächeren – Schülerinnen und Schülern widmen können. Dementgegen deuten Studien zur Unterstützung durch Schulassistentenzen darauf hin, dass diese von den Lehrpersonen häufig in der Arbeit mit Leistungsschwächeren eingesetzt werden, mit potenziell ungünstiger Wirkung auf die Qualität pädagogischer Handlungen und den Lernerfolg (Radford, Blatchford & Webster, 2011; Zumwald, 2014). Von einer Erhöhung der Kompetenzen der Schulassistentenzen (z.B. durch Weiterbildungen oder den intensiven Austausch mit der Lehrperson) lassen sich dabei zwar positive Effekte erwarten, doch bleibt das Potenzial des Instruments für die Bewältigung pädagogisch anspruchsvoller Situationen begrenzt, wenn hohe professionelle Standards gewahrt werden sollen (vgl. z.B. Zumwald, 2014, 2018). Die oben angesprochenen Tätigkeitsbereiche und die Äusserungen einzelner Schulassistentenzen aus dem Pilotprojekt können dementsprechend als Hinweis darauf interpretiert werden, dass die Schulassistentenzen zum Teil noch zielgerichteter eingesetzt hätten werden können, um den Lehrpersonen Freiräume für die Arbeit mit den Schülerinnen und Schülern zu verschaffen, die der individuellen Förderung besonders bedürfen.

Im Hinblick auf die Qualität der Zusammenarbeit zeigten sich sowohl die befragten Schulassistentenzen als auch Lehrpersonen in beiden Fächern überwiegend sehr zufrieden. Dies betrifft die Aufgabenteilung, die Klarheit der Zuständigkeiten und darüber hinaus – aus der Sicht der Schulassistentenzen – das Führungsverhalten der Lehrpersonen. Die berichteten Inhalte der Zusammenarbeit deuten dabei auf einen meist intensiven Austausch zwischen Schulassistentenzen und Lehrpersonen zu einem breiten Spektrum von Themen hin, das von disziplinarischen Problemen bis zu Lernschwierigkeiten der Schülerinnen und Schülern reicht. Verhaltenszeigten sich einige Schulassistentenzen und Lehrpersonen nur bezüglich der verfügbaren Zeitfenster zur Zusammenarbeit, insbesondere mehrere Deutsch-Lehrpersonen nahmen diese als unzureichend wahr.

Diesem insgesamt sehr positiven Eindruck stehen einige Aussagen aus den Interviews mit Schulleitungen und Lehrpersonen entgegen. Die Äusserungen der Lehrpersonen lassen dabei erkennen, dass das Verhalten einzelner Schulassistentenzen im Unterricht, vor allem aber seitens der Lehrpersonen wahrgenommene fehlende Kompetenzen der Schulassistentenzen zu negativen Einschätzungen und Unmut führten. So merkten Lehrpersonen an, dass eine Schulassistentenz wenig hilfreich ist, die «einen zusätzlichen Schüler» darstellt oder nicht die notwendige Professionalität für ihre Tätigkeit mitbringt. Dass Schulassistentenzen mit bestimmten Erfahrungen und Qualifikationen den Erwartungen der Lehrpersonen nicht entsprechen, geht auch aus dem Bericht einer Schule hervor, die für die ALLE-Projektleitung die Umsetzung der Massnahme so dokumentierte: «In der Schule

[...] stehen 1-2 Zivildienstleistende zur Verfügung. Diese werden jedoch von der Oberstufe wenig genutzt. Als Hindernis werden dabei der geringe Altersunterschied zu den Schülerinnen und Schülern der Oberstufe, sowie die fachlich anspruchsvollen Unterrichtsinhalte gesehen.»

Dabei ist zu beachten, dass eine Beurteilung von Kompetenzen sehr von den jeweiligen Tätigkeitsbereichen abhängt – und diese vermutlich nicht immer den vorgesehenen Aufgaben für Schulassistenten entsprachen (vgl. oben). Auch die unterschiedlichen Qualifikationen und Beschäftigungsbedingungen von Lehrpersonen und Schulassistenten könnten zu ungünstigen Einschätzungen einiger Lehrpersonen beigetragen haben. So dürften viele Lehrpersonen daran gewöhnt sein, mit anderen Fachpersonen (z.B. mit schulischen Heilpädagoginnen und Heilpädagogen, DaZ-Lehrpersonen) zusammenzuarbeiten, weniger mit Personen aus anderen beruflichen Sparten, die einen anderen Blick auf Schule und Unterricht mitbringen und andere professionelle Kompetenzen aufweisen. Dies mag zu unklaren oder zu hohen Erwartungen mancher Lehrpersonen und teilweise zu Enttäuschungen geführt haben. Manche Zitate deuten schliesslich darauf hin (vgl. oben, «einen zusätzlichen Schüler»), dass aus der Sicht mancher Lehrpersonen auch zuviel Aufwand mit dem Einsatz der Schulassistenten verbunden gewesen sein könnte. In dieser Zusammenarbeit übernimmt die Lehrperson eine Führungsrolle, mit der – zusätzlich zu ihren sonstigen Aufgaben – die Verantwortung für die Planung und Strukturierung des Einsatzes der Schulassistenten verbunden ist.

In der Gesamtschau legen diese Erwägungen nahe, dass bei einem Teil der Lehrpersonen eine noch bessere Vorbereitung auf die Zusammenarbeit mit einer Schulassistenten unterstützend gewesen wäre. Damit übereinstimmend empfehlen einschlägige Studien die gezielte Weiterbildung von Lehrpersonen, damit diese die Schulassistenten professionell anleiten und fördern können (vgl. Zumwald, 2015, 2018). Vor diesem Hintergrund ist es als Manko anzusehen, dass die im Pilotprojekt angebotene Weiterbildung zur Zusammenarbeit mit den Schulassistenten – die unter anderem auf die oben angesprochenen Aspekte abzielte (z.B. Rollenklärung; Tätigkeitsbereiche der Schulassistenten) – nur von wenigen Lehrpersonen besucht wurde.

Entgegen der kritischen Einschätzung mancher Lehrpersonen nahmen die Schülerinnen und Schüler in beiden Fächern, die durch die Schulassistenten im Unterricht unterstützt wurden, diese Hilfe in fast allen Fällen positiv wahr. Sie berichteten von wünschenswerten Veränderungen durch den Einsatz der Schulassistenten, am häufigsten ein besseres Lernen und eine stärkere individuelle Unterstützung durch die Lehrperson. Dies bestätigt – in Übereinstimmung mit dem Urteil der Mehrheit der Lehrpersonen –, dass die Schulassistenten ein wertvolles Instrument für die Förderung der Schülerinnen und Schüler sein kann. Ein für alle Beteiligten gewinnbringender Einsatz ist jedoch anspruchsvoll und konnte im Rahmen des Pilotprojekts nicht in allen Schulen erzielt werden.

5.2.2 Aufgabenhilfe

Eine weitere Massnahme im Pilotprojekt ALLE, mit der vor allem leistungsschwächere Schülerinnen und Schüler in Deutsch und Mathematik gezielt gefördert werden sollten, war die Aufgabenhilfe. Die Integration der Aufgabenhilfe in den bestehenden Stundenplan wurde von den Schulleitungen überwiegend als gelungen bewertet. Sie gaben an, die Umsetzung der Massnahme vor allem durch die Organisation von qualifiziertem Lehrpersonal und die Kommunikation der Wichtigkeit der Aufgabenhilfe an Lehrpersonen und Schülerinnen und Schüler zu unterstützen. Die von den Schulen selbst erstellte und von der ALLE-Projektleitung bereitgestellte Dokumentation zu dieser Massnahme deutet darauf hin, dass die Formen der Umsetzung der Aufgabenhilfe in den Schulen sehr unterschiedlich waren. Drei Beispiele sollen dies illustrieren:

- An einer Schule wurde ein *Lernraum* eingerichtet, der «als ruhige[r] Ort des Arbeitens im normalen Schulalltag» dienen sollte und täglich besucht werden konnte. Die Schülerinnen und Schüler schrieben sich – oft in Absprache mit der Lehrperson, oder auch nach einem Standortgespräch oder Elterngespräch mit entsprechender Zielvereinbarung – für je einen Monat für den Lernraum ein. Betreut wurde der Lernraum von der Schulassistentin.
- In einer weiteren Schule fand die Aufgabenhilfe im *Lernatelier* statt, in dem die «Schülerinnen und Schüler [...] bei Verständnisschwierigkeiten Fachberatungen in Anspruch nehmen» konnten. Das Lernatelier wurde von Lehrpersonen, Schulassistenten, aber auch von fachlich fortgeschrittenen Schülerinnen und Schülern in Form eines Peer-Tutoring angeboten. Geplant war (zum Zeitpunkt der Dokumentation) zudem ein wöchentlich angebotenes *Lernstudio*, das allen Schülerinnen und Schülern offen stehen sollte, um dort Hausaufgaben zu erledigen und/oder zusätzliches Fördermaterial zu bearbeiten.
- An einer Schule wurden wöchentlich *IL (Individuelles Lernen) Lektionen* angeboten, die je nach Lernstand der Schülerinnen und Schüler für bis zu sechs Lektionen verpflichtend waren. Ziel war es, in dieser Zeit nicht nur Aufgaben bearbeiten zu lassen, sondern eine Förderung mit «Vor- und Nachbereitung» sowie entsprechenden «Förderberichte[n]» zu realisieren. Die Verantwortung lag dabei offenbar bei den Lehrpersonen; über die Mittagsstunden wurde zudem eine DaZ-Lehrperson für die Aufgabenhilfe eingesetzt.

Gemäss der angesprochenen Dokumentation wurde die Aufgabenhilfe von allen Schulen umgesetzt. Demgegenüber gab nur etwa die Hälfte der Lehrpersonen in beiden Fächern an, dass an ihrer Schule das Angebot einer betreuten Aufgabenhilfe bestehe. Der Aussage stimmten nur in vier (Deutsch) beziehungsweise zwei (Mathematik) der Schulen *alle* befragten Lehrpersonen zu. In zwei (Deutsch) beziehungsweise vier (Mathematik) der Schulen stimmte dagegen *keine* Lehrperson zu, dass es eine Aufgabenhilfe gebe. Eine mögliche Erklärung ist, dass die Aufgabenhilfe nicht von allen Schulen über den gesamten Zeitraum des Pilotprojekts bis zur Befragung der Lehrper-

sonen (gegen Ende der Evaluation) umgesetzt wurde, die Massnahme zum Zeitpunkt der Befragung also bereits beendet war. Hinzu kommt, dass an den Schulen teilweise von «Aufgabenhilfe» abweichende Begriffe (vgl. oben) verwendet wurden, was zur ausbleibenden Zustimmung beziehungsweise zu widersprüchlichen Aussagen innerhalb der Schulen beigetragen haben könnte.

Dort, wo das Angebot einer Aufgabenhilfe zum Zeitpunkt der Befragung bestand, wurde es nach Aussagen sowohl der Schülerinnen und Schüler als auch der Lehrpersonen nur begrenzt in Anspruch genommen, im Fach Mathematik noch seltener als im Fach Deutsch. So gab etwa die Hälfte der Lehrpersonen beider Fächer an, dass keine ihrer Schülerinnen und Schüler die Aufgabenhilfe besuchen. Nach eigenen Angaben besuchte nur ein kleiner Anteil der Schülerinnen und Schüler die Aufgabenhilfe Deutsch oder Mathematik. Diese geringe Nutzung könnte darin begründet sein, dass die Teilnahme an der Aufgabenhilfe zum Teil als freiwilliges Angebot konzipiert – oder zumindest so aufgefasst und umgesetzt – wurde. Die Dokumentation durch die Schulen weist darauf hin, dass der Besuch der Aufgabenhilfe in einigen Schulen tatsächlich grundsätzlich fakultativ war (z.B. ein fachlich betreuter «Hausaufgabenclub»), in anderen Schulen zumindest für bestimmte Schülerinnen- und Schülergruppen obligatorisch vorgesehen wurde (z.B. bei Förderbedarf in Mathematik). Dem entsprechend sah die Mehrheit der befragten Lehrpersonen (je Fach ca. 60 Prozent) das Angebot als vollständig freiwillig an, fast alle weiteren Lehrpersonen beurteilten das Angebot zwar als verpflichtend, aber nur für bestimmte – insbesondere leistungsschwächere – Schülerinnen und Schüler. Dass entsprechende Angebote in der Praxis nur begrenzt besucht wurden, könnte gemäss den Äusserungen von mehreren Lehrpersonen in den Interviews auch durch das mangelnde Interesse oder Engagement der leistungsschwächeren Schülerinnen und Schüler bedingt gewesen sein. Diesen wurde die Nutzung des Angebots zwar nahegelegt oder sie wurde mit ihnen vereinbart, aber zum Teil erfolgte sie nicht oder nicht regelmässig. Ob in solchen Fällen Sanktionsmöglichkeiten fehlten (wie die Äusserung einer Lehrperson im Interview nahelegt) oder auf deren Einsatz bewusst verzichtet wurde, lässt sich nicht feststellen. Auch hier muss das Ergebnis der Befragung jedoch mit einer gewissen Vorsicht interpretiert werden: Zu vermuten ist, dass durch die Verwendung alternativer Bezeichnungen für die Aufgabenhilfe an den Schulen nicht alle teilnehmenden Schülerinnen und Schüler in der Befragung erfasst wurden. Es ist also zu vermuten, dass die Nutzung der Massnahme anhand der Befragungsdaten eher unterschätzt wird.

Für die Einsatzgebiete der Aufgabenhilfe zeigte sich ein recht heterogenes Bild. Am häufigsten wurde sie in beiden Fächern zur Erledigung von Hausaufgaben genutzt. Nach Auskunft der Deutsch-Lehrpersonen ebenfalls relativ häufig war der Einsatz für die Wiederholung von Unterrichtsstoff und das Begleiten leistungsschwächerer Schülerinnen und Schüler. Für alle anderen Einsatzgebiete variierte die berichtete Nutzung sehr deutlich zwischen den befragten Lehrpersonen. Durchgeführt wurde die Aufgabenhilfe am häufigsten durch Lehrpersonen, gefolgt von schulischen Heilpädagoginnen und Heilpädagogen. Nach eigenen Angaben waren zudem knapp zwei

Drittel der befragten Schulassistenzen in der Aufgabenhilfe tätig. Dass auch Schulassistenzen das Angebot betreuten, kann als Widerspruch zur Forderung nach professioneller Betreuung der Aufgabenhilfe angesehen werden, zumindest, wenn die Tätigkeit nicht vor allem unterstützend erfolgte. Gemäss der zusätzlichen Dokumentation zur Massnahme, welche die Schulen anfertigten und von der Projektleitung zur Verfügung gestellt wurde, wurden die Schulassistenzen in vier der neun ALLE-Schulen explizit für die Aufgabenhilfe eingeplant, allerdings mit nur einer Ausnahme immer im Verbund mit Lehrpersonen und gegebenenfalls schulischen Heilpädagoginnen beziehungsweise Heilpädagogen (von zwei der Schulen fehlten entsprechende Angaben).

Die teilnehmenden Schülerinnen und Schüler nahmen die Aufgabenhilfe überwiegend positiv oder sogar sehr positiv wahr, im Fach Mathematik noch etwas mehr als im Fach Deutsch. Relativ häufig berichtete günstige Veränderungen durch die Aufgabenhilfe waren das bessere Folgenkönnen im Unterricht sowie verbesserte Leistungen und ein höheres Selbstvertrauen der Schülerinnen und Schüler. Auch Lehrpersonen, deren Schülerinnen und Schüler die Aufgabenhilfe nutzten, bewerteten das Angebot überwiegend positiv als eine lernförderliche Unterstützung. So stimmten die Lehrpersonen in beiden Fächern zu einem Grossteil zu, dass der Einsatz der Aufgabenhilfe zu einer Stärkung der Kompetenzen der Schülerinnen und Schüler geführt habe.

Die Aussagen vor allem der Schülerinnen und Schüler, aber auch vieler Lehrpersonen, unterstreichen das Potenzial des Angebots der Aufgabenhilfe als individuelle Lernunterstützung. Dass die Schulleitungen die Integration der Aufgabenhilfe überwiegend positiv bewerteten, deutet darauf hin, dass die den Schulen gewährte Freiheit bei der Umsetzung der Massnahme eine Voraussetzung für ihren Erfolg war. Damit übereinstimmend zeigt die Dokumentation zur Massnahme durch die Schulen, dass deren Ausgestaltung in Bezug auf den formalen Rahmen, die Inhalte oder die Betreuung durch das schulische Personal sehr unterschiedlich erfolgte. Dabei deutet sich zumindest für einige Schulen ein Bruch zwischen der Konzeption der Massnahme durch die ALLE-Projektleitung und deren Umsetzung in der Praxis an. So lässt sich der an mehreren Schulen praktizierte Grad an Freiwilligkeit des Besuchs der Aufgabenhilfe kaum mit der empfohlenen Verbindlichkeit dieser Massnahme in Deckung bringen (vgl. Bildungsdirektion Kanton Zürich/Volksschulamt, 2016). Die berichtete geringe Nutzung des Angebots könnte zumindest teilweise darauf zurückzuführen sein beziehungsweise auf eine Unklarheit darüber, für wen, in welchem Umfang und mit welcher Verbindlichkeit die Aufgabenhilfe angeboten werden sollte. Daneben könnte eine Rolle gespielt haben, dass die Schulen bezüglich der verfügbaren Finanzmittel von den Schulbehörden abhängig waren (vgl. Kap. 5.1 und Kap. 5.2.1). In manchen Schulen könnten begrenzte finanzielle Mittel einer noch breiteren Implementation der Massnahme entgegengestanden haben. Einschränkend ist schliesslich nochmals darauf hinzuweisen, dass unterschiedliche verwendete Begrifflichkeiten an den Schulen – abweichend vom Begriff der «Aufgabenhilfe» – die Aussagekraft insbesondere der Schülerangaben beeinträchtigt haben könnten.

5.2.3 Weiterbildungen der Lehrpersonen

Eine weitere wichtige Massnahme, vor allem zu Beginn des Pilotprojekts ALLE, war die Durchführung von Weiterbildungen für die Lehrpersonen, welche teils verpflichtend, teils freiwillig waren. Die fachspezifischen Weiterbildungen sowie die Weiterbildung zur Einführung in EMU wurden von der deutlichen Mehrheit der Lehrpersonen besucht. An der Weiterbildung zum Lerncoaching nahm nach eigenen Angaben demgegenüber nur etwa die Hälfte, an jener zur Kooperation zwischen Lehrpersonen und Schulleistungen nur eine Minderheit der Lehrpersonen teil. Eine Teilnahme an den beiden letztgenannten Weiterbildungen war zwar empfohlen worden, diese war jedoch nicht obligatorisch wie bei den anderen Veranstaltungen. Die geringere berichtete Partizipation dürfte sich entsprechend teilweise durch die Freiwilligkeit des Angebots erklären. Daneben ist davon auszugehen, dass die tatsächliche Teilnahme der Lehrpersonen aufgrund des späten Befragungszeitpunkts (gegen Ende der Evaluation) eher unterschätzt wird. So übernahm eine erhebliche Anzahl Lehrpersonen erst im Lauf des Pilotprojekts eine ALLE-Klasse und löste die bis dahin unterrichtenden Fachlehrpersonen ab. Naheliegender ist demnach, dass ein Anteil der abgelösten Lehrpersonen die Weiterbildungen besucht hatte, aber nicht mehr befragt wurde, während ein Anteil der neuen Lehrpersonen – zumindest in der späteren Phase des Pilotprojekts – nicht mehr teilnehmen konnte oder gegebenenfalls eine Teilnahme für nicht mehr notwendig erachtete.

Die Weiterbildungen wurden von den teilnehmenden Lehrpersonen fast durchweg als «eher nützlich» oder «nützlich» beurteilt. Dementsprechend berichteten die Lehrpersonen in beiden Fächern tendenziell positive Wirkungen des Besuchs der Weiterbildungen. Eine Steigerung der eigenen diagnostischen Kompetenzen und der Fähigkeit zur gezielten Förderung leistungsschwächerer Schülerinnen und Schüler im Hinblick auf die Grundkompetenzen, eine Sensibilisierung für besondere Schwierigkeiten der Schülerinnen und Schüler und eine Verbesserung des didaktischen Wissens sowie der Passung des eigenen Unterrichts zu den Bedürfnissen der Schülerinnen und Schüler nahmen je Fach rund die Hälfte bis zwei Drittel der Lehrpersonen wahr. Zudem gaben in beiden Fächern rund vier Fünftel der Lehrpersonen, die Weiterbildungen besucht hatten, an, im Hinblick auf mindestens einen der erfragten Aspekte von den Weiterbildungen profitiert zu haben. Ein nochmals positiveres Urteil wurde von den Schulleitungen abgegeben, welche die Angebote allesamt als eher nützlich bis sehr nützlich einschätzten. Dennoch ist festzustellen, dass nach eigenem Empfinden nicht alle Lehrpersonen von einer Teilnahme an der jeweiligen Weiterbildung profitieren konnten. Entsprechend resultierte auch keine breite Zufriedenheit der Teilnehmenden mit den Weiterbildungen – in beiden Fächern äusserte sich nur jeweils etwa die Hälfte aller befragten Lehrpersonen als «eher zufrieden» oder «sehr zufrieden».

Eine mögliche Erklärung für die ausbleibende Zufriedenheit mancher Lehrpersonen findet sich in ihren Aussagen in den Gruppeninterviews, die auf eine geringe Nachhaltigkeit der Weiterbildungen im weiteren Projektverlauf aufgrund fehlender Rückmeldungen und nachfolgender Impulse und

Aufträge verweisen. Zudem wird deutlich, dass Anschlussangebote – insbesondere die nachfolgend diskutierten Beratungen – zumindest nicht von allen Lehrpersonen als solche verstanden wurden. Für eine allfällige Anpassung der Angebote ergeben sich eine Reihe von Hinweisen aus den Interviews mit den Weiterbildungspersonen (vgl. Kap. 4.4.4). Die von den Lehrpersonen genannten Einschränkungen aufgreifend sollen hier vor allem zwei Punkte erwähnt werden:

- eine Bereitstellung zusätzlicher Ressourcen für *mehr und fortlaufende Weiterbildungsveranstaltungen*, um eine kontinuierliche Arbeit insbesondere auch an anspruchsvollen Themen sicherzustellen;
- ein Angebot mit *individuellen Coachings* anstelle von Gruppen-Weiterbildungen, um besser auf die Bedürfnisse der einzelnen Lehrpersonen eingehen zu können.

5.2.4 Beratungsangebote für die Lehrpersonen

Die am Pilotprojekt ALLE teilnehmenden Lehrpersonen hatten die Möglichkeit, fachbezogene Beratungsangebote durch Fachdidaktikerinnen und Fachdidaktiker in Anspruch zu nehmen. Die Aussagen der Lehrpersonen deuten darauf hin, dass diese Möglichkeit nur selten wahrgenommen wurde – im Fach Deutsch gaben vier Lehrpersonen, im Fach Mathematik lediglich zwei Lehrpersonen an, ein Beratungsangebot genutzt zu haben. Positiv fällt dabei auf, dass zwei Lehrpersonen im Fach Deutsch berichteten, das Angebot mehrfach in Anspruch genommen zu haben. Anzunehmen ist allerdings, dass die Teilnahmhäufigkeit auf Basis der Lehrpersonenangaben eher unterschätzt wird. Gemäss der Dokumentation der Beratungsangebote durch die ALLE-Projektleitung, die dem Evaluationsteam zur Verfügung gestellt wurde, fanden schulinterne Fachberatungen des Mathematikteams, schulinterne Fachberatungen des Deutschteams und schulinterne Prozessbegleitungen/Coachings je mindestens dreimal statt. Dies legt eine höhere Teilnehmendenzahl nahe, auch wenn nicht angenommen werden kann, dass jeweils alle Lehrpersonen aus den Klassen der ALLE-Schulen die Angebote besucht haben. Dass sehr wenige Lehrpersonen berichteten, Beratungsangebote genutzt zu haben, dürfte zumindest teilweise auf Lehrpersonenwechsel im Projektverlauf und eine begrenzte Teilnahme an der Befragung zurückgehen.

Als ein möglicher Grund für die geringe Nutzung der Beratungsangebote ergab sich aus den Interviews mit den Lehrpersonen und den Beratungspersonen die fehlenden zeitlichen Ressourcen. Dass Lehrpersonen nach eigener Einschätzung nicht genügend Zeit für die Nutzung des Angebots zur Verfügung stand, ist jedoch nicht speziell auf die Beratungen zu beziehen, sondern deutet eher auf eine hohe Beanspruchung im Beruf hin, die ein bedeutsamer Anteil der Lehrpersonen in der Schweiz berichtet (Brägger, 2019; Sandmeier, Kunz Heim, Windlin & Krause, 2017). Zu vermuten ist demnach, dass viele Lehrpersonen generell einen Zeitmangel wahrnehmen, der es im Kontext des Pilotprojekts erschwerte, die für ALLE vorgesehenen Massnahmen noch neben den Anforde-

rungen des alltäglichen Unterrichtens vollumfänglich realisieren zu können. Eine solche Wahrnehmung könnte sich vor allem bei denjenigen Massnahmen negativ auf das Nutzungsverhalten ausgewirkt haben, die für die alltagsnahe Durchführung des Unterrichts keine unmittelbaren Implikationen besitzen, sondern eher indirekter zu einer Verbesserung des Unterrichts beitragen können. Weitere Überlegungen zur geringen Inanspruchnahme des Angebots steuerten die Beratungspersonen in den Interviews bei. Angesprochen wurde eine möglicherweise fehlende Aktualität oder Dringlichkeit von Themen im Rahmen des täglichen Unterrichtens, die zu einer entsprechend geringen Nachfrage geführt haben könnte. Auch eine zu geringe Vertrautheit mit dem Angebot und eine nicht ausreichende Passung zu den Bedürfnissen der Lehrpersonen lassen sich den Äusserungen als mögliche Gründe entnehmen. Schliesslich wird auch eine fehlende Bereitschaft, den Unterricht zu öffnen, und somit auch allfällige Schwächen des eigenen Unterrichtens für andere sichtbar werden zu lassen, als mögliche Ursache des eher geringen Interesses angeführt.

Für eine mögliche Weiterentwicklung der Angebote lassen sich den Interviews mit den Beratungspersonen eine Reihe von Optimierungsvorschlägen entnehmen (vgl. Kap. 4.5.3). Besonders zwei Thematiken fallen dabei ins Auge:

- ein *vermehrter Einbezug der Lehrpersonen* durch zusätzliche Informationen über die Beratungen und durch eine gemeinsame Planung von Inhalten und Aktivitäten;
- mehr *kontinuierliche Begleitung* der Schulen, beispielsweise durch ein kombiniertes Angebot von Weiterbildungen und Beratung.

Der letztgenannte Punkt lässt sich in Verbindung mit Äusserungen der Lehrpersonen zu den Weiterbildungen bringen, in denen eine fehlende Nachhaltigkeit beziehungsweise ausbleibende Anschlussangebote zur Sprache gebracht wurden (vgl. Kap. 5.2.3).

Befragt nach ihrer allgemeinen Zufriedenheit mit der Umsetzung der Beratungen, äusserte sich eine knappe Mehrheit aller befragten Lehrpersonen als «weder zufrieden noch unzufrieden», die übrigen Urteile waren stark polarisiert, insbesondere im Fach Mathematik. In dieses Urteil gingen jedoch auch Einschätzungen vieler Lehrpersonen ein, die nach eigener Aussage keine der Beratungen besuchten. Entsprechend können diese Einschätzungen nicht die Durchführung der Beratungen selbst betreffen, sondern höchstens andere Gesichtspunkte, zum Beispiel die Rahmenbedingungen des Angebots, widerspiegeln. Wichtiger ist das Urteil der (wenigen) Lehrpersonen, die angeben, eine Beratung in Anspruch genommen zu haben – diese äusserten sich je zur Hälfte als «eher zufrieden» oder «weder zufrieden noch unzufrieden». Dieses tendenziell positive Urteil bestätigt sich bei den Schulleitungen, welche die Durchführung der Beratungen überwiegend als «gut» oder «sehr gut» beurteilten.

5.2.5 EMU

Eine weitere zentrale Massnahme im Pilotprojekt war die Implementation der «Evidenzbasierten Methoden der Unterrichtsdiagnostik und -entwicklung» (EMU) an den ALLE-Schulen. Betrachtet man zunächst die Rahmenbedingungen der Umsetzung von EMU, so deuten die Aussagen aus den Interviews darauf hin, dass die Schulleitungen zumindest anteilig besondere Herausforderungen für einen gewinnbringenden Einsatz von EMU (z.B. einen erhöhten Aufwand) wahrnahmen (vgl. Kap. 4.6.1). Dies mag auch dazu beigetragen haben, dass sie die Umsetzung an ihrer Schule im Vergleich mit den anderen Massnahmen etwas zurückhaltender – am häufigsten als «befriedigend» – beurteilten. Im gleichen Zusammenhang fällt auf, dass sich die Schulleitungen bezüglich der von ihnen geleisteten Unterstützung für EMU tendenziell verhalten äusserten. Zwar gaben die Schulleitungen fast durchgehend an, dass EMU in das Schulprogramm aufgenommen worden sei. Die weiteren erfragten Aktivitäten (z.B. die Organisation von Vertretungslösungen bei kollegialen Hospitationen) wurden dagegen höchstens von der Hälfte der Schulleitungen berichtet. Dabei mag eine Rolle gespielt haben, dass diese noch mehr konkretes und praktisches – und damit zeitintensives – Engagement für die Umsetzung von EMU erfordert hätten. Diese Interpretation legt auch die Auskunft der ALLE-Projektleitung nahe, wonach die Schulleitungen regelmässig im Verlauf des Pilotprojekts dazu angehalten werden mussten, die Durchführung von EMU sicherzustellen.

Von einer Umsetzung der im Rahmen von EMU vorgesehenen kollegialen Hospitationen und dem Einholen von Feedback zum eigenen Unterricht – seitens der hospitierenden Kolleginnen und Kollegen sowie den Schülerinnen und Schülern der eigenen Klasse oder Lerngruppe – berichtet die Mehrheit der befragten Lehrpersonen überwiegend zwei- bis dreimal (beim Schülerfeedback bis zu viermal). Summiert man je Lehrperson über die drei Bestandteile von EMU auf, zeigt sich, dass EMU-Elemente mehrheitlich sechs- bis neunmal eingesetzt wurden, in einzelnen Fällen auch deutlich häufiger. Da die Fragen auf die zurückliegenden zwei Jahre bezogen waren, bedeutet dies, dass die Lehrpersonen im Mittel etwa drei- bis fünfmal jährlich Elemente von EMU nutzten. Nur wenige Lehrpersonen gaben dabei an, einzelne von EMU vorgesehene Methoden gar nicht eingesetzt zu haben. Demnach erhielten im betreffenden Zeitraum die meisten Lehrpersonen mehrfach Feedback zur Qualität des eigenen Unterrichts – und damit auf empirischer Evidenz basierende Hinweise zu der Entwicklung ihres Unterrichts und den Erfolgen der von ihnen umgesetzten unterrichtsbezogenen Massnahmen. Dennoch entsprach die Umsetzung damit nicht durchweg den Vorgaben der ALLE-Projektleitung, die einen Einsatz von EMU zumindest zweimal pro Jahr vorsahen. Denn etwa jede siebte Lehrperson (13.0 Prozent der Befragten) gab an, im Mittel weniger als zweimal jährlich einen der Bestandteile von EMU genutzt zu haben.

Die Befragung der Schulleitungen deutet zwar darauf hin, dass ein weiterer Erfahrungsaustausch zu EMU auf der Ebene der Jahrgangsguppen, Fachgruppen oder auf Schulebene eher sporadisch blieb. Einen kollegialen Austausch zwischen einzelnen Lehrpersonen zu EMU nahmen jedoch fast

alle befragten Schulleitungen oft oder zumindest gelegentlich wahr. Hinweise auf die möglichen Gründe einer nicht nochmals intensiveren Nutzung und Auseinandersetzung mit EMU geben die Aussagen von Schulleitungen im Interview, die eine mögliche «Überforderung» bei häufigerem Einsatz und den «zusätzliche[n] Aufwand» hervorheben. Daneben wird von einer Schulleitung auch der «Widerstand» seitens der Lehrpersonen angesprochen, auf Kritik am Unterricht (in diesem Fall der Schülerinnen und Schüler und ihrer Eltern) einzugehen. Darin könnte eine noch fehlende Bereitschaft mancher Lehrpersonen vermutet werden, den eigenen Unterricht dem Urteil anderer zu öffnen, der ähnlich auch von Beratungspersonen im Pilotprojekt angesprochen wurde (vgl. Kap. 5.2.4).

Mit Blick auf positive Wirkungen von EMU betonten die Lehrpersonen vor allem die Möglichkeit, Hinweise auf gezielte Massnahmen zur Unterrichtsverbesserung zu erhalten. Hervorgehoben wurde zudem die Chance, mithilfe von EMU auf «blinde Flecken» in der Wahrnehmung des eigenen Unterrichts aufmerksam zu werden und eigene Stärken und Schwächen besser erkennen zu können. Mehrere Lehrpersonen äusserten sich in diesem Zusammenhang sehr positiv über die durchgeführten Unterrichtshospitationen. Aus den Interviewaussagen der Lehrpersonen wird jedoch auch deutlich, dass der wahrgenommene Nutzen des kollegialen Feedbacks als abhängig von den jeweils beteiligten Personen und den Kontextbedingungen angesehen wurde. So beschrieb beispielsweise eine der Lehrpersonen den Besuch des Unterrichts der eigenen Klasse in einem anderen Fach als besonders anregend; eine andere Lehrperson dagegen bezweifelte, von den vorgesehenen stufenübergreifenden Unterrichtsbesuchen profitieren zu können (siehe dazu auch Kreis, 2015). Gemäss den Befragungsergebnissen noch mehr geschätzt als die Rückmeldungen der Kolleginnen und Kollegen wurden die Feedbacks der Schülerinnen und Schüler. Dass auch hier die Art und Weise des Einsatzes der Massnahme wichtig für den wahrgenommenen Nutzen gewesen sein dürfte, deutet die zitierte Äusserung einer Lehrperson an (vgl. Kap. 4.6.1: «[...] je öfter man es macht desto mehr finden sie es ... [...] Wir haben uns nur noch durchgeklickt, es ist immer das Gleiche, es ist nicht spannend.» Dies könnte darauf hinweisen, dass das Instrument vereinzelt relativ mechanisch, möglicherweise eher in zu hoher Frequenz statt in der notwendigen Tiefe, genutzt wurde – und Bedarf bestehen könnte, die Voraussetzungen einer gewinnbringenden Nutzung zukünftig noch mehr zu betonen beziehungsweise diese noch konsequenter gegenüber allen Lehrpersonen zu kommunizieren.

Die Interviewaussagen mehrerer Lehrpersonen deuten darauf hin, dass das Einholen von Schülerfeedback für sie keine neue Massnahme darstellte, sondern in Eigeninitiative oder durch Regelungen des Kollegiums/der Schule schon regelmässig vor dem Pilotprojekt stattfand. Auch Unterrichtshospitationen waren gemäss den Aussagen in den Interviews an einigen Schulen bereits vor ALLE fest etabliert. Wie die Umsetzung der Massnahme im Rahmen von EMU auch in diesem Fall einen Gewinn darstellen konnte, wird aus der zitierten Aussage einer Schulleitung deutlich: «Wir

haben gesagt, ihr lasst Unterrichtsqualitätsmerkmale von EMU bewusst in die Unterrichtshospitation einfließen. Das heisst, ihr könnt nicht einfach allgemein Rückmeldung geben, sondern ihr müsst von EMU her Unterrichtsqualitätsmerkmale aussuchen und im Hinblick auf diese Merkmale diese Hospitation abwickeln.» Ein Nutzen von EMU ist demnach nicht nur in der blossen Einführung oder regelmässigen Durchführung der Hospitationen zu sehen, sondern auch in einer stärkeren Fokussierung auf bedeutsame Aspekte von Unterrichtsqualität. Die Äusserung einer weiteren Schulleitung deutet an, wie EMU auch bei Lehrpersonen, die bereits vorher Hospitationen durchführten, zu einem Einstellungswandel führen konnte: «[...] dort hat EMU insofern Qualitätsveränderungen gebracht, weil es eigentlich das Erforschen des Unterrichtes in den Vordergrund gestellt hat und nicht das Kritisieren.» Da die Schaffung einer vertrauenswürdigem Umgebung eine zentrale Bedingung für die Akzeptanz von kollegialen Hospitationen darstellt (vgl. Kreis, 2015), kann auch diese veränderte Feedbackkultur als positive Wirkung des Einsatzes von EMU gesehen werden.

5.2.6 Individualisierung im Fachunterricht

Eine Intention des Pilotprojekts ALLE war es, die Passung des Unterrichts zu den individuellen Lernvoraussetzungen zu steigern, und so insbesondere die leistungsschwächeren Schülerinnen und Schüler verstärkt zu fördern. Die Lehrpersonen, Schulleitungen und Schulassistenten an den ALLE-Schulen wurden dazu befragt, wie sie die Wirkung des Pilotprojekts auf die Individualisierung im Unterricht einschätzen. Für verschiedene erfragte Aspekte gaben jeweils rund zwei Drittel der Deutsch-Lehrpersonen und die Hälfte der Mathematik-Lehrpersonen an, dass die Massnahmen im Pilotprojekt zu einer positiven Veränderung geführt hätten. Wahrgenommen wurden von diesen Lehrpersonen einerseits eine bessere Lernzeitnutzung und Lernwirksamkeit des Unterrichts, andererseits eine gesteigerte Lernmotivation und mehr Selbstvertrauen bei leistungsschwächeren Schülerinnen und Schülern.

Dieser positive Eindruck zumindest grösserer Anteile der Lehrpersonen wird gestützt durch die Schulleitungen der ALLE-Schulen, die ausnahmslos einer Veränderung in Richtung einer höheren Lernwirksamkeit des Unterrichts und einer Optimierung der Lernzeit für leistungsschwächere Schülerinnen und Schüler zustimmten. Sie gaben zudem an, dass das Pilotprojekt zu einer Zunahme des Wissens der Lehrpersonen über die Möglichkeiten zur Lernförderung geführt habe. Wie positive professionelle Erfahrungen durch das Pilotprojekt gefördert wurden, illustriert die Aussage einer Schulleitung, die von Lehrpersonen berichtet, «die sich überhaupt nicht vorstellen konnten, dass Niveaudurchmischung funktionieren kann [Sie] haben sich freiwillig dazu bereit erklärt eine A-B-Klasse zu führen, die [...] eigentlich eine A-B-C-Klasse ist und sind hellauf begeistert» (vgl. Kap. 4.7). Ein verändertes Verständnis für die Förderung leistungsschwächerer Schülerinnen und Schüler wird exemplarisch auch in der Aussage dieser Lehrperson erkennbar: «Und dann hast du eben die, die halt eben sonst normalerweise hinten runterfallen, auf die ist natürlich schon der

Fokus zu richten [...]. Und das braucht Ressourcen und es braucht das Know-how [...]. Und das ist schon auch durch ALLE noch mehr pointiert worden, aufgedeckt worden [...].» (vgl. ebd.)

Befragt, wer von den Massnahmen des Pilotprojekts profitiert habe, berichteten die Schulleitungen und Schulassistenzen übereinstimmend, dass ALLE zu einer besseren Förderung der leistungsschwächeren Schülerinnen und Schüler geführt habe. Während nach Einschätzung der meisten befragten Schulassistenzen auch die leistungsstärkeren Schülerinnen und Schüler von einer besseren Förderung in ALLE profitiert haben, geht davon nur etwa die Hälfte der Schulleitungen aus. Aus den Interviews wird ein besonderes Bemühen, auch die leistungsstärkeren Schülerinnen und Schüler zu fördern, zumindest punktuell erkennbar. So berichtete beispielsweise eine Schulleitung von dem Vorhaben der Deutsch-Lehrpersonen, zielgerichtete Lernstandserhebungen durchzuführen, um «auch etwas für die Stärkeren anzubieten».

In den Interviews mit den Lehrpersonen und Schulleitungen der ALLE-Schulen wurde ein grösseres Spektrum an Massnahmen angesprochen, um den Unterricht besser auf die individuellen Bedürfnisse der Schülerinnen und Schüler abzustimmen. Diese umfassten beispielsweise eine Differenzierung und Flexibilisierung beim Einsatz von Lehrmitteln und Lernmaterialien; ein gemeinsames Unterrichten von Schülerinnen und Schülern aus unterschiedlichen Abteilungen, verknüpft Möglichkeiten zum individuellen Lernen; die Einrichtung zusätzlicher Lektionen und Gefässe zur Förderung und Vertiefung; und die individuelle Anpassung der Lernziele, verbunden mit der Entwicklung passender Lernkontrollen. Für die leistungsschwächeren Schülerinnen und Schüler berichteten Lehrpersonen, mehr als bislang auf den Erwerb von Grundkompetenzen zu fokussieren. Dieser «Mut zur Lücke» beziehungsweise das «weniger ist mehr» – wie dies Lehrpersonen in den Interviews formulierten – deutet darauf hin, dass von den Lehrpersonen das Aufgabenmerk mehr auf die Lernintensität und Verstehenstiefe für ausgewählte, für das weitere Lernen zentrale Gegenstände gelenkt wurde, anstatt viele verschiedene Inhalte eher oberflächlich abzuarbeiten.

5.3 Wirkungen des Pilotprojekts ALLE

Ziel der Evaluation war es zu untersuchen, ob mit dem an den ALLE-Schulen umgesetzten Massnahmenbündel positive Wirkungen auf den Unterricht und die Lernergebnisse der Schülerinnen und Schüler erzielt werden konnten. Im Folgenden werden die Befunde zur Wirkung von ALLE auf die professionellen Überzeugungen und Kompetenzen der Lehrpersonen (Kap. 5.3.1), die Qualität des Deutsch- und Mathematikunterrichts (Kap. 5.3.2), die motivational-affektiven Lernergebnisse der Schülerinnen und Schüler (Kap. 5.3.3) und die fachlichen Lernergebnisse der Schülerinnen und Schüler (Kap. 5.3.4) zusammengefasst und diskutiert.

5.3.1 Professionelle Kompetenzen der Lehrpersonen

Verschiedene Massnahmen im Pilotprojekt ALLE – wie die angebotenen Weiterbildungen und Beratungen – zielten auch auf eine Verbesserung der professionellen Kompetenzen der teilnehmenden Lehrpersonen ab. In der Evaluation ALLE wurde die Entwicklung dieser Kompetenzen im Projektverlauf im Hinblick auf zwei Aspekte professioneller Handlungskompetenz (vgl. Baumert & Kunter, 2006) in den Blick genommen, die Überzeugungen der Lehrpersonen zum Lehren und Lernen sowie die Urteilsakkuratheit der Lehrpersonen.

Vergleichbar mit den Befunden anderer Untersuchungen für die Deutschschweiz ergab sich insgesamt für die Lehrpersonen der ALLE- und Vergleichsschulen eine deutlich höhere Zustimmung zu einer konstruktionsorientierten gegenüber einer transmissionsorientierten Perspektive auf das Lernen der Schülerinnen und Schüler (vgl. Biedermann, Brühwiler, Oser, Affolter & Bach, 2015). Damit stimmten die Lehrpersonen relativ stark zu, dass Wissen seitens der Schülerinnen und Schüler durch aktive und individuelle Konstruktionsprozesse aufgebaut werden muss. Weniger Zustimmung fand dagegen die Auffassung, dass Wissen durch einen gerichteten Vermittlungsprozess von der Lehrperson an die Schülerinnen und Schüler weitergegeben wird. Betrachtet man die Entwicklung im Verlauf des Projekts ALLE, zeigte sich im Fach Deutsch eine tendenziell abnehmende Zustimmung zur Transmissionsorientierung, allerdings beschränkt auf die Vergleichsschulen. In den ALLE-Schulen blieb der Grad der Zustimmung praktisch unverändert, dieser war aber bereits zum ersten Messzeitpunkt eher niedrig ausgeprägt. Demgegenüber wurde der konstruktionsorientierten Perspektive sowohl in den ALLE- als auch den Vergleichsschulen vergleichsweise stark zugestimmt und nahm nur in den Vergleichsschulen noch leicht zu.

Bei der Interpretation dieser Ergebnisse ist zu beachten, dass die Beziehung zwischen beiden Orientierungen von Lehrpersonen und dem Lernen der Schülerinnen und Schüler komplex ist. Weder eine hohe Transmissions- noch eine hohe Konstruktionsorientierung sind per se lernförderlich oder dem Lernen abträglich. Tatsächlich kann auch eine etwas stärkere Transmissionsorientierung – die in der Vergangenheit teils pauschal abgelehnt wurde, da sie einem vermeintlich zu sehr auf die Lehrperson zentrierten Unterrichtsverständnis entspricht – in bestimmten Anforderungssituationen zu besserem Unterricht beitragen. Entscheidend ist die angemessene Abstimmung beider Perspektiven, um eine situationsadäquate Auswahl von schüler- und lehrpersonen-zentrierten Unterrichtsformen zu erzielen (Biedermann, Steinmann & Oser, 2015). Ein klar strukturierter – eher direkter – Unterricht kann beispielsweise für leistungsschwächere Schülerinnen und Schüler eher lernförderlich sein, während leistungsstärkere Schülerinnen und Schüler von vergleichsweise offenen Unterrichtsformen profitieren können (Helmke & Weinert, 1997). Als problematisch wäre vor diesem Hintergrund vor allem eine extreme Ausprägung einer der beiden Sichtweisen bei den Lehrpersonen zu bewerten, die allerdings bereits zu Beginn des Pilotprojekts weder in den ALLE-Schulen noch in den Vergleichsschulen vorzufinden war.

Die Zuschreibung von Leistungen auf die Begabung beziehungsweise Begabungsdefizite der Schülerinnen und Schüler wurde von den Lehrpersonen in beiden Fächern tendenziell abgelehnt. Auch wenn es eher unrealistisch erschiene, für schulische Leistungen gar keinen Einfluss gewisser Begabungen anzunehmen, gilt eine deutlich ausgeprägte Zustimmung zu dieser Zuschreibung als ungünstig für den Umgang mit leistungsschwächeren Schülerinnen und Schülern. Werden die Leistungen dieser Schülerinnen und Schüler auf fehlende Begabung zurückgeführt – und damit auf eine Ursache, die der Kontrolle der Lehrperson entzogen ist – liegt nahe, dass die Lehrperson weniger zur Förderung dieser Schülerinnen und Schüler unternommen wird (Georgiou, Christou, Stavrinides & Panaoura, 2002; Weiner, 1979). Betrachtet man die Zuschreibung auf die Begabung im zeitlichen Verlauf, fällt auf, dass diese im Fach Deutsch in den ALLE-Schulen zu Beginn der Evaluation deutlich höher, am Ende der Evaluation dagegen niedriger ausgeprägt war als in den Vergleichsschulen. Dies deutet darauf hin, dass die Zuschreibung auf die Begabung in den ALLE-Schulen im längsschnittlichen Vergleich eher abnahm, in den Vergleichsschulen dagegen eher noch zunahm. Dies kann hinsichtlich der ALLE-Schulen als positives Ergebnis gewertet werden. Im Fach Mathematik blieb die Zuschreibung auf die Begabung im Zeitverlauf in den ALLE-Schulen nahezu unverändert, in den Vergleichsschulen deutete sich wiederum eine leichte Zunahme an.

Neben den Überzeugungen zum Lehren und Lernen wurden die Urteilsakkuratheit der Lehrpersonen sowie die damit einhergehende Urteilssicherheit untersucht. Vergleichbar mit den Befunden anderer Untersuchungen zeigte sich sowohl für die ALLE- als auch die Vergleichsschulen, dass die Lehrpersonen eher wenig bis moderat dazu in der Lage waren, die Lösungshäufigkeiten von Aufgaben in ihrer Klasse beziehungsweise Lerngruppe einzuschätzen (vgl. Karing, Pfost & Artelt, 2013; McElvany et al., 2009). Dieses Ergebnis ist mit Blick auf die Unterrichtspraxis nicht zufriedenstellend, da in den meisten Fächern die Bearbeitung von schriftlichen Unterrichtsmaterialien, speziell auch Aufgaben, eine zentrale Rolle spielt. Theoretisch angenommen wird, dass die aufgabenbezogenen Einschätzungen der Lehrpersonen handlungsleitend sind und wichtig für die Auswahl angemessen schwieriger und kognitiv herausfordernder Lehrmaterialien in der Unterrichtsplanungsphase (McElvany et al., 2009). Darüber hinaus ergaben sich deutliche Unterschiede in der Akkuratheit zwischen den Lehrpersonen. Dies weist auf erheblich divergente Voraussetzungen für eine adäquate Unterrichtsplanung und Unterrichtsdurchführung hin (Rogalla & Vogt, 2008; Schrader & Praetorius, 2018). Bei der Urteilssicherheit lagen die Mittelwerte für alle Aufgaben unterhalb des theoretischen Mittelwerts der Antwortskala, was auf vergleichsweise grosse subjektive Unsicherheiten bei der Einschätzung der Aufgabenschwierigkeit hinweist.

Bei der personenbezogenen Akkuratheit, die sich auf die Einschätzung der individuellen Schülerleistungen bezieht, ergab sich sowohl für die ALLE- als auch die Vergleichsschulen eine moderate bis eher hohe Genauigkeit (vgl. Hoge & Coladarci, 1989; Südkamp et al., 2012), die sich jedoch

deutlich zwischen den Lehrpersonen und den Kompetenzbereichen innerhalb der Fächer unterschied. Daneben hing die erzielte Genauigkeit im Urteil auch vom Erhebungszeitpunkt ab, am Ende des Evaluationszeitraums resultierte eine tendenziell höhere Akkuratheit als zu Beginn des Pilotprojekts, bei allerdings deutlichen Disparitäten zwischen den Kompetenzbereichen. Eine positive Tendenz fand sich auch für die Urteilssicherheit: Anders als in der ersten Erhebung lag am Ende des Evaluationszeitraums die mittlere Urteilssicherheit meist über dem theoretischen Mittelwert der Antwortskala. Auffallend ist dabei der deskriptive Befund, wonach sich die Deutsch-Lehrpersonen der ALLE-Schulen zu beiden Erhebungszeitpunkten und in allen Kompetenzbereichen unsicherer im Urteil waren als die Deutsch-Lehrpersonen der Vergleichsschulen. Eine mögliche Erklärung dafür könnte sein, dass sich unter den teilnehmenden Deutsch-Lehrpersonen der ALLE-Schulen – anders als in den Vergleichsschulen – ein nennenswerter Anteil befindet, der erst wenige Jahre Berufserfahrung mitbringt. Ein Unterschied zwischen den ALLE- und Vergleichsschulen in der Urteilsakkuratheit oder Urteilssicherheit am Ende der Evaluation – unter Kontrolle der Urteilsakkuratheit beziehungsweise Urteilssicherheit zu Beginn der Evaluation – konnte schliesslich für keinen der Kompetenzbereiche in Deutsch oder Mathematik nachgewiesen werden. Entsprechend ergibt sich kein belastbarer Hinweis, dass die Teilnahme der Lehrpersonen an ALLE mit einer verbesserten Urteilsakkuratheit oder Urteilssicherheit einherging.

Dabei ist einschränkend zu beachten, dass Daten für beide Erhebungszeitpunkte nur von einer vergleichsweise kleinen Stichprobe von Lehrpersonen zur Verfügung stand, und die Schlussfolgerungen mit entsprechend hoher Unsicherheit behaftet sind. Auch muss berücksichtigt werden, dass die Urteilsakkuratheit nur ein einzelner Indikator für die Qualität der von einer Lehrperson erbrachten diagnostischen Leistungen darstellt, in welchen sich wiederum deren diagnostische Kompetenz widerspiegelt. Für akkurate Urteile scheint relevant zu sein, auf welche wahrnehmbare Hinweisreize (engl. *cues*) sich die Urteilsbildung stützt (z.B. die konkreten Antworten von Schülerinnen und Schülern auf Prüfungsaufgaben). Entsprechend scheinen vor allem solche Massnahmen zur Förderung der Urteilsakkuratheit erfolgreich zu sein, die auf den Urteilsprozess und auf die Auswahl relevanter Hinweise fokussieren (Thiede et al., 2015). Vereinzelt Hinweise für mögliche Wirkungen des ALLE-Massnahmenbündels in diesem Bereich ergaben sich nach den Aussagen von Lehrpersonen beispielsweise hinsichtlich der Prüfung von Grundkompetenzen (als Folge des Besuchs der Weiterbildungen) oder das Erkennen der Heterogenität von Lernvoraussetzungen (als Folge des Einsatzes von EMU).

5.3.2 Qualitätsmerkmale des Unterrichts

Als ein weiterer Bestandteil der Evaluation wurde die Wirkung von ALLE auf eine breite Palette an Merkmalen der fachübergreifenden Unterrichtsqualität aus Sicht der Schülerinnen und Schüler sowie der Lehrpersonen untersucht. Festgestellt wurde, dass der Unterricht von den Schülerinnen

und Schülern bezüglich aller Qualitätsmerkmale und zu allen Messzeitpunkten tendenziell positiv wahrgenommen wurde. Dies gilt sowohl für die ALLE- als auch die Vergleichsschulen und für beide Fächer. Eine günstigere Entwicklung der Unterrichtsqualität an den ALLE-Schulen liess sich jedoch für keines der Merkmale statistisch absichern. Augenfällig ist bezüglich der Entwicklung an den ALLE-Schulen, dass die berichtete Regelerorientierung und ferner die Allgegenwärtigkeit der Lehrpersonen in beiden Fächern relativ deutlich zurückgingen. Diese Entwicklung ging jedoch nicht – wie sich aus theoretischen Erwägungen zur Klassenführung befürchten liesse (z.B. Doyle, 1986) – mit einer vergleichbaren Abnahme der berichteten Lernzeitnutzung in den Klassen einher.

Aus der Beurteilung des Unterrichts durch die Lehrpersonen selbst lässt sich kein konsistentes Bild der Entwicklungen an den ALLE-Schulen ableiten. Noch am deutlichsten zeigt sich, dass die Lehrpersonen beider Fächer an den ALLE-Schulen von einer verbesserten individuellen Förderung und höheren Leistungserwartungen berichteten, während dies an den Vergleichsschulen nicht oder in geringerem Umfang erkennbar war. Bezogen auf die individuelle Förderung lässt sich dies gut in Einklang bringen mit den Aussagen der Lehrpersonen, Schulleitungen und Schulassistenten, welche von einer besseren Förderung, einer erhöhten Lernzeitnutzung und einer Steigerung von Lernmotivation und Selbstvertrauen bei leistungsschwächeren Schülerinnen und Schülern durch ALLE berichteten (vgl. Kap. 5.2.6). In den Urteilen der Schülerinnen und Schüler in den ALLE-Schulen findet sich darauf allerdings kein Hinweis, sie schätzen den Unterricht im Wesentlichen gleich ein wie die Schülerinnen und Schüler der Vergleichsschulen (vgl. oben).

5.3.3 Motivational-affektive Lernergebnisse der Schülerinnen und Schüler

Ähnlich wie in früheren Studien zeigte sich in beiden Fächern am Ende der 9. Klassenstufe sowohl in den ALLE- als auch den Vergleichsschulen ein alterstypisches Absinken bei den motivationalen Orientierungen der Schülerinnen und Schüler (vgl. z.B. die Resultate der Zürcher Lernstandserhebung; Angelone et al., 2013). Der geringste Rückgang war dabei in beiden Fächern für die intrinsische Motivation zu verzeichnen, daneben für die zukunftsbezogene instrumentelle Motivation im Fach Mathematik. Der deutlichste Rückgang ergab sich für die anerkennungs- und leistungsbezogene Motivation der Schülerinnen und Schüler. Das Niveau der Selbstwirksamkeit blieb in beiden Fächern weitgehend unverändert.

Zwischen den ALLE- und Vergleichsschulen ergaben sich für das Fach Deutsch keine Unterschiede in der Veränderung der motivational-affektiven Lernergebnisse über den Evaluationszeitraum. Im Fach Mathematik wiesen die Schülerinnen und Schüler in den ALLE-Schulen eine günstigere Entwicklung bei der anerkennungs- und leistungsbezogenen Motivation auf als die Schülerinnen und Schüler in den Vergleichsschulen. Auch für die intrinsische Motivation ergab sich ein Hinweis auf eine günstigere Entwicklung in den ALLE-Schulen, dieser Befund ist allerdings als weniger belastbar anzusehen (tendenziell statistisch signifikant; vgl. Kap. 3.6). Die Unterschiede zwischen

den ALLE- und Vergleichsschulen erwiesen sich auch nach Kontrolle von Hintergrundmerkmalen als (tendenziell) statistisch bedeutsam mit einer mittleren beziehungsweise kleinen Effektgrösse. Vergleichbare Ergebnisse zugunsten der ALLE-Schulen resultierten auch in einer separaten Auswertung für die Schülerinnen und Schüler mit ungünstiger Lernausgangslage.

Die unterschiedliche Entwicklung in den ALLE- und Vergleichsschulen bei der anerkennungs- und leistungsbezogenen Motivation in Mathematik könnte durch eine höhere Leistungsorientierung im Unterricht an den ALLE-Schulen erklärt werden, da davon auszugehen ist, dass durch die Implementierung des ALLE-Massnahmenbündels das fachliche Lernen noch stärker in den Vordergrund gerückt wurde. Die – allerdings nur tendenziell signifikante – günstigere Entwicklung der intrinsischen Motivation im Fach Mathematik in den ALLE-Schulen entspricht noch mehr den Erwartungen. Die im Pilotprojekt implementierten Initiativen zur Unterrichtsentwicklung (z.B. die Möglichkeit für die Schülerinnen und Schüler, Feedback zu geben) sowie die Massnahmen zur individuellen Förderung sollten sich auch positiv in der Lernfreude der Schülerschaft niederschlagen. Dies lässt sich gut mit den Rückmeldungen der Schülerinnen und Schüler zu den ALLE-Massnahmen in Einklang bringen. So stimmte jeweils eine deutliche Mehrheit der Schülerinnen und Schüler, die an den entsprechenden Massnahmen teilgenommen hatten, einer Verbesserung ihrer Motivation und ihres Selbstvertrauens durch die Aufgabenhilfe und durch den Einsatz der Schulassistenten zu. Die Aussagen fielen dabei in Übereinstimmung mit der Veränderung in den motivational-affektiven Merkmalen für das Fach Mathematik noch günstiger aus als für das Fach Deutsch. Schliesslich wird gegenüber den Vergleichsschulen deutlich, dass der leichte Rückgang bei den motivationalen Orientierungen der Schülerinnen und Schüler kein Indiz für allfällige ungünstige Effekte der ALLE-Massnahmen darstellt, sondern der allgemeinen Tendenz zur Abnahme der schulbezogenen Motivation in dieser Altersstufe entspricht.

5.3.4 Fachliche Lernergebnisse der Schülerinnen und Schüler

Die Analysen zur Entwicklung der Fachleistungen ergaben, dass das ALLE-Massnahmenbündel zumindest punktuell mit günstigen Wirkungen auf die fachlichen Lernergebnisse der Schülerinnen und Schüler verknüpft war. Hinweise darauf resultierten für das Fach Deutsch im Gesamttest sowie für den Kompetenzbereich «Form und Raum» im Fach Mathematik, wobei die Unterschiede zwischen den ALLE- und Vergleichsschulen als eher gering einzustufen sind. Für das Fach Mathematik im Gesamttest zeigten sich keine Unterschiede zwischen den ALLE- und Vergleichsschulen.

In weiteren Analysen wurde auf die Schülerinnen und Schüler mit ungünstiger Lernausgangslage und in den unteren Leistungsniveaus fokussiert, es liessen sich jedoch keine Unterschiede in den Veränderungen der Fachleistungen zwischen den ALLE- und Vergleichsschulen feststellen. Dies

könnte als Hinweis darauf gewertet werden, dass die Förderung des fachlichen Lernens leistungsschwächerer Schülerinnen und Schüler in den Vergleichsschulen auf einem ähnlichen Niveau wie in den ALLE-Schulen erfolgt – es sich also auch bei den Vergleichsschulen eher um eine Positivselektion besonders engagierter Schulen und Lehrpersonen handelt, die Massnahmen zur individuellen Förderung bereits erfolgreich umsetzen. Dafür spricht, dass das Urteil der Schülerinnen und Schüler über die Unterrichtsqualität im Fachunterricht allgemein günstig ausfällt, und sich nur wenige Hinweise auf Unterschiede zwischen den ALLE- und Vergleichsschulen ergeben (vgl. Kap. 5.3.2). Es erscheint auch mit Blick auf die Stichprobenrekrutierung in der vorliegenden Evaluation plausibel, welche bei den Vergleichsschulen die Bereitschaft zur freiwilligen Teilnahme und ein entsprechendes Engagement voraussetzte (vgl. Kap. 3.3.1). Aus einem Vergleich bezüglich wichtiger soziodemografischer Merkmale der Schülerinnen und Schüler wird weiterhin deutlich, dass die Schülerschaft an den ALLE- und Vergleichsschulen über ähnliche Lernvoraussetzungen verfügte (vgl. Kap. 3.3.2.1). Sicher ist, dass die verstärkten Bemühungen in den ALLE-Schulen um die Förderung der leistungsschwächeren Schülerinnen und Schüler nicht zulasten der leistungsstärkeren Schülerinnen und Schüler gingen. Denn auch für die Subgruppe der leistungsstärkeren Schülerinnen und Schüler liessen sich keine Unterschiede in den Veränderungen der Fachleistungen zwischen ALLE- und Vergleichsschulen feststellen.

Diese Befundlage mag insgesamt nicht durchgängig den hohen Erwartungen an die Wirkungen des Pilotprojekts entsprechen. Bei der Einordnung dieser Befunde ist jedoch zu berücksichtigen, dass die Umsetzung der ALLE-Massnahmen an den beteiligten Schulen mit erheblichen Herausforderungen verknüpft war, beispielsweise mit der Rekrutierung und Einbindung geeigneter Schulassistenten, der Integration der Aufgabenhilfe in die bestehenden Stundenpläne oder der Bereitstellung zeitlicher Ressourcen für den Besuch von Weiterbildungen und Beratungsangeboten. Zudem resultierte für die Schulen ein erheblicher Koordinationsaufwand und die Notwendigkeit zur Einrichtung neuer Kommunikationskanäle und Kooperationsanlässe. Gerade bei der Implementation umfassender, anspruchsvoller Entwicklungsprogramme besteht die Gefahr, dass das Aufbrechen existierender Praktiken zunächst auch gut funktionierende Prozesse in Mitleidenschaft zieht und damit kurzfristig sogar negative Wirkungen mit sich bringen kann. Demgegenüber konnte das Leistungsniveau an den ALLE-Schulen gehalten, punktuell sogar leicht gesteigert werden. Nicht ausser Acht gelassen werden darf auch (vgl. Kap. 5.2), dass die Implementierung des ALLE-Massnahmenbündels durch die Schulen nicht durchweg so gelang, wie dies von der Projektleitung intendiert und vorgegeben war. So war offenbar – trotz der erheblichen Bemühungen aller Schulen – die Umsetzung der Massnahmen durch eine grosse Heterogenität gekennzeichnet, wodurch eine entsprechend breite Wirkung auf den Unterricht und die Lernergebnisse der Schülerinnen und Schüler wohl nicht durchweg im erhofften Ausmass entfaltet werden konnte. Schliesslich muss hinzugefügt werden, dass eine inhaltliche Passung der eingesetzten Kompetenztests zu den

Schwerpunktsetzungen der ALLE-Massnahmen (z.B. fachspezifische Weiterbildungen) nur eingeschränkt zu realisieren war, wodurch sich möglicherweise nicht alle Wirkungen in den getesteten Fachleistungen abbildeten (vgl. Kap. 5.4.3).

5.3.5 Fazit

Mit dem Einsatz des Massnahmenbündels an den ALLE-Schulen wurde darauf abgezielt, die Qualität des Unterrichts in den Fächern Deutsch und Mathematik weiterzuentwickeln und eine Verbesserung der fachlichen und motivational-affektiven Lernergebnisse, insbesondere bei den leistungsschwächeren Schülerinnen und Schülern, zu erreichen. Es zeigte sich, dass dieser Anspruch zumindest teilweise eingelöst werden konnte. Im Vergleich der ALLE-Schulen mit den Vergleichsschulen fanden sich Hinweise auf eine günstigere Leistungsentwicklung für das Fach Deutsch (im Gesamttest) sowie für den Kompetenzbereich «Form und Raum» im Fach Mathematik. Bezüglich der motivational-affektiven Merkmale ergaben sich Unterschiede zugunsten der ALLE-Schulen bei der anerkennungs- und leistungsbezogenen Motivation sowie (der Tendenz nach) bei der intrinsischen Motivation, die allerdings auf das Fach Mathematik beschränkt blieben.

Hinzugefügt werden muss, dass die erzielten Effektgrössen eher gering, die zwischen ALLE- und Vergleichsschulen festgestellten Unterschiede also nicht gross sind. Allerdings sind zum einen die Lernergebnisse von Schülerinnen und Schülern von einer Vielzahl von Bedingungsfaktoren abhängig (Brühwiler & Helmke, 2018; Helmke & Weinert, 1997). Auch ein umfassend angelegtes Schul- und Unterrichtsentwicklungsprogramm wie ALLE kann nicht alle diese Faktoren gleichermassen beeinflussen. Zum anderen beziehen sich die berichteten Effekte auf die gesamte Schülerschaft der beteiligten Schulen. Zieht man diese breite Wirkung in Betracht, sind auch vergleichsweise kleine Effekte als praktisch bedeutsam einzustufen (vgl. auch Ditton, 1990). Nicht festgestellt werden konnte, dass die Wirkungen von ALLE insbesondere bei den leistungsschwächeren Schülerinnen und Schülern auftraten, deren Förderung im Pilotprojekt einen besonderen Stellenwert einnahm. Ebenfalls nicht schlüssig nachgewiesen werden konnten positive Wirkungen auf die erfassten fachübergreifenden Qualitätsmerkmale des Unterrichts.

Für die Einordnung dieser Befunde ist die Umsetzung der Massnahmen an den ALLE-Schulen von entscheidender Bedeutung. Ob die berichteten Effekte tatsächlich die möglichen Wirkungen der ALLE-Massnahmen vollständig reflektieren, ist davon abhängig, ob diese wirklich wie intendiert und in ihrer Gesamtheit implementiert wurden. Trotz vereinzelter Einschränkungen lässt sich festhalten, dass die Evaluation keine Hinweise auf grundlegende Schwierigkeiten in den finanziellen und organisationalen Rahmenbedingungen – und daraus resultierende Einschränkungen in der Implementation des Pilotprojekts – ergaben. Demgegenüber fällt das Bild hinsichtlich der konkreten Umsetzung der Massnahmen an den ALLE-Schulen heterogener aus. Deutlich zeigte sich gerade in den Interviews mit den teilnehmenden Lehrpersonen, dass ALLE vielerorts mit grossem Einsatz

für die Weiterentwicklung des Unterrichts und die Förderung der Schülerinnen und Schüler genutzt wurde. Auch muss betont werden, dass die Urteile der Lehrpersonen, die an den jeweiligen Massnahmen Anteil hatten – und mehr noch der Schülerinnen und Schüler, der Schulleitungen und Schulassistenzen – überwiegend positiv bis sehr positiv ausfielen. Dies betrifft die gesamte Palette der Massnahmen des Pilotprojekts mit nur punktuellen Ausnahmen.

Betrachtet man die Aussagen der Beteiligten zu den einzelnen Massnahmen, wird jedoch ebenfalls deutlich, dass die Implementation von ALLE wohl nicht immer mit der vollen erhofften Wirkung realisiert werden konnte. Worauf diese Einschränkungen zurückzuführen sind, lässt sich aus den Daten der vorliegenden Evaluation nicht mit letzter Sicherheit erschliessen, doch ergaben sich Hinweise darauf, dass unter anderem begrenzte zeitliche Ressourcen, die teilweise gewährte Freiwilligkeit beziehungsweise geringe Verbindlichkeit in der Umsetzung der Massnahmen und zum Teil auch Schwierigkeiten in der Zusammenarbeit eine Rolle gespielt haben dürften. Daneben deutet sich an, dass nicht alle Massnahmen immer so verstanden und umgesetzt wurden, wie dies seitens der ALLE-Projektleitung intendiert war.

Schliesslich wird an einigen Stellen Optimierungsbedarf bei den angebotenen Massnahmen selbst sichtbar. Dies betrifft grob unterteilt einerseits die wahrgenommene Kontinuität beziehungsweise Verschränktheit der Massnahmen über die Projektlaufzeit. So konnten beispielsweise die Beratungen nicht durchweg so gestaltet oder offeriert werden, dass diese an die vorher besuchten Weiterbildungen anknüpften und von den Lehrpersonen als hilfreich empfunden wurden. Andererseits lässt sich hinterfragen, ob die Verankerung von Verantwortlichkeiten an den Schulen immer in der notwendigen Breite und Verbindlichkeit erfolgte, um den unterschiedlichen Bedürfnissen der Beteiligten gerecht zu werden. So äusserten sich zum Beispiel die Schulleitungen allgemein zufrieden mit der Rekrutierung der Schulassistenzen, während ein Anteil der Lehrpersonen die Eignung der gewonnenen Schulassistenzen eher kritisch sah.

Einige Hinweise darauf, welche Bedingungen für das Gelingen der ALLE-Massnahmen auf der Ebene der Klassen beziehungsweise Lerngruppen bedeutsam sein könnten, ergaben sich aus einer zusätzlichen explorativen Analyse der Evaluationsdaten (vgl. Kap. 4.11). Für eine günstige Entwicklung der anerkennungs- und leistungsbezogenen Motivation im Fach Deutsch zeigten sich Verbindungen zu einer verbesserten individuellen Förderung im Unterricht, zu qualitativ hochwertigen (fachbezogenen) Weiterbildungen und zu einer bereits zu Beginn des Pilotprojekts hohen Unterrichtsqualität. Darüber hinaus liess sich eine verbesserte Collective Self-efficacy – das heisst die Überzeugung des Fachkollegiums, gemeinsam erfolgreich handeln zu können –, mit einer verbesserten individuellen Förderung im Deutschunterricht und mit einer günstigeren Entwicklung der anerkennungs- und leistungsbezogenen Motivation im Fach Mathematik in Beziehung bringen. Die letztgenannte Motivationsfacette konnte zudem mit einer positiven Bewertung von Weiterbildung im Bereich der Kooperation zwischen Lehrpersonen und Schulassistenzen in Verbindung gebracht

werden. Auch wenn diese Befunde interessante Denkanstösse geben können – beispielsweise im Rahmen einer möglichen Weiterentwicklung von ALLE (vgl. Kap. 5.5) –, muss nochmals auf den explorativen Charakter der Analyse wie auch die erheblichen Einschränkungen der zugrunde liegenden Datenbasis (vgl. Kap. 4.11.1.2) hingewiesen werden.

Abschliessend ist hervorzuheben, dass ALLE als *Pilotprojekt* konzipiert war, und somit Erfahrungen – positive wie gegebenenfalls negative – bei der Umsetzung eines umfangreichen Schul- und Unterrichtsentwicklungsprogramms gesammelt werden sollten. Dass sich selbst bei nicht immer idealer Umsetzung konkrete Hinweise auf Wirkungen auf die Lernergebnisse ergaben, legt nahe, dass bei einer noch kohärenteren Implementation der intendierten Massnahmen entsprechend grössere Effekte erzielt werden könnten.

5.4 Grenzen der Evaluation

Für die Evaluation des Pilotprojekts ALLE wurde ein anspruchsvolles Erhebungsdesign mit vier Messzeitpunkten realisiert, in dem die ALLE-Schulen mit einer Gruppe von Vergleichsschulen systematisch hinsichtlich der Entwicklung zentraler Lernergebnisse verglichen wurden. Es wurde ein umfangreiches Instrumentarium mit standardisierten Tests, Fragebögen und Interviews eingesetzt, um Daten bei allen wichtigen an ALLE beteiligten Personengruppen zu gewinnen. Aus der Anlage der Untersuchung ergibt sich damit eine hohe Aussagekraft zur Beantwortung der für die Evaluation formulierten Fragestellungen (vgl. Kap. 2). Dennoch muss auf verschiedene Einschränkungen hingewiesen werden, welche bei der Interpretation der Ergebnisse zu berücksichtigen sind.

5.4.1 Erhebungsdesign

Alle Schulen, die am Pilotprojekt partizipierten oder als Vergleichsschule an der Evaluation teilnahmen, entschieden sich freiwillig für ihre Mitarbeit in ALLE. Wollten Schulen am Pilotprojekt teilnehmen, mussten sie die Bereitschaft mitbringen, die in ALLE vorgesehenen umfassenden Massnahmen umzusetzen. Die Vergleichsschulen erklärten sich einverstanden, an der Evaluation und den damit verbundenen Tests und Befragungen teilzunehmen, obwohl sie selbst nicht von den ALLE-Massnahmen profitieren konnten. Da dies bestimmte Bedingungen an den Schulen voraussetzt (z.B. Einverständnis im Kollegium), muss davon ausgegangen werden, dass es sich bei der untersuchten Stichprobe von Schulen um keine für den Kanton Zürich repräsentative Auswahl handelt. Anzunehmen ist demnach, dass sich die Befunde zur Wirkung von ALLE nicht vollständig auf die Gesamtheit der Schulen im Kanton verallgemeinern lassen (Einschränkung der *externen Validität*).

Aus der Freiwilligkeit der Teilnahme als ALLE-Schule oder Vergleichsschule resultiert ausserdem, dass sich Unterschiede in den Evaluationsergebnissen zwischen den beiden Gruppen nicht ohne

Vorbehalte auf die Wirkung der ALLE-Massnahmen zurückführen lassen (Einschränkung der *internen Validität*). Eine zufällige Zuweisung der Schulen zu den beiden Gruppen – nach dem Vorbild randomisierter Erhebungsdesigns, wie sie beispielsweise in klinischen Studien eingesetzt werden (Donner & Klar, 2000) – konnte im Rahmen von ALLE nicht umgesetzt werden. Entsprechend kann trotz der verwendeten statistischen Methoden zur Wirkungsanalyse nicht ausgeschlossen werden, dass die berichteten Ergebnisdifferenzen zwischen den ALLE- und Vergleichsschulen auf bereits zuvor bestehende Unterschiede zwischen den beiden Gruppen (z.B. in den Lernvoraussetzungen der Schülerschaft) zurückzuführen sind. Dass sich die ALLE- und Vergleichsschulen auf zentralen soziodemografischen Merkmalen der Schülerinnen und Schüler nur geringfügig unterschieden (vgl. Kap. 3.3.2.1) deutet darauf hin, dass lern- und leistungsrelevante Voraussetzungen in beiden Gruppen relativ ähnlich waren, sodass eine hohe Belastbarkeit der Befunde vermutet werden kann.

Aufgrund des Erhebungsdesigns nicht möglich war es, differenzierte Aussagen über die Wirkung der einzelnen Bestandteile des ALLE-Massnahmenbündels zu treffen. Grundsätzlich wurde das Massnahmenbündel in den teilnehmenden Schulen entweder als Ganzes umgesetzt (ALLE-Schulen) oder nicht umgesetzt (Vergleichsschulen), sodass die Gegenüberstellung der ALLE- und Vergleichsschulen nur Hinweise darauf ergibt, ob das Massnahmenbündel in seiner Gesamtheit wirksam war. Ob diese Wirkung vor allem auf einzelne Massnahmen (z.B. den Einsatz von Schulassistenten) zurückgeht oder Abhängigkeiten in der Wirkung zwischen einzelnen Massnahmen bestehen, muss weitgehend offen bleiben. Wie in Kapitel 3.3.1 erläutert wurde, wurde ein Teil der Vergleichsschulen auf der Grundlage ausgewählt, dass EMU als Instrument der datenbasierten Unterrichtsreflexion bereits eingesetzt wurde. Dadurch könnte die Wirkung des ALLE-Massnahmenbündels unterschätzt worden sein. Darüber hinaus ist zu beachten, dass in den Vergleichsschulen das ALLE-Massnahmenbündel zwar nicht als Teil des Pilotprojekts umgesetzt wurde. Allerdings stand es den Schulen selbstverständlich frei, unabhängig von ALLE Massnahmen zur Weiterentwicklung des Unterrichts einzuleiten (z.B. Weiterbildungen zur Unterrichtsdiagnostik im Zusammenhang mit der Einführung des Lehrplans 21).

5.4.2 Datengrundlage

Im Verlauf der Evaluation ergaben sich Herausforderungen, wie sie für längsschnittliche Studien im Bildungsbereich typisch sind (Blossfeld, von Maurice, Bayer & Skopek, 2016; Creemers, Kyriakides & Sammons, 2010). So nahmen zwei der Vergleichsschulen nicht bis zum letzten Messzeitpunkt der Evaluation teil, und für eine Vergleichsschule lagen keine Daten für den ersten Messzeitpunkt vor. Auch fanden sich gegen Ende des Evaluationszeitraums weniger Schülerinnen und Schüler sowie Lehrpersonen bereit, die Tests und Fragebögen zu bearbeiten. Um Schülerinnen und Schüler in die Analysen einbeziehen zu können, die nicht für alle Messzeitpunkte Daten aufwiesen, wurden fehlende Werte durch multiple Imputation ersetzt (vgl. Kap. 3.5.3). Auf der Ebene

der Lehrpersonen und Schulen war dies jedoch aufgrund der geringen Fallzahlen nicht möglich. Infolgedessen reduzierte sich die verfügbare Stichprobe insbesondere in Auswertungen mit den Lehrpersonendaten zum Teil erheblich, und die Ergebnisse (z.B. zur Urteilsakkuratheit, Kap. 4.8.2; zur Unterrichtsqualität, Kap. 4.9.3) sind entsprechend mit Vorsicht zu interpretieren.

Eine weitere Herausforderung stellte die vergleichsweise grosse Anzahl von Veränderungen in der organisationalen Zuordnung von Schülerinnen und Schülern sowie Lehrpersonen dar. So wechselte eine erhebliche Anzahl von Schülerinnen und Schülern die Klasse oder die Abteilung oder schied zum Beispiel durch Schulwechsel vollständig aus der Evaluation aus (vgl. Kap. 3.3.2.1). Weiterhin erhielten einige Klassen im Verlauf der Evaluation neue Fachlehrpersonen, sodass sich weder die Unterrichtsmasse noch die Lernergebnisse der Schülerinnen und Schüler in den betreffenden Klassen eindeutig mit bestimmten Lehrpersonen in Beziehung setzen liessen. Einschränkungen durch Wechsel von Lehrpersonen und Schulleitungen ergaben sich auch in den Befragungen zu den ALLE-Massnahmen und in den Interviews, die gegen Ende der Evaluation durchgeführt wurden. Jene Personen, die erst später zu ALLE hinzustiegen, konnten nicht immer zuverlässig Auskunft über den gesamten Zeitraum des Pilotprojekts geben.

5.4.3 Erhebungsinstrumente

Zur Erfassung der fachlichen Lernergebnisse der Schülerinnen und Schüler wurden Tests aus den Aufgabenpools von *Lernpass plus* zusammengestellt (vgl. Kap. 3.4.1). Zu Beginn der Evaluation befanden sich diese Aufgabenpools gerade in einer umfassenden Revision, um in Lernpass plus eine Kompetenzmessung nach dem Lehrplan 21 zu ermöglichen. Für die Evaluation hatte dies zur Folge, dass die Erfassung von Lernergebnissen primär bezogen auf die in Lernpass plus getesteten Kompetenzbereiche des Lehrplans 21 möglich war. Während in Mathematik Tests für alle Kompetenzbereiche des Lehrplans zur Verfügung standen (d.h. für «Zahl und Variable», «Form und Raum» und «Grössen, Funktionen, Daten und Zufall»), konnten in Deutsch Kompetenzen in den Bereichen «Sprechen», «Schreiben» sowie «Literatur im Fokus» nicht getestet werden. Dies stellt besonders im Hinblick auf den Bereich Schreiben eine Einschränkung dar, da die Schreibförderung in ALLE einen hohen Stellenwert hatte. Zwar konnten die Daten des Schreibenlasses im Kanton Zürich analysiert werden (vgl. Kap. 4.10.2.4), die Aussagekraft der Ergebnisse ist aufgrund der querschnittlichen Messung und des frühen Erhebungszeitpunkts (Mitte der 8. Klassenstufe) jedoch begrenzt. Ebenso ermöglichen es die eingesetzten Tests nicht, eine Beurteilung der Leistungen bezogen auf fachliche Kriterien vorzunehmen. Die Tests erlauben den Vergleich des Leistungsniveaus zwischen ALLE- und Vergleichsschulen, aber keine unmittelbaren Rückschlüsse beispielsweise auf das Erreichen basaler Kompetenzen.

Die Qualität des Unterrichts konnte in der Evaluation mithilfe von Einschätzungen durch die Schülerinnen und Schüler und durch die Lehrpersonen erfasst werden. Zwar weisen Studien darauf hin,

dass Schülerinnen- und Schülereinschätzungen sowie Lehrpersoneneinschätzungen valide und reliable Messungen der Unterrichtsqualität ermöglichen können (z.B. Fauth, Decristan, Rieser, Klieme & Büttner, 2014), ihre Eignung ist jedoch unter anderem von Eigenschaften des Konstrukts und der Itemformulierung abhängig (Clausen, 2002; Fauth, Göllner, Lenske, Praetorius & Wagner, 2020). So fällt es beispielsweise Schülerinnen und Schülern schwer, Aspekte der Unterrichtsqualität zu bewerten, wenn für die Beurteilung profunde fachliche beziehungsweise fachdidaktische Kenntnisse Voraussetzung sind. Im Rahmen der Evaluation ALLE nicht möglich war eine Einschätzung der Unterrichtsqualität durch externe Beobachterinnen und Beobachter, welche in der einschlägigen Forschung meist als die beste Methode zur Messung von Unterrichtsqualität angesehen wird (Praetorius, 2013).

Eine weitere Einschränkung ergab sich beim Einsatz der Instrumente im Längsschnitt. Zwar wurden in der Evaluation Daten an vier Messzeitpunkten erhoben, die einzelnen Instrumente wurden jedoch in der Regel nicht an allen Messzeitpunkten eingesetzt (vgl. Kap. 3.1). Differenziertere Aussagen zu Entwicklungsverläufen waren damit nicht für alle Ergebnisvariablen und nicht immer sowohl für die ALLE- als auch die Vergleichsschulen möglich. Gegenüber dem ursprünglichen Erhebungsplan wurde der Umfang der Instrumente nachträglich reduziert (z.B. Streichen der Tests am dritten Erhebungszeitpunkt in den Vergleichsschulen), um die Belastung der Schulen zu begrenzen. Zudem erfolgte der Einsatz der Tests nicht an allen Messzeitpunkten mit den gleichen Rahmenbedingungen. So wurde der Test am zweiten Messzeitpunkt mit der Durchführung der Normierung für das Lernfördersystem Lernpass plus verknüpft, wobei ein deutlich anderes, komplexeres Testheftdesign Verwendung fand als an den anderen Messzeitpunkten. Auch hinsichtlich anderer Voraussetzungen könnte die Einbettung in die Normierung Einfluss auf die Testergebnisse genommen haben (z.B. bezüglich der Testmotivation der Schülerinnen und Schüler). Weiterhin gab es auch bei den Tests am dritten beziehungsweise vierten Messzeitpunkt einen Unterschied gegenüber den Tests am ersten Messzeitpunkt. Um die Belastung der Schulen zu reduzieren, wurde der Umfang der Tests (Anzahl Testaufgaben) ab dem dritten Messzeitpunkt reduziert.

Die Bedingungen der Testdurchführung an den Schulen konnten in der Evaluation nur eingeschränkt kontrolliert werden. Im Rahmen der Evaluation war kein Einsatz von Testadministratorinnen beziehungsweise Testadministratoren möglich, die Tests wurden von den Lehrpersonen vor Ort auf Basis einer Instruktion administriert. Mit hoher Sicherheit kann zumindest angenommen werden, dass die Testdurchführung als solche für die meisten Lehrpersonen keine grosse Herausforderung darstellte. Die Tests wurden über eine technische Plattform durchgeführt (vgl. Kap. 3.2.2), die den Lehrpersonen der Sekundarschulen im Kanton Zürich von den obligatorisch durchgeführten *Stellwerk*-Tests bekannt ist. Schliesslich war es nicht möglich, alle interessierenden Auskünfte und Informationen bereits im Zuge der Implementierung des Pilotprojekts in den ALLE-Schulen zu erheben. Die auf die Umsetzung von ALLE bezogenen Fragen und Interviews wurden am

Ende des Evaluationszeitraums eingesetzt und setzten damit eine retrospektive Beurteilung durch die beteiligten Personen voraus. Die Qualität retrospektiver Urteile kann durch verschiedene Einflüsse eingeschränkt sein, beispielsweise durch das Vergessen von Einzelheiten des interessierenden Phänomens oder zwischenzeitliche Interaktionen mit anderen Personen (Trochim, 2005). Daher kann die Zuverlässigkeit von Aussagen über die Abläufe an den ALLE-Schulen und die Qualität der Umsetzung des ALLE-Massnahmenbündels beeinträchtigt sein.

5.4.4 Sonstige Einschränkungen

Aus sozialwissenschaftlichen Studien ist bekannt, dass sich Veränderungen im Verhalten von Personen nicht nur aus den geplanten Massnahmen ergeben können, sondern bereits aus der Tatsache, dass die Personen beobachtet werden («Hawthorne-Effekte»). Bezogen auf das Pilotprojekt bedeutet dies, dass festgestellte Wirkungen des ALLE-Massnahmenbündels auch durch das Wissen der beteiligten Personen an den ALLE-Schulen mit bedingt sein könnte, dass von ihnen bestimmte Verhaltensweisen erwartet werden und abweichendes Verhalten registriert werden könnte. Unter «normalen» Bedingungen, das heisst bei einer Umsetzung von Massnahmen im Schulalltag und ohne begleitende Evaluation, würde dann keine vergleichbare Wirkung erzielt.

Eine letzte Einschränkung ergibt sich bezogen auf die theoretischen Grundlagen der Umsetzung der ALLE-Massnahmen. Zwar liegen aus Studien zur Schul- und Unterrichtseffektivität eine Vielzahl von Erkenntnissen über Einflussfaktoren auf schulische Lernergebnisse vor (Brühwiler & Helmke, 2018; Creemers & Kyriakides, 2008). Allerdings lässt sich aus diesem Wissen nicht unmittelbar ableiten, wie sich förderliche Bedingungen durch zielgerichtete Massnahmen in der Praxis herstellen lassen (Berliner, 2009; Bischof, Feldhoff, Hochweber & Klieme, 2017; Terhart, 2015). Noch nicht ausreichend verfügbar sind validierte Theorien der Schul- und Unterrichtsentwicklung, welche die zugrunde liegenden Mechanismen detailliert beschreiben und bei der Planung von Projekten wie ALLE handlungsleitend sein können.

5.5 Handlungsfelder und Empfehlungen

Basierend auf den Ergebnissen der Evaluation des Pilotprojekts ALLE lassen sich verschiedene Handlungsfelder identifizieren und Empfehlungen ableiten, die für die weitere Schul- und Unterrichtsentwicklung von Bedeutung sind. Die Empfehlungen sind nachfolgend zu acht Handlungsfeldern zusammengefasst. Dabei wird im Handlungsfeld 8 auf der Basis der Erkenntnisse aus dem Pilotprojekt die Weiterführung und Ausweitung des Schul- und Unterrichtsentwicklungsprojekts ALLE empfohlen.

Handlungsfeld 1: Überprüfung des theoretischen Rahmenmodells

Als konzeptionelle Grundlage und Begründungsrahmen liegt dem Pilotprojekt ALLE ein theoretisches Rahmenmodell (Helmke & Wolf, 2014) zugrunde, in welchem die vermuteten Wirkmechanismen und Gelingensbedingungen für eine erfolgreiche Schul- und Unterrichtsentwicklung dargestellt sind. Das Modell ist eine wesentliche Grundlage für die Ausarbeitung von Zielen und Massnahmen, die in den lokalen Schulentwicklungskonzepten verankert werden sollen (siehe Handlungsfeld 2). Im Hinblick auf eine noch wirksamere Implementierung der Massnahmen von ALLE sollte das Rahmenmodell aufgrund der Erkenntnisse aus dem Pilotprojekt nochmals reflektiert und gegebenenfalls aktualisiert werden. Es wird angeregt, sich insbesondere mit folgenden Bereichen des Rahmenmodells auseinanderzusetzen:

- Ausdifferenzierung der theoretischen Annahmen zur spezifischen Förderung der leistungsschwächeren Schülerinnen und Schüler (z.B. Zusammenhang zwischen «Unterrichtsqualität» und «Lerngelegenheiten und Lernzeit»);
- Reflexion des Verhältnisses zwischen generischen und fachspezifischen Unterrichtsmerkmalen;
- Bedeutung des Kontexts für die Umsetzung eines Schulentwicklungsprogramms (z.B. die spezifische Rolle der Schulleitung und der innerschulischen Kooperation, die Rolle der Klassen/Lerngruppen und des Elternhauses);
- konsequente Verortung aller zentralen Schulentwicklungsmassnahmen (z.B. Aufgabenhilfe und Beratung).

Handlungsfeld 2: Rolle der Schulleitungen und Schulentwicklungskonzept

Aus der Evaluation ALLE wird die zentrale Rolle der Schulleitung für eine erfolgreiche Schulentwicklung deutlich. Dabei lässt sich Schulentwicklung nicht zentral verordnen, sondern muss einem Entwicklungsplan folgen, der durch die Schulleitung auf die lokalen Gegebenheiten einer Schule abgestimmt wird. Grundlegende Elemente sind die Organisations- und Personalentwicklung, die in erster Linie der Unterrichtsentwicklung dienen sollen, damit lernförderliche Bedingungen für die Schülerinnen und Schüler geschaffen beziehungsweise aufrechterhalten werden können. Soll die Wirkung von ALLE verstärkt werden, sind bezogen auf die Schulleitung insbesondere die folgenden Aufgaben wesentlich:

- Ausarbeitung beziehungsweise Überarbeitung eines lokalen Schulentwicklungskonzepts, das mit den Zielsetzungen von ALLE übereinstimmt (z.B. Ressourcenplanung, Einsatz von Schulassistenten, Aufgabenhilfe, Förderung der professionellen Entwicklung der Lehrpersonen);

- Herstellen eines gemeinsamen Verständnisses für die Schulentwicklungsmassnahmen bei allen Beteiligten als Voraussetzung und Gelingensbedingung für die Schulentwicklung;
- Einfordern einer hohen Qualität bei der Umsetzung der Schulentwicklungsmassnahmen;
- Schaffung ausreichender Ressourcen für die pädagogisch-inhaltliche Arbeit der Lehrpersonen und des gesamten Personals (z.B. durch die Etablierung von unterstützenden Strukturen und den Einsatz von IT-Instrumenten).

Der wesentliche Ansatzpunkt von ALLE, um die Lernleistungen der Schülerinnen und Schüler zu verbessern, ist die Weiterentwicklung des Unterrichts. Davon ausgehend ist vor allem jenes Schulleitungshandeln relevant, welches den Unterricht und in der Folge davon das Lernen der Schülerinnen und Schüler verbessert. Das damit angesprochene Konzept einer unterrichtswirksamen Führung bezieht sich auf ein Führungsverhalten seitens der Schulleitung, das seinen «Hauptschwerpunkt auf die Schaffung eines störungsfreien Lernklimas, auf ein System klarer Lernziele und auf hohe Erwartungen an Lehrpersonen und Lernende richtet» (Hattie, 2013, zitiert nach Helmke & Helmke, 2014, S. 12; siehe z.B. Bonsen, 2016a, 2016b; Robinson, Hohepa & Lloyd, 2009). Ein solches Handeln bei den Schulleitungen zu unterstützen – und die dafür notwendigen Kompetenzen und Bereitschaften zu fördern – zählt demnach zu den wichtigsten Massnahmen, um an den Schulen die Grundlagen für eine lernwirksame Unterrichtsentwicklung zu schaffen.

Handlungsfeld 3: Zeitliche Ressourcen für inhaltliche Tätigkeiten

Die zeitlichen Ressourcen des pädagogischen Personals sind an den Schulen so zu organisieren, dass die Lehrpersonen ausreichend Zeit für die pädagogische Arbeit und ihre eigene professionelle Entwicklung einsetzen können. Die Lehrpersonen müssen insbesondere Zeit haben für den Besuch von Weiterbildungen, für die Nutzung der Beratungsangebote sowie für die Zusammenarbeit und den inhaltlichen Austausch im Team. Für eine zeitliche Entlastung der Lehrpersonen, beispielsweise im administrativ-organisatorischen Bereich, können Schulassistenzen eine wichtige Rolle übernehmen.

Handlungsfeld 4: Rollenklärung und adäquater Einsatz von Schulassistenzen

Die Evaluation hat Verbesserungsbedarf beim Einsatz von Schulassistenzen aufgezeigt. Vielen befragten Lehrpersonen aus dem Pilotprojekt standen Schulassistenzen wenig oder gar nicht zur Verfügung. Zudem gibt es Hinweise darauf, dass die tatsächlichen Einsatzbereiche der Schulassistenzen nicht immer den Empfehlungen der Bildungsdirektion beziehungsweise der Bildungsexpertinnen und Bildungsexperten entsprachen. Problematisch erscheint insbesondere, wenn Schulassistenzen (nicht angeleitet) zur Lernförderung von leistungsschwachen Schülerinnen und Schülern eingesetzt werden. Primär sollte die Schulassistentin so eingesetzt werden, dass sie der

Lehrperson Freiraum schafft, gerade solche pädagogisch anspruchsvolle Tätigkeiten selbst auszuüben. Auf eine entsprechende Ausgestaltung der Zusammenarbeit müssen Lehrpersonen und Schulassistenzen vorbereitet sein, beispielsweise durch Weiterbildungen. Die im Rahmen des Pilotprojekts angebotenen Weiterbildungen zur Zusammenarbeit zwischen Lehrpersonen und Schulassistenzen wurden allerdings nur begrenzt genutzt.

Da Schulassistenzen bei ALLE einen wichtigen Beitrag zur Umsetzung eines individualisierenden und lernförderlichen Unterrichts leisten sollen, empfiehlt es sich, den Einsatz von Schulassistenzen verpflichtend einzufordern. Aufgrund der teilweise kritischen Bewertung des Einsatzes von Schulassistenzen wird zudem empfohlen, dass der Rekrutierungsprozess verbessert wird, eine Rollenklärung stattfindet, und damit verbunden eine den fachlichen Voraussetzungen angemessene Aufgabenteilung zwischen Lehrpersonen und Schulassistenzen erfolgt. Dies erfordert sowohl eine gute Abstimmung zwischen der Schulleitung und den Lehrpersonen als auch zwischen den Lehrpersonen und den Schulassistenzen. Eine kontinuierliche Kommunikation mit systematischen Gelegenheiten zum Austausch und zur Reflexion der Zusammenarbeit zwischen den Lehrpersonen und Schulassistenzen dürfte zu einer besseren Abstimmung beitragen. Zudem besteht Weiterbildungsbedarf für Lehrpersonen in der Zusammenarbeit mit Schulassistenzen. Idealerweise besuchen die Lehrpersonen und Schulassistenzen die Weiterbildungen gemeinsam.

Handlungsfeld 5: Aufgabenhilfe als Potenzial für die Individualisierung erkennen und deren Verbindlichkeit erhöhen

Die Aufgabenhilfe kann vor allem für leistungsschwächere Schülerinnen und Schüler eine bedeutende Unterstützung sein. An den Schulen des Pilotprojekts ALLE wurde die Aufgabenhilfe sehr unterschiedlich und weitgehend zur freiwilligen Nutzung angeboten. Dies dürfte dazu beigetragen haben, dass das Angebot zur betreuten Aufgabenhilfe durch die Schülerinnen und Schüler eher begrenzt in Anspruch genommen wurde. Jene, die das Angebot genutzt haben, bewerteten die Aufgabenhilfe überwiegend als positiv und lernunterstützend.

Es ist deshalb zu empfehlen, die Konzeption der Aufgabenhilfe zu überprüfen und die Verbindlichkeit zu erhöhen, damit das Potenzial für eine höhere Individualisierung ausgeschöpft werden kann. Dazu gehören auch die Klärung des Zwecks der Aufgabenhilfe an Schulen (z.B. Aufgabenhilfe für alle oder Fokus auf Unterstützung leistungsschwacher Schülerinnen und Schüler), eine sinnvolle Einbettung in die schulhauspezifischen Angebote (z.B. Koordination mit anderen Angeboten wie Lernateliers oder Freifächern) sowie eine gute Abstimmung mit anderen Schulentwicklungsmaßnahmen. Bei der Ausgestaltung der Aufgabenhilfe ist auch die zeitliche Belastung der Schülerinnen und Schüler zu beachten.

Handlungsfeld 6: Nutzung der Expertise des Schulhausteams

Für eine gelingende Implementierung von Schulentwicklungsmaßnahmen ist die innerschulische Zusammenarbeit eine wesentliche Voraussetzung. Die Evaluation hat Hinweise darauf ergeben, dass zwar bereits vor dem Pilotprojekt eine teils intensive Zusammenarbeit innerhalb der Schulen etabliert war, das Pilotprojekt ALLE selbst aber an vielen Schulen nicht zu einer grundsätzlichen und nachhaltigen Weiterentwicklung der innerschulischen Kooperation geführt hat.

Im Hinblick auf eine erhöhte Wirksamkeit bezüglich der Lehr- und Lernprozesse empfiehlt es sich, die Expertise des Schulhausteams noch gezielter für die Unterrichtsentwicklung zu nutzen. Beispielsweise lassen sich Informationen zum Unterricht, etwa aus dem datengestützten Feedback von Schülerinnen und Schülern (vgl. EMU) oder aus gegenseitigen Hospitationen für kollegiale Unterrichtsreflexionen nutzen. Eine vermehrte Öffnung des Unterrichts oder die Nutzung von Schülerfeedback zur Weiterentwicklung des Unterrichts erfordert ein offenes und vertrauensvolles Schulhausklima sowie einen verantwortungsvollen Umgang mit Daten. Begünstigt wird eine lernwirksame gemeinsame Unterrichtsentwicklung zudem von weithin geteilten Überzeugungen im Kollegium, zusammen erfolgreich handeln und das Lernen der Schülerinnen und Schüler verbessern zu können («Collective Teacher Efficacy»; vgl. Goddard, Hoy & Woolfolk Hoy, 2000; Hattie & Zierer, 2018). Diese von den Lehrpersonen geteilten Überzeugungen sollten dementsprechend – unter anderem seitens der Schulleitung – unterstützt und gefördert werden.

Handlungsfeld 7: Förderung der professionellen Entwicklung der Lehrpersonen durch Weiterbildung und Beratung

Eine der zentralen Schulleitungsaufgaben besteht in der Personalentwicklung, insbesondere mit dem Ziel, die professionelle Entwicklung der Lehrpersonen zu unterstützen. Eine gute Personalentwicklungsstrategie ist langfristig ausgerichtet und berücksichtigt sowohl individuelle als auch schulhausspezifische Bedürfnisse. Demzufolge sind Weiterbildungen einzelner Lehrpersonen mit schulinternen Weiterbildungen als Schulhausteam aufeinander abzustimmen. Dabei sind Weiterbildungen wirksamer, wenn nicht einfach isolierte Kurse besucht werden, sondern wenn sie kontinuierlich über längere Zeiträume erfolgen und durch individuelle Beratungen begleitet werden. Demzufolge bilden die Beratungsangebote für die Lehrpersonen – wie im Pilotprojekt ALLE angedacht, aber nur vergleichsweise selten genutzt – ein wichtiges Element für eine nachhaltige professionelle Entwicklung. Es ist zu empfehlen, die Verbindlichkeit und die Steuerung der Weiterbildung zu erhöhen. Konkret sollen die Schulleitungen die Nutzung von Weiterbildungen und Beratungen in Abstimmung mit den individuellen Bedürfnissen der Lehrpersonen und den Zielsetzungen von ALLE stärker einfordern. Ein grosses Potential wird darin gesehen, dass Weiterbildungs-

tage mit weiterführenden Impulsen oder individuellem Coaching verbunden werden. Aus der Evaluation ALLE ergeben sich Hinweise, dass ein besonderer Weiterbildungsbedarf für die Lehrpersonen zur Zusammenarbeit mit den Schulassistenzen besteht.

Handlungsfeld 8: Weiterführung und Ausweitung der Schul- und Unterrichtsentwicklungsmassnahmen auf der Basis des Pilotprojekts ALLE

Die Evaluation des Pilotprojekts ALLE hat einige ermutigende Ergebnisse zur Wirksamkeit des eingesetzten Massnahmenbündels hervorgebracht. Zu nennen sind hier insbesondere die positiven Leistungsentwicklungen in Deutsch und im Kompetenzbereich «Form und Raum» in der Mathematik. Hervorzuheben ist zudem die positive Einschätzung der am Pilotprojekt ALLE beteiligten Akteure, insbesondere der Schülerinnen und Schüler, der Schulleitungen und Schulassistenzen, aber mehrheitlich auch der Lehrpersonen. Die Zustimmung der beteiligten Akteure stellt gerade im Hinblick auf eine Weiterführung der mit dem Pilotprojekt ALLE intendierten Massnahmen zur Schul- und Unterrichtsentwicklung eine wichtige Gelingensbedingung dar. Demgegenüber weist die Evaluation auch auf kritische Aspekte hin, vor allem mit der teilweise eingeschränkten oder verzögerten Umsetzung der ALLE-Massnahmen. Aus den vorliegenden Ergebnissen lassen sich für die Weiterführung von ALLE wichtige Anknüpfungspunkte ableiten, die für eine verbesserte Implementierung der ALLE-Massnahmen wertvoll sind.

Grundsätzlich sind mit dem Abschluss des Pilotprojekts ALLE drei Szenarien möglich: (1) Beendigung des Pilotprojekts ALLE, (2) Weiterführung von ALLE in einem neuen Projekt oder (3) flächendeckende Implementierung von ALLE im Regelbetrieb. Von diesen drei Szenarien schlagen wir basierend auf den Erkenntnissen der Evaluation das zweite Szenario vor. Dabei soll ausgehend von den Erfahrungen des Pilotprojekts ALLE ein weiterführendes Projekt zur Schul- und Unterrichtsentwicklung ausgearbeitet werden, in welches die Empfehlungen aus der Evaluation einfließen. Weil Schulentwicklungsprozesse viel Zeit brauchen, um Wirkung zu entfalten, ist zu überlegen, ob das Nachfolgeprojekt für einen längeren Zeitraum und mit einer nochmals höheren Verbindlichkeit seitens der Akteure geplant werden sollte. Gegebenenfalls könnten auch bisherige ALLE-Schulen, für welche die Fortführung des Pilotprojekts auf der Grundlage weiterentwickelter Massnahmen erfolgversprechend erschiene, ALLE langfristig beziehungsweise dauerhaft implementieren. Im Rahmen eines Nachfolgeprojekts wäre zu untersuchen, inwieweit solche Schulen auch als Beispiele für gute Praxis (Good-Practice-Schulen) eingesetzt werden könnten. Im Hinblick auf eine allfällige breite Implementierung in der Zukunft wird überdies empfohlen, auch das weiterführende Projekt zu evaluieren. Sofern detailliertere Erkenntnisse zur Wirksamkeit einzelner Massnahmen interessieren, wäre es notwendig, ein Projektdesign mit verschiedenen Varianten auszuarbeiten, welches eine separate Untersuchung von Massnahmen ermöglicht.

Im Rahmen der Ausarbeitung des Nachfolgeprojekts wäre ebenfalls zu prüfen, inwieweit bei den Professionalisierungsmassnahmen auch andere Fächer als Deutsch und Mathematik in den Blick genommen werden könnten, und ob ALLE künftig auf die Primarstufe ausgedehnt werden sollte. Eine Ausweitung von ALLE könnte dazu führen, dass mehr Schülerinnen und Schüler erreicht werden und potenziell von den Massnahmen profitieren können. Erfolgversprechend erscheint dies allerdings nur dann, wenn die in diesem Bericht beschriebenen Einschränkungen und Entwicklungspotenziale des Pilotprojekts ALLE bei der Anlage und Umsetzung eines solchen weiterführenden Vorhabens systematisch berücksichtigt werden.

6 Literatur

- Abdi, H. (2010). Partial least squares regression and projection on latent structure regression (PLS Regression). *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(1), 97–106.
<https://dx.doi.org/10.1002/wics.51>
- Angelone, D., Keller, F. & Moser, U. (2013). *Entwicklung schulischer Leistungen während der obligatorischen Schulzeit: Bericht zur vierten Zürcher Lernstandserhebung zuhanden der Bildungsdirektion des Kantons Zürich*. Zürich: Bildungsdirektion Kanton Zürich.
- Bacher, J., Pfaffenberger, M. & Pöschko, H. (2007). *Arbeitssituation und Weiterbildungsbedarf von Schulassistent/innen. Endbericht* (Korrigierte Fassung). Hagenberg: Pfaffenberger und Pöschko Sozialforschung OG.
- Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9(4), 469–520. <https://dx.doi.org/10.1007/s11618-006-0165-2>
- Baumert, J. & Kunter, M. (2011). Das Kompetenzmodell von COACTIV. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 29–53). Münster: Waxmann.
- Baumert, J., Lehmann, R. H., Lehrke, M., Schmitz, B., Clausen, M., Hosenfeld, I., Köller, O. & Neubrand, J. (1997). *TIMSS – Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich. Deskriptive Befunde*. Opladen: Leske + Budrich.
- Berliner, D. C. (2009). Research, policy, practice: The great disconnect. In S. D. Lapan & M. T. Quartaroli (Eds.), *Research essentials: An introduction to designs and practices* (Vol.1, pp. 295–326). San Francisco, CA: Jossey-Bass.
- Biedermann, H., Brühwiler, C., Oser, F., Affolter, B. & Bach, A. (2015). Überzeugungen zur Mathematik und zum Erwerb mathematischen Wissens. In F. Oser, H. Biedermann, C. Brühwiler & S. Steinmann (Hrsg.), *Zum Start bereit? Vertiefende Ergebnisse aus TEDS-M zur schweizerischen Lehrerbildung im internationalen und nationalen Vergleich* (S. 339–376). Opladen: Barbara Budrich.
- Biedermann, H., Steinmann, S. & Oser, F. (2015). «Glaubensbestände und Glaubenswandel»: Zur Transformation von konstruktions- und transmissionsorientierten Lehr-Lern-Überzeugungen in der Lehrpersonenausbildung. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 33(1), 46–68.
- Bildungsdirektion Kanton Zürich/Volksschulamt (2014). *Aktive Lernzeit und Lernerfolg für ALLE. Rahmenkonzept*. Verfügbar unter: <https://www.zh.ch/content/dam/zhweb/bilder-dokumente/>

themen/bildung/informationen-fuer-schulen/informationen-fuer-die-volksschule/unterricht/
unterrichtsentwicklung/alle/rahmenkonzept.pdf (abgerufen am 02.09.2020)

Bildungsdirektion Kanton Zürich/Volksschulamt (2016). *Pilotprojekt Aktive Lernzeit und Lernerfolg für ALLE: Pilotschulen erproben das Förderprogramm ALLE*. Verfügbar unter:
https://www.zh.ch/content/dam/zhweb/bilder-dokumente/themen/bildung/informationen-fuer-schulen/informationen-fuer-die-volksschule/unterricht/unterrichtsentwicklung/alle/pilotschulen_alle_kurzfassung.pdf (abgerufen am 14.09.2020)

Bildungsdirektion Kanton Zürich/Volksschulamt (2018). *Schulassistenz*. Verfügbar unter:
https://www.zh.ch/content/dam/zhweb/bilder-dokumente/themen/bildung/informationen-fuer-schulen/informationen-fuer-die-volksschule/fuehrung/personal-fuehren/kantonales-kommunales-personal/schulassistenz_empfehlungen.pdf (abgerufen am 02.09.2020)

Bischof, L. M., Feldhoff, T., Hochweber, J. & Klieme, E. (2017). Untersuchung von Schulentwicklung anhand von Schuleffektivitätsdaten – „Yes, we can“?! In U. Steffens, K. Maag Merki & H. Fend (Hrsg.), *Schulgestaltung. Aktuelle Befunde und Perspektiven der Schulqualitäts- und Schulentwicklungsforschung* (S. 287-308). Münster: Waxmann.

Blossfeld, H.-P., von Maurice, J., Bayer, M. & Skopek, J. (2016). *Methodological issues of longitudinal surveys: The example of the National Educational Panel Study*. Wiesbaden: Springer VS.

Bonsen, M. (2016a). Schulleitung und Führung in der Schule. In H. Altrichter & K. Maag Merki (Hrsg.), *Handbuch Neue Steuerung im Schulsystem* (2. Aufl., S. 301-323). Wiesbaden: Springer VS.

Bonsen, M. (2016b). Wirksame Schulleitung. In H.-G. Rolff & H. Buchen (Hrsg.), *Professionswissen Schulleitung* (4. Aufl., S. 193-228). Weinheim und Basel: Beltz.

Brägger, M. (2019). *LCH Arbeitszeiterhebung 2019*. Dübendorf: Büro Brägger.

Brennan, R. L. & Prediger, D. J. (1981). Coefficient Kappa: Some uses, misuses, and alternatives. *Educational and Psychological Measurement*, 41(3), 687–699.
<https://dx.doi.org/10.1177/001316448104100307>

Bruneforth, M., Oberwimmer, K. & Robitzsch, A. (2016). Reporting und Analysen. In S. Breit & C. Schreiner (Hrsg.), *Large-Scale Assessment mit R: Methodische Grundlagen der österreichischen Bildungsstandardüberprüfung* (S. 333–362). Wien: facultas.

Brühwiler, C. & Helmke, A. (2018). Determinanten der Schulleistung. In D. H. Rost, J. R. Sparfeldt & S. R. Buch (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (5. Aufl., S. 78-91). Weinheim: Beltz PVU.

Bundesministerium Bildung, Wissenschaft und Forschung (2018). *Handbuch VWA von A bis Z*. Verfügbar unter: <https://static.uni-graz.at/fileadmin/gewi/vwa/VWA-Gesamtbroschuere.pdf> (abgerufen am 06.09.2020)

- Clausen, M. (2002). *Unterrichtsqualität: Eine Frage der Perspektive? Empirische Analysen zur Übereinstimmung, Konstrukt- und Kriteriumsvalidität*. Münster: Waxmann.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46. <https://dx.doi.org/10.1177/001316446002000104>
- Cohen J. (1988). *Statistical power analysis for the behavioral sciences*. New York, NY: Routledge.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155–159. <https://doi.org/10.1037/0033-2909.112.1.155>
- Creemers, B. P. M. & Kyriakides, L. (2008). *The dynamics of educational effectiveness: A contribution to policy, practice and theory in contemporary schools*. London: Routledge.
- Creemers, B. P. M., Kyriakides, L. & Sammons, P. (2010). *Methodological advances in educational effectiveness research*. London: Routledge.
- Deppermann, A. & Schütte, W. (2008). Data and transcription. In G. Atnos, E. Ventola & T. Weber (Eds.), *Handbook of interpersonal communication* (pp. 179–213). Berlin: Walter de Gruyter.
- DESI-Konsortium (Hrsg.) (2008). *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie*. Weinheim und Basel: Beltz.
- Dresing, T. & Pehl, T. (2011). *Praxisbuch Transkription. Regelsysteme, Software und praktische Anleitungen für qualitative ForscherInnen*. Marburg: Eigenverlag.
- Ditton, H. (1990). Anmerkungen zum praktischen Umgang mit Anteilen erklärter Varianz. *Empirische Pädagogik*, 4(3), 289–309.
- Ditton, H. (2002). Lehrkräfte und Unterricht aus Schülersicht. Ergebnisse einer Untersuchung im Fach Mathematik. *Zeitschrift für Pädagogik*, 48(2), 262–286.
- Donner, A. & Klar, N. (2000). *Design and analysis of cluster randomization trials in health research*. London: Arnold.
- Doyle, W. (1977). Paradigms for research on teacher effectiveness. In L. S. Shulman (Ed.), *Review of research in education* (Vol. 5, pp. 163–198). Washington, DC: American Educational Research Association.
- Doyle, W. (1986). Classroom organization and management. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 392–431). New York, NY: Macmillan.
- Eder, F., Altrichter, H., Hofmann, F. & Weber, C. (Hrsg.) (2015). *Evaluation der Neuen Mittelschule (NMS). Befunde aus den Anfangskohorten. Forschungsbericht*. Graz: Leykam.
- Efron, B. & Tibshirani, R. (1986). The bootstrap method for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, 1, 1–35. <https://dx.doi.org/10.1214/ss/1177013815>

- Ehmke, T., Blum, W., Neubrand, M., Jordan, A. & Ulfig, F. (2006). Wie verändert sich die mathematische Kompetenz von der neunten zur zehnten Klassenstufe? In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, J. Rost & U. Schiefele (Hrsg.), *PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres* (S. 63–86). Münster: Waxmann.
- Enders, C. (2017). Multiple imputation as a flexible tool for missing data handling in clinical research. *Behaviour Research and Therapy*, *98*, 4–18. <https://dx.doi.org/10.1016/j.brat.2016.11.008>
- Epskamp, S., Cramer, A. O., Waldorp, L. J., Schmittmann, V. D. & Borsboom, D. (2012). qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software*, *48*(4), 1–18. <https://dx.doi.org/10.18637/jss.v048.i04>
- Fauth, B., Decristan, J., Rieser, S., Klieme, E. & Büttner, G. (2014). Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learning and Instruction*, *29*, 1–9. <https://dx.doi.org/10.1016/j.learninstruc.2013.07.001>
- Fauth, B., Göllner, R., Lenske, G., Praetorius, A.-K. & Wagner, W. (2020). Who sees what? Conceptual considerations on the measurement of teaching quality from different perspectives. *Zeitschrift für Pädagogik*, *66. Beiheft*, 138–155.
- Fend, H. (1998). *Qualität im Bildungswesen. Schulforschung zu Systembedingungen, Schulprofilen und Lehrerleistung*. Weinheim: Juventa.
- Fisher, R. A. (1958). *Statistical methods for research workers*. New York, NY: Hafner.
- Frey, A., Hartig, J. & Rupp, A. (2009). An NCME instructional module on booklet designs in large-scale assessments of student achievement: Theory and practice. *Educational Measurement: Issues and Practice*, *28*(3), 39–53. <https://dx.doi.org/10.1111/j.1745-3992.2009.00154.x>
- Friedman, J., Hastie, T. & Tibshirani, R. (2011). *glasso: Graphical lasso-estimation of Gaussian graphical models*. R package version 1.7. Verfügbar unter: <http://CRAN.R-project.org/package=glasso>
- Gailberger, S. & Willenberg, H. (2008). Leseverstehen Deutsch. In DESI-Konsortium (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (S. 60–71). Weinheim und Basel: Beltz.
- Georgiou, S. N., Christou, C., Stavrinides, P. & Panaoura, G. (2002). Teacher attributions of student failure and teacher behavior toward the failing student. *Psychology in the Schools*, *39*(5), 583–596. <https://dx.doi.org/10.1002/pits.10049>
- Goddard, R. D., Hoy, W. K. & Woolfolk Hoy, A. (2000). Collective teacher efficacy: Its meaning, measure, and impact on student achievement. *American Educational Research Journal*, *37*, 479–507.

- Gonzalez, E. & Rutkowski, L. (2010). Principles of multiple matrix booklet designs and parameter recovery in large-scale assessments. *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments*, 3, 125–156.
- Hardt, J., Herke, M. & Leonhart, R. (2012). Auxiliary variables in multiple imputation in regression with missing X: A warning against including too many in small sample research. *BMC Medical Research Methodology*, 12(184), 1–13. <https://dx.doi.org/10.1186/1471-2288-12-184>
- Hattie, J. & Zierer, K. (2018). *Visible Learning: Auf den Punkt gebracht*. Baltmannsweiler: Schneider.
- Helmke, A. (2017). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts* (7. Aufl.). Seelze: Klett-Kallmeyer.
- Helmke, A. & Helmke, T. (2014). Wie wirksam ist gute Klassenführung? Effiziente Klassenführung ist nicht alles, aber ohne sie geht alles andere gar nicht. *Lernende Schule*, 65, 9-12.
- Helmke, A., Hosenfeld, I. & Schrader, F.-W. (2004). Vergleichsarbeiten als Instrument zur Verbesserung der Diagnosekompetenz von Lehrkräften. In R. Arnold & C. Griesse (Hrsg.), *Schulleitung und Schulentwicklung* (S. 119-144). Hohengehren: Schneider.
- Helmke, A. & Schrader, F.-W. (1987). Interactional effects of instructional quality and teacher judgement accuracy on achievement. *Teaching and Teacher Education*, 3(2), 91–98. [https://dx.doi.org/10.1016/0742-051X\(87\)90010-2](https://dx.doi.org/10.1016/0742-051X(87)90010-2)
- Helmke, A. & Weinert, F. E. (1997). Bedingungsfaktoren schulischer Leistungen. In F. E. Weinert (Hrsg.), *Psychologie des Unterrichts und der Schule* (Enzyklopädie der Psychologie, Serie Pädagogische Psychologie, Bd. 3, S. 71–176). Göttingen: Hogrefe.
- Hill, C. J., Bloom, H. S., Black, A. R. & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, 2(3), 172–177. <https://dx.doi.org/10.1111/j.1750-8606.2008.00061.x>
- Hoge, R. D. & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research*, 59(3), 297–313. <https://dx.doi.org/10.3102/00346543059003297>
- Holland, P. W. & Wainer, H. (Eds.) (1993). *Differential item functioning: Theory and practice*. Hillsdale, NJ: Erlbaum.
- Hox, J. J. (2010). *Multilevel analysis: Techniques and applications*. New York, NY: Routledge.
- Karing, C., Pfof, M. & Artelt, C. (2013). Is secondary school teacher judgment accuracy related to the development of students' reading literacy? In Pfof, M., Weinert, S. & Artelt, C. (Eds.), *The development of reading literacy from early childhood to adolescence* (pp. 279–311). Bamberg: University of Bamberg Press.

- Kolen, M. J. & Brennan, R. L. (2014). *Test Equating, scaling, and linking: Methods and practices* (2nd ed.). New York, NY: Springer.
- Krauss, S., Kunter, M., Brunner, M., Baumert, J., Blum, W., Neubrand, M., ... Löwen, K. (2004). COACTIV: Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung von mathematischer Kompetenz. In J. Doll & M. Prenzel (Hrsg.), *Bildungsqualität von Schule. Lehrerprofessionalisierung, Unterrichtsentwicklung und Schülerförderung als Strategien der Qualitätsverbesserung* (S. 31–53). Münster: Waxmann.
- Kreis, A. (2015). Kollegiale Hospitation – Chancen und Realisierungsmöglichkeiten. In K. Kansteiner & C. Stamann (Hrsg.), *Zwischen Fremdsteuerung und Selbstentwicklung – Erwartungen, Realitäten, Bedarfe und Entwicklungspotential der Personalentwicklung in der Schule*. Bad Heilbrunn: Klinkhardt.
- Kruse, J. (2009). *Einführung in die qualitative Interviewforschung*. Unveröffentlichter Reader.
- Kruse, J. (2014). *Qualitative Interviewforschung. Ein integrativer Ansatz*. Weinheim: Beltz-Juventa.
- Kuckartz, U. (2010). *Einführung in die computergestützte Analyse qualitativer Daten*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Kuckartz, U. (2016). *Qualitative Inhaltsanalyse: Methoden, Praxis, Computerunterstützung*. Weinheim und Basel: Beltz Juventa.
- Kuckartz, U., Dresing, T., Stefer, C. & Rädiker, S. (2007). *Qualitative Evaluation – Der Einstieg in die Praxis* (2. Aufl.). VS Verlag für Sozialwissenschaften: Wiesbaden.
- Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement für categorical data. *Biometrics*, 33(1), 159–174. <http://dx.doi.org/10.2307/2529310>.
- Langer, A. (2010). Transkribieren – Grundlagen und Regeln. In B. Friebertshäuser, A. Langer & A. Prengel (Hrsg.), *Handbuch Qualitative Forschungsmethoden in der Erziehungswissenschaft* (S. 515–526). Weinheim: Juventa.
- Lehmann, R., Peek, R., Gänsfuß, R. & Husfeldt, V. (2002). *LAU 9. Aspekte der Lernausgangslage und der Lernentwicklung – Klassenstufe 9. Ergebnisse einer Längsschnittuntersuchung in Hamburg*. Hamburg: Behörde für Bildung und Sport.
- Lipowsky, F. (2006). Auf den Lehrer kommt es an. Empirische Evidenzen für Zusammenhänge zwischen Lehrerkompetenzen, Lehrerhandeln und dem Lernen der Schüler. *Zeitschrift für Pädagogik*, 51, 47–70.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum.

- Lüdtke, O., Robitzsch, A., Trautwein, U. & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung. *Psychologische Rundschau*, 58, 103–117.
<https://dx.doi.org/10.1026/0033-3042.58.2.103>
- McElvany, N., Schroeder, S., Hachfeld, A., Baumert, J., Richter, T., Schnotz, W., ... Ullrich, M. (2009). Diagnostische Fähigkeiten von Lehrkräften bei der Einschätzung von Schülerleistungen und Aufgabenschwierigkeiten bei Lernmedien mit instruktionalen Bildern. *Zeitschrift für Pädagogische Psychologie*, 23, 223–235. <https://dx.doi.org/10.1024/1010-0652.23.34.223>
- McNeish, D. M. (2015). Using lasso for predictor selection and to assuage overfitting: A method long overlooked in behavioral sciences. *Multivariate Behavioral Research*, 50, 471–483.
<https://dx.doi.org/10.1080/00273171.2015.1036965>
- Meier, U. & Wolf, M. (2015). *Klassenassistenzen: Rolle und Funktion. Erläuterungen aus Sicht des Volksschulamtes* [Präsentation]. Zürich: Bildungsdirektion Kanton Zürich..
- Mislevy, R. J. (1991). Randomization-based inference about latent variables from complex samples. *Psychometrika*, 56, 177–196. <https://dx.doi.org/10.1002/j.2330-8516.1988.tb00310.x>
- Moosbrugger, H. & Kelava, A. (Hrsg.). (2012). *Testtheorie und Fragebogenkonstruktion* (2. Aufl.). Berlin und Heidelberg: Springer.
- Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, 16, 159–176. <https://dx.doi.org/10.1177/014662169201600206>
- OECD. (2016). *PISA 2015 assessment and analytical framework. Science, reading, mathematics and financial literacy*. Paris: OECD. <http://dx.doi.org/10.1787/9789264255425-9-en>
- Pham, H. G. (2018). *Complex interplay effects of multiple classroom instructional and context factors on student growth in EFL in Vietnam: The case of nonlinear, regularized regression and the relevance of the scaling model*. Dissertation, Universität Koblenz-Landau. Verfügbar unter: <https://kola.opus.hbznrw.de/frontdoor/index/index/docId/1617>
- Pham, G., Hebling, L., Verner, M. & Ambrosetti, A. (2019). *ÜGK – COFO – VeCoF 2017 results: Technical appendices*. St.Gallen und Genf: Pädagogische Hochschule St.Gallen (PHSG) und Service de la recherche en éducation (SRED).
- Pham, G., Robitzsch, A., George, A. C. & Freunberger, R. (2016). Fairer Vergleich in der Rückmeldung. In S. Breit, & C. Schreiner (Hrsg.), *Large-Scale Assessment mit R: Methodische Grundlagen der österreichischen Bildungsstandardüberprüfung* (S. 295–332). Wien: facultas.
- Pohl, S., Gräfe, L., & Rose, N. (2014). Dealing with omitted and not-reached items in competence tests: Evaluating approaches accounting for missing responses in item response theory models. *Educational and Psychological Measurement*, 74(3), 423–452.
<https://dx.doi.org/10.1177/0013164413504926>

- Praetorius, A.-K. (2013). Einschätzung von Unterrichtsqualität durch externe Beobachterinnen und Beobachter. Eine kritische Betrachtung der aktuellen Vorgehensweise in der Schulpraxis. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 31(2), 174–185.
- Praetorius, A.-K., Klieme, E., Herbert, B. & Pinger, P. (2018). Generic dimensions of teaching quality: The German framework of three basic dimensions. *ZDM*, 50(3), 407–426.
<https://dx.doi.org/10.1007/s11858-018-0918-4>
- Praetorius, A.-K. & Südkamp, A. (2017). Eine Einführung in das Thema der diagnostischen Kompetenz von Lehrkräften. In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften: Theoretische und methodische Weiterentwicklungen* (Pädagogische Psychologie und Entwicklungspsychologie, Vol. 94, S. 13–18). Münster: Waxmann.
- Radford, J., Blatchford, P. & Webster, R. (2011). Opening up and closing down: how teachers and TAs manage turntaking, topic and repair in mathematics lessons. *Learning and Instruction*, 21(5), 625–635.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Kopenhagen: Danish Institute for Educational Research.
- Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.
- Robitzsch, A. & Grund, S. (2020). *miceadds: Some additional multiple imputation functions, especially for mice*. R package version 3.9-14.
Verfügbar unter: <https://cran.r-project.org/package=miceadds>
- Robitzsch, A., Grund, S. & Henke, T. (2018). *miceadds: Some additional multiple imputation functions, especially for mice*. R package version 2.15-22.
Verfügbar unter: <https://cran.r-project.org/web/packages/miceadds/miceadds.pdf>
- Robinson, V. M. J., Hohepa, M. K. & Lloyd, C. (2009). *School leadership and student outcomes: Identifying what works and why. Best evidence synthesis iteration*. Wellington: New Zealand Ministry of Education.
- Robitzsch, A., Pham, G. & Yanagida, T. (2016). Fehlende Daten und Plausible Values. In S. Breit & C. Schreiner (Hrsg.), *Large-Scale Assessment mit R: Methodische Grundlagen der österreichischen Bildungsstandardüberprüfung* (S. 259–293). Wien: facultas.
- Rogalla, M. & Vogt, F. (2008). Förderung adaptiver Lehrkompetenz: eine Interventionsstudie. *Unterrichtswissenschaft*, 36, 17–36.
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York, NY: Wiley.

- Sandmeier, A., Kunz Heim, D., Windlin, B., & Krause, A. (2017). Negative Beanspruchung von Schweizer Lehrpersonen. Trends von 2006 bis 2014. *Schweizerische Zeitschrift für Bildungswissenschaften*, 39(1), 75–94.
- Schrader, F.-W. (1989). *Diagnostische Kompetenzen von Lehrern und ihre Bedeutung für die Gestaltung und Effektivität des Unterrichts*. Frankfurt a.M.: Peter Lang.
- Schrader, F.-W. & Helmke, A. (1987). Diagnostische Kompetenz von Lehrern: Komponenten und Wirkungen. *Empirische Pädagogik*, 1(1), 27–52.
- Schrader, F.-W. & Praetorius, A.-K. (2018). Diagnostische Kompetenz von Eltern und Lehrern. In D. H. Rost, J. Sparfeldt & S. R. Buch (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (5. Aufl., S. 92–98). Weinheim: Beltz/PVU.
- Snijders, T. & Bosker, R. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed., pp. 109–118). Thousand Oaks, CA: SAGE Publications.
- Südkamp, A., Kaiser, J. & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, 104(3), 743–762.
<https://dx.doi.org/10.1037/a0027627>
- Terhart, E. (2015). Theorie der Unterrichtsentwicklung: Inspektion einer Leerstelle. In H.-G. Rolff (Hrsg.), *Handbuch Unterrichtsentwicklung* (S. 62-76). Weinheim: Beltz.
- Thiede, K. W., Brendefur, J. L., Osguthorpe, R. D., Carney, M. B., Bremner, A., Strother, S., ... Jesse, D. (2015). Can teachers accurately predict student performance? *Teaching and Teacher Education*, 49, 36–44. <https://dx.doi.org/10.1016/j.tate.2015.01.012>
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society B*, 58(1), 267–288. <https://dx.doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- Trochim, W. (2005). *Research methods: The concise knowledge base*. Mason, Ohio: Thomson.
- Tutz, G. (2012). *Regression for categorical data*. Cambridge: Cambridge University Press.
- Van Buuren, S. (2012). *Flexible imputation of missing data*. Boca Raton, FL: CRC press.
- Van Ginkel, J. R., Van der Ark, A., Sijtsma, K. & Vermunt, J. K. (2007). Two-way imputation: A Bayesian method for estimating missing scores in tests and questionnaires, and an accurate approximation. *Computational Statistics & Data Analysis*, 51(8), 4013–4027.
<https://dx.doi.org/10.1016/j.csda.2006.12.022>
- VERBI Software (2017). MAXQDA 2018 [computer software]. Berlin: VERBI Software. Verfügbar unter <https://maxqda.com>
- Von Davier, M., Gonzalez, E. & Mislevy, R. J. (2009). What are plausible values and why are they useful? *IERI Monograph Series*, 2, 9–36.

- Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, 54(3), 427–450. <https://dx.doi.org/10.1007/BF02294627>
- Weber, C., Bacher, J., Altrichter, H. & Leitgöb, H. (2015). Sind Leistungsverbesserungen an den NMS-Modellschulen zu beobachten? In F. Eder, H. Altrichter, F. Hofmann & C. Weber (Hrsg.), *Evaluation der Neuen Mittelschule (NMS). Befunde aus den Anfangskohorten. Forschungsbericht* (S. 241–264). Graz: Leykam.
- Weiner, B. (1979). A theory of motivation for some classroom experiences. *Journal of Educational Psychology*, 71(1), 3–25. <https://dx.doi.org/10.1037/0022-0663.71.1.3>
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. London: Routledge.
- Winne, P. H. (1982). Minimizing the black box problem to enhance the validity of theories about instructional effects. *Instructional Science*, 11, 13–28. <https://dx.doi.org/10.1007/BF00120978>
- Wu, M. (2005). The role of plausible values in large-scale assessments. *Studies in Educational Evaluation*, 31, 114–128. <https://dx.doi.org/10.1016/j.stueduc.2005.05.005>
- Zumwald, B. (2014). Spannungsfelder beim Einsatz von Klassenassistenten. *Schweizerische Zeitschrift für Heilpädagogik*, 20(4), 21–27.
- Zumwald, B. (2015). Professionalisierung von Lehrpersonen und Fachpersonen Sonderpädagogik für den Einsatz von Assistenzpersonal in inklusiven Schulmodellen. In H. Redlich, L. Schäfer, G. Wachtel, K. Zehbe & V. Moser (Hrsg.), *Veränderungen und Beständigkeit in Zeiten der Inklusion. Perspektiven sonderpädagogischer Professionalisierung* (S. 44–54). Bad Heilbrunn: Klinkhardt.
- Zumwald, B. (2018). Weiterbildung der Lehrpersonen für den professionellen Einsatz von Assistenzpersonen. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 36(1), 63–70.
- Zwick, R. (2012). *A review of ETS differential item functioning assessment procedures: Flagging rules, minimum sample size requirements, and criterion refinement*. Princeton, NJ: ETS.

7 Anhänge

7.1 Anhang A: Kennwerte der Skalen für die Kapitel 4.8 bis 4.9

Tabelle A1. Cronbachs α für die Skalen zu Überzeugungen zum Lehren und Lernen in den Fächern Deutsch und Mathematik für die Messzeitpunkte der ALLE-Evaluation

Skala	Beispielitem	Deutsch	Mathematik
Transmissionsorientierung	SuS lernen Mathematik am besten, indem sie den Erklärungen der Lehrperson aufmerksam folgen.	0.72/0.74	0.64/0.63
Konstruktionsorientierung	Lehrpersonen sollten SuS ermutigen, eigene Lösungen für mathematische Aufgaben zu finden, auch wenn diese nicht effizient sind.	0.66/0.75	0.86/0.74
Schülerleistung als Begabung bzw. angeborene Fähigkeit	Deutsch/Mathematik ist ein Fach, in dem angeborene Fähigkeiten viel wichtiger sind als Anstrengung.	0.80/0.83	0.77/0.77

Anmerkungen. Vor dem Schrägstrich stehen die Werte für T1 (Anfang 7. Klasse), nach dem Schrägstrich die Werte für T4 (Ende 9. Klasse 9).

Tabelle A2. Cronbachs α für die Skalen der Unterrichtsqualitätsmerkmale nach Angaben der Lehrpersonen in den Fächern Deutsch und Mathematik für die Messzeitpunkte der ALLE-Evaluation

Skala/Item	Beispielitem	Deutsch	Mathematik
Verständlichkeit und Klarheit	Ich bringe gute Beispiele, damit die SuS den Stoff besser verstehen.	0.53/0.82	0.91/0.82
Kognitive Aktivierung	Mir ist es wichtig, dass Behauptungen auch gut begründet werden.	0.65/0.62	0.65/0.62
Thematische Motivierung ^a	Ich finde das Fach Deutsch / Mathematik sehr spannend.		
Hohe Leistungserwartungen ^a	Für besonders gute Beiträge zum Deutschunterricht/Mathematikunterricht lobe ich die Schülerinnen und Schüler.		
Unterrichtsklima	Mit den Beiträgen der SuS gehe ich wertschätzend um.	0.83/0.83	0.94/0.83
Individuelle Förderung	Ich gebe die Möglichkeit, zwischen verschiedenen Aufgaben zu wählen.	0.73/0.54	0.63/0.54
Störungsprävention ^a	Bei Störungen gebe ich ein klares STOPP-Signal.		
Nutzung der Lernzeit (Time on Task)/Zeitmanagement ^a	Ich regle vieles mit kurzen Blicken und knappen Zeichen.		

Anmerkungen. Vor dem Schrägstrich stehen die Werte für T1 (Anfang 7. Klasse), nach dem Schrägstrich die Werte für T4 (Ende 9. Klasse). ^aEinzelitem.

Tabelle A3. Cronbachs α für die Skalen der Unterrichtsqualitätsmerkmale nach Angaben Schülerinnen und Schüler im Fach Deutsch für die Messzeitpunkte der ALLE-Evaluation

Skala/Item	Beispielitem	Messzeitpunkt	Wert
Verständlichkeit und Klarheit	Meine Lehrperson bringt gute Beispiele, so dass ich den Stoff besser verstehe.	T1	0.80
		T3	0.82
		T4	0.89
Lernzielorientierung	Meine Lehrperson weist immer wieder auf die Lernziele hin.	T1	0.74
		T3	0.77
		T4	0.82
Konsolidierung	Im Deutschunterricht /Mathematikunterricht kann ich das Gelernte üben und wiederholen, bis es sitzt.	T1	0.76
		T3	0.78
		T4	0.88
Kognitive Aktivierung	Meine Lehrperson stellt uns Aufgaben, bei denen ich gründlich nachdenken muss.	T1	0.70
		T3	0.73
		T4	0.85
Thematische Motivierung	Meine Lehrperson kann mich richtig für die Unterrichtsthemen begeistern.	T1	0.78
		T3	0.80
		T4	0.83
Instrumentelle Motivierung	Meine Lehrperson betont, dass gutes Deutsch/Mathematik in vielen Berufen eine grosse Rolle spielt.	T1	0.83
		T3	0.83
		T4	0.89
Hohe Leistungserwartungen	Meine Lehrperson traut mir zu, dass ich meine Leistungen noch verbessern kann.	T1	0.70
		T3	0.75
		T4	0.84
Unterrichtsklima	Mit meinen Beiträgen geht die Lehrperson wertschätzend um.	T1	0.82
		T3	0.83
		T4	0.89
Positive Fehlerkultur	Meine Lehrperson findet, dass Fehlermachen zum Deutschunterricht/Mathematikunterricht dazu gehört.	T1	0.71
		T3	0.77
		T4	0.82

(Fortsetzung)

Tabelle A3 (Fortsetzung). Cronbachs α für die Skalen der Unterrichtsqualitätsmerkmale nach Angaben Schülerinnen und Schüler im Fach Deutsch für die Messzeitpunkte der ALLE-Evaluation

Skala/Item	Messzeitpunkt	Messzeitpunkt	Wert
Lehrer-Schüler-Beziehung	Ich komme mit meiner Lehrperson gut aus.	T1	0.89
		T3	0.89
		T4	0.92
Diagnostische Kompetenz	Meine Lehrperson merkt, wenn ich etwas nicht richtig verstanden habe.	T1	0.85
		T3	0.87
		T4	0.90
Individuelle Förderung	Meine Lehrperson gibt mir die Möglichkeiten, zwischen verschiedenen Aufgaben zu wählen.	T1	0.86
		T3	0.88
		T4	0.91
Allgegenwärtigkeit	Meine Lehrperson merkt sofort, wenn ich nicht aufpasse.	T1	0.80
		T3	0.80
		T4	0.84
Regelorientierung	Bei uns im Deutschunterricht/Mathematikunterricht gibt es klare Regeln, an die ich mich halten muss	T1	0.81
		T3	0.82
		T4	0.88
Nutzung der Lernzeit (Time on Task)/Zeitmanagement	Ich bin während des Deutschunterrichts/Mathematikunterrichts aufmerksam und konzentriert.	T1	0.79
		T3	0.82
		T4	0.89
Störungsprävention ^a	Bei uns im Deutschunterricht/Mathematikunterricht gibt es bei Störungen ein klares STOPP-Signal.	T1	
		T3	
		T4	

Anmerkungen. T1 = Anfang 7. Klasse, T3 = Ende 8. Klasse, T4 = Ende 9. Klasse. ^aEinzelitem.

Tabelle A4. Cronbachs α für die Skalen der Unterrichtsqualitätsmerkmale nach Angaben Schülerinnen und Schüler im Fach Mathematik für die Messzeitpunkte der ALLE-Evaluation

Skala/Item	Beispielitem	Messzeitpunkt	Wert
Verständlichkeit und Klarheit	Meine Lehrperson bringt gute Beispiele, so dass ich den Stoff besser verstehe.	T1	0.86
		T3	0.90
		T4	0.91
Lernzielorientierung	Meine Lehrperson weist immer wieder auf die Lernziele hin	T1	0.77
		T3	0.81
		T4	0.86
Konsolidierung	Im Deutschunterricht/Mathematikunterricht kann ich das Gelernte üben und wiederholen, bis es sitzt.	T1	0.83
		T3	0.85
		T4	0.88
Kognitive Aktivierung	Meine Lehrperson stellt uns Aufgaben, bei denen ich gründlich nachdenken muss.	T1	0.83
		T3	0.86
		T4	0.91
Thematische Motivierung	Meine Lehrperson kann mich richtig für die Unterrichtsthemen begeistern.	T1	0.83
		T3	0.84
		T4	0.86
Instrumentelle Motivierung	Meine Lehrperson betont, dass gutes Deutsch/Mathematik in vielen Berufen eine grosse Rolle spielt.	T1	0.88
		T3	0.87
		T4	0.90
Hohe Leistungserwartungen	Meine Lehrperson traut mir zu, dass ich meine Leistungen noch verbessern kann.	T1	0.76
		T3	0.81
		T4	0.86
Unterrichtsklima	Mit meinen Beiträgen geht die Lehrperson wertschätzend um.	T1	0.87
		T3	0.89
		T4	0.91
Positive Fehlerkultur	Meine Lehrperson findet, dass Fehlermachen zum Deutschunterricht/Mathematikunterricht dazu gehört.	T1	0.70
		T3	0.80
		T4	0.85

(Fortsetzung)

Tabelle A4 (Fortsetzung). Cronbachs α für die Skalen der Unterrichtsqualitätsmerkmale nach Angaben Schülerinnen und Schüler im Fach Mathematik für die Messzeitpunkte der ALLE-Evaluation

Skala/Item	Messzeitpunkt	Messzeitpunkt	Wert
Lehrer-Schüler-Beziehung	Ich komme mit meiner Lehrperson gut aus.	T1	0.88
		T3	0.91
		T4	0.93
Diagnostische Kompetenz	Meine Lehrperson merkt, wenn ich etwas nicht richtig verstanden habe.	T1	0.89
		T3	0.91
		T4	0.93
Individuelle Förderung	Meine Lehrperson gibt mir die Möglichkeiten, zwischen verschiedenen Aufgaben zu wählen.	T1	0.87
		T3	0.90
		T4	0.92
Allgegenwärtigkeit	Meine Lehrperson merkt sofort, wenn ich nicht aufpasse.	T1	0.82
		T3	0.85
		T4	0.89
Regelorientierung	Bei uns im Deutschunterricht/Mathematikunterricht gibt es klare Regeln, an die ich mich halten muss	T1	0.86
		T3	0.88
		T4	0.92
Nutzung der Lernzeit (Time on Task)/Zeitmanagement	Ich bin während des Deutschunterrichts/Mathematikunterrichts aufmerksam und konzentriert.	T1	0.82
		T3	0.86
		T4	0.91
Störungsprävention ^a	Bei uns im Deutschunterricht/Mathematikunterricht gibt es bei Störungen ein klares STOPP-Signal.	T1	
		T3	
		T4	

Anmerkungen. T1 = Anfang 7. Klasse, T3 = Ende 8. Klasse, T4 = Ende 9. Klasse. ^aEinzelitem.

Tabelle A5. Cronbachs α für die Skalen zu den motivationalen Orientierungen und zur Selbstwirksamkeit der Schülerinnen und Schüler im Fach Deutsch für die Messzeitpunkte der ALLE-Evaluation

Skala/Item	Beispielitem	Messzeitpunkt	Wert
Intrinsische Motivationa	Ich habe Spass an Deutsch/Mathematik.	T1	0.91
		T3	0.92
		T4	0.93
Anerkennungs- und leistungsbezogene Motivation	Ich möchte in Deutsch/Mathematik gut sein, damit ich ein gutes Zeugnis bekomme.	T1	0.74
		T4	0.79
Zukunftsbezogene instrumentelle Motivation	Ich strenge mich im Deutsch-/Mathematikunterricht an, weil es für meinen späteren Beruf nützlich sein wird.	T1	0.90
		T4	0.90
Selbstwirksamkeit	Texte lesen und verstehen/für einen Würfel mit gegebenem Rauminhalt (z.B. 1 m ³) die Kantenlänge berechnen	T1	0.90
		T4	0.89

Anmerkungen. T1 = Anfang 7. Klasse, T3 = Ende 8. Klasse, T4 = Ende 9. Klasse. Die Erfassung der aner kennungs- und leistungsbezogenen Motivation, der zukunftsbezogenen instrumentellen Motivation und der Selbstwirksamkeit erfolgte nur zum ersten und vierten Messzeitpunkt (T1 und T4).

Tabelle A6. Cronbachs α für die Skalen zu den motivationalen Orientierungen und zur Selbstwirksamkeit der Schülerinnen und Schüler im Fach Mathematik für die Messzeitpunkte der ALLE-Evaluation

Skala/Item	Beispielitem	Messzeitpunkt	Wert
Intrinsische Motivationa	Ich habe Spass an Deutsch/Mathematik.	T1	0.91
		T3	0.94
		T4	0.94
Anerkennungs- und leistungsbezogene Motivationa	Ich möchte in Deutsch/Mathematik gut sein, damit ich ein gutes Zeugnis bekomme.	T1	0.72
		T4	0.81
Zukunftsbezogene instrumentelle Motivation	Ich strenge mich im Deutsch-/Mathematikunterricht an, weil es für meinen späteren Beruf nützlich sein wird.	T1	0.85
		T4	0.94
Selbstwirksamkeit	Texte lesen und verstehen/für einen Würfel mit gegebenem Rauminhalt (z.B. 1 m ³) die Kantenlänge berechnen	T1	0.80
		T4	0.89

Anmerkungen. T1 = Anfang 7. Klasse, T3 = Ende 8. Klasse, T4 = Ende 9. Klasse. Die Erfassung der aner kennungs- und leistungsbezogenen Motivation, der zukunftsbezogenen instrumentellen Motivation und der Selbstwirksamkeit erfolgte nur zum ersten und vierten Messzeitpunkt (T1 und T4).

7.2 Anhang B: Deskriptive Statistiken zu den Abbildungen in den Kapiteln 4.2 bis 4.7

Tabelle B1: Abbildung 2. Angaben der Schulleitungen zum Erhalt finanzieller Mittel als Rahmenbedingung für die Umsetzung des ALLE-Massnahmenbündels

Variable	N	Min	Max	M	SD
Erhalt finanzieller Mittel für Schülernessen	9	3	5	3.9	0.9
Erhalt finanzieller Mittel für Aufgabenhilfe	10	3	5	4.1	0.9
Erhalt finanzieller Mittel für Weiterbildung	11	4	5	4.6	0.5
Erhalt finanzieller Mittel für Beratung	11	3	5	4.4	0.7

Tabelle B2: Abbildung 3. Verbesserung der Lehrerkooperation nach Angaben der Deutsch-Lehrpersonen

Variable	N	Min	Max	M	SD
mehr Zeitfenster	25	1	4	1.9	1.0
bessere Koordinationsstrukturen	25	1	3	1.7	0.8
bessere Regeln der Zusammenarbeit	25	1	3	1.6	0.9
mehr Unterstützung durch Schulleitung	24	1	3	1.9	0.9
mehr Austausch mit anderen Lehrpersonen	24	1	4	2.0	0.8
mehr Arbeitsteilung mit anderen Lehrpersonen	24	1	4	1.9	0.9
mehr Zusammenarbeit mit anderen Lehrpersonen	23	1	4	2.1	1.0

Tabelle B3: Abbildung 4. Verbesserung der Lehrerkooperation nach Angaben der Mathematik-Lehrpersonen

Variable	N	Min	Max	M	SD
mehr Zeitfenster	32	1	4	1.9	1.0
bessere Koordinationsstrukturen	32	1	3	1.8	0.8
bessere Regeln der Zusammenarbeit	32	1	4	1.9	0.9
mehr Unterstützung durch Schulleitung	32	1	4	1.9	0.9
mehr Austausch mit anderen Lehrpersonen	32	1	4	2.4	1.0
mehr Arbeitsteilung mit anderen Lehrpersonen	32	1	4	2.0	1.0
mehr Zusammenarbeit mit anderen Lehrpersonen	32	1	4	1.9	1.0

Tabelle B4: Abbildung 5. Fachinhaltliche Tätigkeiten der Schülerversitäten während des Unterrichts im Fach Deutsch nach eigenen Angaben und nach Angaben der Lehrpersonen

Variable	N	Min	Max	M	SD
SA: Im Fach Deutsch unterrichten.	11	1	3	1.5	0.7
LP: Im Fach Deutsch unterrichten.	9	1	3	1.4	0.9
SA: Mit SuS Fehler oder Missverständnisse besprechen.	11	1	4	2.9	0.9
LP: Mit SuS Fehler oder Missverständnisse besprechen.	9	1	4	3.1	1.1
SA: Mit SuS über Lösungswege diskutieren.	11	1	4	3.2	1.0
LP: Mit SuS über Lösungswege diskutieren.	9	3	4	3.2	0.4
SA: Durch gezieltes Nachfragen ein Bild des Lernweges machen.	11	1	4	3.0	1.0
LP: Durch gezieltes Nachfragen ein Bild des Lernweges machen.	9	1	4	2.9	0.9
SA: Fragen der SuS zum Lerninhalt im Fach Deutsch beantworten.	11	1	4	3.1	0.9
LP: Fragen der SuS zum Lerninhalt im Fach Deutsch beantworten.	9	1	4	3.2	1.1

Tabelle B5: Abbildung 6. Unterrichtsorganisatorische Tätigkeiten der Schulassistenten während des Unterrichts im Fach Deutsch nach eigenen Angaben und nach Angaben der Lehrpersonen

Variable	N	Min	Max	M	SD
SA: Für Ruhe in der Klasse sorgen.	11	1	4	3.2	1.0
LP: Für Ruhe in der Klasse sorgen.	9	1	3	1.9	0.8
SA: Störende SuS ermahnen.	11	1	4	2.8	1.0
LP: Störende SuS ermahnen.	9	1	3	1.9	0.8
SA: Ganze Klasse bei Übungssequenzen beaufsichtigen.	11	1	4	3.0	0.9
LP: Ganze Klasse bei Übungssequenzen beaufsichtigen.	9	1	4	2.2	1.1
SA: Kleingruppen bei Übungssequenzen beaufsichtigen.	11	1	4	3.3	0.9
LP: Kleingruppen bei Übungssequenzen beaufsichtigen.	9	2	4	3.2	0.7
SA: SuS bei der Organisation/Strukturierung ihrer Arbeit unterstützen.	11	1	4	3.3	1.0
LP: SuS bei der Organisation/Strukturierung ihrer Arbeit unterstützen.	9	1	4	3.2	1.0

Tabelle B6: Abbildung 7. Unterstützung spezifischer Schülergruppen durch die Schulassistenten während des Unterrichts im Fach Deutsch nach eigenen Angaben und nach Angaben der Lehrpersonen

Variable	N	Min	Max	M	SD
SA: SuS ohne Lernschwierigkeiten beim Lösen von Aufgaben anleiten.	11	1	4	2.7	1.1
LP: SuS ohne Lernschwierigkeiten beim Lösen von Aufgaben anleiten.	9	1	4	2.2	1.1
SA: SuS mit Lernschwierigkeiten beim Lösen von Aufgaben unterstützen.	11	1	4	3.5	1.0
LP: SuS mit Lernschwierigkeiten beim Lösen von Aufgaben unterstützen.	9	2	4	3.4	0.7
SA: SuS mit Deutsch als Zweitsprache (DaZ) unterstützen.	11	1	4	2.4	1.3
LP: SuS mit Deutsch als Zweitsprache (DaZ) unterstützen.	9	1	4	3.0	1.0

Tabelle B7: Abbildung 8. Fachinhaltliche Tätigkeiten der Schulassistenzen während des Unterrichts im Fach Mathematik nach eigenen Angaben und nach Angaben der Lehrpersonen

Variable	N	Min	Max	M	SD
SA: Im Fach Mathematik unterrichten.	9	1	4	1.8	1.1
LP: Im Fach Mathematik unterrichten.	12	1	4	2.0	1.0
SA: Mit SuS Fehler oder Missverständnisse besprechen.	9	2	4	3.2	0.8
LP: Mit SuS Fehler oder Missverständnisse besprechen.	12	3	4	3.6	0.5
SA: Mit SuS über Lösungswege diskutieren.	9	2	4	3.4	0.7
LP: Mit SuS über Lösungswege diskutieren.	12	3	4	3.8	0.5
SA: Durch gezieltes Nachfragen ein Bild des Lernweges machen.	9	2	4	3.3	0.7
LP: Durch gezieltes Nachfragen ein Bild des Lernweges machen.	12	1	4	3.1	0.9
SA: Fragen der SuS zum Lerninhalt im Fach Mathematik beantworten.	9	3	4	3.4	0.5
LP: Fragen der SuS zum Lerninhalt im Fach Mathematik beantworten.	12	1	4	3.1	0.9

Tabelle B8: Abbildung 9. Unterrichtsorganisatorische Tätigkeiten der Schulassistenzen während des Unterrichts im Fach Mathematik nach eigenen Angaben

Variable	N	Min	Max	M	SD
SA: Für Ruhe in der Klasse sorgen.	9	2	4	3.1	0.8
LP: Für Ruhe in der Klasse sorgen.	12	1	4	2.1	1.0
SA: Störende SuS ermahnen.	9	2	4	2.9	0.8
LP: Störende SuS ermahnen.	12	1	4	2.3	0.8
SA: Ganze Klasse bei Übungssequenzen beaufsichtigen.	9	1	4	2.7	1.0
LP: Ganze Klasse bei Übungssequenzen beaufsichtigen.	12	1	4	2.3	1.1
SA: Kleingruppen bei Übungssequenzen beaufsichtigen.	9	3	4	3.3	0.5
LP: Kleingruppen bei Übungssequenzen beaufsichtigen.	12	2	4	3.3	0.8
SA: SuS bei der Organisation/Strukturierung ihrer Arbeit unterstützen.	9	3	4	3.4	0.5
LP: SuS bei der Organisation/Strukturierung ihrer Arbeit unterstützen.	11	2	4	3.3	0.8

Tabelle B9: Abbildung 10. Unterstützung spezifischer Schülergruppen durch die Schulassistenzen während des Unterrichts im Fach Mathematik nach eigenen Angaben und nach Angaben der Lehrpersonen

Variable	N	Min	Max	M	SD
SA: SuS ohne Lernschwierigkeiten beim Lösen von Aufgaben anleiten.	9	1	4	2.3	0.9
LP: SuS ohne Lernschwierigkeiten beim Lösen von Aufgaben anleiten.	12	1	4	2.0	1.0
SA: SuS mit Lernschwierigkeiten beim Lösen von Aufgaben unterstützen.	9	3	4	3.8	0.4
LP: SuS mit Lernschwierigkeiten beim Lösen von Aufgaben unterstützen.	12	1	4	3.4	0.9
SA: SuS mit Deutsch als Zweitsprache (DaZ) unterstützen.	9	1	4	2.6	1.1
LP: SuS mit Deutsch als Zweitsprache (DaZ) unterstützen.	11	1	4	2.4	1.1

Tabelle B10: Abbildung 11. Einschätzungen zur Aufgabenteilung und zu den Zuständigkeiten im Fach Deutsch nach Angaben der Schulassistenzen und der Lehrpersonen

Variable	N	Min	Max	M	SD
SA: Gelungene Aufgabenaufteilung	10	3	4	3.8	0.4
LP: Gelungene Aufgabenaufteilung	9	3	4	3.8	0.4
SA: Klarheit über Zuständigkeiten	10	2	4	3.7	0.7
LP: Klarheit über Zuständigkeiten	9	3	4	3.8	0.4

Tabelle B11: Abbildung 12. Einschätzungen zur Aufgabenteilung und zu den Zuständigkeiten im Fach Mathematik nach Angaben der Schulassistenzen und der Lehrpersonen

Variable	N	Min	Max	M	SD
SA: Gelungene Aufgabenaufteilung	9	3	4	3.8	0.4
LP: Gelungene Aufgabenaufteilung	12	4	4	4.0	0.0
SA: Klarheit über Zuständigkeiten	9	3	4	3.8	0.4
LP: Klarheit über Zuständigkeiten	12	4	4	4.0	0.0

Tabelle B12: Abbildung 13. Zufriedenheit mit der Führung durch die Lehrperson im Fach Deutsch nach Angaben der Schulassistenten

Variable	N	Min	Max	M	SD
Organisation der Zusammenarbeit	10	4	6	5.0	0.7
Arbeitsverhältnis zur Lehrperson	10	4	6	5.3	0.7
Führungsstil der Lehrperson (mir gegenüber)	10	5	6	5.5	0.5
Unterstützung durch die Lehrperson	10	5	6	5.3	0.5
Rolle/Funktion als Schulassistentz	10	4	6	5.2	0.8

Tabelle B13: Abbildung 14. Zufriedenheit mit der Führung durch die Lehrperson im Fach Mathematik nach Angaben der Schulassistenten

Variable	N	Min	Max	M	SD
Organisation der Zusammenarbeit	9	5	6	5.6	0.5
Arbeitsverhältnis zur Lehrperson	10	5	6	5.6	0.5
Führungsstil der Lehrperson (mir gegenüber)	10	4	6	5.4	0.7
Unterstützung durch die Lehrperson	10	5	6	5.4	0.5
Rolle/Funktion als Schulassistentz	10	4	6	5.4	0.7

Tabelle B14: Abbildung 15. Einschätzungen zur Zusammenarbeit im Fach Deutsch nach Angaben der Schulassistenten und der Lehrpersonen

Variable	N	Min	Max	M	SD
SA: Gutes Auskommen miteinander	10	3	4	3.9	0.3
LP: Gutes Auskommen miteinander	9	4	4	4.0	0.0
SA: Gemeinsame Zielsetzungen	10	3	4	3.8	0.4
LP: Gemeinsame Zielsetzungen	9	3	4	3.7	0.5
SA: Nötige Zeitfenster für Zusammenarbeit	10	1	4	3.3	0.9
LP: Nötige Zeitfenster für Zusammenarbeit	9	1	4	2.8	1.5

Tabelle B15: Abbildung 16. Einschätzungen zur Zusammenarbeit im Fach Mathematik nach Angaben der Schulassistenten und der Lehrpersonen

Variable	N	Min	Max	M	SD
SA: Gutes Auskommen miteinander	10	3	4	3.9	0.3
LP: Gutes Auskommen miteinander	12	4	4	4.0	0.0
SA: Gemeinsame Zielsetzungen	10	3	4	3.9	0.3
LP: Gemeinsame Zielsetzungen	12	3	4	3.8	0.4
SA: Nötige Zeitfenster für Zusammenarbeit	10	2	4	3.3	0.8
LP: Nötige Zeitfenster für Zusammenarbeit	12	1	4	3.3	1.0

Tabelle B16: Abbildung 17. Kooperatives Arbeiten mit der Lehrperson im Fach Deutsch nach Angaben der Schulassistenten

Variable	N	Min	Max	M	SD
Gemeinsames Planen von Lernaktivitäten für einzelne SuS	10	1	5	2.8	1.4
Erarbeiten individueller Massnahmen bei sozial auffälligen SuS	10	1	5	3.8	1.1
Austausch über disziplinarische Probleme bei SuS	10	3	5	4.6	0.7
Austausch über unsere Arbeit mit einzelnen SuS oder Gruppen	10	4	6	5.0	0.7
Gegenseitiges Feedback über Fortschritte von SuS	10	3	6	4.7	0.9
Austausch über Lernschwierigkeiten bei SuS	10	4	6	5.0	0.5
Austausch über Zusammenarbeit und mögliche Verbesserung	10	2	6	4.4	1.1

Tabelle B17: Abbildung 18. Kooperatives Arbeiten mit der Lehrperson im Fach Mathematik nach Angaben der Schulassistenzen

Variable	N	Min	Max	M	SD
Gemeinsames Planen von Lernaktivitäten für einzelne SuS	10	1	4	2.8	1.1
Erarbeiten individueller Massnahmen bei sozial auffälligen SuS	10	1	5	3.2	1.2
Austausch über disziplinarische Probleme bei SuS	10	2	6	4.4	1.2
Austausch über unsere Arbeit mit einzelnen SuS oder Gruppen	10	3	6	4.8	1.1
Gegenseitiges Feedback über Fortschritte von SuS	10	2	6	4.5	1.2
Austausch über Lernschwierigkeiten bei SuS	10	3	6	4.9	0.9
Austausch über Zusammenarbeit und mögliche Verbesserung	10	2	6	4.4	1.2

Tabelle B18: Abbildung 19. Kommunikationsformen von Lehrperson und Schulassistentz im Fach Deutsch nach Angaben der Schulassistenzen

Variable	N	Min	Max	M	SD
Austausch bei gemeinsamen Sitzungen	10	1	3	1.7	0.7
Gespräche bei schulischen Anlässen	10	1	4	2.3	0.8
Spontane Gespräche (z.B. in der Pause)	10	2	4	3.2	0.8
Gespräche in regelmässigen Zeitabständen	10	1	3	2.0	0.9
Telefonate, SMS/Mail, soziale Medien	10	1	3	1.6	0.8
Schriftliche Nachrichten in meinem Fach im Lehrerzimmer	10	1	3	1.9	0.7
Gespräche bei gemeinsamen Pausenaufsichten	10	1	2	1.5	0.5
Gespräche während des Unterrichts	10	2	4	2.6	0.7

Tabelle B19: Abbildung 20. Kommunikationsformen von Lehrperson und Schulassistentz im Fach Mathematik nach Angaben der Schulassistenten

Variable	N	Min	Max	M	SD
Austausch bei gemeinsamen Sitzungen	10	1	4	2.1	1.2
Gespräche bei schulischen Anlässen	10	1	4	2.3	0.8
Spontane Gespräche (z.B. in der Pause)	10	2	4	3.5	0.7
Gespräche in regelmässigen Zeitabständen	10	1	4	2.5	0.9
Telefonate, SMS/Mail, soziale Medien	10	1	4	2.2	0.9
Schriftliche Nachrichten in meinem Fach im Lehrerzimmer	10	1	4	2.0	0.8
Gespräche bei gemeinsamen Pausenaufsichten	10	1	3	2.0	0.9
Gespräche während des Unterrichts	10	2	4	2.6	0.8

Tabelle B20: Abbildung 21. Wahrgenommene Unterstützung durch die Schulassistentz im Fach Deutsch nach Angaben der Lehrpersonen

Variable	N	Min	Max	M	SD
Die für die Zusammenarbeit aufgebrauchte Zeit steht in gutem Verhältnis zum Ergebnis/Nutzen.	9	2	4	3.3	0.7
Durch die Zusammenarbeit fühle ich mich bei Aufgaben entlastet, die über den Unterricht hinausgehen.	9	1	4	2.3	1.6
Durch die Zusammenarbeit kann ich mich besser auf das Unterrichten konzentrieren.	9	1	4	3.0	1.3
Durch die offenen Gespräche fühle ich mich emotional entlastet.	9	1	4	3.3	1.1
Durch die Zusammenarbeit fühle ich mich bei Schwierigkeiten mit Schülerinnen und Schülern unterstützt.	9	1	4	3.2	1.2

Tabelle B21: Abbildung 22. Wahrgenommene Unterstützung durch die Schulassistenz im Fach Mathematik nach Angaben der Lehrpersonen

Variable	N	Min	Max	M	SD
Die für die Zusammenarbeit aufgebrauchte Zeit steht in gutem Verhältnis zum Ergebnis/Nutzen.	11	3	4	3.8	0.4
Durch die Zusammenarbeit fühle ich mich bei Aufgaben entlastet, die über den Unterricht hinausgehen.	12	1	4	3.0	1.2
Durch die Zusammenarbeit kann ich mich besser auf das Unterrichten konzentrieren.	12	1	4	3.3	1.2
Durch die offenen Gespräche fühle ich mich emotional entlastet.	12	1	4	3.3	0.9
Durch die Zusammenarbeit fühle ich mich bei Schwierigkeiten mit Schülerinnen und Schülern unterstützt.	12	1	4	3.5	0.9

Tabelle B22: Abbildung 23. Wahrgenommene Unterstützung durch die Schulassistenz im Fach Deutsch nach Angaben der Schülerinnen und Schüler

Variable	N	Min	Max	M	SD
Die Schulassistenz unterstützt mich im Fach Deutsch genau dort, wo ich Hilfe benötige.	87	1	4	3.2	0.6
Die Schulassistenz erleichtert mir das Lernen im Fach Deutsch.	84	1	4	3.1	0.7
Die Unterstützung durch die Schulassistenz im Fach Deutsch ist hilfreich.	86	1	4	3.2	0.7

Tabelle B23: Abbildung 24. Wahrgenommene Unterstützung durch die Schulassistenz im Fach Mathematik nach Angaben der Schülerinnen und Schüler

Variable	N	Min	Max	M	SD
Die Schulassistenz unterstützt mich im Fach Mathematik genau dort, wo ich Hilfe benötige.	94	1	4	3.3	0.7
Die Schulassistenz erleichtert mir das Lernen im Fach Mathematik.	94	1	4	3.1	0.8
Die Unterstützung durch die Schulassistenz im Fach Mathematik ist hilfreich.	90	1	4	3.2	0.7

Tabelle B24: Abbildung 25. Verbesserte Individualisierung im Unterricht durch den Einsatz der Schullassistenz im Fach Deutsch nach Angaben der Lehrpersonen

Variable	N	Min	Max	M	SD
Gezielte Förderung leistungsstärkerer SuS	9	2	4	3.1	0.9
Bereitstellung niveaudifferenzierter Aufgaben für einzelne SuS	9	2	4	3.2	0.8
Schnelleres Weitergehenlassen zur nächsten Aufgabe leistungsstärkerer SuS	9	2	4	3.2	0.8
Aufbringen von mehr Zeit für die Arbeit mit leistungsschwächeren SuS	9	2	4	3.3	0.7
Aufbringen von mehr Zeit für die Arbeit mit leistungsstärkeren SuS	9	2	4	3.2	0.8

Tabelle B25: Abbildung 26. Verbesserte Individualisierung im Unterricht durch den Einsatz der Schullassistenz im Fach Mathematik nach Angaben der Lehrpersonen

Variable	N	Min	Max	M	SD
Gezielte Förderung leistungsstärkerer SuS	12	3	4	3.9	0.3
Bereitstellung niveaudifferenzierter Aufgaben für einzelne SuS	12	3	4	3.8	0.4
Schnelleres Weitergehenlassen zur nächsten Aufgabe leistungsstärkerer SuS	11	4	4	4.0	0.0
Aufbringen von mehr Zeit für die Arbeit mit leistungsschwächeren SuS	12	2	4	3.8	0.6
Aufbringen von mehr Zeit für die Arbeit mit leistungsstärkeren SuS	12	3	4	3.9	0.3

Tabelle B26: Abbildung 27. Veränderungen durch die Schulassistentz (Fach Deutsch) nach Schülerangaben

Variable	N	Min	Max	M	SD
Die LP kann mich in Deutsch dort unterstützen, wo ich Hilfe benötige.	89	1	4	3.1	0.6
Ich kann im Deutschunterricht besser lernen.	88	1	4	3.1	0.6
Ich fühle mich im Deutschunterricht besser begleitet.	89	1	4	3.0	0.7
Die LP kann im Deutschunterricht besser auf mich eingehen.	89	1	4	3.0	0.8
Meine Leistungen in Deutsch haben sich verbessert.	89	1	4	3.0	0.7
Meine Motivation in Deutsch hat sich verbessert.	88	1	4	2.9	0.7
Mein Selbstvertrauen in Deutsch hat sich verbessert.	85	1	4	2.9	0.8
Der Unterricht in Deutsch verläuft störungsfrei.	86	1	4	2.8	0.8
Der Unterricht in Deutsch verläuft reibungslos.	84	1	4	2.9	0.8

Tabelle B27: Abbildung 28. Veränderungen durch die Schulassistentz (Fach Mathematik) nach Schülerangaben

Variable	N	Min	Max	M	SD
Die LP in Mathematik kann mich dort unterstützen, wo ich Hilfe benötige.	94	1	4	3.2	0.6
Ich kann im Mathematikunterricht besser lernen.	95	1	4	3.1	0.8
Ich fühle mich im Mathematikunterricht besser begleitet.	95	1	4	3.1	0.7
Die LP kann im Mathematikunterricht besser auf mich eingehen.	94	1	4	3.2	0.7
Meine Leistungen in Mathematik haben sich verbessert.	92	1	4	3.1	0.7
Meine Motivation in Mathematik hat sich verbessert.	94	1	4	3.0	0.8
Mein Selbstvertrauen in Mathematik hat sich verbessert.	94	1	4	3.0	0.8
Der Unterricht in Mathematik verläuft störungsfrei.	94	1	4	3.1	0.8
Der Unterricht in Mathematik verläuft reibungslos.	91	1	4	3.0	0.8

Tabelle B28: Abbildung 29. Allgemeine Zufriedenheit mit der Umsetzung der Massnahme Schulassistenten im Rahmen des Pilotprojekts nach Angaben der Lehrpersonen

Variable	N	Min	Max	M	SD
Zufriedenheit mit Umsetzung der Massnahmen an der eigenen Schule: Einsatz der Schulassistenten im Deutschunterricht	18	1	5	2.6	1.2
Zufriedenheit mit Umsetzung der Massnahmen an der eigenen Schule: Einsatz der Schulassistenten im Mathematikunterricht	27	1	5	2.6	1.4

Tabelle B29: Abbildung 30. Einsatzgebiete der Aufgabenhilfe für das Fach Deutsch nach Angaben der Lehrpersonen

Variable	N	Min	Max	M	SD
Hausaufgaben erledigen	14	1	4	3.6	0.8
Unterrichtsstoff wiederholen	13	1	4	3.1	1.0
Leistungsschwächere intensiv begleiten	13	1	4	3.2	0.9
Fehler besprechen	13	1	4	2.5	0.8
zielgerichtete Prüfungsvorbereitung	13	1	4	3.0	0.8
Partnerarbeit fördern	13	1	4	2.3	0.9
wirksame Lernstrategien entwickeln	13	1	4	2.7	0.8

Tabelle B30: Abbildung 31. Einsatzgebiete der Aufgabenhilfe für das Fach Mathematik nach Angaben der Lehrpersonen

Variable	N	Min	Max	M	SD
Hausaufgaben erledigen	14	2	4	3.6	0.6
Unterrichtsstoff wiederholen	15	2	4	2.8	0.8
Leistungsschwächere intensiv begleiten	15	1	4	2.7	1.0
Fehler besprechen	15	1	4	2.6	1.0
zielgerichtete Prüfungsvorbereitung	15	1	4	2.5	0.8
Partnerarbeit fördern	15	1	3	2.1	0.7
wirksame Lernstrategien entwickeln	15	1	4	2.5	0.7

Tabelle B31: Abbildung 32. Wahrnehmung der Aufgabenhilfe als lernförderliche Unterstützung im Fach Deutsch nach Angaben der Lehrpersonen

Variable	N	Min	Max	M	SD
Stärkung der Kompetenzen	14	1	4	3.1	1.1
Entwicklung wirksamer Lernstrategien	14	1	4	2.9	1.2

Tabelle B32: Abbildung 33. Wahrnehmung der Aufgabenhilfe als lernförderliche Unterstützung im Fach Mathematik nach Angaben der Lehrpersonen

Variable	N	Min	Max	M	SD
Stärkung der Kompetenzen	15	1	4	2.8	1.0
Entwicklung wirksamer Lernstrategien	15	1	4	2.5	0.9

Tabelle B33: Abbildung 34. Wahrgenommene Unterstützung durch die Aufgabenhilfe im Fach Deutsch nach Angaben der Schülerinnen und Schüler

Variable	N	Min	Max	M	SD
Dank der Aufgabenhilfe in Deutsch kann ich die Hausaufgaben besser lösen.	16	1	4	3.3	1.0
In der Aufgabenhilfe in Deutsch bekomme ich dort Hilfe, wo ich sie benötige.	16	1	4	3.3	1.0
Die Aufgabenhilfe erleichtert mir das Lernen in Deutsch.	16	1	4	3.3	1.0
Die Aufgabenhilfe in Deutsch ist hilfreich.	16	1	4	3.3	1.0
In der Aufgabenhilfe lerne ich, Deutschaufgaben selbstständig zu lösen.	16	1	4	3.1	1.0
In der Aufgabenhilfe in Deutsch werde ich immer mal wieder gelobt.	14	1	4	3.0	1.2

Tabelle B34: Abbildung 35. Wahrgenommene Unterstützung durch die Aufgabenhilfe im Fach Mathematik nach Angaben der Schülerinnen und Schüler

Variable	N	Min	Max	M	SD
Dank der Aufgabenhilfe in Mathematik kann ich die Hausaufgaben besser lösen.	13	2	4	3.6	0.7
In der Aufgabenhilfe in Mathematik bekomme ich dort Hilfe, wo ich sie benötige.	13	2	4	3.7	0.6
Die Aufgabenhilfe erleichtert mir das Lernen in Mathematik.	13	2	4	3.5	0.8
Die Aufgabenhilfe in Mathematik ist hilfreich.	13	2	4	3.6	0.8
In der Aufgabenhilfe lerne ich, Mathematikaufgaben selbständig zu lösen.	13	2	4	3.5	0.7
In der Aufgabenhilfe in Mathematik werde ich immer mal wieder gelobt.	13	2	4	3.7	0.6

Tabelle B35: Abbildung 36. Veränderungen durch die Aufgabenhilfe bezogen auf den Deutschunterricht nach Angaben der Schülerinnen und Schüler

Variable	N	Min	Max	M	SD
Ich kann dem Deutschunterricht besser folgen.	16	1	4	3.3	0.9
Ich habe gelernt, wie ich Aufgaben anpacken muss.	16	1	4	3.1	0.9
Meine Leistungen in Deutsch haben sich verbessert.	16	1	4	3.3	0.9
Meine Motivation in Deutsch hat sich verbessert.	15	1	4	2.9	1.0
Mein Selbstvertrauen in Deutsch hat sich verbessert.	16	1	4	3.3	1.0
Ich habe eigene Stärken erkannt, die mir beim Lernen in Deutsch helfen.	16	1	4	3.3	1.0
Mir gelingt es besser, Hausaufgaben in Deutsch selbständig zu bearbeiten.	16	1	4	3.2	1.0

Tabelle B36: Abbildung 37. Veränderungen durch die Aufgabenhilfe bezogen auf den Mathematikunterricht nach Angaben der Schülerinnen und Schüler

Variable	N	Min	Max	M	SD
Ich kann dem Mathematikunterricht besser folgen.	13	1	4	3.2	1.1
Ich habe gelernt, wie ich Aufgaben anpacken muss.	13	2	4	3.3	0.8
Meine Leistungen in Mathematik haben sich verbessert.	13	2	4	3.4	0.7
Meine Motivation in Mathematik hat sich verbessert.	13	2	4	3.3	0.9
Mein Selbstvertrauen in Mathematik hat sich verbessert.	13	2	4	3.5	0.7
Ich habe eigene Stärken erkannt, die mir beim Lernen in Mathematik helfen.	13	2	4	3.2	0.9
Mir gelingt es besser, Hausaufgaben in Mathematik selbstständig zu bearbeiten.	13	1	4	3.4	1.0

Tabelle B37: Abbildung 38. Allgemeine Zufriedenheit mit der Umsetzung der Massnahme Aufgabenhilfe im Rahmen des Pilotprojekts nach Angaben der Lehrpersonen

Variable Abb. 45 alle	N	Min	Max	M	SD
Zufriedenheit mit Umsetzung der Massnahmen an der eigenen Schule: Einsatz der Aufgabenhilfe für das Fach Deutsch	18	1	5	2.7	0.9
Zufriedenheit mit Umsetzung der Massnahmen an der eigenen Schule: Einsatz der Aufgabenhilfe für das Fach Mathematik	24	1	5	2.5	1.0

Tabelle B38: Abbildung 39. Nützlichkeit der Weiterbildungen nach Angaben der Deutsch-Lehrpersonen

Variable	N	Min	Max	M	SD
Weiterbildung Fach	19	1	4	2.7	1.4
Weiterbildung Einführung zu EMU	17	1	4	2.7	1.0
Weiterbildung gelingende Kooperation LP und SA	2	3	4	3.5	0.7
Weiterbildung Lerncoaching	10	1	4	3.0	1.2

Tabelle B39: Abbildung 40. Nützlichkeit der Weiterbildungen nach Angaben der Mathematik-Lehrpersonen

Variable	N	Min	Max	M	SD
Weiterbildung Fach	28	1	4	3.1	1.0
Weiterbildung Einführung zu EMU	23	1	4	2.5	0.9
Weiterbildung gelingende Kooperation LP und SA	5	1	3	2.4	0.9
Weiterbildung Lerncoaching	13	3	4	3.7	0.5

Tabelle B40: Abbildung 41. Verbesserungen in Bezug auf den Fachunterricht durch die besuchten Weiterbildungen nach Angaben der Deutsch-Lehrpersonen

Variable	N	Min	Max	M	SD
Fachwissen	24	1	4	1.8	1.1
Didaktisches Wissen	24	1	4	2.3	1.1
Diagnostische Fähigkeiten Überprüfung Grundkompetenzen	24	1	4	2.6	1.1
Sensibilisierung für besondere Schwierigkeiten	24	1	4	2.3	1.2
Passung meines Unterrichts	24	1	4	2.2	1.0
Gezielte Förderung Leistungsschwächerer	24	1	4	2.4	1.1

Tabelle B41: Abbildung 42. Verbesserungen in Bezug auf den Fachunterricht durch die besuchten Weiterbildungen nach Angaben der Mathematik-Lehrpersonen

Variable	N	Min	Max	M	SD
Fachwissen	31	1	4	1.4	0.7
Didaktisches Wissen	32	1	4	2.3	1.2
Diagnostische Fähigkeiten Überprüfung Grundkompetenzen	32	1	4	2.8	1.1
Sensibilisierung für besondere Schwierigkeiten	32	1	4	2.9	1.2
Passung meines Unterrichts	32	1	4	2.4	1.2
Gezielte Förderung Leistungsschwächerer	32	1	4	2.6	1.2

Tabelle B42: Abbildung 43. Allgemeine Zufriedenheit mit der Umsetzung der Durchführung der Weiterbildungen im Rahmen des Pilotprojekts nach Angaben der Lehrpersonen

Variable	N	Min	Max	M	SD
Zufriedenheit mit Umsetzung der Massnahmen an der eigenen Schule: Durchführung der Weiterbildungen Deutsch	21	1	5	2.9	1.0
Zufriedenheit mit Umsetzung der Massnahmen an der eigenen Schule: Durchführung der Weiterbildungen Mathematik	29	2	5	2.9	0.9

Tabelle B43: Abbildung 44. Allgemeine Zufriedenheit mit der Durchführung der Beratungen im Rahmen des Pilotprojekts nach Angaben der Lehrpersonen

Variable	N	Min	Max	M	SD
Zufriedenheit mit Umsetzung der Massnahmen an der eigenen Schule: Durchführung der Beratungen Deutsch	20	2	4	3.0	0.6
Zufriedenheit mit Umsetzung der Massnahmen an der eigenen Schule: Durchführung der Beratungen Mathematik	25	1	5	3.0	0.8

Tabelle B44: Abbildung 45. Formen des Erfahrungsaustauschs zu EMU nach Angaben der Schulleitungen

Variable	N	Min	Max	M	SD
Kollegialer Austausch zwischen einzelnen Lehrpersonen (z.B. im Tandem)	11	2	4	3.3	0.8
Rückspiegelung der Tandemerfahrungen auf die Ebene der Jahrgangsguppen	11	2	4	2.6	0.7
Rückspiegelung der Tandemerfahrungen auf die Ebene der Fachgruppen	11	1	4	2.4	0.8
Rückspiegelung der Tandemerfahrungen auf die Schulebene	11	1	4	2.2	1.2

Tabelle B45: Abbildung 46. Wirkungen von EMU auf den eigenen Unterricht nach Angaben der Lehrpersonen

Variable	N	Min	Max	M	SD
Datengestützte Reflexion des Unterrichts	53	1	4	2.7	1.1
Erweiterung des Handlungsrepertoires im Unterricht	53	1	4	2.5	1.2
Gezielte Unterrichtsverbesserung durch kollegiales Feedback	53	1	4	2.6	1.0
Gezielte Unterrichtsverbesserung durch Schülerfeedback	53	1	4	3.0	1.0
Gezieltes Treffen von Massnahmen zur Unterrichtsverbesserung	53	1	4	2.9	1.0

Tabelle B46: Abbildung 47. Wirkungen von EMU auf die eigene Wahrnehmung im Unterricht nach Angaben der Lehrpersonen

Variable	N	Min	Max	M	SD
Erkennen "blinder Flecken" bei der Unterrichtswahrnehmung	53	1	4	2.7	1.1
Gesteigerte Sensibilität für die Heterogenität von Lernvoraussetzungen	53	1	4	2.5	1.2
Besseres Erkennen eigener Stärken und Schwächen	53	1	4	2.9	1.1

Tabelle B47: Abbildung 48. Wirkungen von EMU auf den Schulalltag nach Angaben der Lehrpersonen

Variable	N	Min	Max	M	SD
Zunahme kollegialer Hospitationen	53	1	4	2.5	1.2
Zunahme des datengestützten Austauschs über Unterricht	52	1	4	2.2	1.2
Steigerung der Qualität des kollegialen Austauschs	50	1	4	2.4	1.1
Weiterentwicklung des Unterrichts anderer durch mein Feedback	52	1	4	2.6	1.1
Verbesserung des Lernklimas durch die Diskussion von Schülerfeedback	52	1	4	2.4	1.1

Tabelle B48: Abbildung 49. Allgemeine Zufriedenheit mit der Umsetzung der Massnahme EMU im Rahmen des Pilotprojekts nach Angaben der Lehrpersonen

Variable	N	Min	Max	M	SD
Zufriedenheit mit Umsetzung der Massnahmen an der eigenen Schule: EMU Deutsch	21	1	5	2.8	1.0
Zufriedenheit mit Umsetzung der Massnahmen an der eigenen Schule: EMU Mathematik	29	2	4	2.9	0.8

Tabelle B49: Abbildung 50. Veränderungen durch das Pilotprojekt bezogen auf die Individualisierung im Unterricht nach Angaben der Deutsch-Lehrpersonen

Variable	N	Min	Max	M	SD
Gezieltere Förderung leistungsschwächere SuS in Deutsch	25	1	4	2.5	1.2
Verbesserte Lernzeitnutzung für leistungsschwächere SuS in Deutsch	25	1	4	2.5	1.2
Lernwirksamere Gestaltung des Unterrichts	25	1	4	2.4	1.1
Gesteigerte Lernmotivation der leistungsschwächeren SuS in Deutsch	25	1	4	2.5	1.1
Gesteigertes Selbstvertrauen der leistungsschwächeren SuS in Deutsch	25	1	4	2.5	1.2

Tabelle B50: Abbildung 51. Veränderungen durch das Pilotprojekt bezogen auf die Individualisierung im Unterricht nach Angaben der Mathematik-Lehrpersonen

Variable	N	Min	Max	M	SD
Gezieltere Förderung leistungsschwächere SuS in Mathematik	31	1	4	2.6	1.0
Verbesserte Lernzeitnutzung für leistungsschwächere SuS in Mathematik	31	1	4	2.5	1.0
Lernwirksamere Gestaltung des Unterrichts	31	1	4	2.4	1.1
Gesteigerte Lernmotivation der leistungsschwächeren SuS in Mathematik	31	1	4	2.3	1.1
Gesteigertes Selbstvertrauen der leistungsschwächeren SuS in Mathematik	31	1	4	2.3	1.1

Tabelle B51: Abbildung 52. Veränderungen durch das Pilotprojekt bezogen auf die Individualisierung des Unterrichts nach Angaben der Schulleitungen

Variable	N	Min	Max	M	SD
Lernwirksamere Gestaltung des Unterrichts	11	3	4	3.2	0.4
Bessere Optimierung der Lernzeitnutzung für leistungsschwächere SuS	11	3	4	3.3	0.5
Umfangreicheres professionelles Wissen hinsichtlich Lernförderung	11	3	4	3.5	0.5

Tabelle B52: Abbildung 53. Verbesserte Förderung des Lernens leistungsschwächerer und leistungsstärkerer Schülerinnen und Schüler durch die Lehrpersonen nach Angaben der Schulleitungen und Schulassistenten

Variable	N	Min	Max	M	SD
SL: Bessere Förderung des Lernens der leistungsschwächeren SuS	11	3	4	3.6	0.5
SA: Bessere Förderung des Lernens der leistungsschwächeren SuS	14	2	4	3.4	0.6
SL: Bessere Förderung des Lernens der leistungsstärkeren SuS	11	2	3	2.6	0.5
SA: Bessere Förderung des Lernens der leistungsstärkeren SuS	14	1	4	3.3	0.9

7.3 Anhang C: Messzeitpunktspezifische Ergebnisse zu den Kapiteln 4.8 und 4.9

Tabelle C1. Mittelwerte für die Überzeugungen zum Lehren und Lernen in den ALLE- und Vergleichsschulen im Fach Deutsch für die Messzeitpunkte der ALLE-Evaluation

Merkmal	Messzeitpunkt	Gesamtstichprobe	
		ALLE-Schulen	Vergleichsschulen
Transmissionsorientierung	T1	3.55	3.97
	T4	3.57	3.44
Konstruktionsorientierung	T1	4.87	4.85
	T4	4.64	5.08
Schülerleistung als Begabung bzw. angeborene Fähigkeit	T1	2.77	2.52
	T4	2.55	2.76

Anmerkungen. T1 = Anfang 7. Klasse, T4 = Ende 9. Klasse; ALLE-Schulen: $N_{T1} = 35$, $N_{T4} = 32$ und Vergleichsschulen: $N_{T1} = 22$, $N_{T4} = 16$.

Tabelle C2. Mittelwerte für die Überzeugungen zum Lehren und Lernen in den ALLE- und Vergleichsschulen im Fach Mathematik für die Messzeitpunkte der ALLE-Evaluation

Merkmal	Messzeitpunkt	Gesamtstichprobe	
		ALLE-Schulen	Vergleichsschulen
Transmissionsorientierung	T1	3.16	3.77
	T4	3.41	3.38
Konstruktionsorientierung	T1	4.87	4.48
	T4	4.93	4.93
Schülerleistung als Begabung bzw. angeborene Fähigkeit	T1	2.47	2.66
	T4	2.68	2.45

Anmerkungen. T1 = Anfang 7. Klasse, T4 = Ende 9. Klasse; ALLE-Schulen: $N_{T1} = 38$, $N_{T4} = 36$ und Vergleichsschulen: $N_{T1} = 25$, $N_{T4} = 16 - 17$.

Tabelle C3. Personenbezogene Urteilsakkuratheit (Rangordnungskomponente) und Urteilssicherheit von Lehrpersonen in den ALLE- und Vergleichsschulen im Fach Deutsch

Variable	Messzeitpunkt	Gesamtstichprobe			
		ALLE-Schulen		Vergleichsschulen	
		N	M (SD)	N	M (SD)
<i>Urteilsakkuratheit</i>					
Zahl und Variable	T1	35	0.39 (0.48)	21	0.44 (0.27)
	T4	22	0.55 (0.34)	13	0.59 (0.40)
Form und Raum	T1	35	0.39 (0.27)	22	0.47 (0.30)
	T4	22	0.49 (0.34)	13	0.48 (0.22)
Grössen, Funktionen, Daten und Zufall	T1	36	0.46 (0.35)	21	0.53 (0.23)
	T4	22	0.45 (0.35)	13	0.52 (0.25)
<i>Urteilssicherheit</i>					
Zahl und Variable	T1	36	3.22 (0.83)	22	3.09 (0.92)
	T4	21	3.90 (0.77)	11	3.36 (0.92)
Form und Raum	T1	35	3.11 (0.87)	21	3.05 (0.86)
	T4	20	3.45 (0.94)	11	3.27 (0.65)
Grössen, Funktionen, Daten und Zufall	T1	31	3.03 (0.75)	23	2.77 (0.92)
	T4	21	3.76 (0.62)	11	3.64 (1.03)

Anmerkungen. T1 = Anfang 7. Klasse, T4 = Ende 9. Klasse. Für jede Lehrperson wurde die Korrelation zwischen ihren Urteilen und den tatsächlichen Lösungshäufigkeiten bei den einzelnen Aufgaben berechnet (10 Aufgaben pro Kompetenzbereich). Die Korrelationen wurden Fisher-Z transformiert, über die Lehrpersonen gemittelt und anschliessend wieder in einen Korrelationskoeffizienten transformiert, der in der Tabelle angegeben ist (Spalte *M*). Die Berechnung der Standardabweichung der Urteilsakkuratheit (Spalte *SD*) erfolgte analog zur Berechnung des Mittelwerts. Für die Berechnung des Cohens *d* wurden die Fisher-Z-transformierten Korrelationskoeffizienten verwendet.

Tabelle C4. Mittelwerte der Unterrichtsqualitätsmerkmale aus Sicht der Schülerinnen und Schüler in den ALLE- und Vergleichsschulen im Fach Deutsch für die Messzeitpunkte der ALLE-Evaluation

Unterrichtsmerkmal	Messzeitpunkt	Gesamtstichprobe	
		ALLE-Schulen	Vergleichsschulen
Verständlichkeit und Klarheit	T1	3.20	3.04
	T3	3.10	2.96
	T4	2.88	2.82
Lernzielorientierung	T1	2.97	2.78
	T3	2.72	2.65
	T4	2.70	2.51
Konsolidierung	T1	3.10	2.84
	T3	2.97	2.91
	T4	2.84	2.66
Kognitive Aktivierung	T1	2.98	2.84
	T3	2.94	2.82
	T4	2.79	2.54
Thematische Motivierung	T1	2.93	2.77
	T3	2.83	2.62
	T4	2.63	2.50
Instrumentelle Motivierung	T1	3.11	2.92
	T3	3.18	2.98
	T4	2.94	2.99
Hohe Leistungserwartungen	T1	3.04	2.94
	T3	3.05	2.98
	T4	2.94	2.66
Unterrichtsklima	T1	3.23	3.09
	T3	3.22	2.95
	T4	3.01	2.82
Positive Fehlerkultur	T1	3.29	3.02
	T3	3.23	3.03
	T4	3.06	2.58

(Fortsetzung)

Tabelle C4 (Fortsetzung). Mittelwerte der Unterrichtsqualitätsmerkmale aus Sicht der Schülerinnen und Schüler in den ALLE- und Vergleichsschulen im Fach Deutsch für die Messzeitpunkte der ALLE-Evaluation

Unterrichtsmerkmal	Messzeitpunkt	Gesamtstichprobe	
		ALLE-Schulen	Vergleichsschulen
Lehrer-Schüler-Beziehung	T1	3.29	3.11
	T3	3.15	3.02
	T4	3.04	2.73
Diagnostische Kompetenz	T1	2.98	2.84
	T3	2.89	2.62
	T4	2.81	2.38
Individuelle Förderung	T1	3.03	2.85
	T3	2.94	2.71
	T4	2.80	2.51
Allgegenwärtigkeit	T1	3.11	2.97
	T3	2.99	2.80
	T4	2.71	2.51
Regelorientierung	T1	3.35	3.13
	T3	3.27	3.17
	T4	2.97	3.03
Nutzung der Lernzeit (Time on Task) / Zeitmanagement	T1	2.99	2.80
	T3	2.88	2.75
	T4	2.83	2.54
Störungsprävention	T1	2.85	2.71
	T3	2.82	2.98
	T4	2.88	2.72

Anmerkungen. T1 = Anfang 7. Klasse, T3 = Ende 8. Klasse, T4 = Ende 9. Klasse; ALLE-Schulen $N = 970$, Vergleichsschulen $N = 717$.

Tabelle C5. Mittelwerte der Unterrichtsqualitätsmerkmale aus Sicht der Schülerinnen und Schüler in den ALLE- und Vergleichsschulen im Fach Mathematik für die Messzeitpunkte der ALLE-Evaluation

Unterrichtsmerkmal	Messzeitpunkt	Gesamtstichprobe	
		ALLE-Schulen	Vergleichsschulen
Verständlichkeit und Klarheit	T1	3.19	3.02
	T3	3.02	2.75
	T4	2.78	2.68
Lernzielorientierung	T1	3.06	2.77
	T3	2.91	2.61
	T4	2.77	2.57
Konsolidierung	T1	3.16	2.96
	T3	2.98	2.79
	T4	2.93	2.75
Kognitive Aktivierung	T1	3.15	2.96
	T3	3.12	2.84
	T4	2.91	2.82
Thematische Motivierung	T1	3.11	2.98
	T3	2.85	2.72
	T4	2.72	2.65
Instrumentelle Motivierung	T1	3.08	2.91
	T3	3.08	2.92
	T4	2.91	2.66
Hohe Leistungserwartungen	T1	3.12	2.96
	T3	3.13	2.93
	T4	2.92	2.66
Unterrichtsklima	T1	3.22	3.09
	T3	3.09	2.97
	T4	3.01	2.91
Positive Fehlerkultur	T1	3.25	3.07
	T3	3.15	2.87
	T4	2.97	2.79

(Fortsetzung)

Tabelle C5 (Fortsetzung). Mittelwerte der Unterrichtsqualitätsmerkmale aus Sicht der Schülerinnen und Schüler in den ALLE- und Vergleichsschulen im Fach Mathematik für die Messzeitpunkte der ALLE-Evaluation

Unterrichtsmerkmal	Messzeitpunkt	Gesamtstichprobe	
		ALLE-Schulen	Vergleichsschulen
Lehrer-Schüler-Beziehung	T1	3.29	3.16
	T3	3.05	2.98
	T4	3.02	2.94
Diagnostische Kompetenz	T1	3.08	2.91
	T3	2.86	2.70
	T4	2.79	2.56
Individuelle Förderung	T1	3.08	2.94
	T3	2.88	2.70
	T4	2.77	2.64
Allgegenwärtigkeit	T1	3.09	2.95
	T3	2.98	2.86
	T4	2.74	2.62
Regelorientierung	T1	3.33	3.14
	T3	3.20	3.15
	T4	2.89	3.13
Nutzung der Lernzeit (Time on Task) / Zeitmanagement	T1	3.08	2.94
	T3	2.94	2.85
	T4	2.73	2.63
Störungsprävention	T1	2.96	2.80
	T3	2.99	2.93
	T4	2.90	2.90

Anmerkungen. T1 = Anfang 7. Klasse, T3 = Ende 8. Klasse, T4 = Ende 9. Klasse; ALLE-Schulen: N = 970 und Vergleichsschulen: N = 717.

Tabelle C6. Mittelwerte zu den Unterrichtsqualitätsmerkmalen aus Sicht der Lehrpersonen in den ALLE- und Vergleichsschulen im Fach Deutsch für die Messzeitpunkte der ALLE-Evaluation

Unterrichtsmerkmal	Messzeitpunkt	Gesamtstichprobe	
		ALLE-Schulen	Vergleichsschulen
Verständlichkeit und Klarheit	T1	5.02	5.13
	T4	4.79	4.72
Kognitive Aktivierung	T1	5.32	5.24
	T4	4.92	4.94
Thematische Motivierung	T1	5.57	5.41
	T4	5.28	5.38
Hohe Leistungserwartungen	T1	5.31	5.41
	T4	5.28	5.31
Unterrichtsklima	T1	5.42	5.45
	T4	5.14	5.20
Individuelle Förderung	T1	4.64	4.70
	T4	4.65	4.59
Störungsprävention	T1	5.32	5.23
	T4	4.77	5.06
Nutzung der Lernzeit (Time on Task) / Zeitmanagement	T1	5.29	5.18
	T4	5.23	5.13

Anmerkungen. T1 = Anfang 7. Klasse, T4 = Ende 9. Klasse; ALLE-Schulen: $N_{T1} = 34 - 35$, $N_{T4} = 31 - 32$ und Vergleichsschulen: $N_{T1} = 22$, $N_{T4} = 16$.

Tabelle C7. Mittelwerte zu den Unterrichtsqualitätsmerkmalen aus Sicht der Lehrpersonen in den ALLE- und Vergleichsschulen im Fach Mathematik für die Messzeitpunkte der ALLE-Evaluation

Unterrichtsmerkmal	Messzeitpunkt	Gesamtstichprobe	
		ALLE-Schulen	Vergleichsschulen
Verständlichkeit und Klarheit	T1	4.96	4.56
	T4	4.85	4.56
Kognitive Aktivierung	T1	5.21	5.06
	T4	4.88	4.78
Thematische Motivierung	T1	5.88	5.00
	T4	5.53	5.06
Hohe Leistungserwartungen	T1	5.25	5.17
	T4	5.14	5.06
Unterrichtsklima	T1	5.48	5.17
	T4	5.17	5.25
Individuelle Förderung	T1	4.73	4.33
	T4	4.72	4.48
Störungsprävention	T1	5.38	5.50
	T4	4.83	4.50
Nutzung der Lernzeit (Time on Task) / Zeitmanagement	T1	5.63	5.00
	T4	5.06	5.07

Anmerkungen. T1 = Anfang 7. Klasse, T4 = Ende 9. Klasse; ALLE-Schulen: $N_{T1} = 8$, $N_{T4} = 36$ und Vergleichsschulen: $N_{T1} = 6$, $N_{T4} = 15 - 17$.

8 Abbildungsverzeichnis

Abbildung 1. Theoretisches Rahmenmodell im Pilotprojekt ALLE (Bildungsdirektion Kanton Zürich/ Volksschulamt, 2014)	14
Abbildung 2. Angaben der Schulleitungen zum Erhalt finanzieller Mittel als Rahmenbedingung für die Umsetzung des ALLE-Massnahmenbündels	66
Abbildung 3. Verbesserung der Lehrerinnen- und Lehrerkooperation nach Angaben der Deutsch- Lehrpersonen	67
Abbildung 4. Verbesserung der Lehrerinnen- und Lehrerkooperation nach Angaben der Mathematik- Lehrpersonen	68
Abbildung 5. Fachinhaltliche Tätigkeiten der Schulassistenzen während des Unterrichts im Fach Deutsch nach eigenen Angaben und nach Angaben der Lehrpersonen	79
Abbildung 6. Unterrichtsorganisatorische Tätigkeiten der Schulassistenzen während des Unterrichts im Fach Deutsch nach eigenen Angaben und nach Angaben der Lehrpersonen	80
Abbildung 7. Unterstützung spezifischer Schülergruppen durch die Schulassistenzen während des Unterrichts im Fach Deutsch nach eigenen Angaben und nach Angaben der Lehrpersonen.....	81
Abbildung 8. Fachinhaltliche Tätigkeiten der Schulassistenzen während des Unterrichts im Fach Mathematik nach eigenen Angaben und nach Angaben der Lehrpersonen.....	82
Abbildung 9. Unterrichtsorganisatorische Tätigkeiten der Schulassistenzen während des Unterrichts im Fach Mathematik nach eigenen Angaben und nach Angaben der Lehrpersonen.....	83
Abbildung 10. Unterstützung spezifischer Schülergruppen durch die Schulassistenzen während des Unterrichts im Fach Mathematik nach eigenen Angaben und nach Angaben der Lehrpersonen	83
Abbildung 11. Einschätzungen zur Aufgabenteilung und zu den Zuständigkeiten im Fach Deutsch nach Angaben der Schulassistenzen und der Lehrpersonen	86
Abbildung 12. Einschätzungen zur Aufgabenteilung und zu den Zuständigkeiten im Fach Mathematik nach Angaben der Schulassistenzen und der Lehrpersonen	87
Abbildung 13. Zufriedenheit mit der Führung durch die Lehrperson im Fach Deutsch nach Angaben der Schulassistenzen	88
Abbildung 14. Zufriedenheit mit der Führung durch die Lehrperson im Fach Mathematik nach Angaben der Schulassistenzen	88
Abbildung 15. Einschätzungen zur Zusammenarbeit im Fach Deutsch nach Angaben der Schulassistenzen und der Lehrpersonen.....	89
Abbildung 16. Einschätzungen zur Zusammenarbeit im Fach Mathematik nach Angaben der Schulassistenzen und der Lehrpersonen.....	90

Abbildung 17. Kooperatives Arbeiten mit der Lehrperson im Fach Deutsch nach Angaben der Schuassistenzen	91
Abbildung 18. Kooperatives Arbeiten mit der Lehrperson im Fach Mathematik nach Angaben der Schuassistenzen	91
Abbildung 19. Kommunikationsformen von Lehrperson und Schuassistentz im Fach Deutsch nach Angaben der Schuassistenzen.....	93
Abbildung 20. Kommunikationsformen von Lehrperson und Schuassistentz im Fach Mathematik nach Angaben der Schuassistenzen.....	93
Abbildung 21. Wahrgenommene Untersttzung durch die Schuassistentz im Fach Deutsch nach Angaben der Lehrpersonen.....	94
Abbildung 22. Wahrgenommene Untersttzung durch die Schuassistentz im Fach Mathematik nach Angaben der Lehrpersonen.....	95
Abbildung 23. Wahrgenommene Untersttzung durch die Schuassistentz im Fach Deutsch nach Angaben der Schlerinnen und Schler	96
Abbildung 24. Wahrgenommene Untersttzung durch die Schuassistentz im Fach Mathematik nach Angaben der Schlerinnen und Schler	97
Abbildung 25. Verbesserte Individualisierung im Unterricht durch den Einsatz der Schuassistentz im Fach Deutsch nach Angaben der Lehrpersonen	98
Abbildung 26. Verbesserte Individualisierung im Unterricht durch den Einsatz der Schuassistentz im Fach Mathematik nach Angaben der Lehrpersonen	98
Abbildung 27. Vernderungen durch die Schuassistentz (Fach Deutsch) nach Schlerangaben	99
Abbildung 28. Vernderungen durch die Schuassistentz (Fach Mathematik) nach Schlerangaben	100
Abbildung 29. Allgemeine Zufriedenheit mit der Umsetzung der Massnahme Schuassistentz im Rahmen des Pilotprojekts nach Angaben der Lehrpersonen	101
Abbildung 30. Einsatzgebiete der Aufgabenhilfe fr das Fach Deutsch nach Angaben der Lehrpersonen	107
Abbildung 31. Einsatzgebiete der Aufgabenhilfe fr das Fach Mathematik nach Angaben der Lehrpersonen	108
Abbildung 32. Wahrnehmung der Aufgabenhilfe als lernfrderliche Untersttzung im Fach Deutsch nach Angaben der Lehrpersonen.....	110
Abbildung 33. Wahrnehmung der Aufgabenhilfe als lernfrderliche Untersttzung im Fach Mathematik nach Angaben der Lehrpersonen.....	111
Abbildung 34. Wahrgenommene Untersttzung durch die Aufgabenhilfe im Fach Deutsch nach Angaben der Schlerinnen und Schler	113

Abbildung 35. Wahrgenommene Unterstützung durch die Aufgabenhilfe im Fach Mathematik nach Angaben der Schülerinnen und Schüler	113
Abbildung 36. Veränderungen durch die Aufgabenhilfe bezogen auf den Deutschunterricht nach Angaben der Schülerinnen und Schüler	114
Abbildung 37. Veränderungen durch die Aufgabenhilfe bezogen auf den Mathematikunterricht nach Angaben der Schülerinnen und Schüler	114
Abbildung 38. Allgemeine Zufriedenheit mit der Umsetzung der Massnahme Aufgabenhilfe im Rahmen des Pilotprojekts nach Angaben der Lehrpersonen	115
Abbildung 39. Nützlichkeit der Weiterbildungen nach Angaben der Deutsch-Lehrpersonen	120
Abbildung 40. Nützlichkeit der Weiterbildungen nach Angaben der Mathematik-Lehrpersonen.....	120
Abbildung 41. Verbesserungen in Bezug auf den Fachunterricht durch die besuchten Weiterbildungen nach Angaben der Deutsch-Lehrpersonen	121
Abbildung 42. Verbesserungen in Bezug auf den Fachunterricht durch die besuchten Weiterbildungen nach Angaben der Mathematik-Lehrpersonen.....	122
Abbildung 43. Allgemeine Zufriedenheit mit der Umsetzung der Durchführung der Weiterbildungen im Rahmen des Pilotprojekts nach Angaben der Lehrpersonen	123
Abbildung 44. Allgemeine Zufriedenheit mit der Durchführung der Beratungen im Rahmen des Pilotprojekts nach Angaben der Lehrpersonen	131
Abbildung 45. Formen des Erfahrungsaustauschs zu EMU nach Angaben der Schulleitungen.....	138
Abbildung 46. Wirkungen von EMU auf den eigenen Unterricht nach Angaben der Lehrpersonen...	141
Abbildung 47. Wirkungen von EMU auf die eigene Wahrnehmung im Unterricht nach Angaben der Lehrpersonen	143
Abbildung 48. Wirkungen von EMU auf den Schulalltag nach Angaben der Lehrpersonen	144
Abbildung 49. Allgemeine Zufriedenheit mit der Umsetzung der Massnahme EMU im Rahmen des Pilotprojekts nach Angaben der Lehrpersonen	146
Abbildung 50. Veränderungen durch das Pilotprojekt bezogen auf die Individualisierung im Unterricht nach Angaben der Deutsch-Lehrpersonen	147
Abbildung 51. Veränderungen durch das Pilotprojekt bezogen auf die Individualisierung im Unterricht nach Angaben der Mathematik-Lehrpersonen	148
Abbildung 52. Veränderungen durch das Pilotprojekt bezogen auf die Individualisierung des Unterrichts nach Angaben der Schulleitungen.....	148
Abbildung 53. Verbesserte Förderung des Lernens leistungsschwächerer und leistungsstärkerer Schülerinnen und Schüler durch die Lehrpersonen nach Angaben der Schulleitungen und Schulassistenzen	149

Abbildung 54. Veränderungen der Überzeugungen zum Lehren und Lernen in den ALLE- und Vergleichsschulen im Fach Deutsch 156

Abbildung 55. Veränderungen der Überzeugungen zum Lehren und Lernen in den ALLE- und Vergleichsschulen im Fach Mathematik..... 157

Abbildung 56. Veränderung der Qualitätsmerkmale des Unterrichts aus Sicht der Schülerinnen und Schüler in den ALLE-Schulen und Vergleichsschulen im Fach Deutsch..... 177

Abbildung 57. Veränderungen der Qualitätsmerkmale des Unterrichts aus Sicht der Schülerinnen und Schüler in den ALLE-Schulen und Vergleichsschulen im Fach Mathematik 178

Abbildung 58. Veränderungen der Qualitätsmerkmale des Unterrichts aus Sicht der Lehrpersonen in den ALLE-Schulen und Vergleichsschulen im Fach Deutsch 182

Abbildung 59. Veränderungen der Qualitätsmerkmale des Unterrichts aus Sicht der Lehrpersonen in den ALLE-Schule und Vergleichsschulen im Fach Mathematik 183

Abbildung 60. Veränderung der intrinsischen Motivation in den ALLE- und Vergleichsschulen im Fach Deutsch 188

Abbildung 61. Veränderung der anerkennungs- und leistungsbezogenen Motivation in den ALLE- und Vergleichsschulen im Fach Deutsch 188

Abbildung 62. Veränderung der zukunftsbezogenen instrumentellen Motivation in den ALLE- und Vergleichsschulen im Fach Deutsch 189

Abbildung 63. Veränderung der Selbstwirksamkeit in den ALLE- und Vergleichsschulen im Fach Deutsch 189

Abbildung 64. Veränderung der intrinsischen Motivation in den ALLE- und Vergleichsschulen im Fach Mathematik..... 191

Abbildung 65. Veränderung der anerkennungs- und leistungsbezogenen Motivation in den ALLE- und Vergleichsschulen im Fach Mathematik..... 191

Abbildung 66. Veränderung der zukunftsbezogenen instrumentellen Motivation in den ALLE- und Vergleichsschulen im Fach Mathematik..... 192

Abbildung 67. Veränderung der Selbstwirksamkeit in den ALLE- und Vergleichsschulen im Fach Mathematik 192

Abbildung 68. Entwicklung der Testleistungen in den ALLE- und Vergleichsschulen im Fach Deutsch für die Messzeitpunkte der ALLE-Evaluation 199

Abbildung 69. Entwicklung der Testleistungen in den ALLE- und Vergleichsschulen im Kompetenzbereich Lesen für die Messzeitpunkte der ALLE-Evaluation 199

Abbildung 70. Entwicklung der Testleistungen in den ALLE- und Vergleichsschulen im Kompetenzbereich Hören für die Messzeitpunkte der ALLE-Evaluation 200

Abbildung 71. Entwicklung der Testleistungen in den ALLE- und Vergleichsschulen im Kompetenzbereich Sprache(n) im Fokus für die Messzeitpunkte der ALLE-Evaluation	200
Abbildung 72. Entwicklung der Testleistungen in den ALLE- und Vergleichsschulen im Fach Mathematik für die Messzeitpunkte der ALLE-Evaluation	202
Abbildung 73. Entwicklung der Testleistungen in den ALLE- und Vergleichsschulen im Kompetenzbereich Zahl und Variable für die Messzeitpunkte der ALLE-Evaluation.....	202
Abbildung 74. Entwicklung der Testleistungen in den ALLE- und Vergleichsschulen im Kompetenzbereich Form und Raum für die Messzeitpunkte der ALLE-Evaluation.....	203
Abbildung 75. Entwicklung der Testleistungen in den ALLE- und Vergleichsschulen im Kompetenzbereich Grössen, Funktionen, Daten und Zufall für die Messzeitpunkte der ALLE-Evaluation.....	203
Abbildung 76. Zusammenhangsstruktur zwischen fünf beispielhaften Konstrukten mittels graphischer Netzwerkanalyse	218
Abbildung 77. Bivariate Korrelationen zwischen allen 84 berücksichtigten Variablen im Fach Mathematik.....	220
Abbildung 78. Korrelationsnetzwerks mit der Veränderung der Unterrichtsqualitätsvariable Individuelle Unterstützung (Variable «uq_») im Fach Deutsch	221
Abbildung 79. Partielles Korrelationsnetzwerk mit der Veränderung der anerkennungs- und leistungsbezogenen Motivation im Fach Deutsch (Variable «mt_») als abhängige Variable	223
Abbildung 80. Partielles Korrelationsnetzwerk mit der Veränderung der anerkennungs- und leistungsbezogenen Motivation im Fach Mathematik (Variable «mt_») als abhängige Variable.....	224

9 Tabellenverzeichnis

Tabelle 1. Fragebereiche mit Fragestellungen der wissenschaftlichen Evaluation und ihnen zugeordnete Datenquellen	18
Tabelle 2. Übersicht zum Untersuchungsdesign der Evaluation ALLE: Erhebungsinhalte, Erhebungsinstrumente sowie Messzeitpunkte nach den Zielbereichen der Evaluation	23
Tabelle 3. Gesamtstichprobe der Schülerinnen und Schüler und Teilnahmemuster der ALLE- und Vergleichsschulen	34
Tabelle 4. Messzeitpunktspezifische Darstellung der Schülerstichprobe an den ALLE- und Vergleichsschulen mit Angaben zum Niveau für das Fach Deutsch	35
Tabelle 5. Messzeitpunktspezifische Darstellung der Schülerstichprobe an den ALLE- und Vergleichsschulen mit Angaben zum Niveau für das Fach Mathematik	36
Tabelle 6. Längsschnittliche Darstellung der Schülerstichprobe an den ALLE- und Vergleichsschulen für die Fächer Deutsch und Mathematik	37
Tabelle 7. Soziodemografische Merkmale und Eingangsleistungen in Deutsch und Mathematik der Schülerinnen und Schüler in den ALLE- und Vergleichsschulen	39
Tabelle 8. Messzeitpunktspezifische Darstellung der Stichprobe der Lehrpersonen an den ALLE- und Vergleichsschulen für die Fächer Deutsch und Mathematik	40
Tabelle 9. Längsschnittliche Darstellung der Stichprobe der Lehrpersonen an den ALLE- und Vergleichsschulen für die Fächer Deutsch und Mathematik	41
Tabelle 10. Demografische Merkmale der Deutsch- und Mathematik-Lehrpersonen in den ALLE- und Vergleichsschulen	42
Tabelle 11. Zusammensetzung und statistische Kennwerte der Kompetenztests Deutsch und Mathematik für die Messzeitpunkte der Evaluation ALLE	46
Tabelle 12. Fragebereiche der Interviews mit Zuordnung zu den Personengruppen	50
Tabelle 13. Verteilung der Interraterreliabilität (Cohens Kappa nach Brennan & Prediger, 1981) für die Oberkategorien in der Kodierung der Interviews nach Interviewtengruppen	61
Tabelle 14. Nutzung von Möglichkeiten zur Unterstützung des Einsatzes von Schulassistenten durch die Schulleitung nach eigenen Angaben	71
Tabelle 15. Ausbildungsabschlüsse der Schulassistenten nach eigenen Angaben (mehrfache Nennungen möglich)	72
Tabelle 16. Anzahl der Lektionen mit Schulassistenz im Verhältnis zur Anzahl unterrichteter Lektionen pro Woche im Fach Deutsch nach Angaben der Lehrpersonen	75

Tabelle 17. Anzahl der Lektionen mit Schulassistenz im Verhältnis zur Anzahl unterrichteter Lektionen pro Woche im Fach Mathematik nach Angaben der Lehrpersonen.....	76
Tabelle 18. Lehrpersonen pro Schule (fächerübergreifend) mit Angabe «0 Lektionen mit Schulassistenz», geordnet nach Prozentanteil ohne Schulassistenz.....	77
Tabelle 19. Prozentuale Anteile des Einsatzes der Schulassistenzen in Tätigkeitsbereichen im Fach Deutsch nach eigenen Angaben	78
Tabelle 20. Prozentuale Anteile der Schulassistenzen in Tätigkeitsbereichen im Fach Mathematik nach eigenen Angaben	78
Tabelle 21. Sporadisch ausgeübte Tätigkeiten der Schulassistenzen nach eigenen Angaben	85
Tabelle 22. Einsatz der Schulassistenzen in weiteren Lernfeldern nach eigenen Angaben	86
Tabelle 23. Allgemeine Zufriedenheit mit der Umsetzung der Massnahme Schulassistenz bei Lehrpersonen, die eine Schulassistenz hatten	102
Tabelle 24. Allgemeine Zufriedenheit mit der Umsetzung der Massnahme Schulassistenz bei den Schulleitungen.....	102
Tabelle 25. Nutzung von Möglichkeiten zur Unterstützung des Einsatzes der Aufgabenhilfe durch die Schulleitung nach eigenen Angaben.....	104
Tabelle 26. Verfügbarkeit einer Aufgabenhilfe in den einzelnen Schulen für die Schülerinnen und Schüler im Pilotprojekt nach Angaben der Lehrpersonen.....	105
Tabelle 27. Durchführung der Aufgabenhilfe in den einzelnen Schulen nach Angaben der Deutsch-Lehrpersonen	109
Tabelle 28. Durchführung der Aufgabenhilfe in den einzelnen Schulen nach Angaben der Mathematik-Lehrpersonen	109
Tabelle 29. Schülerinnen und Schüler mit Aufgabenhilfe nach Fach und Schule nach eigenen Angaben	112
Tabelle 30. Allgemeine Zufriedenheit mit der Umsetzung der Massnahme Aufgabenhilfe bei Lehrpersonen, die berichten, das eine Aufgabenhilfe für ihre Schülerinnen und Schüler angeboten wird	116
Tabelle 31. Allgemeine Zufriedenheit mit der Umsetzung der Massnahme Aufgabenhilfe bei den Schulleitungen.....	117
Tabelle 32. Besuchte Weiterbildungen im Rahmen des Pilotprojekts nach Lehrpersonen-Angaben	118
Tabelle 33. Anzahl der insgesamt besuchten Weiterbildungen im Rahmen des Pilotprojekts pro Lehrperson nach eigenen Angaben	119
Tabelle 34. Nutzung von Möglichkeiten zur Unterstützung von EMU durch die Schulleitung nach eigenen Angaben	136

Tabelle 35. Häufigkeit des Einsatzes von EMU in den der Befragung vorausgehenden beiden Jahren nach Angaben der Lehrpersonen..... 137

Tabelle 36. Anzahl der von den Lehrpersonen berichteten Wirkungen von EMU 145

Tabelle 37. Skalen zur Erfassung der Überzeugungen zum Lehren und Lernen der Lehrpersonen . 153

Tabelle 38. Mittelwerte für die Überzeugungen zum Lehren und Lernen in den ALLE- und Vergleichsschulen im Fach Deutsch für die Messzeitpunkte der ALLE-Evaluation..... 154

Tabelle 39. Mittelwerte für die Überzeugungen zum Lehren und Lernen in den ALLE- und Vergleichsschulen im Fach Mathematik für die Messzeitpunkte der ALLE-Evaluation 155

Tabelle 40. Aufgabenbezogene Urteilsakkuratheit und Urteilssicherheit von Lehrpersonen in den ALLE- und Vergleichsschulen im Fach Deutsch für den ersten Messzeitpunkt (7. Klassenstufe) 161

Tabelle 41. Aufgabenbezogene Urteilsakkuratheit und Urteilssicherheit von Lehrpersonen in den ALLE- und Vergleichsschulen im Fach Mathematik für den ersten Messzeitpunkt (7. Klassenstufe) 162

Tabelle 42. Personenbezogene Urteilsakkuratheit und Urteilssicherheit von Lehrpersonen in den ALLE- und Vergleichsschulen im Fach Deutsch 164

Tabelle 43. Personenbezogene Urteilsakkuratheit und Urteilssicherheit von Lehrpersonen in den ALLE- und Vergleichsschulen im Fach Mathematik..... 165

Tabelle 44. Unterschiede zwischen ALLE- und Vergleichsschulen in der personenbezogenen Urteilsakkuratheit und Urteilssicherheit am Ende der Evaluation (9. Klassenstufe; T4)..... 166

Tabelle 45. Skalen zur Erfassung der Unterrichtsqualität aus Sicht der Lehrpersonen 168

Tabelle 46. Skalen zur Erfassung der Unterrichtsqualität aus Sicht der Schülerinnen und Schüler .. 169

Tabelle 47. Unterrichtsqualität im Fach Deutsch aus Sicht der Schülerinnen und Schüler in den ALLE-Schulen und Vergleichsschulen für die Messzeitpunkte der ALLE-Evaluation..... 172

Tabelle 48. Unterrichtsqualität im Fach Mathematik aus Sicht der Schülerinnen und Schüler in den ALLE-Schulen und Vergleichsschulen für die Messzeitpunkte der ALLE-Evaluation..... 175

Tabelle 49. Unterrichtsqualität im Fach Deutsch aus Sicht der Lehrpersonen in den ALLE-Schulen und Vergleichsschulen für die Messzeitpunkte der ALLE-Evaluation..... 180

Tabelle 50. Unterrichtsqualität im Fach Mathematik aus Sicht der Lehrpersonen in den ALLE-Schulen und Vergleichsschulen für die Messzeitpunkte der ALLE-Evaluation..... 181

Tabelle 51. Skalen zu den motivationalen Orientierungen und zur Selbstwirksamkeit der Schülerinnen und Schüler 185

Tabelle 52. Motivational-affektive Lernergebnisse in den ALLE- und Vergleichsschulen im Fach Deutsch für die Messzeitpunkte der ALLE-Evaluation..... 187

Tabelle 53. Motivational-affektive Lernergebnisse in den ALLE- und Vergleichsschulen im Fach Mathematik für die Messzeitpunkte der ALLE-Evaluation 190

Tabelle 54. Unterschiede in der Veränderung zwischen ALLE- und Vergleichsschulen für die gesamte Evaluationsstichprobe in den motivational-affektiven Lernergebnissen im Fach Deutsch 193

Tabelle 55. Unterschiede in der Veränderung zwischen ALLE- und Vergleichsschulen für die gesamte Evaluationsstichprobe in den motivational-affektiven Lernergebnissen im Fach Mathematik..... 194

Tabelle 56. Unterschiede in der Veränderung zwischen ALLE- und Vergleichsschulen bei Schülerinnen und Schülern mit ungünstiger Lernausgangslage in den motivational-affektiven Lernergebnissen im Fach Deutsch..... 195

Tabelle 57. Unterschiede in der Veränderung zwischen ALLE- und Vergleichsschulen bei Schülerinnen und Schülern mit ungünstiger Lernausgangslage in den motivational-affektiven Lernergebnissen im Fach Mathematik 196

Tabelle 58. Lernergebnisse in den ALLE- und Vergleichsschulen im Fach Deutsch für die Messzeitpunkte der ALLE-Evaluation 198

Tabelle 59. Lernergebnisse in den ALLE- und Vergleichsschulen im Fach Mathematik für die Messzeitpunkte der ALLE-Evaluation 201

Tabelle 60. Unterschiede im Lernzuwachs zwischen den ALLE- und Vergleichsschulen für die gesamte Evaluationsstichprobe im Fach Deutsch 205

Tabelle 61. Unterschiede im Lernzuwachs zwischen den ALLE- und Vergleichsschulen für die gesamte Evaluationsstichprobe im Fach Mathematik..... 206

Tabelle 62. Unterschiede im Lernzuwachs zwischen ALLE- und Vergleichsschulen für Schülerinnen und Schüler mit ungünstiger Lernausgangslage im Fach Deutsch und im Fach Mathematik..... 208

Tabelle 63. Unterschiede im Lernzuwachs zwischen ALLE- und Vergleichsschulen für Schülerinnen und Schüler in den Niveaus B/II und C/III (zusammengefasst) in den Fächern Deutsch und Mathematik 210

Tabelle 64. Unterschied in der Testleistung zwischen ALLE- und Vergleichsschulen für den Kompetenzbereich «Texte schreiben»..... 211

Tabelle 65. ICCs der Unterrichtsprozess- und Ergebnisvariablen auf Schul- und Klassenebene..... 213

10 Abkürzungsverzeichnis

Statistische Kennwerte:

<i>d</i>	Cohens <i>d</i> : Mass für die Grösse eines Effekts
<i>M</i>	Arithmetischer Mittelwert
<i>max</i>	Grösster vorkommender Wert
<i>min</i>	Kleinster vorkommender Wert
<i>N</i>	Anzahl der Personen, Stichprobengrösse
<i>p</i>	Kennwert der statistischen Signifikanz
<i>r</i>	Pearson-Korrelationskoeffizient
<i>R</i> ²	Determinationskoeffizient (Mass erklärter Varianz)
<i>SD</i>	Standardabweichung (Streuung)
<i>SE</i>	Standardfehler
<i>α</i>	Cronbachs Alpha (Reliabilitätsmass)

Weitere Abkürzungen:

1. Sek	7. Klassenstufe
2. Sek	8. Klassenstufe
3. Sek	9. Klassenstufe
A/I	Höchste Niveaustufe
AAI IP- Erziehung	AAI: Alfred Adler Institut, IP: Individualpsychologische Erziehung
AH.....	Aufgabenhilfe
ALLE.....	Aktive Lernzeit und Lernerfolg für ALLE
Anm.	Anmerkung
B/II	Zweithöchste Niveaustufe
BP.....	Beratungs- und Weiterbildungspersonen
bspw.	beispielsweise
C/III	Dritthöchste Niveaustufe
CS.....	Credit Suisse
D	Deutsch
d.h.....	das heisst

DaZ	Deutsch als Zweitsprache
DIF	Differential Item Functioning
EFZ	Eidgenössisches Fähigkeitszeugnis
EMU	Evidenzbasierte Methoden der Unterrichtsdiagnostik und -entwicklung
ETS	Educational Testing Service
ff	fortfolgend
GPCM	Generalisiertes Partial-Credit-Modell
HfH	Hochschule für Heilpädagogik
I	Interviewende/-r
IBE	Institut für Bildungsevaluation
IRT	Item-Response-Theorie
ISR	Integrierte Sonderschulung
Kap.	Kapitel
KB	Kompetenzbereich
KV	kaufmännische Grundbildung
LP	Lehrperson
M	Mathematik
OECD	Organisation for Economic Co-operation and Development
OS	Oberstufe
PHSG	Pädagogische Hochschule St.Gallen
PISA	Programme for International Student Assessment
PLS	Partial-Least-Squares-Methode
PS	Primarstufe
PTT	Post-, Telefon- und Telegrafienbetriebe
PV	Plausible Value
Q	Jahresquartale
QUIMS	Qualität in multikulturellen Schulen, Projekt der BD Zürich
S	Seite
SA	Schulassistenz
Sek.	Sekundarstufe
SHP	Schulische Heilpädagogin/schulischer Heilpädagoge
SL	Schulleitung
SRK	Schweizerisches Rotes Kreuz
SuS; SUS	Schülerinnen und Schüler

T1 Messzeitpunkt 1: Anfang 7. Klasse
T2 Messzeitpunkt 2: Ende 7. Klasse
T3 Messzeitpunkt 3: Ende 8. Klasse
T4 Messzeitpunkt 4: Ende 9. Klasse
u.a. unter anderem
ÜGK..... Überprüfung der Grundkompetenzen
vgl. vergleiche
WB..... Weiterbildung
z.B. zum Beispiel
ZH..... Zürich