



Die Wirkung von unterschiedlichen Formen von co-teaching auf die von den Schülerinnen und Schülern wahrgenommene Unterrichtsqualität in inklusiven Klassen

Elisabeth Moser Opitz · Katharina Maag Merki · Rico Pfaffhauser · Meret Stöckli · Ariana Garrote

Eingegangen: 8. Mai 2020 / Überarbeitet: 23. März 2021 / Angenommen: 25. März 2021
© Der/die Autor(en) 2021

Zusammenfassung Die Zusammenarbeit der Lehrkräfte und damit verbunden co-teaching gilt als zentrale Bedingung für die Durchführung von gelingendem inklusiven Unterricht. Allerdings fehlen bislang längsschnittliche Studien zur Frage, ob und wie Schüler*innen unterschiedliche Formen von co-teaching sowie deren Auswirkungen auf Merkmale der Unterrichtsqualität wahrnehmen. In der Studie wird längsschnittlich untersucht, inwiefern sich die Wahrnehmung der Lernunterstützung (Beziehung Lehrkraft-Schüler*innen, Lernbegleitung, Lernklima) und der Klassenführung von 621 Lernenden aus 36 Grundschulklassen in Abhängigkeit zweier verschiedener Formen von co-teaching im Verlauf eines Schuljahres verändert. In der Experimentalgruppe ($n=331$, 18 Klassen) wurden die Lehrkräfte zum co-teaching verpflichtet und Unterstützungsmaßnahmen (sonderpädagogische Unterstützung, Deutsch als Zweitsprache DaZ) wurden konsequent unterrichtsintegriert von den Regelklassenlehrkräften angeboten. Fachkräfte (DaZ, Sonderpädagogik) waren ausschließlich beratend tätig. In der Kontrollgruppe ($n=290$, 18 Klassen aus denselben Schulgemeinden wie in der Experimentalgruppe) gab es weder eine Verpflichtung zum gemeinsamen Unterrichten noch Vorgaben für ein bestimmtes Unterrichtsetting. Die direkte Arbeit der Fachkräfte mit den Schüler*innen war möglich. Die Schüler*innen wurden am Ende der zweiten und dritten Klasse zur Wahrnehmung der Lernunterstützung und der Klassenführung befragt. Mehrebenenanalytische Auswertungen zeigten, dass sich die Wahrnehmung der Lernunterstützung und der Klassenführung in der Experimentalgruppe negativer entwickelt hatte als in der Kontroll-

Die Autorinnen E. Moser Opitz und K. Maag Merki teilen sich die Erstautorinnenschaft.

E. Moser Opitz (✉) · K. Maag Merki · R. Pfaffhauser · M. Stöckli
Institut für Erziehungswissenschaft, Universität Zürich, Freiestrasse 36, 8032 Zürich, Schweiz
E-Mail: elisabeth.moseropitz@uzh.ch

A. Garrote

Institut Forschung und Entwicklung, Zentrum Lernen und Sozialisation, Fachhochschule Nordwestschweiz, Obere Sternengasse 7, 4502 Solothurn, Schweiz

gruppe. Der Effekt bei der Klassenführung war am stärksten, jener beim Lernklima am geringsten und nur noch tendenziell signifikant ($p = 0,067$). Diese Ergebnisse könnten auf mangelnde Absprachen und unklare Verantwortlichkeiten, auf Unruhe bei klassenintegriertem Arbeiten, oder auf eine fehlende individuelle und spezifische Förderung der Schüler*innen zurückgeführt werden.

Schlüsselwörter Co-teaching · Inklusiver Unterricht · Klassenführung · Lernunterstützung

How pupils perceive teaching quality in inclusive classrooms: the impact of different co-teaching arrangements

Abstract Co-teaching is considered to be a crucial component of inclusive education. To date there are no longitudinal studies on pupils' perception of co-teaching and the impact of co-teaching on teaching quality. This longitudinal study examines the change in the perception of learning support (teacher-pupil relationship, individual support, classroom environment) and classroom management of 621 third graders over the course of a year in two different co-teaching settings. In the experimental group ($n=331$, 18 classes), two general education teachers who were also responsible for special educational needs support and teaching German as a second language were asked to co-teach together as much as possible. Access to advice from specialists (special education, German as second language) was available. In the control group ($n=290$, 18 classes), there were no set parameters for co-teaching and a second teacher (special education or German as second language), taught some children individually. Pupils answered questionnaires on their perceptions of learning support and classroom management twice, at the end of Grade 2 and at the end of Grade 3. Multi-level analysis showed that students in the experimental group formed a significantly more negative perception of learning support and classroom management than those in the control group. This effect was strong for classroom management, but lower and not significant ($p=0,067$) for classroom climate. This could be due to an unclear division of labor, a restless study atmosphere in the classroom when both teachers were present, or a lack of individual and specific support of the pupils.

Keywords Co-teaching · Inclusive education · Classroom management · Learning support · Longitudinal study

1 Einleitung

Die Umsetzung von inklusivem Unterricht ist eine zentrale bildungspolitische Zielsetzung. Allerdings sind bezüglich seiner Gestaltung und Organisation sowie seiner Auswirkungen auf die Schüler*innen noch viele Fragen offen (z.B. Cook et al. 2017). Unklar ist beispielsweise, welche Faktoren für einen gelingenden inklusiven Unterricht ausschlaggebend sind, in dem sich die Schüler*innen unterstützt fühlen sowie motiviert und erfolgreich lernen können. Viele Autor*innen (z.B. Hell-

mich et al. 2017; Lütje-Klose und Urban 2014; Shaffer und Thomas-Brown 2015) erachten die unterrichtsbezogene Zusammenarbeit der Lehrkräfte – im Folgenden synonym als *co-teaching* bezeichnet – als bedeutsame Voraussetzung für die erfolgreiche Umsetzung von inklusivem Unterricht. Es ist allerdings nicht klar, ob und wie sich unterschiedliche Formen von co-teaching auf Merkmale von Unterrichtsqualität auswirken. Erkenntnisse dazu sind insbesondere deshalb zentral, weil es Hinweise gibt, dass sich die Zusammenarbeitsform „paralleles Unterrichten“ (Cook und Friend 1995), d. h. das regelmäßig getrennte Unterrichten von Lernenden mit und ohne besonderen Förderbedarf in unterschiedlichen Räumen, negativ auf die soziale Integration von Schüler*innen mit Förderbedarf – und auch auf Lernende mit Deutsch als Zweitsprache – auswirken kann (z. B. Grosche und Fleischhauer 2017; Wiener und Tardif 2004).

Zudem ist bislang nicht längsschnittlich untersucht worden, ob und wie die Schüler*innen unterschiedliche bzw. unterschiedlich intensive Formen des co-teachings und deren Auswirkungen auf verschiedene Merkmale von Unterrichtsqualität wahrnehmen. Zur Schließung dieser Forschungslücke soll dieser Artikel einen Beitrag leisten. In einer Längsschnittstudie mit einem quasi-experimentellen Kontrollgruppendesign in Primarklassen wurde untersucht, wie sich die Wahrnehmung der Schüler*innen zur Unterrichtsqualität (Klassenführung, Lernunterstützung) in Abhängigkeit von zwei unterschiedlichen Formen des co-teachings verändert hat.

2 Co-teaching: Bedeutung und Auswirkungen auf die Schülerinnen und Schüler

Die Zusammenarbeit von Lehrkräften – insbesondere jene zwischen der Regelklassenlehrkraft und der sonderpädagogischen Lehrkraft – wird seit Beginn der integrativen Bestrebungen in den 1980er-Jahren als zentrale Bedingung für gelingenden inklusiven Unterricht betrachtet (z. B. Hellmich et al. 2017; Lütje-Klose und Urban 2014; Shaffer und Thomas-Brown 2015), insbesondere hinsichtlich der Unterstützung für Lernende mit sonderpädagogischem Förderbedarf. Gemäß den konzeptionellen Überlegungen von Hinz und Köpfer (2016) sollen im inklusiven Unterricht Teams von Lehrkräften Unterstützung innerhalb der Klasse anbieten, ohne dass eine etikettierende Einteilung von Schüler*innen in verschiedene Kategorien von Förderbedarfen geschieht (ebd.).

Studien, die sich mit der Auswirkung von Kooperation auf die Lernenden befassen, gibt es vor allem im englischen Sprachraum unter dem Begriff *co-teaching*. Meistens ist damit das gemeinsame Unterrichten einer Regelklassen- und einer sonderpädagogischen Lehrkraft in unterschiedlichen und unterschiedlich intensiven Formen von Zusammenarbeit gemeint (Cook und Friend 1995; Scruggs et al. 2007; Cook et al. 2017). Die intensivste Form der Zusammenarbeit ist *team teaching*, bei dem die beteiligten Lehrkräfte die Verantwortlichkeiten gleichmäßig teilen und auch gleichberechtigt und in verschiedenen Formen an der Leitung des Unterrichts beteiligt sind. Beim Lernen an Stationen (*station teaching*) unterstützen beide Lehrkräfte die Schüler*innen, die an verschiedenen Posten arbeiten. Paralleles Unterrichten (*parallel teaching*) bedeutet, dass die Lehrkräfte in unterschiedlichen Gruppen die

selben Inhalte unterrichten. Beim alternativen Unterrichten (*alternative teaching*) arbeitet eine Lehrkraft mit einer Gruppe von Lernenden an anderen Inhalten, im Klassenzimmer oder an einem anderen Lernort. Eine Person unterrichtet, die andere assistiert (*one teach/one assist*) schließlich beinhaltet eine Form der Zusammenarbeit, bei der eine Lehrkraft – in der Regel die Regelklassenlehrkraft – den Unterricht leitet und eine zweite Lehrkraft (die sonderpädagogische Lehrkraft) einzelne Lernende unterstützt. Baeten und Simons (2014) unterscheiden zusätzlich zwischen dem Modell *one teach, one observe*, in dem eine Lehrkraft unterrichtet und die andere beobachtet und dem *coaching* Modell, bei dem eine Lehrkraft durch eine andere angeleitet wird.

Scruggs et al. (2007) haben anhand eines Forschungsüberblicks drei zentrale Erkenntnisse zum co-teaching herausgearbeitet. Erstens erwies sich *one teach/one assist* als die dominante Form der Zusammenarbeit, mit der Regellehrkraft in der Rolle der unterrichtenden Lehrkraft. Diese Dominanz von *one teach/one assist* zeigte sich auch in den Studien von Pancsofar und Petroff (2016) und Spörer et al. (2021). Zweitens wurde co-teaching mit seinen verschiedenen Formen von den Beteiligten (Schüler*innen, Lehrkräften, Schulverantwortlichen) positiv wahrgenommen (Scruggs et al. 2007). Drittens zeigte sich, dass für co-teaching Rahmenbedingungen wie genügend Zeit für die Planung sowie geeignete Anleitungen und Trainings notwendig sind (ebd., vgl. auch Pancsofar und Petroff 2016). Baeten und Simons (2014) nennen als Ergebnis ihrer Literatur-Review mehrere Vorteile von co-teaching, ohne allerdings systematisch zwischen verschiedenen Formen des co-teaching zu unterscheiden: Die Lernenden erhalten mehr Unterstützung und individuelle Aufmerksamkeit, sie erleben verschiedene Lehrstile, es besteht mehr Möglichkeiten zur Differenzierung und zwei Lehrkräfte können besser beobachten als nur eine. Als Nachteil wird genannt, dass es für die Schüler*innen verwirrend sein kann, wenn zwei Lehrkräfte im Zimmer sind und beispielsweise auf Fragen unterschiedliche Antworten geben. Weitere Studien verweisen ebenfalls auf ungünstige Effekte und zeigen insgesamt ein eher negatives Bild. So berichten Vangrieken et al. (2015) in einer systematischen Review zur Zusammenarbeit von Regelklassenlehrkräften, dass im Rahmen von co-teaching häufig eher oberflächliche Formen von Zusammenarbeit wie zum Beispiel Besprechungen realisiert würden. Zu einem ähnlichen Ergebnis kommen Scruggs et al. (2007), bezogen auf die Zusammenarbeit von Regelklassenlehrkräften und sonderpädagogischen Lehrkräften. Sie stellen fest, dass effektive Unterrichtsformen wie zum Beispiel Unterstützung durch Peers oder Strategieinstruktion in Schulstunden mit co-teaching kaum beobachtet werden konnten. Cook et al. (2017) weisen darauf hin, dass Chancen von co-teaching nicht wahrgenommen würden, weil auch beim co-teaching traditionelle Unterrichtsformen dominant seien (Klassenunterricht, der von der Regelklassenlehrkraft geleitet wird, die sonderpädagogische Lehrkraft in einer assistierenden Funktion). Gemäß Chitiyo (2017) gaben in inklusiven Settings kooperierende Lehrkräfte an, dass ihnen die notwendigen Fähigkeiten für co-teaching fehlen würden. In einigen Studien wurden die Lernenden zur Zusammenarbeit der Lehrkräfte im Rahmen von co-teaching befragt. Keeley et al. (2017) und Keeley (2015) berichten, dass die Zusammenarbeit von Regelklassen- und sonderpädagogischer Lehrkraft im Modell *one teach, on assist* nach Aussagen der Schüler*innen ineffektiv ist hinsichtlich der

Verantwortlichkeiten der Lehrkräfte und hinsichtlich einer ausgeglichenen Klassenführung. Positivere, allerdings durchaus auch ambivalente Befunde, zeigt die Studie von Schwab (2017), in der Lernende im vierten Schuljahr befragt wurden, wie sie das gemeinsame Unterrichten von Regelklassenlehrkraft und sonderpädagogischer Lehrkraft wahrnehmen. Zwei Drittel der befragten Lernenden fühlten sich durch die Zusammenarbeit von zwei Lehrkräften unterstützt, ein Drittel formulierte hingegen eine negative Wahrnehmung bezüglich der Unterstützung. Darüber hinaus gab es einen positiven Zusammenhang zwischen co-teaching und der Einstellung zu Gruppen- und Partnerarbeiten. Je mehr die Schüler*innen es schätzten, von einem Team von Lehrkräften unterrichtet zu werden und je unterstützender sie dies wahrnahmen, desto mehr mochten sie Gruppen- und Partnerarbeiten, und je stärker sich Schüler*innen durch Teamarbeit der Lehrkräfte unterstützt fühlten, desto besser fühlten sie sich emotional und sozial integriert.

Zur Wirkung von verschiedenen co-teaching-Modellen auf die Lernenden liegen ebenfalls nur einzelne Ergebnisse vor. In einer schon älteren Studie von Vaughn et al. (1998) wurde der Einfluss von verschiedenen Zusammenarbeitsmodellen auf die soziale Integration (Peer-Akzeptanz, Selbstkonzept, reziproke Freundschaften) von Lernenden mit und ohne besonderen Förderbedarf (als „learning disabilities“ bezeichnet) längsschnittlich analysiert. Verglichen wurden ein Modell, in dem eine Regel- und eine sonderpädagogische Lehrkraft vollzeitlich zusammenarbeiteten (co-teaching) sowie ein Modell, in dem die sonderpädagogische Lehrkraft während ein bis zwei Stunden pro Tag in der Klasse anwesend war, zusätzlich zu einer Klassenassistentin, die während vier Stunden pro Woche mitarbeitete (*collaboration*). Das Setting *collaboration* erwies sich als wirksamer für die soziale Integration der Kinder (insbesondere auch der Kinder mit besonderem Förderbedarf) als das co-teaching. Erklärt wurden die Unterschiede allerdings nicht mit der Variable Zusammenarbeit. Im Modell co-teaching war die Anzahl von Schüler*innen mit Förderbedarf pro Klasse sehr groß und Unterrichtsbeobachtungen und Interviews wiesen darauf hin, dass sich die Lehrkräfte vorwiegend an den schwächeren Lernenden orientierten und der Unterricht von niedrigeren Erwartungen an die Kinder geprägt war. Eine Studie aus dem deutschsprachigen Raum von Spörer et al. (2021) zeigte, dass in Schulstunden mit co-teaching im Vergleich zum Unterricht mit einer Lehrkraft mehr Möglichkeiten zur Interaktion mit den anwesenden Personen bestand. Allerdings kommunizierten Schüler*innen mit besonderem Förderbedarf häufiger mit der sie begleitenden Lehrkraft als mit anderen Kindern. Die Metaanalyse zum Thema co-teaching von Murawski und Swanson (2001) ergab kleine positive Effekte bezüglich der Auswirkungen auf die Leistungen der Schüler*innen sowie auf verschiedene soziale Aspekte. Allerdings beziehen sich die berichteten Effekte auf sehr unterschiedliche Konstrukte und sind deshalb mit Vorsicht zu betrachten. Zudem zeigte die Metaanalyse von Cook et al. (2017), dass die bisherigen Studien Mängel aufweisen und beispielsweise (quasi-)experimentelle Studien fehlen.

Zusammenfassend weisen die Forschungsergebnisse darauf hin, dass das co-teaching im inklusiven Unterricht häufig auf einer eher oberflächlichen bzw. organisatorischen Ebene stattzufinden scheint und nicht als eine evidenz-basierte Praxis für einen erfolgreichen Unterricht bezeichnet werden kann. Zur Auswirkung von verschiedenen Formen von co-teaching auf die Entwicklung von zentralen Merkmalen

von gutem Unterricht wie Klassenführung oder unterstützendes Klima liegen keine empirischen Befunde vor. Zudem fehlen fundierte (quasi-)experimentelle Studien (Cook et al. 2017), um Aussagen über die Wirkung von unterschiedlich intensiven Formen von co-teaching auf die Entwicklung der Unterrichtsqualität machen zu können.

3 Lernunterstützung und Klassenführung als Merkmale von Unterrichtsqualität

Untersuchungen mit Regelklassen haben gezeigt, dass die drei Basisdimensionen Klassenführung, Lernunterstützung bzw. unterstützendes Klima und kognitive Aktivierung zentrale Merkmale von gutem Unterricht und somit wichtig für den Lernerfolg und das Wohlbefinden der Schüler*innen sind (z.B. Fauth et al. 2014; Praetorius et al. 2018). Eine gute Klassenführung gibt den Schüler*innen möglichst viel ungestörte Lernzeit. Bei der Lernunterstützung stehen die Wahrnehmung und die Qualität der sozialen Interaktionen (z.B. Sensibilität der Lehrkräfte für die individuellen Bedürfnisse, konkrete Unterstützung durch die Lehrkraft, Vermeiden von Leistungsdruck) im Zentrum. Bei der kognitiven Aktivierung geht es schließlich um herausfordernde Aufgaben und die Förderung der aktiven Auseinandersetzung mit dem Lerngegenstand (Praetorius et al. 2018).

Klassenführung und Lernunterstützung werden auch für den inklusiven Unterricht als zentral erachtet. Allerdings orientieren sich die entsprechenden Studien nicht ausschließlich an der Systematik der Basisdimensionen und werden mit weiteren Qualitätsmerkmalen ergänzt. Zudem wird deren Wirksamkeit häufig nicht empirisch überprüft. So zeichnet sich nach Jordan und McGhie-Richmond (2014) qualitativ hochstehender inklusiver Unterricht zum einen durch spezifische Fördermaßnahmen aus, die sich an den Prinzipien des *Universal Designs for Learning* (CAST 2011) orientieren. Dazu gehört beispielsweise, dass Informationen – angepasst an die Lernvoraussetzungen der Schüler*innen – auf verschiedene Art und Weise angeboten oder spezifische Hilfsmittel eingesetzt werden. Zum anderen werden von Jordan und McGhie-Richmond (2014) Variablen genannt, die inhaltlich teilweise in Übereinstimmung mit den Basisdimensionen stehen: gutes Klassen- und Zeitmanagement, das Fokussieren auf die zentralen Lerninhalte verbunden mit klaren und klar kommunizierten Zielsetzungen, der Einsatz von Scaffolding, eine Lernbegleitung, welche die Lernenden herausfordert und deren aktive Teilnahme einfordert sowie ein Unterrichtsklima, das einen respektvollen Umgang beinhaltet (ebd.).

Im Diskurs um die Gestaltung von inklusivem Unterricht werden häufig gemeinsame Lernsituationen als wichtiges Qualitätsmerkmal gefordert. Diese zeichnen sich nach Feuser (2018, S. 283 f.) dadurch aus, dass „*alle* Schüler miteinander arbeitsteilig und in Kooperation miteinander an der Realisierung eines gemeinsamen Produktes arbeiten“ (Hervorhebung im Original). Köpfer (2014) beschreibt eine inklusive Schule als einen sozialen Raum, der nicht durch die physische Anwesenheit von Personen, sondern durch gemeinschaftliche Aktivitäten und somit durch soziale Interaktionen definiert ist. Inklusiver Unterricht wird somit als qualitativ gut betrachtet, wenn alle Kinder in Kooperation miteinander lernen können und Schüler*innen mit

besonderem Förderbedarf am Klassenunterricht teilnehmen. Dieses gemeinsame Tun betrifft auch die Lehrkräfte und erfordert nach Feuser (2018) eine Zusammenarbeit im Sinn eines „gleichberechtigten Team-Teachings“ (ebd., S. 280). Grosche et al. (2020, S. 471) beschreiben eine solche Zusammenarbeit als Kokonstruktion, bei der „fortwährende und zyklische *Aushandlungen* und *Reflexionen* von Konkretisierungen einer Innovation“ stattfinden (Hervorhebung im Original).

4 Forschungsdesiderata und Fragestellung

Die Ausführungen haben gezeigt, dass hinsichtlich der Bedeutung und Auswirkungen von co-teaching und insbesondere von unterschiedlich intensiven Formen des co-teaching im inklusiven Unterricht viele Fragen offen sind. Insbesondere fehlen Studien, die mittels eines quasi-experimentellen Designs längsschnittlich untersuchen, ob und wie sich unterschiedliche Formen der Zusammenarbeit der Lehrkräfte auf die Wahrnehmung der Unterrichtsqualität durch die Schüler*innen auswirken. Die vorliegende quasi-experimentelle Studie mit einer Experimental- und einer Kontrollgruppe, die im Rahmen der Evaluation eines Schulversuchs durchgeführt wurde (vgl. Abschn. 5.1), soll einen Beitrag zur Schließung dieser Forschungslücke leisten. In der Experimentalgruppe fand ein hoher Anteil an co-teaching statt und Unterstützungsmaßnahmen (sonderpädagogische Unterstützung, DaZ) wurden konsequent unterrichtsintegriert von den Regelklassenlehrkräften angeboten; diese erhielten Beratung durch eine sonderpädagogische Lehrkraft und eine DaZ-Lehrkraft. In den Kontrollklassen hingegen arbeiteten die Lehr- und Fachkräfte zusammen, ohne dass es eine Verpflichtung zum gemeinsamen Unterrichten oder Vorgaben für ein bestimmtes Unterrichtssetting (integriert oder separiert) gab. Im Gegensatz zur Experimentalgruppe arbeiteten die Fachkräfte auch direkt mit den Schüler*innen.

Im Zentrum steht folgende Fragestellung:

Inwiefern zeigen sich zwischen der Experimental- und der Kontrollgruppe bei den Schüler*innen Unterschiede bezüglich der Veränderung der wahrgenommenen Klassenführung und Lernunterstützung (Beziehung Lehrkraft-Schüler*in, Lernbegleitung, Lernklima)?

Der Fokus liegt mit Lernunterstützung und Klassenführung nur auf zwei Basisdimensionen. Die Dimension „kognitive Aktivierung“ konnte aufgrund der Zielsetzungen des Schulversuchs (Fokus auf Lernunterstützung und Lernbeziehung) nicht untersucht werden bzw. es war nicht möglich, im Rahmen der Evaluation zusätzliche Variablen aufzunehmen.

Aufgrund des Forschungsstandes zu den ambivalenten Auswirkungen des co-teachings auf die Schüler*innen (vgl. Abschn. 2) kann keine eindeutige Hypothese formuliert werden. Entsprechend den bildungspolitischen Vorgaben des Schulversuchs wurde angenommen, dass sich die Wahrnehmung der Lernunterstützung und Klassenführung durch die Schüler*innen in der Experimentalgruppe im Vergleich zu jenen in der Kontrollgruppe positiver entwickelt. Aufgrund der Komplexität des co-teachings könnte aber als alternative Hypothese vermutet werden, dass sich die Wahrnehmung der Lernunterstützung und Klassenführung durch die Schüler*innen

in der Experimentalgruppe im Vergleich zu jenen in der Kontrollgruppe eher negativer entwickelt.

5 Methode

5.1 Rahmenbedingungen der inklusiven Schulung in der Schweiz und Merkmale des Schulversuchs

Für Schüler*innen mit dem Förderschwerpunkt Lernen und emotionale und soziale Entwicklung (z. B. ADHS) sowie mit einer Lese-Rechtschreib- oder Rechenschwäche wird in der Schweiz sonderpädagogische Unterstützung in Form von sogenannt integrativer Förderung (IF) und DaZ-Unterricht angeboten. In der Regel stehen für eine Klasse pro Woche zwei bis drei Schulstunden sonderpädagogische Unterstützung zur Verfügung, ohne dass eine offizielle Diagnosestellung für einzelne Kinder vorliegen muss. Die Mittel dafür werden den Schulen nach einem kantonalen Schlüssel aufgrund der Gesamtzahl der Schüler*innen (und zum Teil eines Sozialindex) zugeteilt und die Schulen bzw. Schulleitungen entscheiden, wie sie die Mittel einsetzen wollen. Zusätzlich stehen DaZ-Stunden zur Verfügung, die auch nach einem entsprechenden Schlüssel zur Verfügung gestellt wurden. Zudem werden an den Schulen Sprachheil- und Psychomotoriktherapie angeboten. Diese Unterstützungsmaßnahmen finden in Mischformen von unterrichtsintegrierter und separierter Förderung statt. Insbesondere wenn wenig Unterstützung zur Verfügung steht, dominiert jedoch die separierte Förderung (Pool Maag und Moser Opitz 2014). Ausgebildete Lehrkräfte für DaZ-Unterricht stehen an den Schulen zur Verfügung. Sonderpädagogische Lehrkräfte fehlen jedoch oft und deren Aufgaben werden häufig von Regellehrkräften wahrgenommen. Im Folgenden wird deshalb für die sonderpädagogische Lehrkraft die Bezeichnung „eine weitere (sonderpädagogisch ausgebildete) Lehrkraft“ verwendet.

Dieses Modell mit der zusätzlichen Unterstützung durch mehrere Fachpersonen führte teilweise dazu, dass sehr viele Lehrkräfte und Fachpersonen an einer Klasse mitarbeiteten. Deshalb wurde in einem Schweizer Kanton ein Schulversuch initiiert. Dessen Ziel war deshalb laut der verantwortlichen Behörde, die Anzahl der Lehrkräfte an einer Klasse zu reduzieren und den Schüler*innen der beteiligten inklusiven Klassen eine konstante und vertiefte Lernbegleitung in Form eines intensiven und klassenintegrierten co-teaching zur Verfügung zu stellen. Es wurde angenommen, dass sich dies positiv auf das Lernverhalten, die Persönlichkeitsentwicklung sowie die Sozial- und Kommunikationskompetenz der Schüler*innen auswirken würde. Zudem wurde angestrebt, die separierte Schulung von Schüler*innen mit sonderpädagogischem Förderbedarf und von Lernenden, die Deutsch als Zweitsprache lernen, zu vermeiden und grundsätzlich eine klassenintegrierte Förderung anzubieten. Um diese Zielsetzung zu erreichen, standen den Klassen der Experimentalgruppe zwei Klassenlehrkräfte zur Verfügung, damit diese während einer bestimmten Anzahl von Stunden pro Woche in Doppelbesetzung unterrichten konnten. Sie hatten explizit die Aufgabe, durch intensives und klassenintegriertes co-teaching gezielte Individualisierungsmaßnahmen zu treffen, insbesondere auch für Lernende mit Deutsch als

Zweitsprache, mit Lern- und Verhaltensschwierigkeiten sowie mit Hochbegabung. Fachkräfte (Sonderpädagogik, DaZ) standen den Klassenlehrkräften ausschließlich beratend zur Seite und sie arbeiteten nicht mit der Klasse oder mit einzelnen Schüler*innen.

Die Klassenteams der Experimentalgruppe wurden im Gegensatz zum im ersten Abschnitt beschriebenen üblichen Modell, in dem es keine Vorgaben zum co-teaching gibt, zu einer intensiven Form des co-teaching verpflichtet. Sie mussten eine Zusammenarbeitsvereinbarung abschließen, in der inhaltliche (pädagogische Ziele, besondere Kompetenzen) und formale Aspekte (Form der Zusammenarbeit) festgehalten wurden. Die Versuchsschulen erhielten zudem ein auf sie zugeschnittenes Fortbildungsangebot, die Fortbildungen wurden von Dozierenden für Sonderpädagogik erteilt. Gemäß der online-Befragung aller Lehrkräfte des Schulversuchs bestand an den Schulen insbesondere Fortbildungsbedarf bezüglich der Themen Förderplanung und unterrichtsintegrierte Sprachförderung. Deutlich weniger notwendig erachtet wurden Kurse zum Unterrichten im Team und zu kooperativen und offenen Unterrichtsmethoden. Die Fortbildungen, die zum Thema Teamteaching stattfanden, wurden insgesamt als nützlich beurteilt. In den Klassen der Kontrollgruppe galten die vorgängig beschriebenen üblichen Rahmenbedingungen. Es standen sowohl sonderpädagogische Unterstützung in der Form von integrativer Förderung als auch DaZ-Stunden nach den kantonalen Richtlinien zur Verfügung. Die Verantwortung übernahmen die jeweiligen Fachkräfte. Es gab keine Vorgaben für die Form oder Intensität des co-teachings, und Fachkräfte hatten auch die Möglichkeit, direkt mit den Kindern zu arbeiten.

5.2 Stichprobe

Die Stichprobe umfasste bei der ersten Erhebung am Ende des zweiten Schuljahres (t1) 569 Schüler*innen aus 36 Klassen in 20 Schulen (Knaben: $n=284$), bei der zweiten Erhebung am Ende des dritten Schuljahres (t2) 557 Schüler*innen (Knaben: $n=275$) (Tab. 1). Die Längsschnittstichprobe umfasste 505 Schüler*innen, die an beiden Zeitpunkten an den Befragungen teilgenommen haben. Die *Experimentalgruppe* bestand aus allen zweiten Klassen ($n=18$) aus 10 Schulen, die sich freiwillig am Schulversuch beteiligt haben. Für die *Kontrollgruppe* wurde zu jeder Klasse aus der Experimentalgruppe eine Klasse aus derselben Schulgemeinde, aber einer anderen Schule gesucht, die nicht am Schulversuch beteiligt war ($n=18$, gleicher Sozialindex, gleicher Mischindex bezüglich Erstsprache, vergleichbare Größe der Schule, vergleichbare Klassenzusammensetzung hinsichtlich des altersdurchmischten Lernens). Die Klassen der Kontrollgruppe nahmen ebenfalls freiwillig an der Untersuchung teil. Diese Lehrkräfte erhielten als Entschädigung CHF 200 für die Klassenkasse.

Vier Schulen der Experimentalgruppe waren zum Zeitpunkt der ersten Datenerhebung seit drei Jahren, drei Schulen seit zwei Jahren und drei Schulen seit einem Jahr am Versuch beteiligt. In 12 der 18 Klassen der Experimentalgruppe haben die Schüler*innen ihre gesamte bisherige Schullaufbahn mit intensivem co-teaching erlebt. Da die erste Erhebung am Ende des zweiten Schuljahres stattfand, verfügten

Tab. 1 Überblick über die Stichprobe

	t1 (Ende Klasse 2)		t2 (Ende Klasse 3)	
	Experimentalgruppe	Kontrollgruppe	Experimentalgruppe	Kontrollgruppe
	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)
Total pro Gruppe	293 (51,5)	276 (48,5)	295 (53,0)	262 (47,0)
Knaben	156 (53,2)	128 (46,4)	156 (52,9)	119 (45,4)
Anzahl SuS mit DaZ	78 (26,6)	78 (28,3)	66 (22,7)	70 (26,8)
Anzahl SuS mit IF	36 (12,3)	61 (22,1)	46 (15,9)	59 (22,6)
Total pro Messzeitpunkt		569		557
Total Fälle	621 (Experimentalgruppe: <i>n</i> = 331 (53,3 %); Kontrollgruppe: <i>n</i> = 290 (46,7 %))			
Stichprobe Längsschnitt	505 (Experimentalgruppe: <i>n</i> = 257 (50,9 %); Kontrollgruppe: <i>n</i> = 248 (49,1 %))			

DaZ Deutsch als Zweitsprache (erfasst über Angaben der Lehrkraft), *IF* Integrative Förderung (erfasst über Angaben der Lehrkraft), *t1/t2* Messzeitpunkte

auch die 6 Klassen, die im ersten Schuljahr nicht am Schulversuch teilnahmen, über umfangreiche Erfahrung mit dem intensiven co-teaching der Lehrkräfte.

Von allen Eltern wurde das Einverständnis für die Teilnahme an der Studie mit zwei Erhebungszeitpunkten (Ende Klasse 2 und Ende Klasse 3) eingeholt. Die Beteiligungsquote, das heißt der Anteil Schüler*innen, bei denen eine solche Einwilligung vorlag, betrug bei t1 83,5 % und bei t2 84,3 % (Experimentalgruppe: 84,1 %/86,4 %; Kontrollgruppe: 82,8 %/82,1 %). Dieser klassenspezifische Anteil der Beteiligung wurde zu beiden Erhebungszeitpunkten als Kontrollvariable in den Analysen berücksichtigt. Der Rücklauf betrug bei t1 98,2 % und bei t2 81,9 % (Experimentalgruppe: 98,4 %/97,5 %; Kontrollgruppe: 81,5 %/80,0 %). Die leicht geringere, allerdings immer noch sehr hohe Rücklaufquote in der Kontrollgruppe, kann dadurch erklärt werden, dass die Teilnahme für die Eltern der Kinder der Kontrollgruppe im Gegensatz zur Experimentalgruppe als weniger verbindlich wahrgenommen wurde.

In der Längsschnittstichprobe war der Anteil der Schüler*innen, die gemäß Angaben der Lehrkräfte Förderbedarf aufwiesen (integrative Förderung IF, DaZ-Unterricht), in der Kontrollgruppe zu beiden Erhebungszeitpunkten leicht höher als in der Experimentalgruppe. Nur beim ersten Erhebungszeitpunkt war der Unterschied des Anteils Schüler*innen mit IF signifikant ($t = 2,54$, $df = 515$, $p = 0,011$). In den durchgeführten Mehrebenenanalysen (siehe Abschn. 5.4) wurde dies kontrolliert. Bedeutsame Gruppenunterschiede bezüglich der Schüler*innen mit DaZ-Förderung zeigten sich zu keinem Zeitpunkt. Die Kinder waren zum ersten Zeitpunkt im Durchschnitt 103,3 Monate alt (8 Jahre, 7 Monate), ohne Unterschiede zwischen den beiden Untersuchungsgruppen.

5.3 Datenerhebung und Messinstrumente

Die Schüler*innen wurden am Ende der 2. Klasse und am Ende der 3. Klasse zu ihrer Wahrnehmung der Klassenführung und der Lernunterstützung im Unterricht befragt. Die Erhebung erfolgte in Gruppen von 4–5 Kindern. Die Items konnten je-

weils auf einer vierstufigen Smiley-Skala (4 = ja, 3 = eher ja, 2 = eher nein, 1 = nein) beurteilt werden. Alle Skalen weisen für beide Erhebungszeitpunkte durchgehend zufriedenstellende Reliabilitätswerte auf. Eingesetzt wurden eine Skala zur Erfassung der wahrgenommenen Klassenführung sowie drei Skalen zur Erfassung der wahrgenommenen Lernunterstützung (Lernbeziehung, Lernbegleitung und Lernklima).

Klassenführung: Fünf Items (z. B. „Hören alle Kinder in deiner Klasse auf Frau/Herr XY?“ „Kannst du in deiner Klasse in Ruhe arbeiten?“) in Anlehnung an Fauth et al. (2014); Cronbachs Alpha $\alpha = 0,76$ für t1, $\alpha = 0,77$ für t2.

Lernbeziehung Lehrkraft – Schüler*innen: Sechs Items (z. B. „Gehst du gerne zu Frau/Herr XY in die Schule?“ „Kannst du Frau/Herr XY sagen, wenn es dir nicht so gut geht?“) in Anlehnung an Kemna (2012); Cronbachs Alpha $\alpha = 0,86$ für t1, $\alpha = 0,88$ für t2.

Lernbegleitung: Sechs Items (z. B. „Erklärt Frau/Herr XY etwas immer wieder, bis du es verstanden hast?“ „Merkt Herr/Frau XY, wenn etwas zu schwierig ist für dich?“) in Anlehnung an Bos et al. (2005); Cronbachs Alpha $\alpha = 0,78$ für t1, $\alpha = 0,76$ für t2.

Lernklima: Sechs Items (z. B. „Ist Frau/Herr XY auch dann nett zu dir, wenn du einen Fehler gemacht hast?“ „Ermutigt dich Frau/Herr XY, wenn du eine Aufgabe schwierig findest?“) in Anlehnung an Fauth et al. (2014); Cronbachs Alpha $\alpha = 0,77$ für t1, $\alpha = 0,80$ für t2.

Aus Zeitgründen war es nicht möglich, die Kinder zu beiden Lehrkräften zu befragen. Wenn eine Lehrkraft hauptverantwortlich war für die Klasse, wurden die Kinder zu dieser Person befragt. Wenn beide Lehrkräfte zu gleichen Anteilen unterrichteten, wurden die Kinder nach dem Zufallsprinzip zu einer Lehrkraft (die gleiche Person bei t1 und t2) befragt. Dieses Prinzip wurde auch in der Kontrollgruppe angewendet, wenn die Funktion der Klassenlehrkraft von zwei Personen im Job-Sharing wahrgenommen wurde.

In einigen Klassen fand am Ende des zweiten Schuljahres ein Wechsel der Klassenlehrkräfte statt. Da angenommen wird, dass ein vollständiger Wechsel der Klassenlehrkräfte von t1 zu t2 die Entwicklung der Lernbeziehungen und der Lernumgebung bedeutsam beeinflusst, wurden Klassen ausgeschlossen, bei denen es zu einem vollständigen Wechsel der Klassenlehrkräfte gekommen war (vier Klassen der Experimentalgruppe, eine Klasse der Kontrollgruppe). Damit reduzierte sich die Stichprobe um 8,9 % auf $N = 566$ Schüler*innen in 31 Klassen (Experimentalgruppe: $n = 285$, 50,4 %, 14 Klassen; Kontrollgruppe: $n = 281$, 49,6 %, 17 Klassen). Der Wechsel einer einzelnen Klassenlehrkraft im Team wurde in den Analysen statistisch kontrolliert. Der familiäre Bildungshintergrund der Kinder wurde mit der Bücheraufgabe erhoben (Paulus 2009). Hierzu gaben die Schüler*innen anhand von Bildern mit Bücherregalen an, wie viele Bücher es bei ihnen zu Hause ungefähr gibt. Unterschieden wurde: 1 = 0 bis 10 Bücher ... 6 = mehr als 500 Bücher. In Anlehnung an die international vergleichenden Studien wurde die Variable recodiert (Stubbe

et al. 2016). Da nur wenige Schüler*innen angaben, maximal 100 Bücher zu haben, wurde der Split bei 250 Büchern anstatt wie sonst üblich bei 100 Büchern vorgenommen (0 = bis und mit 250 Bücher, 1 = mehr als 250 Bücher). Die Schüler*innen der Experimentalgruppe wiesen einen leicht höheren familiären Bildungshintergrund auf als jene der Kontrollgruppe. Der Unterschied ist signifikant, so dass es wichtig ist, diesen bei der Beantwortung der Fragestellungen zu kontrollieren ($U = 37080,50$, $p = 0,038$).

Wie bisherige Studien zeigen ist es wichtig, die Implementationsqualität von Interventionen zu überprüfen, um die Effekte der Veränderungen angemessen interpretieren zu können und die Qualität der Intervention zu steigern (Sanetti und Fallon 2011; Sanetti et al. 2014). Die Implementation wurde sowohl von der Schulleitung als auch der Behörde kontinuierlich überprüft, beispielsweise fanden regelmäßig Unterrichtsbesuche statt. Ausgehend von diesen Informationen kann davon ausgegangen werden, dass die Implementation in der Experimentalgruppe gemäß den Intentionen realisiert wurde, wobei sich das co-teaching im Verlauf des Schulversuchs aufgrund des regelmäßigen Monitoringprozesses nach Angaben der Projektleitung sogar verbessert hat.

5.4 Analysen

Um die Fragestellung beantworten zu können, wurde zum ersten und zweiten Erhebungszeitpunkt untersucht, wie die Schüler*innen in inklusiven Klassen die Lernunterstützung (Lernbeziehung, Lernbegleitung und Lernklima) und die Klassenführung wahrgenommen haben und wie sich diese Wahrnehmung über ein Schuljahr verändert hat.

Realisiert wurden zwei Analyseschritte. In einem ersten Schritt wurden deskriptive Analysen zu den zwei Erhebungszeitpunkten für die gesamte Stichprobe wie auch für die Experimental- und Kontrollgruppe getrennt durchgeführt.

Um die Effekte der Intervention in Bezug auf die Veränderung der wahrgenommenen Klassenführung und Lernunterstützung zu überprüfen, wurden in einem zweiten Schritt multivariate hierarchische Zwei-Ebenen-Regressionsanalysen mit randomisierten Intercepts mit dem Statistikpaket HLM 7.0 (Raudenbush et al. 2016) realisiert. Diese Analysen ermöglichen es, nicht nur allfällige Drittvariablen wie die individuellen Voraussetzungen der Schüler*innen (z. B. familiärer Bildungshintergrund) und die Klassenkomposition (z. B. Anzahl Schüler*innen mit Förderbedarf) zu kontrollieren, sondern auch die geschachtelte Struktur der Daten (Schüler*innen in Klassen) angemessen zu berücksichtigen.

Als abhängige Variable diente die Wahrnehmung der Lernumgebung durch die Schüler*innen zum zweiten Erhebungszeitpunkt. In einem ersten Schritt wurden die Null-Modelle (intercept-only-model) berechnet, um die Intraklassenkorrelation (*ICC*) zu bestimmen. Diese informiert über den Anteil der Gesamtvarianz einer Variable, die auf Unterschiede zwischen Schulklassen zurückzuführen ist (Lüdtke et al. 2006). Die Resultate verweisen in allen Dimensionen auf deutliche Unterschiede zwischen den Klassen ($ICC > 13\%$), was die Notwendigkeit von Mehrebenenanalysen bestätigt. In einem zweiten Schritt wurden unabhängige Variablen sowohl auf Individual- wie auch auf Klassenebene eingeführt. Es wurden sowohl die fixen wie

auch die zufälligen Effekte berechnet. Die Variablen auf Individualebene wurden um dem Gesamtmittelwert (grand mean) zentriert. Die bestehenden fehlenden Werte (maximal 10 % pro Variable) wurden nach eingehender Prüfung modellbasiert in den Analysen mittels des Parameter-Schätzers Restricted Maximum Likelihood (RLM) berücksichtigt (Graham 2009; Lüdtke et al. 2007). Folgende unabhängige Variablen wurden eingeführt:

- Individuum (Ebene 1): wahrgenommene Lernumgebung zu t1, Geschlecht (0=Mädchen, 1=Knaben), familiärer Bildungshintergrund (Anzahl Bücher: 0=0 bis 250 Bücher, 1= mehr als 250 Bücher), DaZ (Deutsch als Zweitsprache: 0=nein, 1=ja), Integrierte Förderung IF (ohne Begabtenförderung; 0=nein, 1=ja)
- Klasse (Ebene 2): Untersuchungsgruppe (0=Kontrollgruppe, 1=Experimentalgruppe), Anteil Schüler*innen mit DaZ, Anteil Schüler*innen mit IF, Anteil Lernende pro Klasse mit Einverständniserklärung, Teil-Lehrkraftwechsel (0=kein Wechsel, 1=ein Wechsel)

Ein signifikanter Haupteffekt der Untersuchungsgruppe bedeutet, dass sich die beiden Gruppen bedeutsam *in Bezug auf die Entwicklung des erfassten Aspektes der Lernumgebung* unterscheiden. Die Stärke des Effektes wurde anhand von standardisierten Regressionskoeffizienten (Hox 2002) sowie nach ihrer Transformation anhand von Korrelationskoeffizienten (Lenhard und Lenhard 2016) berechnet. Für die Interpretation der Effekte wurden unterschiedliche Quellen herbeigezogen. In Anlehnung an Cohen (1988) können Effekte von 0,10 als klein, von 0,30 als mittel und von 0,50 als groß bezeichnet werden. Allerdings hängt die Interpretation, ob ein Effekt stark oder schwach ist, auch vom Kontext der Untersuchung und vom Forschungsgebiet ab (Field 2013). Darüber hinaus ist zu berücksichtigen, welche Effekte im Schulkontext erwartet werden können. In Anlehnung an die Resultate internationaler Metaanalysen sind dies Effekte bis ca. 0,20 (vgl. Scheerens 2017). Daher werden Effekte von 0,20 und höher inhaltlich als relevant betrachtet. Die Ergebnisse der Mehrebenenanalysen sind zusammenfassend in Tab. 6 präsentiert.

6 Ergebnisse

6.1 Wahrgenommene Lernbeziehung zwischen Lehrkräften und Schülerinnen und Schülern

6.1.1 Deskriptive Analysen

Am Ende der zweiten Klasse (t1) nahm ein großer Teil der Kinder eine positive Lernbeziehung wahr. 93,1 % der Kinder beschrieben eine eher positive oder positive Lernbeziehung, 6,9 % der Kinder schätzten diese eher niedrig ein ($M=3,51$, $SD=0,60$, $n=569$). Zwischen den Klassen gab es große Unterschiede, wobei die Unterschiede von t1 (ICC: 13,1 %) zu t2 (ICC: 20,6 %) zugenommen haben. Zum zweiten Befragungszeitpunkt zeigte sich in der Gesamtstichprobe eine leicht niedrigere Einschätzung der Qualität der Lernbeziehung als bei t1 ($M=3,41$, $SD=0,64$,

Tab. 2 Wahrgenommene Lernbeziehung zwischen Lehrkräften und Schüler*innen – Vergleich der Messzeitpunkte und Untersuchungsgruppen

	t1 (2016)			t2 (2017)			Längsschnitt t1 bis t2 ($n=505$) Cohens d
	n	M	SD	n	M	SD	
Alle	569	3,51	0,60	557	3,41	0,64	-0,20
Exp	293	3,40	0,64	295	3,26	0,71	-0,26
Ko	276	3,63	0,54	262	3,58	0,51	-0,12
Cohens d Unterschied Exp–Ko		-0,38			-0,52		

Antwortformat: Smiley-Skala: 4 = positiv, 3 = eher positiv, 2 = eher negativ, 1 = negativ

ICC Intraklassenkorrelation t1: 13,1 %, ICC Intraklassenkorrelation t2: 20,6 %

Exp Experimentalgruppe, Ko Kontrollgruppe

$n=557$). 92,1 % der Kinder bewerteten die Lernbeziehung eher positiv oder positiv, während 7,9 % der Kinder die Qualität der Lernbeziehung eher gering einschätzten. Die Schüler*innen der Experimentalgruppe beurteilten die Beziehung zu beiden Zeitpunkten weniger positiv als die Schüler*innen der Kontrollgruppe, wobei die Differenz über ein Schuljahr hinweg größer wurde ($d=-0,38$ bzw. $d=-0,52$; Tab. 2).

6.1.2 Längsschnittliche Analysen

Die Einschätzung der Lernbeziehung verschlechterte sich vom ersten zum zweiten Erhebungszeitpunkt ($d=-0,20$; Tab. 2). Allerdings kam dieses Ergebnis hauptsächlich durch die schlechtere Beurteilung der Qualität der Lernbeziehungen in der Experimentalgruppe zustande, wie die Mehrebenenanalysen zeigen ($B=-0,20$, $p<0,05$; Tab. 6). Mit einem standardisierten Regressionskoeffizienten von $\beta=-0,16$ bzw. $r=-0,21$ ist der Unterschied auch inhaltlich bedeutsam (ohne Tabelle). Das Modell kann insgesamt 23,2 % an Varianz zwischen den Schüler*innen erklären.

6.2 Wahrgenommene Lernbegleitung

6.2.1 Deskriptive Analysen

Die erlebte Unterstützung wurde von den Kindern bei t1 im Durchschnitt eher positiv eingeschätzt ($M=3,15$, $SD=0,69$, $n=569$). 82,6 % der Kinder bewerteten die Lernbegleitung als eher positiv oder positiv. 17,4 % der Kinder beurteilten die Förderung eher negativ. Zwischen den Klassen gab es große Unterschiede, wobei die Unterschiede von t1 (ICC: 17,3 %) zu t2 (ICC: 22,0 %) zunahmen (Tab. 3).

Zwischen den beiden Untersuchungsgruppen zeigten sich Unterschiede. Die Kinder der Experimentalgruppe schätzten die Lernbegleitung schon zum ersten Erhebungszeitpunkt negativer ein als die Kinder aus der Kontrollgruppe ($d=-0,32$). Dieser Unterschied war beim zweiten Erhebungszeitpunkt noch größer ($d=-0,43$; Tab. 3). Am Ende der dritten Klasse bewerteten 79,7 % der Schüler*innen der Experi-

Tab. 3 Wahrgenommene Lernbegleitung – Vergleich der Messzeitpunkte und Untersuchungsgruppen

	t1 (2016)			t2 (2017)			Längsschnitt t1 bis t2 (n= 505)
	n	M	SD	n	M	SD	Cohens d
Alle	569	3,15	0,69	557	3,11	0,65	-0,08
Exp	293	3,05	0,72	295	2,98	0,70	-0,12
Ko	276	3,26	0,63	262	3,26	0,56	-0,04
Cohens d Unterschied Exp–Ko		-0,32			-0,43		

Antwortformat: Smiley-Skala: 4= positiv, 3= eher positiv, 2= eher negativ, 1= negativ

ICC Intraklassenkorrelation t1: 17,3%, ICC Intraklassenkorrelation t2: 22,0%

Exp Experimentalgruppe, Ko Kontrollgruppe

mentalklassen im Vergleich zu 90,1 % der Kinder der Kontrollklassen die Förderung und Unterstützung eher positiv oder positiv.

6.2.2 Längsschnittliche Analysen

Die Mehrebenenanalysen zeigten, dass sich die Experimentalgruppe bezüglich der Veränderung der Wahrnehmung der Lernumgebung signifikant von der Kontrollgruppe unterscheidet (vgl. Tab. 6). Die von den Schüler*innen erlebte Lernbegleitung veränderte sich im Jahresvergleich in der Experimentalgruppe signifikant negativer als in den Klassen der Kontrollgruppe ($B = -0,24, p < 0,05$). Der standardisierte Effekt ist mit $\beta = -0,19$ bzw. $r = -0,24$ auch inhaltlich bedeutsam (ohne Tabelle). Darüber hinaus verlief die Entwicklung – in der gesamten Stichprobe – signifikant negativer, je höher der Anteil an Kindern mit DaZ-Förderung in der Klasse war ($B = -0,37, p < 0,05$). Das Modell kann insgesamt 26,0 % an Varianz zwischen den Schüler*innen erklären.

6.3 Wahrgenommenes Lernklima

6.3.1 Deskriptive Analysen

Zum ersten Befragungszeitpunkt schätzten die Kinder das Lernklima im Durchschnitt positiv ein ($M = 3,43, SD = 0,58, n = 569$). Auf der Skala von 1 bis 4 bewerten 92,8 % der Kinder das Lernklima als eher positiv oder positiv (Werte 3 und 4), 7,2 % der Kinder nahmen ein eher negatives Lernklima wahr. Zwischen den Klassen gab es große Unterschiede, wobei die Unterschiede von t1 (ICC: 16,6 %) zu t2 (ICC: 23,8 %) zunahmen (Tab. 4).

Die beiden Untersuchungsgruppen unterschieden sich zum ersten wie auch zum zweiten Befragungszeitpunkt in der durchschnittlichen Einschätzung des Lernklimas (Tab. 4). Die Lernenden der Kontrollklassen schätzten das Lernklima positiver ein als die Lernenden der Experimentalklassen ($d = -0,24$ bzw. $d = -0,47$).

Tab. 4 Wahrgenommenes Lernklima – Vergleich der Messzeitpunkte und Untersuchungsgruppen

	t1 (2016)			t2 (2017)			Längsschnitt t1 bis t2 ($n=505$)
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	Cohens <i>d</i>
Alle	569	3,43	0,58	557	3,35	0,63	-0,14
Exp	293	3,36	0,62	295	3,22	0,70	-0,23
Ko	276	3,50	0,53	262	3,51	0,50	-0,04
Cohens <i>d</i> Unterschied Exp–Ko		-0,24			-0,47		

Antwortformat: Smiley-Skala: 4 = positiv, 3 = eher positiv, 2 = eher negativ, 1 = negativ

ICC Intraklassenkorrelation t1: 16,6%, ICC Intraklassenkorrelation t2: 23,8%

Exp Experimentalgruppe, Ko Kontrollgruppe

6.3.2 Längsschnittliche Analysen

Die Einschätzung des Lernklimas veränderte sich negativ über den Untersuchungszeitraum (Cohens $d = -0,14$). Gemäß den Ergebnissen der Mehrebenenanalysen ist der Gruppeneffekt allerdings nur tendenziell signifikant ($B = -0,16$, $p = 0,067$; Tab. 6), mit einem standardisierten Effekt von $\beta = -0,13$ bzw. $r = -0,18$ (ohne Tabelle). Die Veränderungen über die Zeit verlaufen in den beiden untersuchten Gruppen somit relativ ähnlich. Das Modell kann insgesamt 21,8% an Varianz zwischen den Schüler*innen erklären.

6.4 Wahrgenommene Klassenführung

6.4.1 Deskriptive Analysen

Die Schüler*innen beurteilten die Klassenführung in ihren Klassen zu t1 im Durchschnitt auf einem eher niedrigen Niveau ($M = 2,26$, $SD = 0,63$, $n = 569$). Auf der Skala von 1 bis 4 bewerteten einzig 35,9% der Kinder die Klassenführung eher positiv oder positiv. 64,1% der Lernenden hingegen nahmen eine eher geringe Qualität in

Tab. 5 Wahrgenommene Klassenführung – Vergleich der Messzeitpunkte und Untersuchungsgruppen

	t1 (2016)			t2 (2017)			Längsschnitt t1 bis t2 ($n=505$)
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	Cohens <i>d</i>
Alle	569	2,26	0,63	557	2,32	0,60	0,05
Exp	293	2,15	0,57	295	2,21	0,56	0,04
Ko	276	2,37	0,67	262	2,44	0,61	0,07
Cohens <i>d</i> Unterschied Exp–Ko		-0,35			-0,39		

Antwortformat: Smiley-Skala: 4 = positiv, 3 = eher positiv, 2 = eher negativ, 1 = negativ

ICC Intraklassenkorrelation t1: 29,1%, ICC Intraklassenkorrelation t2: 19,6%

Exp Experimentalgruppe, Ko Kontrollgruppe

der Klassenführung wahr. Zwischen den Klassen gab es große Unterschiede. Die Unterschiede von t1 (ICC: 29,1 %) zu t2 (ICC: 19,6 %) nahmen ab (Tab. 5).

Die Kinder der beiden Untersuchungsgruppen unterschieden sich in ihrer Einschätzung der Klassenführung, wobei die Kontrollgruppe die Klassenführung

Tab. 6 Multiple hierarchische Regressionsanalysen zur Analyse des Effektes der Gruppenzugehörigkeit auf die Entwicklung der wahrgenommenen Lernunterstützung und die wahrgenommene Klassenführung

Indikator	Wahrgenommene Lernbeziehung t2	Wahrgenommene Lernbegleitung t2	Wahrgenommenes Lernklima t2	Wahrgenommene Klassenführung t2
Fixe Effekte				
Intercept	3,38*** (0,24)	3,53*** (0,27)	3,42*** (0,25)	2,48*** (0,20)
<i>Level 1 Schüler*innen</i>				
Geschlecht (0 = Mädchen; 1 = Knaben)	-0,06 (0,05)	-0,04 (0,06)	-0,11+ (0,06)	0,02 (0,05)
Sozio-kultureller Hintergrund (0 = bis 250 Bücher, 1 = mehr als 250 Bücher)	0,00 (0,05)	0,09 (0,06)	0,02 (0,05)	-0,02 (0,06)
DaZ (0 = kein DaZ; 1 = DaZ)	-0,06 (0,06)	0,09 (0,07)	0,09 (0,07)	-0,05 (0,05)
IF (0 = kein IF; 1 = IF)	-0,03 (0,07)	0,11 (0,09)	0,04 (0,08)	-0,04 (0,05)
Aspekt der Lernumgebung zum Zeitpunkt t1	0,43*** (0,06)	0,34*** (0,05)	0,30*** (0,06)	0,21 *** (0,05)
<i>Level 2 Klasse</i>				
Experimental- vs. Kontrollklasse (0 = Ko, 1 = Ex)	-0,20* (0,07)	-0,24* (0,10)	-0,16+ (0,08)	-0,25** (0,07)
Anteil Schüler*innen Teilnahme an Erhebung 1 (Werte 0 bis 1)	0,18 (0,24)	-0,24 (0,30)	0,07 (0,28)	0,17 (0,23)
Teamwechsel Lehrkräfte (0 = kein Wechsel, 1 = ein Wechsel)	-0,06 (0,09)	-0,17 (0,12)	-0,08 (0,10)	-0,02 (0,08)
Anteil DaZ-Schüler*innen (Werte 0 bis 1)	0,16 (0,18)	-0,37* (0,14)	-0,17 (0,16)	-0,41* (0,15)
Anteil IF-Schüler*innen (Werte 0 bis 1)	-0,28 (0,38)	0,12 (0,41)	0,40 (0,41)	-0,66 (0,39)
Zufällige Effekte (Varianzkomponente)				
Intercept	0,04***	0,06***	0,06***	0,03***
Geschlecht	0,02	0,03	0,04	0,02
Sozio-kultureller Hintergrund (SES)	0,02	0,02	0,02	0,03
DaZ	0,00	0,04	0,04	0,00
IF	0,02	0,11	0,10	0,01
Aspekt der Lernumgebung zum Zeitpunkt t1	0,06*	0,02	0,05	0,01
R²	0,232	0,260	0,220	0,268

Unstandardisierte Koeffizienten (Standardfehler)

Irrtumswahrscheinlichkeit: * $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$, + $p < 0,10$

IF Integrierte Förderung, DaZ Deutsch als Zweitsprache

zum ersten Erhebungszeitpunkt positiver einschätzten als die Experimentalgruppe ($d = -0,35$; Tab. 5). Dies ist auch zu t2 der Fall ($d = -0,39$).

6.4.2 Längsschnittliche Analysen

Die multivariaten Mehrebenenanalysen zeigen, dass sich die Klassenführung in der Kontrollgruppe nach Kontrolle der individuellen Voraussetzungen der Schüler*innen und der Klassenzusammensetzung signifikant positiver verändert hat als in der Experimentalgruppe ($B = -0,25$, $p < 0,01$; Tab. 6). Der standardisierte Effekt ist mit $\beta = 0,21$ bzw. $r = -0,27$ inhaltlich bedeutsam. Zusätzlich verweisen die Ergebnisse auf einen Haupteffekt der Variable „DaZ“: Je geringer der Anteil an Schüler*innen mit DaZ-Förderung in der Klasse war, desto positiver verlief insgesamt die Entwicklung der wahrgenommenen Klassenführung ($B = -0,41$, $p < 0,05$). Der Anteil an erklärter Varianz ist 26,8 %.

7 Diskussion

Die Zusammenarbeit der Lehrkräfte in inklusiven Settings wird oft als zentrale Rahmenbedingung für einen gelingenden inklusiven Unterricht betrachtet (z.B. Hellmich et al. 2017; Lütje-Klose und Urban 2014; Shaffer und Thomas-Brown 2015). Aus inklusionspädagogischer Sicht wird angestrebt, dass Teams von Lehrkräften die Schüler*innen gemeinsam und unterrichtsintegriert unterstützen. Wie die Lernenden ein solches Lernsetting wahrnehmen, wurde bis anhin erst mit kleinen Stichproben (Keeley 2015) und querschnittlich untersucht (Keeley et al. 2017; Schwab 2017). In der hier vorliegenden Studie wurde mittels einer quasi-experimentellen Studie analysiert, inwiefern sich die Wahrnehmung von Schüler*innen zu zentralen Aspekten der Unterrichtsqualität (Klassenführung und Lernunterstützung) im Jahresverlauf veränderte und inwiefern diese Veränderungen in einem systematischen Zusammenhang mit der Zusammenarbeitsform der Lehrkräfte stehen. Verglichen wurden eine Experimentalgruppe, in der ein hoher bzw. intensiver Anteil von co-teaching stattfand und Unterstützungsmaßnahmen unterrichtsintegriert von den Regelklassenlehrkräften angeboten wurden, mit einer Kontrollgruppe, in der in der Regel eine Klassen- und eine zweite (sonderpädagogische ausgebildete) Lehrkraft kooperierten, wobei die sonderpädagogische Lehrkraft im Gegensatz zur Experimentalgruppe auch direkt mit den Schüler*innen arbeitete. Damit werden in dieser Studie zwei unterschiedliche Formen des co-teachings untersucht.

Die mehrebenenanalytischen Auswertungen zeigen durchgehend, dass sich aus Sicht der Schüler*innen der Experimentalgruppe, in der eine intensivere Form des co-teaching umgesetzt wurde als in der Kontrollgruppe, das Ausmaß an Lernunterstützung und die Klassenführung und somit Aspekte der Unterrichtsqualität negativer entwickelt haben als in der Kontrollgruppe. Der Interaktionseffekt zeigt sich am stärksten bezüglich der wahrgenommenen Klassenführung. Die Effekte können zwischen klein und mittel eingestuft werden (Cohen 1988), erreichen aber in Anlehnung an die Schuleffektivitätsbefunde inhaltliche Relevanz (Scheerens 2017).

Die Ergebnisse können in Übereinstimmung mit bisherigen Befunden (z.B. Schwab 2017; Baeten und Simons 2014; Vangrieken et al. 2015; Cook et al. 2017) als Hinweis interpretiert werden, dass co-teaching Modelle ambivalente oder negative Effekte haben können. Die hier vorliegende Studie, die zwei verschiedene Formen des co-teaching miteinander verglichen hat, gibt Anlass zur Annahme, dass sich diese Effekte besonders bei intensiveren Formen des co-teaching zeigen. Die Gründe für diese negativeren Entwicklungen der intensiveren Form des co-teachings können anhand der vorliegenden Daten empirisch nicht bestimmt werden, da es leider nicht möglich war, die konkrete Gestaltung des co-teaching im Unterricht der Experimentalgruppe und die konkrete Zusammenarbeit der Regelklassen- und der zweiten (sonderpädagogisch ausgebildeten) Lehrkraft in der Kontrollgruppe vergleichend zu untersuchen.

Interpretationen für diese Ergebnisse ergeben sich aus bisher vorliegenden Erkenntnissen. Wie Cook et al. (2017) zusammenfassend festhalten, weist der im co-teaching geführte Unterricht nicht an sich eine höhere Qualität auf. Eine anspruchsvollere und intensivere individuelle Betreuung der Schüler*innen scheint trotz zweier Lehrkräfte im Klassenzimmer mehrheitlich nicht realisiert zu werden. Vielmehr dominiert das Modell *one teach/one assist* (Scruggs et al. 2007; Pancsofar und Petroff 2016; Spörer et al. 2021). Ein weiterer Grund für die negativere Entwicklung in der Experimentalgruppe könnte darin liegen, dass gemäß den Forschungsergebnissen von Keeley (2015) und Keeley et al. (2017) für die Schüler*innen im Unterricht mit co-teaching die Verantwortlichkeiten nicht klar sind oder die beiden Lehrkräfte die Regeln nicht konsistent anwenden, was zu Unsicherheiten bezüglich der Regeln und Strukturen des Unterrichts und damit zu einer negativeren Bewertung der Klassenführung durch die Schüler*innen führen könnte.

Darüber hinaus ist zu vermuten, dass mehr Unruhe herrscht, wenn zwei Lehrkräfte im Klassenzimmer sind und die Konzentration der Schüler*innen somit in höherem Masse gefordert ist. Dies ist im traditionellen Unterrichtsmodell weniger ein Problem, insbesondere dann, wenn die zweite (sonderpädagogisch ausgebildete) Lehrkraft mit Kleingruppen außerhalb des Klassenzimmers arbeitet. Allerdings ist die Gestaltung von erfolgreichem inklusivem Unterricht eng verknüpft mit dem Anspruch, klassenexterne Förderung und damit verbunden die Etikettierung von Schüler*innen zu vermeiden (z.B. Hinz und Köpfer 2016; Köpfer 2014) und es hat sich auch gezeigt, dass sich eine häufige klassenexterne Förderung negativ auf die soziale Partizipation auswirken kann (Wiener und Tardif 2004). Die Ergebnisse der hier vorliegenden Studie weisen jedoch darauf hin, dass eine klassenintegrierte Förderung von den Lernenden auch negativ wahrgenommen werden kann. In diesem Zusammenhang sind die Ergebnisse der Studie von Spörer et al. (2021) interessant. Diese ergab, dass Schüler*innen mit besonderem Förderbedarf in Schulstunden mit co-teaching häufig mit der sie begleitenden Lehrkraft interagierten, was auch zu einer Etikettierung führen kann. Eine weitere Erklärung für die identifizierten Befunde könnte in der Qualität, inhaltlichen Fokussierung und Intensität der Absprache zwischen den beiden Lehrkräften liegen. Für die Vorbereitung des Unterrichts in einer intensiven Form des co-teaching sind umfangreichere Absprachen notwendig als für Lektionen, die hauptsächlich von einer Lehrkraft geführt werden bzw. für Situationen, in denen die Verantwortlichkeiten – z.B. für eine bestimmte Gruppe von

Schüler*innen – stärker getrennt sind. Anzustreben wären ein co-planning (Johnson 2015) und eine kokonstruktive Aufgabenbewältigungen (Gräsel et al. 2006; Grosche et al. 2020). Es könnte sein, dass die Absprachen zum co-teaching – wie in der Review von Vangrieken et al. (2015) oder von Gräsel et al. (2006) berichtet – auch in den untersuchten Klassen eher oberflächlich in Form von Austausch erfolgten und sich dies in der Experimentalgruppe, wo es eine Verpflichtung zum co-teaching gab, stärker ausgewirkt hat als in der Kontrollgruppe. Dort kann davon ausgegangen werden, dass die Rollen und die Verantwortlichkeiten zwischen Regelklassenlehrkraft und zweiter (sonderpädagogisch ausgebildeter) Lehrkraft eindeutiger geklärt und getrennt waren und eventuell auch häufiger eine klar arbeitsteilige Kooperation (Grosche et al. 2020) erfolgte. Zudem zeigen verschiedene Studien, dass co-teaching anspruchsvoll ist und von den Lehrkräften hohe Kompetenzen erfordert, die nicht unbedingt vorhanden sind (Chitiyo 2017). Zwar wurden im Schulversuch verschiedene Fortbildungen zum Thema Zusammenarbeit angeboten, diese wurden jedoch von den am Schulversuch teilnehmenden Lehrkräften als weniger notwendig erachtet als Angebote zu den Themen Förderplanung und unterrichtsintegrierte Sprachförderung und demzufolge auch weniger in Anspruch genommen. Zudem weisen die Daten von Gräsel et al. (2006) darauf hin, dass es grundsätzlich schwierig zu sein scheint, die Kooperation von Lehrkräften durch Fortbildungen zu verbessern. Die negativere Entwicklung in der Experimentalgruppe könnte auch dadurch bestimmt sein, dass die Fachkräfte nicht mehr direkt mit den Kindern gearbeitet haben, dass spezifische Fördermaßnahmen, wie von Jordan und McGhie-Richmond (2014) vorgeschlagen, nicht oder wenig eingesetzt wurden und die Schüler*innen sich deshalb nur ungenügend unterstützt fühlten.

Abschließend ist auf einen weiteren interessanten Befund hinzuweisen. Der Anteil an DaZ-Schüler*innen in der Klasse ist unabhängig von der Zugehörigkeit zur Experimental- oder Kontrollgruppe ein Prädiktor für die Wahrnehmungsveränderungen der Schüler*innen während des Schuljahres. Je mehr Schüler*innen mit DaZ in einer Klasse waren, desto negativer war die Entwicklung der Wahrnehmung der Lernbegleitung und der Klassenführung. Dieses Ergebnis könnte dadurch bedingt sein, dass aus der Wahrnehmung der Schüler*innen zu wenig individuelle Förderung angeboten wurde. Um diese Vermutungen zu überprüfen, müssten konkrete Informationen zum Umfang und zur Ausgestaltung der Fördermaßnahmen (unterrichtsintegriert, klassenexterne Förderung usw.) erhoben werden. Differenzierte Erkenntnisse zu dieser Thematik wären sehr bedeutsam. Bezüglich DaZ-Sprachförderung wird nämlich davon ausgegangen, dass diese einerseits in der Form einer durchgängigen Sprachförderung und eines sprachsensiblen Fachunterrichts unterrichtsintegriert geschehen sollte, dass aber – wenn erforderlich – auch eine additive Sprachförderung in Form von zusätzlichen Förderstunden notwendig sein kann (z. B. Riemer 2017).

Die hier präsentierte Untersuchung trägt zur Erweiterung des Forschungsstands bei, da es bis anhin kaum Studien gibt, die co-teaching aus der Wahrnehmung der Lernenden längsschnittlich untersucht haben. Dennoch weist sie verschiedene Limitationen auf: Zwischen der Kontroll- und der Experimentalgruppe gab es erstens bereits zum Zeitpunkt t1 Unterschiede zugunsten der Schüler*innen der Kontrollgruppe. Dies ist in quasi-experimentellen Studien im schulischen Kontext ein häufig beschriebenes Problem (Borman 2009; Slavin 2010). Über mögliche Gründe

können nur Vermutungen angestellt werden. Es könnte sein, dass sich zum Schulversuch besonders jene Schulen gemeldet haben, die in ihrer Unterrichtspraxis besondere Probleme wahrgenommen haben. Allerdings waren die Unterschiede eher gering. Zudem interessierte im Rahmen dieser Studie die längsschnittliche Veränderung, sodass durch die Mehrebenenanalysen das Ausgangsniveau kontrolliert werden konnte. Beachtet werden muss zweitens, dass die Beteiligungsquote der Experimentalgruppe höher war als in der Kontrollgruppe, wahrscheinlich bedingt durch die größere Verbindlichkeit im Schulversuch. Der Unterschied beträgt jedoch 1–2 Kinder pro Klasse, und die Variable wurde in den Analysen kontrolliert. Zudem ist auch mit einer Beteiligungsquote von 80 %, wie sie in der Kontrollgruppe realisiert werden konnte, eine sehr hohe Teilnahme erreicht worden. Somit kann davon ausgegangen werden, dass durch die leicht unterschiedliche Beteiligungsquote keine inhaltlichen Verzerrungen in der Beurteilung der Lernumgebung entstanden sind. Zu kritisieren ist drittens, dass das Leistungsniveau der Schüler*innen nicht kontrolliert werden konnte, weil es aufgrund der Vorgaben des Auftraggebers der Evaluation nicht möglich war, die Leistungsentwicklung der Schüler*innen zu erfassen. Auch wenn die Kontrollklassen aus derselben Gemeinde stammten wie die Experimentalklassen und bezüglich verschiedener Merkmale (Mischindex bezüglich Erstsprache, Größe der Schule, Klassenzusammensetzung hinsichtlich altersdurchmischten Lernens) vergleichbar waren, kann nicht ausgeschlossen werden, dass das Leistungsniveau der Schüler*innen der beiden Gruppen unterschiedlich war und die Ergebnisse beeinflusst hat. Aufgrund der substanziellen Unterschiede zwischen den Klassen wäre es viertens wünschbar gewesen, wenn eine größere Stichprobe hätte untersucht werden können. Dies war aufgrund der Vorgaben des Schulversuchs nicht möglich. Ebenso ist fünftens zu berücksichtigen, dass nicht alle Schulen zum gleichen Zeitpunkt in den Schulversuch eingestiegen sind, sodass in Betracht gezogen werden muss, dass die identifizierten negativen Effekte zumindest teilweise durch den unterschiedlichen Erfahrungshintergrund der Akteure in den drei Staffeln erklärt werden können. Sechstens ergeben sich Limitationen aufgrund der Erhebungen. Aus Zeitgründen konnten die Kinder jeweils den Fragebogen nur bezogen auf eine der beiden Lehrkräfte ausfüllen. Obwohl verschiedene Maßnahmen getroffen wurden, um dieser Einschränkung zu begegnen (zufällige Auswahl der Lehrkräfte, wenn es keine Hauptlehrkraft gab, Beurteilung derselben Lehrkraft im Längsschnitt, gleiches Vorgehen für Experimental- und Kontrollgruppe) wäre es wünschbar gewesen, wenn die Kinder den Unterricht beider Lehrkräfte hätten beurteilen können. Siebtens erlaubte es die Untersuchungsanlage nicht, die konkrete Umsetzung des co-teachings in den beiden Gruppen zu untersuchen. Nach Vangrieken et al. (2015) ist das eine Herausforderung, die sich in vielen Studien zum co-teaching zeigt und die dazu führt, dass es nur ansatzweise möglich ist, die Qualität der Implementation zu kontrollieren (Sanetti und Fallon 2011). Aufgrund der Informationen der Schulversuchsverantwortlichen und der Schulleitungen, von denen die Implementationsqualität regelmäßig und kontinuierlich überprüft und reflektiert wurde (vgl. Abschn. 5.3), kann davon ausgegangen werden, dass das co-teaching im Verlauf des Schulversuchs gemäß den Vorgaben umgesetzt worden ist und sich sogar verbessert hat. Achtens konnten die Ausbildung der Lehrkräfte oder die Zusammensetzung der Teams als unabhängige Variablen in den Analysen nicht berücksichtigt werden.

Für zukünftige Studien wäre es daher beispielsweise wichtig, die Kooperationsform (Austausch, Arbeitsteilung, Kokonstruktion) näher zu analysieren. Grosche et al. (2020) vermuten, dass die kokonstruktive Kooperation die beabsichtigten Wirkungen auf die Schüler*innen besser erreicht als die anderen Formen.

Trotz dieser Einschränkungen kann festgestellt werden, dass im deutschsprachigen Raum erstmals eine Längsschnittstudie zur Wahrnehmung von unterschiedlich intensiven Formen von co-teaching durch die Schülerinnen und Schüler vorliegt, in der verschiedene mögliche Drittvariablen kontrolliert werden konnten. Um die hier präsentierten eher negativen Auswirkungen von co-teaching auf die Schüler*innen weiter zu überprüfen, wäre es daher für zukünftige Studien zentral, co-teaching in den unterschiedlichen Ausprägungen unter Berücksichtigung der oben formulierten Limitationen differenziert zu erfassen und dessen Wirkung auf die Unterrichtsqualität zu untersuchen. Dies ist ein wichtiges Forschungsdesiderat mit Blick auf die Entwicklung des inklusiven Unterrichts.

Funding Open access funding provided by University of Zurich

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

Literatur

- Baeten, M., & Simons, M. (2014). Student teachers' team teaching: Modells, effets and conditions for implementation. *Teaching and Teacher Education*, *41*, 92–110.
- Borman, G. D. (2009). The use of randomized trials to inform education policy. In G. Sykes, B. Schneider & D. N. Plank (Hrsg.), *Handbook of policy research* (S. 129–138). New York, London: Routledge.
- Bos, W., Lankes, E.-M., Prenzel, M., Schwippert, K., Valtin, R., Voss, A., & Walther, G. (Hrsg.). (2005). *IGLU. Skalenhandbuch zur Dokumentation der Erhebungsinstrumente*. Münster: Waxmann.
- CAST (2011). *Universal design for learning guidelines version 2.0*. Wakefield: Center for Applied Special Technology.
- Chitiyo, J. (2017). Challenges to the use of co-teaching by teachers. *International Journal of whole Schooling*, *13*(3), 55–66.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, New York: Erlbaum.
- Cook, L., & Friend, M. (1995). Co-teaching: Guidelines for creative effective practices. *Focus on Exceptional Children*, *28*(3), 1–16.
- Cook, S. C., McDuffie-Landrum, K. A., Oshita, L., & Cook, B. G. (2017). Co-teaching for students with disabilities: A critical and updated analysis of the empirical literature. In J. M. Kauffman, D. P. Hallahan & P. C. P. C. Pullen (Hrsg.), *The handbook of special education* (2. Aufl. S. 233–248). New York: Routledge.

- 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.
- Feuser, G. (2018). *Wider die Integration der Inklusion in die Segregation. Zur Grundlegung einer Allgemeinen Pädagogik und entwicklungslogischen Didaktik*. Berlin: Peter Lang.
- Field, A. (2013). *Discovering statistics using SPSS IBM Statistics: (and sex and drugs and rock, n' roll)* (4. Aufl.). London: SAGE.
- Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60*, 549–576.
- Gräsel, C., Fußangel, K., & Pröbstel, C. (2006). Lehrkräfte zur Kooperation anregen – eine Aufgabe für Sisyphos? *Zeitschrift für Pädagogik, 52*(2), 205–219.
- Grosche, M., & Fleischhauer, E. (2017). Implikationen der Theorien der schulischen Inklusion für das Konzept der Förderung von Deutsch als Zweitsprache. In M. Becker-Mrotzek, P. Rosenberg, C. Schroeder & A. Witte (Hrsg.), *Deutsch als Zweitsprache in der Lehrerbildung* (S. 155–169). Münster: Waxmann.
- Grosche, M., Fussangel, K., & Gräsel, C. (2020). Kokonstruktive Kooperation zwischen Lehrkräften. Aktualisierung und Erweiterung der Kokonstruktionstheorie sowie deren Anwendung am Beispiel schulische Inklusion. *Zeitschrift für Pädagogik, 66*(4), 461–479.
- Hellmich, F., Hoya, F., Görel, G., & Schwab, S. (2017). Unter welchen Voraussetzungen kooperieren Grundschullehrkräfte im inklusiven Unterricht? – Eine Studie zu den Bedingungen der Kooperationsbereitschaft von Grundschullehrerinnen und -lehrern im inklusiven Unterricht. *Empirische Sonderpädagogik, 9*(1), 36–51.
- Hinz, A., & Köpfer, A. (2016). Unterstützung trotz Dekategorisierung? Beispiele für Unterstützung durch Dekategorisierung. *Vierteljahresschrift für Heilpädagogik und ihre Nachbargebiete, 85*(1), 36–47.
- Hox, J. (2002). *Multiple analysis. Techniques and applications*. Mahwah: Erlbaum.
- Johnson, M. (2015). Co-teaching: Voraussetzung und Garant für eine Schule für Alle – Erfahrungen aus den USA. *Zeitschrift für Inklusion, 1*(1). <https://www.inklusion-online.net/index.php/inklusion-online/article/view/262>. Zugegriffen: 15. Sept. 2020.
- Jordan, A., & McGhie-Richmond, D. (2014). Identifying effective teaching practices in inclusive classrooms. *International Perspectives on Inclusive Education, 3*, 133–162.
- Keeley, R. G. (2015). Measurement of student and teacher perceptions of co-teaching models. *The Journal of Special Education Apprenticeship, 4*(1), 1–15.
- Keeley, R. G., Brown, M. R., & Knapp, D. (2017). Evaluation of the student experience in the co-taught classroom. *International Journal of Special Education, 32*(3), 520–537.
- Kemna, P. W. (2012). Effektive Lehrer-Schüler-Beziehung. Empirische Analyse eines Konstrukts. In K.-O. L. Bauer & N. Logemann (Hrsg.), *Effektive Bildung. Zur Wirksamkeit und Effizienz pädagogischer Prozesse* (S. 77–99). Münster: Waxmann.
- Köpfer, A. (2014). Raum und Stigma – eine raumtheoretische Annäherung an die kritische Rolle von Integrationshelfer/innen in inklusiven Settings. *Vierteljahresschrift für Heilpädagogik und ihre Nachbargebiete, 83*(4), 295–304.
- Lenhard, W., & Lenhard, A. (2016). Computation of effect sizes. *Psychometrica*. https://www.researchgate.net/publication/336836189_Computation_of_Effect_Sizes. Zugegriffen: 15. Sept. 2020.
- Lüdtke, O., Robitzsch, A., Trautwein, U., & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung. Probleme und Lösungen. *Psychologische Rundschau, 58*(2), 103–117.
- Lüdtke, O., Trautwein, U., Kunter, M., & Baumert, J. (2006). Analyse von Lernumwelten. *Zeitschrift für Pädagogische Psychologie, 20*, 85–96.
- Lütje-Klose, B., & Urban, M. (2014). Professionelle Kooperation als wesentliche Bedingung inklusiver Schul- und Unterrichtsentwicklung. Teil 1: Grundlagen und Modell inklusiver Kooperation. *Vierteljahresschrift für Heilpädagogik und ihre Nachbargebiete, 83*(2), 112–123.
- Murawski, W. M., & Swanson, H. L. (2001). A meta-analysis of co-teaching research. Where are the data? *Remedial and Special Education, 22*(5), 258–267.
- Pancsofar, N., & Petroff, J. G. (2016). Teachers' experiences with co-teaching as a model for inclusive education. *International Journal of Inclusive Education, 20*(10), 1043–1053.
- Paulus, C. (2009). Die „Bücheraufgabe“ zur Bestimmung des kulturellen Kapitals bei Grundschulern. <http://psydok.psycharchives.de/jspui/handle/20.500.11780/3344>. Zugegriffen: 19. März 2019.
- Pool Maag, S., & Moser Opitz, E. (2014). Inklusiver Unterricht – grundsätzliche Fragen und Ergebnisse einer explorativen Studie. *Empirische Sonderpädagogik, 2*, 133–149.
- 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.

- Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R. T., & Du Toit, M. (2016). *HLM7 hierarchical linear and nonlinear modeling user manual*. Lincolnwood: Scientific Software International.
- Riemer, C. (2017). DaZ und Inklusion – Gemeinsamkeiten und Unterschiede. Ein fachpolitischer Positionierungsversuch aus der Perspektive des Fachs DaF/DaZ. In M. Becker-Mrotzek, P. Rosenberg, C. Schroeder & A. Witte (Hrsg.), *Deutsch als Zweitsprache in der Lehrerbildung* (S. 171–186). Münster: Waxmann.
- Sanetti, L. M. H., & Fallon, L. M. (2011). Treatment integrity assessment: How estimates of adherence, quality, and exposure influence interpretation of implementation. *Journal of Educational & Psychological Consultation*, 21(3), 209–232.
- Sanetti, L. M. H., Collier-Meek, M. A., Long, A. C. J., Kim, J., & Kratochwill, T. R. (2014). Using implementation planning to increase teachers' adherence and quality to behavior support plans. *Psychology in the Schools*, 51(8), 879–895.
- Scheerens, J. (2017). How results of school effectiveness research could be used for school improvement. In U. Steffens, K. Maag Merki & H. Fend (Hrsg.), *Schulgestaltung. Aktuelle Befunde und Perspektiven der Schulqualitäts- und Schulentwicklungsforschung*. Grundlagen der Qualität von Schule, (Bd. 2, S. 309–330). Münster: Waxmann.
- Schwab, S. (2017). Interprofessionelle Lehrkraftkooperation im inklusiven Unterricht aus der Perspektive der Schülerinnen und Schüler. *Unterrichtswissenschaft*, 45(4), 262–279.
- Scruggs, T. E., Mastropieri, M. A., & McDuffie, K. (2007). Co-teaching in inclusive classrooms: A meta-synthesis of qualitative research. *Exceptional Children*, 73(4), 392–416.
- Shaffer, L., & Thomas-Brown, K. (2015). Enhancing teacher competency through co-teaching and embedded professional development. *Journal of Education and Training Studies*, 3(3), 117–125.
- Slavin, R. E. (2010). Experimental studies in education. In B. P. M. Creemers, L. Kyriakides & P. Sammons (Hrsg.), *Methodological advances in educational effectiveness research* (S. 102–114). New York, London: Routledge.
- Spörer, N., Henke, T., & Bosse, S. (2021). Is there a dark side of co-teaching? A study on the social participation of primary school students and their interactions with teachers and classmates. *Learning and Instruction*. <https://doi.org/10.1016/j.learninstruc.2020.101393>.
- Stubbe, T. C., Schwippert, K., & Wendt, H. (2016). Soziale Disparitäten der Schülerleistungen in Mathematik und Naturwissenschaften. In H. Wendt, W. Bos, C. Selter, O. Köller, K. Schwippert & D. Kasper (Hrsg.), *TIMSS 2015. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 299–316). Münster: Waxmann.
- Vangrieken, K., Dochy, F., Raes, E., & Kyndt, E. (2015). Teacher collaboration: A systematic review. *Educational Research Review*, 15, 17–40.
- Vaughn, S., Elbaum, B. E., Schumm, J. S., & Hughes, M. T. (1998). Social outcomes for students with and without learning disabilities in inclusive classrooms. *Journal of Learning Disabilities*, 31(5), 428–436.
- Wiener, J., & Tardif, C. Y. (2004). Social and emotional functioning of children with learning disabilities: Does special education placement make a difference? *Learning Disabilities Research & Practice*, 19(1), 20–32.